Deep learning-based precise prediction and early detection of radiation-induced temporal lobe injury for nasopharyngeal carcinoma

Pu-Yun OuYang,[a,i] Bao-Yu Zhang,[a,i] Jian-Gui Guo,[b,i] Jia-Ni Liu,[c,i] Jiajian Li,[d,i] Qing-He Peng,[a] Shan-Shan Yang,[a,e] Yun He,[f] Zhi-Qiao Liu,[a] Ya-Nan Zhao,[a] Anwei Li,[d] Yi-Shan Wu,[g] Xue-Feng Hu,[b] Chen Chen,[a,j] Fei Han,[a,j] Kai-Yun You,[h,j] and Fang-Yun Xie[a,j,*]

[a]Department of Radiation Oncology, Sun Yat-sen University Cancer Center; State Key Laboratory of Oncology in South China; Collaborative Innovation Center for Cancer Medicine; Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangzhou, Guangdong, China
[b]Department of Radiation Oncology, The First People's Hospital of Foshan, Foshan, Guangdong, China
[c]Department of Head and Neck Oncology, The Cancer Center of the Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai, Guangdong, China
[d]CVTE Research, Guangzhou, Guangdong, China
[e]Department of Radiation Oncology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, China
[f]Department of Radiology, Sun Yat-sen University Cancer Center; State Key Laboratory of Oncology in South China; Collaborative Innovation Center for Cancer Medicine; Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangzhou, Guangdong, China
[g]Department of Nasopharyngeal Carcinoma, Sun Yat-sen University Cancer Center; State Key Laboratory of Oncology in South China; Collaborative Innovation Center for Cancer Medicine; Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangzhou, Guangdong, China
[h]Department of Radiation Oncology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China

eClinicalMedicine
2023;58: 101930
Published Online xxx
https://doi.org/10.1016/j.eclinm.2023.101930

Summary

Background Radiotherapy is the mainstay of treatment for nasopharyngeal carcinoma. Radiation-induced temporal lobe injury (TLI) can regress or resolve in the early phase, but it is irreversible at a later stage. However, no study has proposed a risk-based follow-up schedule for its early detection. Planning evaluation is difficult when dose-volume histogram (DVH) parameters are similar and optimization is terminated.

Methods This multicenter retrospective study included 6065 patients between 2014 and 2018. A 3D ResNet-based deep learning model was developed in training and validation cohorts and independently tested using concordance index in internal and external test cohorts. Accordingly, the patients were stratified into risk groups, and the model-predicted risks were used to develop risk-based follow-up schedules. The schedule was compared with the Radiation Therapy Oncology Group (RTOG) recommendation (every 3 months during the first 2 years and every 6 months in 3–5 years). Additionally, the model was used to evaluate plans with similar DVH parameters.

Findings Our model achieved concordance indexes of 0.831, 0.818, and 0.804, respectively, which outperformed conventional prediction models (all $P < 0.001$). The temporal lobes in all the cohorts were stratified into three groups with discrepant TLI-free survival. Personalized follow-up schedules developed for each risk group could detect TLI 1.9 months earlier than the RTOG recommendation. According to a higher median predicted 3-year TLI-free survival (99.25% vs. 99.15%, $P < 0.001$), the model identified a better plan than previous models.

Interpretation The deep learning model predicted TLI more precisely. The model-determined risk-based follow-up schedule detected the TLI earlier. The planning evaluation was refined because the model identified a better plan with a lower risk of TLI.

*Corresponding author. Department of Radiation Oncology, Sun Yat-sen University Cancer Center; State Key Laboratory of Oncology in South China; Collaborative Innovation Center for Cancer Medicine; Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, No. 651 Dongfeng East Road, Guangzhou, 510060, China.
E-mail address: xiefy@sysucc.org.cn (F.-Y. Xie).
[i]Joint first authors.
[j]Joint corresponding authors.

Funding The Sun Yat-sen University Clinical Research 5010 Program (2015020), Guangdong Basic and Applied Basic Research Foundation (2022A1515110356), Medical Scientific Research Foundation of Guangdong Province (A2022367), and Guangzhou Science and Technology Program (2023A04J1788).

Keywords: Deep learning; Early detection; Nasopharyngeal carcinoma; Planning evaluation; Radiation-induced temporal lobe injury

Research in context

Evidence before this study

We searched PubMed on February 24, 2020 for the first time and refreshed on December 20, 2022 using the terms "temporal lobe injury" AND ("nasopharyngeal carcinoma" OR "nasopharyngeal cancer") without language restriction. This search yielded 34 results, but no study focused on risk-based follow-up schedule to detect temporal lobe injury early.

Added value of this study

First, we developed and externally tested a deep learning model to precisely predict the monthly risk of temporal lobe injury based on 6065 patients with nasopharyngeal carcinoma from four hospitals. The model stratified temporal lobes into different risk groups. A subsequent risk-based follow-up schedule developed based on the model could detect temporal lobe injury 1.9 months earlier than the Radiation Therapy Oncology Group recommendation. The model could identify a better radiotherapy plan even though the plans had similar dose-volume histogram parameters and were mixed with previous models.

Implications of all the available evidence

Radiotherapy planning evaluation was improved using the deep learning model with a specific risk of temporal lobe injury instead of delphic dose–volume histogram parameters. An optimal radiotherapy plan is more feasible via replanning and reevaluation, possibly reducing the risk of injury. If there is still a high risk of injury or injury is inevitable, we suggest examining the temporal lobe on the model-determined risk-based schedule for early detection.

Introduction

Nasopharyngeal carcinoma is an epithelial carcinoma that arises from the nasopharyngeal mucosal lining. It is highly sensitive to ionizing radiation, and radiotherapy with or without chemotherapy instead of surgery is the mainstay treatment modality for this disease.[1] Survivors are always at risk of radiation-induced temporal lobe injuries. Studies[2,3] showed that edema develops first in the temporal lobes followed by contrast enhancement. Cysts or necrosis are the least frequent manifestations that occur at the late stage of injury. The injury may be asymptomatic in the early phase, but most patients may have headache (41.7%) and dizziness (38.9%).[3] These non-specific symptoms might poorly prompt a magnetic resonance imaging (MRI) if physicians do not have a high index of suspicion for temporal lobe injury. Therefore, timely diagnosis of injury is still difficult, even though patients have correlated symptoms. Unfortunately, when life-threatening epilepsy, convulsions, intracranial hemorrhage, and herniation are present,[4] contrast enhancement possibly extends to the lateral border of the temporal lobe and cysts or necrosis occur.[3] However, there are a few drugs to treat the injury, including anti-inflammatory and antioxidant stress medications (e.g., corticosteroids,[4] edaravone,[5] and biotin[6]), antiapoptotic medications (e.g., Kukoamine A[7]), and anti-microvascular injury medications (e.g., bevacizumab[8]). Non-medical therapeutics, such as hyperbaric oxygen therapy,[9] may also improve the symptoms of injury. Reducing edema is effective, but regressing the lesions and recovering impaired function is difficult or impossible when the injury progresses to cysts or necrosis. Early detection of MRI abnormalities, such as edema, made it possible to prevent the injury evolution and delay the impairment of function by early intervention.

A rational imaging follow-up schedule would be invaluable for early detection, except for improved diagnostic accuracy due to advances in imaging techniques, such as resting-state functional magnetic resonance.[10] Currently, temporal lobes are only examined incidentally in the plan to surveil tumor recurrences, such as the Radiation Therapy Oncology Group (RTOG) recommendation of 14 times and most intensive National Comprehensive Cancer Network suggestion of 27 visits. However, the injury risk may be discrepant across patients and vary over time for identical patients, and the rhythms of tumor recurrence and temporal lobe injury may differ. Therefore, early detection of temporal lobe injury, discovering the regularity of temporal lobe injury over time, stratifying patients into different risk groups, and developing a risk-based follow-up schedule may be helpful. The Wen's nomogram (age, T-stage, and $D_{0.5cc} < 65.06$ Gy),[11] MRI-derived radiomics

nomogram,[12] and our dosiomics model that utilized the spatial dose distribution and dose-volume histogram (DVH) parameters[13] showed high predictive efficiency and could stratify patients. However, they failed to predict the monthly risk of temporal lobe injury and impossibly guided individualized follow-up. A new model with precise predictions is required to solve these problems.

Additionally, optimizing a radiotherapy plan is the basis for minimizing injuries. Currently, planners' empirical effort in reducing injury is evaluated by the DVH parameters of the temporal lobes, such as the QUANTEC model (D_{max}<60 Gy)[14] or Wen's nomogram,[11] in daily clinical practice. Corresponding plan modifications may result in the optimization of DVH parameters, but the temporal lobe injury rate or grade remains unknown. Hence, the new model must be competent in the daily evaluation of treatment plans, especially when DVH parameters are similar.

Herein, we developed and tested a deep learning model based on clinical variables, DVH parameters, spatial dose distribution, and computed tomography (CT)-based features of the temporal lobes to precisely predict the monthly risk of radiation-induced temporal lobe injury. We then used the model to stratify patients and establish a risk-based follow-up schedule to detect injuries early. We also investigated whether the model could identify a better radiotherapy plan in daily clinical practice of planning evaluation.

Methods

Ethics statement

This study was approved by the Institutional Review Board of Sun Yat-sen University Cancer Center (No. B2020-264-01 and No. B2017-013-01). Patient informed consent was waived in the retrospective cohort, but was obtained in the prospective cohort. All clinical investigations were conducted in accordance with the principles of the Declaration of Helsinki.

Patients

Based on the inclusion criteria elucidated in the Supplementary Methods, 5100 patients with histopathology-confirmed nasopharyngeal carcinoma were retrospectively and consecutively included from four hospitals in China: the Sun Yat-sen University Cancer Center (SYSUCC cohort, 3710 patients, 2014/01–2015/12), Sun Yat-sen Memorial Hospital (SYSMH test cohort, 201 patients, 2015/12–2018/10), First People's Hospital of Foshan (Foshan test cohort, 1061 patients, 2014/01–2016/12), and Cancer Center of the Fifth Affiliated Hospital of Sun Yat-sen University (Zhuhai test cohort, 128 patients, 2015/11–2017/07). We randomly divided patients in the SYSUCC cohort into training (3146 patients) and validation (564 patients) cohorts to develop a deep learning model. Patients from the other three hospitals were used as external test cohorts (1390 patients) to evaluate the external generalization ability of the model. Additionally, 965 participants included in the prospective observational study (NCT03003182) were enrolled into the internal test cohort (2017/05–2017/12) (Fig. 1a). All the patients received intensity-modulated radiotherapy (IMRT), with or without systemic treatment. The details are specified in the Supplementary Methods section.

Surrogate assessment of functional temporal lobe injury

Similar to previous studies,[11,12,15] MRI-based temporal lobe radiologic changes,[2] including edema (finger-like lesions of increased signal intensity in T2-weighted images), contrast enhancement (enhanced lesions with or without necrosis in T1-weighted [T1-w] images), and cyst or necrosis (very high signal intensity on T2-weighted (T2-w) images with round or oval lesions), were used as a surrogate assessment of functional temporal lobe injury in all patients in the training, validation, internal, and external test cohorts. A radiation oncologist (OYPY, over 10 years of experience in nasopharyngeal carcinoma) and radiologist (HY, interpretation of MRI images in over 500 patients with nasopharyngeal carcinoma per month for over 5 years) independently evaluated the MRI images at each visit after radiotherapy. Disagreements in diagnosis were resolved by consensus. Representative examples are presented in Supplementary Fig. S1.

Planning CT and MRI protocols

All patients were immobilized in the supine position using a head-neck-shoulder thermoplastic mask. Planning CT images with and without contrast were obtained from the head to 2 cm below the sternoclavicular joint with 3 mm slices, a matrix size of 512 × 512, and voxel resolution of 0.97 × 0.97 × 3.0 mm in the left-right, anteroposterior, and craniocaudal directions.

All the patients underwent head and neck MRI with 1.5-T or 3.0-T unit (Siemens, Germany; Philips, Netherlands; GE, USA) for cancer diagnosis, staging, and imaging follow-up. The images were scanned using a head-neck combined coil. The scan ranged from the suprasellar cistern to the inferior margin of the sternal end of the clavicle. The MRI examinations included unenhanced T1-and T2-w sequences, and contrast-enhanced T1-weighted (CET1-w) and fat-suppressed CET1-w sequences. Contrast-enhanced sequences were obtained after intravenous injection of 0.1 mmol/kg of body weight of gadodiamide (Omniscan, GE Healthcare Ireland, Cork, Ireland) or gadolinium-diethylenetriamine penta-acetic acid (Magnevist; Bayer Schering Pharma AG, Germany).

Deep learning prediction model

Univariate analysis filtered the clinical variables (age [≤20, 21–30, 31–40, 41–50, 51–60, and >60 years],

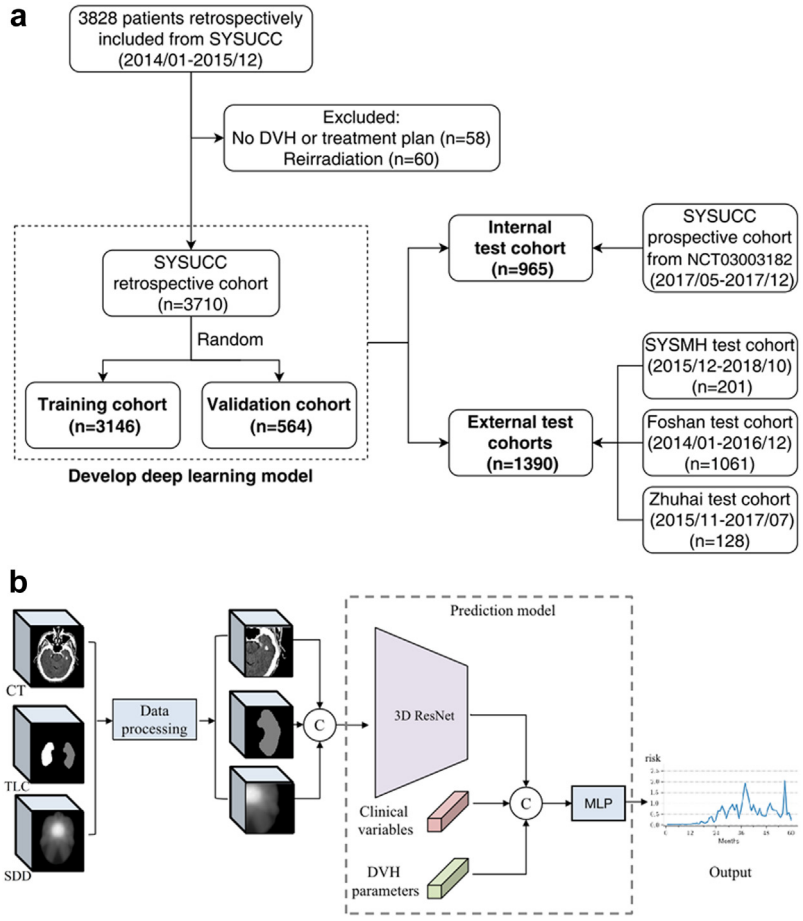

Fig. 1: Study design diagram. Patient flowchart (a) and the framework of the deep learning model for temporal lobe injury risk prediction (b). CT, computed tomography; DVH, dose-volume histogram; MLP, multi-layer perceptron; SDD, spatial dose distribution to temporal lobe; TLC, temporal lobe contours.

T-stage, and concurrent chemotherapy) and 18 DVH parameters (standardized to an equivalent dose of 2 Gy per fraction using the linear quadratic model[16]) that were predictive of temporal lobe injury-free survival (Supplementary Table S1). Baseline images of planning CT and spatial dose distribution of the temporal lobe were cropped according to their contours segmented by YSS and verified by OYPY as previously recommended.[17] After normalization, baseline images of planning CT, dose distribution, and shapes of the temporal lobe were concatenated and added into 3D ResNet[18] to extract image feature *f*. Image feature *f*, clinical variables, and 18 DVH parameters were concatenated and simultaneously added into the multi-layer perceptron to develop the deep learning model after comparing their performance in predicting the injury in the validation cohort (Supplementary Table S2). Each temporal lobe's injured or unwounded state, latencies of MRI-based temporal lobe injury after radiotherapy for the injured lobes, and intervals between the last MRI examination and radiotherapy for the unwounded lobes were used as training outcomes. The details are shown in the Supplementary Methods, and the trained model was uploaded to GitHub (https://github.com/genhao3/TLI). Finally, the prediction model outputs the monthly chance of temporal lobe injury over 60 months (Fig. 1b). The total risk R^i of *i*th instance was calculated as $R^i = \sum_{m=1}^{60} P(m) \cdot r_m$, where r_m denoted the risk of m^{th} month, and $P(.)$ was a linear function, which defined that the risk of temporal lobe injury occurring earlier was more significant than that occurring later (Supplementary Fig. S2).

Risk-based follow-up schedule

The temporal lobes in the training cohort were stratified into low-, intermediate-, and high-risk groups based on the 50% and 85% quantiles of the predicted 5-year total risk R^i by the deep learning model. The threshold corresponding to the quantiles was then tested in the other cohorts for risk stratification.

Generally, the more likely the injury would occur in a specific month, the more follow-up visits should be distributed, regardless of the risk groups. Therefore, the monthly follow-up was the product of the monthly predicted temporal lobe injury probability and total number of visits. Follow-up visits were scheduled according to cumulative monthly follow-up values. If a total of 14 visits, as recommended by the RTOG, were set for a temporal lobe, a visit would be scheduled in the month when the cumulative follow-ups exceeded 0.74, or else no visit was scheduled. Once a visit was planned for a given month, cumulative follow-ups returned to zero and follow-ups were accumulated from the next month. A temporal lobe in the high-risk group of the training cohort was used as an example, and its schedule of 14 visits is presented in Supplementary Table S3. The total number of visits ranged from the least five times (once every year) to a maximum of 27. The schedules established for each risk group in the training cohort were tested in other cohorts.

To evaluate the effectiveness of different schedules, we calculated the delayed months for temporal lobes with injury, defined as the duration from the actual injury time to the next-nearest planned follow-up visit. For example, if a temporal lobe experienced temporal lobe injury in the 38th month, while the next most immediate planned visit was in the 42nd month, temporal lobe injury follow-up was delayed by 4 months, and the delayed months were four. As we limited the analysis to 60 months, the temporal lobes were excluded if temporal lobe injury occurred beyond 60 months. Delayed months were considered the duration from the temporal lobe injury time to the 60th month if the temporal lobe injury was diagnosed after the scheduled visit. The RTOG recommendation of 14 visits (every 3 months during the first 2 years and every 6 months in years 3–5) was set as the control for comparison in the validation, internal test, and external test cohorts because it was applied at our center.

Statistical analysis

Similar to the QUANTEC model[14] and Wen's nomogram,[11] our endpoint was survival free from temporal lobe injury (temporal lobe injury-free survival). It was calculated from the start of radiotherapy to the occurrence of MRI-based temporal lobe injury or the final examination, if negative. Bilateral temporal lobes were evaluated separately, and temporal lobe-based rather than patient-based analyses were conducted. Therefore, a patient had two different temporal lobe injury-free survival values if the unilateral lobe had temporal lobe injury, or if the temporal lobe injury of the bilateral temporal lobes occurred at different times. Alternatively, a patient's temporal lobe injury-free survival value was the same for both temporal lobes.

The performance of the models in the validation, internal test, and external test cohorts was evaluated using Harrell's concordance index (C-index)[19] and compared using the Log likelihood ratio test. Calibration curves and decision curve analyses were used to evaluate the consistency between the predicted and observed risks and clinical usefulness of the models. Temporal lobe injury-free survival rates of the different risk groups were compared using the Kaplan–Meier with the log-rank test. We reported the median predicted survival rates of the two radiotherapy plans and median delayed months on different schedules given the skewed distribution identified by the Kolmogorov–Smirnov test and then compared them using the related samples Wilcoxon rank test.

Statistical analysis was performed using the R software (https://cran.r-project.org/). Statistical significance was set at a P value of <0.05 was significant.

Role of the funding source

The study funders had no role in the study design, data collection, analysis, interpretation, or writing of the report. The corresponding author had full access to all data and took final responsibility for the decision to submit for publication.

Results

Patients

As shown in Fig. 1a and Table 1, 6065 patients were included in the study. In the training, validation, internal test cohorts, SYSMH, Foshan, and Zhuhai test cohorts, 8.2% (519/6292), 7.5% (85/1128), 9.0% (174/1930), 4.0% (16/402), 6.4% (135/2122), and 5.1% (13/256) of the temporal lobes had radiation-induced injuries, respectively, and the 3-year temporal lobe injury-free survival rates were 94.4%, 95.2%, 93.0%, 97.6%, 96.3%, and 97.6%, respectively. Owing to the small number of injured temporal lobes, the SYSMH, Foshan, and Zhuhai test cohorts were grouped as external test cohorts for subsequent analysis.

Development and test of the deep learning model

The deep learning model achieved C-indexes of 0.793 (95% CI, 0.774–0.813), 0.831 (95% CI, 0.786–0.869), 0.818 (95% CI, 0.785–0.849), and 0.804 (95% CI, 0.768–0.840) in the training, validation, internal test, and external test cohorts, respectively. Calibration curves indicated consistency between the predicted and actual temporal lobe injury-free survival rates in all the cohorts (Supplementary Fig. S3). The C-indexes of the deep learning model were higher than those of Wen's model (0.769 [95% CI, 0.725–0.814], 0.769 [95% CI, 0.733–0.801], and 0.787 [95% CI, 0.754–0.816]), and the QUANTEC model (0.520 [95% CI, 0.500–0.534], 0.539 [95% CI, 0.532–0.547], and 0.551 [95% CI, 0.537–0.562]) in the validation, internal test, and external test cohorts, which showed its superiority in predicting temporal lobe injury-free survival (all P < 0.001, Fig. 2). The decision

Characteristics	Training cohort (n = 3146)	Validation cohort (n = 564)	Internal test cohort (n = 965)	SYSMH test cohort (n = 201)	Foshan test cohort (n = 1061)	Zhuhai test cohort (n = 128)
Male sex	2235/3146 (71.0%)	391/564 (69.3%)	704/965 (73.0%)	140/201 (69.7%)	747/1061 (70.4%)	90/128 (70.3%)
Median age (IQR)	45 (37–53)	45 (37–52)	45 (37–53)	45 (38–53)	48 (40–57)	48 (41–54)
Age group						
≤20	47/3146 (1.5%)	12/564 (2.1%)	15/965 (1.6%)	5/201 (2.5%)	7/1061 (0.7%)	1/128 (0.8%)
21–30	272/3146 (8.6%)	59/564 (10.5%)	82/965 (8.5%)	15/201 (7.5%)	74/1061 (7.0%)	4/128 (3.1%)
31–40	765/3146 (24.3%)	126/564 (22.3%)	235/965 (24.4%)	45/201 (22.4%)	194/1061 (18.3%)	23/128 (18.0%)
41–50	1058/3146 (33.6%)	191/564 (33.9%)	309/965 (32.0%)	69/201 (34.3%)	321/1061 (30.3%)	50/128 (39.1%)
51–60	711/3146 (22.6%)	126/564 (22.3%)	201/965 (20.8%)	45/201 (22.4%)	289/1061 (27.2%)	33/128 (25.8%)
>60	293/3146 (9.3%)	50/564 (8.9%)	123/965 (12.7%)	22/201 (10.9%)	176/1061 (16.6%)	17/128 (13.3%)
T-stage						
T1	186/3146 (5.9%)	42/564 (7.4%)	45/965 (4.7%)	30/201 (14.9%)	62/1061 (5.8%)	24/128 (18.8%)
T2	530/3146 (16.8%)	97/564 (17.2%)	115/965 (11.9%)	105/201 (52.2%)	139/1061 (13.1%)	25/128 (19.5%)
T3	1694/3146 (53.8%)	281/564 (49.8%)	573/965 (59.4%)	43/201 (21.4%)	425/1061 (40.1%)	52/128 (40.6%)
T4	736/3146 (23.4%)	144/564 (25.5%)	232/965 (24.0%)	23/201 (11.4%)	435/1061 (41.0%)	27/128 (21.1%)
N-stage						
N0	354/3146 (11.3%)	69/564 (12.2%)	73/965 (7.6%)	21/201 (10.4%)	120/1061 (11.3%)	4/128 (3.1%)
N1	1187/3146 (37.7%)	217/564 (38.5%)	331/965 (34.3%)	45/201 (22.4%)	286/1061 (27.0%)	25/128 (19.5%)
N2	1133/3146 (36.0%)	214/564 (37.9%)	410/965 (42.5%)	99/201 (49.3%)	477/1061 (45.0%)	83/128 (64.8%)
N3	472/3146 (15.0%)	64/564 (11.3%)	151/965 (15.6%)	36/201 (17.9%)	178/1061 (16.8%)	16/128 (12.5%)
Concurrent chemotherapy						
Yes	2588/3146 (82.3%)	458/564 (81.2%)	836/965 (86.6%)	160/201 (79.6%)	810/1061 (76.3%)	126/128 (98.4%)
No	558/3146 (17.7%)	106/564 (18.8%)	129/965 (13.4%)	41/201 (20.4%)	251/1061 (23.7%)	2/128 (1.6%)
Median follow-up (IQR)	40 (13–60)	41 (13–60)	37 (16–50)	32 (17.5–46)	40 (13–59)	49 (25–62)
Temporal lobe injury[a]						
Yes	519/6292 (8.2%)	85/1128 (7.5%)	174/1930 (9.0%)	16/402 (4.0%)	135/2122 (6.4%)	13/256 (5.1%)
No	5773/6292 (91.8%)	1043/1128 (92.5%)	1756/1930 (90.9%)	386/402 (96.0%)	1987/2122 (93.6%)	243/256 (94.9%)
Temporal lobe injury-free survival[a]						
3-year (95% CI)	94.4% (93.6%-95.0%)	95.2% (93.3%-96.5%)	93.0% (91.4%-94.3%)	97.6% (95.2%-98.8%)	96.3% (95.1%-97.2%)	97.6% (94.3%-99.0%)
5-year (95% CI)	86.0% (84.7%-87.2%)	87.2% (84.2%-89.6%)	76.5% (69.7%-82.0%)	91.4% (84.3%-95.4%)	88.8% (86.7%-90.6%)	93.1% (87.6%-96.2%)

CI = confidence interval; IQR = interquartile range; NA = not available. [a]Bilateral temporal lobes of each patient were evaluated separately.

Table 1: **Basic characteristics of patients in different cohorts.**

curve analysis supported the considerable net benefit of the deep learning model (Supplementary Fig. S4).

Deep learning model in risk-based follow-up schedule

The training cohort was stratified into three groups, 2.0% (66/3146), 10.8% (240/2202), and 22.5% (213/944), of which the temporal lobes had radiologic changes, and the 5-year temporal lobe injury-free survival rate of group I was the highest (96.81%), followed by group II (80.64%, $P < 0.001$) and group III (53.85%, $P < 0.001$) (Fig. 3a). Based on the same stratification threshold, significant differences in the 5-year temporal lobe injury-free survival rates were also observed between the three groups in the validation, internal, and external test cohorts (all $P < 0.001$, Fig. 3b–d).

As predicted by the deep learning model, the incidence of temporal lobe injury was extremely low for the temporal lobes at low risk, and the curves in all the cohorts were relatively flat with no prominent peak. The incidence of the temporal lobes at intermediate risk was much higher than that at low risk, especially 2 years after radiotherapy. Nonetheless, temporal lobes at increased risk may experience radiologic changes even within 2 years, and they have a much higher incidence of radiation-induced injury than those at intermediate risk. A sharp rise in the probability of temporal lobe injury over time was observed in the high-risk group (Supplementary Fig. S5).

Subsequently, we assigned multiple visits (5–27) to specific months according to the predicted monthly temporal lobe injury probabilities. Consequently, the median number of delayed months gradually decreased as the total number of assigned visits increased in all the risk groups, as shown in Fig. 4. Significantly, if the high-risk group's temporal lobes were followed up 14 times on the deep learning-based schedule, the median delayed months were shortened to 1.3, 1.0, and 1.4

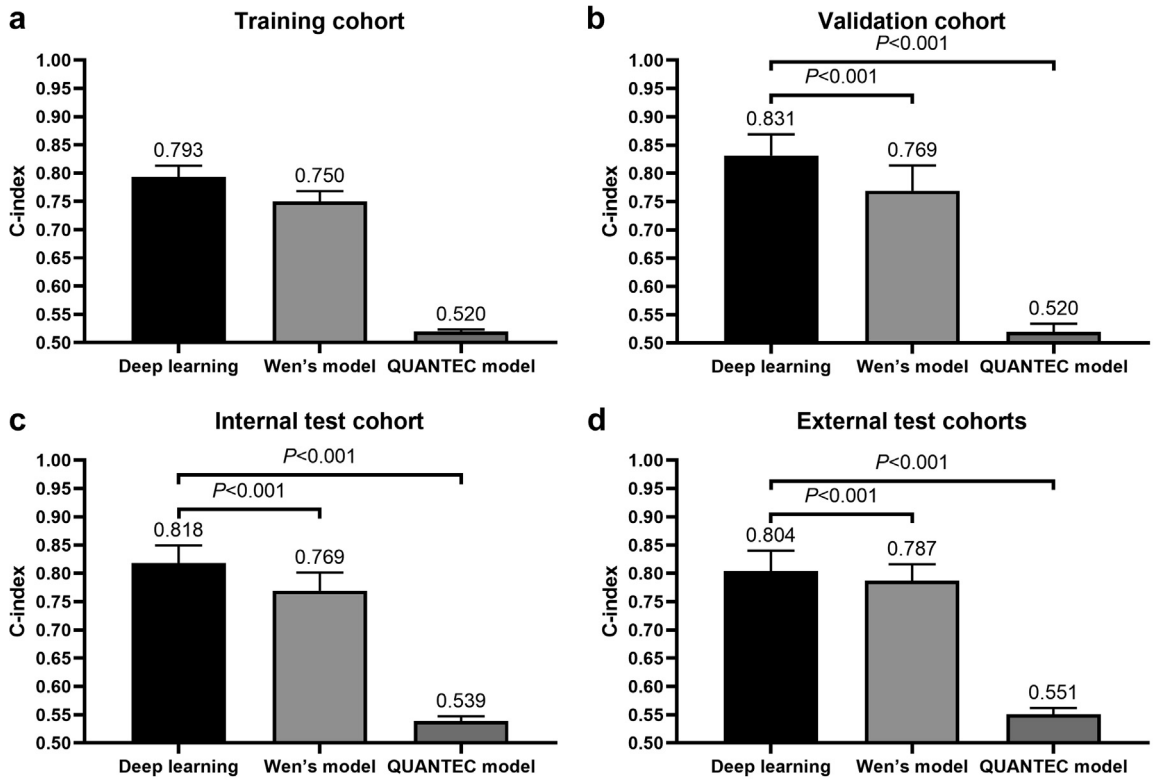

Fig. 2: Harrell's C-index of different models in predicting the risk of temporal lobe radiologic changes in the training (a), validation (b), internal test (c), and external test (d) cohorts.

months in the validation, internal test, and external test cohorts, respectively. In contrast, the median delayed months were 2.8, 2.0, and 3.1 months on the RTOG-recommended schedule ($P \leq 0.002$). Additionally, the superiority of the deep learning-based schedule in earlier detection of temporal lobe injury was also observed for temporal lobes at intermediate and low risks, despite not being significantly distinct from that of the RTOG schedule in the external test cohorts (1.6 vs. 1.9, $P = 0.071$ and 1.8 vs. 2.3, $P = 0.084$, respectively). Therefore, the deep learning model highly suggests imaging follow-up on the risk-based schedule, as depicted in Fig. 5.

Deep learning model in daily radiotherapy plan evaluation

We randomly selected 100 patients from the internal test cohort to mimic the application of the deep learning model in daily planning evaluation. Their basic characteristics are listed in Supplementary Table S4. A new radiotherapy plan was simulated with the same dose prescription for the tumor and dose constraints to the organs at risk as the clinically used plan for each patient. As shown in Fig. 6a, the two plans had similar standardized D_{max} and $D_{0.5cc}$ but not the same dose distribution for the identical temporal lobe of the representative example. Consequently, both the Wen's and QUANTEC models failed to choose the right plans for the 100 patients because of the unchangeable age and T-stage and similar classification of the standardized $D_{0.5cc}$ (</≥65.06 Gy) or D_{max} (</≥60 Gy) to the same temporal lobe. However, as predicted by the deep learning model, the monthly risk of temporal lobe injury differed between the plans (Fig. 6b). As the accuracy of the deep learning model in predicting 3-year temporal lobe injury-free survival rates had been demonstrated above, the 100 clinically used plans generated 200 model-predicted values for these patients, as did the 100 simulated plans. Given that the median predicted values by the 100 clinically used plans were significantly higher than those by the 100 simulated plans using the related samples' Wilcoxon signed-rank test (99.25% [96.86–99.90%] vs. 99.15% [97.12–99.90%, $P < 0.001$) (Fig. 6c), the deep learning model recommended the clinically used plans for the 100 patients.

Discussion

Based on 6065 patients from four hospitals, we developed and tested a deep learning model that could precisely predict the risk of radiation-induced temporal lobe injury with a C-index of 0.793 (95% CI 0.774–0.813).

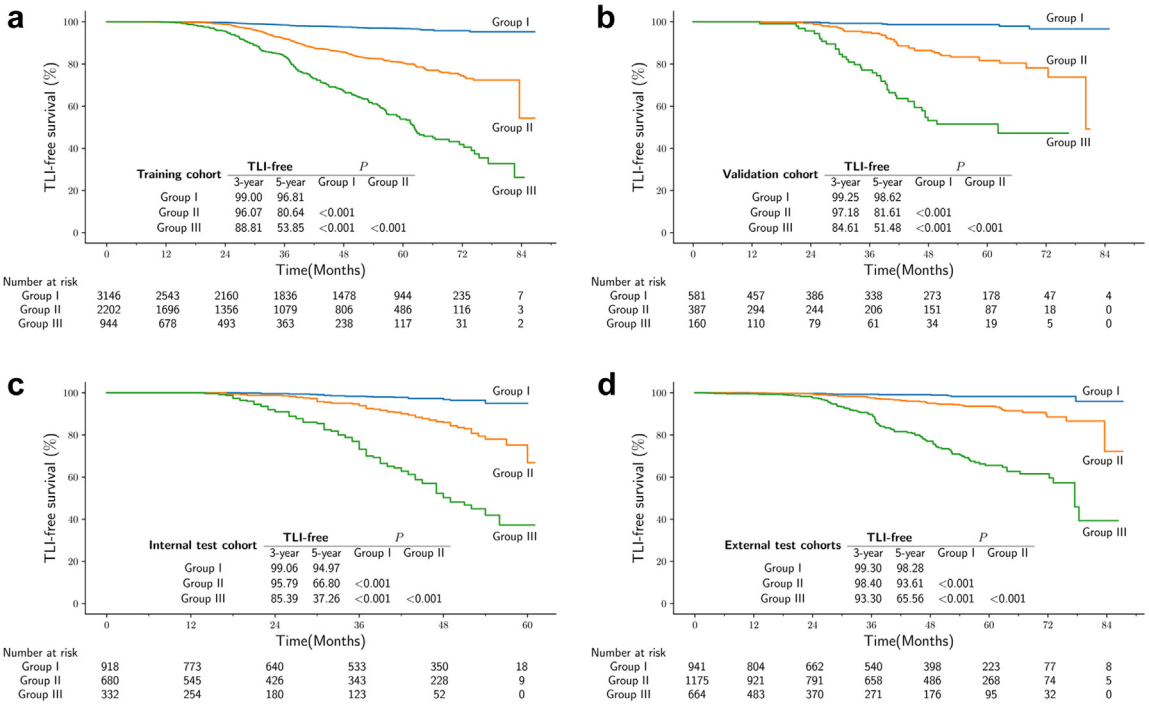

Fig. 3: Kaplan–Meier curves for temporal lobe injury (TLI)-free survival in the different risk groups stratified by the deep learning model in the training (a), validation (b), internal test (c), and external test (d) cohorts. Group I, low-risk group; Group II, intermediate-risk group; Group III, high-risk group.

The developed model performed much better than the clinically used QUANTEC model and Wen's nomogram ($P < 0.001$) in the validation, internal test, and external test cohorts. The temporal lobes were stratified into three groups with discrepancies in the temporal lobe injury-free survival (e.g., 96.81% vs. 80.64% vs. 53.85%, $P < 0.001$). On the follow-up schedules developed for each risk group, temporal lobe injury could be detected earlier than using the RTOG recommendation with shortened median delayed months, especially for temporal lobes at intermediate (e.g., 1.3 vs. 3.1, $P < 0.001$) and high (e.g., 1.4 vs. 3.1, $P < 0.001$) risks. Besides, this model could also identify a better radiotherapy plan with a higher median predicted 3-year temporal lobe injury-free survival rate (99.25% vs. 99.15%, $P < 0.001$), which current models mixed.

Radiation mainly determines the occurrence and latency of temporal lobe injuries in nasopharyngeal carcinoma patients. A patient's bilateral lobes may have injuries at different times, or only a unilateral temporal lobe injury is observed. Therefore, bilateral lobe injuries do not totally depend on each other, even though they come from the same patient. Radiation oncologists independently evaluated DVH parameters of the bilateral temporal lobes by default in clinical practice. Hence, the temporal lobe injury-free survival rates for bilateral temporal lobes were analyzed separately in our study, similar to all previous studies.[3,11–13,15,16,20,21] Notably, the training and validation cohorts were randomly divided by patients instead of the temporal lobes to avoid the repeated use of clinical variables in both training and validation. Age, chemotherapy, intrinsic susceptibility to radiation, and other diseases of a patient may simultaneously affect injury of the bilateral temporal lobes and result in correlated data. However, this effect is secondary and impossible to be eliminated. We selected the clinical variables in the univariate analysis and then added them into the deep learning model. The temporal lobe-based analysis showed outcomes consistent with observations in clinical practice; thus, the resultant deep learning model was more likely to be applied by clinicians.

Similar to radiotherapy-associated pulmonary toxicity,[22] the probability of temporal lobe injury gradually increased over time, as shown in Supplementary Fig. S5. In contrast, the peak time of tumor recurrence or metastasis was around 15th month after radiotherapy.[23] Rhythms of the temporal lobe injury and tumor recurrence were quite different. As shown in a prior study,[11] we also observed heterogeneous injury risks across the temporal lobes of different patients. Consistent with the findings of a surveillance strategy for tumor recurrence,[23] delays in detecting temporal lobe injury would decrease with increment of follow-up

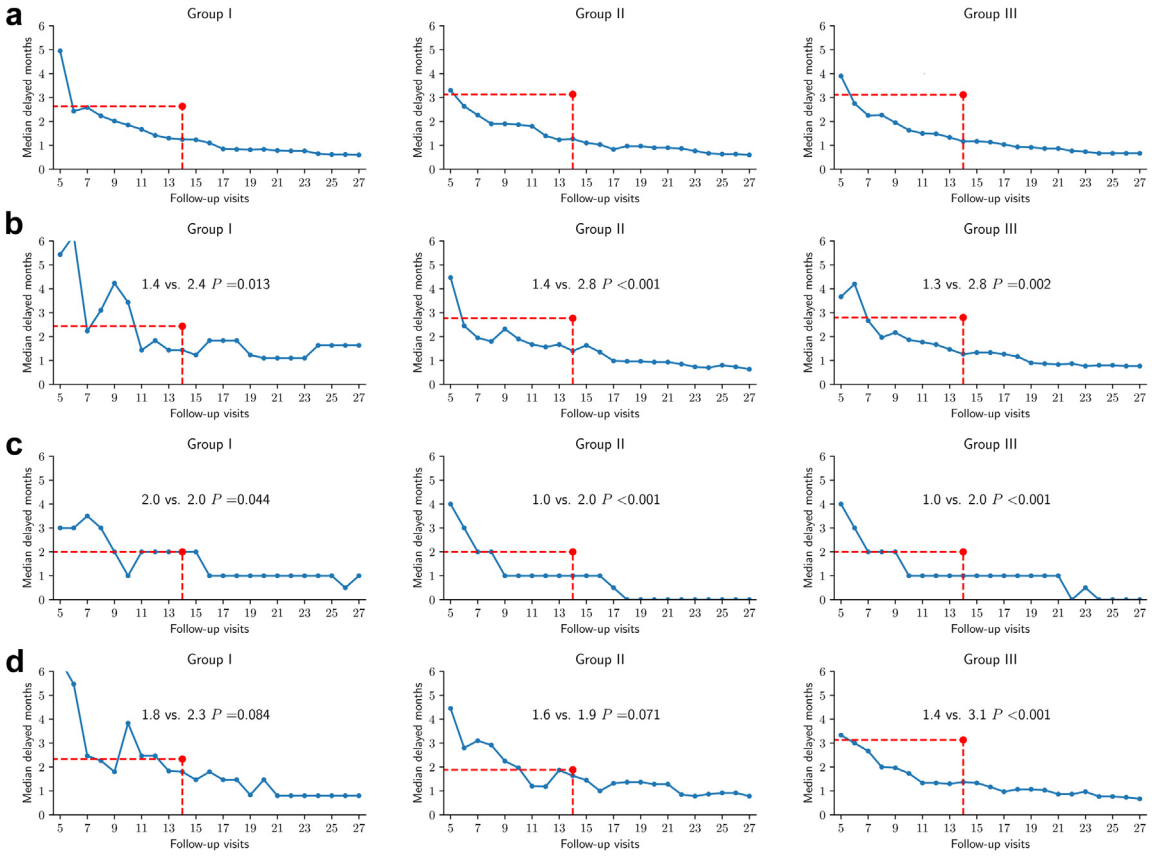

Fig. 4: Median delays in the detection of temporal lobe injury (TLI) by different follow-up schedules in the other risk groups of the training (a), validation (b), internal test (c), and external test (d) cohorts. Blue points, deep learning model-based schedules with the number of visits ranging between 5 and 27; red point, Radiation Therapy Oncology Group recommended schedule with 14 visits.

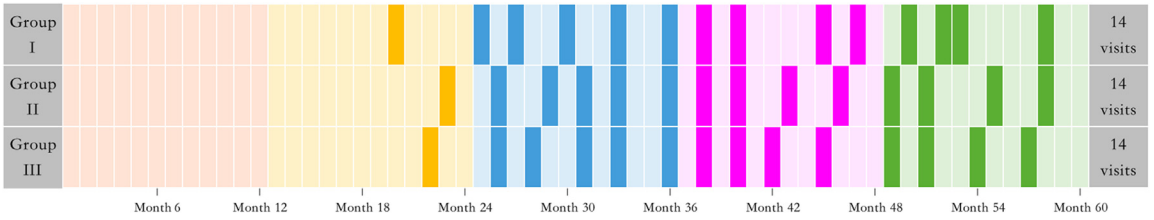

Fig. 5: Recommended imaging follow-up schedules for temporal lobes at different risks based on the deep learning model. Twelve boxes of the grid side by side were painted with the same color to represent 12 months of a year. The boxes with deepened color represent the months when imaging follow-up was recommended for the patient.

visits, regardless of the risk group. Importantly, our deep learning model, for the first time, precisely predicted the monthly risk of temporal lobe injury, and then specially formulated the risk-based follow-up schedule for early injury detection. Concerning the currently adopted RTOG follow-up recommendation, our deep learning model-based schedule could significantly reduce delays in detecting temporal lobe injury at the same 14 visits for temporal lobes, especially for those at intermediate and high risk. Consequently, the injury was likely to be seen 1.9 months earlier if the risk-based schedule was followed, independent of the schedule for surveilling tumor recurrence. Indeed, the superiority in earlier detection was slight in the low-risk group, possibly because of the rare occurrence within 60 months after radiotherapy. Notably, if two independent schedules were executed to survey the injury and tumor recurrence, the inconvenience of frequent visits and

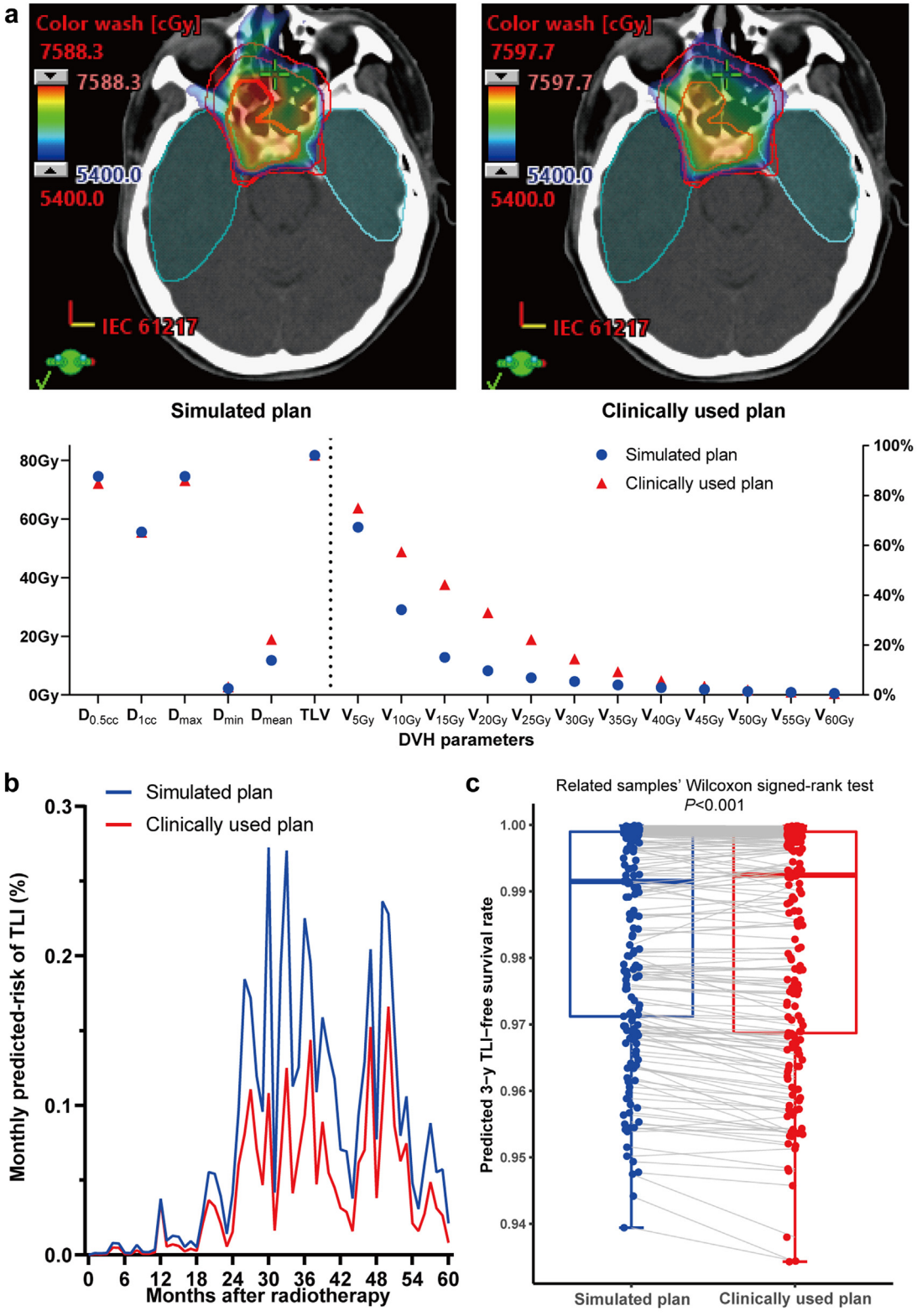

Fig. 6: Two plans with similar D_{max} and $D_{0.5cc}$, but different spatial dose distribution (a), and discrepant monthly risks of temporal lobe injury (TLI) predicted by the deep learning model (b) for an example, and comparison of their median 3-year TLI-free survival rate predicted by the deep learning model for the 100 randomly selected patients (c).

additional costs would be inevitable. Perhaps it would be more advisable for high-risk patients to consider socioeconomic issues.

Since 93.8% (11,372/12,130) of the temporal lobes were irradiated with D_{max} >60 Gy, as observed in our study, the QUANTEC model is incompetent in evaluating and optimizing the plans in the clinic. Other DVH parameters, such as $D_{0.5cc}$, D_{1cc}, and V_{60Gy},[16,20,21,24,25] or their combinations with clinical variables (e.g., age, T-stage)[11,26] performed better in forecasting the risk, despite their cutoff values varying across different studies and lacking external verification. Recently, MRI-based radiomics features of the temporal lobes were extracted into the prediction model, which showed persuasive prediction performance.[12,15] Radiomics features may correspond to the intrinsic susceptibility of the temporal lobe to radiation, as demonstrated by the Genome-Wide Association Study.[27] Nonetheless, only DVH parameters and other information derived from the radiotherapy plan were factors that clinicians could tune. Studies have indicated that dose distribution could complement DVH parameters in predicting radiation-induced xerostomia[28] and gastrointestinal and genitourinary toxicities.[29] Its role in predicting temporal lobe injury was investigated in a recent study.[13] In contrast, we used the deep learning algorithm in the present study to integrate multidimensional data, including clinical variables, DVH parameters,[11,16,30,31] CT-derived radiomics features, and spatial dose distribution. As a result, our deep learning model outperformed both the Wen's and QUANTEC models in prediction. As shown in Fig. 6a, the traditional planning optimizers would generate plans with similar DVH parameters. With the unchangeable age, T-stage, or other clinical and tumor factors for the same patient, previous models[11,12,14] failed to identify the best plan and impossibly guided corresponding modifications. Instead, our deep learning model significantly predicted the different median temporal lobe injury-free survival rates for the two plans, testified again the value of utilizing the ignored spatial dose distribution and CT-based radiomics, and showed how to assist clinicians in choosing the optimal plan. Although it is still unclear how CT-based radiomics features and spatial dose distribution should be implemented in the conventional planning optimizer, it is an essential step in the right direction for automated planning. With the development of deep learning-based automated planning, patient-specific dose distribution is available,[32–34] and so is dose distribution-based auto-optimization.[35]

The main limitation was that the deep learning model was not available for predicting injuries beyond 60 months. However, 88.4% (459/519), 91.8% (78/85), 98.9% (172/174), and 87.8% (144/164) of injuries in the training, validation, internal test, and external test cohorts, respectively, occurred within 60 months. Some patients were followed-up for <3 years or even 18 months, which is a limitation of the retrospective design of this study. However, the median follow-up time for these cohorts was >36 months. According to a previous study with a median follow-up time of 76 months,[16] 93% (27/29) of patients developed injury within 60 months, and the median latency from IMRT until the first temporal lobe injury was 33 months. Second, due to the shallow flat risk of injury over time and insufficient representative cases in group I, the schedule for this group was greatly scattered and possibly had considerable bias. Third, temporal lobe injury was defined by radiologic changes rather than functional deficits, which may have overestimated the incidence of this toxicity, but benefited early diagnosis and screening. Moreover, temporal lobe injury was not stratified by severity to simplify the analysis on one hand and the other because edema and contrast enhancement are likely to remain static, regress, or even resolve if intervention is in time.[2] Computed tomography-based radiomics features were extracted into the model instead of MRI-based radiomics features because MRI images were unavailable in the external cohorts. Since it was unknown which parameters were actually utilized by the deep learning model, the results of comparing the multimodal models using the Log likelihood ratio test might be biased. But C-indexes already indicated the merit of these models. Most patients came from endemic regions; therefore, the findings need further confirmation in other non-endemic areas.

In conclusion, the deep learning model based on clinical variables, DVH parameters, spatial dose distribution, and CT-based radiomics features predicted the risk of MRI-based temporal lobe injury more precisely. The personalized risk-based imaging follow-up schedule determined by the model can detect the injury earlier. According to the model prediction, patients should be examined on the model-determined schedule if the injury has a high risk. The model provides a new way to evaluate radiotherapy plans in daily clinical practice with a specific risk of temporal lobe injury instead of delphic dose–volume histogram parameters. Therefore, a better radiotherapy plan with a reduced injury risk would be more feasible via corresponding replanning and reevaluation by planners and radiation oncologists.

Contributors

Study design: PYOY and FYX.
Data collection: PYOY, BYZ, JGG, JNL, SSY, ZQL, YNZ, QHP, and KYY.
Data analysis: PYOY, SSY, YH, AWL, and JJL.
Writing original draft: PYOY.
Literature search: PYOY, CC, YSW.
Resources: JNL, XFH, FH, KYY, and FYX.
Funding acquisition: FYX, PYOY, and KYY.
Supervision: FYX.
Review & editing: All authors.

Data sharing statement
The data underlying this article are available in the Research Data Deposit (RDD) public platform (www.researchdata.org.cn).

Declaration of interests
None exist.

Acknowledgements
This work was funded by the Sun Yat-sen University Clinical Research 5010 Program (2015020), Guangdong Basic and Applied Basic Research Foundation (2022A1515110356), Medical Scientific Research Foundation of Guangdong Province (A2022367), and Guangzhou Science and Technology Program (2023A04J1788).

Appendix A. Supplementary data
Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2023.101930.

References
1 Chen YP, Chan ATC, Le QT, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. *Lancet*. 2019;394:64–80.
2 Wang YX, King AD, Zhou H, et al. Evolution of radiation-induced brain injury: MR imaging-based study. *Radiology*. 2010;254:210–218.
3 Mao YP, Zhou GQ, Liu LZ, et al. Comparison of radiological and clinical features of temporal lobe necrosis in nasopharyngeal carcinoma patients treated with 2D radiotherapy or intensity-modulated radiotherapy. *Br J Cancer*. 2014;110:2633–2639.
4 Lam TC, Wong FC, Leung TW, Ng SH, Tung SY. Clinical outcomes of 174 nasopharyngeal carcinoma patients with radiation-induced temporal lobe necrosis. *Int J Radiat Oncol Biol Phys*. 2012;82:e57–e65.
5 Tang Y, Rong X, Hu W, et al. Effect of edaravone on radiation-induced brain necrosis in patients with nasopharyngeal carcinoma after radiotherapy: a randomized controlled trial. *J Neuro Oncol*. 2014;120:441–447.
6 Abdel-Magied N, Shedid SM, Ahmed AG. Mitigating effect of biotin against irradiation-induced cerebral cortical and hippocampal damage in the rat brain tissue. *Environ Sci Pollut Res Int*. 2019;26:13441–13452.
7 Zhang Y, Cheng Z, Wang C, Ma H, Meng W, Zhao Q. Neuroprotective effects of kukoamine a against radiation-induced rat brain injury through inhibition of oxidative stress and neuronal apoptosis. *Neurochem Res*. 2016;41:2549–2558.
8 Zhuang H, Shi S, Yuan Z, Chang JY. Bevacizumab treatment for radiation brain necrosis: mechanism, efficacy and issues. *Mol Cancer*. 2019;18:21.
9 Co J, De Moraes MV, Katznelson R, et al. Hyperbaric oxygen for radiation necrosis of the brain. *Can J Neurol Sci*. 2020;47:92–99.
10 Fu G, Xie Y, Pan J, et al. Longitudinal study of irradiation-induced brain functional network alterations in patients with nasopharyngeal carcinoma. *Radiother Oncol*. 2022;173:277–284.
11 Wen DW, Lin L, Mao YP, et al. Normal tissue complication probability (NTCP) models for predicting temporal lobe injury after intensity-modulated radiotherapy in nasopharyngeal carcinoma: a large registry-based retrospective study from China. *Radiother Oncol*. 2021;157:99–105.
12 Hou J, Li H, Zeng B, et al. MRI-based radiomics nomogram for predicting temporal lobe injury after radiotherapy in nasopharyngeal carcinoma. *Eur Radiol*. 2022;32:1106–1114.
13 Yang SS, OuYang PY, Guo JG, et al. Dosiomics risk model for predicting radiation induced temporal lobe injury and guiding individual intensity modulated radiation therapy. *Int J Radiat Oncol Biol Phys*. 2023;115:1291–1300.
14 Bentzen SM, Constine LS, Deasy JO, et al. Quantitative analyses of normal tissue effects in the clinic (QUANTEC): an introduction to the scientific issues. *Int J Radiat Oncol Biol Phys*. 2010;76:S3–S9.
15 Bao D, Zhao Y, Li L, et al. A MRI-based radiomics model predicting radiation-induced temporal lobe injury in nasopharyngeal carcinoma. *Eur Radiol*. 2022;32:6910–6921.
16 Zeng L, Huang SM, Tian YM, et al. Normal tissue complication probability model for radiation-induced temporal lobe injury after intensity-modulated radiation therapy for nasopharyngeal carcinoma. *Radiology*. 2015;276:243–249.
17 Sun Y, Yu XL, Luo W, et al. Recommendation for a contouring method and atlas of organs at risk in nasopharyngeal carcinoma patients receiving intensity-modulated radiotherapy. *Radiother Oncol*. 2014;110:390–397.
18 He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition, 27-30 June 2016. In: *2016 IEEE conference on computer vision and pattern recognition (CVPR)*. 2016:770–778.
19 Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982;247:2543–2546.
20 Su SF, Huang Y, Xiao WW, et al. Clinical and dosimetric characteristics of temporal lobe injury following intensity modulated radiotherapy of nasopharyngeal carcinoma. *Radiother Oncol*. 2012;104:312–316.
21 Lu L, Sheng Y, Zhang G, et al. Temporal lobe injury patterns following intensity modulated radiotherapy in a large cohort of nasopharyngeal carcinoma patients. *Oral Oncol*. 2018;85:8–14.
22 Briere TM, Agrusa JE, Martel MK, et al. Acute and late pulmonary effects after radiation therapy in childhood cancer survivors: a PENTEC comprehensive review. *Int J Radiat Oncol Biol Phys*. 2022. https://doi.org/10.1016/j.ijrobp.2022.01.052.
23 Zhou GQ, Wu CF, Deng B, et al. An optimal posttreatment surveillance strategy for cancer survivors based on an individualized risk-based approach. *Nat Commun*. 2020;11:3872.
24 Hsiao KY, Yeh SA, Chang CC, Tsai PC, Wu JM, Gau JS. Cognitive function before and after intensity-modulated radiation therapy in patients with nasopharyngeal carcinoma: a prospective study. *Int J Radiat Oncol Biol Phys*. 2010;77:722–726.
25 Zeng L, Tian YM, Sun XM, et al. Late toxicities after intensity-modulated radiotherapy for nasopharyngeal carcinoma: patient and treatment-related risk factors. *Br J Cancer*. 2014;110:49–54.
26 Zhang P, Guan W, Huang X, Xie K, Kang L. Development and validation of a nomogram for predicting radiation-induced temporal lobe injury in nasopharyngeal carcinoma. *Int J Radiat Oncol Biol Phys*. 2019;105:E398.
27 Wang TM, Shen GP, Chen MY, et al. Genome-wide association study of susceptibility loci for radiation-induced brain injury. *J Natl Cancer Inst*. 2019;111:620–628.
28 Men K, Geng H, Zhong H, Fan Y, Lin A, Xiao Y. A deep learning model for predicting xerostomia due to radiation therapy for head and neck squamous cell carcinoma in the RTOG 0522 clinical trial. *Int J Radiat Oncol Biol Phys*. 2019;105:440–447.
29 Rossi L, Bijman R, Schillemans W, et al. Texture analysis of 3D dose distributions for predictive modelling of toxicity rates in radiotherapy. *Radiother Oncol*. 2018;129:548–553.
30 Miao Y, Ou X, Wang J, et al. Development and validation of a model for temporal lobe necrosis based on 749 nasopharyngeal carcinoma patients following IMRT. *Int J Radiat Oncol Biol Phys*. 2017;99:S165–S166.
31 Feng M, Huang Y, Fan X, Xu P, Lang J, Wang D. Prognostic variables for temporal lobe injury after intensity modulated-radiotherapy of nasopharyngeal carcinoma. *Cancer Med*. 2018;7:557–564.
32 Chen X, Men K, Li Y, Yi J, Dai J. A feasibility study on an automated method to generate patient-specific dose distributions for radiotherapy using deep learning. *Med Phys*. 2019;46:56–64.
33 Liu Z, Fan J, Li M, et al. A deep learning method for prediction of three-dimensional dose distribution of helical tomotherapy. *Med Phys*. 2019;46:1972–1983.
34 Fan J, Wang J, Chen Z, Hu C, Zhang Z, Hu W. Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique. *Med Phys*. 2019;46:370–381.
35 Li G, Wu X, Ma X. Artificial intelligence in radiotherapy. *Semin Cancer Biol*. 2022;86(Pt2):160–171.