Clinical and Translational Radiation Oncology 52 (2025) 100946

Contents lists available at ScienceDirect

Clinical and Translational Radiation Oncology

journal homepage: www.sciencedirect.com/journal/clinical-and-translational-radiation-oncology

Original Research Article

Predicting radiation-induced hypothyroidism in nasopharyngeal carcinoma patients using a deep learning model

Yichen Mao [a], Mingjun Ding [a], Dan Zong [a], Zhongde Mu [a,*], Xia He [a,b,*]

[a] The Affiliated Cancer Hospital of Nanjing Medical University & Jiangsu Cancer Hospital & Jiangsu Institute of Cancer Research, Nanjing, China
[b] Department of Environmental Genomics, Jiangsu Key Laboratory of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, Nanjing Medical University, Nanjing, China

ARTICLE INFO

Keywords:
Nasopharyngeal carcinoma
Radiation-induced hypothyroidism
Deep learning

ABSTRACT

Background: Radiation-induced hypothyroidism (RIHT) is a common complication in nasopharyngeal carcinoma patients. Predicting its onset is crucial for effective management and early intervention. This study aims to develop a model based on deep learning survival analysis to predict RIHT in nasopharyngeal carcinoma patients.
Methods: This retrospective study included 535 nasopharyngeal carcinoma patients between January 2015 and October 2020. Cox regression, LASSO-Cox analyses and Spearman correlation test were employed to identify significant predictors. Two deep learning and two machine learning algorithms were trained, tuned, and compared against traditional Cox and NTCP models by C-index, Brier score, and decision curve analysis.
Results: The study observed a 41.7 % incidence of RIHT, with a median time to onset of 15 months. AJCC N stage, thyroid volume and specific dose-volume parameters were identified as potential predictors. DeepSurv model outperformed traditional ones (C-index: DeepSurv 0.75, traditional models $\leq$ 0.63). While other models were competitive at early post-treatment intervals, deep learning models demonstrated superior performance over time. Calibration and decision curve analysis corroborated the enhanced predictive capability of DeepSurv. Feature importance analysis highlighted thyroid V30 and V50 as the most significant predictors.
Conclusions: DeepSurv demonstrated superior predictive performance for RIHT in nasopharyngeal carcinoma patients compared to traditional models. Deep learning-based predictions offer high accuracy, which may enable personalized patient management and have great potentials in mitigating the risk of RIHT. These findings suggested that incorporating such model into clinical practice could be beneficial for the management of RITH.

1. Introduction

Radiation-induced hypothyroidism (RIHT) is a common complication following treatment for nasopharyngeal carcinoma, occurring in an estimated 10–50 % of patients treated with radiation in the head and neck area [1]. Despite the widespread adoption of intensity-modulated radiotherapy (IMRT) as the preferred treatment for nasopharyngeal carcinoma, the anticipated decline in the incidence of RIHT has not been observed compared to those with three-dimensional conformal radiotherapy [2]. While hypothyroidism could be managed with thyroxine supplementation, the clinical significance of such complication is underappreciated due to its insidious onset.

Personalized treatment planning and dosimetric evaluation *via* dose–volume histograms (DVHs) is integral to mitigating the risk of radiation-induced late effect. Several dosimetric predictors for hypothyroidism have been identified, including the mean thyroid dose and various volume threshold, specifically V30, V40, V45, V50 (where 'Vx' denotes the percentage of the thyroid volume receiving at least 'x' Grays [Gy]) [3–13], and VS45, VS50, VS60 (where 'VSx' denotes the volume of the thyroid spared from exceeding doses of 'x' Gy) [14–16]. Besides, it was also reported that clinical factors such as sex, age, AJCC N stage, chemotherapy, and pre-treatment thyroid volume are also implicated in the development of RIHT [17–20]. The identification of such predictors had enabled the development of normal tissue complication probability (NTCP) models for RIHT. Notably, the Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC) report has yet to endorse specific guidelines for thyroid dose constraint. The quest for optimal dosimetric predictors continues to be a subject of debate.

It is crucial to accurately predict hypothyroidism and optimize treatment regimens in patients with nasopharyngeal carcinoma.

* Corresponding authors.
E-mail addresses: mouzhongde@jszlyy.com.cn (Z. Mu), hexiabm@163.com (X. He).

https://doi.org/10.1016/j.ctro.2025.100946
Received 21 May 2024; Received in revised form 5 October 2024; Accepted 9 March 2025
Available online 13 March 2025

However, this pursuit faces several challenges in clinical practice. For example, RIHT often manifests as a long-term complication, and surfaces years after treatment. Studies with abbreviated follow-up periods risk prematurely categorizing patients as free from RIHT. A recent systematic review, which found a median follow-up of 32 months in contemporary studies, highlighted the potential bias from such inadequate follow-up durations [21]. Furthermore, NTCP models based on logistic regression, although with the features of simplicity and practicality, might not sufficiently address the intricacies of time-dependent outcomes, or capture complex non-linear relationships between predictors and outcomes.

To address the above issues, our study sought to compile a cohort with an extended follow-up period and to employ survival analysis to investigate the relationship between covariates and the onset of RIHT. Survival analysis could be a complement to traditional algorithms by incorporating the temporal aspects of risk, and thereby provide a more sophisticated approach to understand and manage RIHT. The rapid development of deep learning has demonstrated substantial capabilities for discerning non-linear patterns and facilitating individualized prognostic predictions in medical fields. To the best of our knowledge, the prediction of RIHT using deep learning survival analysis is underreported. The aim of our study was to develop a deep learning survival model that elevates the accuracy of predicting RIHT and offer insights into clinical management strategies.

2. Material and methods

This study received ethical approval (KY-2024-036) from our hospital's Institutional Review Board. Owing to the retrospective character of the study, the necessity for informed consent was waived.

2.1. Study design and patient cohort

Patients subjected to this study were diagnosed with nasopharyngeal carcinoma via biopsy and treated with definitive IMRT at our facility, with and without systemic therapy. Exclusion criteria included patients with pre-existing thyroid disorders, those under the age of 18, those with a history of previous surgical intervention, and those who had previously received irradiation to the head and neck area. Furthermore, individuals with a history of other malignancies were also excluded due to the potential impair of platinum-based chemotherapy and immunotherapy to the hypothalamic-pituitary-thyroid axis. Additionally, those with a follow-up period shorter than 36 months were omitted to avoid the bias of insufficient follow-up time. The remaining patients were then randomly allocated to either a training or a test cohort at a ratio of 4:1 (Fig. 1).

2.2. Treatment and follow-up

All participants underwent definitive IMRT, which was delivered in daily fractions for five days per week by a linear accelerator (6 MV). For accurate and reproducible patient positioning, thermoplastic head-and-neck masks extending to the shoulders were used for immobilization. Computed tomography (CT) simulations were conducted for each patient, and the CT images were featuring a slice thickness of 1.5 mm. These images were then transferred to the treatment planning system (TPS) for delineation.

The prescribed radiotherapy doses were as follows: 66–72 Gy over 28–33 fractions targeted at the primary tumor (GTVnx) and 64–70 Gy over 28–33 fractions directed at the involved lymph nodes (GTVnd). The high-risk clinical target volume (CTV1) received 60–63 Gy, and the low-risk clinical target volume (CTV2) received 54–56 Gy, both over 28–33 fractions. A margin of 3–5 mm was added to the CTV to create the planning target volume (PTV). No stringent dose constraints were applied to the thyroid and pituitary glands in TPS. Each IMRT plan was meticulously crafted using a standard coplanar nine-field technique, with the gantry angles evenly distributed. Patients with stage I disease generally underwent radiotherapy alone, whereas those with stage II-IV disease were treated with platinum-based concurrent chemotherapy, supplemented with induction or adjuvant chemotherapy based on the National Comprehensive Cancer Network (NCCN) guideline.

Clinical follow-up was scheduled at intervals of every three months for the initial two years post-radiation, biannually from the third to the fifth year, and thereafter once annually. Thyroid function test, including serum thyroid-stimulating hormone (TSH), free triiodothyronine (fT3), and free thyroxine (fT4) levels, were conducted prior to radiotherapy and at each follow-up. The normal ranges were 0.27–4.2 mIU/L for TSH, 12–22 pmol/L for fT4, and 3.1–6.8 pmol/L for fT3.

Primary hypothyroidism was characterized by an elevated serum TSH level, with or without a reduction in serum fT4 concentration. Central hypothyroidism was diagnosed when a low serum TSH level occurred alongside a fT4 level below the normal range. The onset of RIHT was delineated by the interval from the completion of radiotherapy to the first instance of abnormal thyroid function meeting the aforementioned diagnostic criteria as endpoint. Patients were also censored at the date of re-irradiation, initiation of treatment for recurrence or metastasis, or at the date of the diagnosis of an additional malignancy.

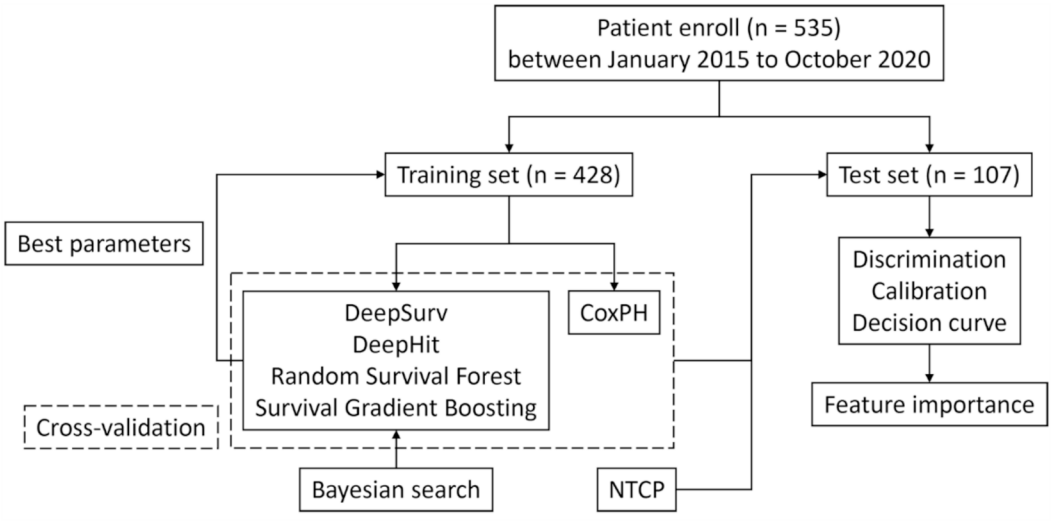

Fig. 1. Workflow diagram of the study.

2.3. Feature collection and model development

Three feature categories were analyzed: clinical, thyroid, and pituitary. Clinical features encompassed sex, age, AJCC T stage, N stage, clinical stage, chemotherapy regimens, thyroid volume and pituitary volume. Dose-volume metrics of thyroid and pituitary glands were extracted from the TPS, included the minimum, maximum, modal, median, and mean doses to the thyroid and pituitary. Vx and VSx of the thyroid and pituitary were also recorded at 5-Gy intervals (e.g., V5, V10, V15 up to V70, and VS5, VS10, VS15 up to VS70).

The feature selection process commenced with the elimination of irrelevant and redundant variables to avoid overfitting. Clinical features underwent univariate and multivariate Cox regression analysis for initial selection. For the dosimetric features, the least absolute shrinkage and selection operator method integrated with Cox regression (LASSO-Cox) was utilized to find the most valuable predictors [22]. The tuning parameter (λ) was optimized using 10-fold cross-validation based on the minimum criteria. Subsequently, Spearman correlation coefficients were then calculated to further evaluate pairwise feature correlations. Features exhibiting a correlation coefficient greater than 0.8 were identified as highly correlated, and the feature with the highest mean absolute correlation with other features was removed to mitigate collinearity.

A total of five survival prediction models were constructed in this study, specifically DeepSurv [23], DeepHit [24], random survival forest (RSF) [25], survival gradient boosting machine (GBM) [26], and the classical statistical Cox proportional hazards model (CoxPH), in which DeepSurv and DeepHit were deep learning-based algorithms. DeepSurv is a Cox proportional hazards deep neural network designed for survival analysis. It extends traditional Cox regression by using deep neural networks to learn complex, non-linear relationships between covariates and hazard rates. DeepHit, on the other hand, is a deep neural network for survival analysis that directly predicts the probability of an event occurring within a specific time frame. Both models leverage the power of deep learning to capture complex relationships in survival data, potentially offering improved predictive performance over traditional statistical methods. Each model was selected as a representative of its respective class of survival analysis algorithms. Furthermore, two NTCP models were adapted to our cohort according to literatures. One model was selected for its well-performing characteristics [27,28], while the other was specifically developed for nasopharyngeal carcinoma patients [29].

To optimize the performance of DeepSurv, DeepHit, RSF and GBM, Bayesian optimization coupled with 5-fold cross-validation was utilized to determine the optimal set of hyperparameters. The discriminative capabilities of the models were quantified using the concordance index (C-index) and time-dependent area under the receiver operating characteristic curves (tAUC). Calibration of the models was evaluated by the Brier score and calibration plot. Additionally, decision curve analysis (DCA) was employed to assess the clinical utility of each model in supporting medical decision-making processes. Feature importance was calculated by permutation importance, which is a model-agnostic method [30]. This method could be suitable for any kind of model, especially black-box ones such as deep learning. By randomly shuffling the values of one feature in the test set, such method could reveal the feature's importance to the model's predictive performance. The model with superior performance metrics was then embedded into a user-friendly application in order to elevate accessibility and facilitate clinical application.

2.4. Statistical analysis

All statistical analysis was conducted in R (https://cran.r-project.org/) and Python (https://www.python.org/). Chi-square test and *t*-test were used to compare categorical and continuous variables, respectively. A two-sided p value < 0.05 was considered statistically significant. The cumulative hazard curves were compared by log-rank test.

3. Results

3.1. Basic characteristics

Between January 2015 and October 2020, 535 patients were recruited for the study with a median age of 50.5 years. The median follow-up period was 52 months, within which 223 (41.7 %) patients developed RIHT (Fig. 2). Of those affected, 67 (30.0 %) individuals developed clinical hypothyroidism, while 163 (73.1 %) patients experienced subclinical hypothyroidism. No case of central hypothyroidism was observed. Notably, 24 patients who were initially diagnosed with subclinical hypothyroidism progressed to clinical hypothyroidism during subsequent follow-ups. The median time to RIHT onset was 15 months. Cumulative incidences of primary hypothyroidism at 1, 3, and 5 years were observed to be 17.3 %, 34.6 %, and 41.3 %, respectively. Significant statistical differences were found in the AJCC N stage and thyroid volume between patients with hypothyroidism and those without. The division into a training set (n = 428) and a test set (n = 107) was balanced (Table 1).

3.2. Model development and evaluation

After univariate and multivariate Cox regression analyses of clinical factors, AJCC N stage and thyroid volume showed as potential predictors of hypothyroidism (Table 2). Six dose-volume parameters with non-zero coefficients were selected by LASSO-Cox analysis, which were thyroid V30, V50, VS10, VS40, VS65 and pituitary max dose (Appendix Table A1). Spearman's correlation analysis on these parameters confirmed the absence of significant correlations between any pairs, and thus supported their independence in the model (Fig. 3). Consequently, a total of eight features were ultimately selected for the development of the model. The two NTCP models were composed of thyroid volume, mean thyroid dose, V50, maximum pituitary dose, and chemotherapy status (received = 1, not received = 0), according to the literature [28,29].

Firstly, hyperparameters were obtained by pre-tuning from two deep learning and two machine learning survival algorithms using Bayesian optimization with 5-fold cross-validation (Appendix Table A2). Then, the models were retrained on the training set with the optimal hyperparameters and tested on the test set. All deep learning and machine learning models outperformed the traditional statistical Cox model and two NTCP models in terms of the C-index. DeepSurv achieved the highest C-index at 0.75, followed by RSF at 0.72, Deephit at 0.72, and GBM at 0.69, while the Cox model at 0.63, NTCP1 [28] at 0.62, and

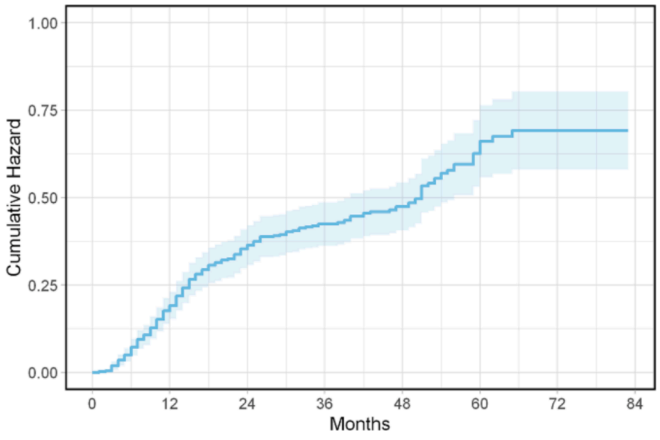

Fig. 2. RIHT cumulative hazard curve for all study participants.

Table 1
Baseline characteristics of study participants.

Variables	Levels	Overall	nHT (N = 312)	HT (N = 223)	p	Training set (N = 428)	Test set (N = 107)	p
Sex	Female	150 (28.0 %)	79 (25.3 %)	71 (31.8 %)	0.119	119 (27.8 %)	31 (29 %)	0.392
	Male	385 (72.0 %)	233 (74.7 %)	152 (68.2 %)		309 (72.2 %)	76 (71 %)	
Age	Mean ± SD	49.0 ± 12.6	48.6 ± 12.6	49.6 ± 12.6	0.389	49.2 ± 12.5	48.4 ± 13.3	0.473
T stage	1	57 (10.7 %)	31 (9.9 %)	26 (11.7 %)	0.15	43 (10 %)	14 (13.1 %)	0.289
	2	126 (23.6 %)	72 (23.1 %)	54 (24.2 %)		106 (24.8 %)	20 (18.7 %)	
	3	192 (35.9 %)	104 (33.3 %)	88 (39.5 %)		151 (35.3 %)	41 (38.3 %)	
	4	160 (29.9 %)	105 (33.7 %)	55 (24.7 %)		128 (29.9 %)	32 (29.9 %)	
N stage	0	21 (3.9 %)	18 (5.8 %)	3 (1.3 %)	0<.001	15 (3.5 %)	6 (5.6 %)	0.918
	1	195 (36.4 %)	129 (41.3 %)	66 (29.6 %)		158 (36.9 %)	37 (34.6 %)	
	2	218 (40.7 %)	117 (37.5 %)	101 (45.3 %)		170 (39.7 %)	48 (44.9 %)	
	3	101 (18.9 %)	48 (15.4 %)	53 (23.8 %)		85 (19.9 %)	16 (15 %)	
Clinical stage	1	4 (0.7 %)	4 (1.3 %)	0 (0 %)	0.316	4 (0.9 %)	0 (0 %)	0.23
	2	75 (14.0 %)	46 (14.7 %)	29 (13 %)		57 (13.3 %)	18 (16.8 %)	
	3	217 (40.6 %)	122 (39.1 %)	95 (42.6 %)		171 (40 %)	46 (43 %)	
	4	239 (44.7 %)	140 (44.9 %)	99 (44.4 %)		196 (45.8 %)	43 (40.2 %)	
Chemotherapy	C	9 (1.7 %)	5 (1.6 %)	4 (1.8 %)	0.246	7 (1.6 %)	2 (1.9 %)	0.507
	I + C	6 (1.1 %)	3 (1 %)	3 (1.3 %)		5 (1.2 %)	1 (0.9 %)	
	C + A	315 (58.9 %)	196 (62.8 %)	119 (53.4 %)		249 (58.2 %)	66 (61.7 %)	
	I + C + A	194 (36.3 %)	101 (32.4 %)	93 (41.7 %)		159 (37.1 %)	35 (32.7 %)	
	RT alone	11 (2.1 %)	7 (2.2 %)	4 (1.8 %)		8 (1.9 %)	3 (2.8 %)	
Thyroid volume	Mean ± SD	14.6 ± 6.4	15.2 ± 6.2	13.8 ± 6.7	0.012	14.7 ± 6.2	14.3 ± 7.3	0.225
Pituitary volume	Mean ± SD	0.4 ± 0.2	0.4 ± 0.2	0.4 ± 0.3	0.495	0.4 ± 0.2	0.4 ± 0.3	0.536

Abbreviation: nHT, no hypothyroidism. HT, hypothyroidism. C, concurrent chemotherapy. I, induction chemotherapy. A, adjuvant chemotherapy. RT, radiotherapy.

Table 2
Univariate and multivariate Cox regression analysis of clinical factors.

Variables	**Levels**	**Overall**	**Univariable (HR)**	**p**	**Multivariable (HR)**	**p**
Sex	Female	119 (27.8 %)				
	Male	309 (72.2 %)	0.79 (0.57–1.08)	p = 0.143	0.83 (0.60–1.15)	p = 0.256
Age	Mean ± SD	49.2 ± 12.5	1.00 (0.99–1.01)	p = 0.971		
T stage	≤2	149 (34.8 %)				
	> 2	279 (65.2 %)	0.95 (0.70–1.30)	p = 0.756		
N stage	≤1	173 (40.4 %)				
	> 1	255 (59.6 %)	1.48 (1.08–2.04)	p = 0.015	1.51 (1.09–2.07)	p = 0.012
Clinical stage	≤II	61 (14.3 %)				
	> III	367 (85.7 %)	1.21 (0.77–1.92)	p = 0.407		
Chemotherapy	C	7 (1.6 %)				
	I + C	5 (1.2 %)	0.76 (0.13–4.55)	p = 0.763		
	C + A	249 (58.2 %)	0.97 (0.31–3.05)	p = 0.953		
	I + C + A	159 (37.1 %)	1.07 (0.34–3.41)	p = 0.906		
	RT alone	8 (1.9 %)	0.99 (0.20–4.90)	p = 0.989		
Thyroid volume	Mean ± SD	14.7 ± 6.2	0.97 (0.95–1.00)	p = 0.034	0.97 (0.95–1.00)	p = 0.046
Pituitary volume	Mean ± SD	0.4 ± 0.2	1.26 (0.70–2.27)	p = 0.450		

NTCP2 [29] at 0.60 (Fig. 4a). Although DeepSurv model demonstrated superior discrimination capacity, the performance was not uniformly superior across all time points (Appendix Table A3). Traditional models were on par with DeepSurv at early post-treatment intervals, while the latter excelled with extended follow-up periods, maintaining better predictive performance at most time points. The clinical utility of our models was further appraised using decision curve analysis, demonstrating the superiority of deep learning and machine learning models over Cox and NTCP models in informing decisions. (Fig. 4b).

In terms of calibration, traditional NTCP models showed higher Brier score compared to the deep learning survival algorithms at all assessed time points (Fig. 4c). The integrated Brier score of each model was DeepSurv at 0.17, Deephit at 0.19, RSF at 0.18 and GBM at 0.20. Calibration curves at third-year post treatment revealed improved correlation between the predicted probabilities and actual outcomes for the DeepSurv model compared to traditional models (Fig. 4d). DeepSurv could effectively stratify patients into high-risk and low-risk groups of RIHT (Fig. 5).

3.3. Feature importance

Feature importance of DeepSurv, DeepHit, RSF and GBM was assessed by permutation importance, in which features achieving higher permutation scores were more significant. In our dataset, thyroid-related features played dominant roles across all four models, whereas features of the pituitary and clinical N stage had less impact. Notably, thyroid V30 was identified as the most valuable feature in most models, followed by thyroid V50 (Fig. 6).

3.4. Model deployment

We created a user-friendly web application to make the DeepSurv model more accessible for practical use, considering its superior performance, accessible at https://huggingface.co/spaces/mmyycc/demo.

4. Discussion

This study reported one of the largest cohorts of patients with nasopharyngeal carcinoma who were treated with definitive IMRT and platinum-based chemotherapy. Besides, this study takes the lead in RIHT prediction using deep learning survival analysis. Our findings suggested that the proposed deep learning model incorporated both clinical and dose-volume features, outperformed traditional models.

AJCC N stage and thyroid volume were underscored as significant

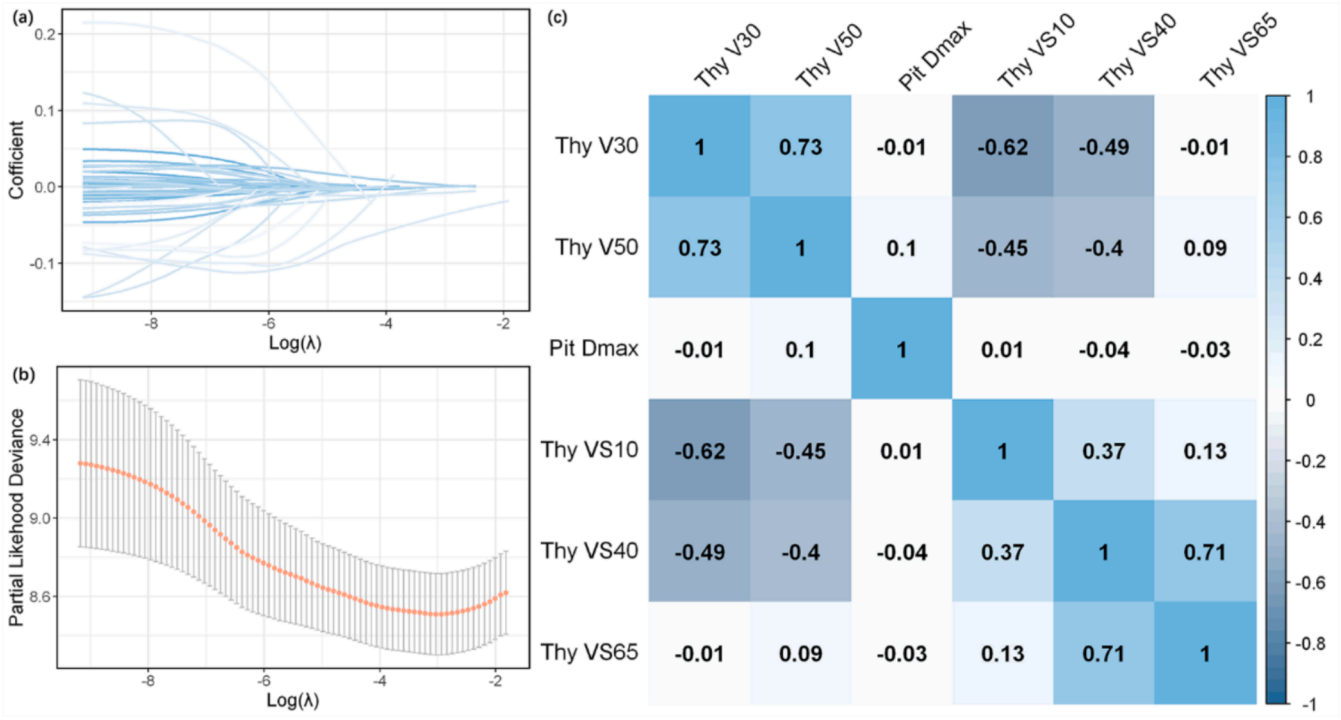

Fig. 3. Feature selection process using LASSO-Cox analysis and spearman correlation test. (a) LASSO-Cox regression for variable selection. (b) Optimization of penalty parameter (λ) in LASSO-Cox. (c) Spearman correlation plot of selected features. Pit: pituitary, Thy: thyroid.

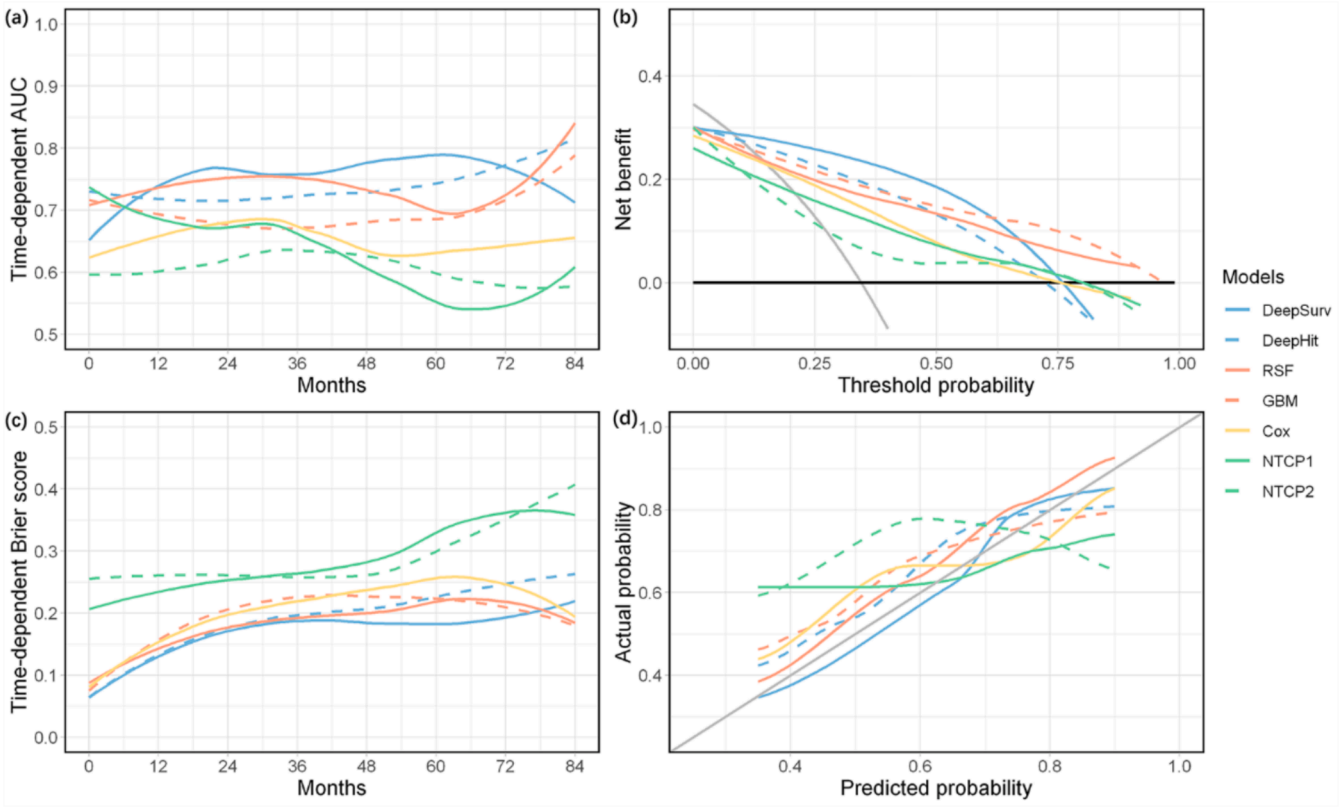

Fig. 4. Performance evaluation of prognostic models over time. (a) Time-dependent AUC of 7 prognostic models. (b) Decision curve analysis of 7 prognostic models. (c) Time-dependent Brier score of 7 prognostic models. (d) Calibration plot of HT-free probability at the third-year post treatment. NTCP1: NTCP model published by Rønjom [28]. NTCP2: NTCP model published by Luo [29].

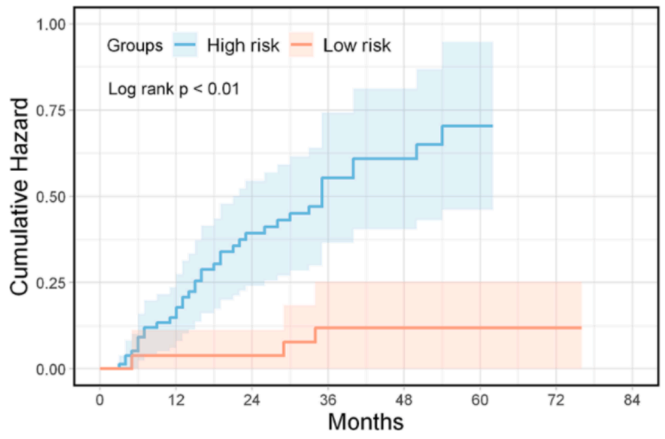

Fig. 5. Cumulative hazard curves of high- and low-risk groups in the test set stratified by DeepSurv.

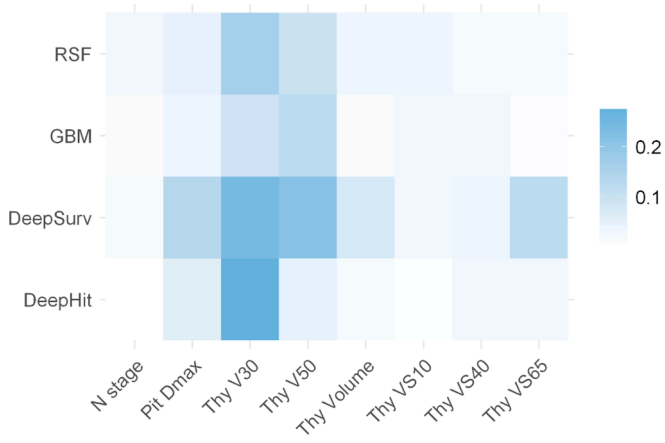

Fig. 6. Permutation feature importance across survival models. Pit: pituitary, Thy: thyroid.

clinical predictors for RIHT, and the possible reason might be due to the increased radiation exposure to the thyroid region in patients with advanced nodal involvement [31]. Higher N stages in nasopharyngeal carcinoma often required broader or more intensive irradiation fields to encompass the affected lymph nodes, which increases the likelihood of incidental thyroid irradiation and subsequent hypothyroidism. The association of higher N stages with RIHT may also be compounded using platinum-based chemotherapy in order to enhance the radiosensitivity of the thyroid gland. Patients with larger thyroids may benefit from a greater functional reserve because of more tissue could be intact and functional [1].

With regards to address the challenge of multicollinearity among thyroid-related dose-volume parameters, this study employed LASSO-Cox regression method due to its capability for variable selection and regularization. LASSO-Cox effectively shrinks less critical predictors to zero, thereby retaining the most significant variables. Spearman's rank correlation was then calculated to validate the LASSO-Cox findings, offered a non-parametric measure of association that is robust to outliers and further elucidated the relationships between selected features. Other dimensionality reduction methods, such as principal component analysis (PCA), were also considered. However, PCA transforms the original variables into a new set of uncorrelated variables, making it challenging to ascertain the individual influence of each original variable on the outcome and limiting its practical value. Therefore, it was not chosen for this study.

In this study, we compared several models for predicting RIHT in nasopharyngeal carcinoma patients, including traditional statistical models, machine learning models, and deep learning models. Each approach has its strengths and limitations. Traditional statistical models offer interpretability and are well-established in medical literature, but they may struggle to capture complex non-linear relationships in the data. Machine learning models and deep learning models excel at learning complex patterns from high-dimensional data. However, they typically require larger datasets, and are less interpretable. The investigation of RIHT has been an active area of research in radiation oncology, with numerous studies employing traditional statistical models to analyze DVH parameters and other clinical factors. Despite these efforts, a consensus on the most reliable predictors have not been firmly established. Given the limitations of traditional models in reaching a common conclusion, there is a pressing need to explore RIHT prediction using more advanced methodologies. Our study suggested that deep learning survival model such as DeepSurv, may offer incremental improvements in prediction relative to established traditional models like the CoxPH and NTCP models. These advanced algorithms have shown high accuracy in reflecting the complex and nonlinear interactions among variables, a nuanced understanding that traditional models sometimes approximate inadequately. The capacity of deep learning survival algorithms to process the temporal dimension of survival data where both the occurrence and timing of events are essential, highlighted its adaptability in modeling patient risk profiles that evolve over time. This is a subtle yet important distinction from the traditional statistical models. The calibration and decision curve analyses also provided evidence for the enhanced predictive utility of deep learning model. They point towards the possibility that these models could support improved patient management strategies.

The efficacy of the models in predicting patient outcomes within our study exhibited variability at distinct time points. This observation suggests that the utility of these models may be contingent upon the length of follow-up. RIHT progression and the impact of therapeutic interventions are subject to temporal shifts, and thereby alter the prognostic significance of various predictors throughout the disease course. Immediate post-treatment outcomes might frequently be influenced by acute reactions to therapy, and in the meanwhile outcomes over an extended period might be shaped by a confluence of persistent treatment effects, patient-specific attributes, and other enduring health determinants. As such, a model with high classification accuracy in the short term may not necessarily maintain this performance for longer-term prognostications, and vice versa. Despite the existing challenge, deep learning survival algorithms with their inherent flexibility and capacity in learning processing large-scale data, hold promise in addressing this temporal variance. It could be expected that by enlarging the data scales, these models could have better performances in capturing and differentiating the patterns relevant to both short-term and long-term prognoses. Therefore, clinicians can effectively harness the predictive power of deep learning to advance personalized medicine and improve the quality of care.

The permutation importance analysis offered insight into the relative significance of features within the models. The consistent prominence of thyroid V30 and V50 across most of the models strengthens the robustness of our findings. V30 and V50 were frequently reported in previous studies as they might precisely capture the threshold at which radiation exposure leads to significant thyroid damage [4–6,12,13]. Understanding the importance of such features could have profound clinical practices. Moreover, the consistency of these parameters across models suggests that they could be fundamental indicators of RIHT risk, regardless of the specific predictive model employed. Clinicians and medical physicists might aim to minimize these specific dosimetry parameters during the planning stage to preserve thyroid function.

Our study is not without limitations. The single-center, retrospective design may limit the generalizability of our findings, and external validation in multicenter cohorts is necessary to confirm the applicability of our predictive model. Meanwhile, the dosimetrics collected in

this study were physical absorbed doses. Converting dosimetrics to equivalent doses in 2 Gy fractions (EQD2) may provide a more comprehensive assessment of radiation damage biologically [32]. At last, multi-omics data, such as radiomics and dosiomics, might offer a deeper understanding of the molecular and biological mechanisms driving treatment responses and toxicities, as well as provide a more comprehensive analysis of dose distribution, potentially to improve prediction accuracy. The inclusion of multi-omics data in our future research is under consideration.

5. Conclusions

In conclusion, our research offered a comprehensive analysis of clinical and dose-volume predictors for RIHT, and highlighted the significance of certain parameters. Deep learning survival algorithm could predict RIHT and facilitate more informed clinical decision-making. Once a patient was predicted at high risk of RIHT in advance, clinicians could deliberately tailor treatment plans to better preserve thyroid function. It is hoped that this study could bring the treatment for nasopharyngeal carcinoma one step closer to the realization of personalized management and mitigate the suffering of RIHT.

CRediT authorship contribution statement

Yichen Mao: Data curation, Formal analysis, Software, Writing – original draft. **Mingjun Ding:** Methodology, Writing – review & editing. **Dan Zong:** Conceptualization, Supervision. **Zhongde Mu:** Conceptualization, Software, Writing – original draft, Project administration. **Xia He:** Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study was supported by the Key Program of the Jiangsu Commission of Health (K2019028), the Jiangsu Postdoctoral Research Funding Program (2021K422C), the National Natural Science Foundation of China (81801834), and the China Postdoctoral Science Foundation (2018M632263).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ctro.2025.100946.

References

[1] Chow JCH, Cheung KM, Cheung GTC, Tam AHP, Lui JCF, Lee FKH, et al. Dose-volume predictors of post-radiation primary hypothyroidism in head and neck cancer: a systematic review. Clin Transl Radiat Oncol 2022;33:83–92.
[2] Xu Y, Shao Z, Tang T, Liu G, Yao Y, Wang J, et al. A dosimetric study on radiation-induced hypothyroidism following intensity-modulated radiotherapy in patients with nasopharyngeal carcinoma. Oncol Lett 2018;16:6126–32.
[3] Zhu MY, Wu HJ, Miao JJ, Di MP, Chen BY, Huang HG, et al. Radiation-induced hypothyroidism in patients with nasopharyngeal carcinoma treated with intensity-modulated radiation therapy with or without chemotherapy: Development of a nomogram based on the equivalent dose. Oral Oncol 2021;120:105378.
[4] Zhou L, Chen J, Shen W, Chen ZL, Huang S, Tao CJ, et al. Thyroid V(50) is a risk factor for hypothyroidism in patients with nasopharyngeal carcinoma treated with intensity-modulated radiation therapy: a retrospective study. Radiat Oncol 2020;15:68.
[5] Zhai RP, Kong FF, Du CR, Hu CS, Ying HM. Radiation-induced hypothyroidism after IMRT for nasopharyngeal carcinoma: Clinical and dosimetric predictors in a prospective cohort study. Oral Oncol 2017;68:44–9.
[6] Sachdev S, Refaat T, Bacchus ID, Sathiaseelan V, Mittal BB. Thyroid V50 highly predictive of hypothyroidism in head-and-neck cancer patients treated with intensity-modulated radiotherapy (IMRT). Am J Clin Oncol 2017;40:413–7.
[7] Ren W, Liang B, Sun C, Wu R, Men K, Xu Y, et al. Dosiomics-based prediction of radiation-induced hypothyroidism in nasopharyngeal carcinoma patients. Phys Med 2021;89:219–25.
[8] Prpic M, Kruljac I, Kust D, Suton P, Purgar N, Bilos LK, et al. Dose-volume derived nomogram as a reliable predictor of radiotherapy-induced hypothyroidism in head and neck cancer patients. Radiol Oncol 2019;53:488–96.
[9] Lian CL, Zhuo RG, Zhou R, Yu YF, Zhou P, Lin Q, et al. Risk factors of early thyroid dysfunction after definitive radiotherapy in nasopharyngeal carcinoma patients. Head Neck 2023;45:2344–54.
[10] Huang CL, Tan HW, Guo R, Zhang Y, Peng H, Peng L, et al. Thyroid dose-volume thresholds for the risk of radiation-related hypothyroidism in nasopharyngeal carcinoma treated with intensity-modulated radiotherapy-A single-institution study. Cancer Med 2019;8:6887–93.
[11] Chyan A, Chen J, Shugard E, Lambert L, Quivey JM, Yom SS. Dosimetric predictors of hypothyroidism in oropharyngeal cancer patients treated with intensity-modulated radiation therapy. Radiat Oncol 2014;9:269.
[12] Akgun Z, Atasoy BM, Ozen Z, Yavuz D, Gulluoglu B, Sengoz M, et al. V30 as a predictor for radiation-induced hypothyroidism: a dosimetric analysis in patients who received radiotherapy to the neck. Radiat Oncol 2014;9:104.
[13] Peng L, Mao YP, Huang CL, Guo R, Ma J, Wen WP, et al. A new model for predicting hypothyroidism after intensity-modulated radiotherapy for nasopharyngeal carcinoma. Front Oncol 2020;10:551255.
[14] Lertbutsayanukul C, Kitpanit S, Prayongrat A, Kannarunimit D, Netsawang B, Chakkabat C. Validation of previously reported predictors for radiation-induced hypothyroidism in nasopharyngeal cancer patients treated with intensity-modulated radiation therapy, a post hoc analysis from a Phase III randomized trial. J Radiat Res 2018;59:446–55.
[15] Lee V, Chan SY, Choi CW, Kwong D, Lam KO, Tong CC, et al. Dosimetric predictors of hypothyroidism after radical intensity-modulated radiation therapy for non-metastatic nasopharyngeal carcinoma. Clin Oncol (r Coll Radiol) 2016;28:e52–60.
[16] Fujiwara M, Kamikonya N, Odawara S, Suzuki H, Niwa Y, Takada Y, et al. The threshold of hypothyroidism after radiation therapy for head and neck cancer: a retrospective analysis of 116 cases. J Radiat Res 2015;56:577–82.
[17] Ling S, Bhatt AD, Brown NV, Nguyen P, Sipos JA, Chakravarti A, et al. Correlative study of dose to thyroid and incidence of subsequent dysfunction after head and neck radiation. Head Neck 2017;39:548–54.
[18] Cella L, Liuzzi R, Conson M, D'Avino V, Salvatore M, Pacelli R. Development of multivariate NTCP models for radiation-induced hypothyroidism: a comparative analysis. Radiat Oncol 2012;7:224.
[19] Kamal M, Peeler CR, Yepes P, Mohamed ASR, Blanchard P, Frank S, et al. Radiation-Induced hypothyroidism after radical intensity modulated radiation therapy for oropharyngeal carcinoma. Adv Radiat Oncol 2020;5:111–9.
[20] Fan CY, Lin CS, Chao HL, Huang WY, Su YF, Lin KT, et al. Risk of hypothyroidism among patients with nasopharyngeal carcinoma treated with radiation therapy: a population-based cohort study. Radiother Oncol 2017;123:394–400.
[21] Rooney MK, Andring LM, Corrigan KL, Bernard V, Williamson TD, Fuller CD, et al. Hypothyroidism following radiotherapy for head and neck cancer: a systematic review of the literature and opportunities to improve the therapeutic ratio. Cancers (Basel) 2023;15.
[22] Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997;16:385–95.
[23] Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol 2018;18:24.
[24] Lee C, Yoon J, Schaar MV. Dynamic-deephit: a deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. IEEE Trans Biomed Eng 2020;67:122–33.
[25] Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. Ann Appl Stat 2008;2(841–60):20.
[26] Chen Y, Jia Z, Mercola D, Xie X. A gradient boosting algorithm for survival analysis via direct optimization of concordance index. Comput Math Methods Med 2013;2013:873595.
[27] Nowicka Z, Tomasik B, Papis-Ubych A, Bibik R, Graczyk L, Latusek T, et al. Radiation-induced hypothyroidism in patients with oropharyngeal cancer treated with IMRT: independent and external validation of five normal tissue complication probability models. Cancers (Basel) 2020;12.
[28] Ronjom MF, Brink C, Bentzen SM, Hegedus L, Overgaard J, Johansen J. Hypothyroidism after primary radiotherapy for head and neck squamous cell carcinoma: normal tissue complication probability modeling with latent time correction. Radiother Oncol 2013;109:317–22.
[29] Luo R, Wu VWC, He B, Gao X, Xu Z, Wang D, et al. Development of a normal tissue complication probability (NTCP) model for radiation-induced hypothyroidism in nasopharyngeal carcinoma patients. BMC Cancer 2018;18:575.
[30] Breiman L. Random forests. Mach Learn 2001;45:5–32.
[31] Yang SS, Peng QH, Wu AQ, Zhang BY, Liu ZQ, Chen EN, et al. Radiomics and dosiomics for predicting radiation-induced hypothyroidism and guiding intensity-modulated radiotherapy. iScience 2023;26(12):108394. https://doi.org/10.1016/j.isci.2023.108394.
[32] Shen G, Peng Y, Li J, Wu H, Zhang G, Zhao C, et al. Multivariate NTCP model of hypothyroidism after intensity-modulated radiotherapy for nasopharyngeal carcinoma. Front Oncol 2021;11:714536.