

[1] Fan Zhang#, Lianzhen Zhong#, Xun Zhao#, Di Dong#, Jijin Yao, Siyang Wang, Ye Liu, Ding Zhu, Yin Wang, Guojie Wang, Yiming Wang, Dan Li, Jiang Wei\*, Jie Tian\*, Hong Shan\*, A deep learning-based prognostic nomogram integrating microscopic digital pathology and macroscopic magnetic resonance images in nasopharyngeal carcinoma: a multi-cohort study, *Therapeutic Advances in Medical Oncology*, 2020, DOI: 10.1177/1758835920971416 (SCI IF: 6.852)

**Title Page**

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

**Short title:** Deep learning-based prognostic nomogram in nasopharyngeal carcinoma

Fan Zhang M.D.,<sup>1,2\*</sup> Lian-Zhen Zhong M.S.,<sup>3,4\*</sup> Xun Zhao B.S.,<sup>3,4\*</sup> Di Dong Ph.D.,<sup>4\*</sup> Ji-jin Yao M.D.,<sup>1,2</sup> Si-Yang Wang M.D.,<sup>1</sup> Ye Liu M.D.,<sup>5</sup> Ding Zhu M.D.,<sup>5</sup> Yin Wang M.D.,<sup>6</sup> Guo-Jie Wang M.D.,<sup>6</sup> Yi-Ming Wang Ph.D.,<sup>2</sup> Dan Li Ph.D.,<sup>2</sup> Jiang Wei, M.D.,<sup>7†</sup> Jie Tian Ph.D.,<sup>4†</sup> Hong Shan M.D.<sup>2,8†</sup>

**Authors and affiliations:**

1. Department of head and neck oncology, The cancer center of the Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, Guangdong Province 519000, P. R. China
2. Guangdong Provincial Key Laboratory of Biomedical Imaging, The Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, Guangdong Province 519000, P. R. China.
3. School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, 100049, P. R. China
4. 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
5. Department of Pathology, The Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, Guangdong Province 519000, P. R. China
6. Department of Radiology, The Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, Guangdong Province 519000, P. R. China

1 7. Department of Radiation Oncology, Affiliated Hospital of Guilin Medical University, Guilin, Guangxi Province 541000, P. R.  
2 China

3 8. Department of Interventional Medicine, The Fifth Affiliated Hospital Sun Yat-sen University, Zhuhai, Guangdong Province 519000,  
4 P. R. China.

5 **\*Fan Zhang, Lian-Zhen Zhong, Xun Zhao, and Di Dong contributed equally to this work.**

6 **†Corresponding authors:**

7 **Wei Jiang**, M.D., Department of Radiation Oncology, Affiliated Hospital of Guilin Medical University, Guilin, Guangxi Province  
8 541000, P. R. China

9 **Telephone:** +86- 0773-2867295

10 **E-mail:** weijiang@glmc.edu.cn

11

12 **Jie Tian**, Ph.D, CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, P.  
13 R. China

14 **Telephone:** +86-10-82618465

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

16

17 **Hong Shan**, M.D., Guangdong Provincial Key Laboratory of Biomedical Imaging, Department of Interventional Medicine, The Fifth  
18 Affiliated Hospital, Sun Yat-sen University, Zhuhai, Guangdong Province 519000, P. R. China

19 **Telephone:** +86-0756-2528947

20 **Fax:** +86-0756-2528947

21 **E-mail:** shanhong@mail.sysu.edu.cn

## Abstract

**Background:** To explore the prognostic value of radiomics-based and digital pathology-based imaging biomarkers from macroscopic magnetic resonance images 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 testing ( $n=44$ ), external testing ( $n=44$ ) 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 testing 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 to the clinical model in the training (C-index: 0.817 vs. 0.730,  $p < 0.050$ ), internal testing (C-index: 0.828 vs. 0.602,  $p < 0.050$ ) and external testing (C-index: 0.834 vs. 0.679,  $p < 0.050$ ) cohorts. Furthermore, patients were successfully stratified 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 ( $n = 16$ ).

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

**Keywords:** nasopharyngeal carcinoma; survival analysis; radiomics; digital pathology; multi-scale features.

## 1    **Introduction**

2        Nasopharyngeal carcinoma (NPC) is the most common head and neck cancer with an extremely uneven distribution of disease  
3    burden worldwide, with about 120,000 new cases worldwide, and >50% of which are in the east and southeast Asia.<sup>1, 2</sup> Generally, NPC  
4    is a radiosensitive tumor but with locoregional and distant aggressiveness. Despite the introduction of intensity modulated radiotherapy  
5    and chemotherapy, 30-50% of patients still experience disease relapse after radical radio-chemotherapy.<sup>3, 4</sup> Variant prognoses were also  
6    found in patients with similar stage, pathologic subtype, and treatment.<sup>5</sup> Based mainly on a manual anatomical and histopathologic  
7    qualitative assessment of the extent of disease, the current TNM system and histopathologic classification are inadequate to guide  
8    research and clinical practice. Therefore, it is urgently in need for a more comprehensive and tailored biomarker.

9        Radiomics, an emerging field in medicine, applies computer-aided algorithms to extract and analyze quantitative macroscopic  
10    features from radiological images, and has achieved successful applications for the clinical diagnosis and treatment especially for  
11    malignancies.<sup>6-8</sup> Several prognostic models based on macroscopic radiomic features have been proposed to improve the risk  
12    stratification ability for NPC<sup>9-12</sup>. Digital pathology driven artificial intelligence enables the mining of microscopic morphometric  
13    phenotypes from digitizing whole-slide images (WSI).<sup>13-15</sup> Histopathologic properties such as histopathologic classification,<sup>16</sup>  
14    immunological phenotype (such as tumor-infiltrating lymphocytes, TILs)<sup>17</sup> and handcrafted microscopic features from WSI<sup>18</sup> are found  
15    to be independently prognostic for NPC. However, there have been no published reports on the integration of imaging features from  
16    radiological images and WSI for the prognostic prediction of patients with cancer including NPC. Therefore, it is necessary and  
17    reasonable to integrate histopathologic and radiomic features, as well as the known prognostic factors, to assess the morphological  
18    features at multiple scale (cell, tissue, and patient level), thus enabling better characterization of the aggressiveness of disease.<sup>8, 19</sup>

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

## 23    **Methods**

### 24    **Patient selection**

The study workflow is displayed in **Figure 1**. We retrospectively included 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 a training cohort (n=132) and an internal testing cohort (n=44). We also enrolled 44 patients between January 2013 and November 2016 in the Guilin Medical University Affiliated hospital (GMH), Guilin, China (external testing 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) contained 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 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 **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 testing 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 (**Text S2**). Of all patients, 81.3% (13/16) had pathological slides available. All the 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 **Text S3** and **Table S1**. The axial T1WI, T2WI FS and CE T1WI FS MRI sequences were retrieved from the institutional picture archiving and communication system and loaded into ITK-SNAP software (version 2.2.0; www.itksnap.org) for manually segmenting the primary tumor on the axial MRI slices by the 2 radiologists with 5 years (G.J.W.) and 10 years (Y.W.) of experience with MRI.

The diagnostic glass slides for each patient were collected from hematoxylin-eosin 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 of 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,<sup>16</sup> and 4) TILs grading according to the criteria proposed by the International Immuno-Oncology Biomarker Working Group and by previous researches.<sup>17, 20</sup> 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 manually captured 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 (L.Y.) with 15 years of experience.

#### **Imaging features extraction/selection and signatures construction**

The radiomic features computation and extraction was performed according to guideline of Image Biomarker Standardisation Initiative <sup>21</sup>. Based on the open platform (PyRadiomics, <https://pypi.org/project/pyradiomics/>),<sup>22</sup> a total of 2364 radiomic features were extracted from MR images (788 features for each MR 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 cooccurrence 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);<sup>6</sup> 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 testing cohort are close to average performance among 100 repetitions. The maximum value of the selected CPHs was set as the radiomic signature (detailed in **Text S4 and Figure S1**).

1 A histopathologic signature was developed in the training cohort as follows: 1) we standardized the stain color of ROIs of WSI to  
2 reduce undesirable staining deviation;<sup>23</sup> 2) a deep convolutional neural network (DCNN), ResNet-18,<sup>24</sup> was used to end-to-end predict  
3 FFS based on the standardized ROIs under the multiple instance learning assumption,<sup>25</sup> with DeepSurv as a loss function.<sup>26</sup> We used  
4 ImageNet-pretrained ResNet-18 for transfer learning, to get a good starting of weight values. In training phase, we adopted label  
5 smoothing, early stopping and dropout layers to avoid overfitting. The output value of the model was set to be the histopathologic  
6 signature. Training details and network structure are shown in the **Text S4, Table S2 and Figure S2**.

## 7 **Development and validation of an individualized multi-scale nomogram**

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

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

16 The performance of the multiscale nomogram was assessed in the training cohort and tested in the internal and external cohorts  
17 using Harrell's concordance index (C-index) and time-independent receiver operating characteristic (TI-ROC) analysis. Comparisons  
18 between C-indices were conducted using Student t-test.<sup>27</sup> The calibration curves and Hosmer-Lemeshow test were used to evaluate the  
19 agreement between nomogram predicted FFS and observed FFS. In addition, risk stratification and stratified analysis were conducted to  
20 test the discriminability and stability of the multiscale nomogram. The biological basis of the key radiomic features and histopathologic  
21 signature was evaluated using Wilcoxon signed-rank test on the biological testing cohort.

## 22 **Statistical analysis**

24 Image preprocessing for MRI and WSI was conducted in MATLAB R2018a (MathWorks, Natick, MA, USA). ResNet-18 survival  
25 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 the **Text S6**.

## **Results**

### **Baseline clinic-pathological characteristics**

The distribution of host, stage, radiologic, pathologic characteristics were well balanced between the three cohorts (Table 1), except worse anemia occurred in the external testing cohort (61.36% *vs.* 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 testing cohorts, respectively. At the last follow-up (September 30, 2019), we respectively found 32/132 (24%), 11/44 (25%), and 11/44 (25%) number of patients who had experienced a confirmed treatment failure in these three cohorts ( $p = 0.99$ ).

### **Development and validation of macroscopic radiomic signature**

Twelve key radiomic features were finally selected after (**Table S3**), and incorporated into the radiomic signature (**Table S4**). The radiomic signature exhibited strong association with FFS in the training cohort (concordance index [C-index]: 0.721, 95% CI: 0.630-0.812] that was validated in the internal testing (C-index: 0.777, 95% CI: 0.657-0.896) and external testing (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 testing cohort (C-index, 0.779, 95% CI, 0.679-0.880) and external testing cohort: C-index, 0.689, 95% CI, 0.552-0.826). For comparison, a cox proportional hazard (CPH) model (Model<sup>pathologist</sup>) was constructed according to manual measurements for WSIs by the pathologists. Model<sup>pathologist</sup> 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 testing cohort (C-index: 0.737, 95% CI: 0.493-0.982) and external testing cohort (C-index: 0.679, 95% CI: 0.411-0.947).



## 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 testing cohort and 0.679 [95% CI: 0.442-0.917] in the internal testing 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 testing cohort, and 0.834 (95% CI: 0.736-0.932) in the external testing 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 to 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 testing cohorts (internal testing cohort: HR 2.771 [95% CI 1.584-4.850]; external testing cohort: HR 2.331 [95% CI 1.302-4.174], **Figure 2c2-2c3**). Similarly, the multiscale nomogram successfully allowed risk stratification with secondary endpoints (log-rank: all  $p < 0.050$ , **Figure S3**). When stratified by age ( $< 52$  or  $\geq 52$ ), 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 (**Figure S4**). Furthermore, time-dependent receiver operating characteristic (TD-ROC) analysis also verified its superior performance for predicting FFS compared to other models or variables (**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 only available 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 testing cohort (**Figure S6**) using pathway analysis as described previously<sup>28</sup>. 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 labs in clinic 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 to 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.

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 vs. 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).<sup>18</sup> 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<sup>29</sup> 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.<sup>16</sup> In addition to the prognostic morphology of tumor cells, we also found the attention map also focused on the sites of

1 the tumoral infiltrating lymphocytes (**Figure 4d**). Intriguingly, TIL grading was also included in our nomogram. Given that, we  
2 speculated that the proposed nomogram might also involve immune morphology that influenced disease aggressiveness for NPC.<sup>17</sup>

3 Whereas at macroscopic scale, our model selected a set of key radiomic features from the multi-parametric MRI (Table S1), which  
4 depict the shape (size and border influenced by tumor growth) and texture (density and perfusion heterogeneity caused by the tumor cell  
5 density, inflammation, neovascularization, and necrosis) of the whole tumor tissue.<sup>8</sup> Compared to previous radiomic models,<sup>9</sup> more  
6 features from wavelet category and T2WI sequencing were selected by our model, suggesting texture details still require further mining  
7 for better reflect disease aggressiveness(e.g. necrosis displayed in T2WI sequencing), which is in accord with findings in other two  
8 studies.<sup>10, 12</sup> Several typical images were displayed in **Figure S7**, manifesting different levels of the radiomic signatures in two patients  
9 with similar stage and pathological signature, while their outcomes were distinct.

10 Impressively, we found that clinic, histopathologic and radiomic signatures exhibited complementary value in pretreatment  
11 individualized prediction of prognoses, with the best performance in pretreatment individualized prediction of prognoses in two testing  
12 cohorts (**Table 2**). Compared to previous radiomic<sup>9</sup> and histopathologic<sup>18</sup> models (C-index 0.723-0.761), our results also showed an  
13 impressive improvement (C-index >0.8 in three cohorts). We believed these improvements might come from the maximum mining of  
14 clinic data on different spatial scales. Moreover, the multiscale nomogram maintained good and stable performance in the prediction of  
15 the secondary endpoints (**Figure S3**) and in the stratified analysis (**Figure S4**). These findings might potentially improve clinic practice  
16 for NPC. For example, patients with stage I-II and high risk might need enhancing treatment (the 3 year FFS lower than 60%, **Figure**  
17 **S4**), while intensive treatment like induction chemotherapy should be re-examined in patients with stage III-IV and low risk (the 3 year  
18 FFS higher than 90%, **Figure S4**).

19 To date, the underlying biology of imaging predictors for NPC have sparsely been studied before.<sup>9, 11, 12, 18, 30</sup> Herein, we performed a  
20 preliminary investigation in the association between the gene expression and the imaging features of our model. Notably, we identified  
21 two texture features related to the genetic alterations of chromatin remodeling pathway, which is regarded as a mark of subclones  
22 diversity during tumor evolution and correlated to higher mutational burden, EBV burden and poor prognoses.<sup>28, 31</sup> These findings  
23 revealed that gene-driven tumor heterogeneity can be captured by texture features via radiomic analysis of prognoses. No significant  
24 association was found between the other radiomic features and histopathologic signature and specific mutations. The main reason might  
25 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 tumor with low mutation rate, further radiogenomics studies should involve more comprehensive gene data e.g. transcriptomics.<sup>32</sup> To ?

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, 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 profile 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 in non-Asian cohort needs to be verified.

## 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 of NPC.

## References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018; 68: 394-424.
2. Ferlay J EM, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I. 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. *The Lancet Global Health.* 2018; 6: e555-e67.
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, Mercke C and Bjork-Eriksson T. 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: 482-92.
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: 431-8.
8. Grossmann P, Stringfield O, El-Hachem N, et al. Defining the biological basis of radiomic phenotypes in lung cancer. *Elife.* 2017; 6.
9. 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: 4259-69.
10. 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: 327-35.
11. 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.
12. Wang G, He L, Yuan C, Huang Y, Liu Z and Liang C. Pretreatment MR imaging radiomics signatures for response prediction to induction chemotherapy in patients with nasopharyngeal carcinoma. *Euro J Radiol.* 2018; 98: 100-6.
13. 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.
14. Bera K, Schalper KA, Rimm DL, Velcheti V and Madabhushi A. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. *Nature reviews Clinical oncology.* 2019; 16: 703-15.
15. 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-E9.
16. Wang HY, Chang YL, To KF, et al. A new prognostic histopathologic classification of nasopharyngeal carcinoma. *Chin J Cancer.* 2016; 35: 41.
17. 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: 2558-66.
18. 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.* 2019 Dec 20: 2045-7634
19. 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.
20. 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: 259-71.
21. Zwanenburg A LS, Vallières M., Image biomarker standardisation initiative. *arXiv preprint arXiv.* 2016; 1612.07003.
22. Van Griethuysen JJ, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 2017; 77: e104-e7.
23. 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: 1962-71.
24. He K, Zhang X, Ren S and Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2016, p. 770-8.
25. Kraus OZ, Ba JL and Frey BJ. Classifying and segmenting microscopy images with deep multiple instance learning. *Bioinformatics.* 2016; 32: i52-i9.
26. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T and Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med Res Methodol.* 2018; 18: 24.
27. Haibe-Kains B, Desmedt C, Sotiriou C and 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.
28. Lin DC, Meng X, Hazawa M, et al. The genomic landscape of nasopharyngeal carcinoma. *Nat Genet.* 2014; 46: 866-71.
29. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D and Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision.* 2017, p. 618-26.
30. 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: 4271-9.
31. Tsang CM, Lui VWY, Bruce JP, Pugh TJ and Lo KW. Translational genomics of nasopharyngeal cancer. *Semin Cancer Biol.* 2019 Sep 12: 1096-3650
32. 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.

**Supplementary Materials:**

**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..

**Funding:** 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 (82022036, 81901699, 91959130, 81971776, 81771924, 81930053, 81620108017), 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.

**Acknowledgements:** Not applicable.

**Conflict of interest statements:** The authors declare that they have no competing interests.