BJR https://doi.org/10.1259/bjr.20210340

Received: 15 March 2021 | Revised: 29 August 2021 | Accepted: 08 September 2021

Cite this article as:
Vernuccio F, Arnone F, Cannella R, Verro B, Comelli A, Agnello F, et al. Diagnostic performance of qualitative and radiomics approach to parotid gland tumors: which is the added benefit of texture analysis?. *Br J Radiol* 2021; **94**: 20210340.

FULL PAPER

Diagnostic performance of qualitative and radiomics approach to parotid gland tumors: which is the added benefit of texture analysis?

[1]FEDERICA VERNUCCIO, [1]FEDERICA ARNONE, [1,2]ROBERTO CANNELLA, [1]BARBARA VERRO, [3,4]ALBERT COMELLI, [1]FRANCESCO AGNELLO, [4]ALESSANDRO STEFANO, [1]ROSALIA GARGANO, [2]VITO RODOLICO, [1]GIUSEPPE SALVAGGIO, [1]ROBERTO LAGALLA, [1]MASSIMO MIDIRI and [1]ANTONIO LO CASTO

[1]Department of Biomedicine, Neuroscience and Advanced Diagnostics (BIND), University Hospital of Palermo, Palermo, Italy
[2]Department of Health Promotion, Mother and Child Care, Internal Medicine and Medical Specialties (PROMISE), University of Palermo, Palermo, Italy
[3]Ri.MED Foundation, Palermo, Italy
[4]Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), Cefalù, Italy

Address correspondence to: Dr Federica Vernuccio
E-mail: *federicavernuccio@gmail.com*

Objective: To investigate whether MRI-based texture analysis improves diagnostic performance for the diagnosis of parotid gland tumors compared to conventional radiological approach.
Methods: Patients with parotid gland tumors who underwent salivary glands MRI between 2008 and 2019 were retrospectively selected. MRI analysis included a qualitative assessment by two radiologists (one of which subspecialized on head and neck imaging), and texture analysis on various sequences. Diagnostic performances including sensitivity, specificity, and area under the receiver operating characteristic curve (AUROC) of qualitative features, radiologists' diagnosis, and radiomic models were evaluated.
Results: Final study cohort included 57 patients with 74 tumors (27 pleomorphic adenomas, 40 Warthin tumors, 8 malignant tumors). Sensitivity, specificity, and AUROC for the diagnosis of malignancy were 75%, 97% and 0.860 for non-subspecialized radiologist, 100%, 94% and 0.970 for subspecialized radiologist and 57.2%, 93.4%, and 0.927 using a MRI radiomics model obtained combining texture analysis on various MRI sequences. Sensitivity, specificity, and AUROC for the differential diagnosis between pleomorphic adenoma and Warthin tumors were 81.5%, 70%, and 0.757 for non-subspecialized radiologist, 81.5%, 95% and 0.882 for subspecialized radiologist and 70.8%, 82.5%, and 0.808 using a MRI radiomics model based on texture analysis of T_2 weighted sequence. A combined radiomics model obtained with all MRI sequences yielded a sensitivity of 91.5% for the diagnosis of pleomorphic adenoma.
Conclusion: MRI qualitative radiologist assessment outperforms radiomic analysis for the diagnosis of malignancy. MRI predictive radiomics models improves the diagnostic performance of non-subspecialized radiologist for the differential diagnosis between pleomorphic adenoma and Warthin tumor, achieving similar performance to the subspecialized radiologist.
Advances in knowledge: Radiologists outperform radiomic analysis for the diagnosis of malignant parotid gland tumors, with some MRI qualitative features such as ill-defined margins, perineural spread, invasion of adjacent structures and enlarged lymph nodes being highly specific for malignancy. A radiomic model based on texture analysis of T_2 weighted images yields higher specificity for the diagnosis of pleomorphic adenoma compared to a radiologist non-subspecialized in head and neck radiology, thus minimizing false-positive pleomorphic adenoma diagnosis rate and reducing unnecessary surgical complications.

INTRODUCTION

Salivary gland tumors represent 2–6.5% of all head and neck tumors,[1,2] and 70% of them are located in the parotid gland.[3] Pleomorphic adenomas account for 65% of parotid gland tumors, followed by Warthin tumor in 15–20% and malignant lesions in about 10% of the cases.[1–3] Parotid fine needle aspiration is currently considered the gold-standard for diagnosis prior to surgical management but it is limited by low sensitivity and poor levels of diagnostic accuracy, with

false-positive rates of pleomorphic adenomas and Warthin tumor of 9 and 8%, respectively.[4,5] The differential diagnosis between these three entities is crucial because of the different management. For instance, pleomorphic adenomas exhibit a higher recurrence rate after surgical removal than Warthin tumors and have a higher incidence of malignant transformation.[1–8] Therefore, the recommended treatment for Warthin tumors is enucleation, as it reduces the risk of facial nerve damage,[2] while superficial or total parotidectomy with preservation of mandibular branch of the facial nerve, the hypoglossal nerve, and the lingual nerve is recommended for pleomorphic adenomas. In case of malignancy, complete curative surgical resection of primary tumor, with adequate free margins is the mainstay, and post-operative radiotherapy is considered in patients with high risk factors.

MRI demonstrates a specificity of 86–100% for the diagnosis of malignancy,[9–12] while specificity of MRI is limited to up to 86% for the diagnosis of pleomorphic adenomas.[13,14] Prior studies showed the potential role of some MRI features – *i.e.* multiplicity, margins, ADC values, T-peak enhancement, diffusion kurtosis imaging – in improving the diagnostic performance for the diagnosis of pleomorphic adenomas, but the results are still limited.[15–19] As radiomics mathematically quantifies tumor heterogeneity uncovering tumor texture features that cannot be fully appreciated by qualitative analysis of MRI images, we hypothesize that textural analysis of different signal intensities from parotid gland tumors may improve the diagnostic performance for their differential diagnosis, resulting in improved lesion management. To our knowledge, only one prior study,[20] investigated the use of MRI texture analysis for the diagnosis of primary tumors of salivary glands and identified some texture features that are significantly different between the three entities, but providing only limited results for the differentiation between pleomorphic adenomas and Warthin tumor (*i.e.* correctly classified data vectors: 50–82%), and lacking a comparison with the diagnostic performance of radiologists, which still represent the standard-of-care.

The purpose of our study was to investigate whether MRI-based texture analysis improves diagnostic performance for the diagnosis of malignant parotid tumors and for the differential diagnosis between pleomorphic adenoma and Warthin tumor compared to the conventional radiological approach.

METHODS AND MATERIALS

The Ethical Committee of University Hospital "Paolo Giaccone" approved this retrospective study and waived the requirement to obtain patient approval or written informed consent for the review of medical records or images.

Study cohort

The departmental electronic database of our tertiary referral academic center was searched for consecutive patients with parotid gland tumors submitted to parotid glands MRI between November 2008 and April 2019. Patients were eligible for inclusion if they had: (a) pathologically confirmed parotid gland tumor and (b) underwent contrast-enhanced parotid glands MRI before pathology diagnosis.

This search yielded a target population of 94 consecutive patients. Subjects were considered ineligible for this study in case of: (a) lack of full-protocol preoperative MRI (n = 22); (b) pathology diagnosis other than Warthin tumor, pleomorphic adenoma or primary malignancy (*e.g.* cyst, metastasis, lipoma, inflammation, uncertain pathological diagnosis) (n = 15).

Our final study population (Table 1) consisted of 57 patients (mean age ±SD, 56 ± 18 years; range, 10–88 years), including 24 men (mean age ±SD, 55 ± 21 years; range, 10–88 years) and 33 women (mean age ±SD, 56 ± 16 years; range, 25–80 years). In patients with multiple tumors a maximum of three pathologically proven lesions were selected by the study coordinator to avoid clustering bias. Tumors smaller than 5 mm were excluded. A total of 75 lesions (median maximum diameter, 1.7 cm; interquartile range, 1.4–2.2 cm, range, 0.6–5.5 cm) were analyzed.

MRI technique

All MRI examinations were performed using the same clinical 1.5 T MRI system (Achieva, Philips, Amsterdam, Netherlands). The standard MRI protocol included axial and coronal turbo spin-echo T1 and T2 sequences, three-dimensional T2 sequence, and four sequences after contrast injection in the following order: a three-dimensional T1 ultrafast spoiled gradient echo sequence, axial spectral presaturation with inversion recovery (SPIR) sequence, and axial and coronal turbo spin echo T1 sequence as shown in Supplementary Table 1. All patients received a weight-based intravenous dose of 0.1 ml/Kg mmol/kg of gadobenate dimeglumine or gadoteridol (MultiHance or Prohance, respectively, Bracco Imaging Italia s.r.l., Milan, Italy).

MRI qualitative analysis

Two board-certified radiologists (not subspecialized in head and neck radiology) evaluated the included lesions on an independent workstation using the Horos™ software (<https://www.horosproject.org/>) by consensus, and blinded to the purpose of the study and to the pathology information.

The qualitative analysis of the tumor included morphological and signal intensity features of the tumors – as shown in Supplementary Material 1 – and was similar to a prior study by Christe et al.[21] Furthermore, the readers evaluated the presence of lesion capsule, hemorrhage, cystic degeneration, restricted diffusion, as well as features of adjacent organs invasion including perineural spread, invasion of adjacent structures, and enlarged lymph nodes.[21–23]

Finally, two others independent readers, the first one with 7 years of experience in radiology, but not subspecialized in head and neck radiology, the second with European diploma in radiology on head and neck radiology and 30 years of experience, not involved in the assessment of the qualitative and quantitative features of the tumors, blinded to the purpose of the study and to the pathology information, were asked to provide their

Table 1. Characteristics of the included patients

Characteristics	Number (%)
Patients	57
Sex	
Males	24 (42.1)
Females	33 (57.9)
Age (years)	56 ± 18
Tumors	75
Histopathological diagnosis	
Pleomorphic adenoma	27 (36.0)
Warthin tumor	40 (53.3)
Malignant tumors	8 (10.7)
Carcinoma	6 (8.0)
Squamous cell carcinoma NOS $n = 3$	
Epithelial-myoepithelial carcinoma $n = 1$	
Mucoepidermoid carcinoma $n = 1$	
Carcinoma ex pleomorphic adenoma $n = 1$	
Diffuse large B-cell lymphoma (Mature B-cell neoplasms: Marginal zone lymphoma (sec WHO 2017 Classification of tumors of haematopoietic and lymphoid tissues)	1 (1.3)
Metastasis	1 (1.3)

NOS, Not otherwise specified.

Footnotes: Age is expressed ad mean ± standard deviation, size is expressed as median and interquartile range (25th to 75th percentile), after testing for normal distribution. Categorical variables are expressed as numbers and percentages in parenthesis.

diagnostic suspicion among pleomorphic adenoma, Warthin tumor or malignancy.

MRI texture analysis

A radiology fellow (trained in head and neck radiology), blinded to the pathology result and not involved in prior imaging interpretation, performed tumor segmentation and texture feature extraction. Segmentation was performed in four MRI data sets in axial images, on pre- and post-contrast axial turbo spin echo T1 sequences, SPIR sequences, and T_2 weighted turbo spin echo sequences. Due to the retrospective study design, some sequences were not available in few patients, including postcontrast T1 sequence in 14 patients with 18 lesions and SPIR sequence in 4 patients with 7 lesions.

Tumor segmentation was performed using a freely available texture analysis software (MaZda 4.6, Institute of Electronics, Technical University of Lodz, Lodz, Poland)[24] using both a manual region of interest (ROI) contouring the lesion and covering the largest possible lesion area of the tumor and standardized ROI (*i.e.* fixed morphology, size and area) (Figure 1). Each ROI was performed on two different contiguous slices in which the tumor had the largest diameter in each of the four sequences (*i.e.* pre- and post-contrast axial turbo spin echo T1 sequences, SPIR sequences, and T_2 weighted turbo spin echo sequences).

Figure 1. Manual and standardized regions of interest drawn on each sequence for different parotid gland tumors. Of note, SPIR sequence is T_1 weighted and acquired post-injection. SPIR, spectral presaturation with inversion recovery.

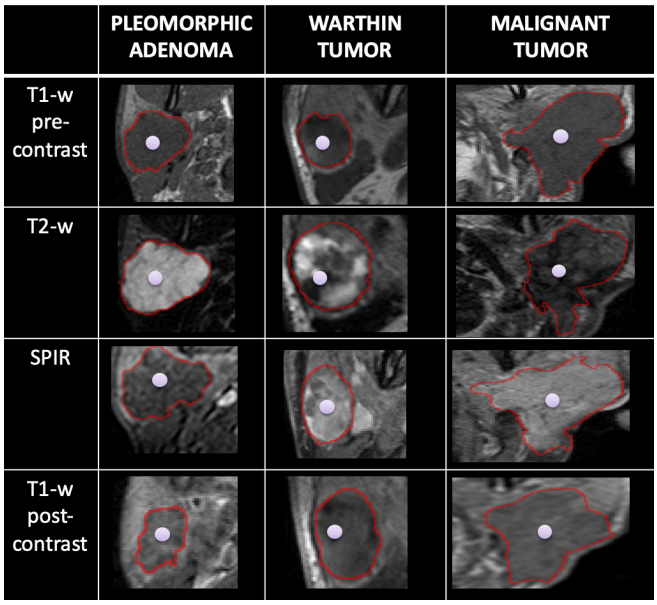


Table 2. Imaging characteristics between different salivary gland tumors

	Pleomorphic adenoma (n = 27)	Warthin tumor (n = 40)	Malignant tumors (n = 8)	*p* values Overall	*p* values Benign *vs* malignant	*p* values Pleomorphic *vs* Warthin
Size (cm)	1.8 (1.4, 2.0)	1.6 (1.3, 1,9)	3.2 (1.9, 4.6)	**0.015**	**0.018**	0.079
Side:				0.673	0.722	0.451
Left	14 (51.9)	17 (42.5)	3 (37.5)			
Right	13 (48.1)	23 (57.5)	5 (62.5)			
Multiple	5 (18.5)	23 (57.5)	2 (25.0)	**0.004**	0.464	**0.002**
Bilateral	2 (7.4)	15 (37.5)	0 (0)	**0.004**	0.186	**0.005**
Location:				0.647	0.638	0.524
Superficial	21 (77.8)	34 (85.0)	6 (75.0)			
Deep	6 (22.2)	6 (15.0)	2 (25.0)			
Signal intensity:				0.054	0.101	0.142
Homogeneous	22 (81.5)	26 (65.0)	3 (37.5)			
Heterogeneous	5 (18.5)	14 (35.0)	5 (62.5)			
T_2 weighted signal:				0.255	1.000	0.267
Hypointense	27 (100)	37 (92.5)	8 (100)			
Isointense	0 (0)	3 (7.5)	0 (0)			
Hyperintense	0 (0)	0 (0)	0 (0)			
T_2 weighted signal:				**<0.001**	**0.009**	**<0.001**
Markedly hyperintense	21 (77.8)	7 (17.5)	0 (0)			
Slightly hyperintense	6 (22.2)	11 (27.5)	1 (12.5)			
Isohypointense	0 (0)	22 (55.0)	7 (87.5)			
Lesions margins:				**<0.001**	**<0.001**	NA
Well-defined	27 (100)	40 (100)	2 (25.0)			
Ill-defined	0 (0)	0 (0)	6 (75.0)			
Lesions border:				**0.004**	0.143	**0.003**
Round	9 (33.3)	28 (70.0)	2 (25.0)			
Lobulated	18 (66.7)	12 (30.0)	6 (75.0)			
Capsule	14 (51.9)	10 (25.0)	1 (12.5)	**0.031**	0.256	**0.025**
Hemorrhage	0 (0)	3 (7.5)	0 (0)	0.255	1.000	0.267
Cystic degeneration	1 (3.7)	8 (20.0)	1 (12.5)	0.156	1.000	0.073
Restricted diffusion	1 (3.7)	5 (12.5)	1 (12.5)	0.454	0.562	0.389
Enhancement:				**0.049**	0.223	**0.035**
Homogeneous	16 (59.3)	33 (82.5)	4 (50.0)			
Heterogeneous	11 (40.7)	7 (17.5)	4 (50.0)			
Enhancement pattern:				0.090	0.595	0.103
Whole	20 (74.1)	36 (90.0)	8 (100)			
Peripheral	7 (25.9)	4 (10.0)	0 (0)			
Early enhancement:				0.186	0.223	0.205
None-mild	22 (81.5)	27 (67.5)	4 (50.0)			
Moderate-marked	5 (18.5)	13 (32.5)	4 (50.0)			

(Continued)

Table 2. (Continued)

	Pleomorphic adenoma (n = 27)	Warthin tumor (n = 40)	Malignant tumors (n = 8)	*p* values Overall	*p* values Benign *vs* malignant	*p* values Pleomorphic *vs* Warthin
Late enhancement:				**0.014**	0.717	**0.004**
None-mild	10 (37.0)	29 (72.5)	4 (50.0)			
Moderate-marked	17 (63.0)	11 (27.5)	4 (50.0)			
Perineural spread	0 (0)	0 (0)	4 (50.0)	**<0.001**	**<0.001**	NA
Invasion of adjacent structure	0 (0)	0 (0)	6 (75.0)	**<0.001**	**<0.001**	NA
Enlarged lymph node	0 (0)	0 (0)	3 (37.5)	**<0.001**	**0.001**	NA

Footnotes: Continuous variable (size) is expressed as median and interquartile range (25th to 75th percentile) in parenthesis, categorical variables are expressed as numbers and percentages in parenthesis. Categorical variables were compared using the Pearson $\chi 2$ or Fisher exact test and continuous variable (size) using the Kuskal–Wallis or Mann–Whitney *U* test. NA: not available since the specific feature was a constant in benign tumors. Statistically significant values ($p < 0.05$) are highlighted in bold.

After segmentation, MaZda software (v. 4.6, Institute of Electronics, Technical University of Lodz, Lodz, Poland) was used to extract features. A total of 289 features were extracted for each ROI, including features that provide information related to the gray-level distribution within the ROI without considering spatial relations between voxels (first-order texture features), considering the spatial relations of voxels (second-order texture features), and evaluating spatial relations among three or more voxels (third-order texture features). Texture features automatically extracted using the Mazda software are calculated from the gray-level histogram analysis (*i.e.* mean, variance, skewness, kurtosis, and percentiles), gray-level co-occurrence matrix (GLCM), grey-level run-length matrix (GLRLM), autoregressive model, and Wavelet transform. Detailed description of extracted texture features is freely available on MaZda website.[24] Each lesion was labelled according to the reference standard.

Reference standard

The reference standard was assessed by a study coordinator, who had access to electronic patient medical records, including histopathological reports, but was not involved in the MRI qualitative or texture assessment of the lesions. Pathological analysis included surgical specimens in all patients. Based on pathology result, lesions were labelled as follows: pleomorphic adenoma for 27 lesions, Warthin tumor for 40 lesions and malignant tumors for 8 lesions, including 7 primary carcinomas and 1 B-cell lymphoma.

Informatics and statistical analysis

Statistical analysis was performed by a group of scientists in medical image processing and medical data analysis, who had access to the collected patients and tumors data, including MRI qualitative and MRI texture features with the corresponding reference standard, but was not involved in qualitative evaluation or lesions segmentations. Data summarized as continuous variables were expressed as mean and standard deviation (SD) or median and interquartile range (IQR) according to their normality distribution, and categorical variables were expressed as numbers and percentages.

Figure 2. MRI of pleomorphic adenoma (top row - arrow), Warthin tumor (middle row - arrow) and malignant tumor (bottom row - arrow) on axial T_1 weighted precontrast, SPIR, T_2 weighted, T_1 weighted post-contrast and DWI sequences. Of note, SPIR sequence is T_1 weighted and acquired post-injection. DWI, diffusion weighted imaging; SPIR, spectral presaturation with inversion recovery.

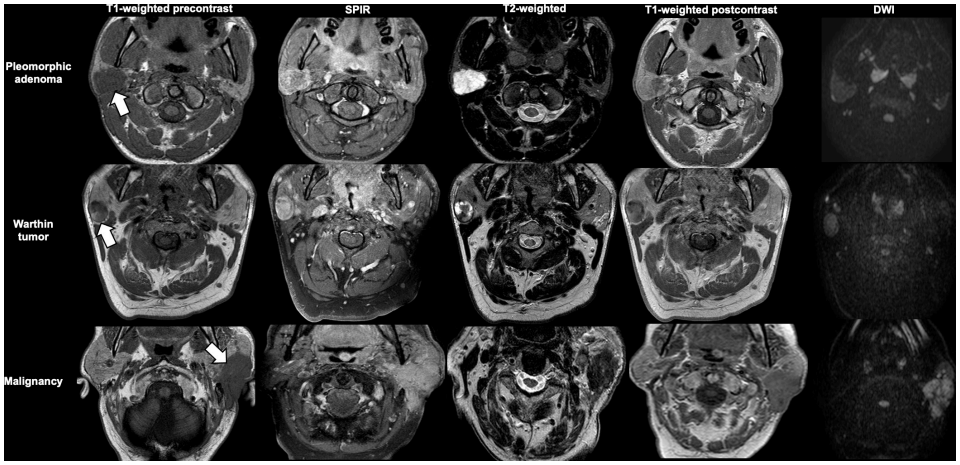


Table 3. Diagnostic performance of the qualitative variable and radiologist evaluation for the differential diagnosis between malignant and benign tumors

	AUROC	95% CI	*p* values	Se (%)	Sp (%)
Size (cut-off >3.5 cm)	0.737	0.552–0.952	0.029	50.0	97.0
Isohypointensty on T2	0.773	0.618–0.928	0.012	87.5	67.2
Ill-defined margins	0.875	0.693–1.000	0.001	75.0	100
Perineural spread	0.750	0.523–0.977	0.021	50.0	100
Invasion of adjacent structures	0.875	0.693–1.000	0.001	75.0	100
Enlarged lymph node	0.688	0.452–0.923	0.085	37.5	100
Non-subspecialized radiologist	0.860	0.678–1.000	0.001	75.0	97.0
Subspecialized radiologist	0.970	0.934–1.000	<0.001	100	94.0

AUROC, Area under the receiver operating characteristic curve; CI, Confidence interval; Se, Sensitivity; Sp, Specificity.

For MRI qualitative assessment, each MRI finding was assessed for the overall differences in parotid gland tumors and for pairwise comparisons between benign and malignant tumors or pleomorphic adenomas and Warthin tumors, respectively. Categorical variables were compared using the Pearson $\chi 2$ or Fisher exact test and continuous variables using the Kruskal–Wallis or Mann–Whitney *U* test, as appropriate. A p value < 0.05 was considered significant.

The diagnostic performance of imaging features was calculated, including odd ratio (OR) with 95% confidence interval (CI) at univariate and multivariate analysis. A binary logistic regression analysis was conducted using the imaging features statistically significant at univariate analysis as covariates and pathological diagnosis as dependent variable. Finally, the diagnostic performance (including sensitivity and specificity) of the radiologists for the diagnosis of malignancy and for the differential diagnosis between pleomorphic adenoma and Warthin tumor were calculated.

An operator-independent statistical system for feature reduction and selection, and the discriminant analysis predictive method for the differential diagnosis between benignity and malignancy and between pleomorphic adenoma and Warthin tumor was employed for MRI texture analysis as shown in Supplementary Material 1.[25–27] Intra reader variability was calculated by analyzing the differences between manual and standard ROIs using the intraclass correlation coefficient for individual measurements, *r*, and the *F* test statistic for individual intraclass correlation coefficient with the corresponding *p*-value for each feature, in each sequence. The diagnostic performance of each prediction model was finally calculated, including sensitivity, specificity, and area under the receiver operating characteristic curve (AUROC).

RESULTS

Study cohort

Of the 57 patients included in this study, 24 patients (42%, mean age 47 ± 18 years) had 27 pleomorphic adenomas (median size 1.8 cm; IQR 1.4–2.5 cm), 26 patients (46%, mean age 61 ± 16 years) had 40 Warthin tumors (median size 1.6 cm, IQR 1.3–1,9 cm), and 7 patients (12%, mean 66 ± 14 years) had 8 malignant tumors (median size 3.2 cm; IQR 1.9–4.6 cm).

Diagnostic performance of MRI qualitative analysis

Larger size ($p = 0.018$), isohypointensity on T_2 weighted images ($p = 0.009$), ill-defined margins ($p < 0.001$), perineural spread ($p < 0.001$), invasion of adjacent structures ($p < 0.001$), and enlarged lymph nodes ($p = 0.001$) were significantly more common in malignant lesions compared to benign ones (Table 2 and Figure 2). These qualitative features had a fair-to-good to high diagnostic performance (AUROC: 0.688–0.875; sensitivity: 37.5–87.5%; specificity: 67.2–100%) for the diagnosis of malignant tumors (Table 3). The diagnostic performance of radiologists for the diagnosis of malignancy had a sensitivity and a specificity of 75 and 97% with an AUROC of 0.860 for the non-subspecialized radiologist and 100 and 94% with an AUROC of 0.970 for the subspecialized radiologist, respectively (Table 3).

Compared to Warthin tumors, pleomorphic adenomas were most commonly solitary ($p = 0.002$), and unilateral ($p = 0.005$) lesions, with marked hyperintensity on T_2 weighted images ($p < 0.001$), lobulated border ($p = 0.003$), capsule ($p = 0.025$), heterogeneous enhancement ($p = 0.035$), and moderate-to-marked late enhancement ($p = 0.004$) (Table 2 and Figure 2) with a diagnostic performance for the identification of pleomorphic adenoma being poor-to-good (AUROC: 0.616–0.801%; sensitivity: 40.7–92.6%; specificity: 37.5–82.5%). At multivariate analysis the only independent predictor of the diagnosis of pleomorphic adenoma was the marked hyperintensity on T_2 weighted images [AUROC: 0.801 (95% CI: 0.687–0.916); OR: 26.12; 95% CI: 3.41–199.61; $p = 0.002$] (Table 4).

The diagnostic performance of radiologists for the diagnosis of pleomorphic adenoma had a sensitivity and a specificity of 81.5 and 70% respectively, with an AUROC of 0.757 (95%CI: 0.638–0.877) for the non-subspecialized radiologist and 81.5 and 95%, with an AUROC of 0.882 (95% CI 0.787–0.987) for the subspecialized radiologist (Table 4), with a moderate inter -reader agreement (Cohen's κ 0.54, 95% CI, 0.37–0.70).

Table 4. Diagnostic performance of the qualitative variable and radiologist evaluation for the differential diagnosis between pleomorphic adenoma and Warthin tumor and univariate and multivariate analyses for the diagnosis of pleomorphic adenoma

	Diagnostic performance					Univariate analysis			Multivariate analysis		
Characteristics	AUROC	95% CI	*p* values	Se (%)	Sp (%)	OR	95% CI	*p* values	OR	95% CI	*p* values
Solitary lesion	0.695	0.567–0.823	0.007	81.5	57.5	5.95	1.87–18.91	0.002	0.61	0.08–4.70	0.639
Unilateral lesion	0.650	0.520–0.781	0.038	92.6	37.5	7.50	1.55–36.2	0.005	12.33	0.75–200.58	0.077
Marked T2-hyperintensity	0.801	0.687–0.916	<0.001	77.8	82.5	16.50	4.87–55.88	<0.001	26.12	3.41–199.61	0.002
Lobulated border	0.683	0.551–0.816	0.011	66.7	70.0	4.66	1.63–13.30	0.003	3.25	0.67–15.70	0.142
Capsule	0.634	0.496–0.773	0.064	51.9	75.0	3.23	1.14–9.14	0.025	1.64	0.36–7.39	0.519
Heterogeneous enhancement	0.616	0.475–0.757	0.109	40.7	82.5	3.24	1.05–9.93	0.035	0.36	0.05–2.54	0.308
Moderate-marked late enhancement	0.677	0.544–0.811	0.014	63.0	72.5	4.48	1.57–12.74	0.004	2.79	0.60–12.86	0.186
Non-subspecialized radiologist	0.757	0.638–0.877	<0.001	81.5	70.0	–	–	–	–	–	–
Subspecialized radiologist	0.882	0.787–0.987	<0.001	81.5	95.0	–	–	–	–	–	–

AUROC, Area under the receiver operating characteristic curve; CI, Confidence interval; OR, Odds ratio; Se, Sensitivity; Sp, Specificity.

Diagnostic performance of MRI texture analysis

The intraclass correlation coefficient ranged from poor to excellent in the T_1 weighed precontrast sequence (r : −0.97–0.86; p-value: from <0.001 to 1), T_2 weighted sequence (r : −0.95–0.81; p-value: from <0.001 to 1), SPIR sequence (r : −0.94–0.96; p-value: from <0.001 to 1), and T_1 weighted post-contrast sequence (r : −0.92–0.98; p value: from <0.001 to 1). Manual segmentation of the tumors resulted in better specificity and AUROC for the majority of predictive models (8 of 10, 80%) both the diagnosis of malignancy (Table 5) and the differential diagnosis between pleomorphic adenomas and Warthin tumors (Table 6) compared to standardized regions of interest. The texture features automatically identified as the most significant per each sequence for the differential diagnosis between benign and malignant tumors and between Warthin tumors and pleomorphic adenomas are included in Supplementary Tables 2 and 3, respectively.

A combination of T_1 weighted precontrast images and T_2 weighted images achieved the best diagnostic performance overall for the differential diagnosis between benign and malignant tumors, including a sensitivity, specificity, and AUROC of 57.2%, 93.4%, and 0.927, respectively (Table 5, Figures 3 and 4). The highest sensitivity, specificity, and AUROC of radiomic approach for the differential diagnosis between Warthin tumors and pleomorphic adenomas was obtained through texture analysis of T_2 weighted sequence and corresponded to 70.8%, 82.5%, e 0.808, respectively (Table 6). By combining texture analysis of all sequences for the differential diagnosis between Warthin tumors and pleomorphic adenomas, there was an increase in sensitivity (91.5%), with a reduction in specificity (34.6%), and AUROC (0.701) (Table 6, Figures 5 and 6).

DISCUSSION

MRI is the gold-standard imaging modality in patients with parotid gland tumors.[28] MRI diagnosis of malignancy is quite straightforward because of the ill-defined margins, heterogeneous enhancement, and common invasion of the adjacent structures, while the differential diagnosis between pleomorphic adenoma and Warthin tumor is still challenging.[13,14] Multiple efforts have been done in the last decade to improve this differential diagnosis by identifying multiple quantitative MRI parameters (*e.g.* T-peak enhancement, diffusion kurtosis imaging, ADC values).[15–18] Fruehwald-Pallamar et al[20] recently evaluated whether some MRI texture features could help for the differentiation between pleomorphic adenomas and Warthin tumors, but they obtained limited results in terms of diagnostic performance. Their study, however, was limited by lack of combination of texture features from different sequences, lack of comparison with radiologist performance, and, above all, lack of texture analysis of T_2 weighted images. This latter was an important limitation because, as demonstrated in our study, the marked T2-hyperintensity is a strong independent predictor for the diagnosis of pleomorphic adenoma and a radiomic model based on T_2 weighted sequence demonstrated a good performance (AUROC: 0.808) for the differential diagnosis between pleomorphic adenomas and Warthin tumors.

Table 5. Diagnostic performance of predictive radiomics models for the diagnosis of malignancy based on texture analysis through manual or standardized segmentation

	AUROC	95% CI	*p* values	Se (%)	Sp (%)
Manual					
T1	0.696	0.527–0.865	0.010	51.2	92.3
T2	0.815	0.671–0.958	<0.001	56.0	94.1
SPIR	0.795	0.702–0.890	<0.001	51.2	94.8
POSTMDC	0.754	0.642–0.849	<0.001	36.9	93.5
T1 +T2	0.927	0.870–0.984	0.004	57.2	93.4
COMBINED	0.702	0.626–0.806	0.030	36.9	94.4
Standardized					
T1	0.880	0.934–1.000	<0.001	74.7	88.2
T2	0.609	0.678–1.000	0.012	25.0	94.2
SPIR	0.720	0.841–0.920	0.014	64.6	67.7
POSTMDC	0.640	0.545–0.674	<0.001	69.9	52.5
T1 +T2	0.845	0.684–0.898	<0.001	68.2	85.9
COMBINED	0.658	0.603–0.798	<0.001	74.3	58.4

AUROC, Area under the receiver operating characteristic curve; CI, Confidence interval; SPIR, spectral presaturation with inversion recovery; Se, Sensitivity; Sp, Specificity.

Our predictive radiomic model based on texture analysis of T_2 weighted images yielded a specificity of 82.5% for the diagnosis of pleomorphic adenomas, which was higher compared to the specificity of 70% achieved by our non-subspecialized radiologist, similar to the highest specificity of 86% reported in the literature for the MRI standard assessment by expert subspecialized radiologists, but lower than 95% specificity yielded by our subspecialized radiologist with 30-year-experience in head and neck radiology.[13,14] Interestingly, our predictive radiomics models showed that the combination of all sequences resulted in higher sensitivity for the diagnosis of pleomorphic adenoma compared to assessment of both non-subspecialized and subspecialized radiologists, or radiomic analysis of T_2 weighted images (sensitivities of 91.5%, 81.5%, 81.5%, and 70.8%, respectively), but without a benefit in terms of specificity (34.6%, 70.0%, 95.0% and 82.5%, respectively). These results are highly relevant from

Table 6. Diagnostic performance of predictive radiomics models for the differential diagnosis between Warthin tumor and pleomorphic adenoma based on texture analysis through manual or standardized segmentation

	AUROC	95% CI	*p* values	Se (%)	Sp (%)
Manual					
T1	0.715	0.625–0.805	0.004	70.3	65.7
T2	0.808	0.736–0.881	<0.001	70.8	82.5
SPIR	0.723	0.692–0.900	0.028	81.7	45.8
POSTMDC	0.676	0.622–0.829	0.008	88.0	38.9
T1 +T2	0.773	0.689–0.857	<0.001	73.8	81.3
COMBINED	0.701	0.624–0.805	0.002	91.5	34.6
Standardized					
T1	0.576	0.506–0.646	0.038	81.9	21.7
T2	0.709	0.644–0.774	<0.001	90.1	27.0
SPIR	0.636	0.591–0.792	<0.001	92.0	22.7
POSTMDC	0.573	0.523–0.722	0.030	76.4	38.6
T1 +T2	0.638	0.569–0.707	<0.001	90.0	25.8
COMBINED	0.622	0.522–0.704	<0.001	87.6	22.2

AUROC, Area under the receiver operating characteristic curve; CI, Confidence interval; SPIR, spectral presaturation with inversion recovery; Se, Sensitivity; Sp, Specificity.

Figure 3. ROC curves of prediction models based on radiomics using discriminant analysis classification with manual segmentation on T_1 weighted precontrast (a), T_2 weighted (b), SPIR (c), T1 post-contrast (d), T1 pre-contrast and T_2 weighted (e) and combined (f) sequences for the differential diagnosis between benign and malignant tumors. ROC, receiver operating characteristic; SPIR, spectral presaturation with inversion recovery.

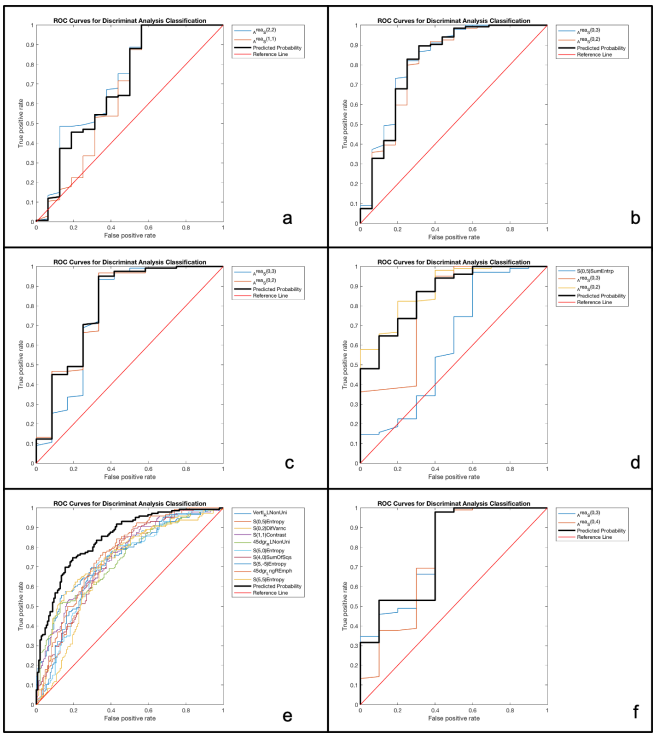
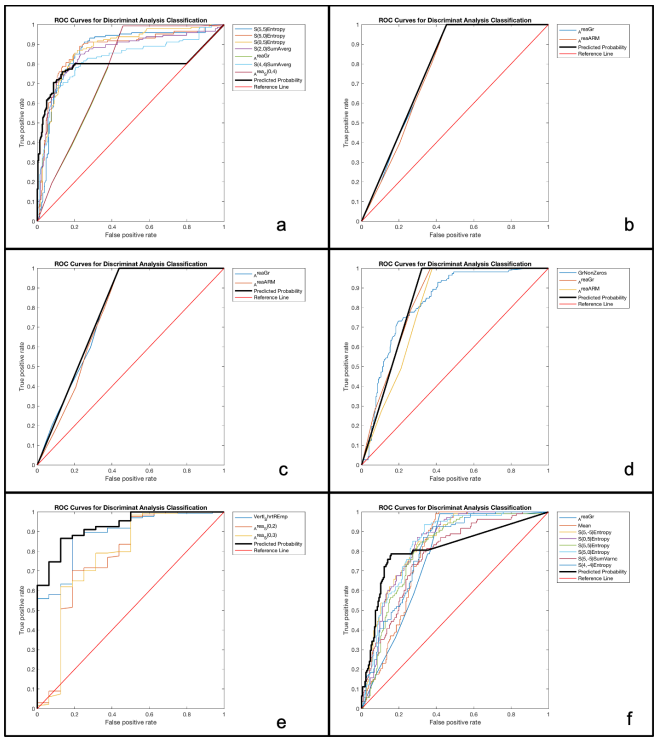


Figure 4. ROC curves of prediction models based on radiomics using discriminant analysis classification with standardized segmentation on T_1 weighted precontrast (a), T_2 weighted (b), SPIR (c), T1 post-contrast (d), T1 pre-contrast and T_2 weighted (e) and combined (f) sequences for the differential diagnosis between benign and malignant tumors. ROC, receiver operating characteristic; SPIR, spectral presaturation with inversion recovery.



a clinical standpoint because the management of pleomorphic adenoma and Warthin tumor is completely different and includes enucleation for Warthin tumors and superficial or total parotidectomy with preservation of mandibular branch of the facial nerve, the hypoglossal nerve, and the lingual nerve is recommended for pleomorphic adenomas. One of the advantages of radiomics, as compared to qualitative MRI assessment, is that it does not require any specific expertise for manual segmentation and feature extraction is automated. If our results are confirmed in external cohorts or prospective study, we would suggest that MRI texture analysis of T_2 weighted sequences should be integrated in clinical practice to improve the differentiation between pleomorphic adenomas and Warthin tumors.

Similarly to Fruehwald-Pallamar et al,[20] our results showed that the discrimination between benign and malignant parotid tumors based on texture analysis achieves, in general, better diagnostic performance than the discrimination between pleomorphic adenomas and Warthin tumors. However, in our study we showed that the translational clinical benefit of these results is negligible. Indeed, our radiologists outperformed MRI radiomic models for the diagnosis of malignancy, with higher sensitivity (75–100% *vs* 36.9–57.2%, respectively) and similar specificity (94–97% *vs* 92.3–94.8%). As already shown in prior studies,[17,22,29] multiple MRI qualitative variables are relevant in making the correct diagnosis of malignancy, and some of these features such as perineural spread, invasion of adjacent structures, and enlarged lymph nodes cannot be captured with the texture analysis of the parotid lesion. In addition, the diagnostic value of conventional MRI seems to be increased when DWI is applied,[29] although conflicting results on the role of DWI exist.[18] Based on the results of Yuan et al[29], parotid lesions with ADC values equal or less than 1.01 × 10–3 mm^2 s–one are more likely to be malignant. The adoption of multiparametric MRI is probably one of the reasons that explains why radiologists outperform radiomic analysis. Moreover, the radiomics models of malignant lesions may have been influenced by the different histopathological grade and differentiation of the included malignant tumors. Therefore, MRI qualitative assessment should still be preferred over radiomics approach for the diagnosis of malignant parotid tumors.

In addition to the small study cohort with the sample size of the malignant group comprising only eight cases, and the retrospective design, other limitations pertain to this study. First, radiomic analysis may be affected by several technical MRI parameters[30,31]; however, all MRI exams of this study were acquired using the same MRI scanner and acquisition protocol, which should have minimized this variability. Second, our radiologists performed a single session to read all MRI exams

Figure 5. ROC curves of prediction models based on radiomics using discriminant analysis classification with manual segmentation on T_1 weighted pre-contrast (a), T_2 weighted (b), SPIR (c), T1 post-contrast (d), T1 pre-contrast and T_2 weighted (e) and combined (f) sequences for the differential diagnosis between Warthin tumor and pleomorphic adenoma. ROC, receiver operating characteristic; SPIR, spectral presaturation with inversion recovery.

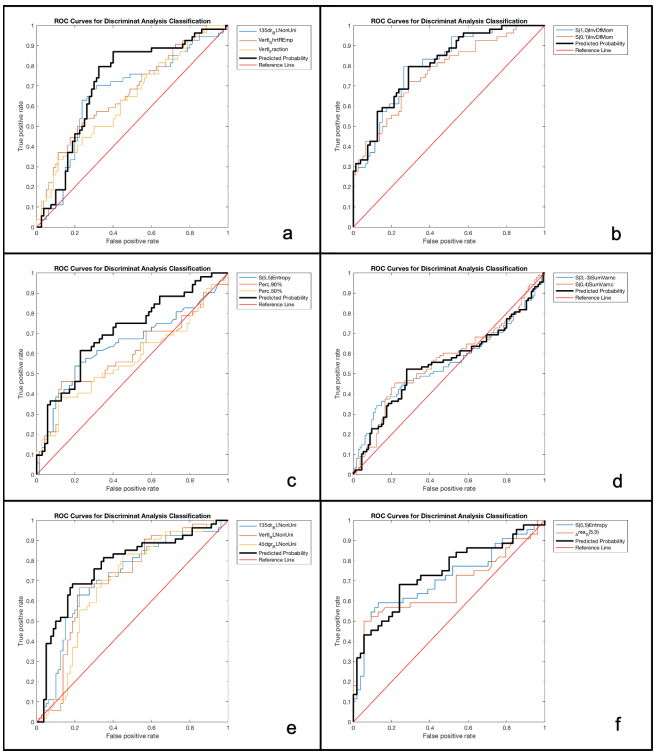
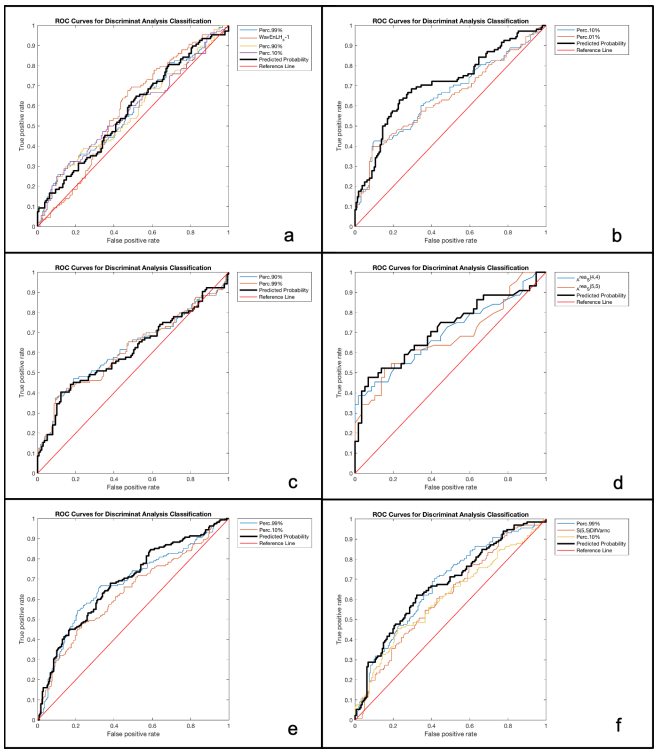


Figure 6. ROC curves of prediction models based on radiomics using discriminant analysis classification with standardized segmentation on T_1 weighted precontrast (a), T_2 weighted (b), SPIR (c), T1 post-contrast (d), T1 pre-contrast and T_2 weighted (e) and combined (f) sequences for the differential diagnosis between Warthin tumor and pleomorphic adenoma. ROC, receiver operating characteristic; SPIR, spectral presaturation with inversion recovery.



and potential bias due to testing or learning effects cannot be ruled out. To minimize this effect, MRI exams were shown using random case order. Third, we did not assess time required for segmentation of the lesions specifically; however, the radiology fellow estimated this time to be about 10 min for placement of all ROIs for the first analyzed cases and less than 5 min for the last analyzed cases. The radiomic analysis of our study is limited by the lack of analysis of DWI images, which were available to radiologists and are known to be helpful to characterize head and neck tumors, and by analysis of a single slide ROI. Another limitation is the variability of measurements in the intra reader assessment and lack of inter-reader agreement for lesion segmentation or MRI reading and the use of only one radiomics software. The intraclass correlation coefficient ranged from poor to excellent and this may be partially explained to the reliance of some of the texture features on the morphology of the ROIs. Reproducibility was beyond the purpose of this study and needs to be further analyzed for radiomics studies overall with a multi reader study to investigate whether manual or standardized ROIs should be preferred for increasing reproducibility of the features. In addition, further prospective studies in external cohorts are also warranted to validate our results considering that Mazda is a free radiomic software, thus facilitating future validation studies.

In conclusion, we demonstrated that radiologists outperform radiomic analysis for the diagnosis of malignancy, with some MRI qualitative features such as ill-defined margins, perineural spread, invasion of adjacent structures and enlarged lymph nodes being highly specific for malignancy. Conversely, a MRI predictive radiomics model based on texture analysis of T_2 weighted sequence improves the diagnostic performance of non-subspecialized radiologists for the differential diagnosis between pleomorphic adenoma and Warthin tumor. Interestingly, the highest specificity for the differential diagnosis between pleomorphic adenoma and Warthin tumor was achieved by the head and neck specialist radiologist, which means that adequate specific training allows to reduce false positives and unnecessarily radical operations.

FUNDING

Open Access Funding provided by Universita degli Studi di Palermo within the CRUI-CARE Agreement.

REFERENCES

1. Luers JC, Guntinas-Lichius O, Klussmann JP, Küsgen C, Beutner D, Grosheva M. The incidence of Warthin tumours and pleomorphic adenomas in the parotid gland over a 25-year period. *Clin Otolaryngol* 2016; **41**: 793-–7. doi: https://doi.org/10.1111/coa.12694
2. Bjørndal K, Krogdahl A, Therkildsen MH, Overgaard J, Johansen J, Kristensen CA, et al. Salivary gland carcinoma in Denmark 1990-2005: a national study of incidence, site and histology. Results of the Danish head and neck cancer Group (DAHANCA. *Oral Oncol* 2011; **47**: 677-–82. doi: https://doi.org/10.1016/j.oraloncology.2011.04.020
3. Bradley PJ, McGurk M. Incidence of salivary gland neoplasms in a defined UK population. *Br J Oral Maxillofac Surg* 2013; **51**: 399-–403. doi: https://doi.org/10.1016/j.bjoms.2012.10.002
4. Jo HJ, Ahn HJ, Jung S, Yoon H-K. Diagnostic difficulties in fine needle aspiration of benign salivary glandular lesions. *Korean J Pathol* 2012; **46**: 569–75. doi: https://doi.org/10.4132/KoreanJPathol.2012.46.6.569
5. Kechagias N, Ntomouchtsis A, Valeri R, Patrikidou A, Kitikidou K, Xirou P, et al. Fine-Needle aspiration cytology of salivary gland tumours: a 10-year retrospective analysis. *Oral Maxillofac Surg* 2012; **16**: 35–40. doi: https://doi.org/10.1007/s10006-011-0291-8
6. Dulguerov P, Todic J, Pusztaszeri M, Alotaibi NH. Why do parotid pleomorphic adenomas recur? A systematic review of pathological and surgical variables. *Front Surg* 2017; **4**: 26. doi: https://doi.org/10.3389/fsurg.2017.00026
7. Andreasen S, Therkildsen MH, Bjørndal K, Homøe P. Pleomorphic adenoma of the parotid gland 1985-2010: a Danish nationwide study of incidence, recurrence rate, and malignant transformation. *Head Neck* 2016; **38 Suppl 1**(Suppl 1): E1364-–9. doi: https://doi.org/10.1002/hed.24228
8. Valstar MH, de Ridder M, van den Broek EC, Stuiver MM, van Dijk BAC, van Velthuysen MLF, et al. Salivary gland pleomorphic adenoma in the Netherlands: a nationwide observational study of primary tumor incidence, malignant transformation, recurrence, and risk factors for recurrence. *Oral Oncol* 2017; **66**: 93-–9. doi: https://doi.org/10.1016/j.oraloncology.2017.01.004
9. Eida S, Sumi M, Nakamura T. Multiparametric magnetic resonance imaging for the differentiation between benign and malignant salivary gland tumors. *J Magn Reson Imaging* 2010; **31**: 673-–9. doi: https://doi.org/10.1002/jmri.22091
10. Yabuuchi H, Matsuo Y, Kamitani T, Setoguchi T, Okafuji T, Soeda H, et al. Parotid gland tumors: can addition of diffusion-weighted MR imaging to dynamic contrast-enhanced MR imaging improve diagnostic accuracy in characterization? *Radiology* 2008; **249**: 909-–16. doi: https://doi.org/10.1148/radiol.2493072045
11. Sumi M, Van Cauteren M, Sumi T, Obara M, Ichikawa Y, Nakamura T. Salivary gland tumors: use of intravoxel incoherent motion MR imaging for assessment of diffusion and perfusion for the differentiation of benign from malignant tumors. *Radiology* 2012; **263**: 770-–7. doi: https://doi.org/10.1148/radiol.12111248
12. Munhoz L, Ramos EADA, Im DC, Hisatomi M, Yanagi Y, Asaumi J, et al. Application of diffusion-weighted magnetic resonance imaging in the diagnosis of salivary gland diseases: a systematic review. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2019; **128**: 280–310. doi: https://doi.org/10.1016/j.oooo.2019.02.020
13. Kashiwagi N, Murakami T, Nakanishi K, Maenishi O, Okajima K, Takahashi H, et al. Conventional MRI findings for predicting submandibular pleomorphic adenoma. *Acta Radiol* 2013; **54**: 511-–5. doi: https://doi.org/10.1177/0284185113475922
14. Abdel Razek AAK, Samir S, Ashmalla GA. Characterization of parotid tumors with dynamic susceptibility contrast perfusion-weighted magnetic resonance imaging and diffusion-weighted MR imaging. *J Comput Assist Tomogr* 2017; **41**: 131-–6. doi: https://doi.org/10.1097/RCT.0000000000000486
15. Al Ajmi E, Forghani B, Reinhold C, Bayat M, Forghani R. 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* 2018; **28**: 2604-–11. doi: https://doi.org/10.1007/s00330-017-5214-0
16. Matsuda E, Fukuhara T, Donishi R, Kawamoto K, Hirooka Y, Takeuchi H. Usefulness of a novel ultrasonographic classification based on Anechoic area patterns for differentiating Warthin tumors from pleomorphic adenomas of the parotid gland. *Yonago Acta Med* 2017; **60**: 220-–6. doi: https://doi.org/10.33160/yam.2017.12.002
17. Elmokadem AH, Abdel Khalek AM, Abdel Wahab RM, Tharwat N, Gaballa GM, Elata MA, et al. Diagnostic accuracy of multiparametric magnetic resonance imaging for differentiation between parotid neoplasms. *Can Assoc Radiol J* 2019; **70**: 264-–72. doi: https://doi.org/10.1016/j.carj.2018.10.010
18. Mikaszewski B, Markiet K, Smugała A, Stodulski D, Szurowska E, Stankiewicz C. Diffusion-weighted MRI in the differential diagnosis of parotid malignancies and pleomorphic adenomas: can the accuracy of dynamic MRI be enhanced? *Oral Surg Oral Med Oral Pathol Oral Radiol* 2017; **124**: 95-–103. doi: https://doi.org/10.1016/j.oooo.2017.03.007
19. Qian W, Xu X-Q, Zhu L-N, Ma G, Su G-Y, Bu S-S, et al. Preliminary study of using diffusion kurtosis imaging for characterizing parotid gland tumors. *Acta Radiol* 2019; **60**: 887-–94. doi: https://doi.org/10.1177/0284185118803784
20. Fruehwald-Pallamar J, Czerny C, Holzer-Fruehwald L, Nemec SF, Mueller-Mang C, Weber M, et al. Texture-based and diffusion-weighted discrimination of parotid gland lesions on Mr images at 3.0 Tesla. *NMR Biomed* 2013; **26**: 1372-–9. doi: https://doi.org/10.1002/nbm.2962
21. Christe A, Waldherr C, Hallett R, Zbaeren P, Thoeny H. MR imaging of parotid tumors: typical lesion characteristics in MR imaging improve discrimination between benign and malignant disease. *AJNR Am J Neuroradiol* 2011; **32**: 1202-–7. doi: https://doi.org/10.3174/ajnr.A2520
22. Tartaglione T, Botto A, Sciandra M, Gaudino S, Danieli L, Parrilla C, et al. Differential diagnosis of parotid gland tumours: which magnetic resonance findings should be taken in account? *Acta Otorhinolaryngol Ital* 2015; **35**: 314-–20. doi: https://doi.org/10.14639/0392-100X-693
23. Lee YYP, Wong KT, King AD, Ahuja AT. Imaging of salivary gland tumours. *Eur J Radiol* 2008; **66**: 419-–36. doi: https://doi.org/10.1016/j.ejrad.2008.01.027
24. Szczypiński PM, Strzelecki M, Materka A, Klepaczko A. MaZda--a software package for image texture analysis. *Comput Methods Programs Biomed* 2009; **94**: 66–76. doi: https://doi.org/10.1016/j.cmpb.2008.08.005
25. Giambelluca D, Cannella R, Vernuccio F, Comelli A, Pavone A, Salvaggio L, et al. PI-RADS 3 lesions: role of prostate MRI texture analysis in the identification of prostate cancer. *Curr Probl Diagn Radiol* 2021; **50**: 175–85. doi: https://doi.org/10.1067/j.cpradiol.2019.10.009
26. Comelli A, Stefano A, Coronnello C. Radiomics: a new biomedical workflow to create a predictive model. In: *Annual Conference on Medical Image Understanding*

and Analysis. Springer, Cham; 2020. pp. 280–93.
27. Comelli A, Stefano A, Bignardi D. Tissue classification to support local active delineation of brain tumors. In: *Annual Conference on Medical Image Understanding and Analysis. Springer, Cham*; 2019. pp. 3–14.
28. Kato H, Kawaguchi M, Ando T, Mizuta K, Aoki M, Matsuo M. Pleomorphic adenoma of salivary glands: common and uncommon CT and MR imaging features. *Jpn J Radiol* 2018; **36**: 463-–71. doi: https://doi.org/10.1007/s11604-018-0747-y
29. Yuan Y, Tang W, Tao X. Parotid gland lesions: separate and combined diagnostic value of conventional MRI, diffusion-weighted imaging and dynamic contrast-enhanced MRI. *Br J Radiol* 2016; **89**: 20150912. doi: https://doi.org/10.1259/bjr.20150912
30. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: a systematic review. *Int J Radiat Oncol Biol Phys* 2018; **102**: 1143-–58. doi: https://doi.org/10.1016/j.ijrobp.2018.05.053
31. Meyer M, Ronald J, Vernuccio F, Nelson RC, Ramirez-Giraldo JC, Solomon J, et al. Reproducibility of CT radiomic features within the same patient: influence of radiation dose and CT reconstruction settings. *Radiology* 2019; **293**: 583-–91. doi: https://doi.org/10.1148/radiol.2019190928