Computer Methods and Programs in Biomedicine 219 (2022) 106785

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Deep learning for locally advanced nasopharyngeal carcinoma prognostication based on pre- and post-treatment MRI

Song Li[a,1], Yu-Qin Deng[a,1], Hong-Li Hua[a], Sheng-Lan Li[b], Xi-Xiang Chen[b], Bao-Jun Xie[b], Zhiling Zhu[c], Ruoyun Liu[d], Jin Huang[d,*], Ze-Zhang Tao[a,e,**]

[a] *Department of of Otolaryngology-Head and Neck Surgery, Renmin Hospital of Wuhan University, 238 Jie-Fang Road, Wuhan, Hubei 430060, PR China*
[b] *Department of of Radiology, Renmin Hospital of Wuhan University, 238 Jie-Fang Road, Wuhan, Hubei 430060, PR China*
[c] *Department of of Otolaryngology-Head and Neck Surgery, Tongji Hospital Affiliated to Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei 430030, PR China*
[d] *College of Mathematics and Computer Science, Wuhan Textile University, No.1 Fangzhi road, Wuhan, Hubei 430200, PR China*
[e] *Department of Otolaryngology-Head and Neck Surgery, Central Laboratory, Renmin Hospital of Wuhan University, 238 Jie-Fang Road, Wuhan, Hubei 430060, PR China*

ARTICLE INFO

Article history:
Received 16 January 2021
Revised 7 March 2022
Accepted 29 March 2022

Keywords:
Deep learning
Transfer learning
Nasopharyngeal carcinoma
Prognosis
Post-treatment MRI

ABSTRACT

Purpose: We aimed to predict the prognosis of advanced nasopharyngeal carcinoma (stage III-IVa) using Pre- and Post-treatment MR images based on deep learning (DL).

Methods: A total of 206 patients with primary nasopharyngeal carcinoma who were diagnosed and treated at the Renmin Hospital of Wuhan University between June 2012 and January 2018 were retrospectively selected. A rectangular region of interest (ROI), which included the tumor area, surrounding tissues and organs, was delineated on each Pre- and Post-treatment MR image. Two Inception-Resnet-V2 based transfer learning models, named Pre-model and Post-model, were trained with the Pre-treatment images and the Post-treatment images, respectively. In addition, an ensemble learning model based on the Pre-model and Post-models was established. The three established models were evaluated by receiver operating characteristic curve (ROC), confusion matrix, and Harrell's concordance indices (C-index). High-risk-related gradient-weighted class activation mapping (Grad-CAM) images were developed according to the DL models.

Results: The Pre-model, Post-model, and ensemble model displayed a C-index of 0.717 (95% CI: 0.639 to 0.795), 0.811 (95% CI: 0.745–0.877), 0.830 (95% CI: 0.767–0.893), and AUC of 0.741 (95% CI: 0.584–0.900), 0.806 (95% CI: 0.670–0.942), and 0.842 (95% CI: 0.718–0.967) for the test cohort, respectively. In comparison with the models, the performance of Post-model was better than the performance of Pre-model, which indicated the importance of Post-treatment images for prognosis prediction. All three DL models performed better than the TNM staging system (0.723, 95% CI: 0.567–0.879). The captured features presented on Grad-CAM images suggested that the areas around the tumor and lymph nodes were related to the prognosis of the tumor.

Conclusions: The three established DL models based on Pre- and Post-treatment MR images have a better performance than TNM staging. Post-treatment MR images are of great significance for prognosis prediction and could contribute to clinical decision-making.

Abbreviations: NPC, nasopharyngeal carcinoma; MR, magnetic resonance; DL, deep learning; ROI, region of interest; ROC, receiver operating characteristic; C-index, Harrell's concordance indices; AI, artificial intelligence; PFS, progression-free survival; AUC, area under the curves; CI, confidence interval; CNN, convolution neural network; NCCN, the National Comprehensive Cancer Network; RECIST, the Response Evaluation Criteria in Solid Tumors; CT, computed tomography; PET, positron emission tomography.

[*] Corresponding author.

[**] Corresponding author at: Department of of Otolaryngology-Head and Neck Surgery, Renmin Hospital of Wuhan University, 238 Jie-Fang Road, Wuhan, Hubei 430060, PR China.

E-mail addresses: jhuang@wtu.edu.cn (J. Huang), taozezhang@hotmail.com (Z.-Z. Tao).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.cmpb.2022.106785

1. Introduction

Nasopharyngeal carcinoma (NPC) is a malignant tumor that originates from the epithelium of the nasopharynx and has the highest incidence reported in Southeast Asia [1]. Radiotherapy for early NPC and concurrent chemoradiotherapy for advanced NPC is recommended by the National Comprehensive Cancer Network (NCCN) [2]. As a benefit from the improvements in radiotherapy technology and equipment, the overall 5-year survival rate has improved significantly [3,4]. The common TNM staging system, which combines evidence-based findings with empirical knowledge, is adopted for the prognosis assessment of most tumor types [5,6]. However, it is inevitable that accuracy is sacrificed to make the criterion simple and intuitive.

Artificial intelligence (AI) has been rapidly applied to the field of medicine, especially to medical image processing in recent years [7,8]. Many studies have introduced AI to predict the prognosis of patients with cancer and have achieved remarkable results [9–11]. Based on past achievements in this field, more AI-based studies on tumor prognosis using different methods have been reported. Several studies have focused on evaluating the prognosis of NPC [12–16] and predicting the response of NPC to induction chemotherapy [17–19] which is related to our topic. However, the progression of many studies included a step of manual primary tumor segmentation based on radiomics. This manual segmentation, which removes the "noise" from images, however, is complex and inconvenient for practical applications [16], and more importantly, it is questionable that much valuable information could also be removed inadvertently. The standard of T staging in the TNM staging system of NPC only contains information on the relationship between the tumor and the surrounding tissues and organs, not including intertumoral information, and it performs well for prognosis prediction [12,13,16,20]. Therefore, the signals of organs and tissues bordering tumors that are lost by the manual segmentation approach are of great value in predicting tumor prognosis. Furthermore, it is not reliable to predict the prognosis of patients with cervical lymph node metastasis as the main manifestation based on primary tumor segmentation, while the primary tumor is limited or even not presented on MR images. Despite the great advantages of primary tumor segmentation in noise reduction, the removal of valuable information makes it insufficient. Due to the disadvantage of relying on expert knowledge and a predefined algorithm, traditional radiomics that is based on precise segmentation was considered to be gradually replaced by DL [6,21]. However, DL models for MRI classification face a common problem: they are typically limited by low sample size [22]. Transfer learning, which improves learning a new task through the transfer of knowledge from a related task that has already been learned [23,24], is an ideal solution for this problem and was applied to our study.

The purpose of evaluating tumor prognosis is to guide oncologists to formulate a more reasonable and individualized treatment plan. For advanced NPC or other malignant tumors that require multiple courses of treatment, assessment of post-treatment tumor risk is crucial for the improvement of outcome. The Response Evaluation Criteria in Solid Tumors (RECIST) is the common method to evaluate the response of NPC to treatment, which categorizes patients by assessing changes in tumor size over time and the presence or absence of new tumors [25,26]. However, similar to the TNM system, this evaluation system overlooks other valuable information related to prognosis inside the images for better clinical feasibility. Reported studies referring to AI for post-treatment risk assessment are scarce, while Pre-treatment images are used as default input data for prediction. However, we believe that Post-treatment images contain a massive amount of information related to prognosis. It is valuable to incorporate Post-treatment images for AI-assisted prediction to assist in optimizing the treatment plan.

For the reasons mentioned above, we carried out this study of NPC risk assessment based on DL that incorporated Pre- and Post-treatment MR images, including tumor area, as well as tissues and organs around the tumor. We tried to provide a different approach to the prognosis assessment of NPC.

2. Materials and methods

2.1. Patient screening

Our study was approved by the Ethics Committee of the Renmin Hospital of Wuhan University, and the informed consent from patients was exempted. Patients with locally advanced NPC who were admitted to Renmin Hospital of Wuhan University for treatment between June 2012 and January 2018 were retrospectively selected. Information was collected for sex, age, clinical TNM stage (according to the 7th American Joint Committee on Cancer (AJCC)), TNM staging manual) [27], treatment method adopted in the first course of treatment, pathological outcome, MR images before treatment and after one course of treatment, recurrence time (including local progress and distant metastasis), and death events caused by any reason during follow-up. The endpoint was November 20, 2019, to January 5, 2020 (information on more than 1000 patients was collected in these two months). MR images of each listed patient were retrieved and transferred from the picture archiving and communication system. Patients were screened depending on the following inclusion and exclusion criteria. The inclusion criteria were as follows: 1. Primary NPC diagnosed by pathology and treated in our hospital; 2. No distant metastasis occurred before diagnosis; 3. MRI before treatment and after the first course of treatment could be obtained; 4. A regular review was performed; 5. Patients were classified into stages III and IV by the TNM staging system. The exclusion criteria were as follows: 1. Recurrent NPC after treatment; 2. Distant metastasis was observed at the time of diagnosis; 3. No regular review was performed; 4. Surgery was included in the treatment procedure; 5. The required demographic information and imaging data were not available. The flowchart of patient screening is shown in Fig. 1.

2.2. Concept definition

Definition of the course of treatment for NPC: The formulated treatment plan was completed once. The standard for regular review was at least every 3 months within 2 years after treatment, at least once every six months after treatment for 2–5 years, and annually 5 years after treatment. The requisite items for the review were nasopharyngeal MRI and lung CT, CT/MRI of the head and abdomen, or contingent PET/CT. Telephone follow-up was performed on all patients who stopped review after at least 2 years of regular review to obtain information on whether the patient was still alive. If the patient survived but stopped review or performed a local review, the data were distributed to censored data. Recurrence was ascertained based on nasopharyngeal MRI or pathology obtained endoscopically. The criteria for the MRI to determine recurrence were progressive local bone erosion, abnormal soft tissue areas larger than the previous review, and the newly found intensive shadows in the previous neck review increased progressively during this review. Distant metastasis was determined by the results of lung CT, head and abdomen CT/MRI, or PET/CT. Data on fatalities were obtained by telephone follow-up. To reduce the proportion of censored data due to the small proportion of fatalities at the endpoint, we set progression-free survival (PFS) as the observation endpoint. The PFS period was defined as the time from

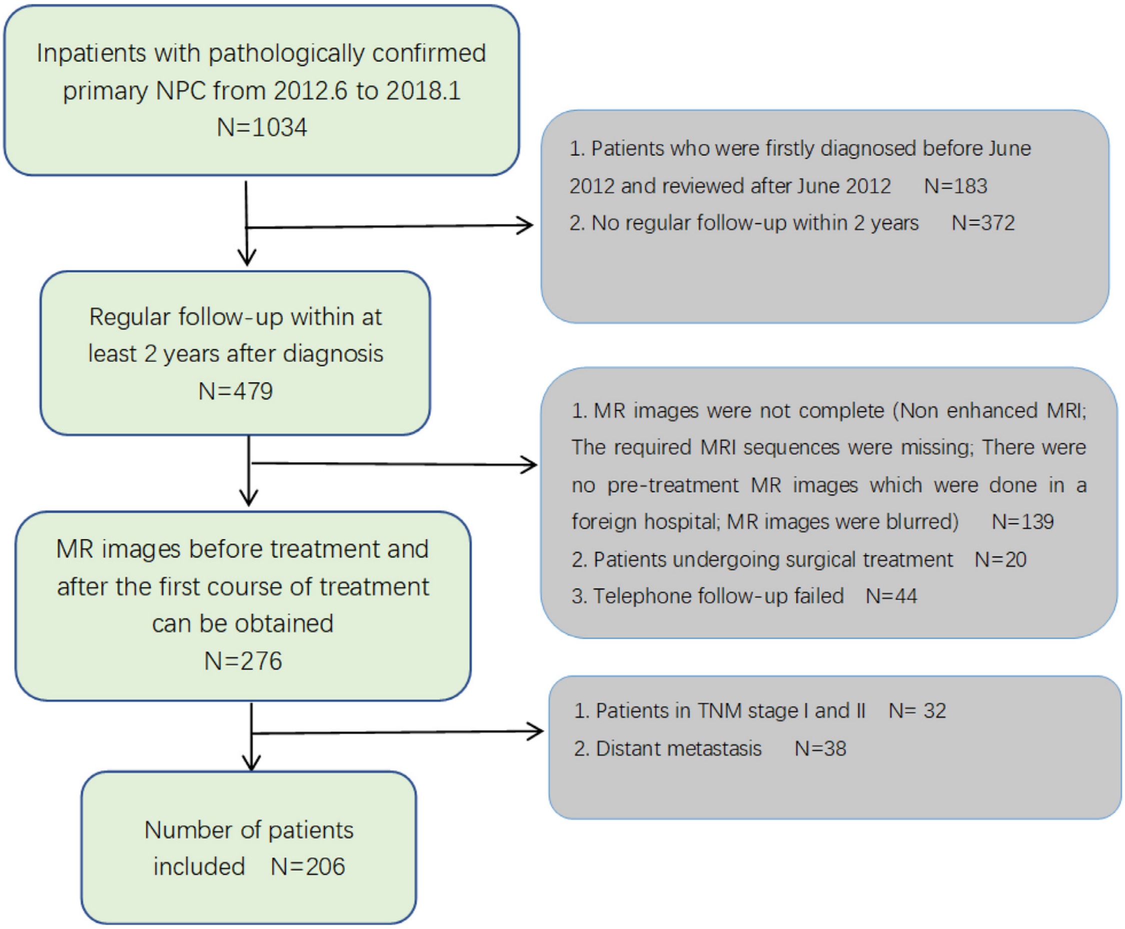
Inpatients with pathologically confirmed primary NPC from 2012.6 to 2018.1 N=1034
1. Patients who were firstly diagnosed before June 2012 and reviewed after June 2012 N=183
2. No regular follow-up within 2 years N=372
Regular follow-up within at least 2 years after diagnosis N=479
1. MR images were not complete (Non enhanced MRI; The required MRI sequences were missing; There were no pre-treatment MR images which were done in a foreign hospital; MR images were blurred) N=139
2. Patients undergoing surgical treatment N=20
3. Telephone follow-up failed N=44
MR images before treatment and after the first course of treatment can be obtained N=276
1. Patients in TNM stage I and II N= 32
2. Distant metastasis N=38
Number of patients included N=206

Fig. 1. Flowchart of the patient selection procedure.

treatment to disease progression (recurrence or distant metastasis) or the occurrence of death caused by any reason or the last review.

2.3. Datasets

The nasopharyngeal and neck MR images were obtained using a 1.5-T MR scanner and were saved in the DICOM format with a size of 512 × 512. Axial images of WATER T2 and enhanced WATER T1 before treatment and after the first treatment course were included for imaging processing. Rectangular regions of interest (ROI) with an approximate size of 400 × 200, including all tumor areas, metastatic cervical lymph nodes, and tissues and organs surrounding the tumor were delineated on each primary image (the size of the rectangular ROIs were allowed to be different for the different sizes of tumors, and the convolution neural network in the established model was set to transform all pictures into the size of 512 × 512 by the "padding" operation which filling "0" pixels around the images). Patients were divided into the high-risk and low-risk groups according to the median progression-free survival (PFS), and patients on the point of median PFS were classified into the high-risk group. Each MR image was labeled according to the patient's outcome. The prediction was made depending on the weighted average of all image classification probabilities of a patient.

Finally, 206 NPC patients were included (male: 147, average age: 50.0 years, range: 16–74 years; female: 59, average age: 50.5 years, range: 18–75 years) and randomly divided into a training cohort and a random cohort (by random number table) according to a ratio of 4:1.

2.4. Network architecture and model training

Our DL models were compiled and trained by Professor Huang, who majored in AI for medical imaging for more than 10 years. Keras 2.2.0, with TensorFlow 2.0, in Python version 3.6, was used as the compiling platform. Inception-Resnet-V2 was introduced as the basis of our transfer learning model to optimize the initial parameters and reduce the training cycle. Inception-Resnet-V2 is a convolutional neural network that achieved top accuracy on the ImageNet Large Scale Visual Recognition Challenge image classification benchmark [28]. The model draws on the strengths of the ResNet network, with the ResNet module in the Inception-Resnet-V2 architecture both speeding up training and improving performance (preventing gradient dispersion); the inception module allows sparse or non-sparse features to be obtained on the same layer. The top layer of Inception-Resnet-V2 was removed, and a full connection layer with two neurons was added at the tail end to achieve the classification function in our study. An ensemble learning model based on a soft voting classifier was established by adding a decision layer to the output layer of the Pre-model and Post-model to integrate the results of the two models (Fig. 2). The soft voting classifier takes the average of the probabilities of all model predictions for a particular class as the criterion, and the category with the highest probability is used as the final prediction result.

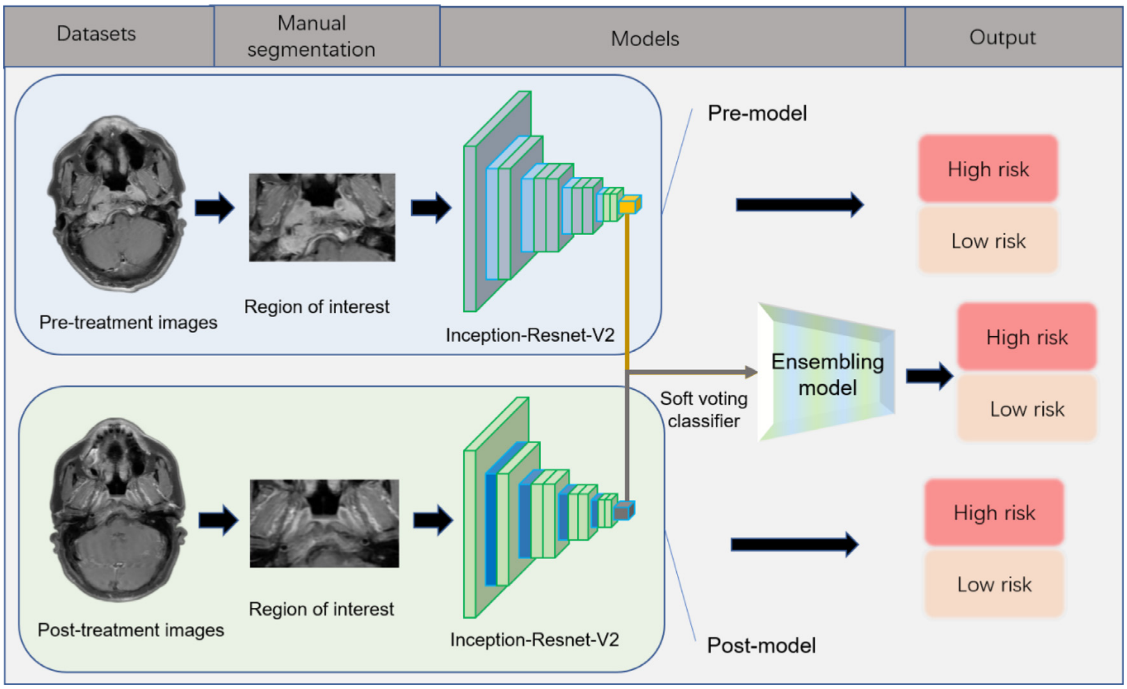

Fig. 2. Pre-model and post-model based on Inception-Resnet-V2 were trained and integrated by soft voting classifier to established the ensembling model. Inception-Resnet-V2: the top layer of Inception-Resnet-V2 was removed, and a full connection layer with two neurons was added at the tail end to achieve the classification function. Ensembling model: the ensemble learning model was established by adding a decision layer to the output layer of the pre-model and Post-model to integrate the results of the two models based on a soft voting classifier.

The training cohort contained approximately 80% of the datasets (163/206). Pre- and Post-treatment images were used as input data to train the modified Inception-Resnet-V2 model, named Pre-model and Post-model, respectively. The Adam Optimizer was used to train the network with a batch size of 32. The initial learning rate was set to 0.0001, and the training procedure was terminated when the accuracy did not improve further for 10 continuous epochs. The dropout in the fully connected layers was set with a probability of 0.5, and the L2 regularization strategy was used to prevent overfitting.

2.5. Model testing and statistical analysis

Approximately 20% of patients' datasets were used as the test cohort (43/206) The probability of patients with low risk and high risk was given according to the patient's MR image dataset, and the category with high probability was selected as the prediction result. In the ensembling model, a new prediction result was obtained by integrating the output of the Pre-model and Post-model based on the soft voting algorithm.

ROC curves were constructed, and the area under the curves (AUC) was calculated to compare the performances of the Pre-model, Post-model, ensembling model, and TNM staging system. The concordance index (C-index) with 95% CIs was reported for the three established models and the TNM staging system. The number of true-positives, false-positives, true-negatives, and false-negatives of the established models were listed in a 2 × 2 table to calculate the confusion matrix. Differences were considered statistically significant at $P < 0.05$. Statistical analyses were performed using SPSS 24.0 and Python, version 3.6. For a better interpretation of the DL models' prediction process, gradient-weighted class activation mapping (Grad-CAM) images for visualizing the areas of the image most indicative of high risk were produced by extracting feature maps from the final convolutional layers. The Matplotlib package in Python was used to produce Grad-CAM images.

3. Results

3.1. Clinical characteristics of patients

A total of 206 patients with locally advanced NPC, including 132 patients with stage III and 74 patients with stage IVa, were included in our study. The median time and interquartile time of follow-up were 36.00 months, 26.25 months, and 48.75 months, respectively. Patients were randomly divided into a training cohort (163 cases) and a test cohort (43 cases). The clinical characteristics of the patients are shown in Table 1. There was no significant difference in age, sex, TNM staging, T staging, N staging, and treatment between the two cohorts

3.2. Performance of DL models

The three DL models (Pre-model, Post-model, and ensembling model) achieved better performance in predicting the prognosis of NPC, with AUCs of 0.741 (95% CI: 0.584–0.900), 0.806 (95% CI: 0.670–0.942), and 0.842 (95% CI: 0.718–0.967), respectively, when compared to TNM staging (AUC = 0.723, 95% CI: 0.567, 0.879) in the test cohort. The ROC curve based on the outcomes of individual patients obtained from the DL models and TNM staging is shown in Fig. 3.

A confusion matrix that reports the number of true-positives, false-positives, true-negatives, and false-negatives was arranged as shown in Table 2. The positive and negative predictive values were 73.9% (17/23) and 75.0% (15/20) for the Pre-model, 73.9% (17/23) and 90.0% (2/20) for the Post-model, 78.2% (18/23) and 90.0% (2/20) for the ensembling model, 69.6% (16/23), and 75.0% (15/20) for TNM staging, respectively. The accuracies of the Pre-model, Post-model, ensembling model and TNM staging were

Table 1
Clinical characteristics of patients in the training cohort and testing cohort.

	training cohort	test cohort	*P value*
Age(years)			*P* = 0.351
Mean±SD	51.63 ± 10.21	49.82 ± 11.59	
< 45	40	11	
45–55	73	15	
> 55	50	17	
Gender			*P* = 0.380
Male	114	33	
Female	49	10	
Staging [1]			*P* = 0.204
III	108	24	
IVa	55	19	
T stage [1]			*P* = 0.806
T1	15	5	
T2	53	15	
T3	54	11	
T4	41	12	
N stage [1]			*P* = 0.570
N0	9	2	
N1	19	3	
N2	113	29	
N3	22	9	
Treatment			*P* = 0.169
CCR [2]	30	12	
IC+CCR/R [3]	133	31	

SD: standard deviation; [1]: based on the 7th edition of the American Joint Committee on Cancer (AJCC)/International Union Against Cancer staging systems; [2]: concurrent chemoradiation; [3]: Including induction chemotherapy + radiotherapy, induction chemotherapy + concurrent chemoradiation.

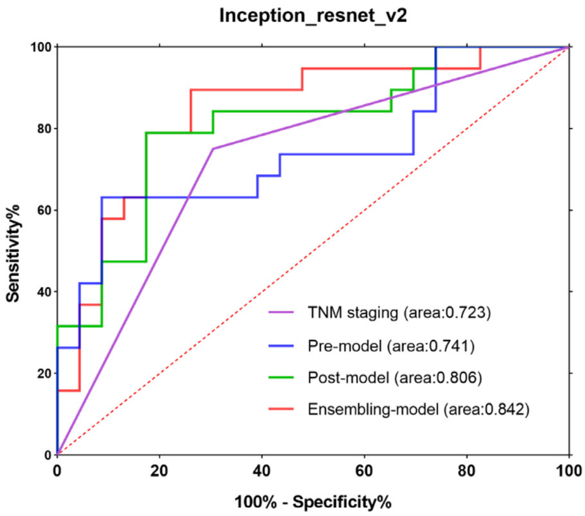

Fig. 3. ROC curves of the Pre-model, post-model, ensembling model and TNM staging in the test cohort.

74.4% (32 /43), 81.4% (35/43), 83.7% (36/43), and 72.1% (31/43), respectively. The C-index was calculated based on the category probability obtained from DL models. As reported in Table 2, the Pre-model, Post-model, ensembling model and TNM staging obtained C-indexes of 0.717 (95% CI: 0.639–0.795), 0.811 (95% CI: 0.762–0.861), 0.830 (95% CI: 0.784–0.877), and 0.709 (95% CI: 0.620-0.788), respectively. The performance of the Post-model based on Post-treatment images was better than that of the Pre-model, which was based on Pre-treatment images whenever evaluated with AUC, accuracy or C-index, and all the three DL models were better than TNM staging, while the ensembling model, which integrated the results of the two DL models, performed best.

Grad-CAM images, which are produced using the class activation mapping method, are composed of four colors(red, yellow, green, and blue), which represent areas that have different predictive significance. The red area represents the greatest correlation with the classification, followed by the yellow region. The green and blue areas indicate a weaker predictive significance (Figs 4–7). For patients classified as high risk, the red and yellow areas represent features related to poor prognosis. The presentation of Grad-CAM images can help in better understanding of how DL networks capture image features for prediction, and to dispel our doubts about the black box of the CNN. We observed that our DL models suggested that the areas around the tumor and some cervical lymph nodes were strongly related to the prognosis of the tumor (marked in red). In many cases, the relationship between the signal of the tumor area and the prognosis was not as strong as expected (Figs. 4–7).

4. Discussion

With the increasing subspecialization of medical fields, the demand for more accurate and informative image reports is booming, challenging radiologists, and medical imaging specialists to know everything about all exams and regions [29]. The purpose of image examination today is not only qualitative diagnosis but also obtaining rich quantitative information such as the severity of the disease, prognosis, therapeutic effect of drugs, etc. [30,31], in which artificial intelligence will make an important difference. The pathological processes of tumor occurrence, growth, and invasion are affected by gene regulation and tumor microenvironment and will show corresponding manifestations in medical images [31]. The "common" manifestations of some tumors can be identified by the naked eye and empirically summarized as the image characteristics of specific tumors. However, more "hidden" information that contains personal data, such as the individual prognosis and response to specific drug treatment, cannot be recognized. With the development of algorithms, the efficacy and efficiency of information extraction from images have significantly improved, thus enabling researchers to make more accurate predictions of prognosis, and greatly benefit the clinical management of cancer [32]. It has been reported that DL matches and even surpasses human performance in task-specific applications [33,34].

Table 2
Confusion matrices for DL Models and TNM staging.

	Ensembling model (prediction)		Pre-model (prediction)		Post-model (prediction)		TNM staging (prediction)	
	High risk	Low risk	High risk	Low risk	High risk	Low risk	High risk	Low risk
High risk(true)	18	5	17	6	17	6	16	7
Low risk(true)	2	18	5	15	2	18	5	15
C-index	0.830		0.717		0.811		0.709	

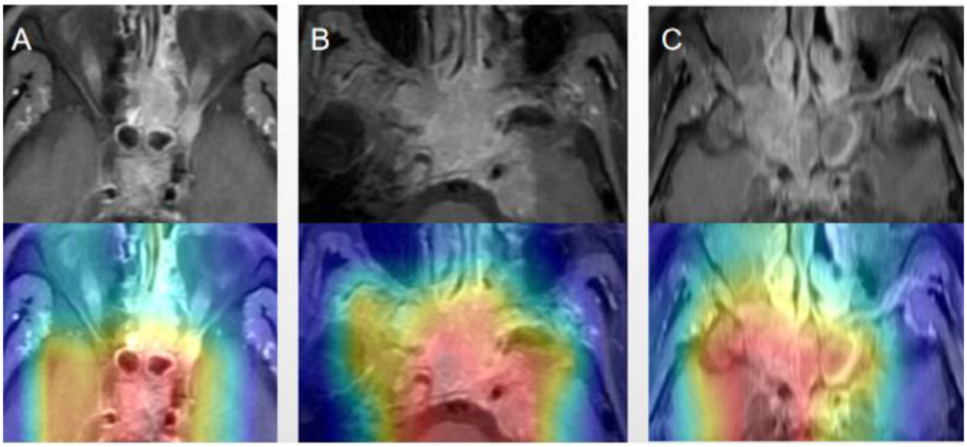

Fig. 4. Original images of A, B and C show that the tumor invade brain. The areas of tumor and surrounding tissues are considered to be related to high risk prognosis in the corresponding Grad-CAM images.

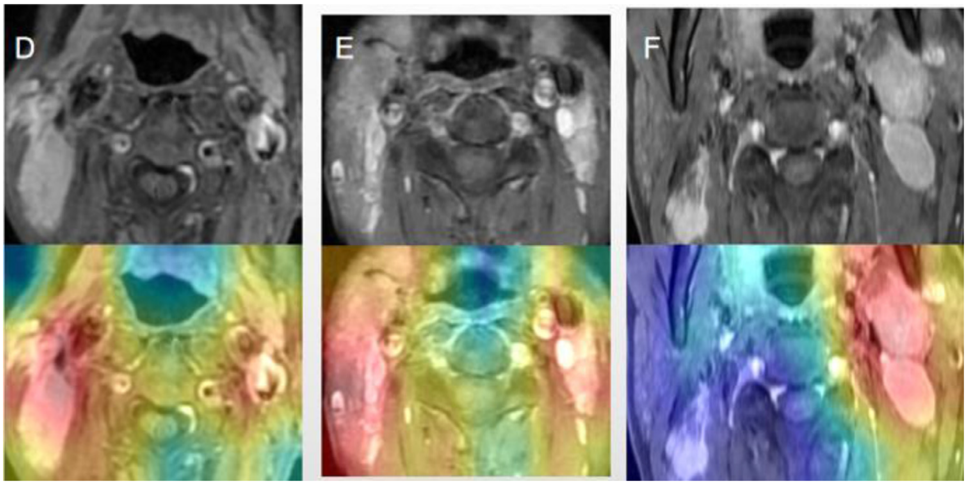

Fig. 5. D, E and F images show metastatic lymph nodes in the neck, and DL models suggests that these lymph node regions contain features associated with high risk prognosis.

Most of the previous studies on tumor prognosis prediction were developed based on the approach of radiomics, and the classic steps generally include image acquisition, manual tumor segmentation, feature extraction, feature filtering, and classification [35]. Although there are several tumor segmentation methods, segmenting images along the edge of the primary tumor were preferred. However, all the criteria for T staging of NPC, which combined evidence-based findings with empirical knowledge, considered the relationship between the tumor and the surrounding tissues and organs [36] and were abandoned by the approach of primary tumor segmentation. The characteristics of these relationships undoubtedly contain much valuable information related to tumor prognosis as the C-index of T staging can reach approximately 60%–70% [20,37]. However, an analysis based on the whole MR image is an indispensable step to realize the clinical practical value of these predictive models. Based on this ideal, we established a model based on the whole MR image for prediction in the pre-experiment stage of our study. Although we obtained an accuracy of nearly 70%, the Grad-CAM images based on the model indicated unreasonable extraction of features, such as cerebrospinal fluid, cerebellum, orbital, and parotid gland, that were considered to be related to prognosis in most cases, even if the tumor was far away from these structures. Excessive image noise and small sample size were considered to be the main reasons; therefore, analysis based on the whole MR image was abandoned. Based on this reason, a rectangular ROI composed of the tumor and the surrounding tissues and organs was included in our study. The Grad-CAM images generated by our DL models indicated that tumor peripheral signals contained very important prognostic information.

The purpose of tumor risk assessment is to guide the development of an appropriate treatment plan. We expect tumors to respond after receiving the treatment. however, it is not uncommon for the “best treatment plan” to yield poor results. Radiotherapy resistance, recurrence, distant metastasis, and complications caused by radiotherapy are analyzed and considered in many studies as the main causes of treatment failure and patient death [38–41]. Oncologists could adjust the predetermined treatment plan according to the obtained evidence to match the estimated tumor risk. As the common RECIST method only evaluates the changes of tumor scope, AI-based analysis is valuable. In the “mind” of AI, medical images are not only pictures, many more prognostic features that are not limited to the tumor scope can be extracted. In our study, we included MR images of NPC before and after a course of treatment for prognosis prediction, and the AUCs of the Pre-model, Post-model, and ensembling models were 0.741, 0.806, and 0.842, respectively. The Post-model shows a better prediction than the Pre-model, while MR images of post-treatment were rarely used for AI-based prognosis prediction. For advanced NPC or other advanced malignant tumors that require multiple courses of treat-

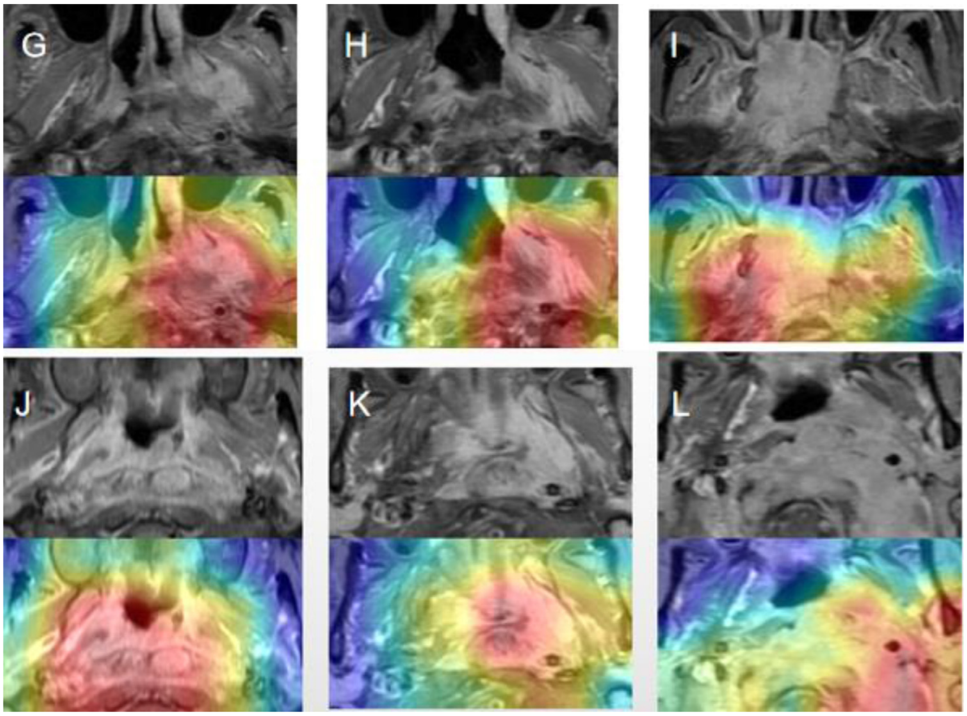

Fig. 6. G, H, I, J, K and L images present the NPCs that mainly invade lateral and posterior structures. The red areas represent the image features that are considered by the DL models to be associated with high risk prognosis.

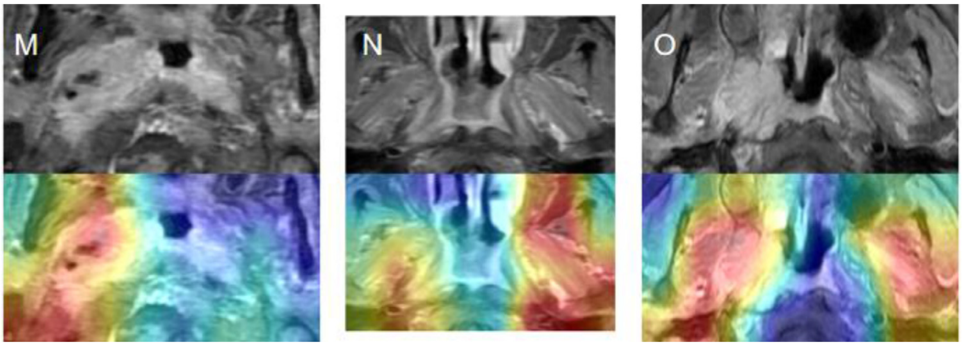

Fig. 7. M, N and O images indicate that not all tumor areas are associated with prognosis.

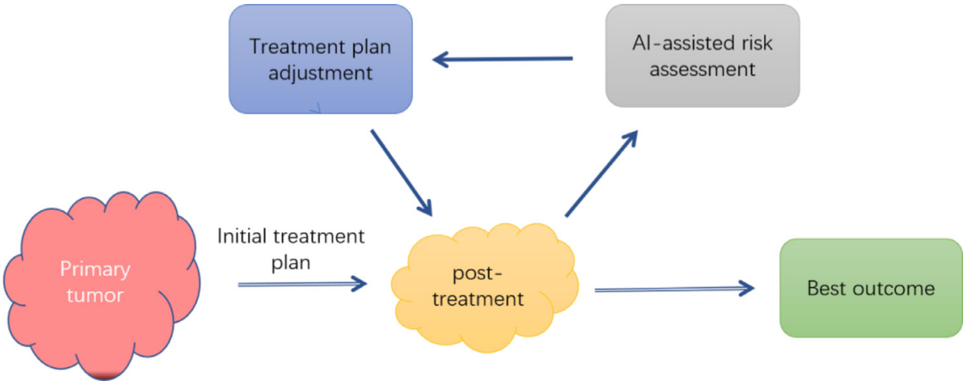

Fig. 8. Cycle assessment of cancer risk after treatment guiding treatment plan adjustment for the best outcome.

ment, it is worth recommending including post-treatment medical images when performing AI-assisted prognosis assessment. In fact, depending on a more mature condition, the best imaginable scenario to assess the images after each treatment course based on AI is to evaluate the real-time risk to guide the optimization of the treatment plan (Fig. 8).

There are several shortcomings in our study. First, the number of cases in our study was limited. The size of the dataset has a complex impact on the performance of DL models that are based on a convolutional neural network. Although transfer learning provides a good solution for small datasets, large samples are expected, especially when confronted with MRI-related tasks. However, to ensure the quality of the dataset, only 206 patients remained in our study after filtering the 1034 patients in the initial list. Second, there are no external datasets for validation in our study. The variety of hospitals has an impact on the outcome of tumors, which cannot be reflected in a single-center dataset. Testing based on multicenter data can provide a better understanding of the generalization ability of the established DL models.

5. Conclusions

The three established DL models based on pre- and post-treatment MR images have a good performance and can accurately capture the image features related to prognosis. Furthermore, post-treatment MR images are of great significance for prognosis prediction, which could assist clinicians in treatment decision optimization.

Declaration of Competing Interest

All authors declared that there were no competing interests.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (Ref. Nos. 81670910 and 81970860). The authors would like to thank all subjects, and their caregivers, for their participation in the study.

References

[1] G. Carioli, E. Negri, D. Kawakita, et al., Global trends in nasopharyngeal cancer mortality since 1970 and predictions for 2020: focus on low-risk areas, Int. J. Cancer 140 (10) (2017) 2256–2264.

[2] R.S. Walters, J.W. Sweetenham, P.J. O'Brien, et al. National Comprehensive Cancer Network About NCCN. (Available at:) https://www.nccn.org/guidelines/guidelines-detail?category=1&id=1437. Date accessed: 15 Feb, 2020.

[3] A.W. Lee, W.T. Ng, L.L. Chan, et al., Evolution of treatment for nasopharyngeal cancer–success and setback in the intensity-modulated radiotherapy era, Radiother. Oncol. 110 (3) (2014) 377–384.

[4] M.L.K. Chua, J.T.S. Wee, E.P. Hui, et al., Nasopharyngeal carcinoma, Lancet 387 (10022) (2016) 1012–1024.

[5] M.K. Gospodarowicz, D. Miller, P.A. Groome, et al., The process for continuous improvement of the TNM classification, Cancer 100 (1) (2004) 1–5.

[6] A. Hosny, C. Parmar, T.P. Coroller, et al., DL for lung cancer prognostication: a retrospective multi-cohort radiomics study, PLoS Med. 15 (11) (2018) e1002711.

[7] A. Hosny, C. Parmar, J. Quackenbush, et al., Artificial intelligence in radiology, Nat. Rev. Cancer 18 (8) (2018) 500–510.

[8] M.K.K. Niazi, A.V. Parwani, M.N. Gurcan, Digital pathology and artificial intelligence, Lancet Oncol. 20 (5) (2019) e253–e261.

[9] O.J. Skrede, S. De Raedt, A. Kleppe, et al., DL for prediction of colorectal cancer outcome: a discovery and validation study, Lancet N. Am. Ed. 395 (10221) (2020) 350–360.

[10] J.N. Kather, J. Krisam, P. Charoentong, et al., Predicting survival from colorectal cancer histology slides using DL: a retrospective multicenter study, PLoS Med. 16 (1) (2019) e1002730.

[11] P. Courtiol, C. Maussion, M. Moarii, et al., DL-based classification of mesothelioma improves prediction of patient outcome, Nat. Med. 25 (10) (2019) 1519–1525.

[12] B. Zhang, J. Tian, D. Dong, et al., Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma, Clin. Cancer Res. 23 (15) (2017) 4259–4269.

[13] E.H. Zhuo, W.J. Zhang, H.J. Li, et al., Radiomics on multi-modalities MR sequences can subtype patients with non-metastatic nasopharyngeal carcinoma (NPC) into distinct survival subgroups, Eur. Radiol. 29 (10) (2019) 5590–5599.

[14] L. Zhang, X. Wu, J. Liu, et al., MRI-based deep-learning model for distant metastasis-free survival in locoregionally advanced nasopharyngeal carcinoma, J. Magn. Reson. Imaging 53 (1) (2021) 167–178.

[15] L.Z. Zhong, X.L. Fang, D. Dong, et al., A deep learning MR-based radiomic nomogram may predict survival for nasopharyngeal carcinoma patients with stage T3N1M0, Radiother. Oncol. 151 (2020) 1–9.

[16] M. Qiang, C. Li, Y. Sun, et al., A prognostic predictive system based on deep learning for locoregionally advanced nasopharyngeal carcinoma, J. Natl. Cancer Inst. 113 (5) (2021) 606–615.

[17] H. Peng, D. Dong, M.J. Fang, 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. 25 (14) (2019) 4271–4279.

[18] D. Dong, F. Zhang, L.Z. Zhong, 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. 17 (1) (2019) 190.

[19] L. Zhao, J. Gong, Y. Xi, et al., MRI-based radiomics nomogram may predict the response to induction chemotherapy and survival in locally advanced nasopharyngeal carcinoma, Eur. Radiol. 30 (1) (2020) 537–546.

[20] P. OuYang, Z. Su, X. Ma, et al., Comparison of TNM staging systems for nasopharyngeal carcinoma, and proposal of a new staging system, Br. J. Cancer 109 (12) (2013) 2987–2997.

[21] A. Hosny, C. Parmar, J. Quackenbush, et al., Artificial intelligence in radiology, Nat. Rev. Cancer 18 (8) (2018) 500–510.

[22] M. Leming, J.M. Górriz, J. Suckling, Ensemble deep learning on large, mixed-site fMRI datasets in autism and other tasks, Int. J. Neural Syst. 30 (7) (2020) 2050012.

[23] L. Shao, F. Zhu, Li XJItonn, Systems l. Transfer learning for visual categorization: a survey, IEEE Trans. Neural Netw. Learn. Syst. 26 (5) (2015) 1019–1034.

[24] C.Q. Tan, F.C. Sun, T. Kong, W.C. Zhang, C. Yang, C.F. Liu, V. Kurkova, Y. Manolopoulos, B. Hammer, L. Iliadis, I. Maglogiannis, A survey on deep transfer learning, in: Artificial Neural Networks and Machine Learning - ICANN 2018, Springer International Publishing Ag, Cham, 2018, pp. 270–279. Pt Iii. Lecture Notes in Computer Science. 11141.

[25] T. Patrick, G.A. Susan, A.E. Elizabeth, et al., New guidelines to evaluate the response to treatment in solid tumors, Natl. Cancer Inst. Canada 92 (3) (2000) 205–216.

[26] E.A. Eisenhauer, P. Therasse, J. Bogaerts, et al., New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1), Eur. J. Cancer 45 (2) (2009) 228–247.

[27] S.B. Edge, CC. Compton, The american joint committee on cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM, Ann. Surg. Oncol. 17 (6) (2010) 1471–1474.

[28] C. Szegedy, S. Ioffe, V. Vanhoucke, et al., Inception-v4, inception-resnet and the impact of residual connections on learning[C], in: Proceedings of the Thirty–First AAAI Conference on Artificial Intelligence, 2017.

[29] M.K. Santos, J.R. Ferreira, D.T. Wada, et al., Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: advances in imaging towards to precision medicine, Radiol. Bras. 52 (6) (2019) 387–396.

[30] S. Huang, J. Yang, S. Fong, et al., Artificial intelligence in cancer diagnosis and prognosis: opportunities and challenges, Cancer Lett. 471 (2020) 61–71.

[31] W.L. Bi, A. Hosny, M.B. Schabath, et al., Artificial intelligence in cancer imaging: clinical challenges and applications, CA Cancer J. Clin. 69 (2) (2019) 127–157.

[32] W. Zhu, L. Xie, J. Han, et al., The application of DL in cancer prognosis prediction, Cancers (Basel) 12 (3) (2020) 603–621.

[33] B.B. Ehteshami, M. Veta, D.P. Johannes van, et al., Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer, JAMA 318 (22) (2017) 2199–2210.

[34] A. Esteva, B. Kuprel, R.A. Novoa, et al., Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (7639) (2017) 115–118.

[35] E. Limkin, R. Sun, L. Dercle, et al., Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology, Ann. Oncol. 28 (6) (2017) 1191–1206.

[36] M. Amin, F. Greene, S. Edge, et al., The eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging, CA Cancer J. Clin. 67 (2) (2017) 93–99.

[37] J.J. Pan, W.T. Ng, J.F. Zong, et al., Proposal for the 8th edition of the AJCC/UICC staging system for nasopharyngeal cancer in the era of intensity-modulated radiotherapy, Cancer 122 (4) (2016) 546–558.

[38] L.L. Tang, W.Q. Chen, W.Q. Xue, et al., Global trends in incidence and mortality of nasopharyngeal carcinoma, Cancer Lett. 374 (1) (2016) 22–30.

[39] I. Karam, S.H. Huang, A. McNiven, et al., Outcomes after reirradiation for recurrent nasopharyngeal carcinoma: North American experience, Head Neck 38 (Suppl 1) (2016) E1102–E1109.

[40] J.J. Yao, Z.Y. Qi, Z.G. Liu, et al., Clinical features and survival outcomes between ascending and descending types of nasopharyngeal carcinoma in the intensity-modulated radiotherapy era: a big-data intelligence platform-based analysis, Radiother. Oncol. 137 (2019) 137–144.

[41] H. Peng, L. Chen, Y. Zhang, et al., The tumour response to induction chemotherapy has prognostic value for long-term survival outcomes after intensity–modulated radiation therapy in nasopharyngeal carcinoma, Sci. Rep. 6 (2016) 24835.