EBioMedicine 70 (2021) 103522

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

Research paper

A deep learning-based radiomic nomogram for prognosis and treatment decision in advanced nasopharyngeal carcinoma: A multicentre study

Lianzhen Zhong[a,b,1], Di Dong[a,b,1], Xueliang Fang[c,1], Fan Zhang[d,e,1], Ning Zhang[f,1], Liwen Zhang[a,b,1], Mengjie Fang[a,b], Wei Jiang[g], Shaobo Liang[h], Cong Li[a,b], Yujia Liu[a,b], Xun Zhao[a,b], Runnan Cao[a,b], Hong Shan[d,e], Zhenhua Hu[a,b,*], Jun Ma[c,*], Linglong Tang[c,*], Jie Tian[a,b,i,*]

[a] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, PR China
[b] CAS Key Laboratory of Molecular Imaging, Beijing Key Laboratory of Molecular Imaging, The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, No. 95 Zhongguancun East Road, Hai Dian District, Beijing 100190, PR China
[c] Department of radiation oncology, Sun Yat-sen University Cancer Centre, State Key Laboratory of Oncology in South China, Collaborative Innovation Centre for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangzhou, Guangdong Province 510060, PR China
[d] Department of head and neck oncology, The cancer centre of the Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, Guangdong Province 519000, PR China
[e] Guangdong Provincial Key Laboratory of Biomedical Imaging, The Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, Guangdong Province 519000, PR China
[f] Department of radiation oncology, First People's Hospital of Foshan Affiliated to Sun Yat-sen University, Foshan, Guangdong Province 528000, PR China
[g] Department of radiation Oncology, Affiliated Hospital of Guilin Medical University, Guilin, Guangxi Province 541000, PR China
[h] Department of radiation oncology, The Third Affiliated Hospital of Sun Yat-Sen University, Guangzhou, Guangdong Province 510000, PR China
[i] Beijing Advanced Innovation Centre for Big Data-Based Precision Medicine, School of Engineering Medicine, Beihang University, Beijing 100191, PR China

ARTICLE INFO

Article History:
Received 7 May 2021
Revised 9 July 2021
Accepted 22 July 2021
Available online xxx

Keywords:
Multi-task deep learning
Radiomic nomogram
Survival analysis
Treatment decision
Advanced nasopharyngeal carcinoma

ABSTRACT

Background: Induction chemotherapy (ICT) plus concurrent chemoradiotherapy (CCRT) and CCRT alone were the optional treatment regimens in locoregionally advanced nasopharyngeal carcinoma (NPC) patients. Currently, the choice of them remains equivocal in clinical practice. We aimed to develop a deep learning-based model for treatment decision in NPC.

Methods: A total of 1872 patients with stage T3N1M0 NPC were enrolled from four Chinese centres and received either ICT+CCRT or CCRT. A nomogram was constructed for predicting the prognosis of patients with different treatment regimens using multi-task deep learning radiomics and pre-treatment MR images, based on which an optimal treatment regimen was recommended. Model performance was assessed by the concordance index (C-index) and the Kaplan-Meier estimator.

Findings: The nomogram showed excellent prognostic ability for disease-free survival in both the CCRT (C-index range: 0.888-0.921) and ICT+CCRT (C-index range: 0.784-0.830) groups. According to the prognostic difference between treatments using the nomogram, patients were divided into the ICT-preferred and CCRT-preferred groups. In the ICT-preferred group, patients receiving ICT+CCRT exhibited prolonged survival over those receiving CCRT in the internal and external test cohorts (hazard ratio [HR]: 0.17, $p<0.001$ and 0.24, $p=0.02$); while the trend was opposite in the CCRT-preferred group (HR: 6.24, $p<0.001$ and 12.08, $p<0.001$). Similar results for treatment decision using the nomogram were obtained in different subgroups stratified by clinical factors and MR acquisition parameters.

Interpretation: Our nomogram could predict the prognosis of T3N1M0 NPC patients with different treatment regimens and accordingly recommend an optimal treatment regimen, which may serve as a potential tool for promoting personalized treatment of NPC.

Funding: National Key R&D Program of China, National Natural Science Foundation of China, Beijing Natural Science Foundation, Strategic Priority Research Program of CAS, Project of High-Level Talents Team Introduction in Zhuhai City, Beijing Natural Science Foundation, Beijing Nova Program, Youth Innovation Promotion Association CAS.

List of abbreviations: NPC, nasopharyngeal carcinoma; CCRT, concurrent chemoradiotherapy; ICT, induction chemotherapy; MR, magnetic resonance; DFS, disease-free survival; pre-EBV DNA, pre-treatment plasma Epstein-Barr virus DNA; CPTDN, Combined prognosis and treatment decision nomogram; C-index, Harrell's concordance index; HR, hazard ratio; MCox, multivariate Cox proportional hazards regression; IQR, interquartile range; WHO, World Health Organization

* Corresponding authors at.
E-mail addresses: zhenhua.hu@ia.ac.cn (Z. Hu), majun2@mail.sysu.edu.cn (J. Ma), tangll@sysucc.org.cn (L. Tang), jie.tian@ia.ac.cn (J. Tian).
[1] these authors contributed equally to this work.

https://doi.org/10.1016/j.ebiom.2021.103522

Research in Context

Evidence before this study

We searched *PubMed, Google Scholar* and *Web of Science* for papers published from database inception to May, 2021, with the terms ("radiomics" OR "deep learning" OR "machine learning" OR "MR-radiomics" OR "radiomic nomogram" OR "imaging predictor") AND ("NPC" OR "nasopharyngeal carcinoma" OR "prognosis and treatment decision" OR "prognosis and treatment response" OR "treatment decision" OR "treatment response" OR "benefit" OR "prognosis"), with no language restrictions and source restrictions. To the best of our knowledge, there are no published studies so far that have used radiomics to predict the prognosis of patients with cancers including NPC with different treatment regimens and accordingly recommend an optimal treatment regimen. Studies in treatment decision of NPC using radiomics have focused on risk stratification of a prognostic model/factor and development of predictive signatures for treatment response, but did not provide the prognosis of patients with different treatment regimen. None of the studies use multi-task deep learning radiomics.

Added value of this study

Our study is the first multi-task deep learning study that could predict prognosis and concurrently recommend treatment regimens for NPC patients, which was validated in both internal and external test cohorts from a multicentre and large-scale dataset, supporting the move toward personalized and optimized management of NPC.

Implications of all the available evidence

Our study suggests the potential of multi-task deep-learning radiomics for developing more comprehensive prognostic model with personalized treatment recommendations for NPC, which also offers a promising start for treatment decisions and prognostic prediction in other cancers. By providing the prognosis of patients with different treatment regimens and accordingly recommending an optimal treatment regimen, integrating our model into routine clinical practice would promote personalized treatment and optimize management of NPC.

1. Introduction

Nasopharyngeal carcinoma (NPC) is a radiosensitive epithelial malignancy associated with the risk of locoregional and distant aggressiveness, with the highest prevalence reported for the populations of East and Southeast Asia, and North Africa [1]. Of the 87,000 newly diagnosed NPC annually, over 70% of them are classified as advanced disease [2]. Patients with advanced NPC have several treatment options, including induction chemotherapy (ICT) plus concurrent chemoradiotherapy (CCRT) and CCRT alone [3,4]. Currently, the choice of individualized treatment regimen remains equivocal in clinical practice [3,5].

In recent decades, some prognostic biomarkers, such as mRNA [6], plasm protein [7], and tumour-infiltrating lymphocytes [8], had been developed for advanced NPC. However, these prognostic signature are not predictive biomarkers for treatment response, as they did not inform therapy strategy before treatment [9,10]. Therefore, it is very necessary to examine the impact of treatment regimen on prognosis of NPC patients to help optimize treatment decision.

Recently, the use of radiomics in oncology, which allows in-depth characterization of tumour phenotypes using a computerized imaging analysis algorithm, has garnered increasing attention [11,12]. The introduce of deep learning enables radiomics to perform end-to-end modelling of medical data and multi-task learning for multiple clinical tasks, which demonstrated excellent results [13-15]. Emerging evidence has confirmed that radiomics had potential value in predicting prognosis and treatment response in NPC [16-18]. However, no previous study has used multi-task learning to simultaneously predict prognosis and make treatment recommendations for NPC patients.

The aim of this study was to use multi-task deep-learning radiomics to develop simultaneously prognostic and predictive signatures from pre-treatment magnetic resonance (MR) images of NPC patients, and to construct a combined prognosis and treatment decision nomogram (CPTDN) for predicting the prognosis of NPC patients with different treatment regimens and accordingly recommending an optimal treatment regimen. In this study, we focused on stage T3N1M0 NPC patients, the largest subgroup of advanced NPC, and collected a large multicentre dataset used for model development and validation.

2. Methods

2.1. Patients

The ethical review board of Institute of Automation, Chinese Academy of Sciences approved this retrospective study (ID: IA-202043) and waived the requirement of informed consent.

The overall study design is shown in **Fig. 1**. We retrospectively reviewed the charts of patients with NPC diagnosed at the four Chinese centres (centre 1: Sun Yat-sen University Cancer Centre, centre 2: First People's Hospital of Foshan Affiliated to Sun Yat-sen University, centre 3: Fifth Affiliated Hospital of Sun Yat-sen University, centre 4: Guilin Medical University Affiliated Hospital) between January 2010 and June 2017. A total of 1872 candidate eligible patients with stage T3N1M0 NPC and available diagnostic MR images were included, according to the inclusion and exclusion criteria.

The inclusion criteria: had diagnosed T3N1M0; treated by either ICT+CCRT or CCRT alone; received intensity-modulated radiotherapy; underwent MRI scans within 2 weeks before treatment.

The exclusion criteria: received adjuvant chemotherapy, targeted therapy, or biotherapy during the course of radical treatment; with previous chemotherapy or radiotherapy or other malignant tumours; had artifacts, blurs, faults, and disordered slices in the MR image.

Data on demographic characteristics, pre-treatment plasma Epstein-Barr virus DNA (pre-EBV DNA) levels, and treatment regimen were gathered from the medical records of each centre. All patients were treated by either CCRT or ICT+CCRT. To reduce the impact of clinicians' judgment on the choice of treatment regimen, a nonparametric matching method was used to screen out patients with similar baseline characteristics in the CCRT and ICT+CCRT groups at a ratio of 1:1 [19] (**Appendix A**). Matched patients from centre 1 were divided into a training cohort and an internal test cohort, and those from the other three centres were divided into an external test cohort.

The primal endpoint is disease-free survival (DFS), the time from the date of initial diagnose until either the date of disease progression or death from any cause. The protocol of disease progression identification, including local and regional recurrence or distant metastasis, was the same as previously reported [17]. Tumour staging was determined

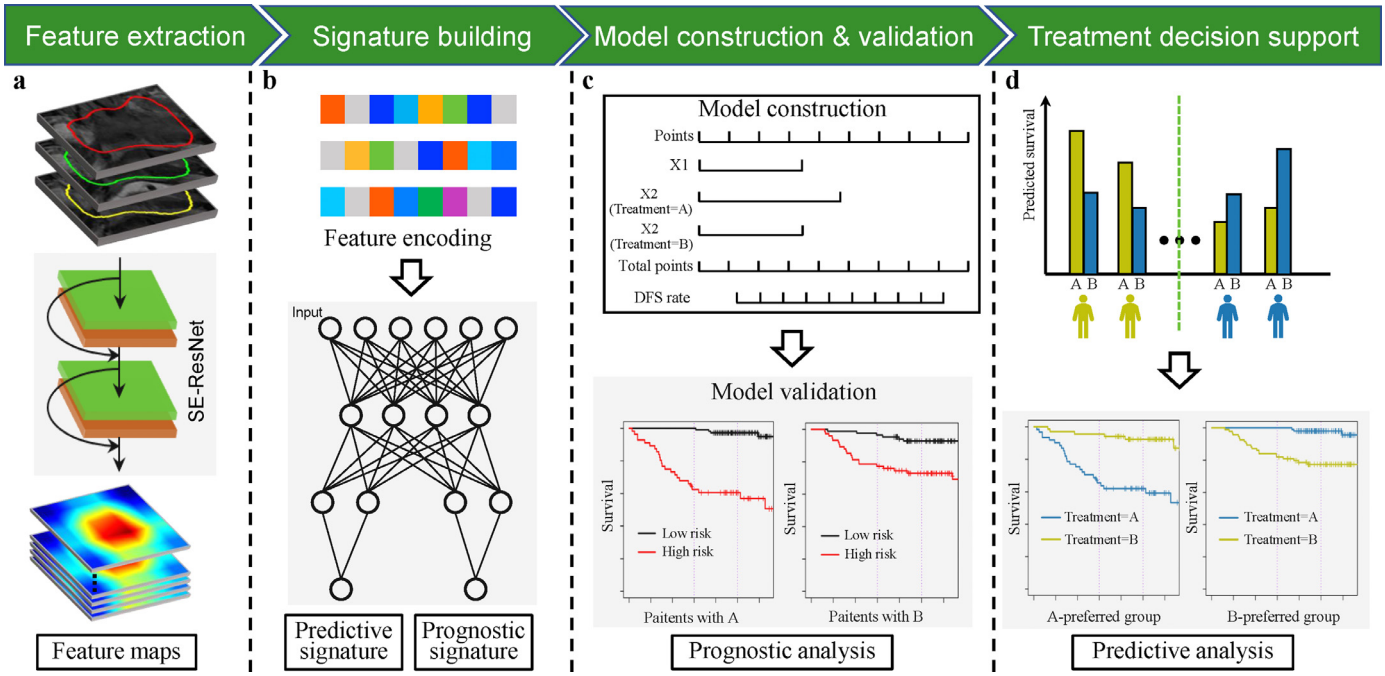

Fig. 1. Deep-learning radiomics workflow. (a) Extracting deep radiomic features from multi-sequence magnetic resonance images. **(b)** Building prognostic and predictive signatures using a unified fully connected neural network. **(c)** Constructing and validating a prognostic nomogram that could predict the prognosis of patients with different treatment regimens. **(d)** Supporting treatment decision using the prognostic difference between treatments based on the nomogram. DFS, disease-free survival.

based on the American Joint Committee on Cancer TNM Staging Manual, eighth Edition [20]. Treatment details are summarised in **Appendix A**.

2.2. MR acquisition and tumour segmentation

In two weeks prior to any anti-tumour therapy, all patients underwent three sequences of MRI (T1-weighted, T2-weighted, and contrast enhanced T1-weighted) scanning. The MR acquisition protocols (including the details of manufacturer and magnetic field strength, among others) are shown in **Appendix B**. For each MR sequence, the primary tumour was manually contoured on each axial MR slice by two experienced radiologists (X.L.F. and F.Z.).

2.3. Feature extraction and radiomic signature building

Considering scanner-dependent bias in image intensity, the MR images were interpolated and normalized to match the reference template [21]. We used the SE-ResNet architecture [22] as the backbone network combined with the multiple instance learning method [23] to extract the deep radiomic features from the MR images. Afterward, taking the deep radiomic features from three MR sequences as input, a unified fully connected neural network, which consists of a shared backbone network and two task-specific subnetworks, was constructed to simultaneously predict prognosis and treatment response. The output of the subnetwork for predicting prognosis (Prognostic-score) was the estimate of the risk of disease progression, whereas that of another subnetwork (Predictive-score) was the estimate of the relative risk of disease progression associated with receiving ICT+CCRT *vs.* receiving CCRT. In the subnetwork that predicted treatment response, interactions between treatments and activation units were considered, using the modified covariate method [18]. Cox partial log likelihood [24] was chosen as the loss function to train feature extraction networks and signature building network in the training cohort. Specific architectures and training details of these networks were shown in **Appendix C**.

2.4. CPTDN construction

Univariate Cox proportional hazards regression analysis was used to evaluate the association between clinical characteristics and DFS and response to treatment regimen in the training cohort. Incorporating independent clinical factors, Prognostic-score, and Predictive-score, a CPTDN was developed using multivariate Cox proportional hazards regression (MCox) method and was plotted for the convenient of clinicians. Using the CPTDN, the prognosis predictions of each patient were calculated when it receives different treatment regimens, based on which the prognostic difference was used to make a personalized treatment recommendation.

To compare to CPTDN in prognostic prediction, three clinical models were built, using MCox method in all patients ($Model^{clin}_{all}$), those receiving CCRT ($Model^{clin}_{ccrt}$) and those receiving ICT+CCRT ($Model^{clin}_{ict+ccrt}$) of the training cohort. In addition, $Model^{clin+prog}_{all}$ was built using MCox method in the training cohort to assess the incremental prognostic value of prognostic-score in clinical practice.

2.5. Performance assessment

Harrell's concordance index (C-index) and hazard ratio (HR) estimates were used to assess and test prognostic performance of the proposed models in all cohorts. The quality of Predictive-score was assessed by its $p_{interaction}$ value, indicative of treatment interaction.

Furthermore, the calibration curve was plotted to evaluate the agreement between the observed survival and the CPTDN-predicted survival. Subgroup analysis using clinical factors and MR acquisition parameters was performed to assess the stability of CPTDN. In addition, we evaluated the prognostic performance of CPTDN in terms of second clinical endpoints, including overall survival, distant metastasis-free survival, and locoregional relapse-free survival.

2.6. Statistical analysis

The survival curves were drawn using Kaplan-Meier method. The *p*-value of HR estimate was computed using the log-rank test. In the

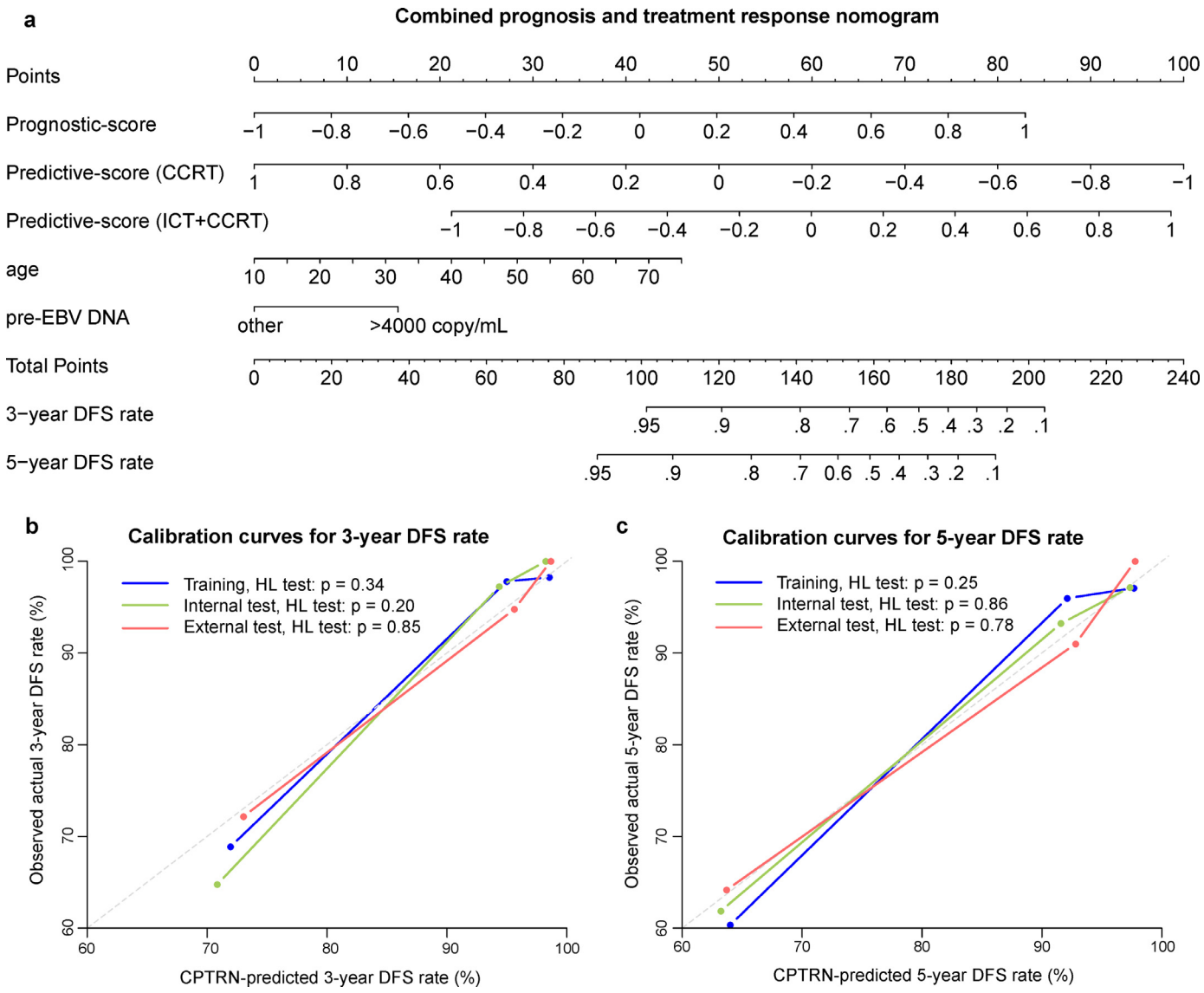

Fig. 2. Combined prognosis and treatment decision nomogram (CPTDN) and its calibration curves. (**a**) A CPTDN was built for predicting disease-free survival (DFS) of patients with different treatment regimens and accordingly recommend an optimal treatment regimen. The calibration curves of the CPTDN were plotted for predicting the 3-year (**b**) and 5-year DFS rate (**c**) in three cohorts (n=1206). The CPTDN could noninvasively recommend an optimal treatment regimen and predict patient prognosis. For example, consider a 40-year-old target patient with stage T3N1M0, a pre-EBV DNA of >4000 copy/mL, a Prognostic-score of 0.0, and a Predictive-score of -0.4; If the patient received ICT+CCRT, the patient's total score would be 122 (21+15+41+45), for which the 5-year DFS rate is estimated at 81%. If the patient received CCRT, the patient's score would be 147 (21+15+41+70), for which the 5-year DFS rate is estimated at 63%. The 5-year prognostic difference is 18%, so the CPTDN would recommend ICT+CCRT and provide a 5-year DFS rate of 81%. CCRT, concurrent chemoradiotherapy; ICT, induction chemotherapy; pre-EBV DNA, pre-treatment plasma Epstein-Barr virus DNA. HL test: Hosmer-Lemeshow test.

MCox analysis, a hybrid mode of forward and backward stepwise selection was applied to select independent features using Akaike's information. A two-sided $p<0.05$ means that the corresponding estimate reaches significant difference.

Open-source R software v3.6.1 was used to conduct statistical analysis, and open-source Python v3.6.8 and TensorFlow v1.12 was used to implement deep learning network models. A detailed description of the R packages used is shown in **Appendix C**.

2.7. Role of the funding source

The funders had no role in the study design, data collection, analysis, patient recruitment, writing of the manuscript, the decision to submit the manuscript for publication, or any aspect pertinent to the study. All authors had full access to the full data in the study and accept responsibility to submit for publication.

3. Results

3.1. Clinical characteristics

According to sex, age, smoking status, drinking status, family history of cancer, haemoglobin levels, tumour volume, and pre-EBV DNA levels, 1008 matched NPC patients with stage T3N1M0 in centre 1 were selected and divided randomly into the training (n=684) and internal test (n=324) cohorts. The external test cohort comprised of 198 matched target patients from three other centres. All matched patients had balanced baseline clinical characteristics between treatments (**Table A1**). Baseline clinical characteristics are summarized in **Table 1**; they did not show significant differences between the cohorts, except for the pre-EBV DNA levels. Median follow-up was 64.0 months (interquartile range [IQR]: 53.4-78.3), 65.7 months (IQR: 52.9-77.6), and 63.3 months (IQR: 50.9-76.6) in the training, internal test, and external test cohorts, respectively. At the last follow-up,

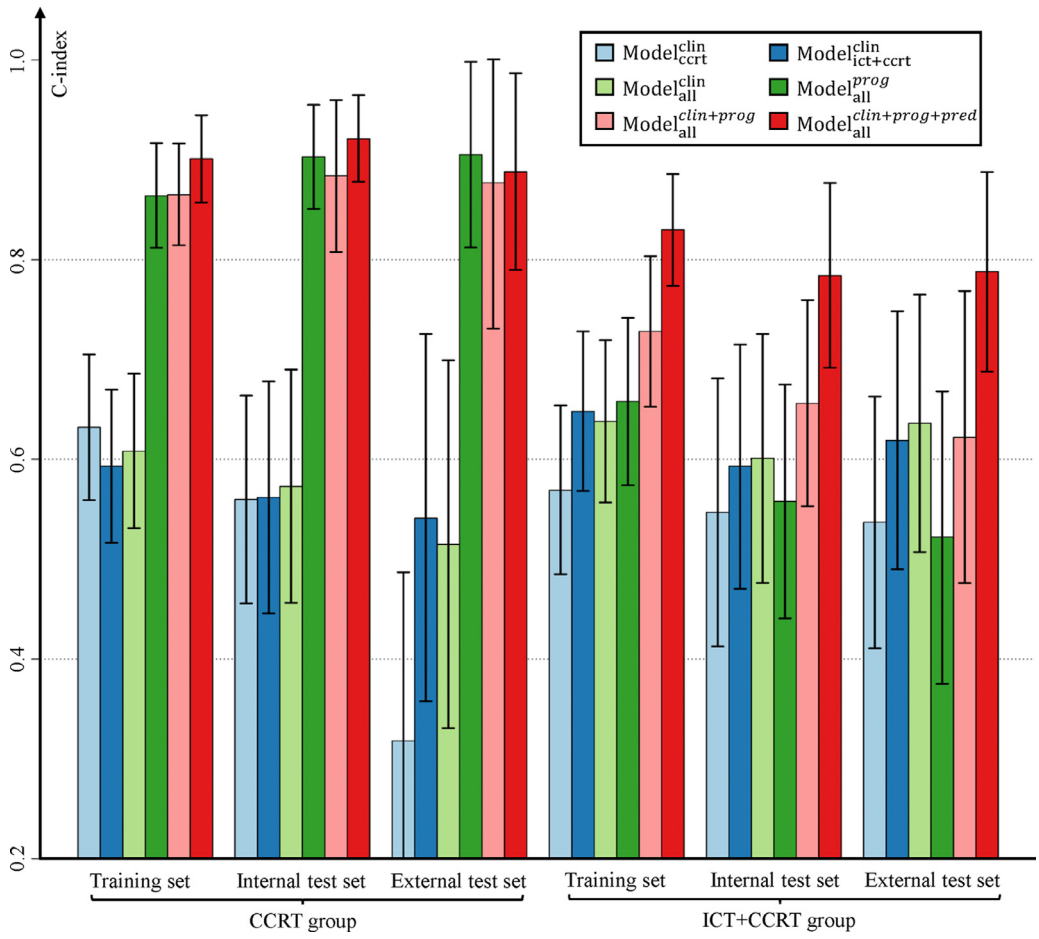

Fig. 3. Comparison of prognostic models' performance on different treatment-specific subgroups. The vertical line over the bar presents the confidence interval of the C-index. CCRT, concurrent chemoradiotherapy; ICT, induction chemotherapy.

107/684 (15.6%), 55/324 (17.0%), and 35/198 (17.7%) patients had experienced a confirmed disease progression in each of the cohort, respectively.

3.2. Development and validation of radiomic signatures

Prognostic-score and Predictive-score were developed simultaneously by the unified fully connected neural network. Prognostic-score exhibited a good predictive value for DFS in the training cohort (C-index: 0.772, 95% confidence interval [CI]: 0.721-0.822). Similar prognostic ability of Prognostic-score was obtained in the internal test (C-index: 0.733, 95% CI: 0.657-0.809), and external test (C-index: 0.681, 95% CI: 0.568-0.793) cohorts. Predictive-score exhibited a strong association with treatment response in all cohorts (all $p_{interaction} < 0.001$ [Wald test]). Furthermore, Predictive-score did not exhibit a consistently significant association with any clinical factor in all cohorts (**Table D1**).

3.3. Construction and prognostic performance of CPTDN

Univariate Cox proportional hazards regression analysis (**Fig. D1**) demonstrated that age (HR: 1.25 [per 10 years], log-rank test: $p = 0.020$) and pre-EBV DNA levels (HR: 2.08, log-rank test: $p < 0.001$) were significantly associated with DFS, and that no clinical factor was significantly associated with treatment response (all $p_{interaction} > 0.05$ [Wald test]). Incorporating Prognostic-score, interaction item of treatment with Predictive-score, pre-EBV DNA levels, and age, CPTDN was constructed for DFS prediction in the training cohort using the MCox analysis and is plotted in **Fig. 2a**. Meanwhile, *Mode* l^{clin}_{all} was constructed using age and pre-EBV DNA levels; *Model*$^{clin}_{ccrt}$ was constructed using tumour volume and pre-EBV DNA levels; *Mode* $l^{clin}_{ict+ccrt}$ was constructed using age and pre-EBV DNA levels; and *Mode* $l^{clin+prog}_{all}$ was constructed using age, Prognostic-score and pre-EBV DNA levels. CPTDN obtained the best prognostic ability in the treatment-specific subgroup of all cohorts (**Fig. 3**). Furthermore, the C-index of CPTDN (training: 0.868 [95% CI 0.833-0.903], internal test: 0.856 [0.807-0.905], external test: 0.851 [0.787-0.914]) was also highest compared to other models in all cohorts (**Table D2**).

CPTDN presents good agreement with the actual observed DFS rate at 3-year and 5-year (**Fig. 2b** and **2c**). Risk stratification analysis demonstrated that CPTDN could successfully classify patients into the low-risk and high-risk groups with significant differences in DFS (log-rank test: $p < 0.05$, **Fig. 4**). Similarly, CPTDN could identify patients with good prognosis for second clinical endpoints in the combined internal and external test cohort (**Fig. D2**). In addition, subgroup analysis demonstrated that CPTDN retained the capacity for risk stratification for DFS (**Fig. D3**) when considering clinical factors and MR parameters, including sex, C-reaction protein levels (normal or abnormal), smoking status (yes or no), family history of cancer (yes or no), tumour volume (≤29 mL or >29 mL), magnetic field strength and manufacturer of MR scanner (<3.0T or =3.0T).

3.4. Performance for treatment decision of CPTDN

According to the CPTDN, the prognostic difference was calculated using the difference in DFS rate of receiving ICT+CCRT *vs.* receiving CCRT. The 5-year prognostic difference of receiving ICT+CCRT *vs.* receiving CCRT divided patients into the ICT-preferred group and

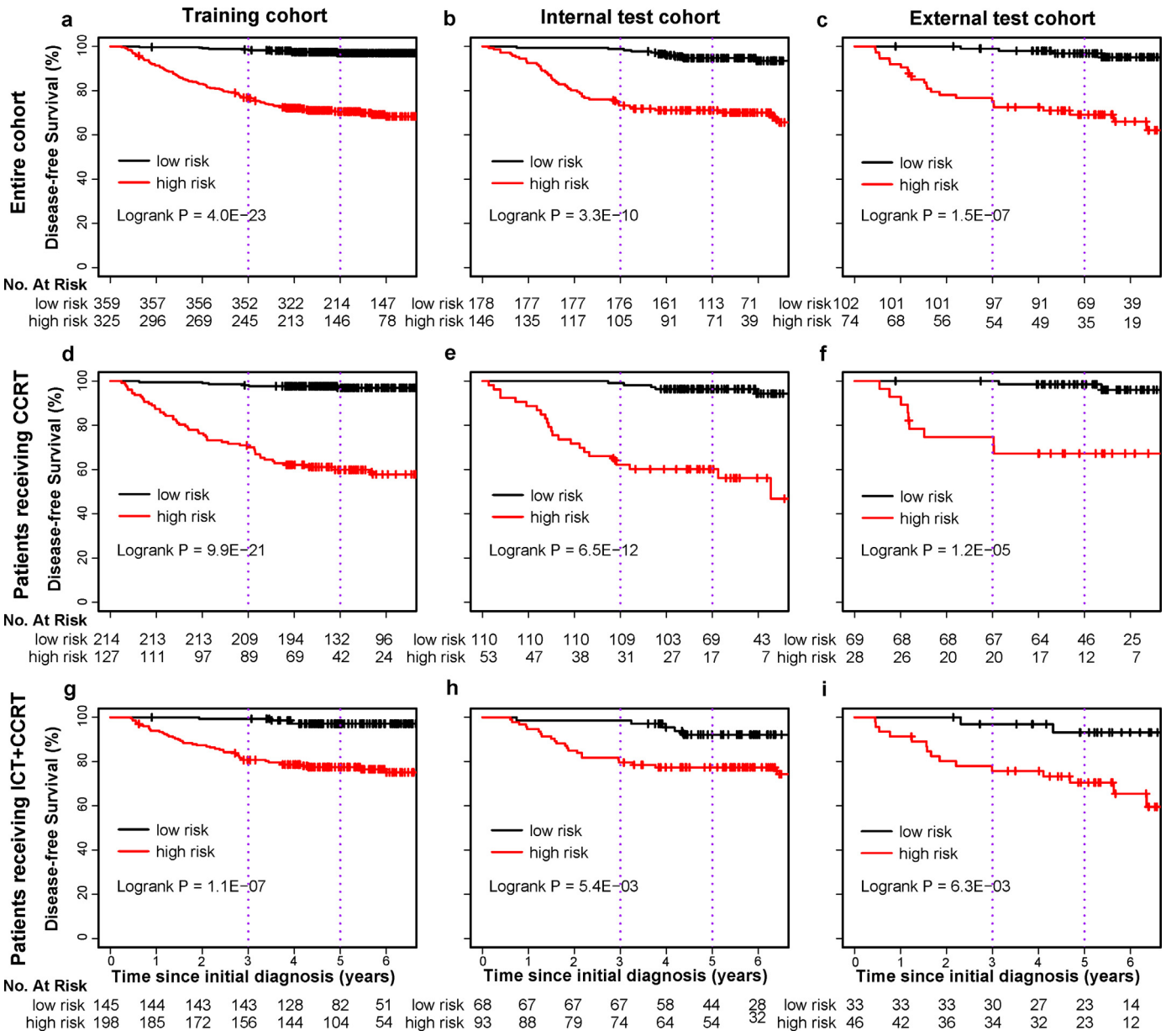

Fig. 4. Risk stratification analysis of CPTDN for predicting disease-free survival in three cohorts (n=1206). **(a)** Kaplan-Meier curves of disease-free survival in the entire group, **(d)** patients receiving CCRT, and **(g)** patients receiving ICT+CCRT in the training cohort. **(b)** Kaplan-Meier curves of disease-free survival in the entire group, **(e)** patients receiving CCRT, and **(h)** patients receiving ICT+CCRT in the internal test cohort. **(c)** Kaplan-Meier curves of disease-free survival in the entire group, **(f)** patients receiving CCRT, and **(i)** patients receiving ICT+CCRT in the external test cohort. All Kaplan-Meier curves were stratified by the mean total point (100) of CPTDN outputs. CCRT, concurrent chemoradiotherapy; ICT, induction chemotherapy.

CCRT-preferred group, using a threshold of 0% (Fig. 5). In the ICT-preferred group (difference >0%), ICT+CCRT was a preferred treatment regimen with an improved DFS compared to CCRT in the training (HR: 0.21, 95% CI: 0.11-0.38, log-rank test: $p < 0.001$), internal test (HR: 0.17, 95% CI: 0.07-0.42, log-rank test: $p < 0.001$) and external test (HR: 0.24, 95% CI: 0.07-0.90, log-rank test: $p = 0.022$) cohorts. While in the CCRT-preferred group (difference ≤0%), CCRT was a preferred treatment regimen compared to ICT+CCRT in the training (HR: 5.34, 95% CI: 2.49-11.47, log-rank test: $p < 0.001$), internal test (HR: 6.24, 95% CI: 2.14-18.19, log-rank test: $p < 0.001$) and external test (HR: 12.08, 95% CI: 2.74-53.19, log-rank test: $p < 0.001$) cohorts. Moreover, in the treatment decision analysis of the 5-year prognostic difference with 2% and 5% as cut-off points, it remained an effective indicator for identifying patients who were suitable for CCRT or ICT+CCRT (**Fig. D4** and **D5**). Furthermore, CPTDN could identify patients who were suitable for CCRT or ICT+CCRT for second clinical endpoints in the combined internal and external test cohort (Fig. 6). When stratified by sex, age (≤42 years or >42 years), tumour volume (≤29 mL or >29 mL), and magnetic field strength of MR scanner (<3.0T or =3.0T), the 5-year prognostic difference still identified patients who were suitable for CCRT or ICT+CCRT (**Fig. D6**).

4. Discussion

In this large multicentre study, we developed a nomogram (CPTDN) for prognosis and treatment decision in NPC patients with stage T3N1M0. The CPTDN was constructed by the predictive radiomic signature and independent prognostic factors, which could predict the prognosis of patients with different treatment regimens and accordingly recommend an optimal treatment regimen. Our CPTDN identified patients most likely to benefit from specific treatment regimens and presented better prognostic ability than the clinical models

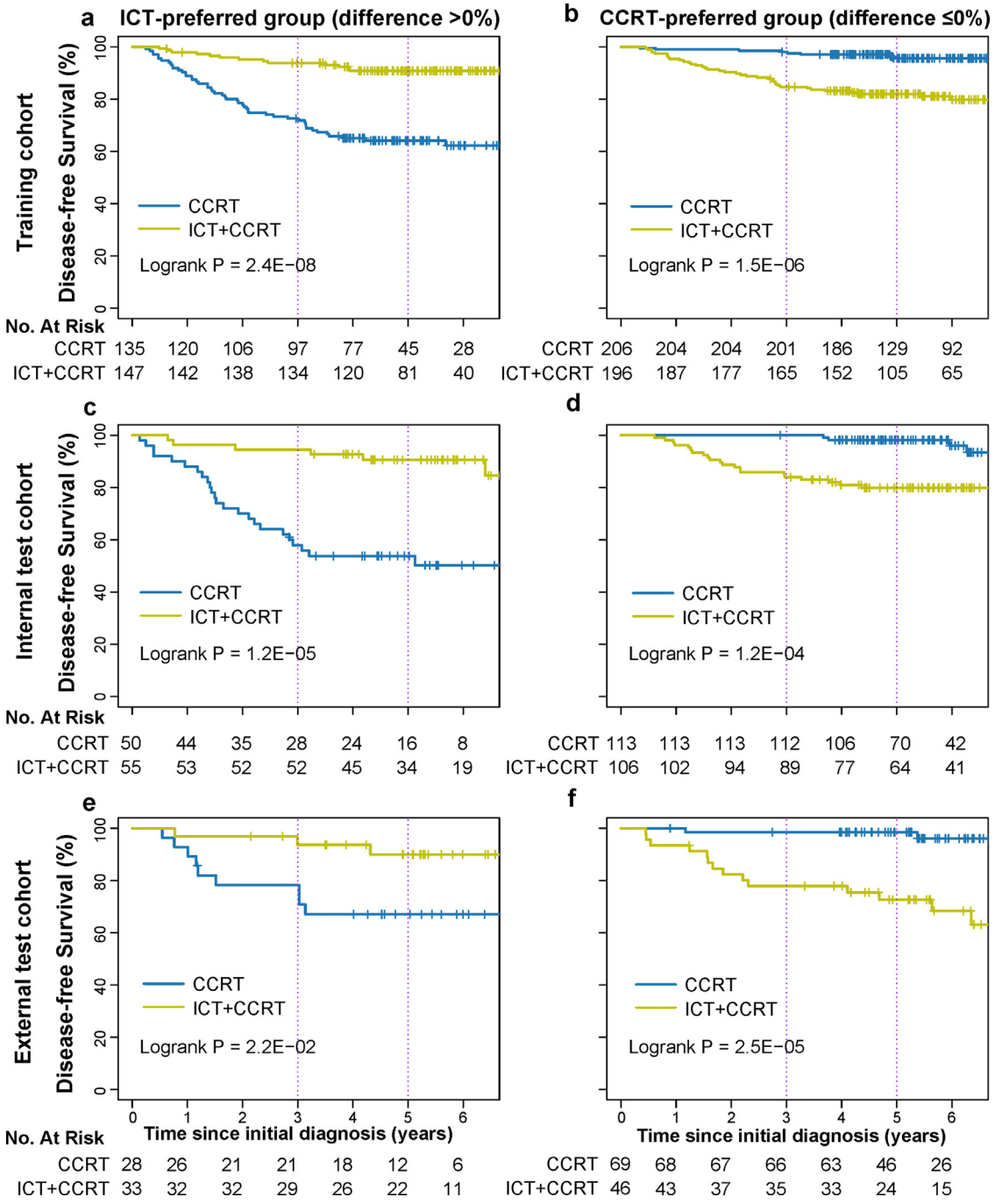

Fig. 5. Kaplan-Meier curves of disease-free survival according to dichotomized 5-year prognostic difference in all cohorts (n=1206). **(a)** Comparisons of the ICT+CCRT and CCRT groups in the ICT-preferred group (difference >0%), and **(b)** the CCRT-preferred group (difference ≤0%) in the training cohort. **(c)** Comparisons of the ICT+CCRT and CCRT groups in the ICT-preferred (difference >0%) group, and **(d)** CCRT-preferred group (difference ≤0%) in the internal test cohort. **(e)** Comparisons of the ICT+CCRT and CCRT groups in the ICT-preferred group (difference >0%), and **(f)** CCRT-preferred group (difference ≤0%) in the external test cohort. CCRT, concurrent chemoradiotherapy; ICT, induction chemotherapy.

developed on different treatment-specific groups. These findings indicate that the proposed model may serve as a tool for promoting personalised treatment and optimizing the management of NPC patients. We have developed a user-friendly web browser-based tool for our CPTDN on an open-access website (http://www.radiomics.net.cn/post/135) to facilitate its validation and application.

Previous radiomics/deep learning studies in NPC separately explored the prognosis value or predictive value of clinical factors/models. Although some studies found that imaging-based prognostic signatures were predictive of treatment outcomes, they did not take treatment factor into account in the development of prognostic models [16,25]. Given the fact that treatment regimen may largely affect clinical outcomes, it is necessary to consider treatment factor when analysing prognosis. Some tailored predictive signatures could identify the patients who would benefit most from adjuvant therapy or induction therapy, but they could not give the individualized prognostic evaluation for a patient receiving the specific treatment regimen [10,18]. Therefore, these studies could not predict the prognosis of patients with different treatment regimen. In contrast, Our CPTDN incorporating prognostic and predictive factors could provide individualized evaluation of DFS in different treatment regimen and accordingly recommend optimal treatment regimens, CCRT alone or

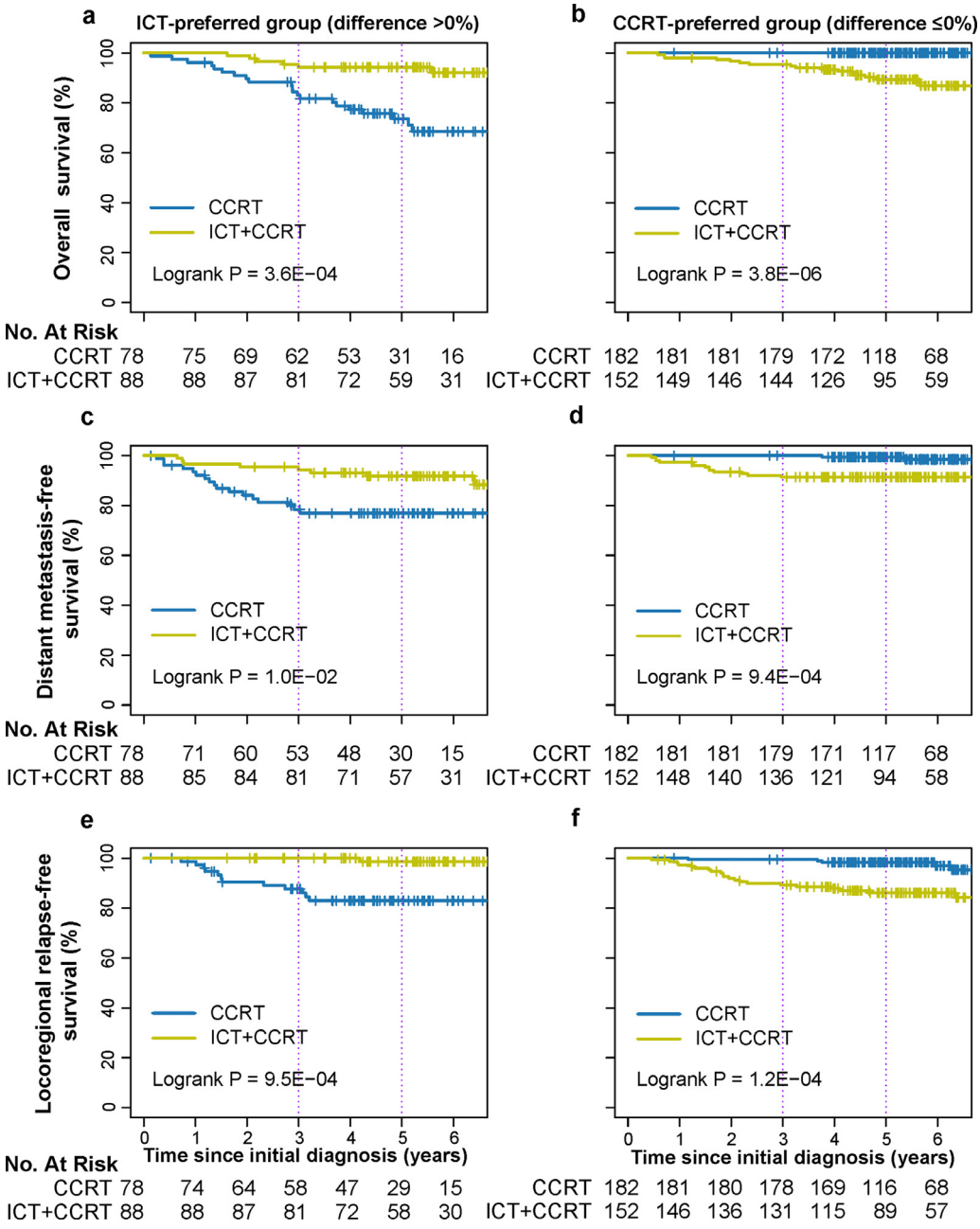

Fig. 6. Kaplan-Meier curves for second clinical endpoints according to dichotomized 5-year prognostic difference in the combined internal and external test cohort (n=500). **(a)** Comparisons of overall survival between the ICT+CCRT and CCRT groups in the ICT-preferred group (difference >0%), and **(b)** CCRT-preferred group (difference ≤0%). **(c)** Comparisons of distant metastasis-free survival between the ICT+CCRT and CCRT groups in the ICT-preferred (difference >0%) group, and **(d)** CCRT-preferred group (difference ≤0%). **(e)** Comparisons of locoregional relapse-free survival between the ICT+CCRT and CCRT groups in the ICT-preferred group (difference >0%), and **(f)** CCRT-preferred group (difference ≤0%). CCRT, concurrent chemoradiotherapy; ICT, induction chemotherapy.

ICT+CCRT, to NPC patients. Our CPTDN remained good prognostic and predictive ability for second clinical endpoints. Moreover, the prognostic and predictive stability of our CPTDN was further validated in different subgroups stratified by clinical factors and MR acquisition parameters. To the best of our knowledge, this is the first multi-task deep learning study aimed at simultaneously predicting prognosis and recommending treatment regimens for NPC patients.

In our study, we found that Prognostic-score and Predictive-score showed a large incremental value in prognostic prediction of DFS, which validated the prognostic and predictive value of multi-task deep-learning radiomics and emphasized the necessity of considering treatment factor in prognostic prediction. Compared the single-task radiomic model (C-index: 0.759) [17], the multi-task radiomic model integrating Prognostic-score and Predictive-score obtained better prognostic results (C-index: 0.836 in the combined internal and external test cohort). To provide insights into what the multi-task deep-learning model learned, we performed gradCAM analysis [26] on patients' MR images (**Fig. D7**). This analysis revealed that the primary tumour regions of the MR images are most responsible for the Prognostic-score in high-risk patients, whereas the regions adjacent to the primary tumour are most responsible for this signature in low-risk patients. As past studies indicate, both intra-tumoral and peritumoral regions offer information valuable for prognosis [27]. In the areas surrounding the tumour mass, immune signatures, such as lymphangiogenic activity and lymphocytic infiltration can be detected [28]. In high-risk patients with poor survival, the primary

Table 1
Baseline characteristics in the training, internal validation, and external validation cohorts.

Characteristics	Training (N = 684)	Internal test (N = 324)	External test (N = 198)	*p*-value
Age (years), median (range)	42 (14-72)	41 (15-65)	43 (13-72)	0.33
Sex, No. (%)				0.63
Male	464 (67.8)	229 (70.7)	134 (67.7)	
Female	220 (32.2)	95 (29.3)	64 (32.3)	
Haemoglobin[1] (g/L), No. (%)				
Normal	599 (87.6)	283 (87.3)	152 (86.4)	0.91
Abnormal	85 (12.4)	41 (12.7)	24 (13.6)	
Unknown	0	0	22	
Family history of cancer, No. (%)				0.72
Yes	189 (27.6)	82 (25.3)	44 (27.7)	
No	495 (72.4)	242 (74.7)	115 (72.3)	
Unknown	0	0	39	
Smoking, No. (%)				0.73
Yes	205 (30.0)	105 (32.4)	50 (31.4)	
No	479 (70.0)	219 (67.6)	109 (68.6)	
Unknown	0	0	39	
Drinking, No. (%)				0.33
Yes	72 (10.5)	40 (12.3)	23 (14.5)	
No	612 (89.5)	284 (87.7)	136 (85.5)	
Unknown	0	0	39	
WHO pathology type, No. (%)				0.54
I-II	7 (1.0)	2 (0.6)	3 (1.6)	
III	677 (99.0)	322 (99.4)	179 (98.4)	
Unknown	0	0	16	
Albumin (g/L), No. (%)				0.88
Normal (40-55)	630 (92.1)	301 (92.9)	146 (91.8)	
Abnormal (other)	54 (7.9)	23 (7.1)	13 (8.2)	
Unknown	0	0	39	
C-reaction protein (mg/L), No. (%)				0.97
Normal ($\leq$3)	505 (73.8)	238 (73.5)	116 (73.0)	
Abnormal (other)	179 (26.2)	86 (26.5)	43 (27.0)	
Unknown	0	0	39	
Lactate dehydrogenase (U/L), No. (%)				0.52
Normal (120-250)	634 (92.7)	304 (93.8)	152 (91.0)	
Abnormal (other)	50 (7.3)	20 (6.2)	15 (9.0)	
Unknown	0	0	31	
Treatment				1.00
CCRT	341 (49.9)	163 (50.3)	99 (50.0)	
ICT+CCRT	343 (50.1)	161 (49.7)	99 (50.0)	
pre-EBV DNA (copy/mL), No. (%)				0.0087
$<$ 4000	427 (62.4)	212 (65.4)	91 (51.7)	
$\geq$ 4000	257 (37.6)	112 (34.6)	85 (48.3)	
Unknown	0	0	22	
Tumour volume (mL), mean	29.2	28.6	31.4	0.88

[1] For males, normal haemoglobin levels are in the range of 130-175 g/L; the equivalent values for females are in the range of 115-150 g/L.
WHO, World Health Organization; pre-EBV DNA, pre-treatment plasma Epstein-Barr virus DNA.

tumour tends to be more aggressive and the peritumoral immune response rate tends to be low. In contrast, in low-risk patients with good survival, the primary tumour is less aggressive and the peritumour immune response rate is high. This may explain our findings. When evaluating the Predictive-score, the multi-task deep-learning model focused on the primary tumour regions of the CCRT-preferred group and on the peritumoral regions of the ICT-preferred group. This differentiation may be due to the fact that CCRT-preferred patients tend to have vessel malfunction, which may cause hypoxia inside the tumour and drive chemoresistance [29,30]. However, in ICT-preferred patients, there are more functional vessels around the tumour that guarantee effective blood supply and drug delivery [29].

This study has three main strengths. First, this multicentre study offers a promising start for developing more comprehensive prognostic model with personalized treatment recommendations for NPC. The present study design and methods can be applied to treatment decisions and prognostic prediction in other cancer types. Second, target delineation in radiotherapy ensures the repeatability of the extracted features, as the feature extraction network uses tumour-centred patches as input, which improves the convenience and feasibility of our model in clinical practice. Finally, our study validated the incremental value of multi-task learning in prognosis prediction, paving the path for the exploration of multi-task learning in fields, such as tumour staging and disease surveillance.

Nevertheless, this study has some limitations, which should be considered when interpreting its findings. First, only patients with stage T3N1M0 NPC were included in this study; future studies should consider patients with other disease stages. Second, CCRT plus adjuvant chemotherapy is also an optional treatment regimen in the advanced NPC [31], and the response to CCRT plus adjuvant chemotherapy remains to be investigated. Third, a nonparametric matching method was used to balance baseline characteristics between treatments in the retrospective study; however, selection bias between the two treatment regimens likely remained and may have affected the presented estimates; therefore, validation of the present findings in prospective trials is required.

In conclusion, we developed and validated a multi-task nomogram that could noninvasively predict the prognosis of T3N1M0 NPC patients with different treatment regimens and accordingly recommend an optimal treatment regimen in the multicentre cohorts. Our nomogram may act as a non-invasive and useful tool for promoting personalized treatment and optimizing management of NPC patients.

Declaration of Competing Interest

The authors have no conflicts of interest to disclose.

Contributors

Conceptualization and study design: Jie Tian, Linglong Tang, Jun Ma, Zhenhua Hu, Lianzhen Zhong, Di Dong, Xueliang Fang, Fan Zhang, and Ning Zhang.

Data verification: Jie Tian, Linglong Tang, Jun Ma, Fan Zhang, Ning Zhang, and Wei Jiang.

Funding acquisition: Jie Tian, Zhenhua Hu, and Di Dong.

Project administration: Jie Tian, Linglong Tang, Jun Ma, and Zhenhua Hu.

Supervision: Jie Tian, Linglong Tang, Jun Ma, Zhenhua Hu, and Di Dong.

Methodology: Lianzhen Zhong, Di Dong, Xueliang Fang, Liwen Zhang, and Mengjie Fang.

Investigation: Lianzhen Zhong, Di Dong, Xueliang Fang, Fan Zhang, and Ning Zhang.

Validation: Liwen Zhang, Mengjie Fang, Wei Jiang, Shaobo Liang, Cong Li, Yujia Liu, Xun Zhao, Runnan Cao, and Hong Shan.

Original draft, and writing: all authors.

Final approval of the manuscript: all authors.

Acknowledgments

This work was supported by the National Key R&D Program of China (2017YFC1309100 and 2017YFA0205200), National Natural Science Foundation of China (82022036, 91959130, 81971776, 81771924, 6202790004, 81930053, and 81671759); the Beijing Natural Science Foundation (L182061 and JQ19027); Strategic Priority Research Program of Chinese Academy of Sciences (XDB38040200); Chinese Academy of Sciences (GJJSTD20170004 and QYZDJ-SSW-JSC005); the Project of High-Level Talents Team Introduction in Zhuhai City (HLHPTP201703 Zhuhai); Beijing Nova Program (Z181100006218046); and the Youth Innovation Promotion Association CAS (2017175).

Data sharing statement

Subject to the institutional review boards' ethical approval, the extracted deep learning-based radiomic features and deidentified clinical characteristics including follow-up information and treatment regimens have been public on GitHub (https://github.com/Zlz-shoulder/CPTDN). The raw MR images can be obtained after asking for the corresponding authors and clarifying purpose of use. Moreover, we have included all the codes used for feature extraction, model development and validation on GitHub. All experiments and implementation details are described thoroughly in the Methods section and appendix C so they can be independently replicated with non-proprietary libraries.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2021.103522.

References

[1] De Martel C, Georges D, Bray F, Ferlay J, Clifford GM. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. Lancet Global Health 2020;8(2):e180–e90.
[2] Tang L-L, Chen Y-P, Mao Y-P, Wang Z-X, Guo R, Chen L, et al. Validation of the 8th edition of the UICC/AJCC staging system for nasopharyngeal carcinoma from endemic areas in the intensity-modulated radiotherapy era. J Nat Comprehens Cancer Network. 2017;15(7):913–9.
[3] Pfister DG, Spencer S, Adelstein D, Adkins D, Anzai Y, Brizel DM, et al. Head and neck cancers, version 2.2020, NCCN clinical practice guidelines in oncology. J Nat Comprehens Cancer Network 2020;18(7):873–98.
[4] Chan A, Grégoire V, Lefebvre J-L, Licitra L, Hui EP, Leung S, et al. Nasopharyngeal cancer: EHNS-ESMO-ESTRO clinical practice guidelines for diagnosis, treatment andfollow-up. Ann Oncol 2012;23:vii83–vii5.
[5] Chen Y-P, Chan ATC, Le Q-T, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. Lancet North Am Ed 2019;394(10192):64–80.
[6] 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. Lancet Oncol 2018;19(3):382–93.
[7] Liang Y, Li J, Li Q, Tang L, Chen L, Mao Y, et al. Plasma protein-based signature predicts distant metastasis and induction chemotherapy benefit in Nasopharyngeal Carcinoma. Theranostics 2020;10(21):9767.
[8] Wang Y-Q, Chen L, Mao Y-P, Li Y-Q, Jiang W, Xu S-Y, et al. Prognostic value of immune score in nasopharyngeal carcinoma using digital pathology. J Immunother Cancer 2020;8(2).
[9] Dowsett M, Turner N. Estimating risk of recurrence for early breast cancer: integrating clinical and genomic risk. J Clin Oncol 2019;37:689–97.
[10] Hui E, Li W, Ma B, Lam W, Chan K, Mo F, et al. Integrating post-radiotherapy plasma Epstein-Barr virus DNA and TNM stage for risk stratification of nasopharyngeal carcinoma to adjuvant therapy. Ann Oncol 2020;31(6):769–79.
[11] 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. Ann Oncol 2019;30(3):431–8.
[12] Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. CA Cancer J Clin 2019;69(2):127–57.
[13] Dong D, Fang M-J, Tang L, Shan X-H, Gao J-B, Giganti F, et al. Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multi-center study. Ann Oncol 2020;31(7):912–20.
[14] Wang X, Li Q, Cai J, Wang W, Xu P, Zhang Y, et al. Predicting the invasiveness of lung adenocarcinomas appearing as ground-glass nodule on CT scan using multi-task learning and deep radiomics. Transl Lung Cancer Res 2020;9(4):1397.
[15] Zhang L-W, Dong D, Zhang W, Hao X, Fang M, Wang S, et al. A deep learning risk prediction model for overall survival in patients with gastric cancer: a multicenter study. Radiother Oncol 2020;150:73–80.
[16] Peng H, Dong D, Fang M-J, 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. Clin Cancer Res 2019;25(14):4271–9.
[17] Zhong L-Z, Fang X-L, Dong D, Peng H, Fang M-J, Huang C-L, et al. A deep learning MR-based radiomic nomogram may predict survival for nasopharyngeal carcinoma patients with stage T3N1M0. Radiother Oncol 2020;151:1–9.
[18] 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 Med 2019;17(1):190.
[19] Stuart EA, King G, Imai K, Ho D. MatchIt: nonparametric preprocessing for parametric causal inference. J Statistical Software 2011;42(8).
[20] American Joint Committee on Cancer.. American Joint Committee on Cancer TNM staging manual. In: Cancer AJCo, editor. American Joint Committee on Cancer TNM staging manual. 8th Ed New York: Springer; 2017.
[21] 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. Biomed Eng Online 2015;14(1):1–17.
[22] Squeeze-and-excitation networks. In: Hu J, Shen L, Sun G, editors. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018.
[23] Campanella G, Hanna MG, Geneslaw L, Miraflor A, Silva VWK, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med 2019;25(8):1301–9.
[24] 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 Med Res Method 2018;18(1):24.
[25] Jiang Y, Jin C, Yu H, Wu J, Chen C, Yuan Q, et al. Development and validation of a deep learning ct signature to predict survival and chemotherapy benefit in gastric cancer: a multicenter, retrospective study. Ann Surg 2020;31(6):760–8.
[26] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vision 2020;128(2):336–59.
[27] Braman NM, Etesami M, Prasanna P, Dubchuk C, Gilmore H, Tiwari P, et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. Breast Cancer Res 2017;19(1):1–14.
[28] Ocaña A, Diez-González L, Adrover E, Fernández-Aramburo A, Pandiella A, Amir E. Tumor-infiltrating lymphocytes in breast cancer: ready for prime time? J Clin Oncol 2015;33(11):1298–9.
[29] Carmeliet P, Jain RK. Principles and mechanisms of vessel normalization for cancer and other angiogenic diseases. Nat Rev Drug Discovery 2011;10(6):417–27.
[30] Doktorova H, Hrabeta J, Khalil MA, Eckschlager T. Hypoxia-induced chemoresistance in cancer cells: the role of not only HIF-1. Biomed Papers Med Faculty Palacky Univ Olomouc 2015;159(2).
[31] Setakornnukul J, Thephamongkhol K. Neoadjuvant chemotherapy followed by concurrent chemoradiotherapy versus concurrent chemoradiotherapy followed by adjuvant chemotherapy in locally advanced nasopharyngeal carcinoma. BMC Cancer 2018;18(1):1–8.