

Deep Learning for Locally Advanced Nasopharyngeal Carcinoma Prognostication Based on Pre- and Post-treatment MRI

Song Li

Renmin Hospital of Wuhan University

Yu-Qin Deng

Renmin Hospital of Wuhan University

Hong-Li Hua

Renmin Hospital of Wuhan University

Sheng-Lan Li

Renmin Hospital of University

Xi-Xiang Chen

Renmin Hospital of Wuhan University

Bao-Jun Xie

Renmin Hospital of Wuhan University

Zhiling Zhu

Huazhong University of Science and Technology

Ruoyun Liu

Wuhan Textile University

Jin Huang

Wuhan Textile University

Ze-Zhang Tao (✉ taozezhang@163.com)

Renmin Hospital of Wuhan University

Research

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

Posted Date: November 12th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-103654/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Computer Methods and Programs in Biomedicine on March 1st, 2022. See the published version at <https://doi.org/10.1016/j.cmpb.2022.106785>.

Abstract

Background: Although it has been reported by several studies that using AI to predict the prognosis of nasopharyngeal carcinoma (NPC) based on magnetic resonance (MR) image, the information around the tumor was not valued and the post-treatment MR images were ignored. Herein we aimed to predict the prognosis of advanced NPC (stage \geq IIa) using pre- and post-treatment MR images based on deep learning (DL).

Methods: A total of 206 patients with primary NPC 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 InceptionResnetV2-based transfer learning models, named pre-model and post-model, were trained with the Pre-dataset and the Post-dataset, respectively. In addition, an ensemble learning model based on the pre-model and post-models was trained. The three established models were evaluated by receiver operating characteristic (ROC) analysis, confusion matrix, and Harrell's concordance indices (C-index) after the model test. High-risk-related heat maps were developed according to the DL models.

Results: The pre-model, post-model, and ensemble models 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 0.745 (95% CI: 0.592–0.897), 0.820 (95% CI: 0.687–0.953), and 0.841 (95% CI: 0.715–0.968) for the test cohort, respectively. In comparison with the models, the post-model performance was better than the pre-model, which indicated the importance of post-treatment images for prognosis prediction. All three DL models performed better than the TNM staging system. The captured features presented on heat maps showed 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.

Background

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, except for the study reported by Mengyun Qiang (16), the progression of these studies unanimously 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. 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.

Materials And Methods

Patient screening

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 (PACS). 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.

Figure 1: Flowchart of the patient selection procedure

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

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 ROIs 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, 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.

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. InceptionResnetV2 was introduced as the basis of our transfer learning model to optimize the initial parameters and reduce the training cycle. The top layer of InceptionResnetV2 was removed, and a full connection layer with two neurons was added at the tail end to achieve the classification function. 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.

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 InceptionResnetV2 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. The network architecture of our model and the conventional method based on primary tumor segmentation are shown in Fig. 2.

Figure 2: Models for tumor prognosis prediction

Model Testing and Statistical Analysis

Approximately 20% of patients' datasets were used as the test cohort (43/206) and were imported into the trained Pre-model and Post-model. 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- and post-models 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 C-index with 95% confidence interval (CI) 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, heat maps 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 heat maps.

Results

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

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

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.

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 0.745 (95% CI: 0.592, 0.897), 0.820 (95% CI: 0.687, 0.953), and 0.841 (95% CI: 0.715, 0.968), 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.

Figure 3: Receiver operating characteristic curves of DL Models and TNM staging

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 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, and 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.

Table 2: Confusion matrices for DL Models and TNM staging

Heat maps, which are produced using the class activation mapping method, are composed of four colors, namely, 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 heat maps 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 convolution neural network (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 (Fig. 4–7).

Figure 4: Original images of A, B and C show that the tumor invade brain. The areas of tumor and surrounding tissues were considered to be related to high risk prognosis In the corresponding heat maps.

Figure 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

Figure 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 associated with high risk prognosis

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

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 (28). 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. (29, 30), 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 (30). 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 (31). It has been reported that DL matches and even surpasses human performance in task-specific applications (32, 33).

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 (34). 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 combining evidence-based findings with empirical knowledge were composed of the relationship between tumor and surrounding tissues and organs (35), 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, 36). 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 heatmaps 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 heat maps 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 (37–40). 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 "eye" 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 0.745, 0.820, and 0.841, 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 treatment, 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).

Figure 8: Cycle assessment of cancer risk after treatment guiding treatment plan adjustment for the best outcome

Regarding the topic of predicting the prognosis of NPC based on imaging data, we listed C-indexes and AUCs of several reports which are commonly used as evaluation indicators in Table 3 for comparison. Because it is impossible to analyze all the variables that affect survival, the accuracy given by prediction models inevitably has an upper limit, irrespective if they are based on medical imaging data, clinical data, or both. The C-indexes or AUCs of the prediction models established in the previous studies were between 0.694 and 0.863, which exceeded the predictive ability of TNM staging.

Table 3: Studies predicting the prognosis of nasopharyngeal carcinoma based on medical imaging

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.

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.

Abbreviations

NPC: nasopharyngeal carcinoma; MR: magnetic resonance; DL: deep learning; ROI: region of interest; ROC: receiver operating characteristic; C-index/Harrell's concordance indices; NCCN/the National Comprehensive Cancer Network; AI: Artificial intelligence; RECIST: The Response Evaluation Criteria in Solid Tumors; AJCC: American Joint Committee on Cancer; PACS: picture archiving and communication system; CT: Computed tomography; PET: positron emission tomography; PFS: progression-free survival; AUC: area under the curves; CI: confidence interval; CNN: convolution neural network.

Declarations

Acknowledgements

The authors would like to thank all subjects, and their caregivers, for their participation in the study.

Authors' contributions

ZZT and JH conceived and designed the study, JH established computer models; SL collected data, created image datasets and drafted the manuscript; YQD guided the construction of datasets, conducted data analysis and helped to modify the manuscript; HLH, SLL, XXC and BJX collected data; ZLZ and RYL contributed to discussion and conducted data analysis. All author(s) read and approved the final manuscript.

Funding

This work was funded by the National Natural Science Foundation of China (Reference Numbers 81670910 and 81970860).

Availability of data and materials

The datasets used during this study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Our study was approved by the Ethics Committee of the Renmin Hospital of Wuhan University, and the informed consent from patients was exempted.

Consent for publication

All authors have approved the manuscript and consent publication.

Competing interests

All authors declared that there were no competing interests.

References

1. Carioli G, Negri E, Kawakita D, et al. Global trends in nasopharyngeal cancer mortality since 1970 and predictions for 2020: Focus on low-risk areas. *Int J Cancer*. 2017;140(10):2256–64.
2. National Comprehensive Cancer Network guidelines. About NCCN. Available at <https://www.nccn.org/>. Accessed 15 Feb 2018.
3. Lee AW, Ng WT, Chan LL, et al. Evolution of treatment for nasopharyngeal cancer—success and setback in the intensity-modulated radiotherapy era. *Radiother Oncol*. 2014;110(3):377–84.
4. Chua MLK, Wee JTS, Hui EP, et al. Nasopharyngeal carcinoma. *The Lancet*. 2016;387(10022):1012–24.
5. Gospodarowicz MK, Miller D, Groome PA, et al. The process for continuous improvement of the TNM classification. *Cancer*. 2004 Jan;100(1)(1):1–5.
6. Hosny A, Parmar C, Coroller TP, et al. DL for lung cancer prognostication: A retrospective multi-cohort radiomics study. *PLoS Med*. 2018;15(11):e1002711.
7. Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018 Aug;18(8):500–10.
8. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol*. 2019 May;20(5):e253–61.
9. Skrede O-J, De Raedt S, Kleppe A, et al. DL for prediction of colorectal cancer outcome: a discovery and validation study. *The Lancet*. 2020;395(10221):350–60.
10. Kather JN, Krisam J, Charoentong P, et al. Predicting survival from colorectal cancer histology slides using DL: A retrospective multicenter study. *PLoS Med*. 2019;16(1):e1002730.
11. Courtiol P, Maussion C, Moarii M, et al. DL-based classification of mesothelioma improves prediction of patient outcome. *Nat Med*. 2019;25(10):1519–25.
12. 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(15):4259–69.
13. Zhuo EH, Zhang WJ, Li HJ, et al. Radiomics on multi-modalities MR sequences can subtype patients with non-metastatic nasopharyngeal carcinoma (NPC) into distinct survival subgroups. *Eur Radiol*. 2019;29(10):5590–9.
14. Zhang L, Wu X, Liu J, et al. MRI-Based Deep-Learning Model for Distant Metastasis-Free Survival in Locoregionally Advanced Nasopharyngeal Carcinoma. *J Magn Reson Imaging*. 2020 Aug 9.
15. Zhong LZ, Fang XL, Dong D, et al. A deep learning MR-based radiomic nomogram may predict survival for nasopharyngeal carcinoma patients with stage T3N1M0. *Radiother Oncol*. 2020 Jul;4:151:1–9.
16. Qiang M, Li C, Sun Y, et al. A Prognostic Predictive System Based on Deep Learning for Locoregionally Advanced Nasopharyngeal Carcinoma. *J Natl Cancer Inst*. 2020 Sep;24:djaa149.
17. 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 Jul 15;25(14):4271–4279.

18. 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(1):190.
19. Zhao L, Gong J, Xi Y, et al. MRI-based radiomics nomogram may predict the response to induction chemotherapy and survival in locally advanced nasopharyngeal carcinoma. *Eur Radiol.* 2020;30(1):537–46.
20. OuYang P, Su Z, Ma X, et al. Comparison of TNM staging systems for nasopharyngeal carcinoma, and proposal of a new staging system. *Br J Cancer.* 2013;109(12):2987–97.
21. Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology. *Nat Rev Cancer.* 2018;18(8):500–10.
22. Leming M, Górriz JM, Suckling J. Ensemble Deep Learning on Large, Mixed-Site fMRI Datasets in Autism and Other Tasks. *Int J Neural Syst.* 2020 Jul;30(7):2050012.
23. Shao L, Zhu F, Li X, Jitton, systems I. Transfer learning for visual categorization: a survey. *IEEE Trans Neural Netw Learn Syst.* 2015;26(5):1019-34.
24. Tan CQ, Sun FC, Kong T, Zhang WC, Yang C, Liu CF. A Survey on Deep Transfer Learning. In: Kurkova V, Manolopoulos Y, Hammer B, Iliadis L, Maglogiannis I, editors. *Artificial Neural Networks and Machine Learning - Iccann 2018, Pt Iii.* Lecture Notes in Computer Science. 11141. Cham: Springer International Publishing Ag; 2018. pp. 270–9.
25. SG PT, EA A. E, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States. National Cancer Institute of Canada. 2000;92(3):205–16.
26. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer.* 2009 Jan;45(2):228–47.
27. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol.* 2010 Jun;17(6):1471–4.
28. Santos MK, Ferreira Junior JR, Wada DT, et al. Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: advances in imaging towards to precision medicine. *Radiol Bras.* 2019;52(6):387–96.
29. Huang S, Yang J, Fong S, et al. Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges. *Cancer Lett.* 2020;471:61–71.
30. Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J Clin.* 2019;69(2):127–57.
31. Zhu W, Xie L, Han J, et al. The Application of DL in Cancer Prognosis Prediction. *Cancers (Basel).* 2020;12(3).
32. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA.* 2017 Dec;12(22):2199–210. 318(.

33. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017 Feb 2;542(7639):115–118.
34. Limkin E, Sun R, Dercle L, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol*. 2017;28(6):1191–206.
35. Amin M, Greene F, Edge S, 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*. 2017;67(2):93–9.
36. Pan JJ, Ng WT, Zong JF, et al. Proposal for the 8th edition of the AJCC/UICC staging system for nasopharyngeal cancer in the era of intensity-modulated radiotherapy. *Cancer*. 2016 Feb 15;122(4):546–58.
37. Tang LL, Chen WQ, Xue WQ, et al. Global trends in incidence and mortality of nasopharyngeal carcinoma. *Cancer Lett*. 2016 Apr 28;374(1):22–30.
38. Karam I, Huang SH, McNiven A, et al. Outcomes after reirradiation for recurrent nasopharyngeal carcinoma: North American experience. *Head Neck*. 2016;38(Suppl 1):E1102-9.
39. Yao JJ, Qi ZY, Liu ZG, 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*. 2019 Aug;137:137–44.
40. Peng H, Chen L, Zhang Y, 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*. 2016 Apr;21:6:24835.

Tables

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

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

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	

Table 3
 Studies predicting the prognosis of nasopharyngeal carcinoma based on medical imaging

Studies	Number of patients	Main evaluation indicators	Outcome of test
Zhang B, 2017.3(1)	118	C-index	0.737
		prediction model	0.634
		TNM	
En-Hong Zhuo, 2019.2(2)	658	C-index	0.814
		prediction model	0.765
		TNM	
Lina Zhao, 2019.4(3)	123	C-index	0.863
		prediction model	0.826
		Accuracy	0.863
		prediction model	
		AUC	
Hao Peng, 2019.4(4)	707	C-index	0.722
		prediction model	0.671
		EB-DNA-model	
Lian-Zhen Zhong, 2020.6(5)	638	C-index	0.788
		prediction model	0.625
		clinical-model	
Lu Zhang, 2020.7(6)	233	AUC	0.765
		prediction model	0.649
		clinical-model	

Studies	Number of patients	Main evaluation indicators	Outcome of test
Mengyun Qiang,2020.9(7)	3444	C-index	0.694
		MRI score	0.633
		TNM	0.776
		Overall score	0.690
		AUC	0.674
		MRI score	0.806
		TNM	
		Overall score	
Our study	206	C-index	0.717
		Pre-model	0.811
		Post-model	0.830
		Ensembling model	0.709
		TNM	0.745
		AUC	0.820
		Pre-model	0.841
		Post-model	0.723
		Ensembling model	
TNM			

Figures

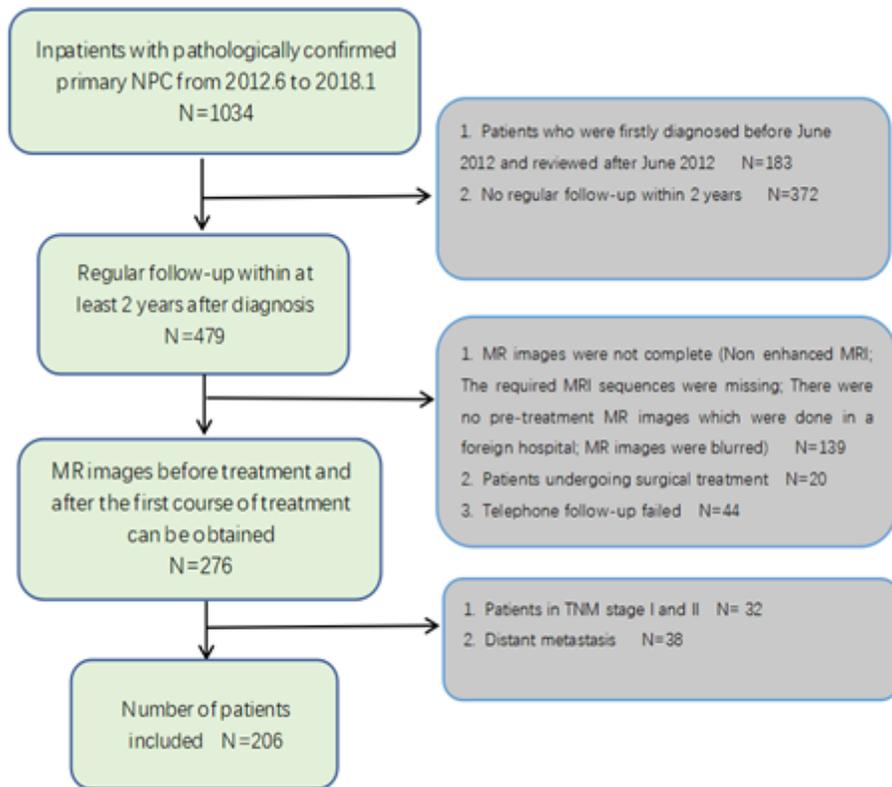


Figure 1

Flowchart of the patient selection procedure

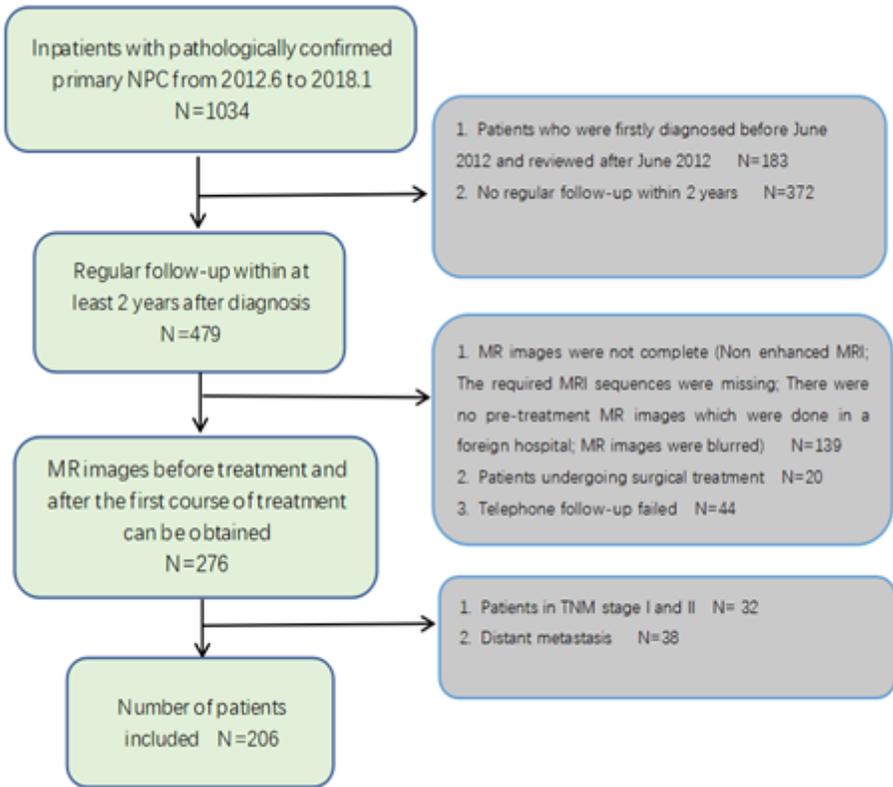


Figure 1

Flowchart of the patient selection procedure

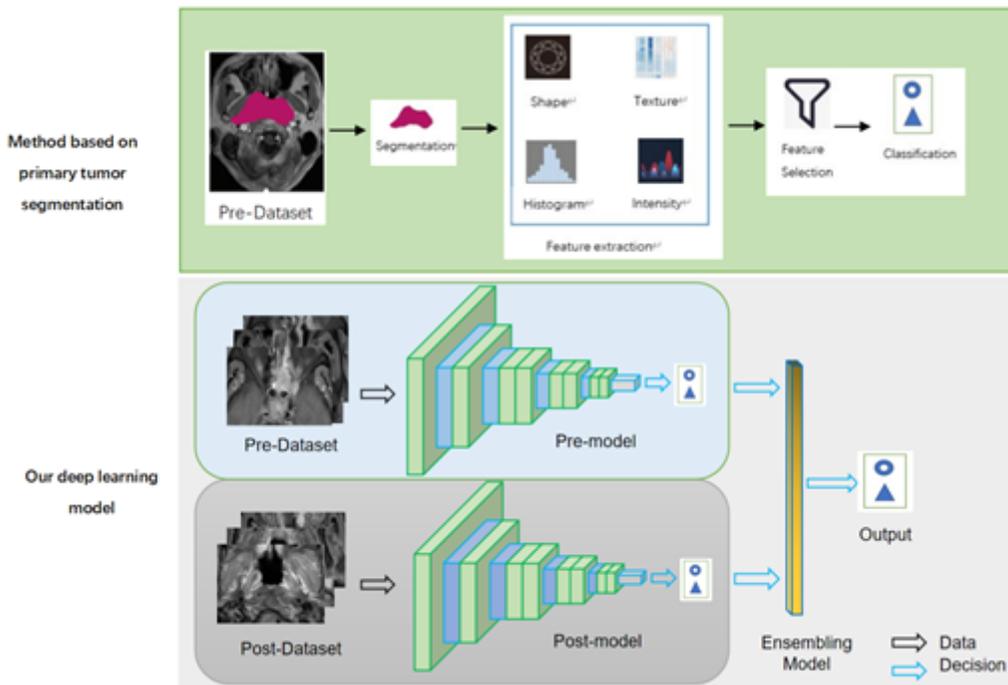


Figure 2

Models for tumor prognosis prediction

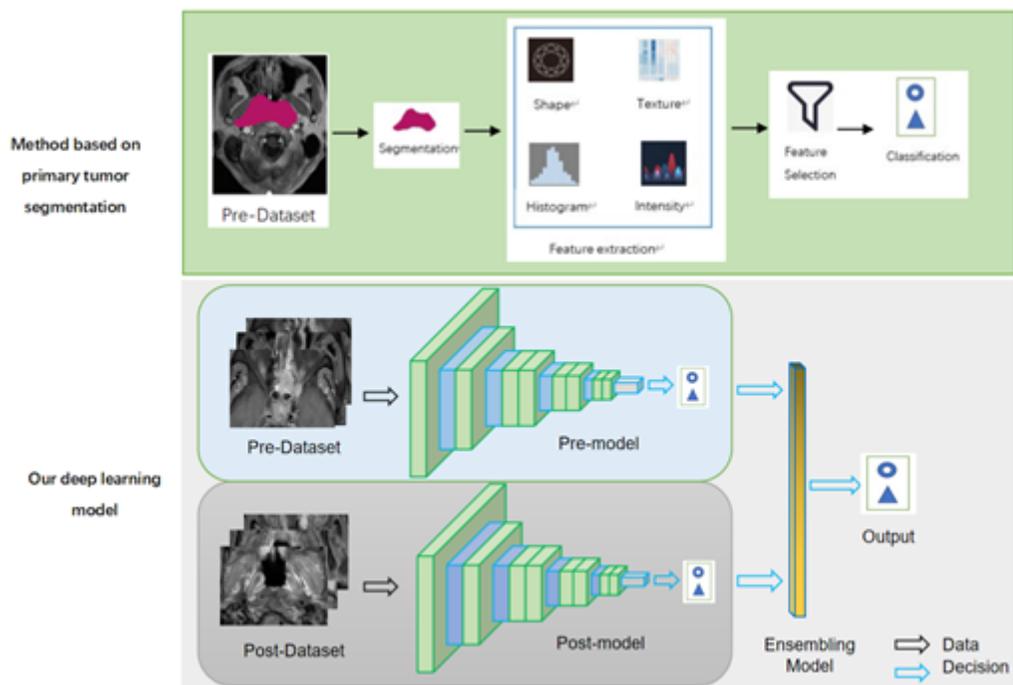


Figure 2

Models for tumor prognosis prediction

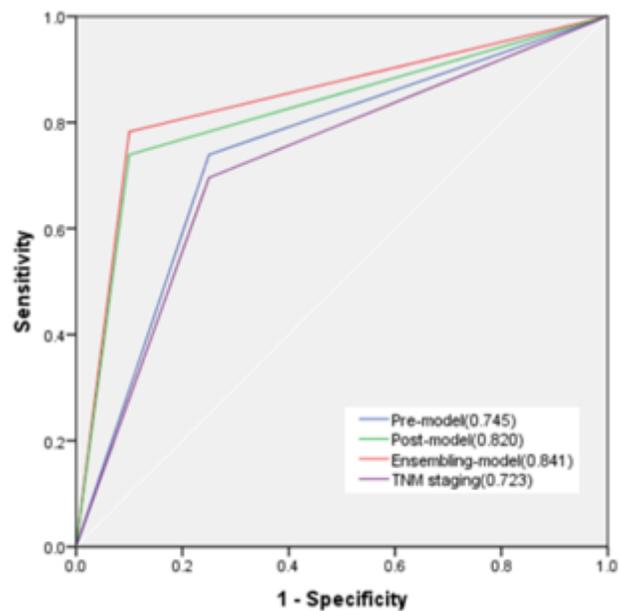


Figure 3

Receiver operating characteristic curves of DL Models and TNM staging

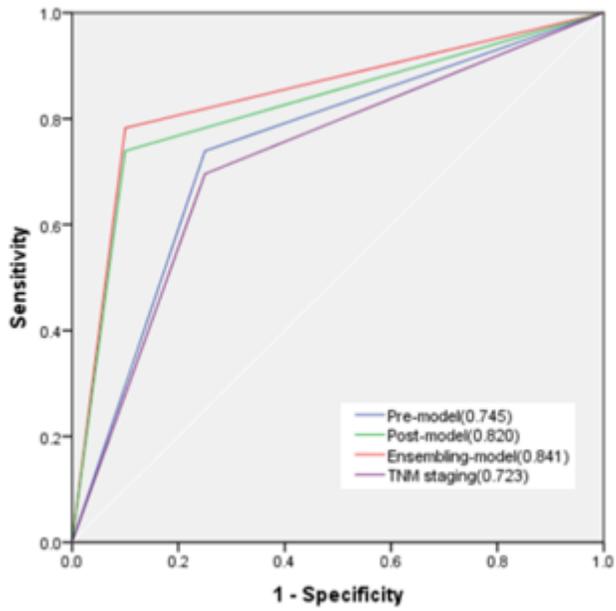


Figure 3

Receiver operating characteristic curves of DL Models and TNM staging

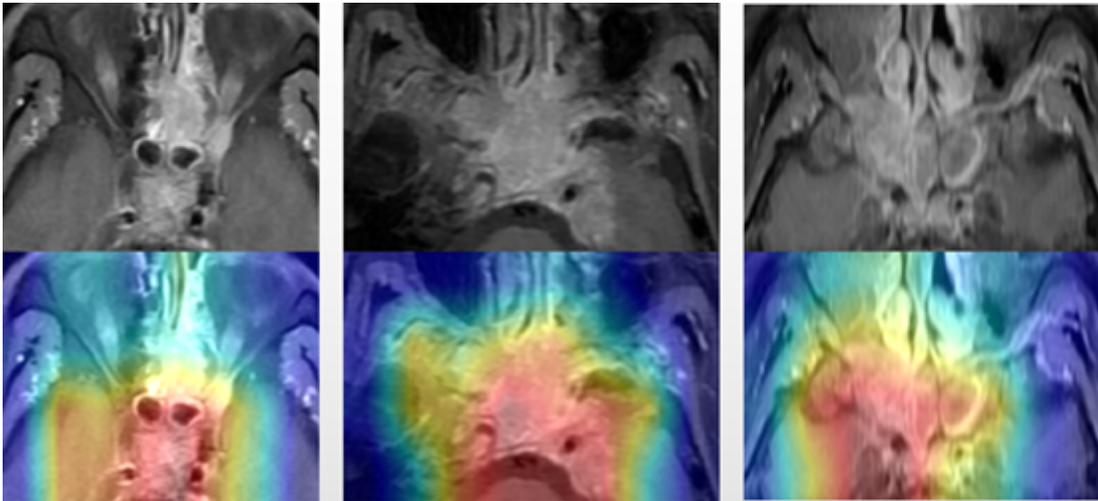


Figure 4

Original images of A, B and C show that the tumor invade brain. The areas of tumor and surrounding tissues were considered to be related to high risk prognosis In the corresponding heat maps.

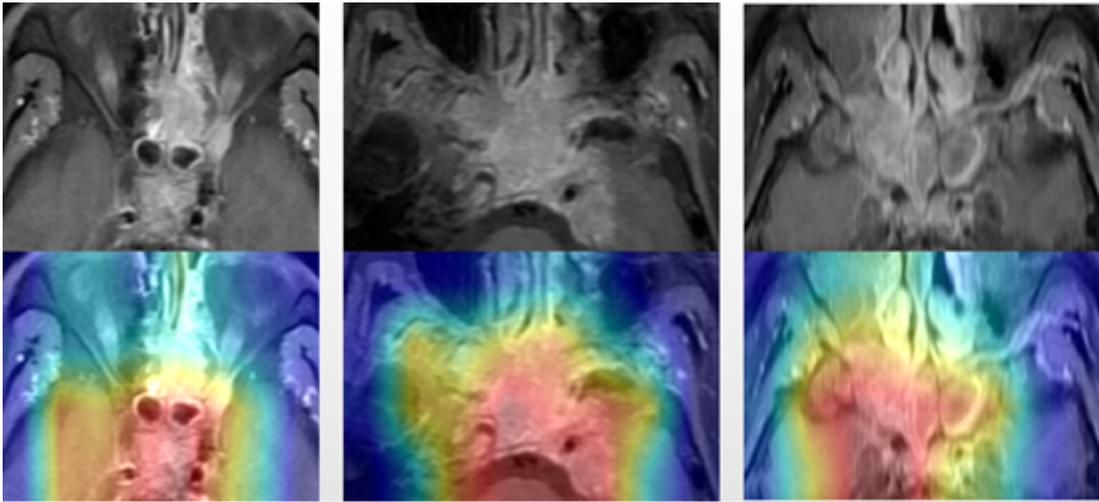


Figure 4

Original images of A, B and C show that the tumor invade brain. The areas of tumor and surrounding tissues were considered to be related to high risk prognosis In the corresponding heat maps.

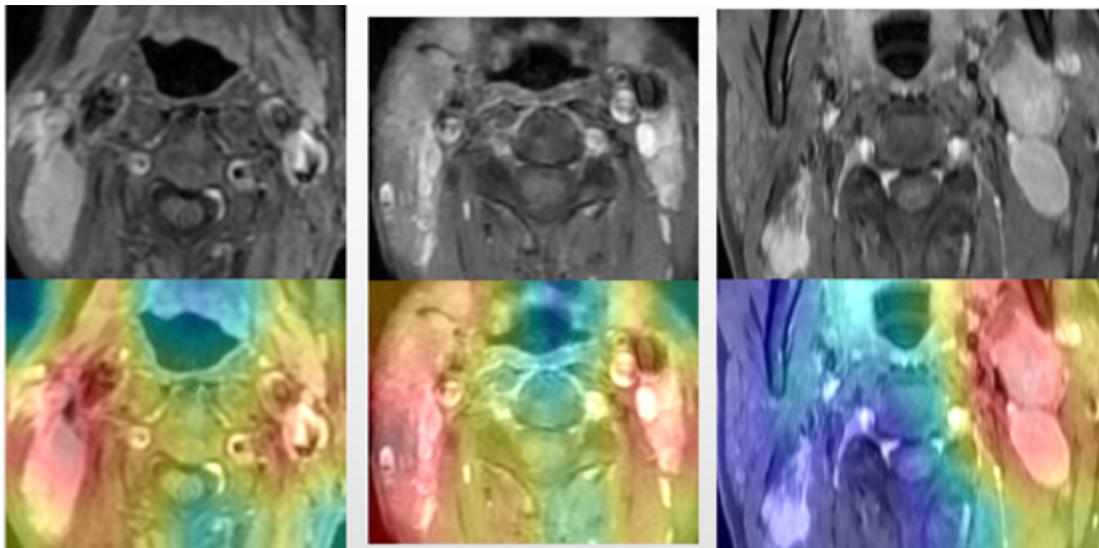


Figure 5

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

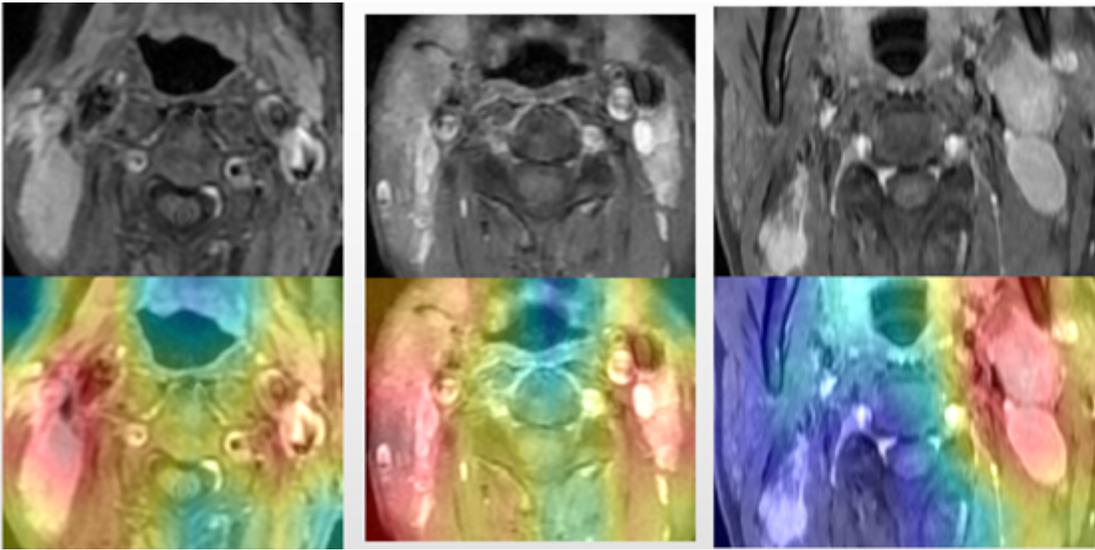


Figure 5

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

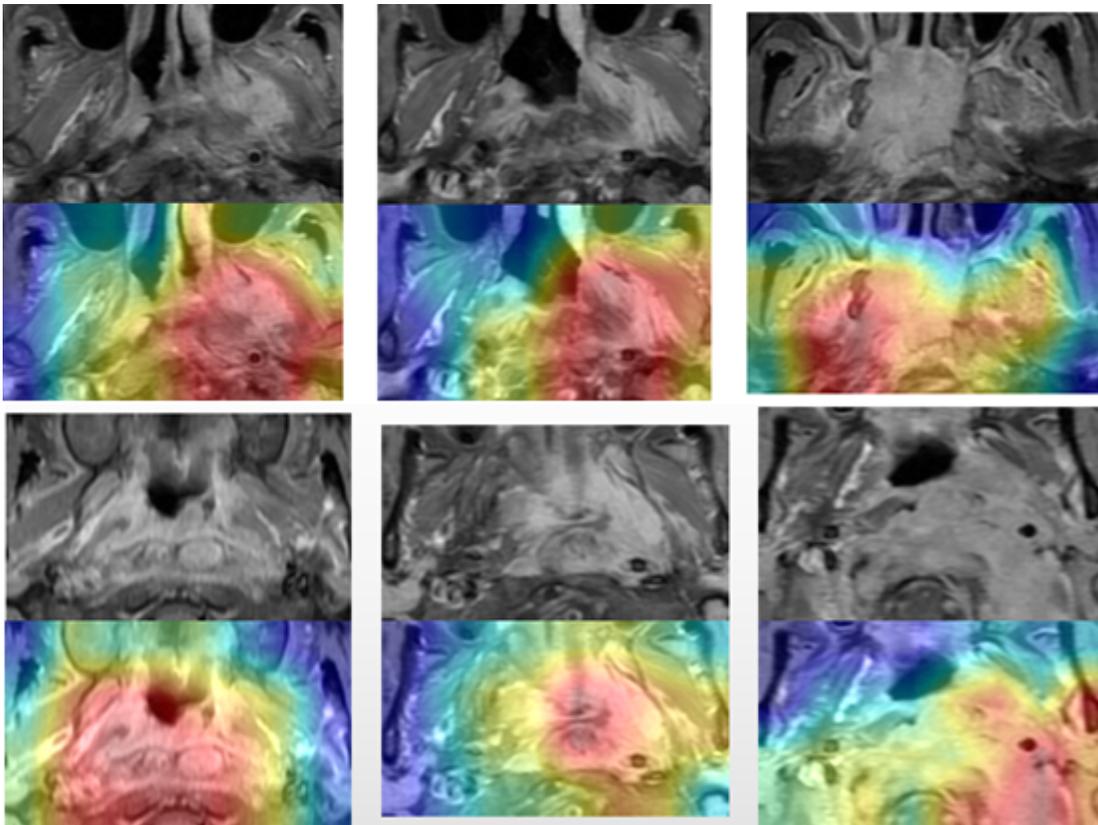


Figure 6

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 associated with high risk prognosis

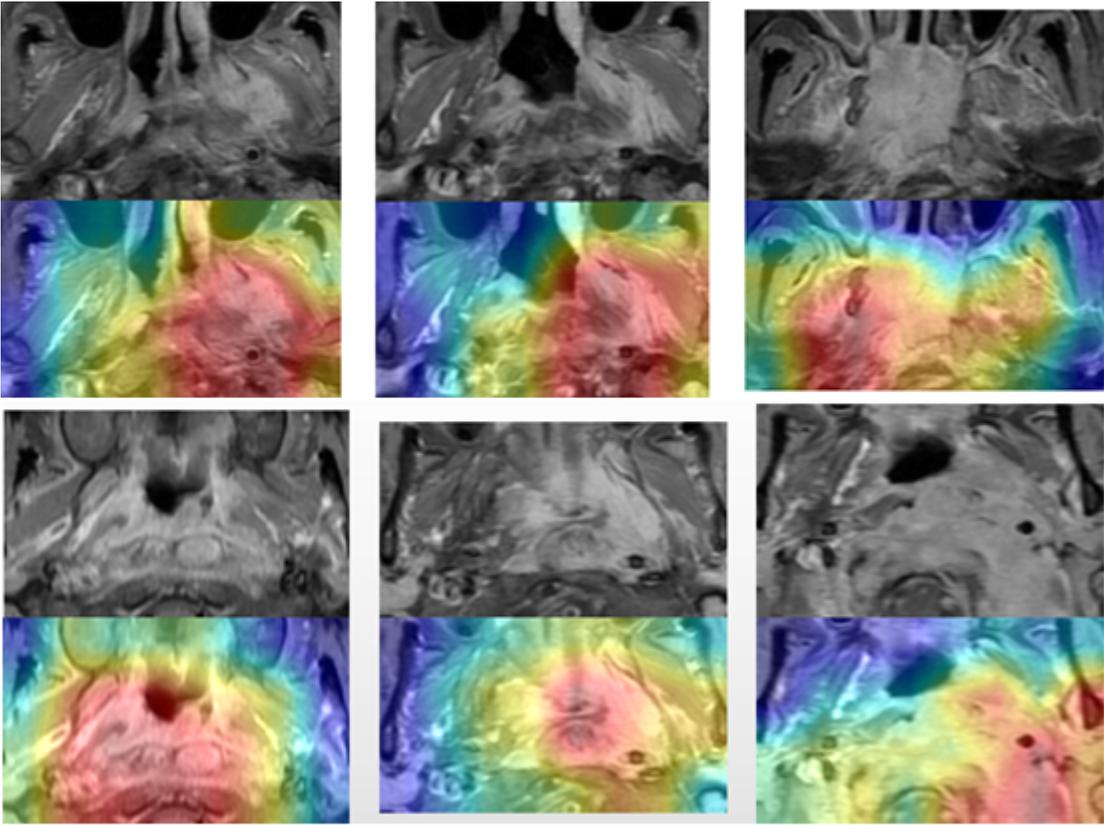


Figure 6

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 associated with high risk prognosis

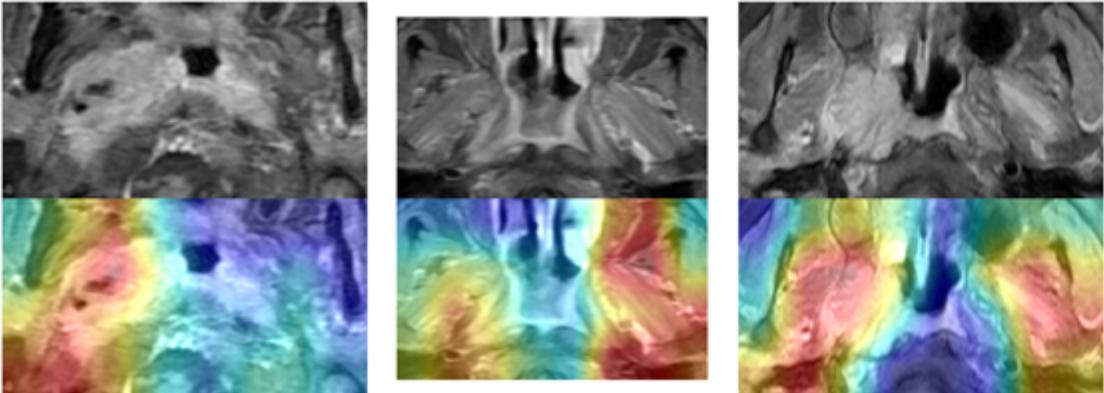


Figure 7

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

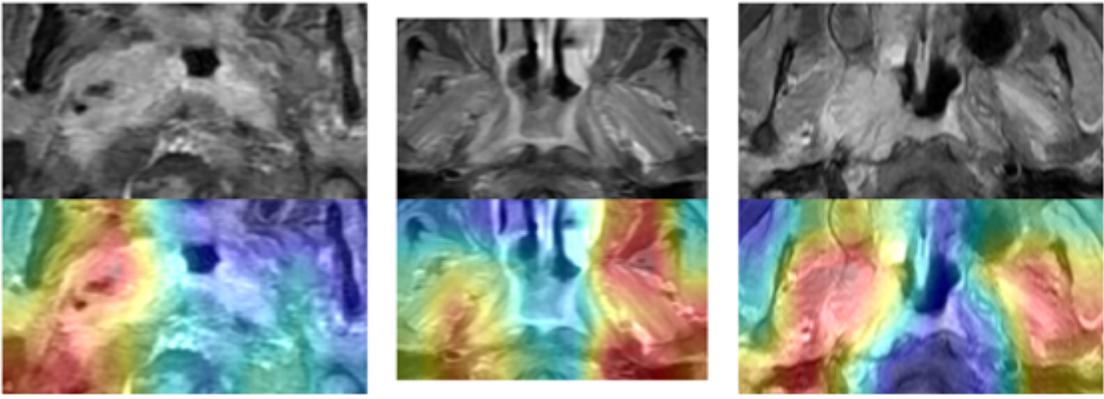


Figure 7

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

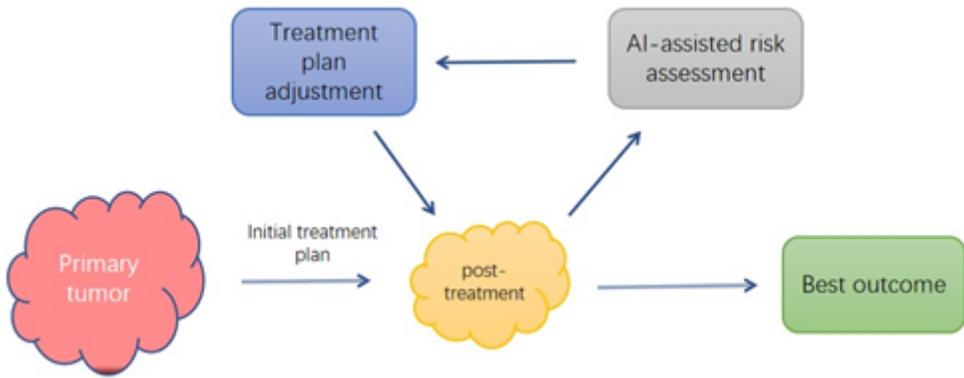


Figure 8

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

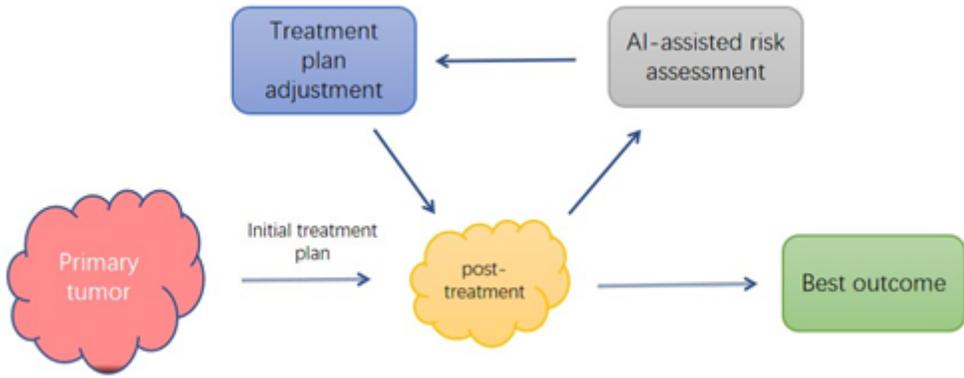


Figure 8

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