Ther Adv Med Oncol

2020, Vol. 12: 1–12

DOI: 10.1177/1758835920971416

A deep-learning-based prognostic nomogram integrating microscopic digital pathology and macroscopic magnetic resonance images in nasopharyngeal carcinoma: a multi-cohort study

Fan Zhang*, Lian-Zhen Zhong*, Xun Zhao*, Di Dong*, Ji-Jin Yao, Si-Yang Wang, Ye Liu, Ding Zhu, Yin Wang, Guo-Jie Wang, Yi-Ming Wang, Dan Li, Jiang Wei, Jie Tian and Hong Shan

Abstract

Background: To explore the prognostic value of radiomics-based and digital pathology-based imaging biomarkers from macroscopic magnetic resonance imaging (MRI) and microscopic whole-slide images for patients with nasopharyngeal carcinoma (NPC).

Methods: We recruited 220 NPC patients and divided them into training ($n = 132$), internal test ($n = 44$), and external test ($n = 44$) cohorts. The primary endpoint was failure-free survival (FFS). Radiomic features were extracted from pretreatment MRI and selected and integrated into a radiomic signature. The histopathological signature was extracted from whole-slide images of biopsy specimens using an end-to-end deep-learning method. Incorporating two signatures and independent clinical factors, a multi-scale nomogram was constructed. We also tested the correlation between the key imaging features and genetic alternations in an independent cohort of 16 patients (biological test cohort).

Results: Both radiomic and histopathologic signatures presented significant associations with treatment failure in the three cohorts (C-index: 0.689–0.779, all $p < 0.050$). The multi-scale nomogram showed a consistent significant improvement for predicting treatment failure compared with the clinical model in the training (C-index: 0.817 *versus* 0.730, $p < 0.050$), internal test (C-index: 0.828 *versus* 0.602, $p < 0.050$) and external test (C-index: 0.834 *versus* 0.679, $p < 0.050$) cohorts. Furthermore, patients were stratified successfully into two groups with distinguishable prognosis (log-rank $p < 0.0010$) using our nomogram. We also found that two texture features were related to the genetic alternations of chromatin remodeling pathways in another independent cohort.

Conclusion: The multi-scale imaging features showed a complementary value in prognostic prediction and may improve individualized treatment in NPC.

Keywords: digital pathology, multi-scale features, nasopharyngeal carcinoma, radiomics, survival analysis

Received: 18 July 2020; revised manuscript accepted: 14 October 2020.

Correspondence to:
Wei Jiang
Department of Radiation Oncology, Affiliated Hospital of Guilin Medical University, Guilin, Guangxi Province 541000, P. R. China
weijiang@glmc.edu.cn

Jie Tian
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, Beijing, 100190, P. R. China

Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine, Beihang University, Beijing, China
jie.tian@ia.ac.cn

Hong Shan
Guangdong Provincial Key Laboratory of Biomedical Imaging, The Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, Guangdong Province 519000, P. R. China

Department of Interventional Medicine, The Fifth Affiliated Hospital Sun Yat-sen University, No.52 Meihua East Road, Xiangzhou District, Zhuhai, Guangdong Province 519000, P. R. China
shanhong@mail.sysu.edu.cn

Fan Zhang
Ji-Jin Yao
Department of Head and Neck Oncology, The Cancer Center of the Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, Guangdong Province, P. R. China

Guangdong Provincial Key Laboratory of Biomedical Imaging, The Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai,

Introduction

Nasopharyngeal carcinoma (NPC) is the most common head and neck cancer, with an extremely uneven distribution of disease burden worldwide. There are about 120,000 new cases worldwide annually, >50% of which are in east and southeast Asia.[1,2] Generally, NPC is a radiosensitive tumor, but with locoregional and distant aggressiveness. Despite the introduction of intensity modulated radiotherapy and chemotherapy,

Guangdong Province, P. R. China

Lian-Zhen Zhong
Xun Zhao
School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, P. R. China

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, Beijing, P. R. China

Di Dong
School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, P. R. China

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, Beijing, P. R. China

Si-Yang Wang
Department of Head and Neck Oncology, The Cancer Center of the Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, Guangdong Province, P. R. China

Ye Liu
Ding Zhu
Department of Pathology, The Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, Guangdong Province, P. R. China

Yin Wang
Guo-Jie Wang
Department of Radiology, The Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, Guangdong Province, P. R. China

Yi-Ming Wang
Dan Li
Guangdong Provincial Key Laboratory of Biomedical Imaging, The Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, Guangdong Province, P. R. China

*These authors contributed equally

30–50% of patients still experience disease relapse after radical radio-chemotherapy.[3,4] Variant prognoses were also found in patients with similar stage, pathologic subtype, and treatment.[5] Based mainly on a manual anatomical and histopathologic qualitative assessment of the extent of disease, the current tumor, nodes, and metastases (TNM) system and histopathologic classification are inadequate to guide research and clinical practice. Therefore, there is an urgent need for a more comprehensive and tailored biomarker.

Radiomics, an emerging field in medicine, applies computer-aided algorithms to extract and analyze quantitative macroscopic features from radiological images, and has achieved successful applications for the clinical diagnosis and treatment especially for malignancies.[6–9] Several prognostic models based on macroscopic radiomic features have been proposed to improve the risk stratification ability for NPC.[10–13] Digital pathology driven artificial intelligence enables the mining of microscopic morphometric phenotypes from digitizing whole-slide images (WSI).[14–16] Histopathologic properties such as histopathologic classification,[17] immunological phenotype [such as tumor-infiltrating lymphocytes (TILs)][18] and handcrafted microscopic features from WSI are found to be independently prognostic for NPC.[19] However, there have been no published reports on the integration of imaging features from radiological images and WSI for the prognostic prediction of patients with cancer including NPC. Therefore, it is necessary and reasonable to integrate histopathologic and radiomic features, as well as the known prognostic factors, to assess the morphological features at multiple scale (cell, tissue, and patient level), thus enabling better characterization of the aggressiveness of disease.[8,20]

Based on the above, we collected pretreatment multi-parametric MRI images and WSI from multiple cohorts. We aimed to develop a deep learning-based model to explore whether factors at different scales could improve the prognostic predictability of NPC, in order to support clinical decision-making for patients with NPC. In another independent cohort, gene-expression data analysis was further performed for biologic exploration.

Methods

Patient selection

The study workflow is displayed in Figure 1. We included retrospectively 176 consecutive patients with NPC between February 2014 and June 2017 in the Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai, China (SYSU5 cohort). Computer-generated random numbers were used to allocate these patients into training ($n = 132$) and internal test ($n = 44$) cohorts. We also enrolled 44 patients between January 2013 and November 2016 in the Guilin Medical University Affiliated Hospital (GMH), Guilin, China (external test cohort). All these patients were (1) newly histologically confirmed and non-metastatic NPC; (2) underwent pretreatment MRI of head and neck region [including the axial T1-weighted (T1WI), T2-weighted fat-suppressed (T2WI FS), contrast-enhanced T1-weighted fat-suppressed (CE T1WI FS) sequences] and biopsy under nasopharyngofiberscope with hematoxylin-eosin (H&E) staining slide before any anti-cancer treatment in the two institutions; (3) without prior or concomitant malignancy; (4) treated with definitive-intent radiotherapy. Exclusion criteria included (1) incomplete pretreatment MRI of head and neck; (2) notable motion artifacts in MRI; (3) less than 1% tumor cells when assessing the histopathological slides evaluated by the two pathologists (D.Z. and Y.L.). This retrospective study was approved by the ethical review boards of SYSU5 (the approval ID SYSU5-K208-1) and GMH (the approval ID GLMU1A2018062). The requirement for informed consent was waived. Detailed information on the treatment is also summarized in Supplemental Text S1.

To investigate the underlying biology of the image features, we retrospectively enrolled another independent cohort between June 2018 and November 2019 in the SYSU5 (biological test cohort). There were 16 NPC patients with available pretreatment data of head and neck MRI and 508 tumor-related genes panel sequencing of primary tumor designed by Beijing Genomics institution (Supplemental Text S2). Of all patients, 81.3% (13/16) had pathological slides available. All 16 patients signed informed consent for the usage of their anonymized clinical and genetics data.

Imaging preparation

Multi-parametric MRI was obtained within 2 weeks before any anti-cancer treatment for each patient. The acquisition details (contrast agents, image thickness, etc.) is displayed in Supplemental Text S3 and Table S1. The axial T1WI, T2WI FS and CE T1WI FS MRI sequences were retrieved from the institutional picture archiving

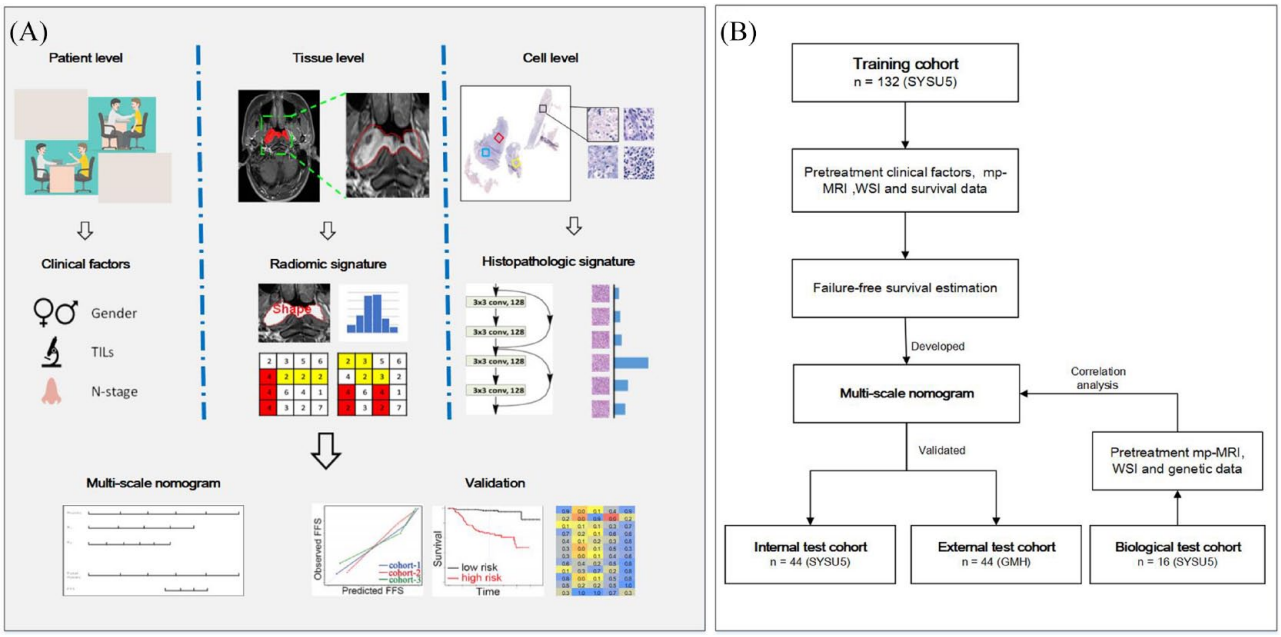

Figure 1. (a) Imaging analysis and (b) data flow of this study. The multi-scale nomogram was developed to predict FFS based on the clinical, radiomic, and histopathologic data from the training cohort, and in the internal and external test cohorts. The biological test cohort was used to test the correlation with the key imaging features and signatures.
FFS, failure-free survival; GMH, Guilin Medical University Affiliated Hospital; MRI, magnetic resonance imaging; SYSU5, Fifth Affiliated Hospital of Sun Yat-sen University; TIL, tumor-infiltrating lymphocytes; WSI, whole-slide images.

and communication system and loaded into ITK-SNAP software (version 2.2.0; www.itk-snap.org) for manually segmenting the primary tumor on the axial MRI slices by two radiologists with 5 years (G.J.W.) and 10 years (Y.W.) experience with MRI.

The diagnostic glass slides for each patient were collected from H&E-stained sections of the formalin-fixed, paraffin-embedded tumor tissue blocks acquired from nasopharyngeal lesion biopsy. Then, we converted these into WSI at 40× objective magnification (0.23 μm/pixel) by a digital slide scanner (Pannoramic 1000, Budapest, Hungary). All the WSI were evaluated by pathologist A (D.Z.) with 5 years experience and who was blinded to patients' treatment and outcomes. A manual assessment included (1) intra-tumor necrosis, (2) 2003 World Health Organization (WHO) pathologic classification, (3) presence of sarcomatoid tumor cells described with poor prognosis in a previous study,[17] and (4) TILs grading according to the criteria proposed by the International Immuno-Oncology Biomarker Working Group and by previous research.[18,21] Representative regions of interest (ROIs, size: 25 × 25 μm) containing (1) tumor area with predominant tumor nuclei, (2) intra-tumoral necrosis (if found), (3) intra-tumoral lymphocytes infiltrating, and (4) stromal lymphocytes infiltrating area were also captured manually for each WSI avoiding artifacts (bubbles, significant tissue section folds, overstaining, understaining) *via* CaseViewer (version 2.12; 3DHISTECH, Budapest, Hungary) by pathologist A. To account for heterogeneity, at least two non-overlapping ROIs of the tumor area were selected from each patient. Assessment and capture were reviewed and confirmed by another pathologist (Y.L.) with 15 years experience.

Imaging features extraction/selection and signatures construction

The radiomic features computation and extraction was performed according to guideline of Image Biomarker Standardisation Initiative.[22] Based on the open platform (PyRadiomics, https://pypi.org/project/pyradiomics/),[23] a total of 2364 radiomic features were extracted from MRI images (788 features for each MRI sequence), including five groups: histogram, shape, gray level dependency matrix (GLDM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM) and gray level co-occurrence matrix (GLCM).

A radiomic signature was developed in the SYSU5 cohort as follows: (1) the intra/inter-class correlation coefficient (ICC) was used to select radiomic features with great reliability (ICC > 0.75) [6]; (2) we employed univariate analysis, minimal redundancy maximum relevance, and random forest to select key features associated with FFS using 3-fold cross validation, which was used to construct CPH models; (3) in order to obtain a robust model, we performed randomly 3-fold cross-validation for 100 times, followed by the repeating construction and validation of the CPH models. We choose the CPH models whose performance on both the training cohort and the holdout test cohort are close to average performance among 100 repetitions. The maximum value of the selected CPHs was set as the radiomic signature (detailed in Supplemental Text S4 and Figure S1).

A histopathologic signature was developed in the training cohort as follows: (1) we standardized the stain color of ROIs of WSI to reduce undesirable staining deviation[24]; (2) a deep convolutional neural network (DCNN), ResNet-18,[25] was used to end-to-end predict FFS based on the standardized ROIs under the multiple instance learning assumption,[26] with DeepSurv as a loss function.[27] We used ImageNet-pretrained ResNet-18 for transfer learning, to get a good starting of weight values. In training process, we adopted label smoothing, early stopping and dropout layers to avoid overfitting. The output value of the model was set to be the histopathologic signature. Training details and network structure are shown in Supplemental Text S4, Table S2 and Figure S2.

Development and validation of an individualized multi-scale nomogram

The primary endpoint was FFS (time to locoregional failure, distant failure, or death from any cause, whichever occurred first). The secondary endpoints included overall survival (OS, time to death from any cause), distant FFS (D-FFS, time to distant failure), and locoregional FFS (LR-FFS, time to local or regional failure or both).

We used the training cohort to explore the association between clinical risk factors and FFS using univariate CPH analysis, and identified independent prognostic factors for predicting FFS using multivariate CPH analysis. Then, model^{CRH} was developed by integrating independent clinical factors, radiomic, and histopathologic signatures. Finally, an individualized multiscale nomogram was constructed for FFS estimation using the above regression coefficients.

The performance of the multiscale nomogram was assessed in the training cohort and tested in the internal and external cohorts using Harrell's concordance index (C-index) and time-independent receiver operating characteristic (TI-ROC) analysis. Comparisons between C-indices were conducted using Student *t* test.[28] The calibration curves and Hosmer-Lemeshow test were used to evaluate the agreement between nomogram predicted FFS and observed FFS. In addition, risk stratification and stratified analysis were conducted to test the discriminability and stability of the multiscale nomogram. The biological basis of the key radiomic features and histopathologic signature was evaluated using Wilcoxon signed-rank test on the biological test cohort.

Statistical analysis

Image preprocessing for MRI and WSI was conducted in MATLAB R2018a (MathWorks, Natick, MA, USA). ResNet-18 survival model were implemented with open-source Python v3.6.5 and Pytorch v1.1.0, and statistical analysis was conducted in open-source R v3.6.1. We used two-sided p-value < 0.050 as the level of statistical significance. A description of the statistical methods is shown in detail in Supplemental Text S6.

Results

Baseline clinic-pathological characteristics

The distribution of host, stage, radiologic, and pathologic characteristics were well balanced between the three cohorts (Table 1), except worse anemia occurred in the external test cohort (61.36% *versus* 21.21–22.73%). Median follow-up was 37.1 months [interquartile range (IQR) 27.5–46.4], 39.1 months (IQR 25.5–47.0), and 37.6 months (IQR 27.5–50.6) in the training, internal, and external test cohorts, respectively. At the last follow up (30 September 2019), we respectively found 32/132 (24%), 11/44 (25%), and 11/44 (25%) patients who had experienced a confirmed treatment failure in these three cohorts ($p = 0.99$).

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

	Training cohort	Internal test cohort	External test cohort	*p* value
	(*n* = 132)	(*n* = 44)	(*n* = 44)	
	No. (%)	No. (%)	No. (%)	
Age (years)				0.095[c]
Median (range)	48 (19–83)	49 (27–78)	44 (24–70)	
Sex				0.89[d]
Male	96 (72.73%)	33 (75.00%)	31 (70.45%)	
Female	36 (27.27%)	11 (25.00%)	13 (29.56%)	
WHO pathological type				0.23[d]
I	0 (0.00%)	0 (0.00%)	0 (0.00%)	
II	7 (5.30%)	1 (2.27%)	0 (0.00%)	
III	125 (94.70%)	43 (97.73%)	44 (100.00%)	
Total stage				0.55[d]
I	1 (0.76%)	1 (2.27%)	0 (0.00%)	
II	19 (14.39%)	6 (13.64%)	4 (9.09%)	
III	74 (56.06%)	21 (47.73%)	21 (47.73%)	
IV	38 (28.79%)	16 (36.36%)	19 (43.18%)	
T stage				0.28[d]
T1	32 (24.24%)	10 (22.73%)	7 (15.91%)	
T2	13 (9.85%)	8 (18.18%)	4 (9.09%)	
T3	69 (52.27%)	22 (50.00%)	22 (50.00%)	
T4	18 (13.64%)	4 (9.09%)	11 (25.00%)	
N stage				0.30[d]
N0	2 (1.52%)	2 (4.55%)	2 (4.55%)	
N1	48 (36.36%)	13 (29.55%)	14 (31.82%)	
N2	59 (44.70%)	15 (34.09%)	16 (36.36%)	
N3	23 (17.42%)	14 (31.82%)	12 (27.27%)	
Anemia				$<$0.0010[d]
Yes	28 (21.21%)	10 (22.73%)	27 (61.36%)	
No	104 (78.79%)	34 (77.27%)	17 (38.64%)	
Intra-tumoral necrosis[a]				0.077[d]
No	104 (78.79%)	34 (77.27%)	41 (93.18%)	
Yes	28 (21.21%)	10 (22.73%)	3 (6.82%)	

(Continued)

Table 1. (Continued)

	Training cohort (*n* = 132) No. (%)	Internal test cohort (*n* = 44) No. (%)	External test cohort (*n* = 44) No. (%)	*p* value
TILs grading[a]				0.44[d]
High	64 (48.49%)	26 (59.09%)	24 (54.55%)	
Low	68 (51.52%)	18 (40.91%)	20 (45.46%)	
Presence of sarcomatoid tumor cells[a]				0.66[d]
No	102 (77.27%)	31 (70.46%)	33 (75.00%)	
Yes	30 (22.73%)	13 (29.55%)	11 (25.00%)	
Lymph node necrosis[b]				0.41[d]
No	94 (71.21%)	32 (72.73%)	27 (61.36%)	
Yes	38 (28.79%)	12 (27.27%)	17 (38.64%)	
The remaining	65 (49.24%)	17 (38.64%)	17 (38.64%)	
pEBV DNA level				NA[e]
<4000	62 (46.97%)	26 (59.09%)	0 (0.00%)	
⩾4000	7 (5.30%)	2 (4.55%)	0 (0.00%)	
Missing data	63 (47.73%)	16 (36.36%)	44 (100.00%)	

[a]Intra-tumoral necrosis, TIL grading, and presence of sarcomatoid tumor cells were manual evaluated by pathologists on WSI.
[b]Lymph node necrosis was manual identified by radiologists on MRI.
[c]*p* values were calculated by Student *t* test.
[d]*p* values were calculated by chi-square test.
[e]*p* values were not calculated because of large amount of missing data.
MRI, magnetic resonance imaging; pEBV DNA, plasma Epstein–Barr Virus DNA; TILs, tumoral infiltrating lymphocytes; WHO, World Health Organization; WSI, whole slide images.

Development and validation of macroscopic radiomic signature

A total of 12 key radiomic features were finally selected after (Supplemental Table S3), and incorporated into the radiomic signature (Supplemental Table S4). The radiomic signature exhibited strong association with FFS in the training cohort [concordance index (C-index): 0.721, 95% confidence interval (CI): 0.630–0.812] that was validated in the internal test (C-index: 0.777, 95% CI: 0.657–0.896) and external test (C-index: 0.711, 95% CI: 0.534–0.889) cohorts.

Development and validation of microscopic histopathologic signature

The histopathologic signature was computed by the DCNN in an end-to-end way. The histopathologic signature yielded a good C-index of 0.741 (95% CI: 0.662–0.819) and remained a statistically significant prognostic factor in the internal test cohort (C-index, 0.779, 95% CI, 0.679–0.880) and external test cohort: C-index, 0.689, 95% CI, 0.552–0.826). For comparison, a Cox proportional hazard (CPH) model (Model^{pathologist}) was constructed according to manual measurements for WSIs by the pathologists. Model^{pathologist} identified sarcomatoid tumor cells and TILs grading as independent variables and obtained a slightly inferior performance than the histopathologic signature in the training cohort (C-index: 0.667, 95% CI: 0.558–0.796), internal test cohort (C-index: 0.737, 95% CI: 0.493–0.982), and external test cohort (C-index: 0.679, 95% CI: 0.411–0.947).

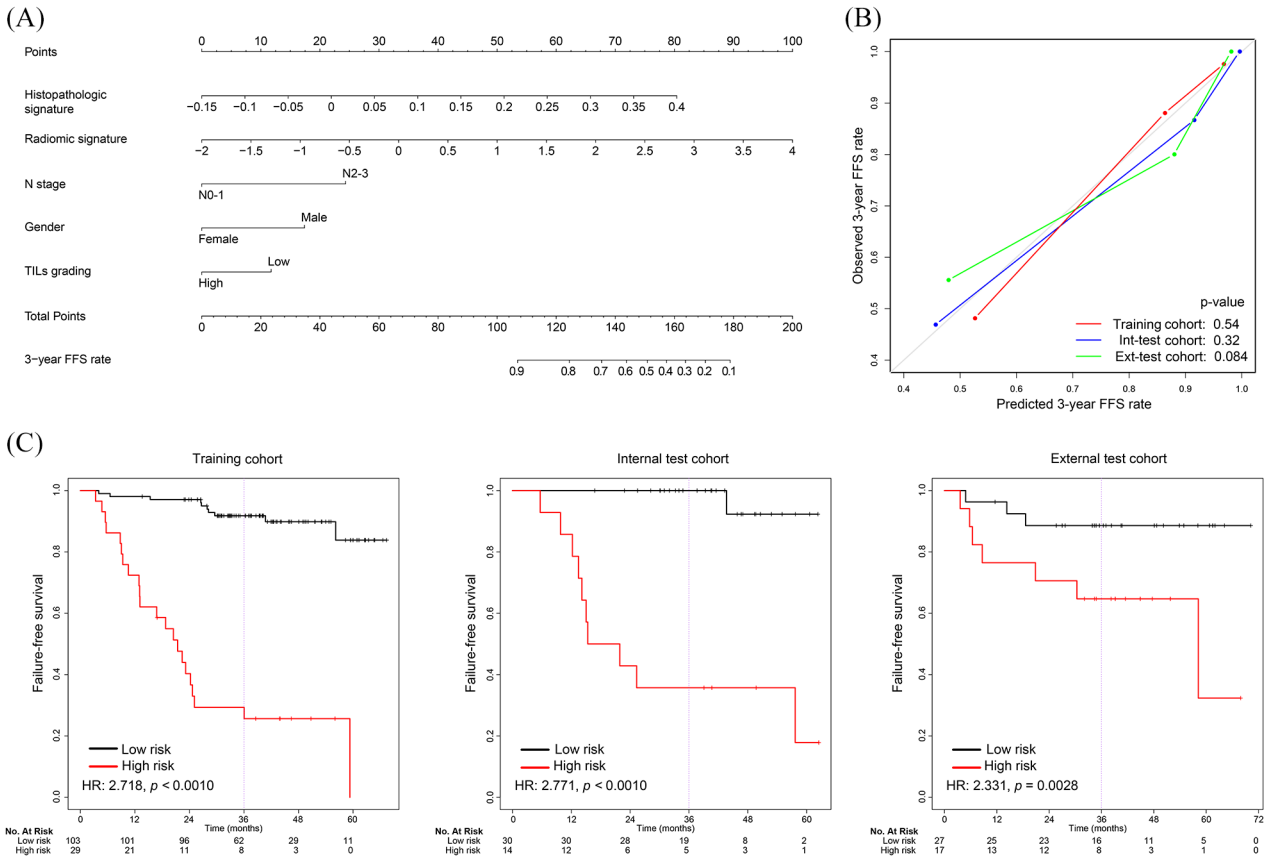

Figure 2. (a) Multiscale nomogram, (b) its calibration curves, and (c) Kaplan–Meier curves in the three cohorts.
Ext-test, external test; FFS, failure-free survival; HR, hazard ratio; Int-test, internal test; TILs, tumor-infiltrating lymphocytes.

Development and validation of the individualized multiscale nomogram

In univariate analysis, clinical parameters in Table 1 were explored, and sex [hazard ratio (HR): 0.34, $p = 0.037$], N-stage (HR: 2.93, $p = 0.013$), and TILs grading (HR: 0.49, $p = 0.049$) were found to be significantly associated with FFS (Table S5). A clinical CPH ($\text{Model}^{\text{clinic}}$) model was built according to sex, N-stage, and TILs grading after adjusting for covariables [C-index 0.730 (95% CI: 0.613–0.847) in the training cohort, 0.602 (95% CI: 0.416–0.788) in the internal test cohort and 0.679 (95% CI: 0.442–0.917) in the internal test cohort]. To integrate the three independent clinical factors, radiomic and histopathologic signature, we built a clinical-radio-histopathologic CPH model ($\text{Model}^{\text{CRH}}$), which was visualized into a multiscale nomogram for clinicians (Figure 2a).

The multiscale nomogram showed the best performance, with a C-index of 0.817 (95% CI: 0.758–0.876) in the training cohort, 0.828 (95% CI: 0.741–0.915) in the internal test cohort, and 0.834 (95% CI: 0.736–0.932) in the external test cohort. Moreover, the multiscale nomogram showed a consistently significant improvement over the TNM staging system, as well as the $\text{Model}^{\text{clinic}}$ in the three cohorts (all $p < 0.050$). The calibration curves of the multiscale nomogram demonstrated good agreement between the nomogram-estimated FFS rate and the observed FFS rate (Figure 2b).

We identified the threshold score of the multiscale nomogram as 0.644 corresponding to a total point of 139.6 in the training cohort. In the training cohort, the high-risk group exhibited short-lived FFS compared with that in the low-risk group (HR 2.718, 95% CI 2.039–3.624, $p < 0.0010$; Figure 2c1). This trend was validated in the two test cohorts [internal test cohort: HR 2.771 (95% CI 1.584–4.850); external test cohort: HR 2.331 (95% CI 1.302–4.174), Figure 2c2–2c3]. Similarly, the multiscale nomogram successfully allowed risk stratification with

Table 2. Performance of models in the three cohorts.

Models	Training cohort		Internal test cohort		External test cohort	
	C-index (95% CI)	*p* value	C-index (95% CI)	*p* value	C-index (95% CI)	*p* value
Histopathologic signature	0.741 (0.662–0.819)	<0.0010	0.779 (0.679–0.880)	<0.0010	0.689 (0.552–0.826)	0.0070
Radiomic signature	0.721 (0.630–0.812)	<0.0010	0.777 (0.657–0.896)	<0.0010	0.711 (0.534–0.889)	0.020
$Model^{TNM}$	0.737 (0.614–0.861)	<0.0010	0.593 (0.358–0.828)	0.44	0.654 (0.411–0.896)	0.22
$Model^{clinic}$	0.730 (0.613–0.847)	<0.0010	0.602 (0.416–0.788)	0.28	0.679 (0.442–0.917)	0.14
$Model^{RH}$	0.801 (0.732–0.870)	<0.0010	0.850 (0.757–0.943)	<0.0010	0.811 (0.698–0.925)	<0.0010
$Model^{CRH}$	0.817 (0.758–0.876)	<0.0010	0.828 (0.741–0.915)	<0.0010	0.834 (0.736–0.932)	<0.0010
$Model^{pathologist}$	0.677 (0.558–0.796)	0.0040	0.737 (0.493–0.982)	0.057	0.679 (0.411–0.947)	0.19

$Model^{TNM}$ was constructed by T-stage and N-stage. $Model^{clinic}$ was constructed by sex, N-stage, TILs grading; $Model^{RH}$ was constructed by radiomic and histopathologic signatures; $Model^{CRH}$, which is the multiscale nomogram, was constructed by radiomic and histopathologic signatures and above three independent clinical factors; $Model^{pathologist}$ was constructed by manual assessments of sarcomatoid tumor cells presence and TILs grading. CI, confidence interval.

secondary endpoints (log-rank: all $p < 0.050$, Supplemental Figure S3). When stratified by age (<52 or ⩾52), T-stage (I–II or III–IV), anemia (yes or no), and overall-stage (I–III or IV); the multiscale nomogram achieved a satisfactory prognostic ability (Supplemental Figure S4). Furthermore, time-dependent receiver operating characteristic (TD-ROC) analysis also verified its superior performance for predicting FFS compared with other models or variables (Supplemental Figure S5). Performance of all models in the three cohorts is shown in Table 2 in detail.

Data on pretreatment plasma Epstein–Barr Virus DNA (pEBV DNA) is available only for 107 patients in the SYSU5 cohort; therefore, we did not include it into the nomogram. However, we found that pEBV DNA slightly improved the performance of the nomogram on the sub-group (C-index from 0.870 to 0.875), but not significantly ($p = 0.74$).

Genetic investigation of the imaging features

To evaluate the biological basis of the imaging features, we further investigated the genetic landscape of the biological test cohort (Supplemental Figure S6) using pathway analysis as described previously.[29] We used a heatmap to display the association of the imaging features with the genetic alternations (Figure 3). Two texture features of radiomics were found to be significantly associated with genetic alternations in chromatin remodeling pathway. No significant association was found between the other radiomic features and histopathologic signature and specific mutations.

Discussion

Based on the hypothesis that treatment outcome is determined by tumoral biology on different spatial scales (e.g., individual, tissue, cell and molecule), we explored currently available factors from physical examination, radiology, pathology, and laboratories in clinics for NPC. Using machine learning methods including deep learning, we developed and validated a multi-scale nomogram for predicting treatment failure of NPC, which consistently exhibited superior performances compared with a clinical model. We also identified significant links between the imaging predictors and mutations of chromatin remodeling pathways in another independent cohort. To the best of our knowledge, this is the first study incorporating multi-scale prognostic features based on digital pathology and radiomics to enhance the prognosis prediction for NPC. Our study also provides a feasible protocol to mine prognostic features from the multi-scale images in the clinic for other malignancies.

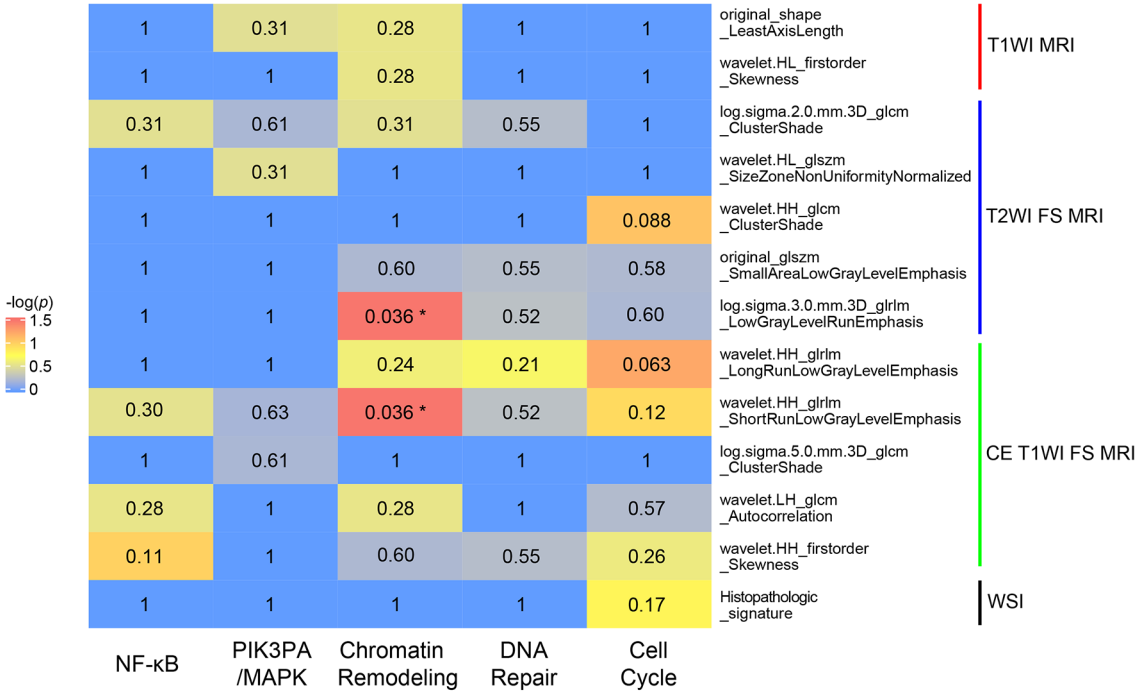

Figure 3. Association between the key imaging features and gene sets. Two texture features were found to be related to genetic alternations of chromatin remodeling pathway mutations (log.sigma.3.0.mm.3D_glrlm_ LowGrayLevelRunEmphasis from T2WI FS sequence and wavelet.HH_glrlm_ShortRunLowGrayLevelEmphasis from CE T1WI FS sequence).

CE-T1WI FS, contrast-enhanced T1-weighted fat-suppressed image; gldm, gray level dependency matrix; glrlm, gray level run length matrix; T1WI, T1-weighted image; T2WI FS, T2-weighted fat-suppressed image; WSI, whole slide images.

At microscopic scale, we used a DCNN to construct the histopathologic signature from WSI, which showed superior performance to the pathologists' assessment (C-index 0.741 *versus* 0.677). A recent study about NPC showed a signature based on handcrafted features from WSI as an independent prognostic factor, with a similar performance (C-index 0.723).[19] In contrast, our histopathologic signature was constructed in an end-to-end method, avoiding incompleteness and instabilities of artificial setting. To interpret the histopathologic signature, we visualized DCNN using the attention map,[30] and found the hotspots located on the tumor cells with multiple nucleoli (Figure 4a), vesicular nuclei (Figure 4b), and spindle shape (Figure 4c, as sarcomatoid tumor cells). These are characteristic appearances of the high proliferation and pleomorphism and of the poor differentiation, and are related to the prognoses, according to Wang's study.[17] In addition to the prognostic morphology of tumor cells, we also found the attention map also focused on the sites of the tumoral infiltrating lymphocytes (Figure 4d). Intriguingly, TIL grading was also included in our nomogram. Given that, we speculated that the proposed nomogram might also involve immune morphology that influenced disease aggressiveness for NPC.[18]

Whereas at macroscopic scale, our model selected a set of key radiomic features from the multi-parametric MRI (Supplemental Table S1), which depict the shape (size and border influenced by tumor growth) and texture (density and perfusion heterogeneity caused by the tumor cell density, inflammation, neovascularization, and necrosis) of the whole tumor tissue.[8] Compared with previous radiomic models,[10] more features from wavelet category and T2WI sequencing were selected by our model, suggesting texture details still require further mining for better reflect disease aggressiveness(e.g., necrosis displayed in T2WI sequencing), which is in accord with findings in other two studies.[11,13] Several typical images were displayed in Figure S7, manifesting different levels of the radiomic signatures in two patients with similar stage and pathological signature, while their outcomes were distinct.

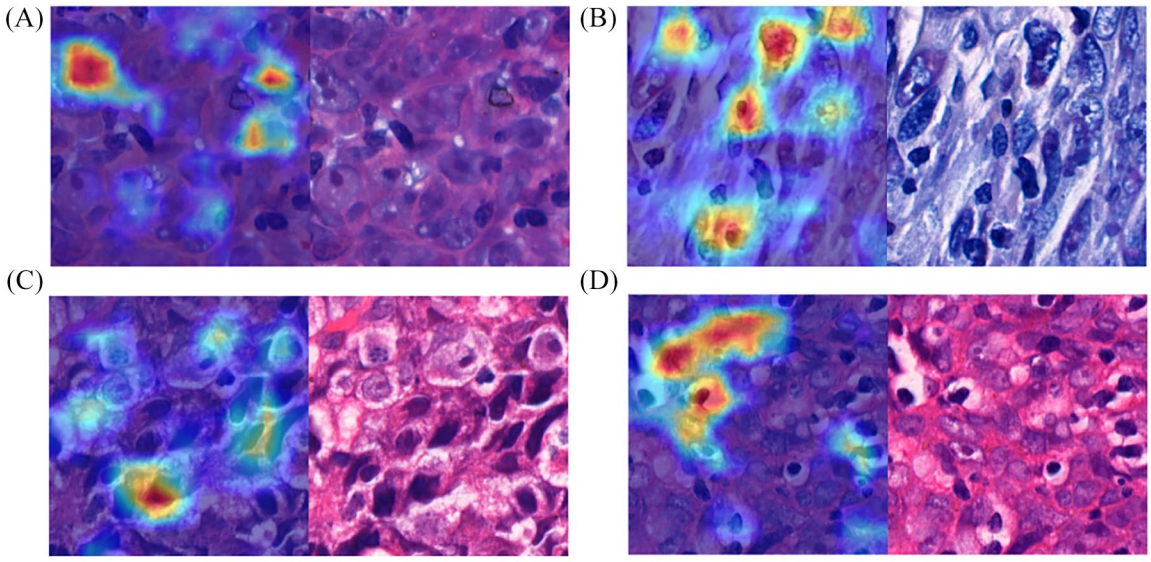

Figure 4. Attention map of the histopathologic signature. The hotspot located at the regions of tumor cells with (a) vesicular nuclei, (b) multiple nucleoli, (c) spindle shape, and (d) TILs.
TILs, tumor-infiltrating lymphocytes.

Impressively, we found that clinic, histopathologic, and radiomic signatures exhibited complementary value in pretreatment individualized prediction of prognoses, with the best performance in pretreatment individualized prediction of prognoses in two test cohorts (Table 2). Compared with previous radiomic and histopathologic models (C-index 0.723–0.761),[10,19] our results also showed an impressive improvement (C-index >0.8 in three cohorts). We believed these improvements might come from the maximum mining of clinic data on different spatial scales. Moreover, the multiscale nomogram maintained good and stable performance in the prediction of the secondary endpoints (Supplemental Figure S3) and in the stratified analysis (Supplemental Figure S4). These findings might potentially improve clinic practice for NPC. For example, patients with stage I–II and high risk might need enhancing treatment (the 3 year FFS lower than 60%, Supplemental Figure S4), while intensive treatment like induction chemotherapy should be re-examined in patients with stage III–IV and low risk (the 3 year FFS higher than 90%, Supplemental Figure S4).

To date, the underlying biology of imaging predictors for NPC have barely been studied bef ore.[10,12,13,19,31] Herein, we performed a preliminary investigation of the association between gene expression and the imaging features of our model. Notably, we identified two texture features related to the genetic alterations of chromatin remodeling pathway, which is regarded as a mark of subclone diversity during tumor evolution, and is correlated to higher mutational burden, EBV burden, and poor prognoses.[29,32] These findings revealed that gene-driven tumor heterogeneity can be captured by texture features *via* radiomic analysis of prognoses. No significant association was found between the other radiomic features and histopathologic signature and specific mutations. The main reason might be the relatively infrequent targetable genetic lesions in NPC (mostly $<10\%$), and, secondly, the limitation of the panel sequencing (only 508 tumor-related genes) and small sample size ($n=16$). For tumors with low mutation rate, further radiogenomics studies should involve more comprehensive gene data, for example, transcriptomics.[33]

Our study has several limitations. First, due to the retrospective nature and relatively small sample size of our data, we were unable to examine the optimal chemotherapy strategy for NPC patients. Future prospective studies with a large sample will be necessary. Secondly, a molecular profile was not included in the multi-scale model. Our sub-analysis showed a trend of complementary value from pEBV DNA. It is needed to fully incorporate molecular profiles like pEBV DNA and genomics. Thirdly, though validated in a multi-cohort setting, the current study was based on subjects that were all Chinese cohorts. Thus,

the applicability of our model needs to be verified in a non-Asian cohort.

Conclusion

By integrating the risk prognostic features from cell to patient level, a multiscale nomogram was established and validated to predict the prognosis of NPC before treatment. Our multi-scale nomogram may serve as a noninvasive, cost-effective, and useful tool for facilitating individualized treatment and future decision-making in NPC.

Author contributions

All authors have read and agree to the published version of the manuscript. Study concept and design, F.Z., L.Z.Z, X.Z., D.D., W.J., J.T., and H.S.; acquisition of data, F.Z., W.J., J.J.Y., S.Y.W., Y.L., D.Z., Y.W., G.J.W., Y.M.W., D.L., J.W.,; analysis and interpretation of data, F.Z., L.Z.Z, X.Z., D.D., J.J.Y., S.Y.W., Y.L., D.Z., Y.W., G.J.W., Y.M.W., D.L.,W.J., J.T., and H.S.; drafting of the manuscript, F.Z., L.Z.Z, X.Z. and D.D.; critical revision of the manuscript for important intellectual content: F.Z., L.Z.Z, X.Z., D.D., W.J., J.T., and H.S.; Supervision, W.J., J.T., and H.S..

Conflict of interest statement

The authors declare that there is no conflict of interest.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by grants from the National Key R&D Program of China (2018YFC0910600, 2017YFA0205200), 2016 Guangdong special support program outstanding talent project, 2017 Zhuhai High-level Health Team Project, the National Natural Science Foundation of China (81620108017, 81901699, 82022036, 91959130, 81971776, 81771924, 81930053), the Beijing Natural Science Foundation (L182061), Strategic Priority Research Program of Chinese Academy of Sciences (XDB 38040200), and the Youth Innovation Promotion Association CAS (2017175). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Supplemental material

Supplemental material for this article is available online.

References

1. Bray F, Ferlay J, Soerjomataram I, *et al*. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; 68: 394–424.
2. Ferlay J, Ervik M, Lam F, *et al*. *Global cancer observatory: cancer today*. Lyon, France: International Agency for Research on Cancer, 2018.
3. Zeng H, Chen W, Zheng R, *et al*. Changing cancer survival in China during 2003–15: a pooled analysis of 17 population-based cancer registries. *Lancet Glob Health* 2018; 6: e555–e567.
4. Chen L, Zhang Y, Lai SZ, *et al*. 10-year results of therapeutic ratio by intensity-modulated radiotherapy versus two-dimensional radiotherapy in patients with nasopharyngeal carcinoma. *Oncologist* 2019; 24: e38–e45.
5. Taheri-Kadkhoda Z, Magnusson B, Svensson M, *et al*. Expression modes and clinical manifestations of latent membrane protein 1, Ki-67, cyclin-B1, and epidermal growth factor receptor in nonendemic nasopharyngeal carcinoma. *Head Neck* 2009; 31: 482–492.
6. Aerts HJ, Velazquez ER, Leijenaar RT, *et al*. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014; 5: 4006.
7. Dong D, Tang L, Li ZY, *et al*. Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer. *Ann Oncol* 2019; 30: 431–438.
8. Grossmann P, Stringfield O, El-Hachem N, *et al*. Defining the biological basis of radiomic phenotypes in lung cancer. *Elife* 2017; 6: e23421.
9. Dong D, Fang MJ, Tang L, *et al*. Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. *Ann Oncol* 2020; 31: 912–920.
10. Zhang B, Tian J, Dong D, *et al*. Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. *Clin Cancer Res* 2017; 23: 4259–4269.
11. Zhang L, Dong D, Li H, *et al*. Development and validation of a magnetic resonance imaging-based model for the prediction of distant metastasis before initial treatment of nasopharyngeal carcinoma: a retrospective cohort study. *EBioMedicine* 2019; 40: 327–335.
12. Dong D, Zhang F, Zhong LZ, *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: 190.

13. Wang G, He L, Yuan C, *et al*. Pretreatment MR imaging radiomics signatures for response prediction to induction chemotherapy in patients with nasopharyngeal carcinoma. *Euro J Radiol* 2018; 98: 100–106.
14. Kather JN, Krisam J, Charoentong P, *et al*. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. *PLoS Med* 2019; 16: e1002730.
15. Bera K, Schalper KA, Rimm DL, *et al*. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol* 2019; 16: 703–715.
16. Mobadersany P, Yousefi S, Amgad M, *et al*. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci U S A* 2018; 115: E2970–E2979.
17. Wang HY, Chang YL, To KF, *et al*. A new prognostic histopathologic classification of nasopharyngeal carcinoma. *Chin J Cancer* 2016; 35: 41.
18. Wang YQ, Chen YP, Zhang Y, *et al*. Prognostic significance of tumor-infiltrating lymphocytes in nondisseminated nasopharyngeal carcinoma: a large-scale cohort study. *Int J Cancer* 2018; 142: 2558–2566.
19. Liu K, Xia W, Qiang M, *et al*. Deep learning pathological microscopic features in endemic nasopharyngeal cancer: prognostic value and protentional role for individual induction chemotherapy. *Cancer Med*. Epub ahead of print 20 December 2019. DOI: 10.1002/cam4.2802.
20. Sanduleanu S, Woodruff HC de Jong EEC,, *et al*. Tracking tumor biology with radiomics: a systematic review utilizing a radiomics quality score. *Radiother Oncol* 2018; 127: 349–360.
21. Salgado R, Denkert C, Demaria S, *et al*. The evaluation of Tumor-Infiltrating Lymphocytes (TILs) in breast cancer: recommendations by an International TILs working group 2014. *Ann Oncol* 2015; 26: 259–271.
22. Zwanenburg A, Leger S, Vallières M, *et al*. Image biomarker standardisation initiative. *arXiv preprint arXiv 1612.07003*, 2016.
23. Van Griethuysen JJ, Fedorov A, Parmar C, *et al*. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 2017; 77: e104–e107.
24. Vahadane A, Peng T, Sethi A, *et al*. Structure-preserving color normalization and sparse stain separation for histological images. *IEEE Trans Med Imaging* 2016; 35: 1962–1971.
25. He K, Zhang X, Ren S, *et al*. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, 27–30 June 2016, pp.770–778.
26. Kraus OZ, Ba JL and Frey BJ. Classifying and segmenting microscopy images with deep multiple instance learning. *Bioinformatics* 2016; 32: i52–i59.
27. Katzman JL, Shaham U, Cloninger A, *et al*. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med Res Methodol* 2018; 18: 24.
28. Haibe-Kains B, Desmedt C, Sotiriou C, *et al*. A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all? *Bioinformatics* 2008; 24: 2200–2208.
29. Lin DC, Meng X, Hazawa M, *et al*. The genomic landscape of nasopharyngeal carcinoma. *Nat Genet* 2014; 46: 866–871.
30. Selvaraju RR, Cogswell M, Das A, *et al*. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 22–29 October 2017, pp.618–626.
31. Peng H, Dong D, Fang MJ, *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: 4271–4279.
32. Tsang CM, Lui VWY, Bruce JP, *et al*. Translational genomics of nasopharyngeal cancer. *Semin Cancer Biol*. Epub ahead of print 12 September 2019. DOI: 10.1016/j.semcancer.2019.09.006.
33. Lee HW, Cho HH, Joung JG, *et al*. Integrative radiogenomics approach for risk assessment of post-operative metastasis in pathological T1 renal cell carcinoma: a pilot retrospective cohort study. *Cancers (Basel)* 2020; 12: 866.