Am J Cancer Res 2026;16(4):1678-1689
www.ajcr.us /ISSN:2156-6976/ajcr0171570

Original Article

Deep-learning-based radiomics of intratumoral and peritumoral CT images to differentiate benign from malignant parotid tumors

Qian Shen[1,2*], Jiuzhou Dai[3*], Yongliang Han[2], Xiaojuan Liu[3*], Kui Huang[4*]

[1]Department of Radiology, The Affiliated Stomatology Hospital of Southwest Medical University, Luzhou 646000, Sichuan, China; [2]Department of Radiology, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China; [3]School of Artificial Intelligence, Chongqing University of Technology, Chongqing 400016, China; [4]Department of Oral and Maxillofacial Surgery, The Affiliated Stomatology Hospital of Southwest Medical University, Luzhou 646000, Sichuan, China. [*]Equal contributors.

Received January 20, 2026; Accepted April 25, 2026; Epub April 25, 2026; Published April 30, 2026

Abstract: This study aimed to explore the value of radiomics and deep learning features derived from intratumoral and peritumoral regions based on CT images in distinguishing benign from malignant parotid gland tumors (PGTs), and to evaluate whether integrating clinical, radiomics, and deep learning features could improve differential diagnostic efficiency. We retrospectively collected CT imaging and clinical data from 357 patients with confirmed parotid gland tumors (PGTs) after surgery. The dataset was randomly divided into two sets at a ratio of 7:3. Radiomics and deep learning features were extracted from both intratumoral and peritumoral regions. The least absolute shrinkage and selection operator (LASSO) regression method was applied for feature selection. Traditional radiomics (RAD), deep learning (DL), and fusion (RAD+DL) models were constructed using logistic regression (LR). A combined predictive model was further developed by integrating the optimal model with clinical and CT features. Model performance was assessed using receiver operating characteristic (ROC) curves, calibration curves, and decision curve analysis (DCA). The fusion model (Tumor+External2_RAD+DL) incorporating both radiomics and deep learning features from intratumoral and peritumoral regions achieved the best performance, with an AUC of 0.881 (95% CI: 0.809-0.943) in the test set. After integrating clinical features (including symptoms, margin, and enlarged lymph nodes), the combined model showed the highest predictive performance, with an AUC of 0.913 (95% CI: 0.842-0.966) and an accuracy of 0.880, significantly outperforming the clinical model alone and models using only intratumoral or peritumoral features from a single region. In conclusion, the fusion model integrating radiomics and deep learning features from intratumoral and peritumoral regions on CT images effectively distinguishes benign from malignant PGTs. The combined model incorporating clinical factors achieves superior predictive performance, demonstrating good potential for clinical application.

Keywords: Parotid gland tumor, radiomics, deep learning, peritumoral, computed tomography

Introduction

Parotid gland tumors account for approximately 80% of salivary gland tumors, among which some 80% are benign, and approximately 20% of them exhibit malignant behavior [1]. The treatment strategies and prognosis vary depending on the different subtypes of PGTs [2]. Therefore, accurate preoperative diagnosis is crucial for formulating clinical treatment decisions in patients with PGTs. Currently, fine-needle aspiration (FNA) biopsy and imaging examinations are the most common preoperative diagnostic methods for parotid gland tumors, but the FNA carries the risks of tumor cell transplantation and inducing parotitis [3], the radiological images can provide valuable information for preoperative assessment of parotid gland tumors. Although MRI provides exceptionally good resolution of soft tissues, CT can provide useful insights into tumor characteristics and surrounding structures at a lower cost, and it takes less time and is more convenient to operate than MRI [4]. It has been widely applied in

https://doi.org/10.62347/QDFD4776

the preoperative localization diagnosis of parotid gland tumors. However, parotid gland tumors have a complex variety of histopathological types, which leads to overlapping imaging features for some tumors, imaging examinations sometimes lead to uncertain results, the conclusion mainly depends on the professional knowledge and subjective experience of radiologists [5].

Radiomics can extract useful information from high-throughput medical images, provide more comprehensive tumor descriptions, and improve the diagnostic performance and clinical predictive ability of diseases [6-8]. It has been widely applied in the early diagnosis, prognosis assessment and therapeutic effect evaluation of tumors [9, 10]. Previous studies have demonstrated that the application of CT radiomics analysis could assist radiologists in the preoperative differential diagnosis of salivary gland tumors [11, 12]. However, the predictive performance of radiomics model depends heavily on the accurate manual delineation of regions of interest. Deep learning leverages convolutional neural networks to automatically extract complex features directly from raw image pixels, achieving a high degree of automation. Early researches have illustrated that the application of deep learning models can significantly improve the overall work efficiency of radiologists [13]. Nevertheless, establishing a deep learning model requires a comprehensive and sizable data set, it is undeniable that, under the current medical conditions, this remains a challenge.

In recent years, the combination of radiomics, with its rich quantitative features, and deep learning, with its powerful feature representation ability, has become an emerging research trend and attracted more and more attention. This complex fusion feature set has been proven effective in several clinical fields, such as breast cancer prediction [14], lymph node metastasis of breast cancer [15], and the auxiliary diagnosis of pulmonary nodules [16]. To our knowledge, there are few literatures on the application of complex fusion feature set for accurate preoperative differentiation of parotid gland tumors. Besides, tumor biology studies have demonstrated the tumor-surrounding micro-environment plays a crucial role in tumor initiation, progression, and metastasis [17, 18]. Integrated analysis of the radiomics information of the intratumoral and peritumoral regions can provide a more comprehensive evaluation of the infiltration characteristics of malignant tumors. Currently, the differentiation between BPGT and MPGT through the extraction of radiomics or deep learning features from multiple regions has not been fully explored.

Accordingly, we proposed to establish a fusion model based on the CT images. The model aimed to enhance the differential diagnostic efficiency of BPGT and MPGT by taking the complementary advantages of both radiomics and deep learning, as well as the combined diagnostic value of the intratumoral and peritumoral regions. By integrating the two types of imaging features from multiple regions, we expect to enhance the accuracy of preoperative predictive diagnosis of parotid gland tumors.

Material and methods

Patients

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and this retrospective study was approved by the ethics committees of The First Affiliated Hospital of Chongqing Medical University (approval number: K2025-674), with patient consent waived. Patient sets were retrospectively collected from January 2014 to January 2023 at The First Affiliated Hospital of Chongqing Medical University. The exclusion criteria and the inclusion criteria were presented in Supplementary Appendix 1. Finally, 357 patients were collected in this study, including 89 MPGT patients and 268 BPGT patients. The cohort was randomly divided into a training set (n= 249; BPGT=187, MPGT=62) and a testing set (n=108; BPGT=81, MPGT=27) at a ratio of 7:3. The patients' data were collected from the Picture Archiving and Communication System (PACS).

Image acquisition

All participants underwent axial contrast-enhanced scanning using multi-slice spiral computed tomography equipment. The scanning range extended from the base of the skull to the thoracic inlet. After the plain scan was completed, the contrast agent was injected at a flow rate of 2 to 4 ml/s, followed by 30 ml of

normal saline, and the dosage of contrast agent was 1-2 ml/kg. Detailed imaging parameters of various scanning devices were available in the Supplementary Table 1.

Clinical and radiological data analysis

Patient clinical data and images were acquired from our hospital's clinical records and PACS system. We retrospectively collected clinical factors including age, sex, smoking/drinking status, and symptoms. All CT images were retrospectively and independently reviewed by two radiologists (with 5 and 10 years of experience, respectively), who were blinded to the histopathological diagnoses. When their evaluations were inconsistent, consensus was reached through consultation. We also collected the clinical features, including age, sex, smoking, drinking, symptoms and the postoperative pathological results. The radiological features including maximum diameter, location, density, cystic, shape, margin, number, calcification, enhanced uniformity, enlarged lymph nodes, and enhancement degree. The detailed definitions of these radiological features were provided in Supplementary Appendix 2.

Image segmentation

All the CT images of the patients were stored in the Digital Imaging and Communications in Medicine (DICOM) format, and the standard soft tissue settings (window width, 400 HU; window level, 40 HU) were used. The region of interest (ROI) was segmented manually from axial contrast-enhanced CT images using ITK-SNAP software (version 3.8.0, http://www.itk-snap. org/) by two experienced radiologists, they were also blinded to the clinical and histopathological data of the tumor and contoured the outer edge of the tumor layer-by-layer, the adjacent normal tissue, bone, and vessels were not included.

To evaluate the consistency of the extracted features, the intra- and inter-observer reproducibility were quantified using the intraclass correlation coefficients (ICC). And 20 cases of arterial-phase enhanced CT images were randomly selected, the tumor VOI segmentation was independently conducted by radiologist-1 and radiologist-2 for assessing the inter-observer agreement of the extracted radiomics features. Radiologist-1 then carried out the same case procedure after two weeks, and the ICC values over 0.9 demonstrated good consistency.

After manual tumor segmentation, 2-mm peritumoral regions were automatically expanded by using Python software (version 3.7.12; http://www. python.org), the bone and air were then filtered from the delineation by setting the maximum (400 HU) and minimum (-200 HU) thresholds, the final ROI border (External2) was manually refined (see Supplementary Figure 1) [19]. A complete schematic was presented in **Figure 1**.

Radiomics feature extraction and classification model building

PyRadiomics in Python was utilized for image preprocessing and feature extraction. To eliminate the potential impact of diverse CT scanning equipment, a $1 \times 1 \times 1$ mm^3 voxel spacing was applied to resample the CT images. And in order to ensure the consistency of the results, all images were preprocessed by the techniques of resampling and standardization.

We utilized the Pyradiomics toolkit to extract 1874 radiomics features from the Tumor and External2, respectively. Among these captures features, the first-order features capture information related to pixel intensity distribution, whereas shape features characterize the contour and spatial structure of the lesions; as well as the texture features, which are used to capture the subtle differences of the lesion. The screening and dimensionality reduction of radiomics features involve multiple stages: firstly, we used analysis of variance (ANOVA) to identify the highly correlated features with ICC scores > 0.9. Next, we reduce the dimension of the feature space by comparing the similarity of each pair of features. If the Pearson correlation coefficient value is higher than 0.95, one of the features is deleted. Then, the least absolute shrinkage and selection operator (LASSO) algorithm with ten-fold cross validation was executed to optimize the penalty parameter (λ) and identify the most robust predictor variables. At the optimal λ, the features with nonzero coefficients are retained and sorted based on the absolute value of the coefficients. According to the empirical rule, each feature corresponds to 10 samples in a binary classifi-

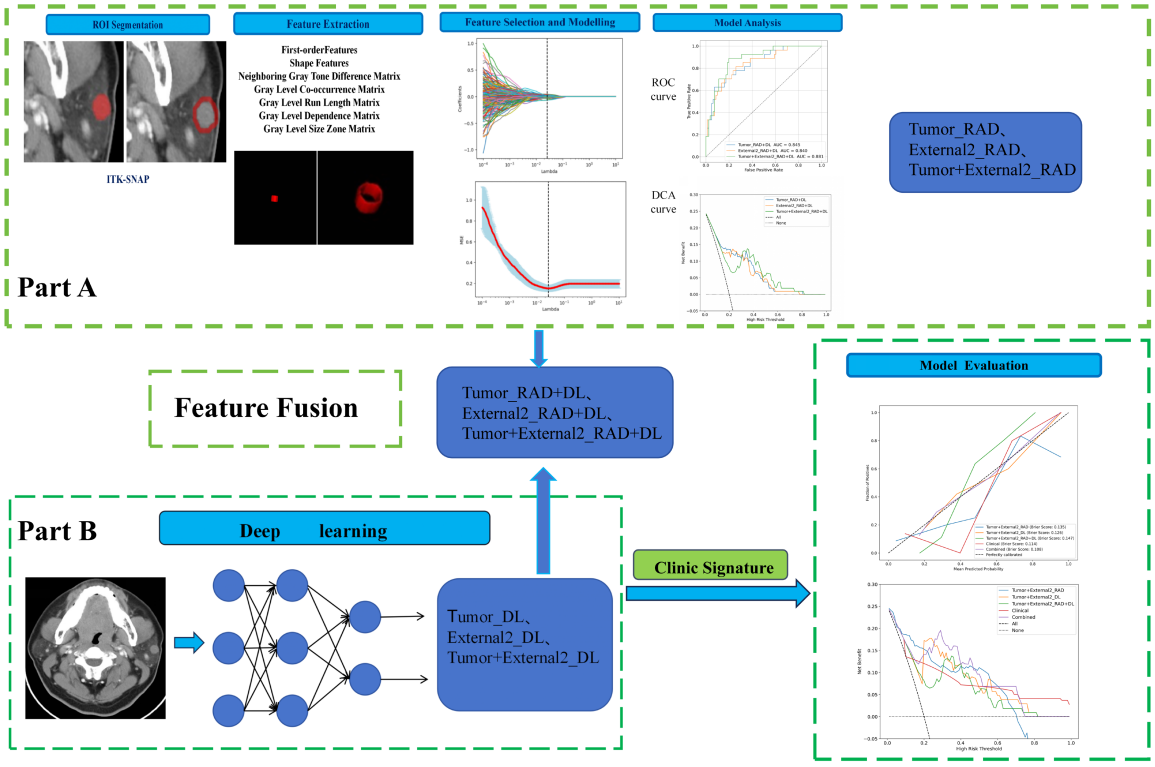


Figure 1. Workflow of this study, Part A the radiomics model, extracting and fusing radiomics features from the Tumor and External2. Part B the deep learning model, extracting and fusing deep learning features from the Tumor and External2; Tumor intra-tumoral region, External2 2-mm peri-tumoral region, Tumor+External2 multi-region.

er. To reduce the risk of over-fitting and avoid the curse of dimensionality, only the top 20 ranked features were ultimately selected as the discriminative features [20]. All feature selection procedures were conducted on the training set and then applied to the test set.

Ultimately, those valuable features extracted from the training datasets were used to develop the radiomics model based on the logistic regression (LR) classifier. We used features from each region to construct single-region radiomics models (Tumor_RAD, External2_RAD). In addition, a fusion radiomics model (Tumor+External2_RAD) was established by combining the radiomics features from the intratumoral (Tumor) and peritumoral (External2) regions.

Deep-learning feature extraction and classification model building

MobileNet V3 was selected to extract image features, because it had the minimum loss value, and exhibited faster convergence than any other CNN, and given the constraints of a limited data-set [13], all data were pretrained on the ImageNet data-set to obtain the initial weight values. An input size of 96 × 96 × 16 pixels for the deep learning model was selected. To enhance the algorithm's efficiency and reduce the degree of over-fitting, we also performed data augmentation on CT images.

We established a multi-input residual network to classify BPGT and MPGT. This network was made up of three components: a Tumor network, an External2 network, and a feature fusion (Tumor+External2) network. We used the External2 network and Tumor network to extract peritumoral and intratumoral image features, respectively. The feature fusion network performed matrix addition to fuse corresponding same-scale features from the Tumor and External2 networks. Next, the convolution operation was employed to capture the differential features between these two regions. We transformed the feature maps at each scale into feature vectors by using adaptive max pooling, and then concatenated the feature

vectors of different scales to acquire deep-learning features for each patient. Then, the Tumor_DL model, the External2_DL model and the Tumor+External2_DL model were established based on MobileNet V3, which was selected to extract image features at different scales.

In the process of constructing the fusion model (Tumor_RAD+DL, External2_RAD+DL, Tumor+External2_RAD+DL), we employed the logistic regression machine learning algorithm to integrate radiomics features and deep learning features. A five-fold cross-validation method was adopted, and the test cohort was determined to ensure fair comparisons. Then, we compared the differential diagnostic efficacy in differentiating BGPT from MPGT of all models, the optimal model and the independent clinical predictors were integrated to establish a combined model.

Performance evaluation of the models

The each model's performance was assessed in the test sets by reporting AUC, ROC analysis, accuracy, positive predictive value (PPV), sensitivity, negative predictive value (NPV) and specificity. Then, for the model with the best performance, we conducted the sensitivity and specificity analyses using the confusion matrix. The observed outcomes and the predicted probabilities were assessed by calibration curves, which were quantified through the utilization of Brier scores [21]. The AUCs between models were compared using the Delong's test. The decision curve analysis (DCA) was applied to evaluate the clinical utility of these models.

Statistical analysis

Statistical analysis was performed using PyRadiomics in Python (version 3.7.12; http://www.python.org) and SPSS version 26.0 (version 26.0; IBM, Armonk, NY, USA). Univariate analysis was used to compare differences in the clinical information and radiological features between the training set and the test set, and the valuable factors were screened out by using univariate and multivariable logistic regression analysis. For quantitative data, normality and homogeneity of variances were first assessed. Data following a normal distribution were presented as mean ± standard deviation (SD) and compared using the independent samples t-test; otherwise, data were presented as median with interquartile range (IQR) and compared using the non-parametric Mann-Whitney U test. Categorical variables were shown as numbers and percentages and analyzed by Chi-square test or Fisher's exact test, as appropriate. One-way ANOVA, Pearson correlation analysis and LASSO regression were used to compare the values of the different radiomics features for distinguishing between BPGT and MPTG. Model performance was evaluated using accuracy, sensitivity, specificity, receiver operating characteristic (ROC) curve analysis, and the area under the curve (AUC). The Delong method was employed to compare AUCs and determine whether the combined model significantly outperformed the other models. $P < 0.05$ was considered statistically significant.

Results

The population and radiological features of patients

Ultimately, this study collected 357 patients (89 MPGT, 268 BPGT), the cohort was randomly split 3:7 into a test set of 108 patients (81 BPGT, 27 MPGT) and a training set of 249 patients (187 BPGT, 62 MPGT). The clinical and CT imaging features of patients in the training and test set were compared in Supplementary Table 2, there were no statistically significant differences between the clinical and radiological characteristics of the two groups, ensuring an unbiased data partition (all $P > 0.05$). Significant differences were noted in distinguishing BPGT from MPGT in terms of clinical symptoms, tumor margin status and enlarged lymph nodes ($P < 0.05$; see Supplementary Table 3), the clinical model was constructed using the above independent predictors, with AUCs of 0.889 and 0.765 in the training and test sets, respectively.

Feature selection and model performance

Radiomics models: For the images of each patient, 1874 radiomics features were extracted from the Tumor and External2, respectively. After feature selection and dimensionality reduction, nineteen, twenty and twenty discriminative radiomics features were selected in the training set, respectively (Supplementary

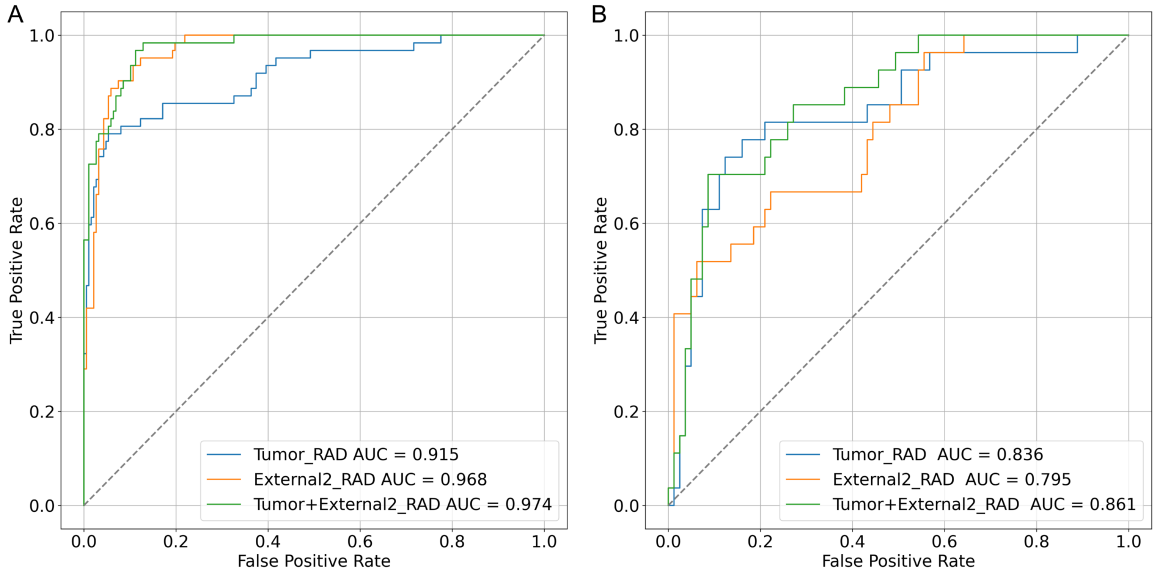


Figure 2. The ROC curves of radiomics models in the training set (A) and test set (B). Among these models, the Tumor+External2_RAD model yielded superior predictive performance compared with the Tumor_RAD model and the External2_RAD model, achieving AUCs of 0.974 and 0.861 in the training and test sets, respectively. AUC, area under the curves.

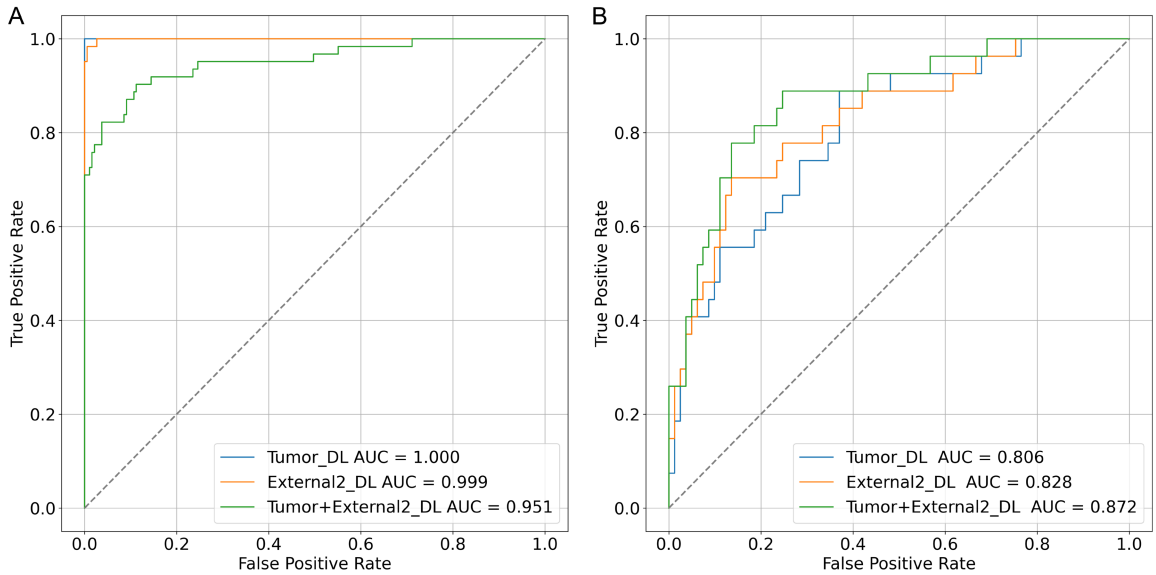


Figure 3. The ROC curves of deep learning models based on different regions in the training set and test set. ROC curves of Tumor_DL model, External2_DL model and Tumor+External2_DL model in (A) the training set and (B) the test set were demonstrated and compared.

Figure 2), and were used to construct the traditional radiomics models. Among these models, the Tumor+External2_RAD model yielded superior predictive performance compared with the Tumor_RAD model and the External2_RAD model, achieving AUCs of 0.974 and 0.861 in the training and test sets, respectively (**Figure 2**).

DL models: The MobileNet V3 was employed to extract deep learning features from the Tumor and External2 regions, respectively. Meanwhile, we used the fusion network to integrate these features of different scales, then obtaining the deep-learning fusion features. Finally, three deep learning models (Tumor_DL model, External2_DL model and Tumor+External2_DL

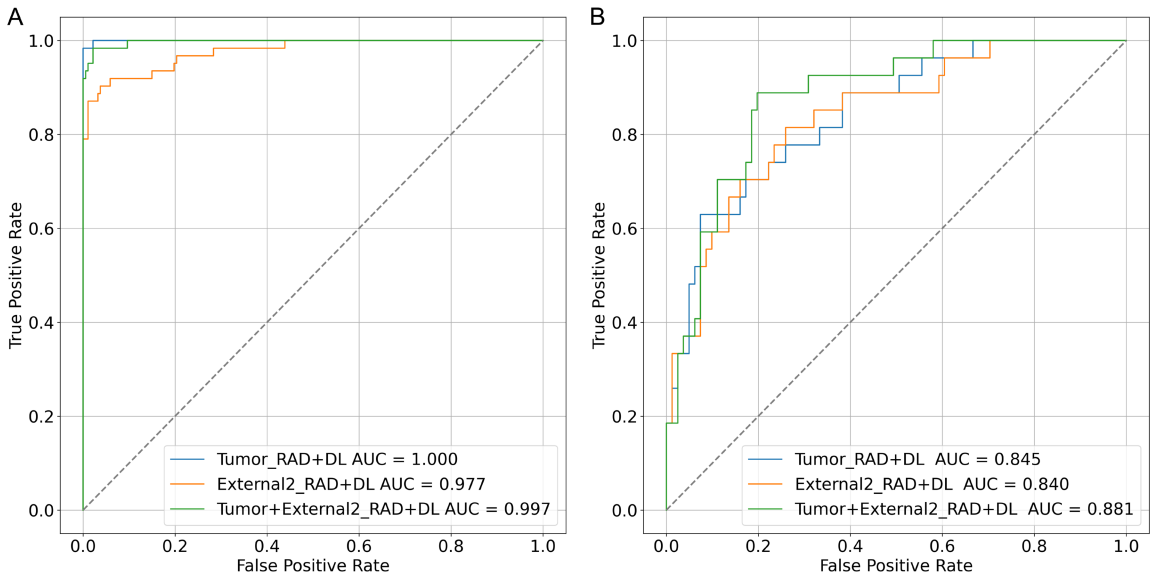


Figure 4. The ROC curves of deep-learning radiomics models based on different regions in the training set and test set. ROC curves of Tumor_RAD+DL model, External2_RAD+DL model and Tumor+External2_RAD+DL model in (A) the training set and (B) the test set were demonstrated and compared.

model) were established, and the Tumor+External2_DL model demonstrated superior performance compared to the Tumor_DL model (AUC=0.806) and the External2_DL model (AUC=0.812), achieving an AUC of 0.872 (95% CI: 0.795-0.938) in the test set (**Figure 3**).

Feature fusion models: We employed the logistic regression machine learning algorithm to integrate radiomics features and deep learning features to establish three deep learning radiomics models (Tumor_RAD+DL, External2_RAD+DL, Tumor+External2_RAD+DL). As was shown in the **Figure 4**, the Tumor+External2_RAD+DL model achieved an AUC of 0.881 (95% CI: 0.809-0.943) in the test set. This performance was further enhancement compared to that of both the Tumor_RAD+DL model (AUC=0.881 vs. 0.845) and the External2_RAD+DL model (AUC=0.881 vs. 0.840).

Construction of combine model and comparison of all models

By integrating radiomics features from both intratumoral and peritumoral regions, the model yielded a higher discriminative ability, surpassing the performance of models utilizing features from only one of these regions, and the Tumor+External2_RAD+DL model achieved a superior AUC of 0.881 (95% [CI]: 0.809-0.943) and was regarded as the optimal model, then combining the meaningful clinical features (symptom, margin, and enlarged lymph nodes), the final combined model was established and illustrated that the deep-learning radiomics features played a significant role in distinguishing BPGT from MPGT.

Table 1 presented the predictive performance of the clinical model, radiomics model, deep learning model, feature fusion model, and combined model, the combined model exhibited the best performance with an AUC of 0.913 (95% [CI]: 0.842-0.966) in the test set (**Figure 5**). According to **Table 1**, Delong test revealed statistically significant differences between combined model compared to the clinical model (AUC=0.765, P=0.015) and the single region's radiomics model (in the test set: AUC=0.836, P=0.032; AUC=0.795, P=0.034) or the single region's deep learning model (in the test set: AUC=0.806, P=0.030; AUC=0.828, P=0.004). Furthermore, the combined model had a higher AUC than the Tumor+External2_RAD model, the Tumor+External2_DL model as well as the deep-learning radiomics models, but there was no significant difference (P > 0.05).

The 2 × 2 confusion matrices were shown in Supplementary Figure 3, the results showed

Table 1. The performance of different models

Model	AUC	*P value*	95% CI	ACC	SEN	SPE
Tumor_RAD	0.836	0.032*	0.738-0.918	0.787	0.741	0.852
External2_RAD	0.795	0.034*	0.690-0.885	0.731	0.667	0.753
Tumor+External2_RAD	0.861	0.255	0.779-0.930	0.796	0.704	0.815
Tumor_DL	0.806	0.030*	0.707-0.893	0.574	0.333	0.963
External2_DL	0.828	0.004*	0.727-0.914	0.759	0.741	0.765
Tumor+External2_DL	0.872	0.342	0.795-0.938	0.806	0.926	0.457
Tumor_RAD+DL	0.845	0.150	0.749-0.917	0.769	0.704	0.790
External2_RAD+DL	0.840	0.186	0.747-0.917	0.685	0.667	0.840
Tumor+External2_RAD+DL	0.881	0.458	0.809-0.943	0.824	0.926	0.605
Clinical	0.765	0.015*	0.657-0.872	0.787	0.593	0.852
Combined	0.913		0.842-0.966	0.880	0.778	0.914

Abbreviations: *P* values were derived from the DeLong test for AUCs between the other ten models and the combined model, AUC: area under curve, CI: confidence interval, ACC: accuracy, SEN: sensitivity, SPE: specificity. *Represents P < 0.05.

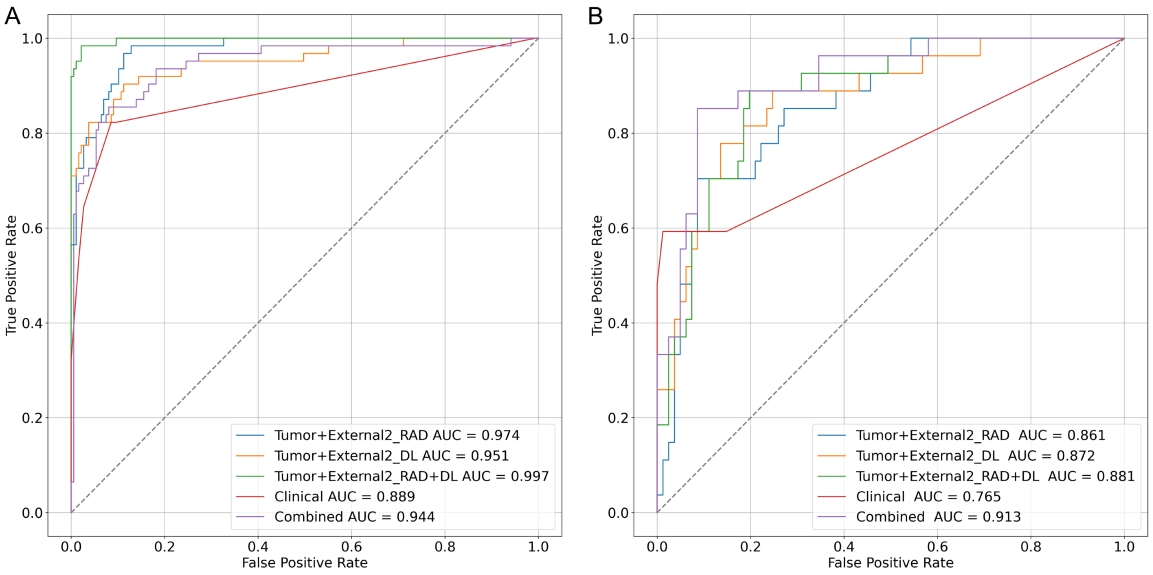


Figure 5. The ROC curves of the top performing model in different datasets. ROC curves of Tumor+External2_RAD model, Tumor+External2_DL model, Tumor+External2_RAD+DL model, Clinical model and Combined model in (A) the training set and (B) the test set were demonstrated and compared.

that our proposed combined model achieved superior diagnostic accuracy compared to senior radiologists. This analysis provides a clearer understanding of the clinical utility of the model, suggesting its potential to support less experienced radiologists in improving diagnostic performance. **Figure 6** presented the calibration curves (A) and DCA curves (B) for the five models in the test set, the calibration curves with Brier scores shown good calibration of the combined model for distinguishing the benign and malignant parotid gland tumors, the DCA curves demonstrated that using the combined model to predict the benign and malignant parotid gland tumors had a high clinical net benefit.

Discussion

In this study, we employed radiomics and deep learning methods to distinguish BPGT from MPTG by analyzing features extracted from different regions. To further enhanced the classification performance, we constructed fusion models by combining radiomics and deep learning features extracted from different regions,

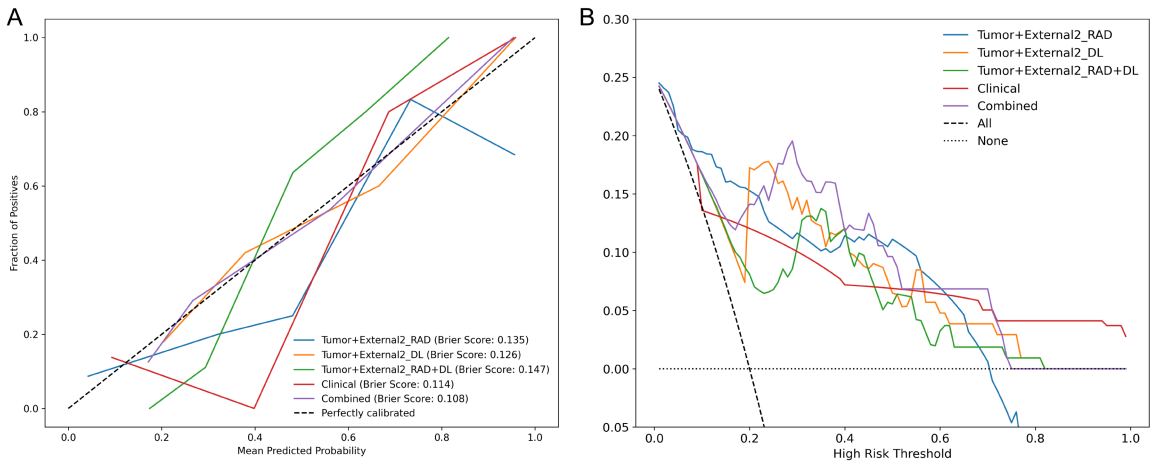


Figure 6. The calibration curves and DCA curves of different models in the test set. A: Calibration curves, the dotted diagonal line represents an ideal evaluation, while the solid lines represent the performance of the model. Closer to the dotted diagonal line indicates better evaluation. B: DCA curves of the clinical, radiomics, deep-learning, deep-learning radiomics and combined model.

we found that the fusion model combining both intratumoral and peritumoral radiomic features achieved a higher AUC compared to the models with the single region's features. The Tumor+External2_RAD+DL model demonstrated better classification performance. Furthermore, the clinical and CT imaging features also provided valuable information for the model construction, then the final combined model, integrating all available data, achieved the best performance (AUC=0.913), as well as the DCA curve indicated that the adoption of the combined model would bring more clinical net benefits to patients.

It is well known that adequate clinical characteristics and radiological information facilitate accurate preoperative differentiation of parotid gland tumors. Studies have shown that irregular shape, diffuse growth pattern, unclear margin, surrounding tissue involvement, and as well as lymphatic metastasis of tumor are more likely to be considered as MPGT [22, 23]. In this study, we found that having symptoms, lymph node enlargement and unclear margins were more common in malignant tumors than that in benign tumors, the AUC of clinical model based on clinical-radiological features were 0.765 in the test set, which suggested that the differentiation model developed only by clinical-radiological features could not accurately distinguish between BPGT and MPGT.

Radiomics and deep learning are non-invasive methods, which construct models based on the original images through intelligent calculation to obtain more information, thereby reflecting the true biological information of tumor [20, 24]. Most previous studies have predominantly concentrated on the tumor parenchym [4, 12, 25], with comparatively little attention paid to the surrounding tissues of the tumor, while it can reveal the important biological information about the tumor, including the possibility of its malignant behavior and its interaction with the surrounding tissues.

Therefore, we explored the significance of the image features of the peritumoral region for differentiating BPGT from MPTG. Our findings indicated that models utilizing a combination of intratumoral and peritumoral features - whether based on radiomics or deep learning - consistently achieved higher accuracy than those models with only intratumoral radiomic features. It revealed that for tumor differential diagnosis, valuable information was contained not only within the tumor itself, but also in the peritumoral region, which provided critical supplementary data. This result was consistent with previous studies, which displayed the important role of peri-tumoral regions in the classification of the parotid gland tumors [26]. The intratumoral and peritumoral features of the tumor represented different imaging phenotypes of the disease. Therefore, integrating the intratumoral and peritumoral radiomic features is a feasible method, which could significantly enhance the classification performance of the model.

Besides, we also developed fusion models that integrating radiomics and deep learning features to distinguish between BPGT and MPTG. Feature extraction plays an important role in radiomics, but conventional radiomics features are often highly standardized, which may ignore the individual differences among patients [27]. Deep learning networks can autonomously extract high-level features specific to research problems, thereby capturing a more comprehensive representation of lesion characteristics. By fusing the radiomics features with deep learning features, a complex feature set has been formed, which has been applied to the diagnosis of various diseases [28, 29]. It has been demonstrated that models integrating deep learning with radiomics features achieved superior performance compared to models using either features alone in diverse clinical problems, including renal cystic lesions [28], meningiomas [30], and tuberculosis [31]. We fused and compared the image features extracted by these two methods, our results demonstrated that the fusion models were superior to the models using either feature alone in distinguishing BPGT from MPTG, among those models, the Tumor+External2_RAD+DL model was the optimal model, the AUC achieved 0.881(95% [CI]: 0.809-0.943) in the test set, and this result was consistent with those of previous studies [32]. Through integrating the complementary strengths of radiomics (rich quantitative features) and deep learning (powerful feature representation), the fusion model can effectively improve the capacity of tumor differentiation. Then, combining the meaningful clinical factors, the final combined model was built, with the AUC of 0.913, ACC of 0.880, which was significantly higher than that of the clinical model and the single region's intratumoral or peritumoral radiomics model or deep learning model (DeLong's test, $P < 0.05$). These results demonstrated that integrating those deep learning, radiomics, and clinical data produced a more accurate predictive tool by leveraging the unique strengths of each individual component. The strategy of integrating multiple types of data information shows considerable potential to improve personalized clinical decision-making.

However, despite the promising results, our study still has some limitations. Firstly, as a single-center retrospective study, our findings are subject to potential selection bias. The models' stability and generalizability could be strengthened through external validation using independent, multi-center patient cohorts. Secondly, the final segmentation of the peritumoral area was determined through a two-step process: automatic software delineation, followed by necessary manual adjustments, which may cause some human errors. Therefore, we tend to use a fully automatic method, which may improve the stability and reproducibility of the results, and could be applied in future studies. In addition, a significant number of patients were lost to follow-up, the association between model predictions and patient prognosis could not be explored, the future research will require further expansion of the follow-up dataset with a larger number of patients. Finally, although our combined model integrates clinical and radiological information, radiomics and deep learning features, the prospective clinical trial studies are still necessarily needed to validate its effectiveness before this diagnostic assistance system can be applied in clinical practice.

Conclusion

The fusion model, which integrating the radiomics features and deep learning features extracted from the both intratumoral and peritumoral regions, demonstrates excellent discriminatory ability for distinguishing between BPGT and MPGT. The integration of these quantitative imaging biomarkers with the independent clinical predictors, we constructed a combined model that was associated with superior diagnostic accuracy and increased clinical net benefit. With continued expansion of the dataset and further model optimization, this model can accurately differentiate BPGT from MPGT and has the potential to serve as a clinical diagnostic aid, contributing to the advancement of precision medicine.

Acknowledgements

We would like to thank ShanXiong Chen, Ph.D., for constructive criticism of this manuscript. This study was supported by the Sichuan Science and Technology Innovation and Entrepreneurship Seedling Project (2024JDRC0050); the Doctoral Initiation Program of the Affiliated Stomatological Hospital of Southwest Medical

University (2023BS02 and 2025BS05); the Youth science climbing program of the Affiliated Stomatological Hospital of Southwest Medical University (2023KQ02); the Specialized Research Project on Dentistry of Southwest Medical University (2025KQZX18).

Disclosure of conflict of interest

None.

Address correspondence to: Xiaojuan Liu, School of Artificial Intelligence, Chongqing University of Technology, No. 69 Red Road, Banan District, Chongqing 400016, China. E-mail: Liuxiaojuan0127@cqut.edu.cn; Kui Huang, Department of Oral and Maxillofacial Surgery, The Affiliated Stomatological Hospital of Southwest Medical University, No. 2, Jiang Yang Nan Road, Luzhou 646000, Sichuan, China. E-mail: huangkui@swmu.edu.cn

References

[1] Feng B, Wang Z, Cui J, Li J, Xu H, Yu D, Zeng Q and Xiu J. Distinguishing parotid polymorphic adenoma and Warthin tumor based on the CT radiomics nomogram: a multicenter study. Acad Radiol 2023; 30: 717-726.

[2] Lewis AG, Tong T and Maghami E. Diagnosis and management of malignant salivary gland tumors of the parotid gland. Otolaryngol Clin North Am 2016; 49: 343-380.

[3] He Z, Mao Y, Lu S, Tan L, Xiao J, Tan P, Zhang H, Li G, Yan H, Tan J, Huang D, Qiu Y, Zhang X, Wang X and Liu Y. Machine learning-based radiomics for histological classification of parotid tumors using morphological MRI: a comparative study. Eur Radiol 2022; 32: 8099-8110.

[4] Xu Y, Shu Z, Song G, Liu Y, Pang P, Wen X and Gong X. The role of preoperative computed tomography radiomics in distinguishing benign and malignant tumors of the parotid gland. Front Oncol 2021; 11: 634452.

[5] Vogl TJ, Albrecht MH, Nour-Eldin NA, Ackermann H, Maataoui A, Stöver T, Bickford MW and Stark-Paulsen T. Assessment of salivary gland tumors using MRI and CT: impact of experience on diagnostic accuracy. Radiol Med 2018; 123: 105-116.

[6] Gong Z, Li J, Han Y, Chen S and Wang L. Nomogram combining dual-energy computed tomography features and radiomics for differentiating parotid warthin tumor from pleomorphic adenoma: a retrospective study. Front Oncol 2025; 15: 1505385.

[7] Chen Y, Xu W, Li YL, Liu W, Sah BK, Wang L, Xu Z, Wels M, Zheng Y, Yan M, Zhang H, Ma Q, Zhu Z and Li C. CT-based radiomics showing generalization to predict tumor regression grade for advanced gastric cancer treated with neoadjuvant chemotherapy. Front Oncol 2022; 25: 758863.

[8] Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P and Cook G. Introduction to radiomics. J Nucl Med 2020; 61: 488-495.

[9] Li MD, Lu XZ, Liu JF, Chen B, Xu M, Xie XY, Lu MD, Kuang M, Wang W, Shen SL and Chen LD. Preoperative survival prediction in intrahepatic cholangiocarcinoma using an ultrasound-based radiographic-radiomics signature. J Ultrasound Med 2022; 41: 1483-1495.

[10] Jiang M, Li CL, Luo XM, Chuan ZR, Chen RX, Tang SC, Lv WZ, Cui XW and Dietrich CF. Radiomics model based on shear-wave elastography in the assessment of axillary lymph node status in early-stage breast cancer. Eur Radiol 2022; 32: 2313-2325.

[11] Zheng M, Chen Q, Ge Y, Yang L, Tian Y, Liu C, Wang P and Deng K. Development and validation of CT-based radiomics nomogram for the classification of benign parotid gland tumors. Med Phys 2023; 50: 947-957.

[12] Zheng Y, Zhou D, Liu H and Wen M. CT-based radiomics analysis of different machine learning models for differentiating benign and malignant parotid tumors. Eur Radiol 2022; 32: 6953-6964.

[13] Yu Q, Ning Y, Wang A, Li S, Gu J, Li Q, Chen X, Lv F, Zhang X, Yue Q and Peng J. Deep learning-assisted diagnosis of benign and malignant parotid tumors based on contrast-enhanced CT: a multicenter study. Eur Radiol 2023; 33: 6054-6065.

[14] Furtney I, Bradley R and Kabuka MR. Patient graph deep learning to predict breast cancer molecular subtype. IEEE/ACM Trans Comput Biol Bioinform 2023; 20: 3117-3127.

[15] Li X, Yang L and Jiao X. Comparison of traditional radiomics, deep learning radiomics and fusion methods for axillary lymph node metastasis prediction in breast cancer. Acad Radiol 2023; 30: 1281-1287.

[16] Hu X, Gong J, Zhou W, Li H, Wang S, Wei M, Peng W and Gu Y. Computer-aided diagnosis of ground glass pulmonary nodule by fusing deep learning and radiomics features. Phys Med Biol 2021; 66: 065015.

[17] Semenza GL. The hypoxic tumor microenvironment: a driving force for breast cancer progression. Biochim Biophys Acta 2016; 1863: 382-391.

[18] Jiang X, Wang J, Deng X, Xiong F, Zhang S, Gong Z, Li X, Cao K, Deng H, He Y, Liao Q, Xiang B, Zhou M, Guo C, Zeng Z, Li G, Li X and Xiong W. The role of microenvironment in tumor angiogenesis. J Exp Clin Cancer Res 2020; 39: 204.

[19] Lin CH, Yan JL, Yap WK, Kang CJ, Chang YC, Tsai TY, Chang KP, Liao CT, Hsu CL, Chou WC, Wang HM, Huang PW, Fan KH, Huang BS, Tung-Chieh Chang J, Tu SJ and Lin CY. Prognostic value of interim CT-based peritumoral and intratumoral radiomics in laryngeal and hypopharyngeal cancer patients undergoing definitive radiotherapy. Radiother Oncol 2023; 189: 109938.

[20] Gillies RJ, Kinahan PE and Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016; 278: 563-577.

[21] Mao N, Shi Y, Lian C, Wang Z, Zhang K, Xie H, Zhang H, Chen Q, Cheng G, Xu C and Dai Y. Intratumoral and peritumoral radiomics for preoperative prediction of neoadjuvant chemotherapy effect in breast cancer based on contrast-enhanced spectral mammography. Eur Radiol 2022; 32: 3207-3219.

[22] Yu Q, Wang A, Gu J, Li Q, Ning Y, Peng J, Lv F and Zhang X. Multiphasic CT-based radiomics analysis for the differentiation of benign and malignant parotid tumors. Front Oncol 2022; 12: 913898.

[23] Aasen MH, Hutz MJ, Yuhan BT and Britt CJ. Deep lobe parotid tumors: a systematic review and meta-analysis. Otolaryngol Head Neck Surg 2022; 166: 60-67.

[24] Wei W, Liu Z, Rong Y, Zhou B, Bai Y, Wei W, Wang S, Wang M, Guo Y and Tian J. A computed tomography-based radiomic prognostic marker of advanced high-grade serous ovarian cancer recurrence: a multicenter study. Front Oncol 2019; 9: 255.

[25] Shen XM, Mao L, Yang ZY, Chai ZK, Sun TG, Xu Y and Sun ZJ. Deep learning-assisted diagnosis of parotid gland tumors by using contrast-enhanced CT imaging. Oral Dis 2023; 29: 3325-3336.

[26] Zhang C, Shi N, Wang Y, Hao M and Ren J. Clinical value of intratumoral and peritumoral CT radiomics models for discriminating benign and malignant parotid gland tumors. Front Oncol 2025; 15: 1650943.

[27] Wei W, Jia G, Wu Z, Wang T, Wang H, Wei K, Cheng C, Liu Z and Zuo C. A multidomain fusion model of radiomics and deep learning to discriminate between PDAC and AIP based on 18F-FDG PET/CT images. Jpn J Radiol 2023; 41: 417-427.

[28] He QH, Feng JJ, Lv FJ, Jiang Q and Xiao MZ. Deep learning and radiomic feature-based blending ensemble classifier for malignancy risk prediction in cystic renal lesions. Insights Imaging 2023; 14: 6.

[29] Yang D, Ren G, Ni R, Huang YH, Lam NFD, Sun H, Wan SBN, Wong MFE, Chan KK, Tsang HCH, Xu L, Wu TC, Kong FS, Wáng YXJ, Qin J, Chan LWC, Ying M and Cai J. Deep learning attention-guided radiomics for COVID-19 chest radiograph classification. Quant Imaging Med Surg 2023; 13: 572-584.

[30] Zhang J, Zhao Y, Lu Y, Li P, Dang S, Li X, Yin B and Zhao L. Meningioma consistency assessment based on the fusion of deep learning features and radiomics features. Eur J Radiol 2024; 170: 111250.

[31] Nijiati M, Guo L, Abulizi A, Fan S, Wubuli A, Tuersun A, Nijiati P, Xia L, Hong K and Zou X. Deep learning and radiomics of longitudinal CT scans for early prediction of tuberculosis treatment outcomes. Eur J Radiol 2023; 169: 111180.

[32] Li L, Zhou X, Cui W, Li Y, Liu T, Yuan G, Peng Y and Zheng J. Combining radiomics and deep learning features of intra-tumoral and peri-tumoral regions for the classification of breast cancer lung metastasis and primary lung cancer with low-dose CT. J Cancer Res Clin Oncol 2023; 149: 15469-15478.

Supplementary Appendix

Supplementary Appendix 1. The inclusion criteria and the exclusion criteria.

The inclusion criteria were as follows: (1) no history of radiotherapy, chemotherapy and FNAB; (2) patients have complete clinical and imaging data; (3) patients underwent preoperative CT enhanced scan.

The exclusion criteria were as follows: (1) evident artifacts or severe noise on CT images; (2) maximum diameter less than 10 mm; (3) patients with other tumour diseases.

Supplementary Appendix 2. The definitions of the imaging features.

Max-diameter: The sizes of the tumors were measured by determining the maximal cross-sectional diameter.

Symptoms (with or without): We assessed the symptoms based on the clinical record, including pain/tenderness or facial nerve palsy.

Number: We observed the lesions of parotid tumor patients on the picture archiving and communication systems (PACS) of our hospital. If there was only one lesion, it was considered single, and if there were two or more lesions, it was multiple.

Location: The lobes were divided by the retromandibular vein into shallow and deep lobes. Therefore, the location of the tumor was defined as in deep lobe if it is medial to the retromandibular vein, or in superficial lobe lateral to the retromandibular vein. If the tumor grows across two lobes, it was defined as in both lobes.

Density: The homogeneous or heterogeneous density of the lesion was assessed on the non-contrast CT.

Calcification: Calcification was defined as the CT value of the foci within the tumor is higher than 100 Hu.

Cystic areas: cystic area was defined as having a CT scan attenuation of 20 HU or less.

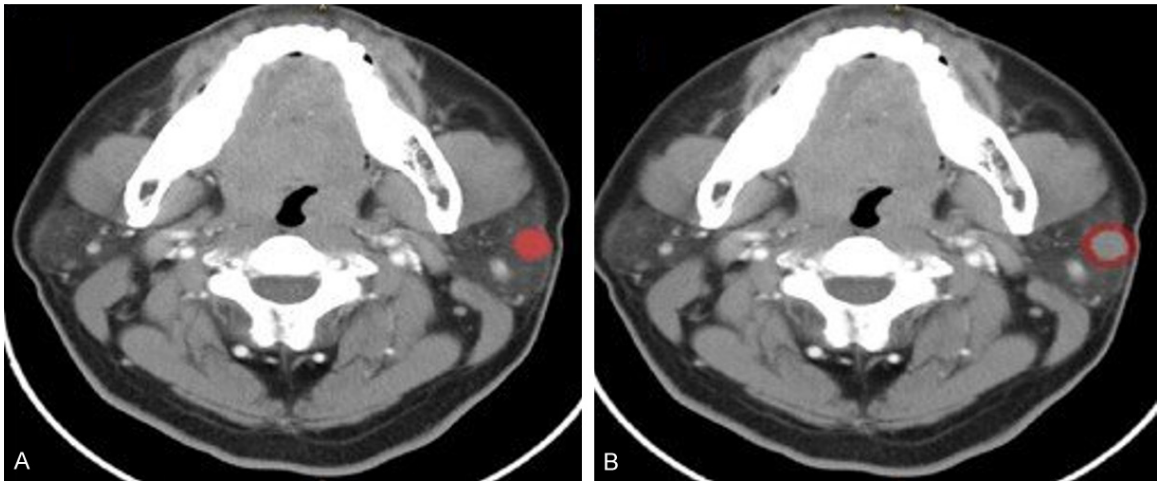
Enhancement degree: Obvious enhancement was defined as the CT value of tumor enhancement on postcontrast CT is 40 Hu higher than it on non-enhanced CT scan. Slight enhancement was defined as the CT value of tumor enhancement on postcontrast CT below 20 Hu on the basis of non-enhanced CT value. Moderate enhanced CT values fell somewhere in between.

Enlarged lymph nodes (with or without): We evaluated ipsilateral lymph node metastases based on imaging features and intraoperative records. The maximal axial dimension criteria for metastatic lymph nodes on imaging were > 15 mm for level I and II nodes, 8 mm for retropharyngeal nodes and 10 mm for all other node levels.

Supplementary Table 1. The scanning parameters of different CT scanners

Parameters	Center1		
CT scanners	Discovery CT750 HD	SOMATOM Definition Flash	SOMATOM Definition Force
Tube voltage	100-120 kV	100-120 kV	100 kV
Tube current	Automatic tube-current	Automatic tube-current	Automatic tube-current
Gantry rotation time	0.6 s	0.5 s	0.28 s
Detector collimation	64 × 0.625 mm	128 × 0.6 mm	128 × 0.6 mm
section thickness	5 mm	5 mm	5 mm
section interval	5 mm	5 mm	5 mm
Image matrix	512×512	512×512	512×512
Contrast agent type	Omnipaque	Ioversol	Ioversol
Contrast agent concentration	300 mgI/mL	320 mgI/mL	320 mgI/mL
Contrast agent dosage	1.5 mL/kg	1.2 mL/kg	1.2 mL/kg
Contrast agent infused rate	3.0-4.0 mL/s	3.0-4.0 mL/s	3.0-4.0 mL/s
Arterial phase scan	30 s after the contrast injection	25 s after the contrast injection	25 s after the contrast injection

Note: CT, computed tomography.



Supplementary Figure 1. An example of manual segmentation of Tumor and External2 on parotid gland tumor CT images. Two ROIs, (A) Inter-tumoral region; (B) 2 mm-peri-tumoral region.

Supplementary Table 2. Clinical and CT imaging features of patients in the training and test set

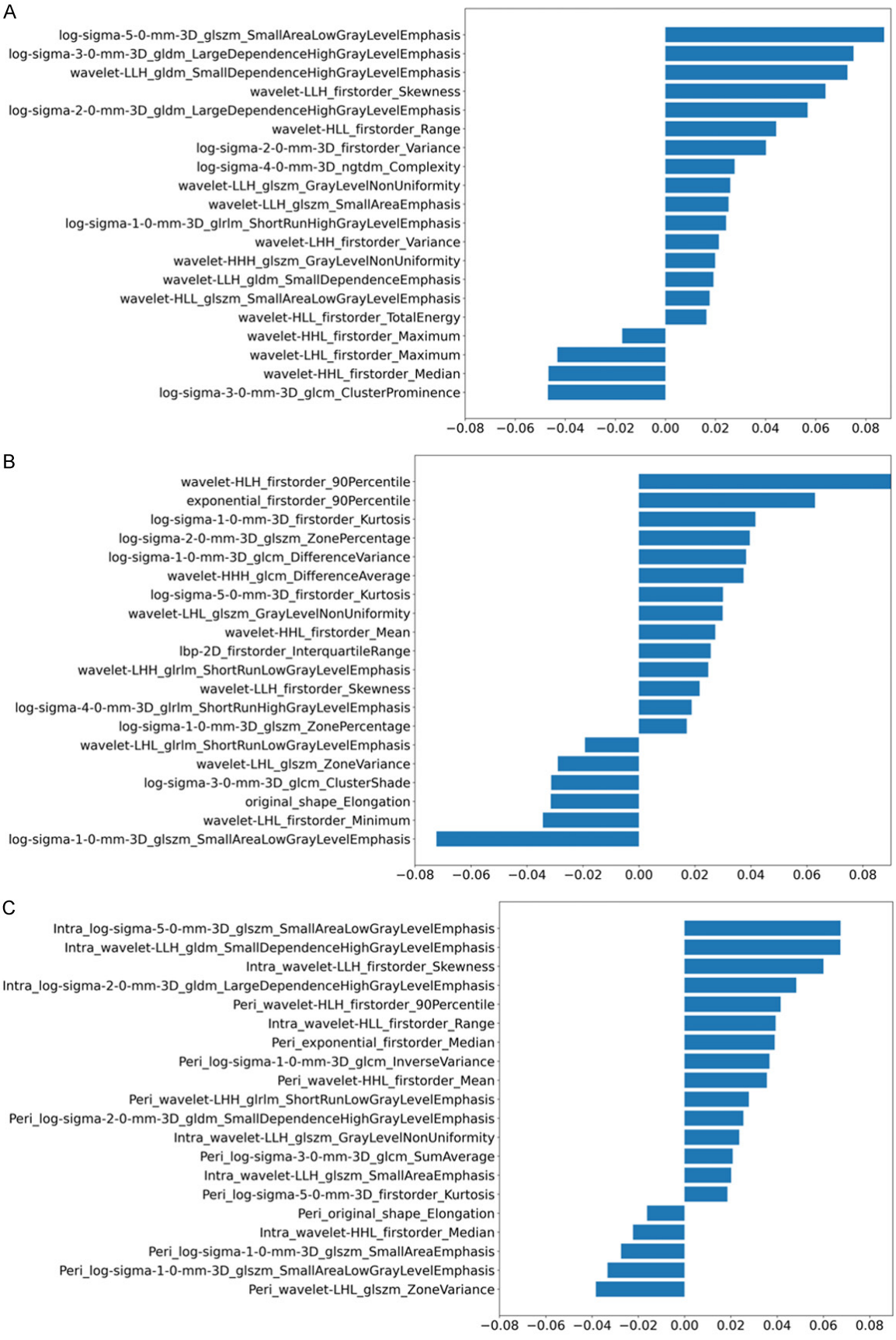
Variables	Training set (n=249)	Test set (n=108)	*P* value
Age[b] (years)	53.181±14.351	51.574±15.248	0.341
Max-diameter[b] (cm)	2.287±0.800	2.281±0.879	0.948
Sex[a] (F/M)	96/153	48/60	0.297
Smoking[a] (Yes/No)	117/132	45/63	0.354
Drinking[a] (Yes/No)	88/161	33/75	0.380
Symptom[a] (With/Without)	36/213	22/86	0.164
Shape[a] (Round/Non-round)	57/192	23/85	0.740
Margin[a] (Clear/Unclear)	214/35	95/13	0.607
Number[a] (Single/Multiple)	229/20	101/7	0.611
Location[a] (Superficial/Deep/Both)	210/5/34	87/3/18	0.649
Density[a] (Homogeneous/Homogeneous)	149/100	64/44	0.918
Calcification[a] (With/Without)	10/239	2/106	0.470
Cystic areas[a] (With/Without)	69/180	29/79	0.867
Enhancement degree[a] (Slight/Moderate/Obvious)	11/77/161	9/36/63	0.259
Enhanced uniformity[a] (Yes/No)	92/157	46/62	0.314
Enlarged lymph nodes[a] (With/Without)	21/228	8/100	0.744

*Represents P < 0.05. [a]Categorical data are presented as numbers (n). [b]Quantitative data are presented as means (standard deviations) or medians (quartiles), *P* value was calculated using the independent samples t-test or Mann-Whitney U test. *P*-value was calculated with the χ^2 or Fisher's exact test.

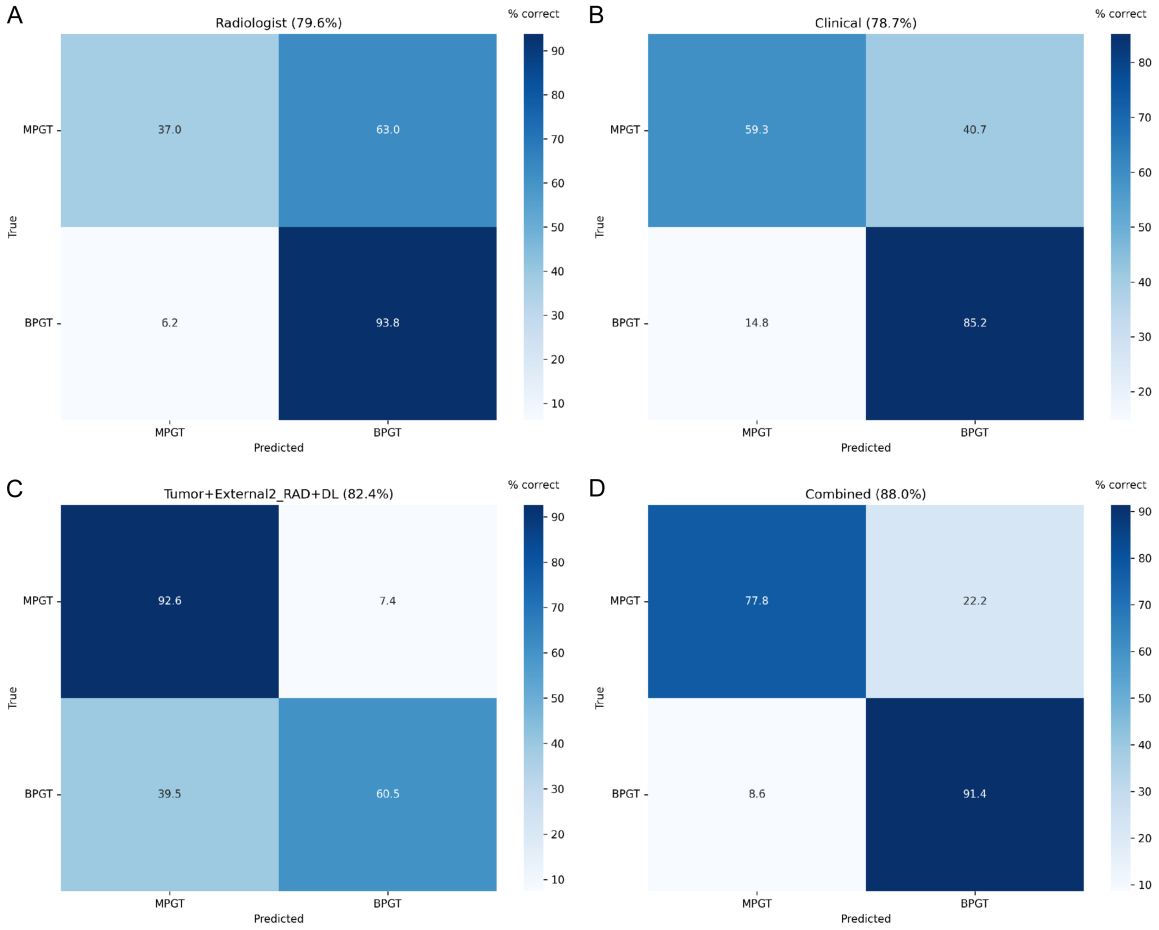
Supplementary Table 3. Univariate and multivariate logistic regression analysis of clinical and imaging features of benign and malignant parotid glands

Variables	*Univariate analysis*		*Multivariate analysis*	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age (years)	1.005 (0.963-1.049)	0.808		
Max-diameter (cm)	0.590 (0.257-1.355)	0.214		
Sex	0.969 (0.240-3.909)	0.964		
Smoking	0.596 (0.139-2.554)	0.486		
Drinking	0.250 (0.062-1.003)	0.050		
Number	0.055 (0.000-74.990)	0.431		
Symptom	40.152 (7.818-206.206)	< 0.001*	0.043 (0.014-0.134)	< 0.001*
Shape	0.288 (0.065-1.269)	0.100		
Margin	0.009 (0.001-0.059)	< 0.001*	100.446 (22.458-449.262)	< 0.001*
Location	3.060 (0.718-13.032)	0.130		
Density	2.105 (0.382-11.590)	0.393		
Calcification	0.222 (0.006-8.600)	0.420		
Cystic areas	0.558 (0.108-2.894)	0.487		
Enhancement degree	19.451 (1.229-307.889)	0.035*	0.145 (0.012-1.693)	0.124
Enhanced uniformity	0.386 (0.095-1.573)	0.184		
Enlarged lymph nodes	384.530 (16.560-7335.452)	< 0.001*	0.014 (0.002-0.092)	< 0.001*

*Represents P < 0.05. OR, odds ratio. CI, confidence interval.


A
log-sigma-5-0-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis
log-sigma-3-0-mm-3D_gldm_LargeDependenceHighGrayLevelEmphasis
wavelet-LLH_gldm_SmallDependenceHighGrayLevelEmphasis
wavelet-LLH_firstorder_Skewness
log-sigma-2-0-mm-3D_gldm_LargeDependenceHighGrayLevelEmphasis
wavelet-HLL_firstorder_Range
log-sigma-2-0-mm-3D_firstorder_Variance
log-sigma-4-0-mm-3D_ngtdm_Complexity
wavelet-LLH_glszm_GrayLevelNonUniformity
wavelet-LLH_glszm_SmallAreaEmphasis
log-sigma-1-0-mm-3D_glrlm_ShortRunHighGrayLevelEmphasis
wavelet-LHH_firstorder_Variance
wavelet-HHH_glszm_GrayLevelNonUniformity
wavelet-LLH_gldm_SmallDependenceEmphasis
wavelet-HLL_glszm_SmallAreaLowGrayLevelEmphasis
wavelet-HLL_firstorder_TotalEnergy
wavelet-HHL_firstorder_Maximum
wavelet-LHL_firstorder_Maximum
wavelet-HHL_firstorder_Median
log-sigma-3-0-mm-3D_glcm_ClusterProminence
−0.08 −0.06 −0.04 −0.02 0.00 0.02 0.04 0.06 0.08
B
wavelet-HLH_firstorder_90Percentile
exponential_firstorder_90Percentile
log-sigma-1-0-mm-3D_firstorder_Kurtosis
log-sigma-2-0-mm-3D_glszm_ZonePercentage
log-sigma-1-0-mm-3D_glcm_DifferenceVariance
wavelet-HHH_glcm_DifferenceAverage
log-sigma-5-0-mm-3D_firstorder_Kurtosis
wavelet-LHL_glszm_GrayLevelNonUniformity
wavelet-HHL_firstorder_Mean
lbp-2D_firstorder_InterquartileRange
wavelet-LHH_glrlm_ShortRunLowGrayLevelEmphasis
wavelet-LLH_firstorder_Skewness
log-sigma-4-0-mm-3D_glrlm_ShortRunHighGrayLevelEmphasis
log-sigma-1-0-mm-3D_glszm_ZonePercentage
wavelet-LHL_glrlm_ShortRunLowGrayLevelEmphasis
wavelet-LHL_glszm_ZoneVariance
log-sigma-3-0-mm-3D_glcm_ClusterShade
original_shape_Elongation
wavelet-LHL_firstorder_Minimum
log-sigma-1-0-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis
−0.08 −0.06 −0.04 −0.02 0.00 0.02 0.04 0.06 0.08
C
Intra_log-sigma-5-0-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis
Intra_wavelet-LLH_gldm_SmallDependenceHighGrayLevelEmphasis
Intra_wavelet-LLH_firstorder_Skewness
Intra_log-sigma-2-0-mm-3D_gldm_LargeDependenceHighGrayLevelEmphasis
Peri_wavelet-HLH_firstorder_90Percentile
Intra_wavelet-HLL_firstorder_Range
Peri_exponential_firstorder_Median
Peri_log-sigma-1-0-mm-3D_glcm_InverseVariance
Peri_wavelet-HHL_firstorder_Mean
Peri_wavelet-LHH_glrlm_ShortRunLowGrayLevelEmphasis
Peri_log-sigma-2-0-mm-3D_gldm_SmallDependenceHighGrayLevelEmphasis
Intra_wavelet-LLH_glszm_GrayLevelNonUniformity
Peri_log-sigma-3-0-mm-3D_glcm_SumAverage
Intra_wavelet-LLH_glszm_SmallAreaEmphasis
Peri_log-sigma-5-0-mm-3D_firstorder_Kurtosis
Peri_original_shape_Elongation
Intra_wavelet-HHL_firstorder_Median
Peri_log-sigma-1-0-mm-3D_glszm_SmallAreaEmphasis
Peri_log-sigma-1-0-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis
Peri_wavelet-LHL_glszm_ZoneVariance
−0.08 −0.06 −0.04 −0.02 0.00 0.02 0.04 0.06 0.08

Supplementary Figure 2. The radiomics model characters of obtaining different regions. A: Tumor intra-tumoral region; B: External2 2-mm peri-tumoral region; C: Tumor+External2 multi-region.



Supplementary Figure 3. Diagnostic confusion matrix analysis for different models in differentiating benign from malignant parotid gland tumors. A: Radiologist; B: Clinical model; C: Deep-learning radiomics model; D: Combined model. Rows represent true labels, and columns represent predicted labels. Diagonal cells indicate correct classifications, and off-diagonal cells indicate misclassifications.The percentage value in parenthesis (beside the name of each model) indicates the total accuracy of the model.