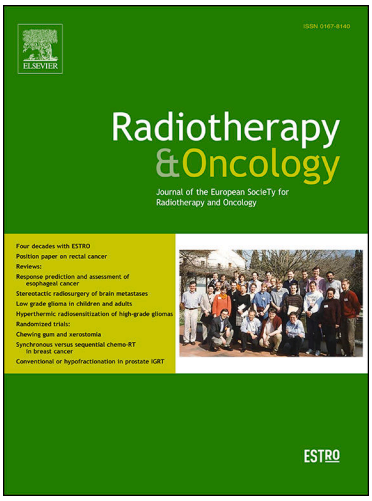
Journal Pre-proofs



Original Article

A deep learning MR-based radiomic nomogram may predict survival for nasopharyngeal carcinoma patients with stage T3N1M0

Lian-Zhen Zhong, Xue-Liang Fang, Di Dong, Hao Peng, Meng-Jie Fang, Cheng-Long Huang, Bing-Xi He, Li Lin, Jun Ma, Ling-Long Tang, Jie Tian

PII: S0167-8140(20)30387-X
DOI: https://doi.org/10.1016/j.radonc.2020.06.050
Reference: RADION 8414

To appear in: *Radiotherapy and Oncology*

Received Date: 9 January 2020
Revised Date: 1 June 2020
Accepted Date: 30 June 2020

Please cite this article as: Zhong, L-Z., Fang, X-L., Dong, D., Peng, H., Fang, M-J., Huang, C-L., He, B-X., Lin, L., Ma, J., Tang, L-L., Tian, J., A deep learning MR-based radiomic nomogram may predict survival for nasopharyngeal carcinoma patients with stage T3N1M0, *Radiotherapy and Oncology* (2020), doi: https://doi.org/10.1016/j.radonc.2020.06.050

A deep learning MR-based radiomic nomogram may predict survival for nasopharyngeal carcinoma patients with stage T3N1M0

Authors and Affiliations:

Lian-Zhen Zhong[†,*,§,] , MS; Xue-Liang Fang[‡,§], MD; Di Dong[†,*,§], PhD; Hao Peng[††], MD; Meng-Jie Fang[†,*], MS; Cheng-Long Huang[‡], MD; Bing-Xi He[*,¦] , MS; Li Lin[‡], MD; Jun Ma[‡,¶], MD; Ling-Long Tang[‡,¶], MD; Jie Tian[*,#,¶], PhD.

[†]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, 100049, P. R. China

[*]CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, P. R. China

[‡]State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Sun Yat-sen University Cancer Center, Guangzhou, 510060, P. R. China

[††]Center for Translational Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, 510080, P. R. China

[¦]School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, P.R. China

[#]Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine, Beihang University, Beijing, 100191, P. R. China

[§]Lian-Zhen Zhong, Xue-Liang Fang, Di Dong contributed equally to this article.

[¶]Corresponding authors:

Jie Tian, PhD

FAIMBE, FIAMBE, FIEEE, FSPIE, FOSA, FIAPR, FISMRM

Director of the CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, 100049, P. R. China.

Telephone: +86-10-82618465

Postal address: No. 95 Zhongguancun East Road, Hai Dian District, Beijing, 100190, P. R. China

Fax: +86-10-82618465

E-mail: jie.tian@ia.ac.cn

Jun Ma, MD

Vice President of Sun Yat-sen University Cancer Center, Guangzhou, 510060, P. R. China

Telephone: +86-20-87343469

Postal address: 651 Dongfeng Road East, Guangzhou 510060, P. R. China

Fax: +86-20-87343469

E-mail: majun2@mail.sysu.edu.cn

Ling-long Tang, MD

Sun Yat-sen University Cancer Center, Guangzhou, 510060, P. R. China

Telephone: +86-20-87343096

Postal address: 651 Dongfeng Road East, Guangzhou 510060, P. R. China

Fax: +86-20-87343096

E-mail: tangll@mail.sysu.edu.cn

Abstract

Purpose: To estimate the prognostic value of deep learning (DL) magnetic resonance (MR)-based radiomics for stage T3N1M0 nasopharyngeal carcinoma (NPC) patients receiving induction chemotherapy (ICT) prior to concurrent chemoradiotherapy (CCRT).

Methods: A total of 638 stage T3N1M0 NPC patients (training cohort: n = 447; test cohort: n = 191) were enrolled and underwent MRI scans before receiving ICT+CCRT. From the pretreatment MR images, DL-based radiomic signatures were developed to predict disease-free survival (DFS) in an end-to-end way. Incorporating independent clinical prognostic parameters and radiomic signatures, a radiomic nomogram was built through multivariable Cox proportional hazards method. The discriminative performance of the radiomic nomogram was assessed using the concordance index (C-index) and the Kaplan-Meier estimator.

Results: Three DL-based radiomic signatures were significantly correlated with DFS in the training (C-index: 0.695-0.731, all $p < 0.001$) and test (C-index: 0.706-0.755, all $p < 0.001$) cohorts. Integrating radiomic signatures with clinical factors significantly improved the predictive value compared to the clinical model in the training (C-index: 0.771 *vs.* 0.640, $p < 0.001$) and test (C-index: 0.788 *vs.* 0.625, $p = 0.001$) cohorts. Furthermore, risk stratification using the radiomic nomogram demonstrated that the high-risk group exhibited short-lived DFS compared to the low-risk group in the training cohort (hazard ratio [HR]: 6.12, $p < 0.001$), which was validated in the test cohort (HR: 6.90, $p < 0.001$).

Conclusions: Our DL-based radiomic nomogram may serve as a noninvasive and useful tool for pretreatment prognostic prediction and risk stratification in stage T3N1M0 NPC.

Introduction

Nasopharyngeal carcinoma (NPC) has an unbalanced geographical distribution, with the highest incidence reported in Southeast Asia [1]. There were 129,000 new cases of NPC in 2018 [2] and approximately 70% of them were locoregionally advanced at diagnosis [3]. For patients with locoregionally advanced NPC (LANPC), concurrent chemoradiotherapy (CCRT) with or without induction chemotherapy (ICT) is recommended by the National Comprehensive Cancer Network guidelines [4].

The TNM staging system is now widely used for risk stratification and treatment decision in NPC. But it is noteworthy that LANPC patients with same TNM stage and similar treatment still have large variations in clinical outcomes, with 30%~40% of them eventually developing distant metastasis [5]. Therefore, there is a need for development of individualized methods to predict prognosis of LANPC patients.

Great efforts have been made to search for prognostic biomarkers for LANPC, such as microRNAs [6], mRNAs [7] and PD-1/PD-L1 expression [8]. However，they require invasive biopsy and specialized equipment for molecular analysis, thus causing high cost and limiting its routine use [9, 10]. Obviously, a convenient and low-cost approach will be of great value.

Radiomics has recently emerged as a promising field in oncology, which assumes that medical imaging can characterize information of the primary tumor, such as genetics, pathology and prognosis [11, 12]. Via transferring medical images into high-dimension quantitative data, it can offer a noninvasive, effective and reliable method to aid clinical diagnosis, staging, treatment planning, and response assessment [13-15]. The rise of deep learning (DL), especially modern deep convolutional neural networks (DCNNs), enables radiomics to extract correlative quantitative representation of tumor phenotype in a fully automated and end-to-end way [16]. Past studies have suggested its potential medical application in a variety of cancers [17-19].

Based on above, we collected multi-parameter magnetic resonance (MR) images and clinical data of NPC before treatment to construct and validate a DL-based radiomic model for exploring whether radiomics could predict prognosis of stage T3N1M0 NPC

patients treated with ICT+CCRT, which is the largest subgroup (> 30%) of LANPC.

Methods and Materials

Patient data

From an NPC-specific big-data intelligence platform at the Sun Yat-sen University Cancer Center (SYSUCC), we reviewed and screened eligible patients diagnosed at our center between January 2010 and March 2016. A detailed description of this database is presented in Appendix A.

The inclusion criteria were as follows: (i) patients who had diagnosed T3N1M0 disease; (ii) treated by ICT+CCRT; (iii) underwent pretreatment MRI scans; (iv) had complete clinical data, including age, sex, history of drinking and smoking, follow-up and pretreatment hematological examination results; (v) did not have other malignancies; and (vi) received intensity-modulated radiotherapy. The exclusion criteria included: (i) history of anticancer treatment before baseline MRI scans, such as radiotherapy, chemotherapy, immunotherapy and surgery, etc.; (ii) had artifacts, blurs, faults, and disordered slices in the MR image. This retrospective study of anonymous data was approved by our institution's ethical review board and the requirement for informed consent was waived.

All clinical data were gathered from the records of the institution's Picture Archiving and Communication System (PACS, Carestream). Epstein-Barr Virus DNA (EBV-DNA) concentration was measured by real-time quantitative polymerase chain reaction assay [20]. The AJCC-TNM Staging System Manual (8th Edition, 2017) was used to perform tumor staging.

Treatment, follow-up and clinical endpoint

All patients received ICT+CCRT. Intensity-modulated radiotherapy lasted for 6 to 7 weeks with 5 daily fractions per week. The prescribed radiation doses were 66 to 72 Gy (2.12 to 2.27 Gy per fraction) for the primary tumor and 64 to 70 Gy (30 to 33 Gy per fraction) for the involved lymph nodes. ICT consisted of every 3-week cisplatin-based regimens for 2 to 4 cycles while CCRT consisted of weekly cisplatin 40 mg/m^2 for 5~7 cycles or triweekly cisplatin 80 mg/m^2 for 2~3 cycles during radiotherapy.

Details of the treatment have been reported before [21]. In the first 2 years of follow-up, patients were examined by routine imaging methods every 3 months, every 6 months from years 3-5, and annually thereafter. Follow-up duration was time from initial diagnosis to last visit or death. All local and regional recurrences were eventually confirmed by pathology. Distant metastasis was mainly diagnosed by the means of imaging examinations such as MRI or PET/CT. Disease-free survival (DFS) was set as the primary endpoint for earlier individual treatment while locoregional relapse-free survival, distant metastasis-free survival, and overall survival set as the secondary endpoints. The definitions of all clinical endpoints are consistent with Peng et al [19].

MR imaging protocol and tumor segmentation

Multi-parametric MRI scans were performed on each patient within 2 weeks before any anti-tumor therapy, which included sequential non-contrast enhanced T1-weighted (axial, coronal, and sagittal planes) and T2-weighted (axial plane) images, and contrast enhanced T1-weighted images (axial, coronal, and sagittal planes). The MR protocols (scanner version, magnetic field strength, etc.) are depicted at length in Appendix B. All MR images extracted from the PACS were loaded into 3.8.0 version of ITK-SNAP software and then manually segmented. The primary tumor on the basilar region was contoured as the region of interest (ROI) on each slice of the three axial MR images. Coronal and sagittal MR images were only used in the need of guiding segmentation of the ROIs in cross-sectional plane. Therefore, every patient in this study had three ROIs. A junior radiologist (H.P.) with 6 years of clinical MR image-reading experience performed all image segmentations and resolved any uncertainty through consultation with another senior radiologist (L.L.T.).

Deep learning-based radiomic signature building

Conventional handcrafted features strongly rely on precise ROIs with good intra/inter-reader agreement, which is difficult to be satisfied and limits its application in routine clinical practice [11]. Therefore, we adopted a DCNN (SE-ResNeXt [22]) to end-to-end quantify the tumor phenotype for predicting DFS using multiple instance learning [23], which only required the approximate central location of ROI. SE-

ResNeXt uses ResNeXt [24] as the backbone architecture and can adaptively retune channel-wise activation responses by explicitly modelling interrelation between channels, which have successfully improved significantly performance of existing state-of-the-art convolutional neural networks. In order to train SE-ResNeXt end-to-end, Cox partial log likelihood [25] was used as the loss function to guide weight update. Before training the survival model, it was pretrained on a 3-year DFS classification task to obtain good initial weights. In addition, augmentation techniques were employed to prevent overfitting. Before the model construction, MR images were interpolated and normalized to compensate for scanner-dependent variability in image intensity [26]. Detailed image processing, multiple instance learning, and the architecture and implementation of SE-ResNeXt are described in the Supplementary Methods (see Appendix C). All operations, network architecture, and parameters were the same for each sequence of MR images. Therefore, we built three DL-based radiomic signatures (the model outputs that predicted hazard of DFS) to reflect the phenotypic characteristics of the primary tumor from three sequences of MR images: the DL_T1_sig, DL_T2_sig and DL_T1C_sig.

Prognostic verification of DL-based radiomic signatures

The predictive performance of each radiomic signature for DFS was firstly assessed in the training cohort and then verified in the test cohort. To compare prognostic performance between DL-based radiomic signatures and routine clinical parameters, a clinical model ($Model^{clinic}$) was built based on independent clinical prognostic factors using multivariable Cox proportional hazards (CPH) method.

Creation and performance of an individualized radiomic nomogram

The association between the clinical risk parameters and DFS of NPCs was firstly analyzed using univariate CPH method in the training cohort. Then, a multivariable CPH model ($Model^{clinic+dl}$) was used to identify independent predictors among clinical factors and DL-based radiomic signatures after adjustment for various covariates. Finally, $Model^{clinic+dl}$ was visualized into a radiomic nomogram for helping clinicians to get conveniently individualized DFS estimation. The quality of $Model^{clinic+dl}$ was

measured quantitatively using hazard ratio (HR) and Harrell's concordance index (C-index). In addition, risk stratification using the radiomic nomogram was conducted, where a specific threshold was identified in the training cohort and locked during the test phase.

Verification of the radiomic nomogram

The agreement between the observed actual DFS rate and the nomogram-predicted DFS rate was evaluated using calibration curves. Time-dependent receiver operating characteristic (TD-ROC) analysis [27], which could accommodate censored data, was adopted to assess the predictive power of $Model^{clinic+dl}$, and to compare $Model^{clinic+dl}$ with DL-based radiomic signatures and $Model^{clinic}$.

Statistical analysis

Kaplan-Meier curves were drawn for DFS, where a month is defined as 30 days. All HRs and C-indices were reported with 95% confidence intervals (CIs) and corresponding p-values. The p-values of HRs were tested using the log-rank test; the p-value of a C-index was tested using the Z-test and measures the significance of the difference between the value of C-index and 0.5; the p-value for comparing two different C-indices were tested using the paired Student t-test based on the assumption of normality for the natural logarithm of the concordance index [28]. In the multivariable CPH method, backward stepwise selection was applied with Akaike's information as the optimization criterion. Note that, all models except radiomic signatures were established using the multivariable CPH method.

SE-ResNeXt was implemented with 3.6.5 version of Python based on the open-source deep learning library Tensorflow (version: 1.10.0; https://www.tensorflow.org). Image processing algorithms were conducted in MATLAB R2017a (MathWorks, Natick, MA, USA) while statistical analysis with 3.6.1 version of R software. The open-source PyRadiomics package (https://pypi.org/project/pyradiomics/) was used to extract handcrafted radiomic features from the segmented tumor volume for a comparison with DL-based radiomic signatures (Appendix C). A two-sided p-value < 0.05 indicated that a comparison reaches a statistically significant difference. A detailed

description of statistical methods was shown in the Supplementary Methods (Appendix C). Figure 1 shows the radiomic workflow.

Results

In total, this study enrolled 638 patients (442 males and 196 females; mean [± SD] age, 41.65 ± 10.21 years; range, 10-69 years). The last follow-up was on October 19, 2019 with a median follow-up of 5.46 years (interquartile range: 4.52-6.52). All patients were divided into the training (n = 447) and test (n = 191) cohorts by computer-generated random numbers (Figure D1). No significant differences were found between groups with respect to any clinical factor (p = 0.18-0.99) in Table 1 summarizing and comparing the participant characteristics in the training and test cohorts. Upon the last follow-up, 81 (18.1%) in the training cohort and 32 (16.8%) patients in the test cohort experienced a confirmed disease progression or death (p = 0.76).

There was a significant association between each DL-based radiomic signature and DFS in the training cohort (C-index: 0.695-0.731, all $p < 0.001$), which was subsequently validated in the test cohort (C-index: 0.706-0.755, all $p < 0.001$). Moreover, we performed a 3-fold cross-validation in each MR sequence of the training cohort using the same network architecture and training strategies. In this experiment, no significant difference was found between the resulted C-indexes, which ranged from 0.690 to 0.746 in the holdout validation sets. Univariate analysis explored clinical factors in Table 1, and found that pretreatment plasma EBV-DNA (pre-EBV DNA), serum level of lactate dehydrogenase (LDH) and C-reaction protein were significantly associated with DFS (Table D1). $\text{Model}^{\text{clinic}}$ started with all clinical parameters and finally identified sex, age, pre-EBV DNA, C-reaction protein and LDH as independent variables. $\text{Model}^{\text{clinic}}$ obtained worse results than any of DL-based radiomic signatures in the training (C-index: 0.640, 95% CI: 0.577-0.703, $p < 0.001$) and test (C-index: 0.625, 95% CI: 0.521-0.729, p = 0.018) cohorts.

When a multivariate CPH model was performed, $\text{Model}^{\text{clinic+dl}}$ screened out three DL-based radiomic signatures, age, LDH and pre-EBV DNA after adjusting covariates (Table D2). Then, a radiomic nomogram for individualized DFS estimation was built

using the regression coefficients of $Model^{clinic+dl}$ (Figure 2A). In the training cohort, $Model^{clinic+dl}$ achieved the strongest prognostic ability for DFS (C-index: 0.771, 95% CI: 0.715-0.827, $p < 0.001$) compared to DL-based radiomic signatures and $Model^{clinic}$. The same trend was found in the test cohort (C-index: 0.788, 95% CI: 0.695-0.882, $p < 0.001$). Good agreement in the calibration curves was observed between the nomogram-estimated DFS rate and the observed DFS rate at 3 and 5 years (Figure 2B and 2C).

We identified the threshold score of the radiomic nomogram as 0.398 corresponding to the total point of 180 in the training cohort and divided patients into low- and high-risk groups (Figure 2A). Consequently, the high-risk group exhibited short-lived DFS compared to the low-risk group (HR: 6.12, 95% CI: 3.72-10.07, $p < 0.001$; 5-year DFS rate: 63.1% *vs.* 93.0%; Figure 2D) in the training cohort, which was verified in the test cohort (HR: 6.90, 95% CI: 3.10-15.36, $p < 0.001$; 5-year DFS rate: 66.9% *vs.* 94.4%; Figure 2E). Similarly, patients in the low-risk group also achieved better locoregional relapse-free survival, distant metastasis-free survival, and overall survival (all $p < 0.001$; Figure 3), demonstrating the good clinical usefulness of our model.

When stratified by family history of cancer (yes or no), sex (male or female), C-reactive protein (normal or abnormal), hemoglobin (normal or abnormal) and smoking (yes or no), the radiomic nomogram remained a statistically significant prognostic model (Figure D2). Furthermore, TD-ROC analysis also validated the superior performance of $Model^{clinic+dl}$ for predicting DFS (Figure 4). The performance of the radiomic nomogram and other models are shown in detail in Table 2.

Discussion

We developed and validated the prognostic value of multiparametric MR-based radiomics for pretreatment individualized evaluation of DFS in stage T3N1M0 NPC patients receiving ICT + CCRT, using an end-to-end deep learning method on this relatively large-scale cohort. The DL-based radiomic signatures from single sequence of MR images demonstrated significant prognostic value for estimating DFS. Integrating radiomic signatures with independent clinical factors adequately improved

predictive value compared to the routinely used clinical model and successfully stratified patients into the low- and high-risk groups, indicating that it could serve as a noninvasive and useful tool for predicting prognosis of stage T3N1M0 NPC patients. We have made our prediction models public on an open-access website (http://www.radiomics.net.cn/post/126) to facilitate its integration into the routine work-up.

In this study, a state-of-the-art DCNN was used to develop the radiomic signature from single sequence of MR images, with Cox partial log likelihood as a loss function under the multiple instance learning assumption. The handcrafted radiomic features require precise ROIs with good intra/inter-reader agreement and showed unsatisfactory results that could not be generalized in our dataset (Appendix C). In contrast, DCNNs take tumor-centered boxes as input without need of precise object segmentation and have the advantages of end-to-end training strategy and powerful feature representation ability [16, 29], which led to our preference for it. The DL-based radiomic signatures were highly associated with DFS in the training and test cohorts (all $p < 0.001$). Impressively, the radiomic nomogram with integrated clinical risk parameters and radiomic signatures revealed a significant improvement for predicting DFS compared to clinical model (training cohort: C-index 0.771 *vs.* 0.640, $p < 0.001$; test cohort: C-index 0.788 *vs.* 0.625, $p = 0.001$). Moreover, the radiomic nomogram presented good clinical usefulness and stability in prediction of secondary endpoints and stratified analysis. Attention map [30] of the radiomic signatures showed that our DCNN focused on different regions of the MR images between the high-risk and low-risk patients (Appendix D and Figure D3).

Since older patients have difficulty tolerating intense radiotherapy and high-intensity chemotherapy, and are more prone to severe comorbidities, the increase in age is associated with poor survival in NPC patients [4]. For NPC patients with liver micro metastasis, hepatocellular membrane injury and enzyme leakage can result in significant elevated baseline LDH levels. Conversely, high serum LDH levels may indicate subsequent liver metastasis [31]. Pre-EBV DNA load correlates with tumor

burden and has become an acknowledged prognostic biomarker [19, 32]. Our study found that they were indeed valuable in predicting the survival of stage T3N1M0 NPC patients and were therefore included in our radiomic nomogram.

Since stage T3N1M0 NPC is the largest subgroup that makes up over 30% of LANPC [33], we conducted the study on patients with stage T3N1M0. To date, LANPC has multiple treatment patterns, including CCRT, ICT+CCRT, and CCRT plus adjuvant chemotherapy [34, 35]. Recent studies indicated that ICT+CCRT could significantly improve survival compared with CCRT alone in LANPC [36-38]. Thus, we selected patients who had received ICT+CCRT to reduce treatment-related bias. Compared to routine molecular biomarkers, our radiomic nomogram is noninvasive, accessible, and financially affordable, which could stratify patients at the same stage into two subgroups according to predicted survival. For patients in the low-risk group, the standard treatment had a good 5-year DFS rate of 93.5% in the entire cohort. For those in the high-risk group, the standard treatment did not bring about expected benefits (only 64.1% in 5-year DFS rate), but superfluous drug toxicities and economic burden. Therefore, the standard treatment was recommended for patients stratified into the low-risk group while novel regimens, such as immunotherapy [39], may be an option for patients stratified into the high-risk group. To our knowledge, this is the first study exploring the feasibility of DL-based radiomics in predicting prognosis of stage T3N1M0 NPC patients.

Nevertheless, our study had some limitations. Given that tumors of basilar regions are usually inaccessible for biopsy, but are aggressive and can result in skull bone invasion, which is associated with poor survival [40], we chose slices of basilar region for tumor segmentation and radiomic analysis. Nasopharyngeal and other regions ought to be considered in our future study. For patient entry, we focused only on stage T3N1M0 patients, so extended research involving other TNM stages should be initiated in the future. Besides, all patients were from SYSUCC, which calls for external multicenter validations. Moreover, there was possible selection bias owing to the retrospective nature of this study. Well-designed prospective studies will be necessary.

In conclusion, the current study identified the value of deep learning on MR-based radiomics for prognostic prediction in stage T3N1M0 NPC patients treated with ICT+CCRT. Our pretreatment radiomic nomogram presented superior discrimination for several clinical endpoints and successfully stratified individual patients into two groups with distinguishable prognosis, which may act as a noninvasive and useful tool for prognostic prediction and risk stratification.

Funding Source:

This work was supported by the National Key R&D Program of China (2017YFC1308700, 2017YFA0205200, 2017YFC1309100, 2017YFA0700401), National Natural Science Foundation of China (91959130, 81971776, 81930072, 81771924, 81930053, 81227901, 81671851, 81527805), Key-Area Research and Development Program of Guangdong Province (2019B020230002), the Natural Science Foundation of Guangdong Province (2017A030312003), the Beijing Natural Science Foundation (L182061), Strategic Priority CAS Project (XDB38040200), Health & Medical Collaborative Innovation Project of Guangzhou City, China (201803040003), Innovation Team Development Plan of the Ministry of Education (IRT_17R110), Overseas Expertise Introduction Project for Discipline Innovation (111 Project, B14035), and the Youth Innovation Promotion Association CAS (2017175).

The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgments: None.

References

[1] Tang LL, Chen W-Q, Xue W-Q, He Y-Q, Zheng R-S, Zeng Y-X, et al. Global trends in incidence and mortality of nasopharyngeal carcinoma. Cancer Letters. 2016, 374(1): 22-30.

[2] 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: a cancer journal for clinicians. 2018;68:394-424.

[3] Pan JJ, Ng WT, Zong JF, Lee SW, Choi HC, Chan LL, et al. Prognostic nomogram for refining the prognostication of the proposed 8th edition of the AJCC/UICC staging system for nasopharyngeal cancer in the era of intensity-modulated radiotherapy. Cancer. 2016;122:3307-15.

[4] National Comprehensive Cancer Network guidelines. 2018.

[5] Hui EP, Leung SF, Au JS, Zee B, Tung S, Chua D, et al. Lung metastasis alone in nasopharyngeal carcinoma: a relatively favorable prognostic group: a study by the Hong Kong Nasopharyngeal Carcinoma Study Group. Cancer. 2004;101:300-6.

[6] Liu N, Chen N-Y, Cui R-X, Li W-F, Li Y, Wei R-R, et al. Prognostic value of a microRNA signature in nasopharyngeal carcinoma: a microRNA expression analysis. The lancet oncology. 2012;13:633-41.

[7] Tang X-R, Li Y-Q, Liang S-B, Jiang W, Liu F, Ge W-X, et al. Development and validation of a gene expression-based signature to predict distant metastasis in locoregionally advanced nasopharyngeal carcinoma: a retrospective, multicentre, cohort study. The Lancet Oncology. 2018;19:382-93.

[8] Ono T, Azuma K, Kawahara A, Sasada T, Matsuo N, Kakuma T, et al. Prognostic stratification of patients with nasopharyngeal carcinoma based on tumor immune microenvironment. Head & neck. 2018;40:2007-19.

[9] Segal E, Sirlin CB, Ooi C, Adler AS, Gollub J, Chen X, et al. Decoding global gene expression programs in liver cancer by noninvasive imaging. Nature biotechnology. 2007;25:675-80.

[10] van Nimwegen KJ, van Soest RA, Veltman JA, Nelen MR, van der Wilt GJ, Vissers

LE, et al. Is the $1000 genome as near as we think? A cost analysis of next-generation sequencing. Clinical chemistry. 2016;62:1458-64.

[11] Lambin P, Leijenaar RT, Deist TM, Peerlings J, De Jong EE, Van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nature reviews Clinical oncology. 2017;14:749.

[12] Zhang B, Tian J, Dong D, Gu D, Dong Y, Zhang L, et al. Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. Clinical Cancer Research. 2017;23:4259-69.

[13] Aerts HJ. The potential of radiomic-based phenotyping in precision medicine: a review. JAMA oncology. 2016;2:1636-42.

[14] Limkin E, Sun R, Dercle L, Zacharaki E, Robert C, Reuzé S, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. Annals of Oncology. 2017;28:1191-206.

[15] Dong D, Zhang F, Zhong L-Z, Fang M-J, Huang C-L, Yao J-J, et al. Development and validation of a novel MR imaging predictor of response to induction chemotherapy in locoregionally advanced nasopharyngeal cancer: a randomized controlled trial substudy (NCT01245959). BMC medicine. 2019;17:190.

[16] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems2012. p. 1097-105.

[17] Dong D, Fang MJ, Tang L, Shan XH, Gao JB, F. Giganti, et al. Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study, Annals of Oncology, 2020, DOI: 10.1016/j.annonc.2020.04.003.

[18] Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nature medicine. 2019;25:954.

[19] Peng H, Dong D, Fang M, Li L, Tang L-L, Chen L, et al. Prognostic Value of Deep Learning PET/CT-based Radiomics: Potential Role for Future Individual Induction

Chemotherapy in Advanced Nasopharyngeal Carcinoma. Clinical Cancer Research. 2019; 3065.2018.

[20] Huang C-L, Sun Z-Q, Guo R, Liu X, Mao Y-P, Peng H, et al. Plasma Epstein-Barr virus DNA load after induction chemotherapy predicts outcome in locoregionally advanced nasopharyngeal carcinoma. International Journal of Radiation Oncology* Biology* Physics. 2019;104:355-61.

[21] Lv JW, Qi ZY, Zhou GQ, He XJ, Chen YP, Mao YP, et al. Optimal cumulative cisplatin dose in nasopharyngeal carcinoma patients receiving additional induction chemotherapy. Cancer science. 2018;109:751-63.

[22] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. p. 7132-41.

[23] Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature Medicine. 2019;25:1301-9.

[24] Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 1492-500.

[25] Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC medical research methodology. 2018;18:24.

[26] Sun X, Shi L, Luo Y, Yang W, Li H, Liang P, et al. Histogram-based normalization technique on human brain magnetic resonance images from different acquisitions. Biomedical engineering online. 2015;14:73.

[27] Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics. 2000;56:337-44.

[28] Haibe-Kains B, Desmedt C, Sotiriou C, Bontempi G. A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all? Bioinformatics. 2008;24:2200-8.

[29] Dong D, Tang L, Li Z-Y, Fang M-J, Gao J-B, Shan X-H, et al. Development and

validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. Annals of Oncology. 2019;30:431-8.

[30] Selvaraju RR, Das A, Vedantam R, Cogswell M, Parikh D, Batra D. Grad-CAM: Why did you say that? arXiv preprint arXiv:1611.07450, 2016.

[31] Zhou G-Q, Tang L-L, Mao Y-P, Chen L, Li W-F, Sun Y, et al. Baseline serum lactate dehydrogenase levels for patients treated with intensity-modulated radiotherapy for nasopharyngeal carcinoma: a predictor of poor prognosis and subsequent liver metastasis. International Journal of Radiation Oncology* Biology* Physics. 2012;82:e359-e65.

[32] Lin J-C, Wang W-Y, Chen KY, Wei Y-H, Liang W-M, Jan J-S, et al. Quantification of plasma Epstein–Barr virus DNA in patients with advanced nasopharyngeal carcinoma. New England Journal of Medicine. 2004;350:2461-70.

[33] Tang L, Xu C, Zhang S, Li W-F, Chen L, Mao Y, et al. Selection and validation of induction chemotherapy beneficiaries among patients with T3N0, T3N1, T4N0 nasopharyngeal carcinoma using Epstein-Barr virus DNA: a joint analysis of real-world and clinical trial data. Frontiers in Oncology. 2019;9:1343.

[34] Cao S-M, Yang Q, Guo L, Mai H-Q, Mo H-Y, Cao K-J, et al. Neoadjuvant chemotherapy followed by concurrent chemoradiotherapy versus concurrent chemoradiotherapy alone in locoregionally advanced nasopharyngeal carcinoma: a phase III multicentre randomised controlled trial. European journal of cancer. 2017;75:14-23.

[35] Hui EP, Ma BB, Leung SF, King AD, Mo F, Kam MK, et al. Randomized phase II trial of concurrent cisplatin-radiotherapy with or without neoadjuvant docetaxel and cisplatin in advanced nasopharyngeal carcinoma. Journal of clinical oncology. 2009;27:242-9.

[36] Sun Y, Li W-F, Chen N-Y, Zhang N, Hu G-Q, Xie F-Y, et al. Induction chemotherapy plus concurrent chemoradiotherapy versus concurrent chemoradiotherapy alone in locoregionally advanced nasopharyngeal carcinoma: a phase 3, multicentre, randomised controlled trial. The lancet oncology. 2016;17:1509-

20.

[37] Zhang Y, Chen L, Hu G-Q, Zhang N, Zhu X-D, Yang K-Y, et al. Gemcitabine and Cisplatin Induction Chemotherapy in Nasopharyngeal Carcinoma. The New England journal of medicine. 2019;381:1124.

[38] Chen Y-P, Tang L-L, Yang Q, Poh S-S, Hui EP, Chan AT, et al. Induction chemotherapy plus concurrent chemoradiotherapy in endemic nasopharyngeal carcinoma: individual patient data pooled analysis of four randomized trials. Clinical Cancer Research. 2018;24:1824-33.

[39] Fang W, Yang Y, Ma Y, Hong S, Lin L, He X, et al. Camrelizumab (SHR-1210) alone or in combination with gemcitabine plus cisplatin for nasopharyngeal carcinoma: results from two single-arm, phase 1 trials. The Lancet Oncology. 2018;19:1338-50.

[40] Lu TX, Mai WY, Teh BS, Yong-Hong Hu, Hsin H Lu, J.KamChiu, et al. Important prognostic factors in patients with skull base erosion from nasopharyngeal carcinoma after radiotherapy. Int J Radiat Oncol Biol Phys. 2001;51(3):589-598.

Figure Captions

Figure 1. Radiomics workflow in this study.

Abbreviations: T1w, T1-weighted; T2w, T2-weighted; CE-T1w, contrast enhanced T1-weighted; MR, magnetic resonance; DL, deep learning.

Figure 2. The radiomic nomogram, calibration curves and risk stratification in the training and test cohorts. The radiomic nomogram built for predicting DFS in this study (A). The calibration curves of the radiomic nomogram in the training (B) and test (C) cohorts. Risk stratification using the radiomic nomogram in the training (D) and test (E) cohorts. Calibration curves were assessed using the Hosmer-Lemeshow test. The radiomic nomogram predicts the probability that a patient will not experience disease progression within 3 and 5 years. For example, a 50-year-old NPC patient with stage T3N1M0 had a primary tumor on the nasopharynx, a LDH of 500 U/L, and a pre-EBV DNA of <4000 copy/mL; its radiomic signature scores were 0.4 (DL_T1_sig), -0.4 (DL_T2_sig), and 0 (DL_T1C_sig) separately. Therefore, the total point of the patient was about 212 (40+60+0+60+22+30), and the corresponding 3-year DFS rate and 5-year DFS rate were about 78% and 70%.

Abbreviations: LDH, lactate dehydrogenase; pre-EBV DNA, pretreatment plasma Epstein–Barr virus DNA; DFS, disease-free survival; HR, hazard ratio; CI, confidence interval.

Figure 3. Overall survival (A and D), distant metastasis-free survival (B and E) and locoregional relapse-free survival (C and F) stratified by the radiomic nomogram in the training and test cohort.

Abbreviations: HR, hazard ratio; CI, confidence interval.

Figure 4. Comparison of survival models using TD-ROC curves. $Model^{clinic}$ was built based on sex, age, pre-EBV DNA, LDH and CRP while $Model^{clinic+dl}$ based on three radiomic signatures, age, LDH and pre-EBV DNA.

Abbreviations: TD-ROC, time-dependent receiver operating characteristic; AUC, area under receiver operating characteristic curve; pre-EBV DNA: pretreatment plasma Epstein–Barr virus DNA; LDH: lactate dehydrogenase; CRP: C-reaction protein.

A deep learning MR-based radiomic nomogram may predict survival for nasopharyngeal carcinoma patients with stage T3N1M0

Authors and Affiliations:
Lian-Zhen Zhong[†,*,§,] , MS; Xue-Liang Fang[‡,§], MD; Di Dong[†,*,§], PhD; Hao Peng[††], MD; Meng-Jie Fang[†,*], MS; Cheng-Long Huang[‡], MD; Bing-Xi He[*,¦] , MS; Li Lin[‡], MD; Jun Ma[‡,¶], MD; Ling-Long Tang[‡,¶], MD; Jie Tian[*,#,¶], PhD.
[†]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, 100049, P. R. China
[*]CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, P. R. China
[‡]State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Sun Yat-sen University Cancer Center, Guangzhou, 510060, P. R. China
[††]Center for Translational Medicine, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, 510080, P. R. China
[¦]School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, P.R. China
[#]Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine, Beihang University, Beijing, 100191, P. R. China
[§]Lian-Zhen Zhong, Xue-Liang Fang, Di Dong contributed equally to this article.
[¶]Corresponding authors:
Jie Tian, PhD
FAIMBE, FIAMBE, FIEEE, FSPIE, FOSA, FIAPR, FISMRM
Director of the CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences.
Telephone: +86-10-82618465
Postal address: No. 95 Zhongguancun East Road, Hai Dian District, Beijing, 100190, P. R. China
E-mail: jie.tian@ia.ac.cn

Jun Ma, MD
Vice President of Sun Yat-sen University Cancer Center
Telephone: +86-20-87343469
Postal address: 651 Dongfeng Road East, Guangzhou 510060, P. R. China
E-mail: majun2@mail.sysu.edu.cn

Ling-long Tang, MD
Telephone: +86-20-87343096
Postal address: 651 Dongfeng Road East, Guangzhou 510060, P. R. China
E-mail: tangll@mail.sysu.edu.cn

Conflict of Interest:
The authors of this study have no conflicts of interest to disclose.

National Key Research and Development Program of China
Funding Source:
This work was supported by the National Key R&D Program of China (2017YFC1308700, 2017YFA0205200, 2017YFC1309100, 2017YFA0700401),

National Natural Science Foundation of China (91959130, 81971776, 81930072, 81771924, 81930053, 81227901, 81671851, 81527805), Key-Area Research and Development Program of Guangdong Province (2019B020230002), the Natural Science Foundation of Guangdong Province (2017A030312003), the Beijing Natural Science Foundation (L182061), Strategic Priority CAS Project (XDB38040200), Health & Medical Collaborative Innovation Project of Guangzhou City, China (201803040003), Innovation Team Development Plan of the Ministry of Education (IRT_17R110), Overseas Expertise Introduction Project for Discipline Innovation (111 Project, B14035), and the Youth Innovation Promotion Association CAS (2017175).
The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.
Acknowledgments: None.

Table 1. Baseline characteristics in the training and test cohorts

Characteristics	Primary Cohort (N = 447)	Validation cohort (N = 191)	*p*-value†
Age, median (range), years	41 (10-69)	41 (16-68)	0.78
Sex, No. (%)			
Male	308 (68.9)	134 (70.2)	0.83
Female	139 (31.1)	57 (29.8)	
HGB*, median (range), g/L	141.0 (67.0-174.0)	143.0 (91.0-179.0)	0.18
Normal	388 (86.8)	167 (87.4)	0.93
Abnormal	59 (13.2)	24 (12.6)	
Family history of cancer, No. (%)			
Yes	133 (29.8)	52 (27.2)	0.58
No	314 (70.2)	139 (72.8)	
Smoking, No. (%)			
Yes	129 (28.9)	62 (32.5)	0.42
No	318 (71.1)	129 (67.5)	
Drinking, No. (%)			
Yes	40 (8.9)	19 (9.9)	0.80
No	407 (91.1)	172 (90.1)	
WHO pathology type, No. (%)			
I-II	9 (2.0)	7 (3.7)	0.34
III	438 (98.0)	184 (96.3)	
ALB, median (range), g/L	44.5 (33.0-129.0)	44.8 (36.1-52.2)	0.99
Normal (≥ 40), No. (%)	410 (91.7)	176 (92.1)	0.98
Abnormal (< 40), No. (%)	37 (8.3)	15 (7.9)	
CRP, median (range), mg/L	1.76 (0.14-246.58)	1.79 (0.03-81.17)	0.71
Normal (≤ 3), No. (%)	313 (70.0)	131 (68.6)	0.79
Abnormal (> 3), No. (%)	134 (30.0)	60 (31.4)	
LDH, median (range), U/L	176.0 (76.6-753.0)	173.6 (109.0-554.0)	0.26
Normal (≤ 250), No. (%)	415 (92.8)	182 (95.3)	0.33
Abnormal (> 250), No. (%)	32 (7.2)	9 (4.7)	
pre-EBV DNA, median (range), copy/mL	3140 (0-1,070,000)	2560 (0-9,240,000)	0.22
< 4000, No. (%)	234 (52.3)	105 (55.0)	0.60
≥ 4000, No. (%)	213 (47.7)	86 (45.0)	

*For male, HGB < 130 is defined abnormal; for female, HGB < 115 is defined abnormal.

†Continuous variables were tested by either the t-test (for those with normal distribution) or the Wilcoxon rank sum test while categorical variables by either the Pearson's $\chi 2$ test or the Fisher's exact test.

Note that, the Shapiro-Wilk test was used to decide if variables followed normal distribution. For the categorical variables, the default was the Pearson's $\chi 2$ test, which was not recommended when some of the cell has small counts (<5).

Abbreviations: WHO, World Health Organization; LDH, lactate dehydrogenase; HGB, hemoglobin; ALB, albumin; CRP, C-reaction protein; pre-EBV DNA, pretreatment plasma Epstein–Barr Virus DNA.

Table 2. Performance of deep learning-based radiomic signatures and models for predicting DFS

Signature or model	Training cohort (n = 447)			Test cohort (n = 191)		
	C-index	95% CI	p-value*	C-index	95% CI	p-value*
DL_T1_sig	0.731	0.675-0.786	< 0.001	0.712	0.612-0.812	< 0.001
DL_T2_sig	0.695	0.632-0.758	< 0.001	0.706	0.609-0.802	< 0.001
DL_T1C_sig	0.701	0.647-0.755	< 0.001	0.755	0.672-0.838	< 0.001
$Model^{clinic}$	0.640	0.577-0.703	< 0.001	0.625	0.521-0.729	0.018
$Model^{clinic+dl}$	**0.771**	0.715-0.827	< 0.001	**0.788**	0.695-0.882	< 0.001

*P-value is used to measure the significance of the difference between the value of a C-index and 0.5 through the Z-test.

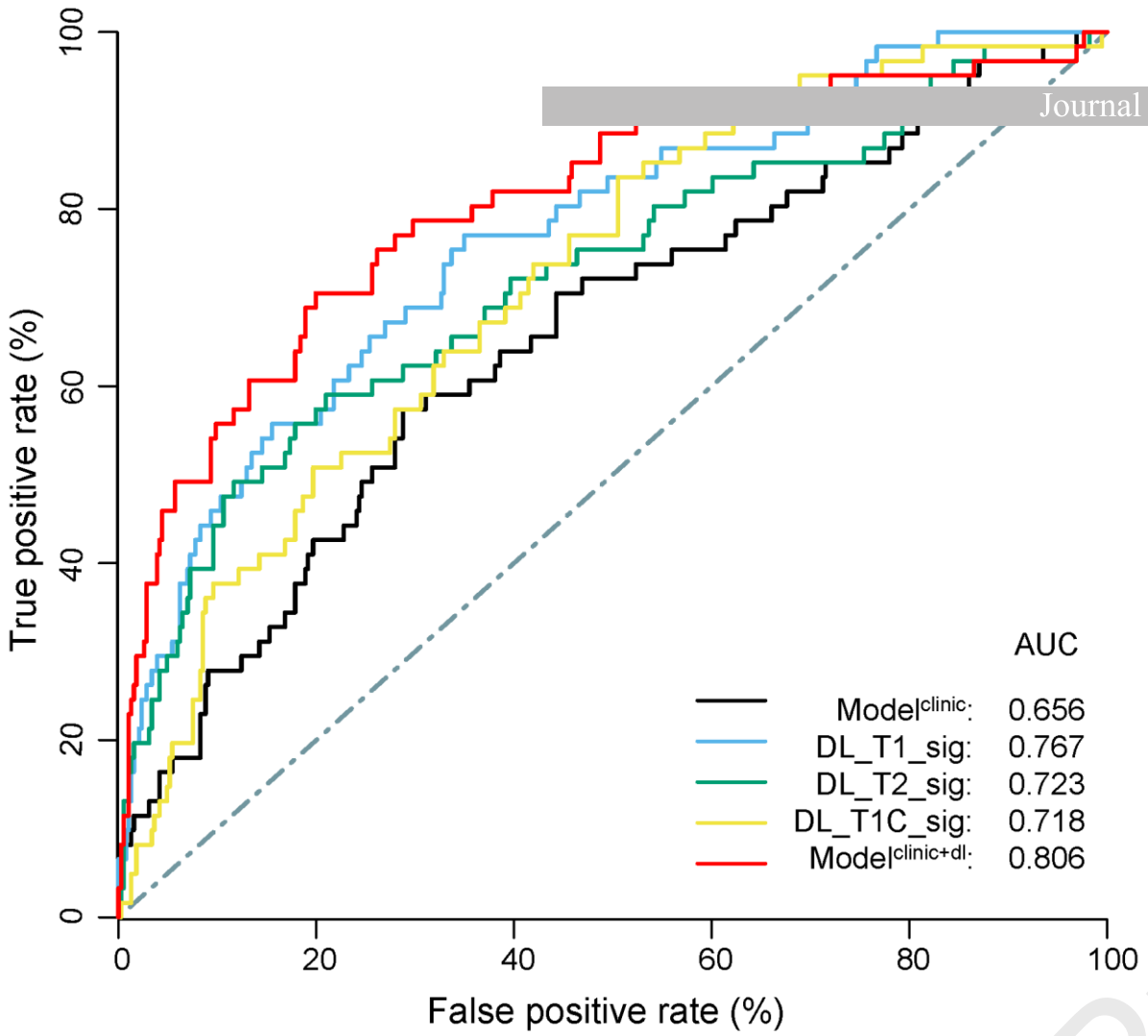
Note that, $Model^{clinic}$ was built based on sex, age, pre-EBV DNA, LDH and CRP while $Model^{clinic+dl}$ based on three radiomic signatures, age, LDH and pre-EBV DNA.

Abbreviations: C-index, Harrell's concordance index; CI, confidence interval.

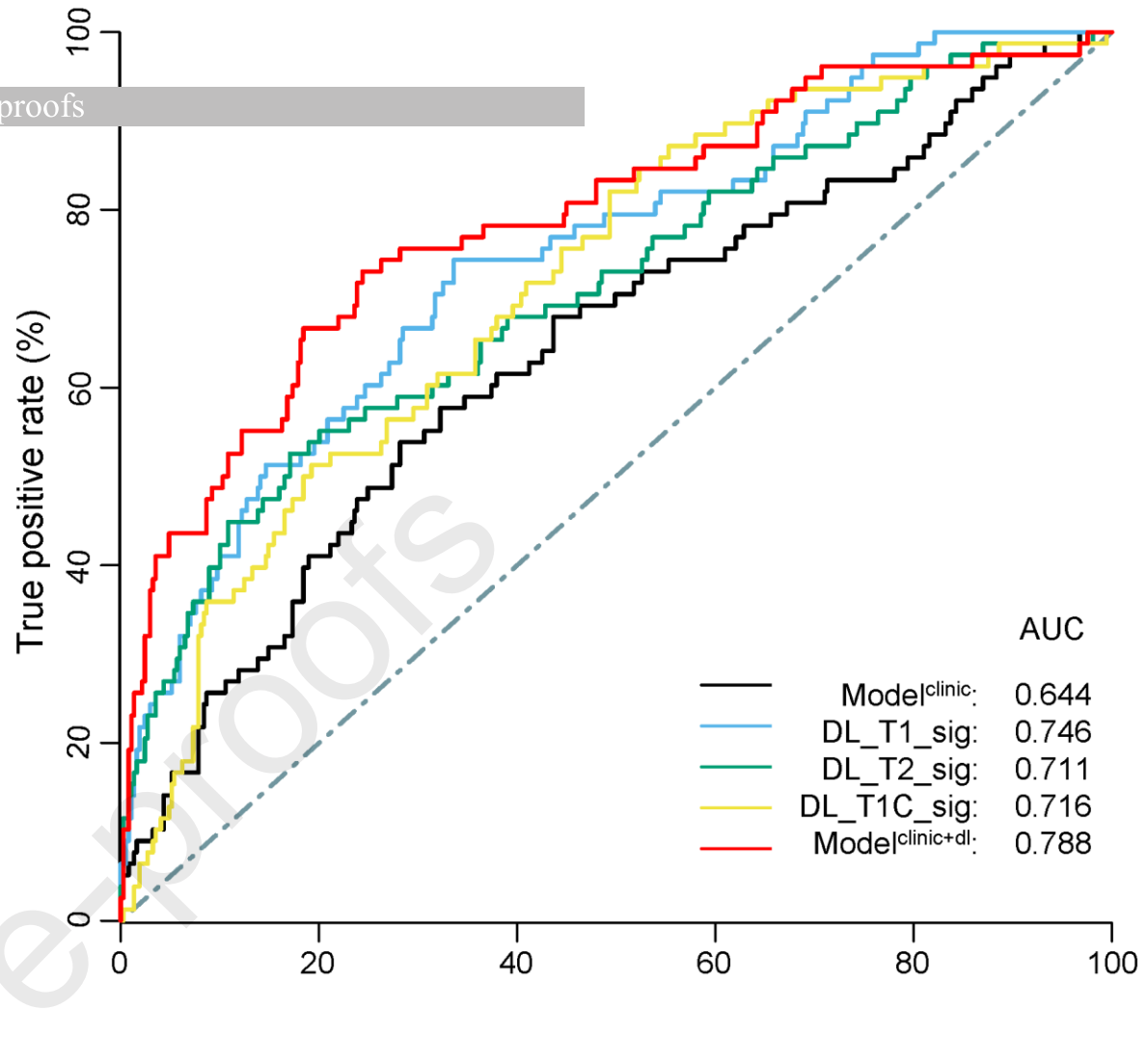
Highlights:

1. DL-based radiomic signatures were significantly correlated with prognosis of NPC.
2. DL-based radiomic signatures were complementary to clinical prognostic factors.
3. The radiomic nomogram improved prediction of DFS, OS, DMFS and LRFS of NPC.
4. The radiomic nomogram may assist in pretreatment risk stratification.

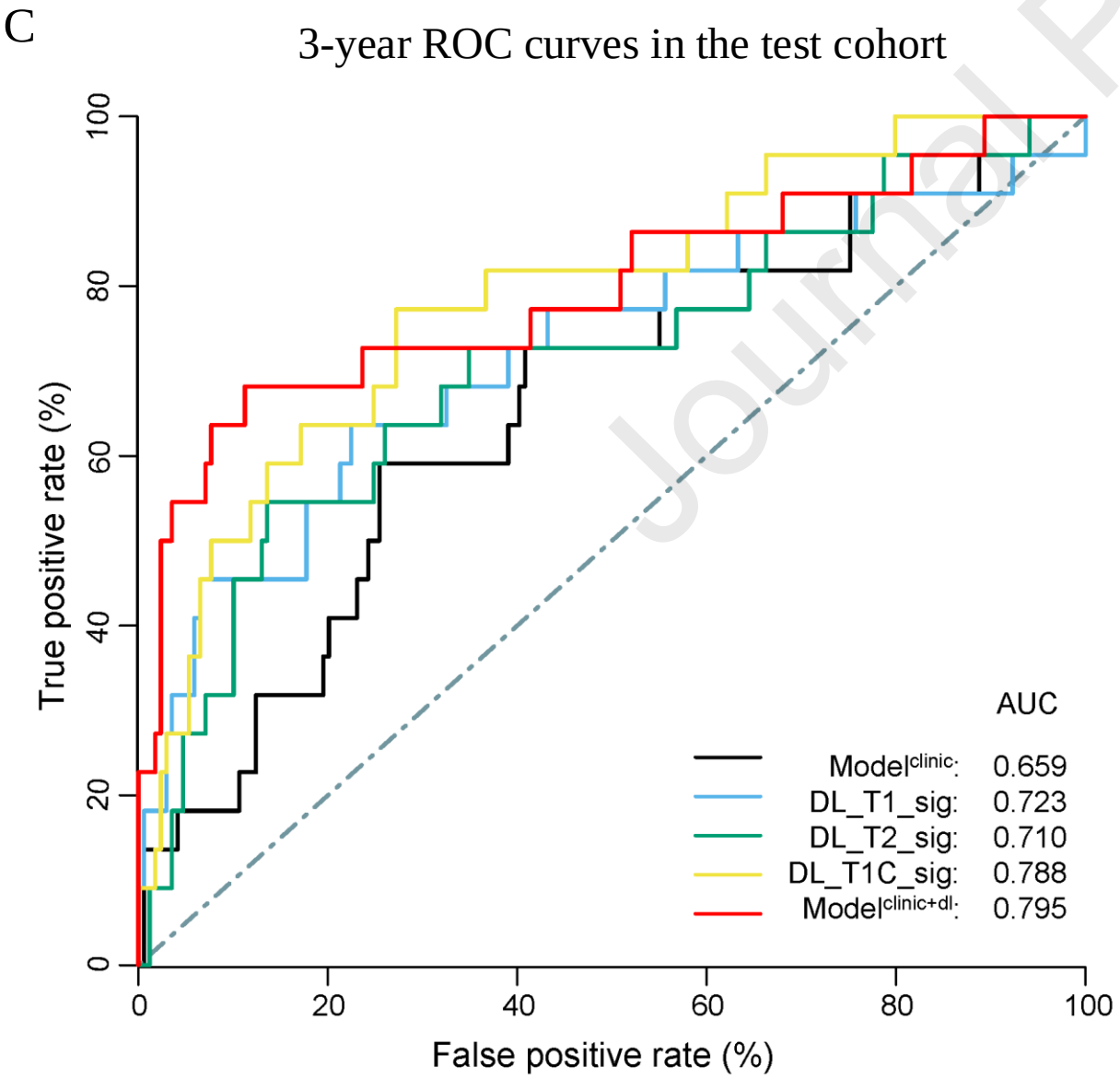
A



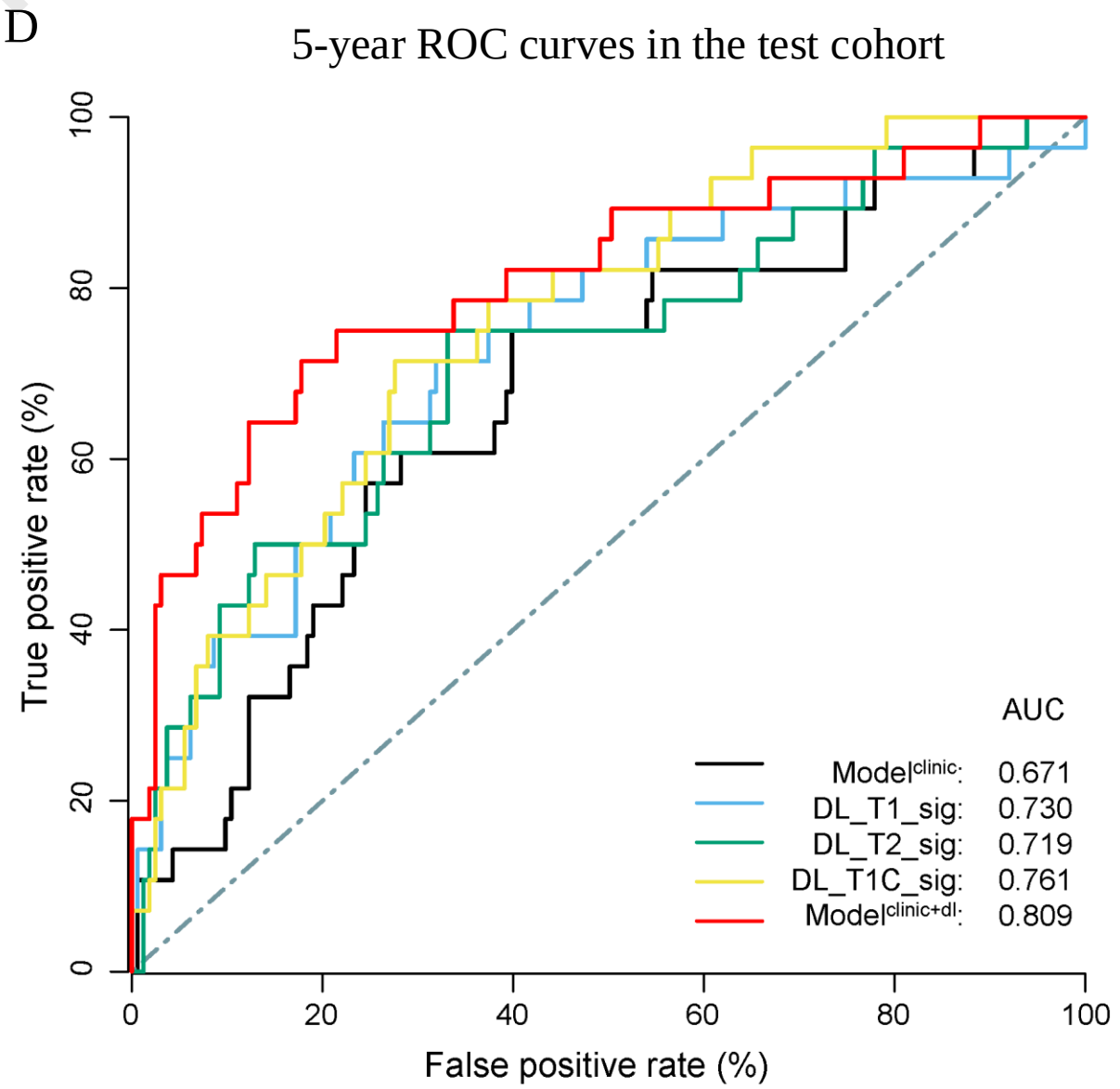
B

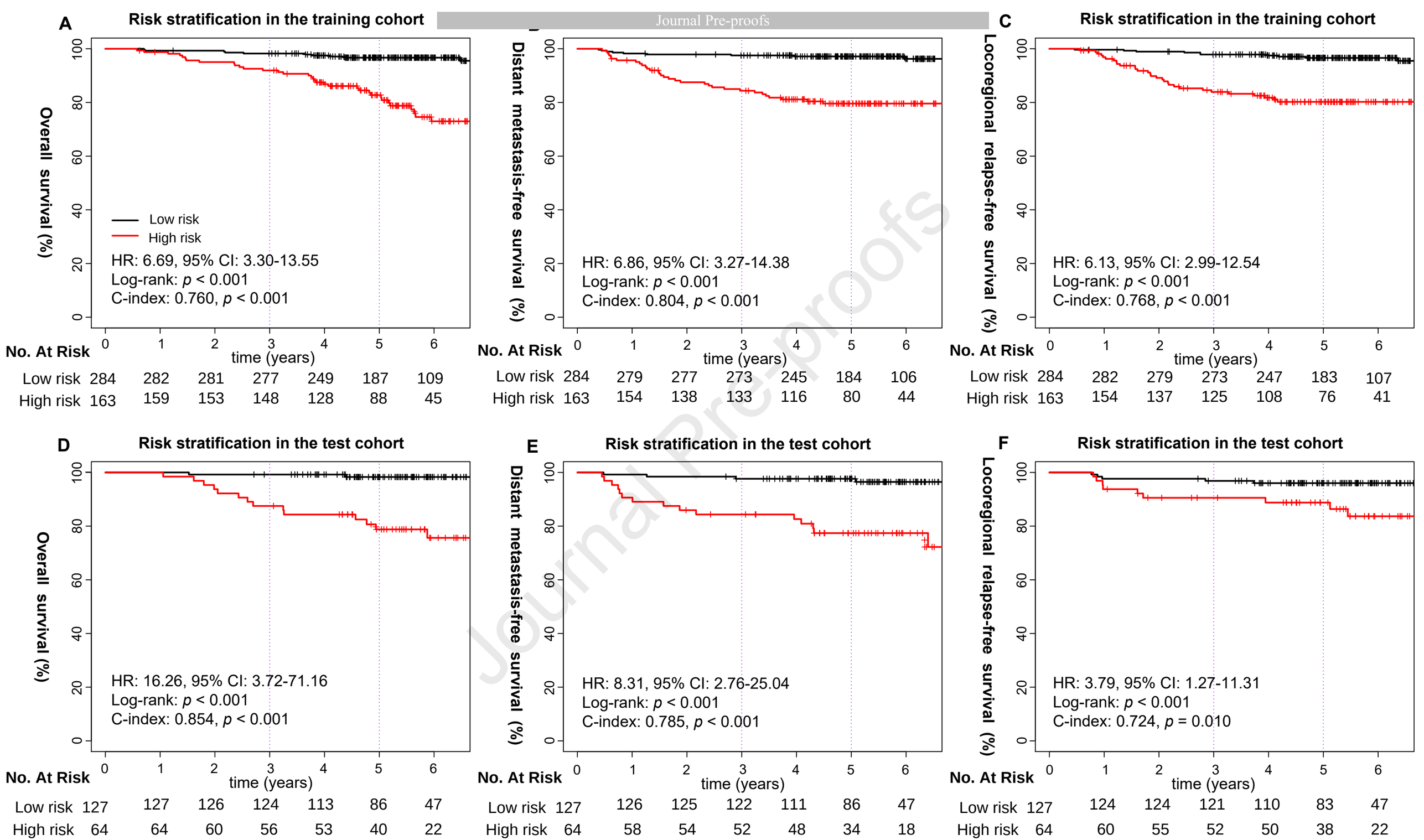


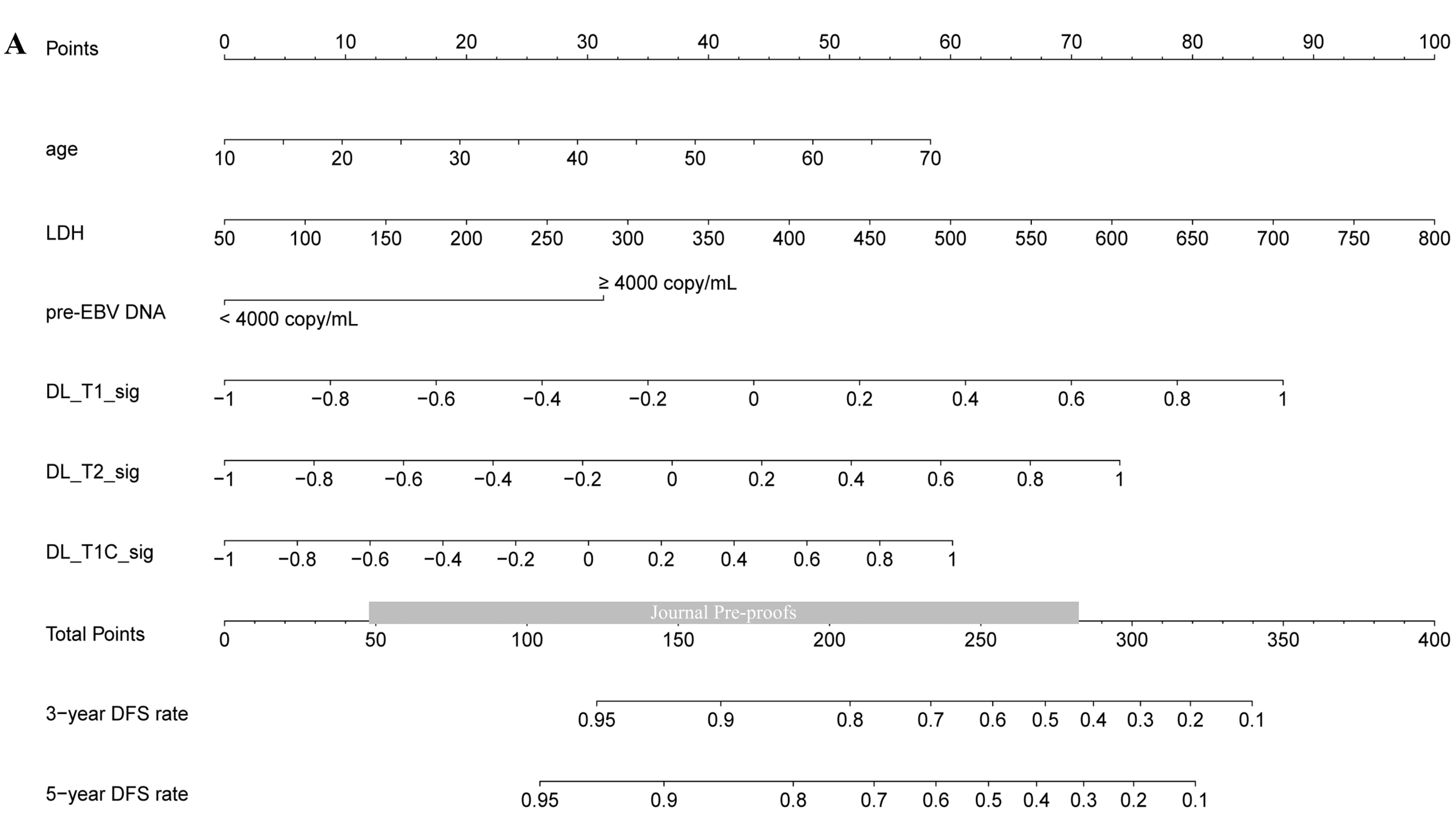
C

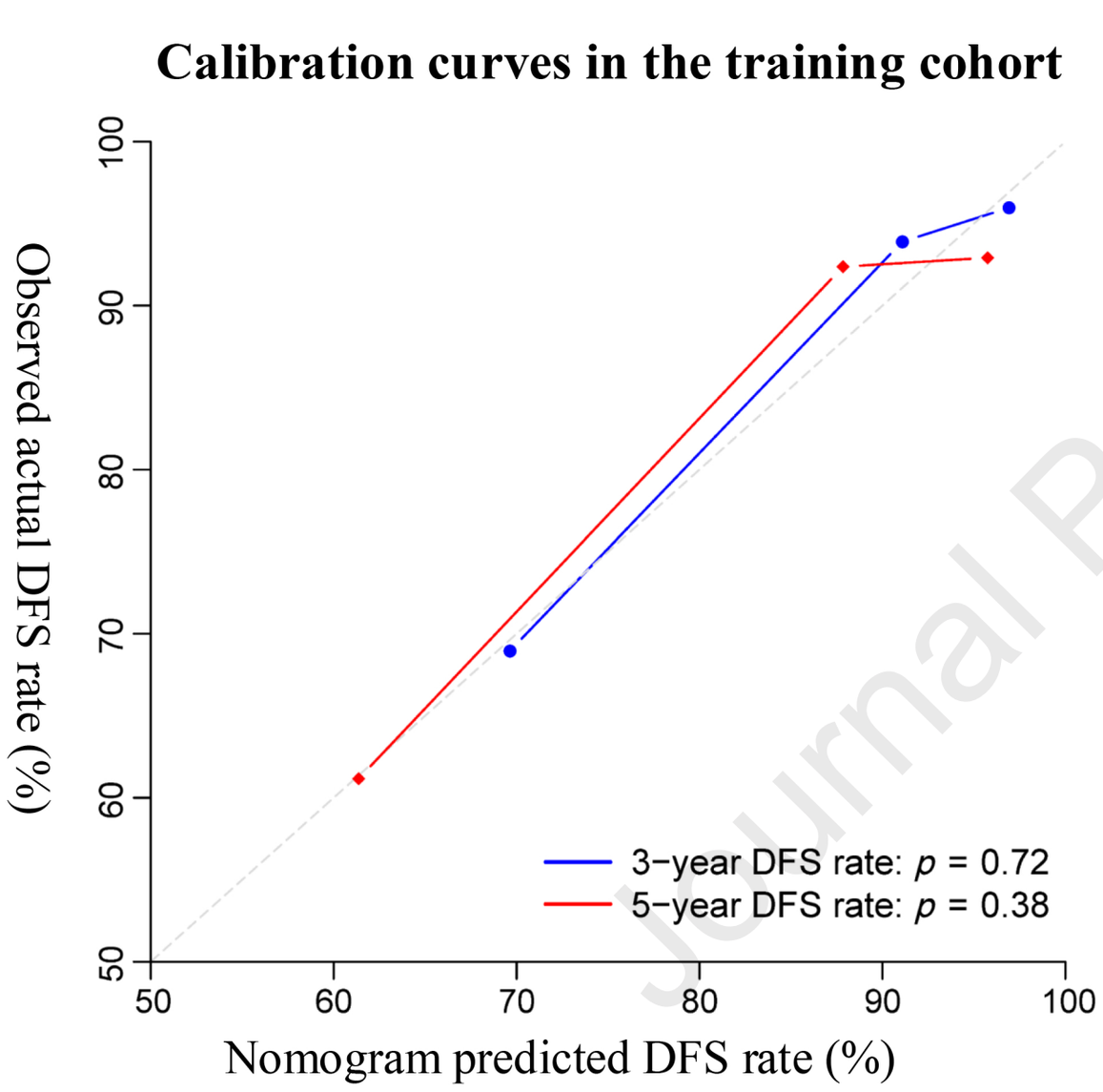


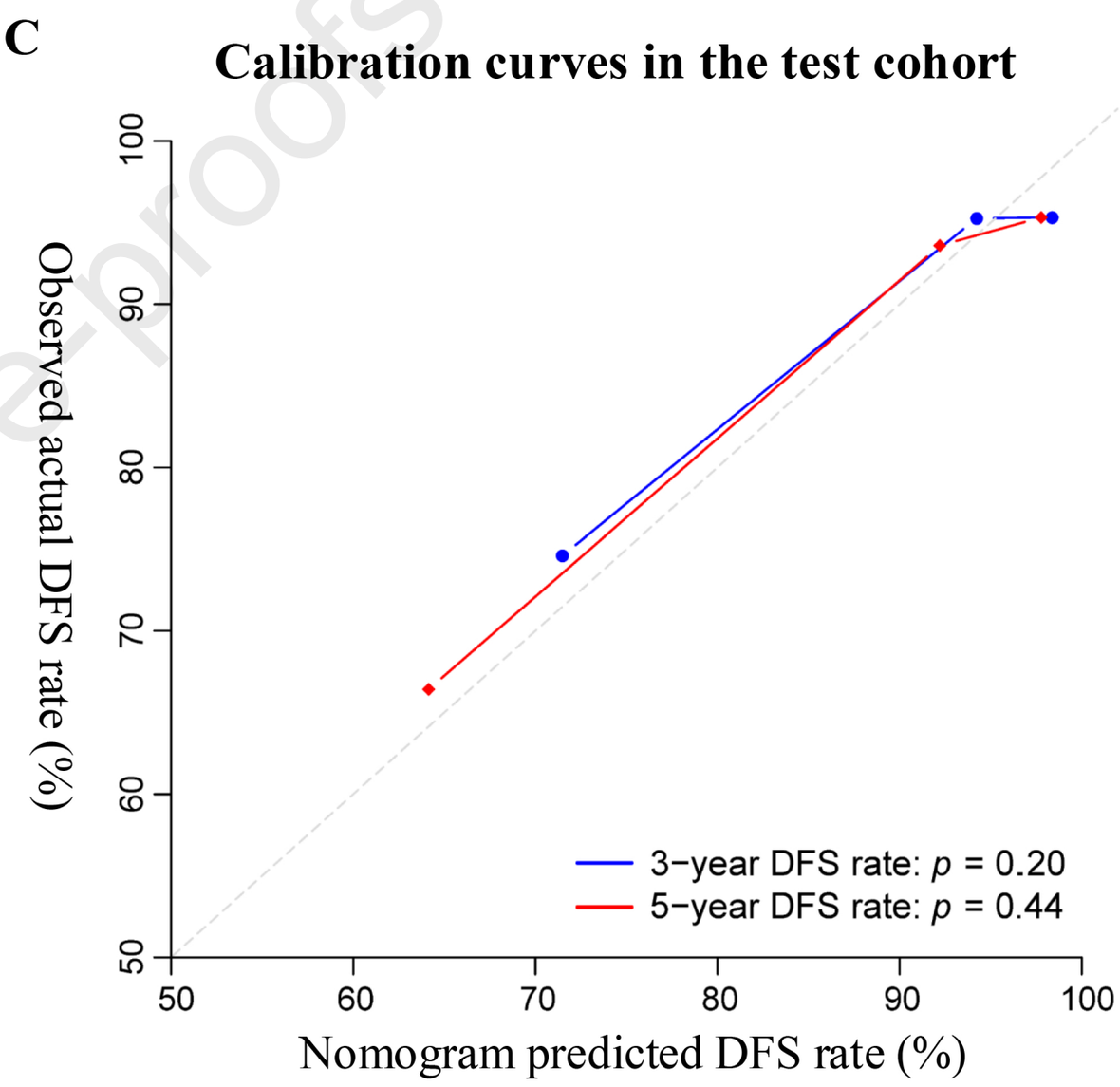
D

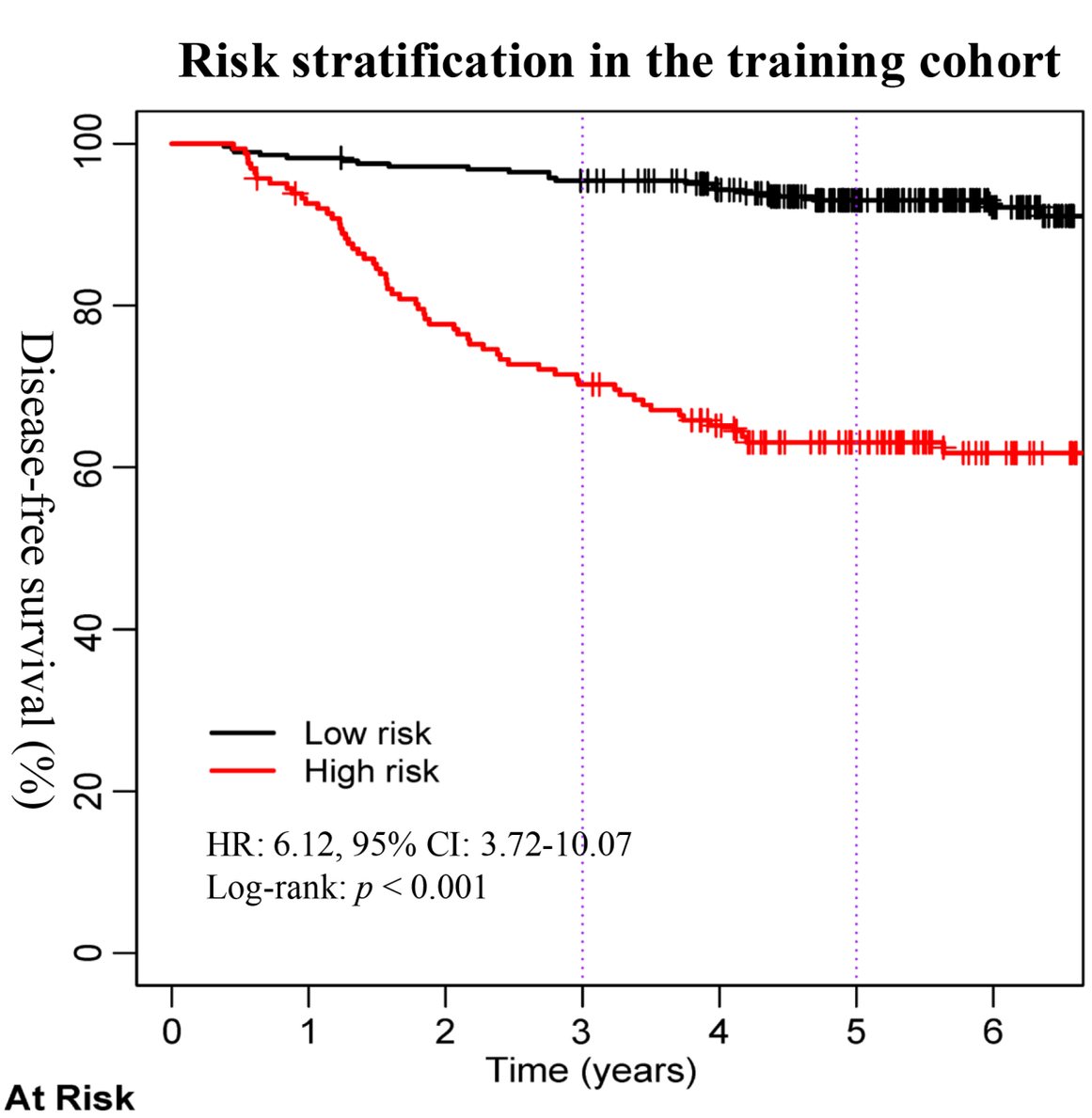


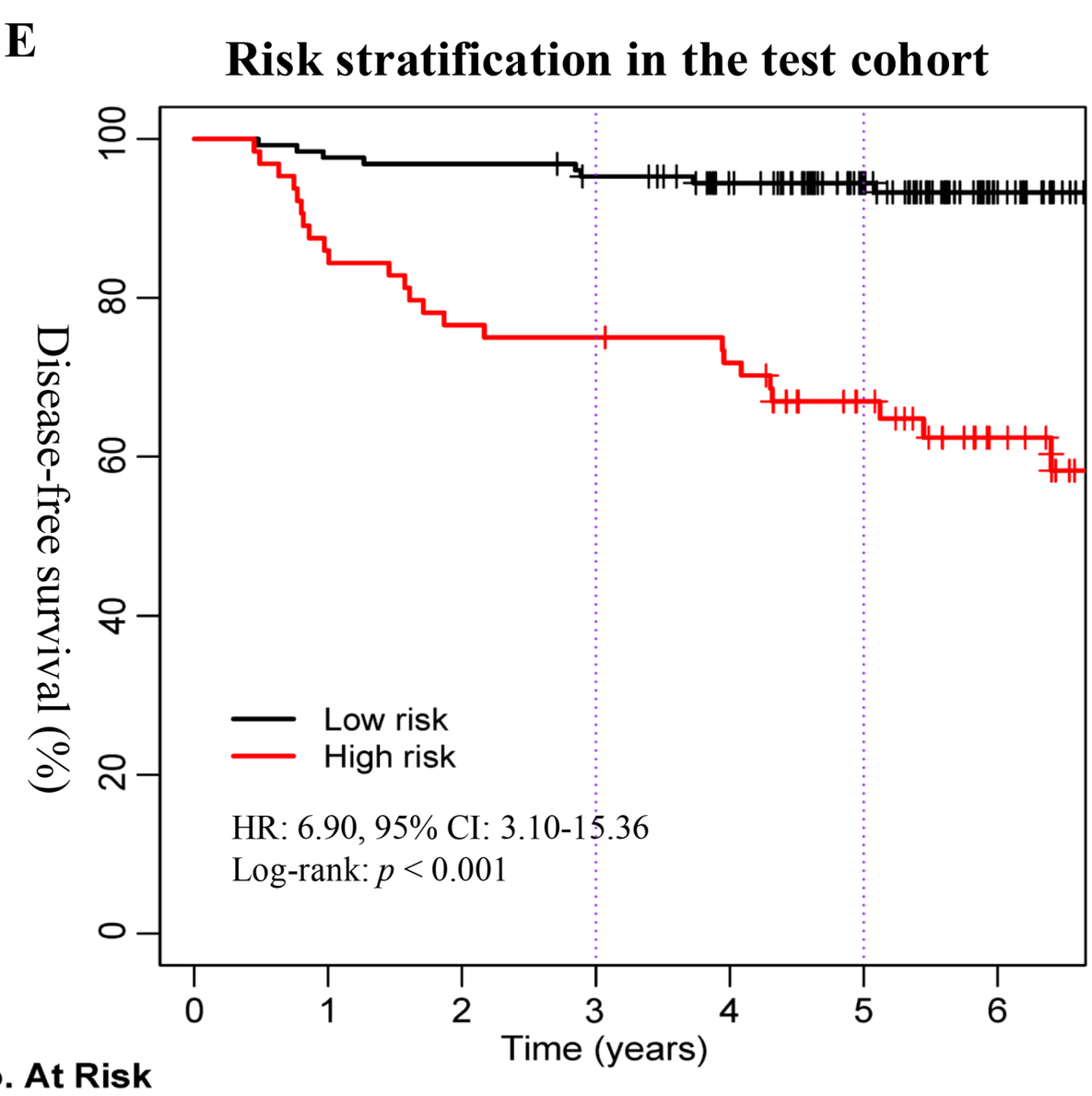












A: Image acquisition

B: Tumor segmentation

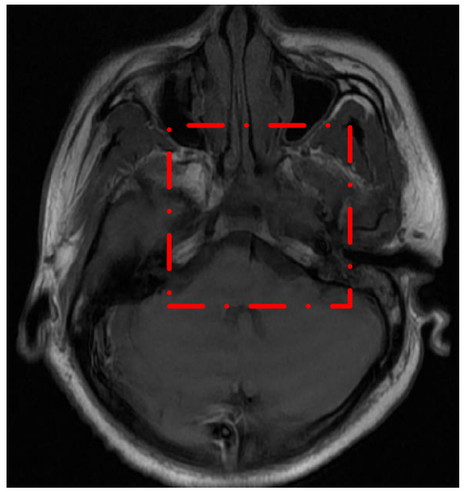
C: Image pre-processing

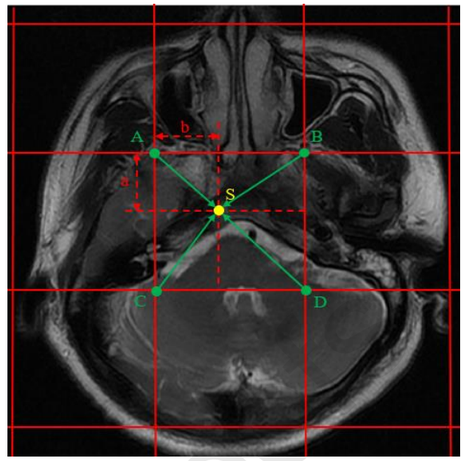
D: Model building and validation

T1W MR

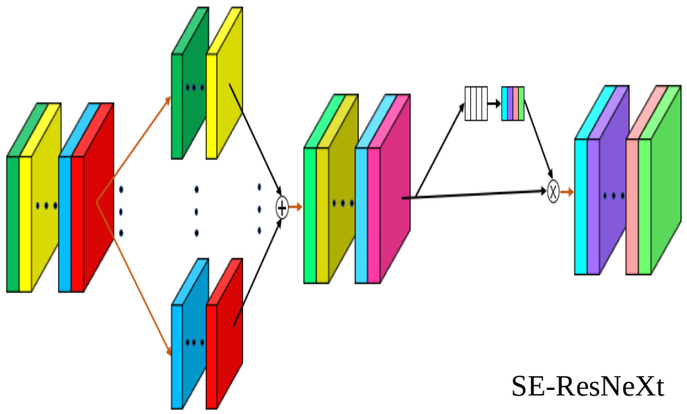
T2W MR

CE-T1W MR

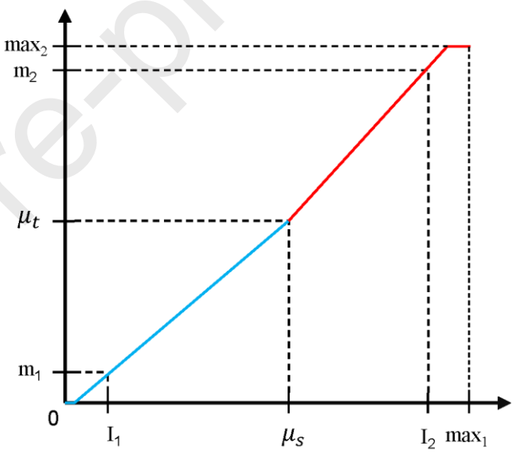




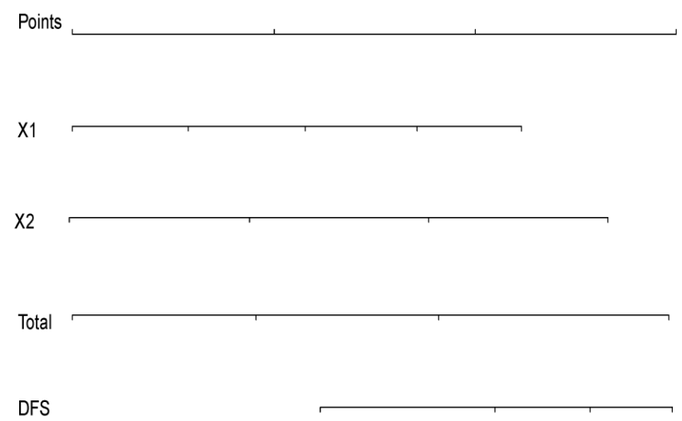
Bilinear interpolation



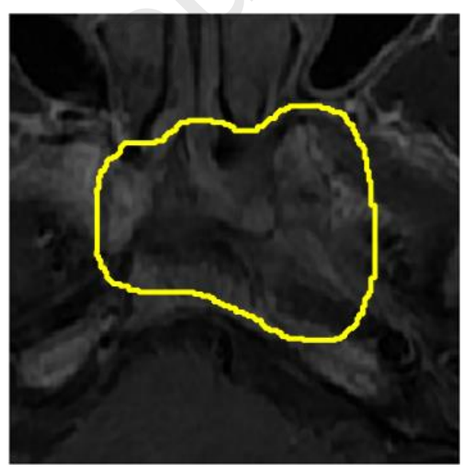
DL-based signature building

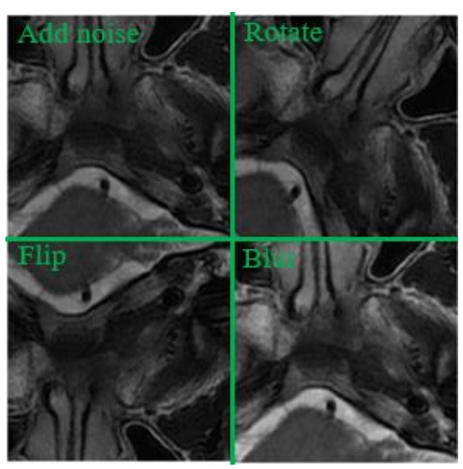


Histogram-based normalization

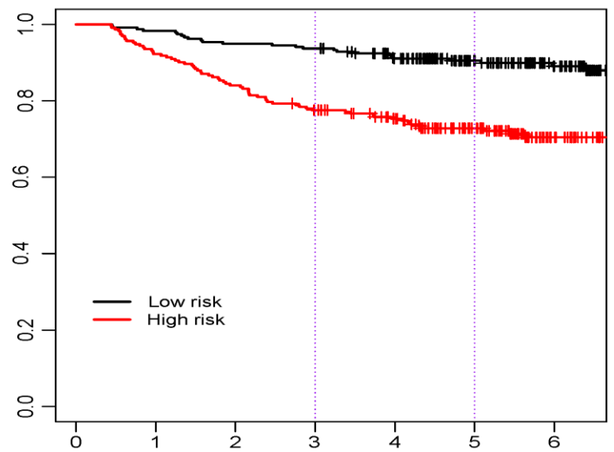


Radiomics nomogram





Data augmentation



Model validation