*Article*

# Radiomics for Discriminating Benign and Malignant Salivary Gland Tumors; Which Radiomic Feature Categories and MRI Sequences Should Be Used?

Rongli Zhang [1], Qi Yong H. Ai [1,2], Lun M. Wong [1], Christopher Green [1], Sahrish Qamar [1], Tiffany Y. So [1], Alexander C. Vlantis [3] and Ann D. King [1,*]

[1] Department of Imaging and Interventional Radiology, Faculty of Medicine, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong SAR, China
[2] Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong SAR, China
[3] Department of Otorhinolaryngology, Head and Neck Surgery, Prince of Wales Hospital, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China
* Correspondence: king2015@cuhk.edu.hk

**Simple Summary:** MRI radiomics shows promise in discriminating salivary gland tumors (SGTs) but a consistent radiomics signature has not emerged, partly due to the multitude of initial features fed into the radiomics pipeline. In this study, we investigated the impact of reducing the number of initial radiomic features on the performance of the radiomic models to discriminate between benign and malignant SGTs, by applying six feature categories separately and all feature categories in combination from three anatomical-based MRI sequences. The best models were built by a combination of T1-weighted + logarithm and fat-suppressed T2-weighted + exponential features, which reduced the initial features by 94.0% (from $1015 \times 3$ to $91 \times 2$) and achieved an average area under the curve of 0.846. Our results show reducing the number of radiomic features initially analyzed improved feature selection stability without compromising performance. This approach may improve future consensus building on a radiomics signature for discriminating SGTs.

**Abstract:** The lack of a consistent MRI radiomic signature, partly due to the multitude of initial feature analyses, limits the widespread clinical application of radiomics for the discrimination of salivary gland tumors (SGTs). This study aimed to identify the optimal radiomics feature category and MRI sequence for characterizing SGTs, which could serve as a step towards obtaining a consensus on a radiomics signature. Preliminary radiomics models were built to discriminate malignant SGTs ($n = 34$) from benign SGTs ($n = 57$) on T1-weighted (T1WI), fat-suppressed (FS)-T2WI and contrast-enhanced (CE)-T1WI images using six feature categories. The discrimination performances of these preliminary models were evaluated using 5-fold-cross-validation with 100 repetitions and the area under the receiver operating characteristic curve (AUC). The differences between models' performances were identified using one-way ANOVA. Results show that the best feature categories were logarithm for T1WI and CE-T1WI and exponential for FS-T2WI, with AUCs of 0.828, 0.754 and 0.819, respectively. These AUCs were higher than the AUCs obtained using all feature categories combined, which were 0.750, 0.707 and 0.774, respectively ($p < 0.001$). The highest AUC (0.846) was obtained using a combination of T1WI + logarithm and FS-T2WI + exponential features, which reduced the initial features by 94.0% (from $1015 \times 3$ to $91 \times 2$). CE-T1WI did not improve performance. Using one feature category rather than all feature categories combined reduced the number of initial features without compromising radiomic performance.

**Keywords:** radiomics; salivary gland neoplasms; conventional magnetic resonance imaging

## 1. Introduction

Salivary gland tumors (SGTs) account for 2–6.5% of all head and neck tumors [1,2]. Around 80% of SGTs arise from the parotid gland of which about 80% are benign (BSGT), mainly pleomorphic adenoma (PA) and Warthin's tumor (WT) while the remaining 15–30% are malignant (MSGTs) [3–5], mainly comprising mucoepidermoid carcinoma, acinic cell carcinoma, adenoid cystic carcinoma, carcinoma ex-pleomorphic adenoma, and adenocarcinoma [6]. Magnetic resonance imaging (MRI) is often the preferred imaging modality in patients with a histologically proven SGT, to map the extent of disease and image deep surrounding structures with advantages over CT for the avoidance of ionizing radiation, better contrast resolution especially in the depiction of local spread including peri-neural disease, and visualization of the relationship of the tumor to the facial nerve and branches [5,7]. However, there are situations when the pathology of the SGT is unknown, such as incidentally found SGTs on a head and neck MRI examination or when tumor location limits access to biopsy or fine needle aspiration cytology. Furthermore, malignant SGTs may be missed or over-diagnosed using parotid fine-needle aspiration cytology and there may be sampling difficulties related to SGT heterogeneity [8,9].

Morphological features of MSGTs on MRI include irregular margins, invasion beyond the gland into adjacent structures, perineural extension, and metastatic regional lymph nodes. Signal intensity and markers from functional MRI techniques such as diffusion-weighted imaging and dynamic contrast-enhanced MRI also help to discriminate MSGTs and BSGTs [9,10]. However, there is still overlap in appearances, notably some benign tumors have irregular margins while low grade malignancies may have benign features [11–13]. Moreover, MRI-based tumor diagnosis by visual qualitative assessment relies primarily on the radiologist's experience, which could lead to objective variations, especially in the complex head and neck region [14,15]. Therefore, quantitative methods to characterize SGTs on MRI could facilitate the clinical workflow by improving diagnostic accuracy and reducing inter-observer variability.

Radiomic analysis involves high-throughput quantitative imaging features extracted from a region of interest in medical images [16]. The radiomic signature produced by combining the best-performing features could be used to discriminate between different tumor types. Such radiomic signatures from MRI have shown early promise in discriminating SGTs [17–23]. Nonetheless, in keeping with most other radiomics oncology studies, a consistent MRI radiomic signature has not emerged in the literature, limiting the widespread clinical application of radiomics for the discrimination of SGTs.

Radiomic packages commonly involve thousands of features, comprising six main feature categories: shape, first-order features, texture features as well as filtered base derivatives (i.e., exponential, logarithm and wavelet) [24]. These numerous features are commonly extracted from relatively small-sized MRI samples, which causes dimensionality problems [25–27] that must be tackled by additional methods and steps such as least absolute shrinkage and selection operator (LASSO), ridge and elastic net, ..., etc. to shrink the dimensionality and identify potential features [28–30]. However, when the initial pool of features is very large, it is unlikely that these techniques will select the same combination of features every time to produce a stable radiomics model [28–33]. With so many features to evaluate at the start of the analysis, it is not surprising that radiomic signatures for discriminating SGTs rarely contain the same combination of features which limits the wider use of radiomics in clinical practice. To take a step towards improving stability of the selected radiomic signature, it would be advantageous to limit the number of initial feature categories fed into the radiomic pipeline while still providing acceptable diagnostic performance.

The aim of this study was to determine if we could improve feature selection stability by restricting the initial number of radiomic features fed into the pipeline without compromising the performance of MRI radiomic features in discriminating between BSGTs and MSGTs. We evaluated the impact of restricting the number of initial features by narrowing

down to one feature category and MRI sequences, comparing between T1-weighted (T1WI), fat-suppressed T2WI (FS-T2WI) and contrast-enhanced T1WI (CE-T1WI) images.

## 2. Materials and Methods

### 2.1. Patient Characteristics

This retrospective study was approved by our Institutional Review Board, and the requirement for informed consent was waived. The enrolled patients had histologically confirmed SGTs (34 MSGTs, 57 BSGTs) and had undergone three MR examinations that included axial T1-weighted (T1WI), fat-suppressed T2WI (FS-T2WI) and contrast-enhanced T1WI (CE-T1WI) images. The patient demographics and tumor distributions are detailed in Table 1.

**Table 1.** General characteristics of the patient cohort. Two-tailed unpaired student *t*-test was used to evaluate differences in characteristics across MSGTs and BSGTs.

| Characteristics | MSGT (*n* = 34) | BSGT (*n* = 57) | *p*-Value |
|---|---|---|---|
| Tumor histology | Lymphoepithelioma-like carcinoma 7 (20.6%) Myoepithelial carcinoma 2 (5.9%) Salivary duct carcinoma 4 (11.8%) Adenoid cystic carcinoma 5 (14.7%) Mucoepidermoid carcinoma 8 (21.6%) Metastatic carcinoma 2 (5.9%) Acinic cell carcinoma 1 (2.9%) Poorly differentiated carcinoma 2 (5.9%) Basal cell adenocarcinoma 1 (2.9%) Other carcinomas 2 (5.9%) | Pleomorphic adenoma 44 (77.2%) Warthin's tumor 13 (22.8%) | |
| Sex (M/F) | 20/14 | 30/27 | 0.57 |
| Age (years) | 57.94 ± 16.76 | 55.35 ± 15.91 | 0.25 |
| Tumor location | Parotid 25 (73.5%) Submandibular 5 (14.7%) Sublingual 4 (11.8%) | Parotid 51 (89.5%) Submandibular 6 (10.5%) Sublingual 0 (0%) | |
| Tumor site | Unilateral 34 (100%) Bilateral 0 (0%) | Unilateral 47 (90.4%) Bilateral 5 (9.6%) | |

Numerical data are presented as means ± standard deviations, categorical data as numbers (*n*). M = male, F = female; MSGT = malignant salivary gland tumor; BSGT = benign salivary gland tumor.

### 2.2. Image Acquisition

All MRI examinations were performed on a Philips 3.0 Tesla scanner (Achieva TX, Phillips Healthcare, Best, The Netherlands) with a 16-channel head and neck coil for radiofrequency pulse transmission and a neurovascular phased-array coil for imaging reception. The data acquisition protocols consisted of axial (1) T1WI: repetition time (TR) = 298–715 ms, echo time (TE) = 10 ms, echo number = 1, slice thickness = 4 mm, flip angle = 90°; (2) FS-T2WI: TR = 1825–5412 ms, TE = 80 ms, slice thickness = 4 mm, fat-suppression technique and spectral attenuation inversion recovery; and (3) CE-T1WI: TR = 298–655 ms, TE = 10 ms, echo number = 1, slice thickness = 4 mm, flip angle = 90°.

### 2.3. Tumor Segmentation

All salivary gland tumors on each MRI sequence were manually segmented by a researcher (Q.Y.H.A.) with seven years of experience with MRI of head and neck tumors, using open-source software ITK-SNAP (version 3.4.0; http://www.itksnap.org, accessed on 1 October 2020) [34]. To assess inter-observer agreement, 30 salivary gland tumors (10 MTs, 10 PAs and 10 WTs) were randomly selected and manually segmented on each MRI sequence by a second researcher (S.Q.) with four years of experience with head and neck MRI who was blind to the patients' diagnoses and the segmentation of the first researcher (Q.Y.H.A.). The volumetric dice similarity coefficient (DSC) [35] was used to calculate the inter-observer segmentation agreement. DSC of <0.6, 0.6–<0.8, 0.8–1.0 and 1.0 indicates inadequate, good, very good and ideal consistency, respectively [36].

### 2.4. Image Pre-Processing

To normalize image intensity and minimize the effects of variation in the weighted MRI scanning parameters, the N4ITK bias field correction algorithm was implemented to

remove the artifacts caused by the inhomogeneity of the scanner's magnetic field [37]. The well-established "$\mu \pm 3\sigma$" algorithm was applied to identify and remove image intensity outliers [38]. Next, the images were resampled to the median spacing of the training cases. Lastly, z-score normalization was implemented in the non-zero areas for intensity normalization [39].

### 2.5. Feature Extraction

From the volume of interest of the segmented tumors on each MRI sequence (T1WI, CE-T1WI and FS-T2WI), 1015 3D quantitative features were extracted using PyRadiomics (version 3.0.1) (available at https://pyradiomics.readthedocs.io/en/latest/, accessed on 5 November 2020) [24]. All quantitative features ($n = 1015$) were divided into six feature categories: (1) shape ($n = 14$), (2) first-order ($n = 18$), (3) texture ($n = 73$), (4) exponential ($n = 91$), (5) logarithm ($n = 91$) and (6) wavelet ($n = 801$) (Datasheet S1). Image pre-processing and feature extraction operation was performed using the open-source packages SimpleITK (version 2.1.1) [40] and PyRadiomics (version 3.0.1) [24] on Python (version 3.7.10) programing language.

### 2.6. Data Augmentation

To reduce potential bias caused by imbalance between the number of positive and negative samples, data augmentation was performed using a synthetic minority oversampling technique (SMOTE) [41,42] in both the training and validation sets. Data augmentation was performed on the MATLAB 2020a (MathWorks, Natick, MA, USA) software platform, using an open-source package available at https://github.com/Nekooeimehr/MATLAB-Source-Code-Oversampling-Methods, accessed on 5 November 2020).

### 2.7. Feature Selection

All features and features grouped by the six feature categories extracted from T1WI, CE-T1WI and FS-T2WI were pooled for the feature selection procedure. Before feature selection, quantitative features were standardized by z-score normalization. Radiomic feature selection was performed on the training dataset. For each cross-validation loop, first, the *p*-values of the individual features of the internal training set were calculated using the two-tailed unpaired *t*-test (for features with normalized distribution) and the Wilcoxon rank-sum test (for features with non-normalized distribution) [43], features with a $p < 0.05$ were enrolled for the next step; second, the LASSO algorithm [44], a well-established dimensionality shrinkage approach, was performed to identify the potential features with the regularization parameter ($\lambda$) determined using the minimum criteria via 10-fold cross-validation; third, the features with non-zero LASSO coefficients were registered as the selected features.

### 2.8. Radiomics Models Construction and Evaluation

A clinically recognized multivariable logistic regression (LR) classifier was used to construct radiomic models with the selected features. We applied 5-fold cross-validation with 100 repetitions to assess model performance. The performance metric was the area under the receiver operating characteristic curve (AUC). The stability strength of the feature selection for the model building was evaluated using Nogueira score [29,31] and Jaccard index [31]. The feature selection, model construction and evaluation were performed using in-house code and the "Glmnet" package [45] (Qian, J. http://www.stanford.edu/~hastie/glmnet_matlab/, accessed on 5 November 2020) on the MATLAB R2020a (MathWorks, Natick, MA, USA). The schematic workflow of the radiomics method is shown in Figure 1.
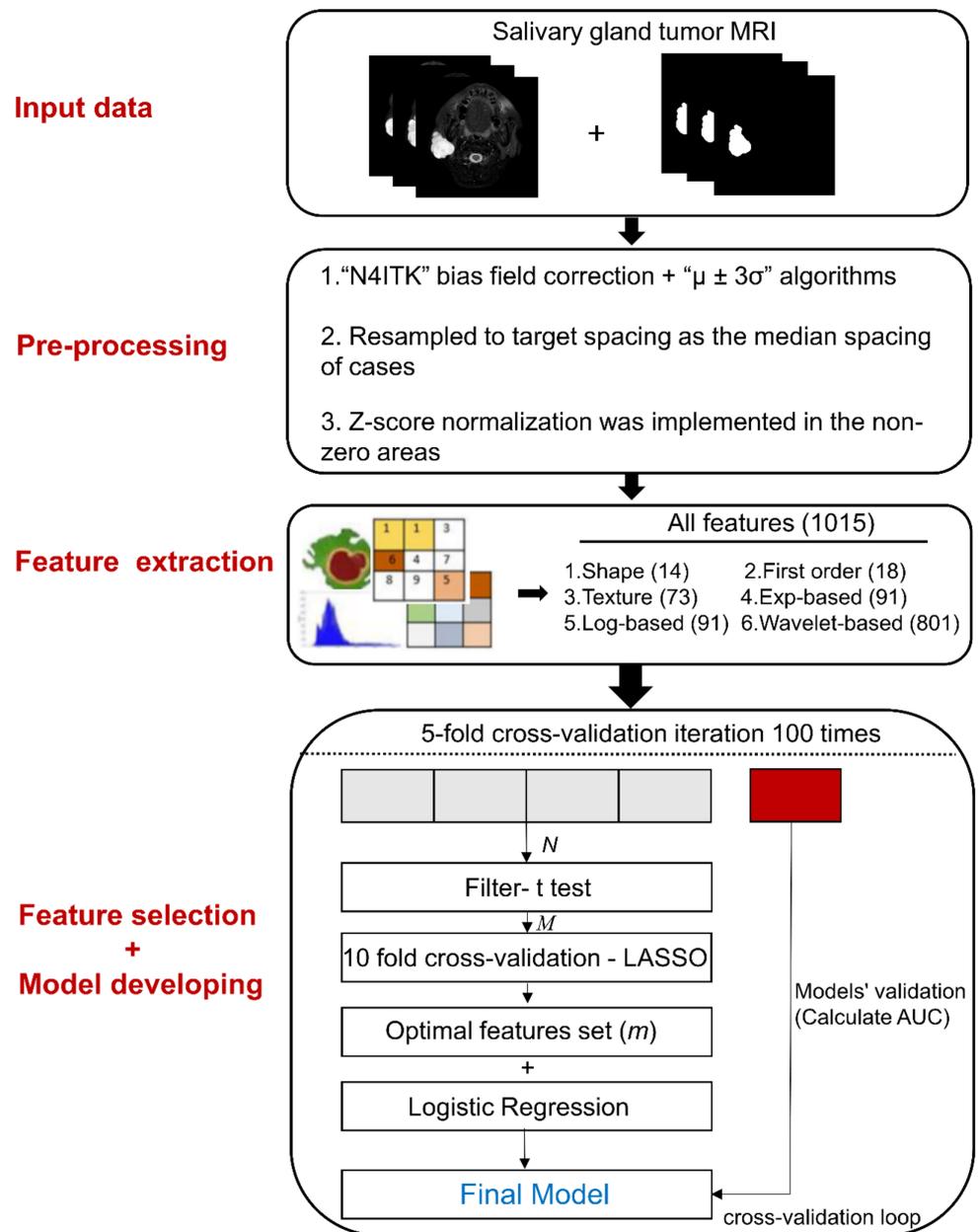
**Figure 1.** Overview of the radiomic analysis framework to differentiate benign and malignant salivary gland tumors. Region of interest segmentation was performed by experienced clinical researchers. After pre-processing, radiomic features were extracted and categorized into six groups. Feature selection was performed using 10-fold cross-validation in the training dataset. Radiomic models were constructed using multivariable logistic regression classifiers for salivary gland tumor differentiation. Finally, models were validated by 5-fold cross-validation with 100 repetitions. Exp = exponential; Log = logarithm; LASSO = least absolute shrinkage and selection operator; AUC = the area under the receiver operating characteristic curve.

*2.9. Selection of the Best Sequences and Feature Categories*

The performances (AUCs) of the models based on (1) shape ($n = 14$), (2) first-order ($n = 18$), (3) texture ($n = 73$), (4) exponential ($n = 91$), (5) logarithm ($n = 91$), (6) wavelet ($n = 801$) and all features ($n = 1015$) for each sequence (T1WI, CE-T1WI and FS-T2WI) were compared to determine the best feature category for each sequence. Next, to investigate whether the various combinations of sequences could improve performance, ra-

diomics models were built by adding sequences one by one according to the performance of each sequence.

### 2.10. Statistical Analysis

The AUCs of all models were compared using a one-way analysis of variance (ANOVA). Differences between the age, sex of the patients and Jaccard index of the different methods were compared using an independent samples *t*-test. The difference between the Nogueira score of the different methods was tested using a technique reported by S. Nogueira [29]. In all conditions, a *p* < 0.05 was considered statistically significant. Statistical analysis was performed using GraphPad Prism software (version 5.01, Dotmatics, San Diego, CA, USA).

## 3. Results

### 3.1. Radiomic Analysis to Discriminate between MSGTs and BSGTs

In the training set, all sequences with the corresponding feature categories and all feature categories combined (except for the first-order features, where the AUC was 0.554 for T1WI and 0.625 for CE-T1WI) showed the potential to discriminate between MSGTs and BSGTs, with AUCs of T1WI: 0.721–0.999, FS-T2WI: 0.833–0.998 and CE-T1WI: 0.706–0.997. Similar to the training set, all sequences in the validation set, other than the first-order features (AUCs of 0.552 for T1WI and 0.605 for CE-T1WI), showed the potential to discriminate between MSGTs and BSGTs, with AUCs of T1WI: 0.718–0.828, FS-T2WI: 0.774–0.819 and CE-T1WI: 0.689–0.754 (Table 2).

**Table 2.** The performance (AUC) of MRI sequence and feature category for discriminating MSGTs from BSGTs.

|  | Shape (n = 14) | First Order (n = 18) | Texture (n = 73) | Exp (n = 91) | Log (n = 91) | Wavelet (n = 801) | All Features (n = 1015) |
|---|---|---|---|---|---|---|---|
| | | | *Validation set* | | | | |
| T1WI | 0.718 ± 0.004 | 0.552 ± 0.003 | 0.801 ± 0.004 | 0.729 ± 0.004 | **0.828 ± 0.004 \*\*\*** | 0.725 ± 0.005 | 0.750 ± 0.004 |
| FS-T2WI | 0.778 ± 0.004 | 0.788 ± 0.004 | 0.785 ± 0.004 | **0.819 ± 0.004 \*\*\*** | 0.806 ± 0.004 | 0.785 ± 0.005 | 0.774 ± 0.004 |
| CE-T1WI | 0.704 ± 0.004 | 0.605 ± 0.003 | 0.729 ± 0.004 | 0.747 ± 0.004 | **0.754 ± 0.005 \*\*\*** | 0.689 ± 0.004 | 0.707 ± 0.004 |
| | | | *Training set* | | | | |
| T1WI | 0.721 ± 0.003 | 0.554 ± 0.003 | 0.871 ± 0.003 | 0.835 ± 0.002 | 0.902 ± 0.001 | 0.996 ± 0.000 | 0.999 ± 0.000 |
| FS-T2WI | 0.833 ± 0.002 | 0.841 ± 0.001 | 0.891 ± 0.001 | 0.924 ± 0.001 | 0.926 ± 0.001 | 0.998 ± 0.000 | 0.998 ± 0.000 |
| CE-T1WI | 0.706 ± 0.002 | 0.625 ± 0.004 | 0.845 ± 0.002 | 0.862 ± 0.001 | 0.866 ± 0.002 | 0.980 ± 0.001 | 0.997 ± 0.000 |

Numerical data are presented as means ± standard errors. AUC = area under the receiver operating characteristic curve; MSGT = malignant salivary gland tumor; BSGT = benign salivary gland tumor; *n* = the initial feature number entered the radiomics procedure; T1WI = T1-weighted imaging; FS-T2WI = fat-suppressed T2-weighted imaging; CE-T1WI = contrast-enhanced T1WI. Exp = exponential; log = logarithm; differences were considered statistically significant for \*\*\* = *p* < 0.001. Bold indicates the highest AUC mean value for each sequence in the validation set.

### 3.2. Performance Comparison of Each Feature Category and All Features Combined

To discriminate between MSGTs and BSGTs on each MRI sequence, the following feature categories with the best discriminate performances were identified according to the validation results shown in Table 2. For T1WI, the logarithm-based features (AUC of 0.828) were compared with other feature subcategories (AUCs of 0.552–0.801, *p* < 0.001). For FS-T2WI, the exponential-based features (AUC of 0.819) were compared with other feature subcategories (AUCs of 0.778–0.806, *p* < 0.001). For CE-T1WI, the logarithm-based features (AUC of 0.754) were compared with other feature subcategories (AUCs of 0.605–0.747, *p* < 0.001). For all sequences, preliminary results built using the best performing feature category achieved significantly higher AUCs than those built using all features combined: T1WI, 0.828 vs. 0.750; FS-T2WI, 0.819 vs. 0.774; and CE-T1WI, 0.754 vs. 0.707, respectively (*p* < 0.001).

### 3.3. Comparison of Stability Strength and Number of Features Based on the Best Features Category and All Combined Features

The feature selection stability based on the best features category achieved higher values than all features combined both in Nogueira score (T1WI: 0.437 vs. 0.360, FS-T2WI: 0.466 vs. 0.292, CE-T1WI: 0.433 vs. 0.331) and Jaccard index (T1WI: 0.330 vs.0.234, FS-T2WI: 0.368

vs.0.184, CE-T1WI: 0.322 vs.0.219) (all $p < 0.001$, Table 3). Using the best feature category (logarithm for T1WI and CE-T1WI, $n = 91$; exponential for FS-T2WI, $n = 91$) for each sequence and for T1WI and FS-T2WI combined reduced the initial input of features from 1015 to 91 (91.0%) and from $1015 \times 3$ to $91 \times 2$ (94.0%), respectively.

**Table 3.** Stability strength of feature selection based on different methods.

|  | All Features | Best Feature Category | *p*-Values |
|---|---|---|---|
| Nogueira score |  |  |  |
| T1WI | 0.360 | **0.437** | <0.001 |
| FS-T2WI | 0.292 | **0.466** | <0.001 |
| CE-T1WI | 0.331 | **0.433** | <0.001 |
| Jaccard index |  |  |  |
| T1WI | $0.234 \pm 0.066$ | **0.330 ± 0.145** | <0.001 |
| FS-T2WI | $0.184 \pm 0.069$ | **0.368 ± 0.150** | <0.001 |
| CE-T1WI | $0.219 \pm 0.067$ | **0.322 ± 0.137** | <0.001 |

Numerical data are presented as means $\pm$ standard deviation. T1WI = T1-weighted imaging; FS-T2WI = fat-suppressed T2-weighted imaging; CE-T1WI = contrast-enhanced T1WI. Differences were considered statistically significant for $p < 0.05$. Bold indicates the highest Nogueira score and Jaccard index for each sequence.

### 3.4. Selection of MRI Sequences to Discriminate between MSGTs and BSGTs

Using the best feature category for each sequence, the preliminary radiomics models built on T1WI-logarithm, T1WI-logarithm combined with FS-T2WI-exponential and T1WI-logarithm combined with FS-T2WI-exponential and CE-T1WI-logarithm achieved AUCs of 0.828, 0.846 and 0.825; accuracies of 0.750, 0.761 and 0.751; sensitivities of 0.730, 0.740 and 0.728; and specificities of 0.769, 0.782 and 0.775, respectively, in validation set (Table 4).

**Table 4.** Assessing the performances of the combined MRI sequence based on the best feature category to discriminate MSGTs from BSGTs.

|  | Validation Set | | | Training Set | | |
|---|---|---|---|---|---|---|
|  | T1WI-Log ($n = 91$) | T1WI-Log + FS-T2WI-Exp ($n = 182$) | T1WI-Log + FS-T2WI-Exp + CE-T1WI-Log ($n = 273$) | T1WI-Log ($n = 91$) | T1WI-Log + FS-T2WI-Exp ($n = 182$) | T1WI-Log + FS-T2WI-Exp + CE-T1WI-Log ($n = 273$) |
| AUC | $0.828 \pm 0.004$ | **0.846 ± 0.004 \*\*** | $0.825 \pm 0.005$ | $0.902 \pm 0.001$ | $0.953 \pm 0.001$ | $0.978 \pm 0.001$ |
| Accuracy | $0.750 \pm 0.004$ | **0.761 ± 0.004** | $0.751 \pm 0.004$ | $0.837 \pm 0.002$ | $0.885 \pm 0.001$ | $0.951 \pm 0.001$ |
| Sensitivity | $0.730 \pm 0.005$ | **0.740 ± 0.005** | $0.728 \pm 0.005$ | $0.846 \pm 0.002$ | $0.893 \pm 0.002$ | $0.950 \pm 0.001$ |
| Specificity | $0.769 \pm 0.005$ | **0.782 ± 0.005** | $0.775 \pm 0.005$ | $0.826 \pm 0.003$ | $0.878 \pm 0.002$ | $0.951 \pm 0.001$ |

Numerical data are presented as means $\pm$ standard errors. MSGT = malignant salivary gland tumor; BSGT = benign salivary gland tumor; $n$ = the initial feature number entered the radiomics procedure; T1WI = T1-weighted imaging; FS-T2WI = fat-suppressed T2-weighted imaging; CE-T1WI = contrast-enhanced T1WI; exp = exponential; log = logarithm; differences were considered statistically significant for \*\*: $0.001 < p < 0.01$. Bold indicates the best performance in the validation set.

### 3.5. Inter-Observer Agreement for Segmentation

The inter-observer agreement for tumor segmentation on T1WI, FS-T2WI and CE-T1WI showed mean DSC values of $0.843 \pm 0.065$, $0.862 \pm 0.059$ and $0.827 \pm 0.067$, respectively (Datasheet S2).

### 3.6. Additional Analysis to Further Reduce Radiomic Features

Having completed the aim of this study and shown that reducing from the number of initial features ($1015 \times 3$ to $91 \times 2$) fed into the radiomics pipeline (restricting feature categories and MRI sequences) improved feature selection stability without compromising performance, we then analyze here a second step to further reduce the number of features that is essential for building a final radiomic model (Supplementary Text S1).

## 4. Discussion

In this study, we investigated the performance of radiomics based on conventional MRI sequences in discriminating between MSGTs and BSGTs. The aim was to determine if the diagnostic performance of MRI radiomic features would be reduced by restricting the initial number of radiomic features fed into the pipeline. We evaluated the impact of reducing features by restricting initial feature categories and MRI sequences.

We found that all radiomic categories for all three sequences, except for first-order features extracted from T1WI and CE-T1WI, could distinguish MSGTs from BSGTs. The diagnostic performance was not affected by restricting analysis to one feature category rather than combining all feature categories. We also showed that feature extraction could be reduced by confining extraction to only two of the three MRI sequences because the CE-T1WI did not outperform either the T1WI or FS-T2WI sequences (Table 2). The best overall performance was obtained by combining T1WI and FS-T2WI to produce an AUC of 0.846, with an accuracy of 0.761, a sensitivity of 0.740 and a specificity of 0.782 (Table 4).

The exponential features were the best categories for FS-T2WI (AUC of 0.819), and the logarithm features were best for T1WI (AUC of 0.828) and CE-T1WI (AUC of 0.754). Although feeding more features into the radiomics pipeline improved the radiomic model performance in the training set, a wider choice of initial features decreased the feature selection stability and model performance in the validation set. Specifically, the feature categories with a smaller number of features (shape and first-order, $n = 14$–18) showed similar results in both the training and validation sets; feature categories with an intermediate number of features (texture, exponential and logarithm, $n = 73$–91) showed a slight drop in performance in the validation sets; and feature categories with the largest number of features (wavelet and all categories combined, $n = 801$–1015) showed the largest drop in performance such that these categories no longer had the best performance (Table 2). This finding was also supported by the stability strength of the best performing feature category in all three MRI sequences. Specifically, reducing the initial input features to the best feature categories to build a radiomic model could also improve the stability strength (Table 3).

Our results indicate that including more radiomic features and categories in analysis might not produce a better model. It has been observed in practice that if the amount of training data used is small compared to the number of features, adding too many features degrades the metrics of the classifier [46]. High dimensionality and a small sample size increase the overfitting risk and decrease the classifiers accuracy, posing challenges to classification techniques [32,33]. Moreover, as classifiers rarely scale well to vast numbers of features, high dimensionality can lead to unreasonably long computation time [33]. Thus, analyzing thousands of radiomic features in a small group of samples usually restricts the use of radiomics in clinical practice [47,48]. Limiting the number of features in the initial pool could overcome some of these problems without compromising diagnostic accuracy.

The categorized features investigated in this study could provide more valuable and efficient potential markers for discriminating SGTs based on radiomic analysis. As there can be large overlaps in the shapes and sizes of malignant and benign tumors, it is not surprising that this category did not perform well. We cannot explain why the logarithm and exponential features performed best. However, studies of other MRI sequences, such as diffusion-weighted imaging (DWI) [49,50], have shown that malignant tumors may have features that fall between those of PA and WT which are the two most common benign tumors. It is possible, therefore, that logarithm and exponential, which are both filtered features, were able to change the distribution of features between these three groups so that there is a greater difference between BSGTs and MSGTs.

The study results indicate that numerous features included in radiomic packages are unnecessary and inefficient as initial input features to construct a radiomics model. Both previous studies and our results indicate shape and first-order features are not primary choices for radiomic analysis based on conventional MRI sequences to discriminate between MSGTs and BSGTs [20,51]. Moreover, although several studies show radiomics features

could discriminate SGTs [19,20,22,51] but there is little overlap of selected features and no consensus on a radiomic signature. Zheng YM [20] selected 15 wavelets, one texture and one first-order feature for the final radiomic signature to discriminate benign from malignant tumors. Not surprisingly, the wavelets then dominated the radiomic signature, given such a number of wavelet features (1488/1702) being fed into the radiomics procedure. Three other studies [19,22,51] which proposed a radiomic analysis also lacked consensus on the selection of features. At present, most discussion about radiomics variability has focused on image scanner instrumentation manufacturers, parameter settings, tumor segmentation, pre-processing methods and radiomic software packages [38,52–57], but little attention has been paid to the problem of variability caused by the abundance of initial features.

This study reduced dimensionality by more than 90% simply by using a single feature category to build radiomics models, and then demonstrated its the effectiveness and potential for radiomics research without reducing performance. Surprisingly, there is little data in the literature on the value of radiomics analysis of the T1WI contrast image. In radiomics, the discrimination of SGTs have mostly reported results from T1WI, T2WI and DWI [18–21]. Only three machine learning study [58–60], using deep learning instead of radiomics, evaluated CE-T1WI images of the parotid gland. In agreement with our study, one of them found that CE-T1WI did not improve the classification of SGTs [59]. It should be emphasized that while contrast enhanced images may not provide additional valuable radiomic features for SGT discrimination they are still a necessity for clinical evaluation of SGTs, especially for documenting extent of spread of MSGTs. The aim of this study was to take the first step in reducing the number of evaluated features in order to facilitate reaching a consensus on the radiomic signature. We have substantially decreased the number of features from $1015 \times 3$ to $91 \times 2$. However, we believe it will require a second step to further reduce the number of features before model building. We do not know the best method for this second step and believe it is an important area for future research, but we have proposed a at present have included a possible direction in the supplementary material that by using repetitive cross-validation to optimize the trade-off between feature stability and model performance, and reduces the features to four, results shown in the supplementary material (Supplementary Text S1). This method takes advantage of resampling to identify features that are not data-specific and, hence, have lower risk of overfitting.

This study has some limitations. First, the study lacks external validation. Nonetheless, the performance of radiomic analysis for discriminating MSGTs and BSGTs was verified by cross-validation. Second, as the focus of this study was conventional anatomical MRI imaging, we did not evaluate other MRI sequences such as DWI, which has yielded variable results and no consensus on the best radiomic signature [18,21]. Third, sampling bias may exist because of the small sample size in our study.

## 5. Conclusions

Reducing the number of features fed into the radiomics pipeline could help researchers take the first step towards achieving a consensus on a radiomic signature for discriminating between BSGTs and MSGTs. In this study we found feature selection stability was improved without compromising performance by restricting the number of initial MRI sequences analyzed to T1WI and FS-T2WI images (avoiding the contrast enhanced T1W images) and restricting the number of initial feature categories to one category per sequence (logarithm for T1WI, and exponential for FS-T2WI) rather than using all feature categories combined. We hope that future studies will evaluate the role of feature category and MRI sequence selection to determine the strongest candidates for improving feature selection stability to facilitate reaching a radiomic signature consensus.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cancers14235804/s1, Datasheet S1: The name list of 1015 3D quantitative features and six feature categories: extracted in this study; Datasheet S2: The volumetric dice similarity coefficient for the 30 randomly selected inter-observer segmentations on T1WI, FS-T2WI and CE-T1WI. Text S1: Additional analysis to further reduce radiomic features.

**Author Contributions:** Conceptualization, R.Z., A.D.K. and L.M.W.; Data curation, R.Z., Q.Y.H.A., C.G. and S.Q.; Formal analysis, R.Z. and A.C.V.; Funding acquisition, Q.Y.H.A. and A.D.K.; Investigation, R.Z., Q.Y.H.A. and L.M.W.; Methodology, R.Z., Q.Y.H.A., L.M.W. and A.D.K.; Project administration, A.D.K.; Resources, A.D.K.; Software, R.Z. and L.M.W.; Supervision, A.D.K.; Validation, R.Z., Q.Y.H.A., L.M.W., T.Y.S. and A.D.K.; Visualization, R.Z. and L.M.W.; Writing—original draft, R.Z. and Q.Y.H.A.; Writing—review and editing, R.Z., Q.Y.H.A., L.M.W., C.G., S.Q., T.Y.S., A.C.V. and A.D.K. All authors have read and agreed to the published version of the manuscript.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the local Institutional Review Board of The Joint Chinese University of Hong Kong—New Territories East Cluster Clinical Research Ethics Committee. (Approval No. CRE-2019.709).

**Informed Consent Statement:** Patient consent was waived by the local Institutional Review Board owing to the retrospective nature of this study.

**Data Availability Statement:** Data and clinical information presented in this study can be provided upon requests from the corresponding authors.

**Conflicts of Interest:** The authors declare no conflict of interest.

# References

1. Meyer, M.T.; Watermann, C.; Dreyer, T.; Ergun, S.; Karnati, S. 2021 Update on Diagnostic Markers and Translocation in Salivary Gland Tumors. *Int. J. Mol. Sci.* **2021**, *22*, 6771. [CrossRef]
2. Razek, A.A.K.A.; Mukherji, S.K. State-of-the-Art Imaging of Salivary Gland Tumors. *Neuroimaging Clin. N. Am.* **2018**, *28*, 303–307. [CrossRef]
3. Lobo, R.; Hawk, J.; Srinivasan, A. A Review of Salivary Gland Malignancies Common Histologic Types, Anatomic Considerations, and Imaging Strategies. *Neuroimaging Clin. N. Am.* **2018**, *28*, 171–182. [CrossRef]
4. Freling, N.; Crippa, F.; Maroldi, R. Staging and follow-up of high-grade malignant salivary gland tumours: The role of traditional versus functional imaging approaches—A review. *Oral Oncol.* **2016**, *60*, 157–166. [CrossRef]
5. Yousem, D.M.; Kraut, M.A.; Chalian, A.A. Major salivary gland imaging. *Radiology* **2000**, *216*, 19–29. [CrossRef]
6. Seethala, R.R.; Stenman, G. Update from the 4th Edition of the World Health Organization Classification of Head and Neck Tumours: Tumors of the Salivary Gland. *Head Neck Pathol.* **2017**, *11*, 55–67. [CrossRef]
7. Afzelius, P.; Nielsen, M.Y.; Ewertsen, C.; Bloch, K.P. Imaging of the major salivary glands. *Clin. Physiol. Funct. I* **2016**, *36*, 1–10. [CrossRef]
8. Schmidt, R.L.; Hall, B.J.; Wilson, A.R.; Layfield, L.J. A Systematic Review and Meta-Analysis of the Diagnostic Accuracy of Fine-Needle Aspiration Cytology for Parotid Gland Lesions. *Am. J. Clin. Pathol.* **2011**, *136*, 45–59. [CrossRef]
9. Zhang, R.; King, A.D.; Wong, L.M.; Bhatia, K.S.; Qamar, S.; Mo, F.K.; Vlantis, A.C.; Ai, Q.Y.H. Discriminating between benign and malignant salivary gland tumors using diffusion-weighted imaging and intravoxel incoherent motion at 3 Tesla. *Diagn. Interv. Imag.* **2022**. [CrossRef]
10. Tao, X.F.; Yang, G.X.; Wang, P.Z.; Wu, Y.W.; Zhu, W.J.; Shi, H.M.; Gong, X.; Gao, W.Q.; Yu, Q. The value of combining conventional, diffusion-weighted and dynamic contrast-enhanced MR imaging for the diagnosis of parotid gland tumours. *Dentomaxillofac. Radiol.* **2017**, *46*, 20160434. [CrossRef]
11. Takumi, K.; Nagano, H.; Kikuno, H.; Kumagae, Y.; Fukukura, Y.; Yoshiura, T. Differentiating malignant from benign salivary gland lesions: A multiparametric non-contrast MR imaging approach. *Sci. Rep.* **2021**, *11*, 2780. [CrossRef] [PubMed]
12. Zhang, Y.F.; Li, H.; Wang, X.M.; Cai, Y.F. Sonoelastography for differential diagnosis between malignant and benign parotid lesions: A meta-analysis. *Eur. Radiol.* **2019**, *29*, 725–735. [CrossRef]
13. Lee, Y.Y.P.; Wong, K.T.; King, A.D.; Ahuja, A.T. Imaging of salivary gland tumours. *Eur. J. Radiol.* **2008**, *66*, 419–436. [CrossRef]
14. Miao, L.Y.; Xue, H.; Ge, H.Y.; Wang, J.R.; Jia, J.W.; Cui, L.G. Differentiation of pleomorphic adenoma and Warthin's tumour of the salivary gland: Is long-to-short diameter ratio a useful parameter? *Clin. Radiol.* **2015**, *70*, 1212–1219. [CrossRef]
15. Gorovitz, S.; Macintyre, A. Toward a Theory of Medical Fallibility. *Hastings Cent. Rep.* **1975**, *5*, 13–23. [CrossRef]
16. Lambin, P.; Leijenaar, R.T.H.; Deist, T.M.; Peerlings, J.; de Jong, E.E.C.; van Timmeren, J.; Sanduleanu, S.; Larue, R.T.H.M.; Even, A.J.G.; Jochems, A.; et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 749–762. [CrossRef]
17. Zheng, Y.M.; Chen, J.; Xu, Q.; Zhao, W.H.; Wang, X.F.; Yuan, M.G.; Liu, Z.J.; Wu, Z.J.; Dong, C. Development and validation of an MRI-based radiomics nomogram for distinguishing Warthin's tumour from pleomorphic adenomas of the parotid gland. *Dentomaxillofac. Radiol.* **2021**, *50*, 20210023. [CrossRef]
18. Shao, S.; Zheng, N.; Mao, N.; Xue, X.; Cui, J.; Gao, P.; Wang, B. A triple-classification radiomics model for the differentiation of pleomorphic adenoma, Warthin tumour, and malignant salivary gland tumours on the basis of diffusion-weighted imaging. *Clin. Radiol.* **2021**, *76*, 472.e11–472.e18. [CrossRef]

19. Piludu, F.; Marzi, S.; Ravanelli, M.; Pellini, R.; Covello, R.; Terrenato, I.; Farina, D.; Campora, R.; Ferrazzoli, V.; Vidiri, A. MRI-Based Radiomics to Differentiate between Benign and Malignant Parotid Tumors With External Validation. *Front. Oncol.* **2021**, *11*, 656918. [CrossRef]

20. Zheng, Y.M.; Li, J.; Liu, S.; Cui, J.F.; Zhan, J.F.; Pang, J.; Zhou, R.Z.; Li, X.L.; Dong, C. MRI-Based radiomics nomogram for differentiation of benign and malignant lesions of the parotid gland. *Eur. Radiol.* **2020**, *31*, 4042–4052. [CrossRef]

21. Shao, S.; Mao, N.; Liu, W.J.; Cui, J.J.; Xue, X.L.; Cheng, J.F.; Zheng, N.; Wang, B. Epithelial salivary gland tumors: Utility of radiomics analysis based on diffusion-weighted imaging for differentiation of benign from malignant tumors. *J. X-ray Sci. Technol.* **2020**, *28*, 799–808. [CrossRef]

22. Gabelloni, M.; Faggioni, L.; Attanasio, S.; Vani, V.; Goddi, A.; Colantonio, S.; Germanese, D.; Caudai, C.; Bruschini, L.; Scarano, M.; et al. Can Magnetic Resonance Radiomics Analysis Discriminate Parotid Gland Tumors? A Pilot Study. *Diagnostics* **2020**, *10*, 900. [CrossRef]

23. Gunduz, E.; Alcin, O.F.; Kizilay, A.; Piazza, C. Radiomics and deep learning approach to the differential diagnosis of parotid gland tumors. *Curr. Opin. Otolaryngol.* **2022**, *30*, 107–113. [CrossRef]

24. van Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.C.; Pieper, S.; Aerts, H.J.W.L. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* **2017**, *77*, E104–E107. [CrossRef]

25. Altman, N.; Krzywinski, M. The curse(s) of dimensionality. *Nat. Methods* **2018**, *15*, 399–400. [CrossRef]

26. Wu, W.M.; Parmar, C.; Grossmann, P.; Quackenbush, J.; Lambin, P.; Bussink, J.; Mak, R.; Aerts, H.J.W.L. Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology. *Front. Oncol.* **2016**, *6*, 71. [CrossRef]

27. Pak, E.; Choi, K.S.; Choi, S.H.; Park, C.K.; Kim, T.M.; Park, S.H.; Lee, J.H.; Lee, S.T.; Hwang, I.; Yoo, R.E.; et al. Prediction of Prognosis in Glioblastoma Using Radiomics Features of Dynamic Contrast-Enhanced MRI. *Korean J. Radiol.* **2021**, *22*, 1514–1524. [CrossRef]

28. Gulgezen, G.; Cataltepe, Z.; Yu, L. Stable and Accurate Feature Selection. *Lect. Notes Artif. Int.* **2009**, *5781*, 455–468.

29. Nogueira, S.; Sechidis, K.; Brown, G. On the Stability of Feature Selection Algorithms. *J. Mach. Learn. Res.* **2018**, *18*, 1–54.

30. Khan, M.H.R.; Bhadra, A.; Howlader, T. Stability selection for lasso, ridge and elastic net implemented with AFT models. *Stat. Appl. Genet. Mol.* **2019**, *18*. [CrossRef]

31. Wong, L.M.; Ai, Q.Y.H.; Zhang, R.L.; Mo, F.; King, A.D. Radiomics for Discrimination between Early-Stage Nasopharyngeal Carcinoma and Benign Hyperplasia with Stable Feature Selection on MRI. *Cancers* **2022**, *14*, 3433. [CrossRef]

32. Krishnaiah, R.R.; Kanal, L.N. Dimensionality and Sample Size Considerations in Pattern Recognition Practice. In *Handbook of Statistics*; North-Holland: Amsterdam, The Netherlands, 1982; pp. 825–855.

33. Dernoncourt, D.; Hanczar, B.; Zucker, J.D. Analysis of feature selection stability on high dimension and small sample data. *Comput. Stat. Data Anal.* **2014**, *71*, 681–693. [CrossRef]

34. Yushkevich, P.A.; Piven, J.; Hazlett, H.C.; Smith, R.G.; Ho, S.; Gee, J.C.; Gerig, G. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage* **2006**, *31*, 1116–1128. [CrossRef] [PubMed]

35. Dine, L.R. Measures of the amount of ecologic association between species. *Ecology* **1945**, *26*, 196–205.

36. Duane, F.; Aznar, M.C.; Bartlett, F.; Cutter, D.J.; Darby, S.C.; Jagsi, R.; Lorenzen, E.L.; McArdle, O.; McGale, P.; Myerson, S.; et al. A cardiac contouring atlas for radiotherapy. *Radiother. Oncol.* **2017**, *122*, 416–422. [CrossRef] [PubMed]

37. Tustison, N.J.; Avants, B.B.; Cook, P.A.; Zheng, Y.J.; Egan, A.; Yushkevich, P.A.; Gee, J.C. N4ITK: Improved N3 Bias Correction. *IEEE Trans. Med. Imaging* **2010**, *29*, 1310–1320. [CrossRef] [PubMed]

38. Collewet, G.; Strzelecki, M.; Mariette, F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magn. Reson. Imaging* **2004**, *22*, 81–91. [CrossRef] [PubMed]

39. Isensee, F.; Jaeger, P.F.; Kohl, S.A.A.; Petersen, J.; Maier-Hein, K.H. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **2021**, *18*, 203–211. [CrossRef] [PubMed]

40. Yaniv, Z.; Lowekamp, B.C.; Johnson, H.J.; Beare, R. SimpleITK Image-Analysis Notebooks: A Collaborative Environment for Education and Reproducible Research. *J. Digit. Imaging* **2018**, *31*, 290–303. [CrossRef] [PubMed]

41. Nekooeimehr, I.; Lai-Yuen, S.K. Adaptive semi-unsupervised weighted oversampling (A-SUWO) for imbalanced datasets. *Expert Syst. Appl.* **2016**, *46*, 405–416. [CrossRef]

42. Blagus, R.; Lusa, L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinform.* **2013**, *14*, 106. [CrossRef]

43. Wilcoxin, F. Probability tables for individual comparisons by ranking methods. *Biometrics* **1947**, *3*, 119–122. [CrossRef]

44. Alhamzawi, R.; Ali, H.T.M. The Bayesian adaptive lasso regression. *Math. Biosci.* **2018**, *303*, 75–82. [CrossRef]

45. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [CrossRef]

46. Ramadan, S.Z. Methods Used in Computer-Aided Diagnosis for Breast Cancer Detection Using Mammograms: A Review. *J. Health Eng.* **2020**, *2020*, 9162464. [CrossRef]

47. Chalkidou, A.; O'Doherty, M.J.; Marsden, P.K. False Discovery Rates in PET and CT Studies with Texture Features: A Systematic Review. *PLoS ONE* **2015**, *10*, e0124165. [CrossRef]

48. Court, L.E.; Fave, X.; Mackin, D.; Lee, J.; Yang, J.Z.; Zhang, L.F. Computational resources for radiomics. *Transl. Cancer Res.* **2016**, *5*, 340–348. [CrossRef]

49. Sumi, M.; Nakamura, T. Head and neck tumours: Combined MRI assessment based on IVIM and TIC analyses for the differentiation of tumors of different histological types. *Eur. Radiol.* **2014**, *24*, 223–231. [CrossRef]

50. 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*, 770–777. [CrossRef]

51. Liu, Y.B.; Zheng, J.B.; Zhao, J.Z.; Yu, L.J.; Lu, X.P.; Zhu, Z.H.; Guo, C.L.; Zhang, T. Magnetic resonance image biomarkers improve differentiation of benign and malignant parotid tumors through diagnostic model analysis. *Oral Radiol.* **2021**, *37*, 658–668. [CrossRef]

52. Traverso, A.; Kazmierski, M.; Welch, M.L.; Weiss, J.; Fiset, S.; Foltz, W.D.; Gladwish, A.; Dekker, A.; Jaffray, D.; Wee, L.; et al. Sensitivity of radiomic features to inter-observer variability and image pre-processing in Apparent Diffusion Coefficient (ADC) maps of cervix cancer patients. *Radiother. Oncol.* **2020**, *143*, 88–94. [CrossRef] [PubMed]

53. 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*, 1143–1158. [CrossRef] [PubMed]

54. Moradmand, H.; Aghamir, S.M.R.; Ghaderi, R. Impact of image preprocessing methods on reproducibility of radiomic features in multimodal magnetic resonance imaging in glioblastoma. *J. Appl. Clin. Med. Phys.* **2020**, *21*, 179–190. [CrossRef] [PubMed]

55. Hoebel, K.V.; Patel, J.B.; Beers, A.L.; Chang, K.; Singh, P.; Brown, J.M.; Pinho, M.C.; Batchelor, T.T.; Gerstner, E.R.; Rosen, B.R.; et al. Radiomics Repeatability Pitfalls in a Scan-Rescan MRI Study of Glioblastoma. *Radiol. Artif. Intell.* **2021**, *3*, e190199. [CrossRef]

56. Korte, J.C.; Cardenas, C.; Hardcastle, N.; Kron, T.; Wang, J.; Bahig, H.; Elgohari, B.; Ger, R.; Court, L.; Fuller, C.D.; et al. Radiomics feature stability of open-source software evaluated on apparent diffusion coefficient maps in head and neck cancer. *Sci. Rep.* **2021**, *11*, 17633. [CrossRef] [PubMed]

57. McHugh, D.J.; Porta, N.; Little, R.A.; Cheung, S.; Watson, Y.; Parker, G.J.M.; Jayson, G.C.; O'Connor, J.P.B. Image Contrast, Image Pre-Processing, and T1 Mapping Affect MRI Radiomic Feature Repeatability in Patients with Colorectal Cancer Liver Metastases. *Cancers* **2021**, *13*, 240. [CrossRef] [PubMed]

58. Gunduz, E.; Alcin, O.F.; Kizilay, A.; Yildirim, I.O. Deep learning model developed by multiparametric MRI in differential diagnosis of parotid gland tumors. *Eur. Arch. Otorhinolaryngol.* **2022**, *279*, 5389–5399. [CrossRef]

59. Chang, Y.J.; Huang, T.Y.; Liu, Y.J.; Chung, H.W.; Juan, C.J. Classification of parotid gland tumors by using multimodal MRI and deep learning. *NMR Biomed.* **2021**, *34*, e4408. [CrossRef]

60. Liu, X.; Pan, Y.; Zhang, X.; Sha, Y.; Wang, S.; Li, H.; Liu, J. A Deep Learning Model for Classification of Parotid Neoplasms Based on Multimodal Magnetic Resonance Image Sequences. *Laryngoscope* **2022**. [CrossRef]