



OPEN

CT-based radiomics and deep learning models for predicting thyroid cartilage invasion and patient prognosis in laryngeal carcinoma

Xinwei Chen^{1,4}, Shuang Xia^{2,4}, Huan Jiang¹, Fajin Lv¹, Qiang Yu¹, Youquan Ning¹, Kai Xie¹, Quanjiang Li¹, Rui Liu¹, Yuan Zhou¹, Guohua Hu³✉ & Juan Peng¹✉

Accurate assessment of thyroid cartilage invasion is crucial for treatment decision-making and prognosis evaluation in laryngeal squamous cell carcinoma (LSCC). This study aimed to compare the performance of the radiomics and deep learning (DL) models for predicting thyroid cartilage invasion in LSCC patients, and evaluate prognostic value of the optimal predictive model. A total of 418 pathologically confirmed LSCC patients from two centers were enrolled and divided into a training cohort ($n=247$), an internal validation cohort ($n=110$), and an external validation cohort ($n=61$). Models were developed based on venous-phase CT images and compared with two radiologists. A nomogram incorporating the optimal model and clinical risk factors was also constructed. Additionally, the prognostic value of the optimal model was assessed regarding disease-free survival (DFS). The 2D DL model showed better performance in predicting thyroid cartilage invasion, and the corresponding nomogram integrating 2D DL signature and clinical risk factors achieved the highest AUCs. However, no differences in AUCs were found in the external validation cohort ($p>0.05$ for all). Additionally, the 2D DL signature and clinical N stage were independent predictors of DFS. The 2D DL-based nomogram demonstrated satisfactory predictive performance for thyroid cartilage invasion and prognosis in patients with LSCC.

Keywords Laryngeal squamous cell carcinoma, Thyroid cartilage invasion, Deep learning, Radiomics, Disease-free survival

Thyroid cartilage invasion serves as a crucial determinant in TNM staging of laryngeal squamous cell carcinoma (LSCC), with medial cortex invasion classified as stage T3 and extralaryngeal tissue penetration as stage T4. Previous studies have reported thyroid cartilage invasion in 12%–43% of laryngeal carcinoma patients, with missed diagnoses possibly resulting in staging errors in 40%–50% of cases^{1,2}. Clinically, thyroid cartilage invasion significantly impacts the choice of treatment strategies. For cases with thyroid cartilage invasion, radiotherapy alone is often insufficient, making concurrent chemoradiotherapy a common option for organ-preserving treatment. However, when the tumor extensively involves the cartilage or invades extralaryngeal tissues, total laryngectomy is necessary to ensure complete resection³. Additionally, thyroid cartilage invasion closely correlates with radiotherapy response and tumor recurrence^{4,5}. Hence, it is of great significance to accurately evaluate thyroid cartilage invasion for making treatment decisions and assessing prognosis in patients with LSCC.

CT is frequently utilized to predict thyroid cartilage invasion, with a sensitivity of approximately 80%⁶. However, its accuracy and specificity are unsatisfactory, which may be attributed to the asymmetric ossification of normal thyroid cartilage, and the similar CT values between tumor tissue and non-ossified cartilage may

¹Department of Radiology, The First Affiliated Hospital of Chongqing Medical University, No. 1 Youyi Road, Yuzhong District, Chongqing 400016, China. ²Department of Radiology, Medical Imaging Institute of Tianjin, School of Medicine, Tianjin First Central Hospital, Nankai University, Tianjin, China. ³Department of Otorhinolaryngology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China. ⁴Xinwei Chen and Shuang Xia contributed equally to this work. ✉email: hghcq@sina.com; pengjuan0717@sina.com

also affect the identification^{7,8}. Although MRI has a higher sensitivity for diagnosing thyroid cartilage invasion, cartilage edema or inflammatory changes may resemble tumor infiltration⁹. Dual-energy CT technology, particularly iodine overlay (IO) image, can help differentiate between iodine-enhanced tumors and non-ossified cartilage^{10,11}. However, IO image has limitations in identifying lesions in previously ossified cartilage, potentially overestimating iodine distribution¹². Currently, there is a lack of robust and precise methods for discerning thyroid cartilage invasion in LSCC.

Recently, artificial intelligence, particularly machine learning and deep learning (DL), has seen extensive use in medical imaging. Radiomics, a key aspect of machine learning, extracts high-dimensional features and offers valid quantitative data for tumor analysis. DL, on the other hand, automatically captures and learns discriminative features in an end-to-end manner, improving prediction accuracy. Radiomics and DL have been reported to be valuable in predicting tumor stage¹³, early recurrence¹⁴, treatment response¹⁵, and prognosis¹⁶ in laryngeal carcinoma. To our knowledge, few studies have investigated radiomics and DL techniques in predicting thyroid cartilage invasion. Given the complexity of laryngeal tumor, we also developed a 3D DL model, which usually requires an extensive training dataset to achieve precision^{17,18}.

Collectively, this study aimed to construct a radiomics model, a 2D DL model, and a 3D DL model using venous-phase CT images to evaluate thyroid cartilage invasion, and compare these models with two radiologists. A nomogram incorporating the optimal predictive model and clinical risk factors was also developed. Additionally, we further explored the prognostic value of the optimal predictive model for thyroid cartilage invasion.

Materials and methods

Patients

This retrospective multicenter study analyzed data from 418 patients with pathologically confirmed LSCC. The study utilized data from two institutions: The First Affiliated Hospital of Chongqing Medical University (center 1), which contributed 357 patients between March 1, 2011, and January 1, 2021, and Tianjin First Central Hospital, School of Medicine, Nankai University (center 2), which provided data from 61 patients between March 1, 2017, and January 1, 2021. Patients were categorized into a training cohort and an internal validation cohort at a 7:3 ratio for center 1, while data from center 2 was used for external validation (Fig. 1). All patients were enrolled via the following inclusion criteria: (1) pathologically confirmed diagnosis of LSCC, (2) comprehensive clinical and imaging data, and (3) preoperative contrast-enhanced CT (CECT) performed within 2 weeks before surgery. The exclusion criteria included: (1) tumor recurrence or history of other malignant tumors, (2) poor quality of CT images, and (3) preoperative anti-tumor therapy.

Baseline clinical data included age, sex, smoking status, alcohol consumption, tumor location, CT-reported anterior commissure (AC) invasion, histological grade, clinical T stage, clinical N stage, and overall clinical stage. All patients in this study underwent surgical resection, and the postoperative pathological findings were considered the gold standard for determining the presence of thyroid cartilage invasion. Representative diagnostic examples are shown in Fig. 2.

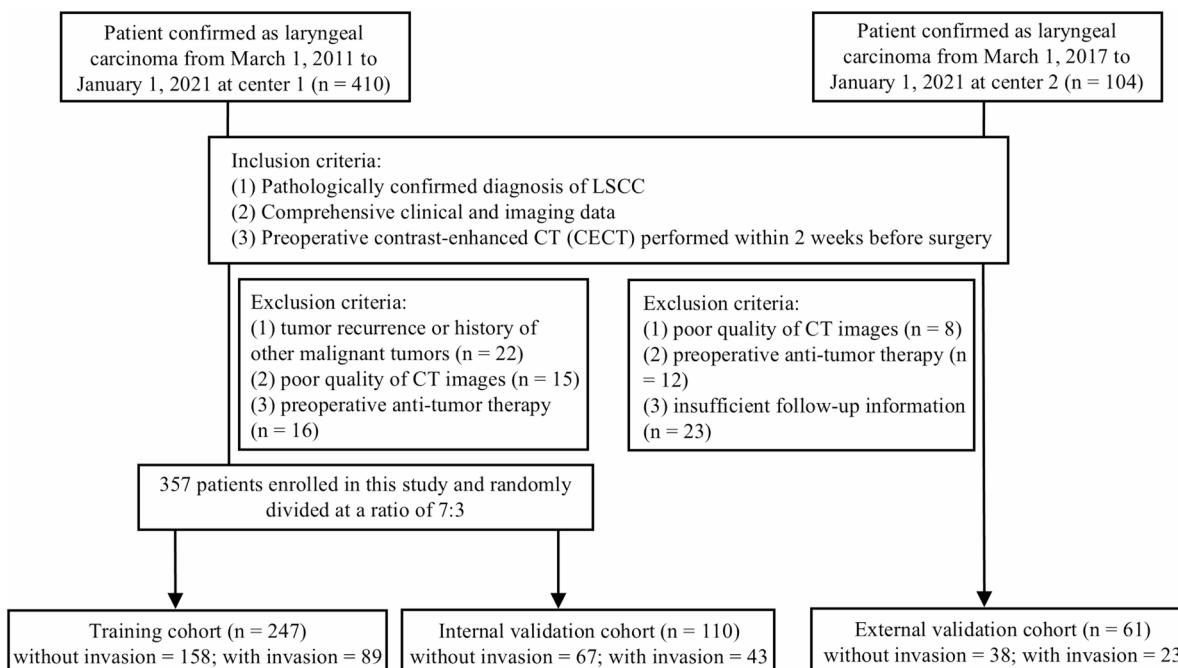


Fig. 1. Flowchart of patient recruitment. LSCC, laryngeal squamous cell carcinoma; center 1, The First Affiliated Hospital of Chongqing Medical University; center 2, Tianjin First Central Hospital, School of Medicine, Nankai University.

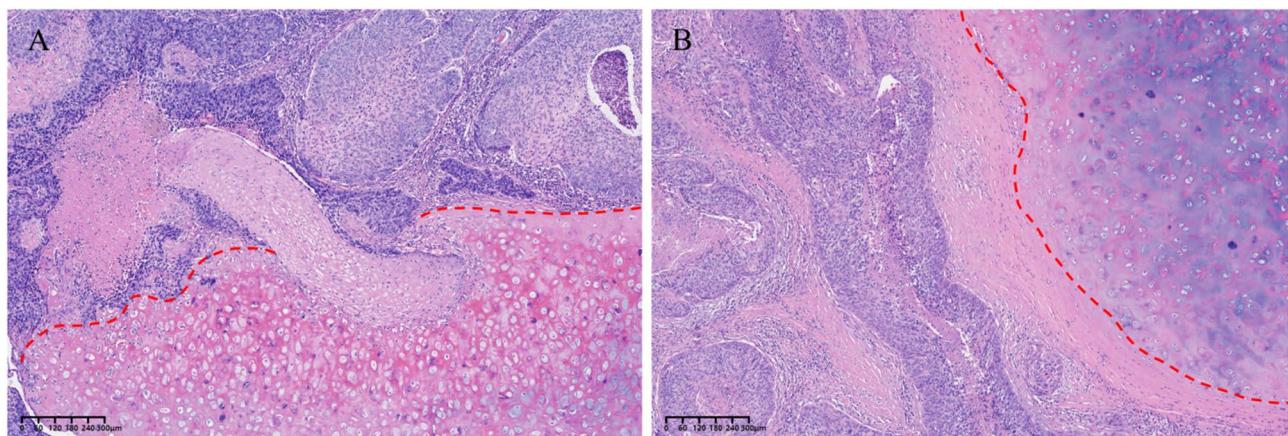


Fig. 2. Thyroid cartilage invasion was evaluated on full sections of LSCC stained with H&E. Representative samples classified as with thyroid cartilage invasion (A) and without thyroid cartilage invasion (B). The red dashed line indicates the margin of the thyroid cartilage.

Endpoints and follow-up

In this study, a comprehensive postoperative follow-up plan was implemented to ensure accurate tracking of patient conditions. The schedule of follow-up was designed every 3 months in the first year, every 6 months from the second to the fifth year, and annually thereafter. Follow-up continued for at least 3 years or until death. For patients lost to follow-up, survival time was calculated from the date of surgery to the last recorded follow-up. The primary objective was to assess disease-free survival (DFS), which measures the time from surgery to disease progression, the last follow-up, or death from any cause.

CT imaging acquisition and tumor segmentation

Each patient underwent CECT scanning using the multislice spiral CT scanners from two centers. After the plain scan, the contrast agent was administered intravenously at a flow rate of 2.5–4.0 mL/s to obtain contrast-enhanced images. The detailed CT acquisition parameters for the two centers are provided in Table S1.

The venous-phase CT images, most commonly used in laryngeal cancer research^{19,20}, were chosen for further analysis. Tumor regions of interests (ROIs) were manually delineated slice-by-slice on each transverse section by radiologist A (C.X.W.). A 3D volume of interest (VOI) was formed by merging the corresponding ROIs. To assess the inter- and intraobserver reproducibility of radiomics features, 30 patients were randomly selected and re-segmented by both radiologist A and radiologist B (J.H.) after one month. Figure 3 shows the overall workflow of this study.

Radiomics feature extraction/selection and radiomics model Building

Utilizing PyRadiomics (version 2.2.0, <https://github.com/Radiomics/pyradiomics>), features were extracted from the VOIs to ensure a robust analysis. The image processing and feature extraction procedures in this study were conducted in accordance with the Image Biomarker Standardization Initiative (IBSI) standards. Prior to extraction, data preprocessing was performed to ensure internal consistency across the dataset, including resampling VOIs to a voxel spacing of $1 \times 1 \times 1 \text{ mm}^3$ and discretizing voxel intensities using a bin width of 25 Hounsfield units (HU). A total of 1,106 radiomics features were initially extracted, encompassing four categories: first-order statistics, shape-based features, texture features, and wavelet-transformed features. The reliability of these features were assessed through intraclass correlation coefficient (ICC) analysis. To refine the feature set, an initial screening procedure was conducted using the Spearman correlation coefficient, which helped in identifying and removing highly correlated features. Subsequently, the remaining features normalized by z-score were followed by a more precise selection process using the least absolute shrinkage and selection operator (LASSO) method, which retained the non-zero coefficient features for modeling. Finally, a radiomics model was constructed using a support vector machine (SVM) classifier, chosen for its effectiveness in binary classification tasks.

2D DL model construction and visualization

In this study, ResNet-50, pretrained on ImageNet, was used for the backbone network of the 2D DL model (see Supplementary Material). The 2D ROIs were semi-automatically cropped by a rectangular box in the largest tumor cross-section. Before model training, all cropped ROIs were resized to a standardized resolution of 224×224 pixels using bilinear interpolation. Data augmentation strategies, including random horizontal and vertical flipping and random cropping, were applied to increase training robustness. The 2D ROIs were then input into ResNet-50 for a binary classification task, with the output aimed at predicting the risk of thyroid cartilage invasion. Model training encompassed forward computation and backpropagation, and categorical cross-entropy served as the loss function. In terms of training parameters, we incorporated a stochastic gradient descent optimizer with a base learning rate of 0.01, a weight decay of 0.01, a batch size of 64, and 50 training epochs. L2 regularization and an early stopping strategy were applied to prevent overfitting. In addition, we

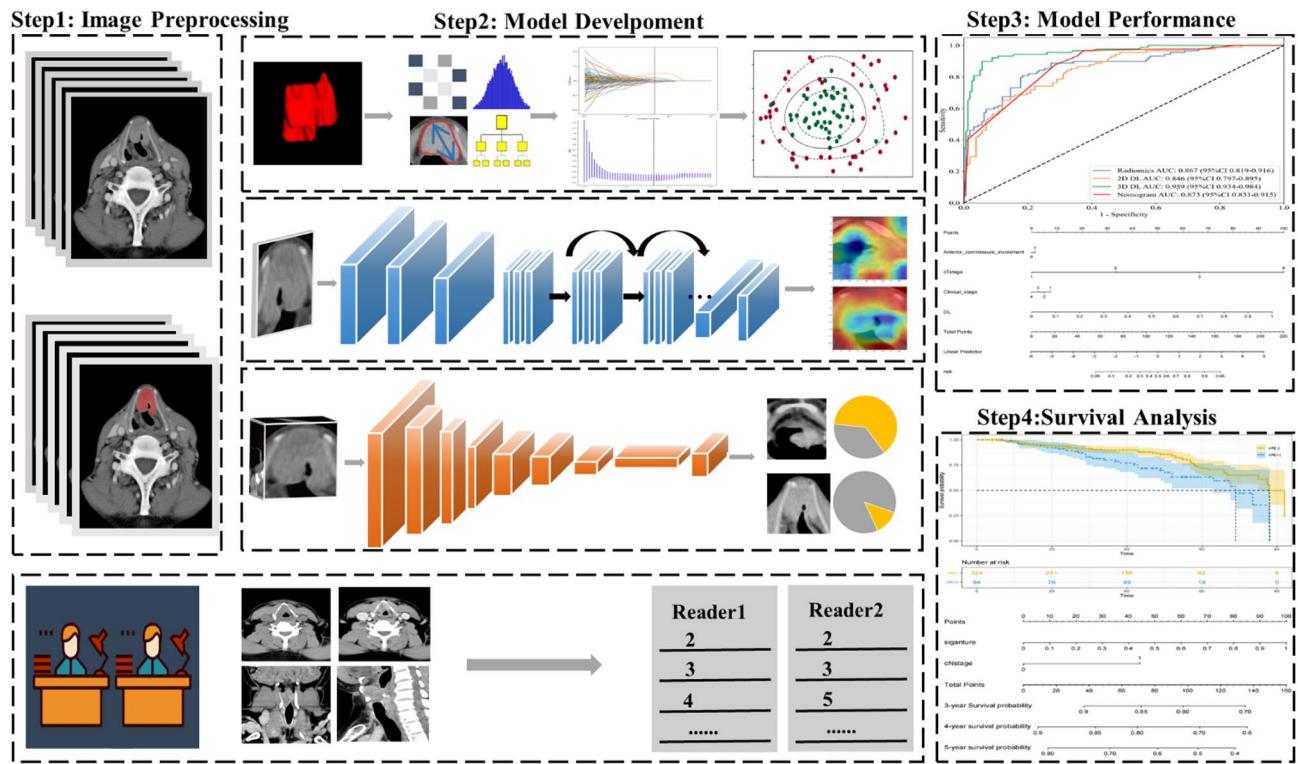


Fig. 3. The pipeline of this retrospective study. In this study, an SVM classifier was applied to construct the radiomics model. The 2D DL model was built using cropped 2D ROIs by ResNet-50, and Grad-CAM was used for model visualization. 3D-ShuffleNet V1 was adopted as the backbone for the 3D DL model, and its input was the bounding cube of the 2D ROIs. Additionally, the performance of two readers was also evaluated for comparison with that of the predictive models. Eventually, the optimal predictive model would be obtained and was utilized to assess patient prognosis.

utilized gradient information from the final convolutional layer to generate gradient-weighted class activation mapping (Grad-CAM), which provides visual explanations for the 2D DL model.

3D DL model development and validation

Considering that 2D ROIs may not fully capture the structural intricacies of laryngeal tumors, we propose a 3D DL model for a more exhaustive analysis (see Supplementary Material). In the context of 3D DL, a rectangular bounding box was used to mark the bounding cube of the 2D ROIs, with linear interpolation standardizing all cubic ROIs to dimensions of $96 \times 96 \times 96$. To improve data diversity and model generalizability, data augmentation strategies such as random flipping along the X, Y, and Z axes were applied. Preprocessed cubic tumor images were subsequently fed into the 3D-ShuffleNet network, where operations such as tridimensional convolutional layers, batch normalization, ReLU activation, and channel shuffling were leveraged for image analysis. Ultimately, the 3D DL model would output the probability of thyroid cartilage invasion for each patient. The 3D network was trained via an adaptive moment estimation optimizer, with a learning rate of 0.02, a batch size of 16, and 100 training epochs.

Radiologists' visual evaluations

Two experienced radiologists (P.J. and L.Q.J.) independently evaluated patients from the two validation cohorts. They referred to the multiphase CT images of each patient for the precise assessment and were encouraged to make a rating on a five-level scale: 1, definitely negative, 2, probably negative, 3, erosion, 4, lysis, and 5, extralaryngeal spread through the cartilage^{12,21}. More detailed information is provided in Supplementary Material. Throughout this procedure, two radiologists were completely blinded to patients' clinical data. To estimate the agreement of the ratings between the two radiologists, Cohen's kappa concordance analysis was employed. Details of the radiologists' ratings are provided in the Supplementary Excel File.

Nomogram for thyroid cartilage invasion prediction

In the training cohort, the candidate clinical factors were tested via univariate analysis to screen out key variables ($p < 0.05$). Meanwhile, the performances of the radiomics model, 2D DL model, and 3D DL model were compared to determine the optimal predictive model. Subsequently, the optimal predictive model collaborating with key clinical variables was subsequently utilized to develop the nomogram. Furthermore, the performance of the nomogram was independently tested in two validation cohorts.

Statistical analysis

The data were analyzed using SPSS software (version 26.0) and R software (version 4.1.0). Differences in clinical features were analyzed by Mann-Whitney U-test (continuous variables) and chi-square test (categorical variables). A p value <0.05 was considered statistically significant. The area under the receiver operating characteristic (ROC) curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated. The calibration curve and Hosmer-Lemeshow (HL) test are utilized to assess the calibration of the model. Decision curve analysis (DCA) curves were utilized to appraise the model's clinical utility.

Results

Patient characteristics

Table 1 presents comprehensive clinical details pertaining to LSCC patients. The rates of thyroid cartilage invasion were 36.0% in the training cohort, 39.1% in the internal validation cohort, and 37.7% in the external validation cohort. No significant differences were found in age, sex, smoking status, histological grade, alcohol consumption, tumor location, or clinical N stage between patients with and without thyroid cartilage invasion in the training cohort, while the differences in CT-reported AC invasion, clinical T stage, and overall clinical stage between the two groups were statistically significant. In addition, Cohen's kappa value of the ratings for thyroid cartilage invasion was 0.784, indicating good consistency between the two radiologists (Table S2).

Clinical characteristics	Training cohort ($n=247$)			Internal validation cohort ($n=110$)			External validation cohort ($n=61$)		
	Without invasion ($n=158$)	With invasion ($n=89$)	p	Without invasion ($n=67$)	With invasion ($n=43$)	p	Without invasion ($n=38$)	With invasion ($n=23$)	p
Age (years, mean \pm SD)	61.39 \pm 9.06	63.17 \pm 7.53	0.205	63.16 \pm 7.83	63.74 \pm 7.32	0.698	61.08 \pm 6.70	63.57 \pm 6.13	0.153
Sex			1.000			0.027*			0.833
Female	8	4		1	6		3	3	
Male	150	85		66	37		35	20	
Smoking status			0.464			0.609			1.000
Never smoked	12	10		2	3		4	3	
Smoker	146	79		65	40		34	20	
Alcohol consumption			0.476			1.000			1.000
No	45	30		17	11		11	7	
Yes	113	59		50	32		27	16	
Tumor location			0.149			0.173			0.002*
Glottic	126	63		54	29		36	15	
Supraglottic	29	21		12	11		1	8	
Subglottic	3	5		1	3		1	0	
CT-reported AC invasion			0.017*			0.028*			0.251
No	48	14		17	3		15	5	
Yes	110	75		50	40		23	18	
Histological grade			0.939			1.000			0.365
Poorly/poorly-middle	34	18		14	9		8	2	
Middle/well	124	71		53	34		30	21	
Clinical T stage			<0.001*			<0.001*			0.007*
cT1	40	2		22	1		14	2	
cT2	59	1		27	1		13	4	
cT3	57	51		13	25		9	13	
cT4	2	35		5	16		2	4	
Clinical N stage			0.651			0.006			0.051
cN0	126	68		55	24		35	16	
cN1-3	32	21		12	19		3	7	
Overall clinical stage			<0.001*			<0.001*			<0.001*
I	38	4		21	1		14	1	
II	55	2		22	0		13	3	
III	46	41		13	19		9	13	
IV	19	42		11	23		2	6	

Table 1. Baseline clinical characteristics of the patients in the datasets. *HR* hazard ratio, *CI* confidence interval, *AC* anterior commissure, *SD* standard deviation. * $p < 0.05$.

Performance analysis of the radiomics model, the 2D DL model, the 3D DL model, and two readers

Thirteen radiomics features were screened to develop the radiomics model (Table S3 and Figure S1), which yielded an AUC of 0.867 (95% confidence interval [CI]: 0.819–0.916) in the training cohort. Additionally, the 2D DL model achieved an AUC of 0.846 (95% CI: 0.797–0.895), and the 3D DL model achieved the highest AUC of 0.959 (95% CI: 0.934–0.984), as shown in Table 2 and Figure S2.

In two validation cohorts, the 2D DL model exhibited superior performance than other models and two readers (P.J. as reader 1 and L.Q.J. as reader 2, respectively) (Table 2; Fig. 4A, B). Specifically, the 2D DL model achieved superior performance in predicting thyroid cartilage invasion, with AUCs of 0.835 (95% CI: 0.758–0.911) in the internal validation cohort and 0.804 (95% CI: 0.696–0.913) in the external validation cohort. In comparison, the 3D DL model yielded AUCs of 0.732 (95% CI: 0.638–0.827) and 0.698 (95% CI: 0.569–0.836), the radiomics model achieved AUCs of 0.727 (95% CI: 0.621–0.823) and 0.705 (95% CI: 0.567–0.843), reader 1 had AUCs of 0.742 (95% CI: 0.644–0.841) and 0.726 (95% CI: 0.598–0.854), and reader 2 had AUCs of 0.727 (95% CI: 0.630–0.824) and 0.715 (95% CI: 0.790–0.944), respectively—all of which were inferior to the 2D DL model. Detailed comparisons of the AUCs are provided in Figure S3.

The calibration curves of the 2D DL model in two validation cohorts had a good fit ($p > 0.05$ for both, HL test) (Fig. 4C, D). The DCA curves indicated that the 2D DL model provided greater net benefit (Fig. 4E, F). In addition, the Grad-CAM heatmaps confirmed the validity of the 2D DL model (Figure S4).

Development and performance of the nomogram

In this study, three key clinical risk factors for predicting thyroid cartilage invasion were identified via univariate analysis, including CT-reported AC invasion, clinical T stage, and overall clinical stage (Table 1). As shown in Fig. 5, we constructed a nomogram that integrates the 2D DL signature (the optimal model) and the three factors, which achieved further improvement and delivered the best predictive performance. In the internal validation cohort, the nomogram achieved an AUC of 0.867 (95% CI: 0.799–0.936), which was significantly higher than that of the 3D DL model, the radiomics model, and both radiologists ($p < 0.05$ for all, DeLong test), with no statistically significant difference from the 2D DL model. In the external validation cohort, the nomogram achieved an AUC of 0.823 (95% CI: 0.714–0.931), with no statistically significant differences compared to other models ($p > 0.05$ for all, DeLong test) (Fig. S3). The calibration curves and DCA curves demonstrated that the nomogram had good consistency ($p > 0.05$, HL test) and excellent clinical utility (Fig. 4C–F).

Survival prediction

The median follow-up time of the whole dataset (418 patients) was 38.6 months, varying from 1.6 to 82.0 months. More detailed information is shown in Table S4. Before performing cox regression analysis, the proportional hazards assumption was assessed for all variables, as shown in Table S5. Ultimately, the 2D DL signature (hazard ratio [HR] = 4.666) and clinical N stage (HR = 2.191) were found to be independent prognostic factors associated with DFS (Table 3; Fig. 6). By utilizing the ‘survminer’ R package, we calculated the optimal cut-off point (0.51) for 2D DL signature, which stratified LSCC patients into high-risk and low-risk groups (Figure S5). Kaplan-Meier survival analysis revealed that the low-risk group with a lower 2D DL signature had significantly better DFS than the high-risk group with a higher 2D DL signature ($p < 0.05$, log-rank test). Similar results were observed for clinical N stage ($p < 0.05$, log-rank test). Eventually, we developed a prognostic nomogram integrating 2D DL signature and clinical N stage to predict the DFS of LSCC patients, yielding a 3-year AUC of 0.620, a 4-year AUC of 0.650, and a 5-year AUC of 0.653 (Fig. 6E).

Cohort		AUC (95%CI)	Accuracy	Sensitivity	Specificity	PPV	NPV
Training cohort	Radiomics model	0.867 [0.819–0.916]	0.810	0.798	0.816	0.710	0.878
	2D DL model	0.846 [0.797–0.895]	0.794	0.640	0.880	0.750	0.813
	3D DL model	0.959 [0.934–0.984]	0.923	0.888	0.943	0.898	0.937
	Nomogram	0.873 [0.831–0.915]	0.749	0.933	0.646	0.597	0.944
Internal validation cohort	Reader 1	0.742 [0.644–0.841]	0.727	0.558	0.836	0.686	0.747
	Reader 2	0.727 [0.630–0.824]	0.700	0.512	0.821	0.647	0.724
	Radiomics model	0.727 [0.621–0.823]	0.682	0.605	0.731	0.591	0.742
	2D DL model	0.835 [0.758–0.911]	0.791	0.744	0.821	0.727	0.833
	3D DL model	0.732 [0.638–0.827]	0.682	0.674	0.687	0.580	0.774
	Nomogram	0.867 [0.799–0.936]	0.836	0.953	0.761	0.719	0.962
External validation cohort	Reader 1	0.726 [0.598–0.854]	0.689	0.565	0.763	0.591	0.843
	Reader 2	0.715 [0.790–0.944]	0.689	0.478	0.816	0.591	0.744
	Radiomics model	0.705 [0.567–0.843]	0.721	0.391	0.921	0.750	0.714
	2D DL model	0.804 [0.696–0.913]	0.705	0.957	0.553	0.564	0.955
	3D DL model	0.698 [0.569–0.836]	0.607	0.870	0.447	0.488	0.850
	Nomogram	0.823 [0.714–0.931]	0.754	0.826	0.711	0.633	0.871

Table 2. The performance of predictive models and readers. *DL* deep learning, *AUC* area under the curve, *CI* confidence interval, *PPV* positive predictive value, *NPV* negative predictive value.

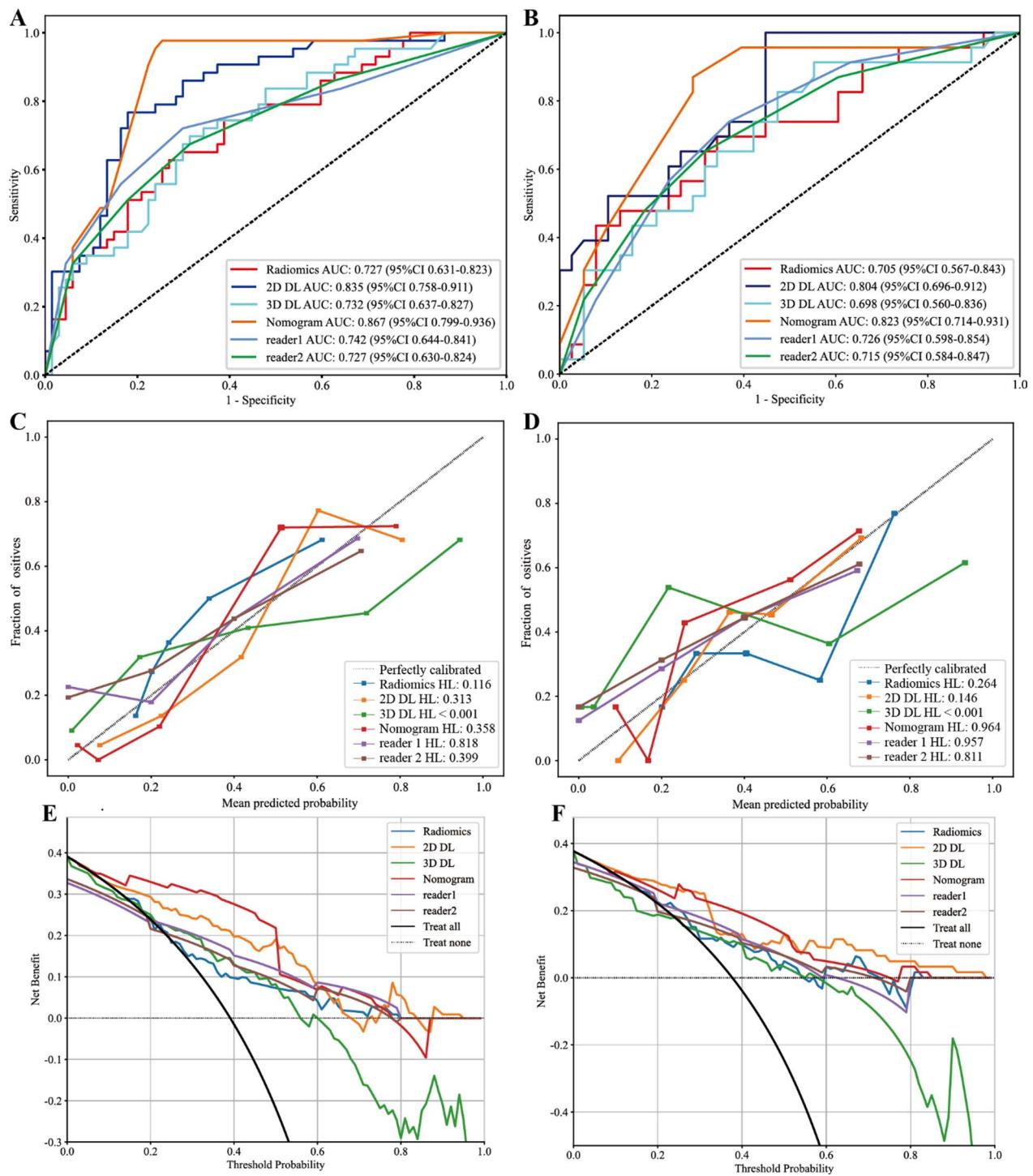


Fig. 4. The ROC curves of different models and readers in the internal validation cohort (A) and the external validation cohort (B), respectively. The calibration curves of different models and readers in the internal validation cohort (C) and external validation cohort (D), respectively. The DCA curves of different models and readers in the internal validation cohort (E) and external validation cohort (F), respectively.

Discussion

In this study, the 2D DL model achieved better performance than the 3D DL model, the radiomics model, and two readers in evaluating thyroid cartilage invasion. In addition, the proposed nomogram incorporating 2D DL signature and clinical risk factors (CT-reported AC invasion, clinical T stage, and overall clinical stage) demonstrated the best classification performance. Furthermore, the 2D DL signature significantly correlated with DFS, offering important information for clinical decision-making and prognosis evaluation for LSCC patients.

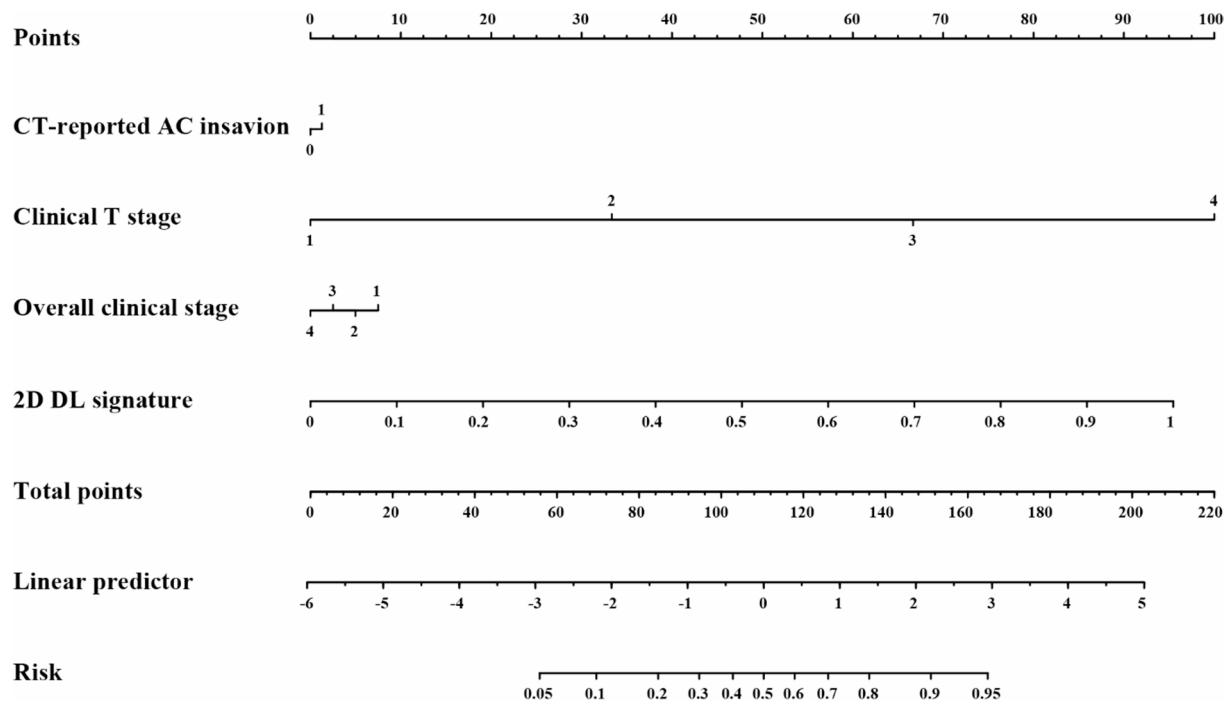


Fig. 5. The nomogram for identifying thyroid cartilage invasion in this study. The comprehensive nomogram was built from 2D DL signature and clinical risk factors, including CT-reported AC invasion, clinical T stage and overall clinical stage.

Clinical characteristics	Univariate cox regression			Multivariate cox regression		
	HR	95% CI	p	HR	95% CI	p
Age	1.941	0.612–6.155	0.260			
Sex	1.016	0.990–1.034	0.236			
Smoking status	1.052	0.457–2.421	0.905			
Alcohol consumption	1.006	0.637–1.589	0.980			
Tumor location						
Glottic	Ref					
Supraglottic	1.517	0.934–2.463	0.092			
Subglottic	0.938	0.285–3.083	0.916			
CT-reported AC invasion	1.205	0.723–2.010	0.475			
Histological grade	1.250	0.676–2.311	0.477			
Clinical T stage						
cT1	Ref			Ref		
cT2	0.868	0.425–1.775	0.699	0.606	0.149–2.466	0.484
cT3	2.134	1.169–3.894	0.014*	1.015	0.221–4.664	0.985
cT4	0.750	0.312–1.806	0.522	0.324	0.059–1.780	0.195
Clinical N stage	2.138	1.384–3.303	0.001*	2.191	1.241–3.869	0.007*
Overall clinical stage						
I	Ref			Ref		
II	0.999	0.478–2.087	0.998	1.206	0.293–4.964	0.795
III	2.112	1.111–4.015	0.023*	1.284	0.263–6.272	0.757
IV	1.593	0.801–3.167	0.184	0.888	0.155–5.074	0.894
2D DL signature	3.961	1.769–8.867	0.001*	4.666	1.861–11.698	0.001*

Table 3. Univariate and multivariate analyses of predictors of DFS. DFS disease-free survival, HR hazard ratio, CI confidence interval, AC anterior commissure, DL deep learning. * $p < 0.05$.

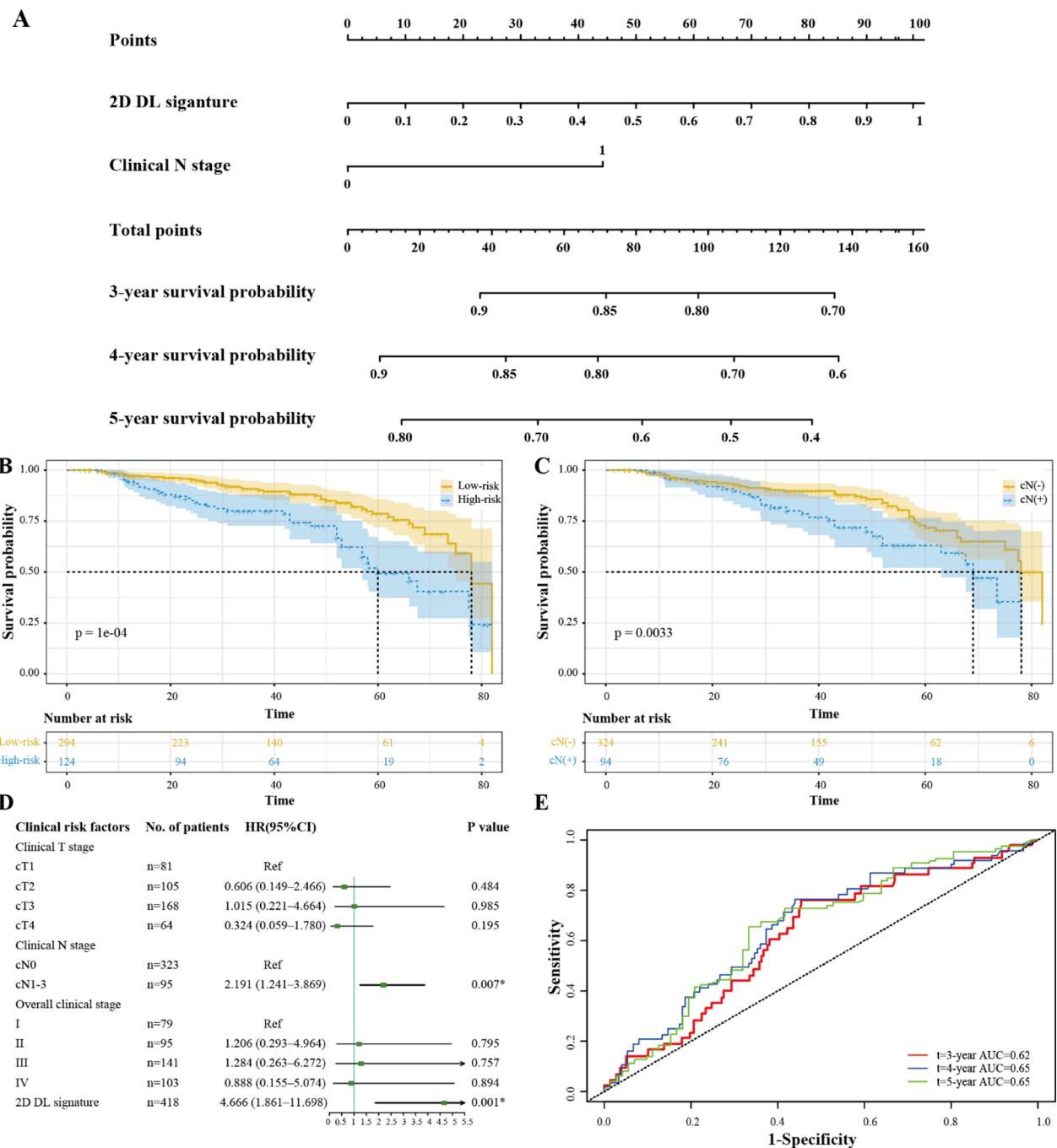


Fig. 6. The relationship between 2D DL signature and patient prognosis in the whole cohort. (A) The 2D DL signature and clinical N stage were found to be independent prognostic factors for LSCC patients, and a corresponding nomogram was developed. Survival analysis stratified by the optimal cut-off value of 2D DL signature (B) and clinical N stage (C) in the whole cohort. Patients with high 2D DL signature or clinical N stage positive (cN+) status had poorer DFS ($p < 0.05$, log-rank test). (D) Forest plot illustrating multivariable cox regression analyses for prognosis in the whole cohort. (E) Time-dependent ROC curves of the nomogram in the whole cohort.

Accurate assessment of thyroid cartilage invasion can assist clinicians in individualized clinical treatment. However, relying on macroscopic appearance of the tumor on CT/MRI to identify thyroid cartilage invasion leads to discrepancies in accuracy. Our findings indicated that the AUCs of the 2D DL model surpassed those of two readers in two validation cohorts. It showed that a considerable proportion of patients were misdiagnosed on the basis of radiologists' visual interpretations^{13,19}, which may arise from inherent limitations of human visual perception. In addition, our study revealed that patients with advanced T stage or overall stage, as well as those with AC involvement, were more susceptible to thyroid cartilage invasion, aligning with other studies^{22,23}.

Recently, radiomics and DL have showcased great potential in tumor staging, molecular typing, and prognosis evaluation^{13,15,24}. Guo et al.¹⁹ established a radiomics model to predict thyroid cartilage invasion in laryngeal and hypopharyngeal carcinoma. By applying a synthetic minority oversampling algorithm to correct data imbalance, they improved the AUC of the model in the training set to 0.905. However, this study did not include an additional validation set. Currently, there still exist many single-center studies on laryngeal cancer, leaving the models' robustness unconfirmed. Moreover, few studies have utilized 3D DL technology in laryngeal carcinoma. This study compared the performance of the radiomics model, 2D DL model, 3D DL model, and two readers in predicting thyroid cartilage invasion, and demonstrated that the 2D DL model and corresponding nomogram displayed outstanding predictive performance. However, the radiomics model revealed unsatisfactory performance, especially since the sensitivity of the external validation set was only 0.391, which may be partially attributed to the differences in CT devices between the two centers^{25,26}. Nevertheless, our study had a relatively large sample size with independent external validation, enhancing the credibility of the findings. Consistent with previous radiomics studies on laryngeal cancer^{13,19,20}, our study employed venous-phase CT images for model construction. This decision was based on the fact that the delayed enhancement allows better visualization of tumor neovasculature and stromal components, thereby enabling more comprehensive and informative features extraction^{13,20}. Although fusing radiomics features from multiple phases may improve model performance, it also can introduce feature redundancy and instability. In certain scenarios, they may fail to provide significant performance gains while increasing model complexity^{27,28}. Given these concerns, we chose to construct our model based solely on venous-phase data.

Considering the complex structure of larynx, this study implemented a 3D DL model to capture the intricate spatial features of laryngeal tumors. The results revealed that the 3D DL model experienced significant overfitting, which may be due to the requirement of a larger dataset for 3D CNNs to analyze effectively²⁹. Additionally, 3D CNNs also necessitate more parameter adjustments and greater storage space^{30,31}. Conversely, the 2D DL model indicated excellent discriminative ability in predicting thyroid cartilage invasion, surpassing other models and two readers. Previous studies have also indicated that the 2D DL model often achieves better results in clinical research^{32,33}. Two possible reasons may account for this preponderance: Firstly, DL can automatically learn more complex and discriminative features from convolutional layers, making it more effective and precise than manually designed radiomics features³⁴. Secondly, the proportion of useful diagnostic information representing lesions in the input data of the 2D DL model may be higher than that in 3D data^{32,35}, rendering it more suitable for clinical diagnostic tasks. Although the rates of thyroid cartilage invasion were comparable across the training, internal, and external cohorts, there were still inherent discrepancies in patient distribution and imaging acquisition protocols between institutions. In addition, the external validation cohort had a relatively limited sample size. These factors likely contributed to the observed performance heterogeneity and, in particular, to the decreased AUC of the 3D DL model on external validation, given the greater sensitivity of 3D networks to sample size and protocol variability. The relatively stable performance of the 2D DL model across cohorts further supported its greater generalizability in the current multi-center setting.

Importantly, the nomogram that fused the 2D DL signature with clinical risk factors (CT-reported AC invasion, clinical T stage and overall clinical stage) achieved the best predictive performance in both validation cohorts; this superiority was statistically significant only in the internal validation cohort, but not in the external validation cohort, which may be attributed to the smaller size of the external validation cohort. From a clinical perspective, such a nomogram could serve as a practical, noninvasive decision-support tool: preoperatively it may help identify patients at high risk of cartilage invasion and thus inform surgical planning, guide multidisciplinary decisions regarding the need for more extensive resection or adjuvant therapy in borderline cases, and enable tailoring of postoperative surveillance intensity according to individualized risk. Moreover, this study observed a close negative correlation between 2D DL signature and DFS, indicating that a lower 2D DL signature was associated with longer DFS. More specifically, patients without thyroid cartilage invasion had significant survival benefits. Our study also identified clinical N stage as another independent prognostic factor, consistent with the findings of previous studies³⁶. Based on these findings, the prognostic nomogram offers a practical tool for clinicians to customize effective therapeutic strategies and evaluate prognosis for LSCC patients.

Notably, some limitations in the study are worth emphasizing. Firstly, this was a retrospective analysis. Although this study included an independent external validation set, the amount of data was limited due to stringent inclusion criteria. Thus, a larger multicenter prospective research is warranted to validate the model presented in this study. Secondly, precise manual delineation of tumor regions required specialized radiologists, and the results were influenced by subjective experience. Therefore, future studies will focus on automatic segmentation algorithms (e.g., U-Net) to improve consistency and reduce human bias. Thirdly, model development in this study was based on the venous-phase CT images. More modalities for modeling may potentially improve the model's performance. Fourthly, there is still a lack of clear understanding on designing network architecture and adjusting training parameters. Therefore, future work should focus on optimizing network architectures and tuning training parameters to enhance model robustness.

Conclusion

In conclusion, we demonstrated that the 2D DL model had notable advantages over the 3D DL model, the radiomics model, and two readers. The integrated nomogram, incorporating 2D DL signature and clinical factors, achieved satisfactory predictive performance for thyroid cartilage invasion and patient prognosis in LSCC. Further investigation is needed to fully generalize the clinical applicability of the proposed nomogram.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 13 January 2025; Accepted: 9 October 2025

Published online: 17 November 2025

References

1. Do, L. et al. Treatment outcomes of T4 locally advanced head and neck cancers with soft tissue invasion or bone and cartilage invasion. *Am. J. Clin. Oncol.* **32**, 477–482. <https://doi.org/10.1097/COC.0b013e31819380a8> (2009).
2. Nakayama, M. & Brandenburg, J. H. Clinical underestimation of laryngeal cancer. Predictive indicators. *Arch. Otolaryngol. Head Neck Surg.* **119**, 950–957. <https://doi.org/10.1001/archotol.1993.01880210038006> (1993).
3. Forastiere, A. A., Ismaila, N. & Wolf, G. T. Use of Larynx-Preservation strategies in the treatment of laryngeal cancer: American society of clinical oncology clinical practice guideline update summary. *J. Oncol. Pract.* **14**, 123–128. <https://doi.org/10.1200/jop.2017.027912> (2018).
4. Dankbaar, J. W. et al. Detection of cartilage invasion in laryngeal carcinoma with dynamic contrast-enhanced CT. *Laryngoscope Investig Otolaryngol.* **2**, 373–379. <https://doi.org/10.1002/lio2.114> (2017).
5. Lefebvre, J. L. et al. Larynx preservation in pyriform sinus cancer: preliminary results of a European organization for research and treatment of cancer phase III trial. EORTC head and neck cancer cooperative group. *J. Natl. Cancer Inst.* **88**, 890–899. <https://doi.org/10.1093/jnci/88.13.890> (1996).
6. Jaipuria, B., Dosemane, D., Kamath, P. M., Sreedharan, S. S. & Shenoy, V. S. Staging of laryngeal and hypopharyngeal cancer: computed tomography versus histopathology. *Iran. J. Otorhinolaryngol.* **30**, 189–194 (2018).
7. Becker, M. et al. Neoplastic invasion of the laryngeal cartilage: reassessment of criteria for diagnosis at CT. *Radiology* **203**, 521–532. <https://doi.org/10.1148/radiology.203.2.9114116> (1997).
8. Hermans, R. Staging of laryngeal and hypopharyngeal cancer: value of imaging studies. *Eur. Radiol.* **16**, 2386–2400. <https://doi.org/10.1007/s00330-006-0301-7> (2006).
9. Preda, L. et al. Diagnostic accuracy of surface coil MRI in assessing cartilaginous invasion in laryngeal tumours: do we need contrast-agent administration? *Eur. Radiol.* **27**, 4690–4698. <https://doi.org/10.1007/s00330-017-4840-x> (2017).
10. Forghani, R. et al. Different spectral Hounsfield unit curve and high-energy virtual monochromatic image characteristics of squamous cell carcinoma compared with nonossified thyroid cartilage. *AJNR Am. J. Neuroradiol.* **36**, 1194–1200. <https://doi.org/10.3174/ajnr.A4253> (2015).
11. Lam, S., Gupta, R., Kelly, H., Curtin, H. D. & Forghani, R. Multiparametric evaluation of head and neck squamous cell carcinoma using a Single-Source Dual-Energy CT with fast kVp switching: state of the Art. *Cancers (Basel.)* **7**, 2201–2216. <https://doi.org/10.3390/cancers7040886> (2015).
12. Kuno, H. et al. Evaluation of cartilage invasion by laryngeal and hypopharyngeal squamous cell carcinoma with dual-energy CT. *Radiology* **265**, 488–496. <https://doi.org/10.1148/radiol.12111719> (2012).
13. Chen, X. et al. A combined model integrating radiomics and deep learning based on Contrast-Enhanced CT for preoperative staging of laryngeal carcinoma. *Acad. Radiol.* **30**, 3022–3031. <https://doi.org/10.1016/j.acra.2023.06.029> (2023).
14. Yao, Y. et al. Applying a nomogram based on preoperative CT to predict early recurrence of laryngeal squamous cell carcinoma after surgery. *J. Xray Sci. Technol.* **31**, 435–452. <https://doi.org/10.3233/xst-221320> (2023).
15. Kang, C. et al. CT radiomics nomogram predicts pathological response after induced chemotherapy and overall survival in patients with advanced laryngeal cancer: A single-center retrospective study. *Front. Oncol.* **13**, 1094768. <https://doi.org/10.3389/fonc.2023.1094768> (2023).
16. Liu, X. et al. CT-based radiomics signature analysis for evaluation of response to induction chemotherapy and progression-free survival in locally advanced hypopharyngeal carcinoma. *Eur. Radiol.* **32**, 7755–7766. <https://doi.org/10.1007/s00330-022-08859-4> (2022).
17. Baek, S. Towards single 2D Image-Level Self-Supervision for 3D human pose and shape estimation. *Appl. Sci.* **11** (2021).
18. Chen, S. et al. Using stacked deep learning models based on PET/CT images and clinical data to predict EGFR mutations in lung cancer. *Front. Med. (Lausanne)* **9**, 1041034. <https://doi.org/10.3389/fmed.2022.1041034> (2022).
19. Guo, R. et al. CT-based radiomics features in the prediction of thyroid cartilage invasion from laryngeal and hypopharyngeal squamous cell carcinoma. *Cancer Imaging* **20**, 81. <https://doi.org/10.1186/s40644-020-00359-2> (2020).
20. Zhao, X. et al. Radiomics analysis of CT imaging improves preoperative prediction of cervical lymph node metastasis in laryngeal squamous cell carcinoma. *Eur. Radiol.* **33**, 1121–1131. <https://doi.org/10.1007/s00330-022-09051-4> (2023).
21. Pietragalla, M. et al. Current role of computed tomography imaging in the evaluation of cartilage invasion by laryngeal carcinoma. *Radiol. Med.* **125**, 1301–1310. <https://doi.org/10.1007/s11547-020-01213-y> (2020).
22. Liu, Q. et al. Machine learning model to preoperatively predict T2/T3 staging of laryngeal and hypopharyngeal cancer based on the CT radiomic signature. *Eur. Radiol.* <https://doi.org/10.1007/s00330-023-10557-8> (2024).
23. Yener, H. M. et al. The incidence of thyroid cartilage invasion in early-stage laryngeal carcinoma: our experience on sixty-two patients. *Clin. Otolaryngol.* **43**, 388–392. <https://doi.org/10.1111/coa.13004> (2018).
24. Chen, L., Wang, H., Zeng, H., Zhang, Y. & Ma, X. Evaluation of CT-based radiomics signature and nomogram as prognostic markers in patients with laryngeal squamous cell carcinoma. *Cancer Imaging* **20**, 28. <https://doi.org/10.1186/s40644-020-00310-5> (2020).
25. Jimenez-Del-Toro, O. et al. The discriminative power and stability of radiomics features with computed tomography variations: Task-Based analysis in an anthropomorphic 3D-Printed CT Phantom. *Invest. Radiol.* **56**, 820–825. <https://doi.org/10.1097/rlr.0000000000000795> (2021).
26. Meyer, M. et al. Reproducibility of CT radiomic features within the same patient: influence of radiation dose and CT reconstruction settings. *Radiology* **293**, 583–591. <https://doi.org/10.1148/radiol.2019190928> (2019).
27. Wang, W. et al. Comparing three-dimensional and two-dimensional deep-learning, radiomics, and fusion models for predicting occult lymph node metastasis in laryngeal squamous cell carcinoma based on CT imaging: a multicentre, retrospective, diagnostic study. *EClinicalMedicine* **67**, 102385. <https://doi.org/10.1016/j.eclim.2023.102385> (2024).
28. Zhao, W., Huang, X., Wang, G. & Guo, J. PET/MR fusion texture analysis for the clinical outcome prediction in soft-tissue sarcoma. *Cancer Imaging* **22**, 7. <https://doi.org/10.1186/s40644-021-00438-y> (2022).
29. Singh, S. P. et al. 3D deep learning on medical images: A review. *Sens. (Basel.)* **20** <https://doi.org/10.3390/s20185097> (2020).
30. Jiang, H. et al. An automatic detection system of lung nodule based on multigroup Patch-Based deep learning network. *IEEE J. Biomed. Health Inf.* **22**, 1227–1237. <https://doi.org/10.1109/jbhi.2017.2725903> (2018).
31. Xie, Z. Z. & Yongdong. Automated pulmonary nodule detection in CT images using deep convolutional neural networks. *Pattern Recognition: J. Pattern Recognit. Soc.* **85** (2019).
32. Kim, H. et al. CT-based deep learning model to differentiate invasive pulmonary adenocarcinomas appearing as subsolid nodules among surgical candidates: comparison of the diagnostic performance with a size-based logistic model and radiologists. *Eur. Radiol.* **30**, 3295–3305. <https://doi.org/10.1007/s00330-019-06628-4> (2020).

33. Yu, Q. et al. Deep learning-assisted diagnosis of benign and malignant Parotid tumors based on contrast-enhanced CT: a multicenter study. *Eur. Radiol.* **33**, 6054–6065. <https://doi.org/10.1007/s00330-023-09568-2> (2023).
34. Avanzo, M. et al. Machine and deep learning methods for radiomics. *Med. Phys.* **47**, e185–e202. <https://doi.org/10.1002/mp.13678> (2020).
35. Wang, T. et al. Prediction of microvascular invasion in hepatocellular carcinoma based on preoperative Gd-EOB-DTPA-enhanced MRI: comparison of predictive performance among 2D, 2D-expansion and 3D deep learning models. *Front. Oncol.* **13**, 987781. <https://doi.org/10.3389/fonc.2023.987781> (2023).
36. Cui, J. et al. Development and validation of nomograms to accurately predict risk of recurrence for patients with laryngeal squamous cell carcinoma: cohort study. *Int. J. Surg.* **76**, 163–170. <https://doi.org/10.1016/j.ijsu.2020.03.010> (2020).

Acknowledgements

We thank onekey AI for technical assistance.

Author contributions

Xinwei Chen: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. Shuang Xia: Conceptualization, Data curation, Funding acquisition, Resources, Validation. Huan Jiang: Conceptualization, Formal analysis, Investigation, Methodology, Software. Fajin Lv: Data curation, Formal analysis, Resources, Supervision. Qiang Yu: Formal analysis, Methodology, Software. Youquan Ning: Methodology. Kai Xie: Investigation. Quanjiang Li: Formal analysis. Rui Liu: Formal analysis. Yuan Zhou: Methodology. Guohua Hu: Conceptualization, Data curation, Methodology. Juan Peng: Conceptualization, Funding acquisition, Resources, Writing – review & editing. All authors contributed to data collection and study design, and approved the final manuscript.

Funding

This work was supported by the Foundation of Science and Technology Bureau of Yuzhong District, Chongqing, China (20190111), the Natural Science Foundation of Chongqing, China (cstc2021jcyj-msxmX0020), the Graduate Supervisor Team Construction Project of the First Clinical College, Chongqing Medical University (CYYY-DSTDXM202507), the Natural Scientific Foundation of Tianjin (21CYBJC01580), and Tianjin Key Medical Discipline (Specialty) Construction Project (TJYXZDXK-041 A).

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval and informed consent

This retrospective study was conducted in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board of the First Affiliated Hospital of Chongqing Medical University (approval number: 2020080). Due to retrospective nature of the study, informed consent was waived by the Institutional Review Board of the First Affiliated Hospital of Chongqing Medical University.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-23809-y>.

Correspondence and requests for materials should be addressed to G.H. or J.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025