European Radiology (2022) 32:6953–6964
https://doi.org/10.1007/s00330-022-08830-3

HEAD AND NECK

CT-based radiomics analysis of different machine learning models for differentiating benign and malignant parotid tumors

Yunlin Zheng[1] · Di Zhou[1] · Huan Liu[2] · Ming Wen[1]

Received: 18 January 2022 / Revised: 3 April 2022 / Accepted: 20 April 2022 / Published online: 29 April 2022

Abstract

Objectives This study aimed to explore and validate the value of different radiomics models for differentiating benign and malignant parotid tumors preoperatively.

Methods This study enrolled 388 patients with pathologically confirmed parotid tumors (training cohort: $n = 272$; test cohort: $n = 116$). Radiomics features were extracted from CT images of the non-enhanced, arterial, and venous phases. After dimensionality reduction and selection, radiomics models were constructed by logistic regression (LR), support vector machine (SVM), and random forest (RF). The best radiomic model was selected by using ROC curve analysis. Univariate and multivariable logistic regression was applied to analyze clinical-radiological characteristics and identify variables for developing a clinical model. A combined model was constructed by incorporating radiomics and clinical features. Model performances were assessed by ROC curve analysis, and decision curve analysis (DCA) was used to estimate the models' clinical values.

Results In total, 2874 radiomic features were extracted from CT images. Ten radiomics features were deemed valuable by dimensionality reduction and selection. Among radiomics models, the SVM model showed greater predictive efficiency and robustness, with AUCs of 0.844 in the training cohort; and 0.840 in the test cohort. Ultimate clinical features constructed a clinical model. The discriminatory capability of the combined model was the best (AUC, training cohort: 0.904; test cohort: 0.854). Combined model DCA revealed optimal clinical efficacy.

Conclusions The combined model incorporating radiomics and clinical features exhibited excellent ability to distinguish benign and malignant parotid tumors, which may provide a noninvasive and efficient method for clinical decision making.

Key Points

- *The current study is the first to compare the value of different radiomics models (LR, SVM, and RF) for preoperative differentiation of benign and malignant parotid tumors.*
- *A CT-based combined model, integrating clinical-radiological and radiomics features, is conducive to distinguishing benign and malignant parotid tumors, thereby improving diagnostic performance and aiding treatment.*

Keywords Computed tomography · Parotid neoplasms · Radiomics · Machine learning

Abbreviations

AIC	Akaike information criterion
ANOVA	Analysis of variance
AUC	Area under the curve
BCA	Basal cell adenoma
CIs	Confidence intervals
DCA	Decision curve analysis
DICOM	Digital Imaging and Communications in Medicine
FNAB	Fine-needle aspiration biopsy
GLCM	Gray-level co-occurrence matrix
GLRLM	Gray-level run-length matrix
GLSZM	Gray-level size zone matrix
IBSI	Image biomarker standardization initiative
ICC	Intraclass correlation coefficient
IST	Infiltration of surrounding tissues
LASSO	Least absolute shrinkage and selection operator
LR	Logistic regression

✉ Ming Wen
liuyucun65@163.com

[1] Department of Radiology, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, People's Republic of China

[2] GE Healthcare, Shanghai 201203, People's Republic of China

mRMR	Minimum redundancy maximum correlation
NPV	Negative prediction value
ORs	Odds ratios
PA	Pleomorphic adenoma
PACS	Picture archiving and communication systems
PPV	Positive prediction value
RF	Random forest
ROC	Receiver operating characteristic
ROI	Region of interest
SVM	Support vector machine
WT	Warthin tumor

Introduction

Parotid tumors are the most dominant salivary gland tumors, of which approximately 80% are benign, and 20% are malignant [1]. It has complex and diverse histopathological types. There are 24 histologic types of malignant parotid tumors, the most common being mucoepidermoid carcinoma, and there are 11 histologic types of benign parotid tumors, the most common being pleomorphic adenoma (PA), followed by Warthin tumor (WT) and basal cell adenoma (BCA) [2]. Although the primary treatment for parotid tumors is surgery, various histological types of parotid tumors determine significantly different treatment decisions and prognoses [3, 4]. Thus, accurate preoperative assessment of benign and malignant parotid tumors is essential for tailoring treatment decisions.

Parotid tumors rarely have apparent clinical symptoms, so the preoperative distinction between benign and malignant parotid tumors mainly relies on fine-needle aspiration biopsy (FNAB) and imaging examination. Although FNAB is the most common preoperative qualitative examination for parotid tumors [5], it carries the risk of tumor cell transplantation and induced mumps, and limited sampling sometimes makes the cytohistologic diagnosis inconclusive [6]. Currently, preoperative imaging for parotid tumors includes ultrasound, magnetic resonance imaging (MRI), and computed tomography (CT). Due to ultrasound examination being easily affected by the tumor location and adjacent bone, its diagnostic efficacy is limited, so CT and MRI examinations are mainly used nowadays [7, 8]. Although MRI demonstrates excellent resolution of soft tissues, multiphase CT provides valuable information on tumor characteristics and surrounding structures at a lower cost and is time-consuming than MRI [9, 10]. However, due to the overlapping radiological features of benign and malignant parotid tumors, imaging sometimes leads to uncertain results, depending mainly on the experience of physicians [11].

Radiomics, an emerging and prospective field, provides comprehensive quantification of tumor types through high-throughput extraction and mining a large number of image features [12]. With the development of algorithms and artificial intelligence, recent studies have described the application of radiomics analysis in parotid tumors [13–16]. These studies demonstrated that radiomics might help patients with benign or malignant parotid tumors solve clinical problems and optimize treatment. The reliability, high accuracy, and efficacy of predictive models are crucial factors promoting radiomics success. Hence, it is meaningful to compare different machine learning models of radiomics-based clinical biomarkers [17]. Nevertheless, to our knowledge, no studies have investigated the application of different radiomics-based models in differentiating benign and malignant parotid tumors.

In this study, we chose different machine learning approaches for radiomics models to develop and validate a CT-based combined model of radiomics and clinical features for distinguishing benign and malignant parotid tumors.

Materials and methods

Patient population

The institutional ethics review board approved this retrospective study of our institution, and the requirement of patient approval or written informed consent for reviewing medical records or images was waived. The data of 388 patients with parotid tumors who underwent parotid surgery in our hospital from January 2014 to September 2021 were included in the study. Inclusion criteria were as follows: (1) patients underwent preoperative CT plain scan and two-phase enhanced scan; (2) no history of FNAB, radiotherapy, and chemotherapy; and (3) PA, WT, BCA, and malignant parotid tumors were diagnosed by postoperative pathology and had complete clinical data. Exclusion criteria were as follows: (1) images with severe motion artifacts or evident noise; (2) maximum tumor diameter less than 1.0 cm; and (3) other tumor diseases. These patients were randomly separated into a training cohort and a test cohort at a 7:3 ratio (see Fig. 1 for details). The tumor distribution is described in Table 1.

Image acquisition

CT scans were performed using two 64-slice CT scanners (Somatom Sensation 64, Siemens Healthcare; Discovery 750, GE Healthcare) with the following parameters: 120 kV tube voltage; automatic tube current modulation (150–200 mAs); 64 × 0.625-mm detector collimation; matrix of 512 × 512; section thickness, 5 mm; section interval, 5 mm. The scanning area ranged from the skull base down to the thorax inlet. Non-enhanced CT scan was performed first, and enhanced images were obtained after intravenous injection of nonionic iodinated contrast medium (Ultravist 370, Bayer Schering Pharma) (1.5 mL/kg, 3.5 mL/s). Contrast-enhanced

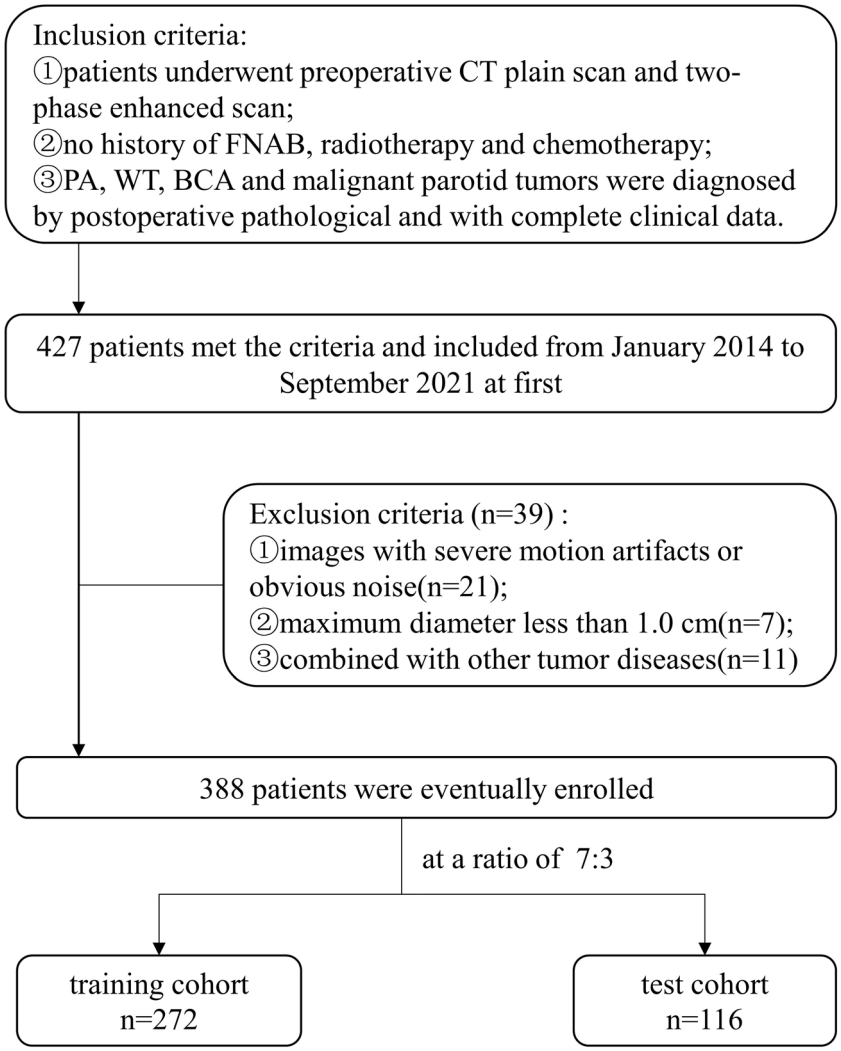

Fig. 1 Flowchart for selecting the study population

CT images were performed at 40 and 100 s after contrast material injection in the arterial phase and venous phase, respectively.

Radiological and clinical data analysis

Patient images and clinical data were obtained from our hospital's regular clinical records and picture archiving and communication systems (PACS). We retrospectively analyzed clinical parameters including age, disease duration, symptoms, sex, smoking/drinking status, hospitalization number, and postoperative pathological diagnosis. All CT images were assessed and agreed upon by two radiologists with 5 and 34 years of experience in head and neck CT, respectively, who were blinded to the histopathological and clinical data. The following parameters were analyzed: maximum diameter, tumor number, distribution, shape, margin, tumor location, density, calcification, cystic degeneration, enhanced peak phase, enhancement degree, enhanced uniformity, enlarged ipsilateral lymph nodes, and infiltration of surrounding tissues (IST). Some radiological feature definitions can be found in the Supplementary Material. When there were multiple lesions in the parotid gland, the largest lesion with confirmed pathology was selected for analysis.

Image segmentation

All images were saved as standard soft tissue settings (window width, 400 HU; window level, 40 HU) [18], which were stored in Digital Imaging and Communications in Medicine (DICOM) format. Blinded to patients' histopathological results, two radiologists (with 5 and 12 years of diagnostic experience) manually segmented the region of interest (ROI) by using ITK-SNAP software (version 3.8.0, http://www.itksnap.org/). The tumors were delineated along margins layer-by-layer on axial multiphase CT images, excluding adjacent normal tissue, calcification, and vessel. A complete schematic is presented in Fig. 2.

Table 1 Histologic analysis of parotid tumors

Histologic type	Patients (*n*)	Male/female	%
Benign tumors	310	180/130	79.9
Pleomorphic adenoma	144	58/86	37.1
Warthin tumor	109	100/9	28.1
Basal cell adenoma	57	22/35	14.7
Malignant tumors	78	42/36	20.1
Mucoepidermoid carcinoma	21	10/11	5.4
Acinic cell carcinoma	16	7/9	4.1
Lymphoma/lymphoepithelial carcinoma	9	3/6	2.3
Adenoid cystic carcinoma	7	4/3	1.8
Squamous cell carcinoma	6	6/0	1.5
Myoepithelial carcinoma	4	3/1	1.0
Adenocarcinoma	3	2/1	0.8
Salivary ductal carcinoma	3	3/0	0.8
Eosinophilic cell carcinoma	3	3/0	0.8
Secretory carcinoma	2	1/1	0.5
Basal cell carcinoma	1	0/1	-
Rhabdomyosarcoma	1	0/1	-
Carcinoma in pleomorphic adenoma	1	1/0	-
Undifferentiated carcinoma	1	0/1	-

Images and data were preprocessed by resampling and standardization to assure repeatability of the results. The intra- and interobserver reproducibility were assessed by the intraclass correlation coefficient (ICC); 116 patients (93 benign tumors and 23 malignant tumors) were chosen at random for ROI segmentation again by radiologist 1 after 1 month, and an ICC greater than 0.9 indicated good consistency.

Radiomics feature extraction/selection

For feature extraction, all images and ROIs were batched into AK software (Artificial Intelligence Kit, version 3.2.2, GE Healthcare). These radiomics features included the shape, histogram, Haralick, gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM),

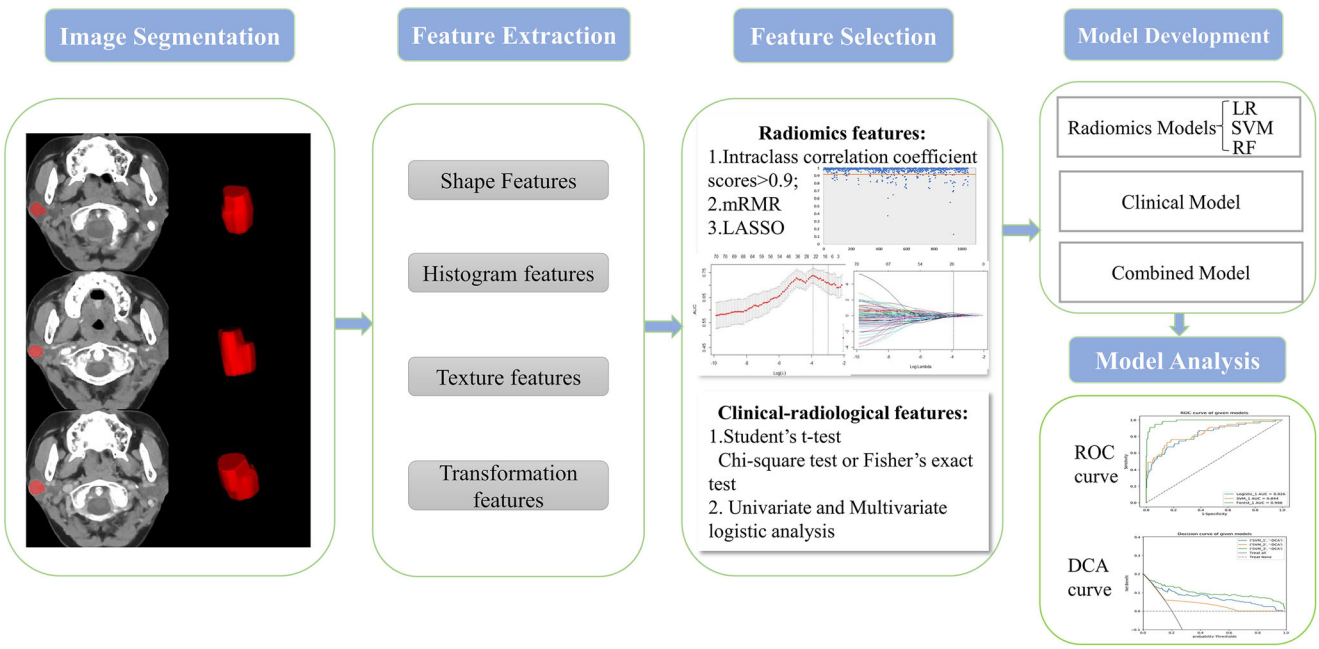

Fig. 2 Workflow of this study

and gray-level size zone matrix (GLSZM). These algorithms for obtaining radiomic features were referenced from the image biomarker standardization initiative (IBSI) [19]. The radiomic features dimensionality reduction and selection in the training cohort were as follows: first, analysis of variance (ANOVA) was performed on the extracted features to select statistically significant features with ICC scores > 0.9. Second, the minimum redundancy maximum correlation (mRMR) algorithm was applied to select the features with the highest relevance with tumor classification and least redundancy. Ultimately, the least absolute shrinkage and selection operator (LASSO) regression model with 10-fold cross-validation was used to select features with nonzero coefficients (Fig. 3). All feature selection procedures were executed on the training cohort and used for the test cohort.

Development of radiomics signature models

The final selected features were applied to construct radiomics models. To select a classifier model that has the greatest recognition of tumor data, our study chose three mainstream machine learning algorithm training models, including logistic regression (LR), support vector machine (SVM), and random forest (RF). The diagnostic performances of the three models were compared by the area under the curve (AUC) of the receiver operating characteristic curve (ROC), accuracy, sensitivity, specificity, positive prediction value (PPV), and negative prediction value (NPV). Then, the best radiomics model was screened.

Clinical and combined models' construction and validation

Univariate logistic regression analysis was performed on each predictor variable, including clinical and radiological characteristics in the training cohort, after which multivariable logistic regression analysis was used on selected statistically significant features to obtain ultimate predictor variables for model development. Odds ratios (ORs) with 95% confidence intervals (CIs) were calculated for each factor. As machine learning can provide highly accurate, objective, and reliable models to assist clinical decisions [17], we chose the best radiomics model to construct a combined model based on the selected predictors. The performances of models were evaluated and validated by the training and test cohorts, including the AUC, accuracy, sensitivity, specificity, PPV, and NPV. Decision curve analysis (DCA) quantified the net benefits with different threshold probabilities in training and test cohorts to evaluate the models' clinical efficiency in tumor classification.

Statistical analysis

SPSS software (version 26.0, IBM), R software (version 3.6.3; https://www.r-project.org), and Python software (version 3.5.6; http://www.python.org) were applied for statistical analysis. Student's *t*-test was used for continuous variables, expressed as $x \pm s$, and the chi-square test or Fisher's exact test was applied for categorical variables, presented as ratios. The forward stepwise selection method was used in multivariable logistic

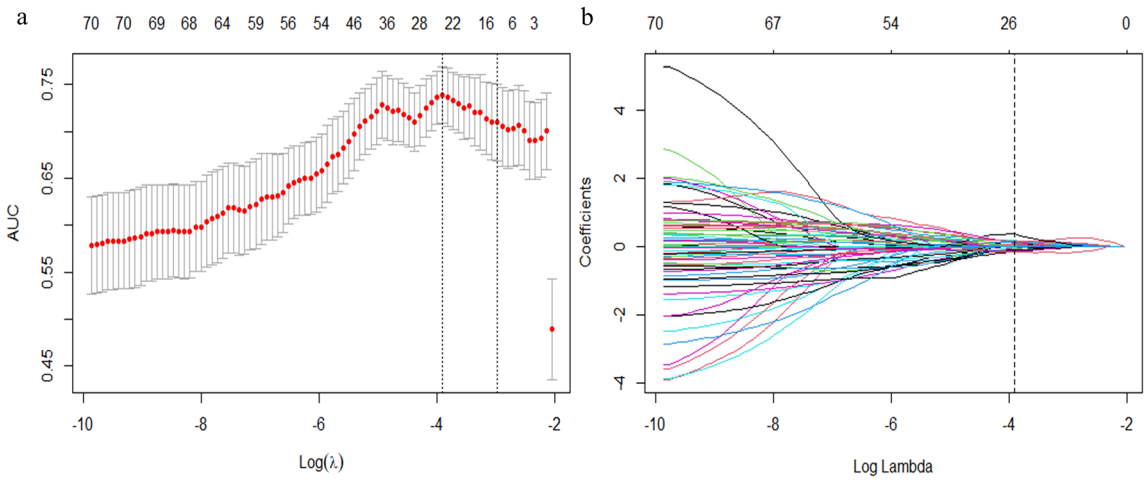

Fig. 3 Feature selection with the least absolute shrinkage and selection operator (LASSO) regression model. **a** The LASSO model's tuning parameter (λ) selection used 10-fold cross-validation via minimum criterion. The vertical lines indicate the optimal value of the LASSO tuning parameter (λ). **b** LASSO coefficient profile plot with different log (λ) was shown. The vertical dashed lines represent 10 radiomics features with nonzero coefficients selected with the optimal λ value

regression analysis based on the Akaike information criterion (AIC). A two-tailed p-value < 0.05 indicated statistical significance.

Results

Patients' population and radiological characteristics

A total of 388 patients (222 men, 166 women) were recruited for our study, including 272 patients in the training cohort and 116 patients in the test cohort. The results indicated that there were no significant differences between the training and test cohorts (see Table 1 in the Supplementary Materials for more information). The details of patients' clinical and radiological features are summarized in Table 2. In the training cohort, symptoms, shape, margin, tumor location, maximum diameter, lymph node status, and IST were significantly different between benign and malignant parotid tumors. There were significant differences in symptoms, margin, tumor location, and IST in the test cohort. Univariate and multivariable logistic regression analyses showed that symptom, shape, margin, and enlarged lymph nodes were independent predictors of malignant tumors (Table 3). A clinical model was constructed by these four features. Malignant tumors were more likely to have symptoms (OR, 8.335; 95% *CI*, 3.504 to 19.829), unclear margin (OR, 7.159; 95% *CI*, 3.455 to 14.833), enlarged ipsilateral lymph nodes (OR, 3.259; 95% *CI*, 1.545 to 6.874), and nonround shape (OR, 2.163; 95% *CI*, 1.176 to 3.980).

Radiomic signature models and performances

A total of 958 radiomic features were extracted from a single ROI; therefore, 2874 radiomic features were extracted from images of three scanning phases. Among these features, 10 features with nonzero coefficients were retained after feature downscaling and selection (Table 4). The ROC curves of three radiomics models (including LR, SVM, RF) in the training and test cohorts are shown in Fig. 4. In the training cohort, the best radiomic model was RF, with an AUC of 0.986, an accuracy of 0.908, a sensitivity of 0.909, a specificity of 0.972, PPV of 1.000, and NPV of 0.897. However, in the test cohort, the best radiomic model was SVM. The AUC, accuracy, sensitivity, specificity, PPV, and NPV of the SVM model were 0.840, 0.853, 0.696, 0.892, 0.800, and 0.858, respectively (see Table 5 for details). The RF model of the training and test cohorts showed a trend of overfitting. To ensure the stability and sustainability of radiomic model, we finally chose the SVM model as the best radiomic model.

Combined model construction and validation

By integrating clinical predictors (symptoms, shape, margin, and enlarged lymph nodes) with the 10 radiomics features, we developed a combined model. The clinical model, radiomic model, and combined model were constructed using SVM to uniformly and objectively evaluate the diagnostic efficacy of different models. The ROC curves of the three models in the training and test cohorts are presented in Fig. 5. In the training and test cohorts, the discrimination ability of the combined model was significantly better than that of the clinical and radiomic signature models (Table 5), with a higher AUC (0.904 in the training cohort; 0.854 in the test cohort). The DCAs of the three models are shown in Fig. 6. The results indicated that the combined model had the most significant net benefit in classifying parotid tumors in training and test cohorts.

Discussion

The present study suggested that CT-based radiomics models may effectually differentiate benign and malignant parotid tumors and improve diagnostic performance combined with clinical features. Thus, we developed a CT-based combined model that integrated clinical predictors and radiomics signatures using SVM for differentiating parotid tumors with satisfactory outcomes. It may provide a promising noninvasive method that is feasible and reliable for parotid tumor evaluation.

Our study demonstrates that benign parotid tumors were commonly observed in females, excluding WT, while the incidence of malignant tumors was similar between males and females, consistent with the results of Comoglu [20]. Consistent with our results, Inaka Y [21] reported that malignant tumors had a significantly higher incidence of several symptoms than benign tumors. Typically, there are nonspecific symptoms when parotid tumors are discovered. Hence, preoperative diagnosis of parotid tumors primarily relies on imaging examinations, especially CT and MRI [4, 5, 9]. This study showed that benign parotid tumors occurred mostly in superficial lobes, with clear margins, whereas malignant tumors occurred primarily in deep lobes or grew across the lobes, with unclear margins, which is similar to some studies [4, 10, 11]. The present study is similar to some studies [21, 22] in that lymph node enlargement is valuable in distinguishing benign from malignant parotid tumors. Although Kato [23] suggested that imaging features of the intratumoral cystic components could help distinguish benign from malignant parotid tumors, this is inconsistent with our findings and may be related to the predisposition of some benign parotid tumors to cystic degeneration. Liu

Table 2 Clinical and radiological characteristics in the training and test cohorts

Variable		Training cohort (n = 272)			Testing cohort (n = 116)		
		Benign (n = 217)	Malignant (n = 55)	p-value	Benign (n = 93)	Malignant (n = 23)	p-value
Age (years)		51.79 ± 14.31	50.02 ± 16.55	0.468	53.58 ± 15.95	53.22 ± 15.75	0.922
Duration (months)		26.67 ± 51.24	36.02 ± 103.53	0.343	32.12 ± 63.62	33.38 ± 74.36	0.935
Max-diameter (cm)		2.95 ± 1.26	3.43 ± 1.08	0.010*	2.90 ± 1.07	3.47 ± 1.45	0.090
Symptoms	With	9 (4.15%)	14 (25.45%)	< 0.001*	4 (4.30%)	5 (21.74%)	0.018*
	Without	208 (95.85%)	41 (74.55%)		89 (95.70%)	18 (78.26%)	
Sex	Female	99 (45.62%)	26 (47.27%)	0.826	31 (33.33%)	10 (43.48%)	0.362
	Male	118 (54.38%)	29 (52.73%)		62 (66.67%)	13 (56.52%)	
Smoking	Yes	96 (44.24%)	20 (36.36%)	0.291	48 (51.61%)	10 (43.48%)	0.485
	No	121 (55.76%)	35 (63.64%)		45 (48.39%)	13 (56.52%)	
Drinking	Yes	82 (37.79%)	16 (29.09%)	0.230	25 (26.88%)	7 (30.43%)	0.733
	No	135 (62.21%)	39 (70.91%)		68 (73.12%)	16 (69.57%)	
Number	Multiple	30 (13.82%)	10 (18.18%)	0.415	10 (10.75%)	1 (4.35%)	0.588
	Single	187 (86.18%)	45 (81.82%)		83 (89.25%)	22 (95.65%)	
Distribution	Left-sided	103 (47.47%)	30 (54.55%)	0.105	40 (43.01%)	12 (52.17%)	0.352
	Right-sided	98 (45.16%)	25 (45.45%)		46 (49.46%)	11 (47.83%)	
	Bilateral	16 (7.37%)	0		7 (7.53%)	0	
Shape	Round	155 (71.43%)	25 (45.45%)	< 0.001*	67 (72.04%)	14 (60.87%)	0.296
	Non-round	62 (28.57%)	30 (54.55%)		26 (27.96%)	9 (39.13%)	
Margin	Clear	187 (86.18%)	26 (47.27%)	< 0.001*	80 (86.02%)	12 (52.17%)	0.001*
	Unclear	30 (13.82%)	29 (52.73%)		13 (13.98%)	11 (47.83%)	
Location	Superficial	155 (71.43%)	26 (47.27%)	0.002*	67 (72.04%)	10 (43.48%)	0.029*
	Deep	34 (15.67%)	19 (34.55%)		18 (19.35%)	8 (34.78%)	
	Both	28 (12.90%)	10 (18.18%)		8 (8.60%)	5 (21.74%)	
Density	Homogeneous	183 (84.33%)	42 (76.36%)	0.163	78 (83.87%)	19 (82.61%)	0.884
	Heterogeneous	34 (15.67%)	13 (23.64%)		15 (16.13%)	4 (17.39%)	
Calcification	With	9 (4.15%)	5 (9.09%)	0.254	2 (2.15%)	0	1.000
	Without	208 (95.85%)	50 (90.91%)		91 (97.85%)	23 (100.00%)	
Cystic areas	NO	171 (78.80%)	41 (74.55%)	0.572	75 (80.65%)	18 (78.26%)	0.740
	Marginal	40 (18.43%)	11 (20.00%)		13 (13.98%)	3 (13.04%)	
	Central	6 (2.76%)	3 (5.45%)		5 (5.38%)	2 (8.70%)	
Enhanced peak phase	Arterial	86 (39.63%)	20 (36.36%)	0.657	40 (43.01%)	9 (39.13%)	0.736
	Venous	131 (60.37%)	35 (63.64%)		53 (56.99%)	14 (60.87%)	
Enhancement degree	Slight	31 (14.29%)	5 (9.09%)	0.609	10 (10.75%)	2 (8.70%)	0.482
	Moderate	63 (29.03%)	20 (36.36%)		31 (33.33%)	5 (21.74%)	
	Obvious	123 (56.68%)	30 (54.55%)		52 (55.91%)	16 (69.57%)	
Enhanced uniformity	Yes	129 (59.45%)	25 (45.45%)	0.061	56 (60.22%)	12 (52.17%)	0.483
	No	88 (40.55%)	30 (54.55%)		37 (39.78%)	11 (47.83%)	
Enlarged lymph nodes	With	18 (8.29%)	18 (32.73%)	< 0.001*	7 (7.53%)	4 (17.39%)	0.294
	Without	199 (91.71%)	37 (67.27%)		86 (92.47%)	19 (82.61%)	
IST	With	19 (8.76%)	14 (25.45%)	0.001*	7 (7.53%)	6 (26.09%)	0.031*
	Without	198 (91.24)	41 (74.55%)		86 (92.47%)	17 (73.91%)	

*Represents $p < 0.05$. Numerical data are presented as mean ± standard deviation. Categorical data as numbers (n%); *IST* infiltration of surrounding tissues

Table 3 Univariate and multivariable logistic regression analyses for selecting clinical features of model development

Variable	Univariate analysis		Multivariate analysis	
	OR (95% *CI*)	*p*-value	OR (95% *CI*)	*p*-value
Age (years)	0.994 (0.978, 1.010)	0.475		
Duration (months)	1.001 (0.998, 1.005)	0.406		
Max-diameter (cm)	1.370 (1.118, 1.678)	0.020*	1.034 (0.813, 1.316)	0.782
Symptom	7.357 (3.445, 15.712)	< 0.001*	8.335 (3.504, 19.829)	< 0.001*
Sex	0.843 (0.512, 1.388)	0.501		
Smoking	0.720 (0.434, 1.197)	0.206		
Drinking	0.793 (0.462, 1.362)	0.401		
Number	0.877 (0.426, 1.802)	0.720		
Distribution	0.673 (0.437, 1.036)	0.072		
Shape	2.523 (1.518, 4.192)	< 0.001*	2.163 (1.176, 3.980)	0.013*
Margin	6.199 (3.594, 10.690)	< 0.001*	7.159 (3.455, 14.833)	< 0.001*
Location	1.788 (1.299, 2.462)	< 0.001*	1.329 (0.885, 1.995)	0.171
Density	1.484 (0.800, 2.754)	0.210		
Calcification	1.862 (0.627, 5.524)	0.263		
Cystic areas	1.254 (0.801, 1.962)	0.322		
Enhanced peak phase	1.157 (0.693, 1.931)	0.577		
Enhancement degree	1.156 (0.806, 1.658)	0.431		
Enhanced uniformity	1.640 (0.996, 2.701)	0.052		
Enlarged lymph nodes	4.479 (2.360, 8.498)	< 0.001*	3.259 (1.545, 6.874)	0.002*
IST	3.021 (1.570, 5.813)	0.001*	0.490 (0.198, 1.210)	0.122

*Represents $p < 0.05$. OR odds ratio, *CI* confidence interval, *IST* infiltration of surrounding tissues

[8] and Liu [24] separately reported no significant differences in diagnostic efficacy and radiomics features for identifying parotid tumors between MRI and CT. The subjective influence of personal clinical experience and a single evaluation index are main limitations of these studies. Currently, a consensus on imaging features to classify benign and malignant parotid tumors has not been established. Consequently, a new noninvasive method is desired to differentiate parotid tumors.

Radiomics is a noninvasive method, which constructs models based on the original images by intelligent calculation to obtain more information and can reflect potentially relevant phenotypic information based on tumor heterogeneity, providing valuable diagnosis, prognosis, and

Table 4 Radiomics features' selection results

CT scanning phase	ID	Radiomics features' name
Non-enhanced phase	1	wavelet.HLH_glcm_ClusterShade
	2	log.sigma.3.0.mm.3D_glszm_SizeZoneNonUniformityNormalized
	3	log.sigma.1.0.mm.3D_glrlm_ RunLengthNonUniformityNormalized
	4	log.sigma.1.0.mm.3D_firstorder_Skewness
Arterial phase	1	wavelet.HHH_glszm_ZoneVariance
	2	original_shape_Sphericity
	3	wavelet.HHL_glcm_Imc1
Venous phase	1	original_glrlm_LongRunLowGrayLevelEmphasis
	2	wavelet.HLL_glcm_Imc1
	3	wavelet.LHH_glcm_JointEntropy

glcm gray-level co-occurrence matrix, *glszm* gray-level size zone matrix, *glrlm* gray-level run-length matrix

Table 5 Diagnostic performance of different models for predicting parotid tumors in training and test cohorts

Model		AUC (95% *CI*)	Accuracy	Sensitivity	Specificity	PPV	NPV
LR	Training	0.826 (0.770, 0.880)	0.809	0.673	0.848	0.522	0.906
	Test	0.798 (0.698, 0.891)	0.810	0.696	0.871	0.516	0.918
SVM*	Training	0.844 (0.788, 0.896)	0.857	0.764	0.816	1.000	0.848
	Test	0.840 (0.761, 0.911)	0.853	0.696	0.892	0.800	0.858
RF	Training	0.986 (0.976, 0.995)	0.908	0.909	0.972	1.000	0.897
	Test	0.779 (0.680, 0.874)	0.845	0.696	0.785	0.727	0.857
Clinical*	Training	0.797 (0.739, 0.851)	0.838	0.818	0.677	0.867	0.837
	Test	0.753 (0.657, 0.843)	0.828	0.739	0.709	1.000	0.823
Combined*	Training	0.904 (0.863, 0.942)	0.890	0.763	0.912	0.931	0.885
	Test	0.854 (0.770, 0.931)	0.871	0.696	0.935	0.750	0.890

*Represents models were constructed using SVM

LR logistic regression, *SVM* support vector machine, *RF* random forest, *AUC* area under the curve, *PPV* positive prediction value, *NPV* negative prediction value

individualized treatment [12]. Our study suggested that texture features, consisting mainly of GLCM features, differed significantly in distinguishing malignant from benign parotid tumors, similar to the results of Xu [25] and Zhang [16]. It can be speculated that malignant tumors grow quickly and easily cause necrosis and microhemorrhage, making their heterogeneity higher than that of benign tumors.

This study indicated that the SVM and RF models were more effective than the LR model. It is speculated that parotid tumor data may be nonlinear or linearly inseparable, so the LR model using a linear algorithm is less efficient than nonlinear RF and SVM. RF is an algorithm that integrates multiple decision trees through ensemble learning, which is excellent in accuracy among current algorithms and can attribute selection while predicting classification [26, 27]. However, it has poor generalization ability when processing small sample cohorts and may overfit. SVM, with satisfactory stability and efficiency, achieved almost the same performance as a large number of training samples with fewer training samples [28, 29]. Our combined model exhibited the best diagnostic performance (AUC, training cohort: 0.904; test cohort: 0.854), followed by the SVM, RF, and LR models alone and then the clinical-radiological features. It suggested that although tumor radiomics models have more excellent predictive capability than clinical features, clinical-radiological information is also significant. Only by

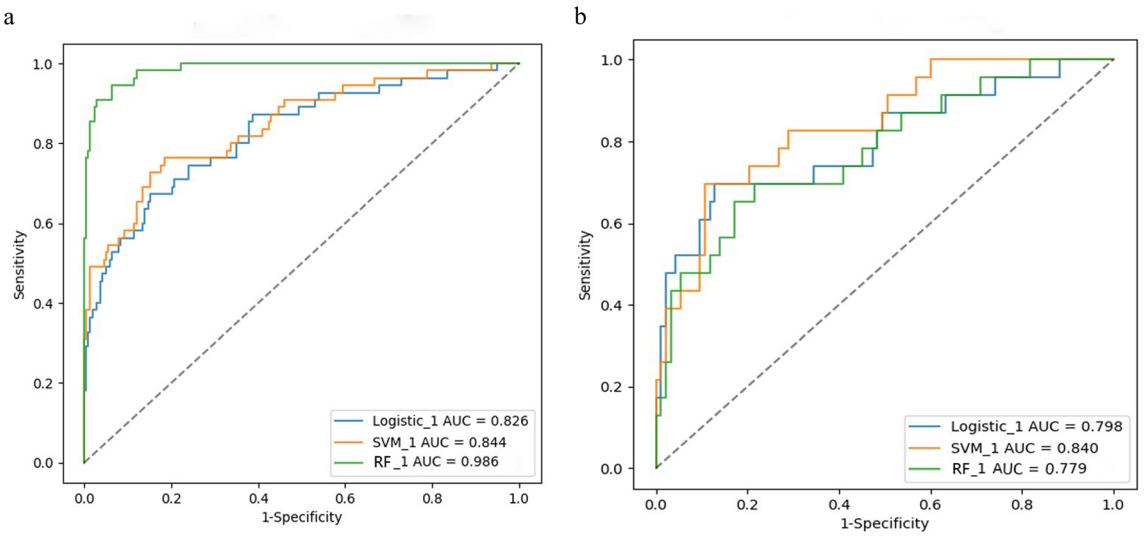

Fig. 4 The receiver operating characteristic (ROC) curves of the SVM, RF, and Logistic in the training (**a**) and validation (**b**) cohorts, respectively. Logistic, logistic regression; SVM, support vector machine; RF, random forest

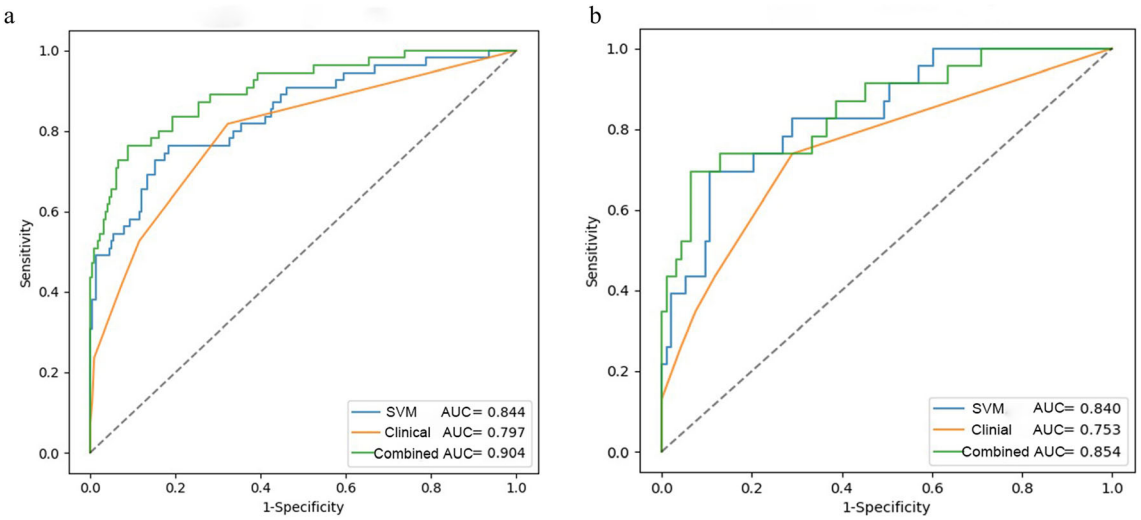

Fig. 5 Receiver operating characteristic (ROC) curves for the SVM model, clinical model, and combined model when predicting malignancy in training cohort (**a**) and test cohort (**b**)

incorporating these features could the model accurately estimate parotid tumors.

This study had several limitations. First, this was a retrospective, single-center study, which may cause potential selection bias. Second, some controversies were associated with subjectivity in defining the boundaries of manual segmentation [30]. We hope to achieve complete automation by deep learning in the future. Third, we adopted CT-enhanced phases of only arterial and venous in this study, while a 5- or 8-min delayed-phase proved valuable [9, 31]. Further prospective studies will investigate and validate whether delayed-phase radiomics could refine model performance.

Conclusion

In summary, our study proposed and verified a CT-based combined model integrating clinical-radiological and radiomics features to differentiate benign and malignant parotid tumors. In our study, the combined model using SVM exhibited the best diagnostic performance and may

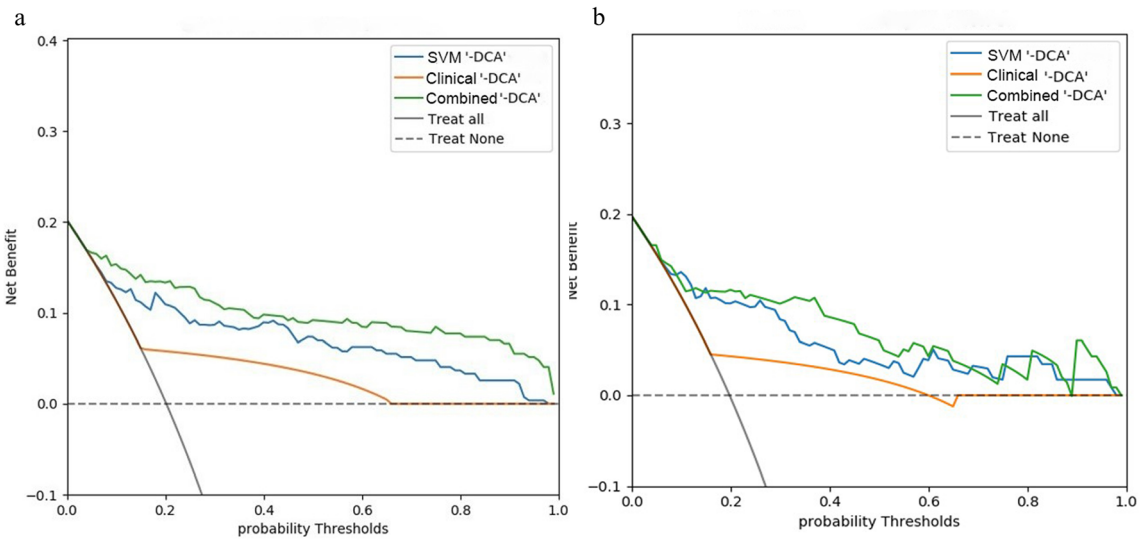

Fig. 6 Decision curve analysis (DCA) for three models in classifying parotid gland tumors in training (**a**) and test (**b**) cohorts, respectively. The graphs show that the combined model has the greatest net benefit for both datasets

serve as a potential approach in precision medicine and improve clinical therapeutic strategies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00330-022-08830-3.

Acknowledgements We thank the American Journal Experts (AJE) for their assistance with language editing.

Funding The authors state that this work has not received any funding.

Declarations

Guarantor The scientific guarantor of this publication is Ming Wen.

Conflict of interest The authors declare that they have no conflict of interest.

Statistics and biometry One of the authors (Huan Liu) has significant statistical expertise and is identified as the statistical guarantor for the statistical analysis used in this study.

Informed consent This study was approved by the institutional review board.

Ethical approval Institutional Review Board approval was obtained.

Methodology
- retrospective
- diagnostic or prognostic study
- performed at one institution

References

1. Gandolfi MM, Slattery W 3rd (2016) Parotid gland tumors and the facial nerve. Otolaryngol Clin North Am 49:425–434
2. Choi SY, Lee E, Kim E et al (2021) Clinical outcomes of bulky parotid gland cancers: need for self-examination and screening program for early diagnosis of parotid tumors. BMC Cancer 21:178
3. Moore MG, Yueh B, Lin DT, Bradford CR, Smith RV, Khariwala SS (2021) Controversies in the workup and surgical management of parotid neoplasms. Otolaryngol Head Neck Surg 164:27–36
4. Lewis AG, Tong T, Maghami E (2016) Diagnosis and management of malignant salivary gland tumors of the parotid gland. Otolaryngol Clin North Am 49:343–380
5. Kato H, Kanematsu M, Watanabe H et al (2015) Perfusion imaging of parotid gland tumours: usefulness of arterial spin labeling for differentiating Warthin's tumours. Eur Radiol 25:3247–3254
6. Espinoza S, Felter A, Malinvaud D et al (2016) Warthin's tumor of parotid gland: surgery or follow-up? Diagnostic value of a decisional algorithm with functional MRI. Diagn Interv Imaging 97:37–43
7. Prasad RS (2016) Parotid gland imaging. Otolaryngol Clin North Am 49:285–312
8. Liu Y, Li J, Tan YR, Xiong P, Zhong LP (2015) Accuracy of diagnosis of salivary gland tumors with the use of ultrasonography, computed tomography, and magnetic resonance imaging: a meta-analysis. Oral Surg Oral Med Oral Pathol Oral Radiol 119:238–245.e2
9. Jin GQ, Su DK, Xie D, Zhao W, Liu LD, Zhu XN (2011) Distinguishing benign from malignant parotid gland tumours: low-dose multi-phasic CT protocol with 5-minute delay. Eur Radiol 21:1692–1698
10. Lam PD, Kuribayashi A, Imaizumi A et al (2015) Differentiating benign and malignant salivary gland tumours: diagnostic criteria and the accuracy of dynamic contrast-enhanced MRI with high temporal resolution. Br J Radiol 88:20140685
11. Vogl TJ, Albrecht MH, Nour-Eldin NA et al (2018) Assessment of salivary gland tumors using MRI and CT: impact of experience on diagnostic accuracy. Radiol Med. 123:105–116
12. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. Radiology 278:563–577
13. Zheng YM, Li J, Liu S et al (2021) MRI-based radiomics nomogram for differentiation of benign and malignant lesions of the parotid gland. Eur Radiol 31:4042–4052
14. Al Ajmi E, Forghani B, Reinhold C, Bayat M, Forghani R (2018) Spectral multi-energy CT texture analysis with machine learning for tissue classification: an investigation using classification of benign parotid tumours as a testing paradigm. Eur Radiol 28:2604–2611
15. Li Q, Jiang T, Zhang C et al (2021) A nomogram based on clinical information, conventional ultrasound and radiomics improves prediction of malignant parotid gland lesions. Cancer Lett 527:107–114
16. Zhang D, Li X, Lv L et al (2020) A preliminary study of CT texture analysis for characterizing epithelial tumors of the parotid gland. Cancer Manag Res 12:2665–2674
17. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL (2015) Machine learning methods for quantitative radiomic biomarkers. Sci Rep 5:13087
18. Wu W, Ye J, Wang Q, Luo J, Xu S (2019) CT-based radiomics signature for the preoperative discrimination between head and neck squamous cell carcinoma grades. Front Oncol 9:821
19. Fornacon-Wood I, Mistry H, Ackermann CJ et al (2020) Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform. Eur Radiol 30:6241–6250
20. Comoglu S, Ozturk E, Celik M et al (2018) Comprehensive analysis of parotid mass: a retrospective study of 369 cases. Auris Nasus Larynx 45:320–327
21. Inaka Y, Kawata R, Haginomori SI et al (2021) Symptoms and signs of parotid tumors and their value for diagnosis and prognosis: a 20-year review at a single institution. Int J Clin Oncol 26:1170–1178
22. Jinnin T, Kawata R, Higashino M, Nishikawa S, Terada T, Haginomori SI (2019) Patterns of lymph node metastasis and the management of neck dissection for parotid carcinomas: a single-institute experience. Int J Clin Oncol 24:624–631
23. Kato H, Kanematsu M, Watanabe H, Mizuta K, Aoki M (2014) Salivary gland tumors of the parotid gland: CT and MR imaging findings with emphasis on intratumoral cystic components. Neuroradiology 56:789–795
24. Liu Y, Zheng J, Lu X et al (2021) Radiomics-based comparison of MRI and CT for differentiating pleomorphic adenomas and Warthin tumors of the parotid gland: a retrospective study. Oral Surg Oral Med Oral Pathol Oral Radiol 131:591–599
25. Xu Y, Shu Z, Song G et al (2021) The role of preoperative computed tomography radiomics in distinguishing benign and malignant tumors of the parotid gland. Front Oncol 11:634452
26. Han L, Yuan Y, Zheng S et al (2014) The Pan-Cancer analysis of pseudogene expression reveals biologically and clinically relevant tumour subtypes. Nat Commun 5:3963
27. Sarica A, Cerasa A, Quattrone A (2017) Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A Systematic Review. Front Aging Neurosci 9:329
28. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W (2018) Applications of support vector machine (SVM) learning in cancer genomics. Cancer Genomics Proteomics 15:41–51

29. Sanz H, Valim C, Vegas E, Oller JM, Reverter F (2018) SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. BMC Bioinformatics 19:432
30. Lohmann P, Bousabarah K, Hoevels M, Treuer H (2020) Radiomics in radiation oncology-basics, methods, and limitations. Strahlenther Onkol 196:848–855
31. Reginelli A, Clemente A, Renzulli M et al (2019) Delayed enhancement in differential diagnosis of salivary gland neoplasm. Gland Surg 8:S130–S135