Topology-based radiomic features for prediction of parotid gland cancer malignancy grade in magnetic resonance images

幾嶋, 宏二郎

Topology-based radiomic features for prediction of parotid gland cancer malignancy grade in magnetic resonance images

Kojiro Ikushima1, 2, Hidetaka Arimura3, Ryuji Yasumatsu4, Hidemi Kamezawa5,Kenta Ninomiya6

1 Department of Health Sciences, Graduate School of Medical Sciences, Kyushu University, 3-1-1, Maidashi, Higashi-ku, Fukuoka 812-8582, Japan

2 Department of Radiological Technology, Yamaguchi University Hospital, 1-1-1 Minami-kogushi, Ube city, Yamaguchi 755-8505, Japan

3 Department of Health Sciences, Faculty of Medical Sciences, Kyushu University, 3-1-1, Maidashi, Higashi-ku, Fukuoka 812-8582, Japan

4 Department of Otorhinolaryngology-Head and Neck Surgery, Faculty of Medicine, Kindai University, 377-2, Onohigashi, Sayama, Osaka, 589-0014, Japan

5 Department of Radiological Technology, Faculty of Fukuoka Medical Technology, Teikyo University, 6-22 misaki-machi, Omuta, Fukuoka 836-8505, Japan

6 Sanford Burnham Prebys Medical Discovery Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

Corresponding author. Hidetaka Arimura, Ph.D.

Division of Quantum Radiation Science, Department of Health Sciences,

Faculty of Medical Sciences, Kyushu University

3-1-1, Maidashi, Higashi-ku, Fukuoka 812-8582, Japan

Tel&Fax: +81-92-642-6719

e-mail: arimura.hidetaka.616@m.kyushu-u.ac.jp

Running title: Topology-based radiomic features for prediction of malignancy grade

Abstract

Purpose: The malignancy grades of parotid gland cancer (PGC) have been assessed for decision of treatment policies. Therefore, we have investigated the feasibility of a topology-based radiomic features for prediction of parotid gland cancer (PGC) malignancy grade in magnetic resonance (MR) images.

Materials and methods: Two-dimensional T1- and T2-weighted MR images of 39 patients with PGC were selected for this study. Imaging properties of PGC can be quantified using the topology, which could be useful for assessing the number of holes or heterogeneity in PGC regions using invariants of the Betti numbers. Radiomic signatures were constructed from 41,472 features obtained after a harmonization using an elastic net model. PGC patients were stratified using a logistic classification into low/intermediate- and high-grade malignancy groups. The training data were increased by four times to avoid the overfitting problem using a synthetic minority oversampling technique. The proposed approach was assessed using a 4-fold cross-validation test.

Results: The highest accuracy of the proposed approach was 0.975 for the validation cases, whereas that of the conventional approach was 0.694.

Conclusion: This study indicated that topology-based radiomic features could be feasible for the noninvasive prediction of the malignancy grade of PGCs.

Keywords Radiomic features, Topology, Parotid gland cancer, Malignancy grade.

Introduction

Parotid gland cancer (PGC) is a rare form of cancer, accounting for approximately 5% of all head and neck cancers [1]. There are many different histological types of PGC, i.e., 23 histopathological types according to the 2017 classification of the World Health Organization [2]. Some tumors are inherently low or high grade, but others (e.g. mucoepidermoid carcinoma) could be difficult to diagnose. The malignancy grade of PGC can be classified into low/intermediate-grade patients and high-grade patients, and help physicians decide treatment options. For instance, the extent of local resection (e.g. whether the facial nerve can be preserved) could be determined depending on PGC grade [3]. In another example, radical surgery and/or postoperative irradiation can be often performed for high-grade PGC, whereas the most standard treatment for low/intermediate-grade PGC would be conservative surgery or minimally invasive surgery for facial nerve preservation [4]. Therefore, ascertaining the malignancy grade of PGC is required for deciding appropriate treatment procedures to be undertaken [3, 5-9].

The malignancy grade of PGC has been assessed using invasive fine-needle aspiration cytology (FNAC). However, the FNAC results could be affected by intra- and inter-observer variability among pathologists. Several studies show histological diagnostic accuracies of FNAC in PGC, which could have a wide range. The three studies on PGC reported the diagnostic accuracy by FNAC was 32% [3], 87.8% [10], and 90% [11]. Zbären *et al*. reported that the accuracy, sensitivity, and specificity of FNAC in detecting PGC tumors were 79%, 74%, and 88%, respectively [12]. Mallon *et al*. reported that the FNAC

had a sensitivity and specificity of 52% and 92%, respectively, in detecting PGC tumors [13]. Furthermore, FNAC may damage facial nerves. Therefore, quantitative and noninvasive approaches are preferable for assessing PGC malignancy.

Radiomics for quantifying the degree of tumor heterogeneity has been widely investigated for the evaluation of malignancy in various cancer [14-16]. Kamezawa *et al.* [17] investigated a noninvasive radiomics approach for predicting the PGC malignancy grade using six conventional machine learning and five deep learning (DL) algorithms. They obtained the highest accuracy of 0.854 using a VGG-16-based DL model. The PGC regions that showed hyperintensity on T1 and T2 weighted images were misclassified in their study. In histology, poorly differentiated cancer cells usually indicate a high-grade malignancy, whereas well-differentiated cancer cells indicate low/intermediate-grade malignancy. In PGC, well-differentiated cancer cells might acquire the characteristics of the parotid gland, such as the ducts. Ho *et al.* [18] reviewed the histological images of various types of parotid gland tumors, including benign (pleomorphic adenoma, Warthin's tumor) and malignant tumors (mucoepidermoid carcinoma, adenoid cystic carcinoma, carcinoma ex pleomorphic adenoma, and acinic cell carcinoma). Ducts in poorly-differentiated PGC tend to be less developed than those in parotid glands with benign tumors. Poorly-differentiated PGC tumors may be more heterogeneous than well-differentiated ones [19]. The ducts in the parotid gland are mathematically modeled as holes or heterogeneous pattern on two-dimensional MR slice images, whose imaging properties could be associated with PGC malignancy. Therefore, the imaging

properties of PGC can be quantified using the topology, which could be useful for assessing the number of holes or heterogeneity in PGC regions.

Computational topology based on Betti number has been employed as an image analysis tool, including the detection, classification, and prognostic evaluation of cancer lesions from pathological or computed tomography (CT) images [20-23]. Betti numbers are topologically invariant quantities that represent the k-dimensional holes of objects, which could quantify pixel value heterogeneity in cancerous regions without considering concrete shapes, and can reveal intrinsic morphological characteristics of cancer [21-23]. Topology-based radiomic features (hereafter referred to as topological features) are obtained as invariants of Betti numbers, which are zero- and one-dimensional holes. In other words, b0 is the number of connected components, and b1 is the number of holes (b1) in an object [23]. We have already reported that Betti-number-based features could have the higher predictability and repeatability on lung cancer [21] and head-and-neck cancer [24] than conventional radiomic features.

We hypothesize that the topological features based on Betti numbers in MR images might demonstrate better performance in predicting malignancy grade compared to the conventional radiomic approach in PGC. Therefore, this study aimed to investigate the feasibility of a topology-based radiomic features for predicting the PGC malignancy grade (low/intermediate-grade, and high-grade) in MR images.

Materials and methods

Clinical cases

This study was approved by the Institutional Review Board of our hospital. From 1983 to 2017, 118 patients with PGC, which were histologically confirmed from surgically resected specimens, were retrospectively selected. PGC patients who did not undergo FNAC (n=16) or T1- and T2 weighted preoperative MR imaging (n=60) were excluded from the 118 original cases. Finally, 42 patients were included in this study. A summary of patient characteristics is shown in Table 1. The 42 PGCs were histologically confirmed as high-grade (n=21; 50%), intermediate-grade (n=3; 7%), or low-grade (n=18; 43%) malignancy according to the 2005 WHO classification [5]. Moreover, the patients were classified into stage Ⅰ (n = 4; 9.5%), II (n = 13; 30.9%), III (n = 7; 16.7%), or IV (n = 18; 42.9%) according to the TNM classification of malignant tumors proposed by the Union for International Cancer Control [8]. The patients were scanned using five different MR scanners, which may allow us to select robust features of the image quality of the MR scanner. MR images were acquired using a 3.0-T MR scanner (Achieva, Philips Medical Systems, Andover, MA, USA, n=20), three 1.5-T MR scanners (Achieva, Philips Medical Systems, Andover, MA, USA, n=9; Symphony, Siemens Healthineers, Erlangen, Germany, n=11; Signa, General Electric Healthcare, Milwaukee, WI, USA, n=1), and a 0.2-T MR scanner (Magnetom Open Viva, Siemens Healthineers, Erlangen, Germany, n=1). The T1- and T2- weighted images were acquired using a spin-echo method with a repetition time (TR) of 400-659 and 1500-5783 ms and echo time (TE) of 8-14 and 80-250 ms,

127 respectively. The other scan parameters were as follows: slice thicknesses of 2-7 mm; matrix size of 256-
128 560; in-plane pixel size of 0.39-0.98 mm; and 16-bit gray levels. The PGC regions on each MR image were
129 contoured using a treatment planning system (Eclipse, ver.15.1.1, Varian Medical Systems, Palo Alto, CA,
130 USA). The regions of interest (ROI) for PGC regions on each MR image were determined based on a
131 consensus between the otorhinolaryngologist (R.Y.) and the medical physicist (H.K.). The
132 otortinolaryngologist and medical physicist have 24 years' experience and 14 years' experience,
133 respectively. Isotropic MR images were produced using cubic interpolation with an isopixel size of 0.45
134 mm, which is the most frequent pixel size of the original image. We performed three validation tests
135 including a 4-fold cross validation test, a hold-out validation test for the comparison with a previous study
136 [17], and an independent test. The three validation tests are explained below:
137 1) The proposed approach was assessed using a 4-fold cross validation test with the 39 PGC patients, who
138 underwent FNAC and had preoperative MR imaging.
139 2) A hold-out validation test was performed with the 39 PGC patients. They were divided into training
140 and test datasets, in which one was for development of the prediction model (training dataset: 70%),
141 and another one was for evaluation of the prediction performance (test dataset: 30%). There were no
142 statistically significant differences ($p > 0.05$) in patients' characteristics between the two groups [17]
143 based on Transparent Reporting of a multivariable prediction model for Individual Prognosis Or
144 Diagnosis (TRIPOD) statement [25]. Therefore, the hold-out test using non-randomly split approach

was used in this study.

3) In an independent test, the 39 PGC patients were employed for the development of the prediction model, and remaining 3 PGC patients, which did not undergo FNAC, were employed for the evaluation of the prediction performance.

Overview of topology-based radiomic features

The overall scheme of this study is illustrated in Figure 1. T1- and T2-weighted preoperative MR images of 42 patients with PGC were selected and preprocessed using a Laplacian of Gaussian (LoG) filter. The PGC regions were contoured on MR images by a medical physicist (H.K.) and confirmed by an otorhinolaryngologist (R.Y.). Topological and conventional features were computed from the cancer regions of the MR images. 41,472 topological features are obtained as invariants of Betti numbers, i.e., b0 and b1 [23], which could represent intrinsic morphological properties of the tumor heterogeneity with respect to connected components and holes [21-23]. On the other hand, 432 conventional features were calculated using wavelet decomposition filters. Radiomic signatures for topological and conventional features were independently constructed for each patient using an elastic net model. The training data were increased by four times to 112 cases to avoid the overfitting problem using a synthetic minority oversampling technique (SMOTE). The malignancy grade of PGCs was predicted using a logistic classification based on radiomic signatures.

163

164 **Preprocessing**

165 The LoG filter was applied to the isotropic MR images to suppress noise and enhance an edge in an image.

166 The LoG filters $LoG(x, y, z, \sigma)$ are as follows:

167 $$LoG(x, y, z, \sigma) = \frac{x^2+y^2+z^2-3\sigma^2}{2\pi\sigma^6} e^{-\frac{x^2+y^2+z^2}{2\sigma^2}},$$

168 where σ denotes the standard deviation of the Gaussian distribution. The LoG-filtered image is the

169 second partial derivative of the original images $f(x, y, z)$, and an edge-enhanced image $f_s(x, y, z)$ is

170 obtained by subtracting the LoG-filtered image from the original image $f(x, y, z)$.

171 $$f_s(x, y, z) = f(x, y, z) - \beta\{LoG(x, y, z, \sigma) * f(x, y, z)\},$$

172 where β is a constant value, which determines the degree of edge enhancement. In this study, the kernel

173 size of the LoG filter and σ in the Gaussian filter were 5×5×5 pixels and 3 pixels, respectively, and β

174 was set to 4. The pixel values of the MR images were re-quantized to 128 levels as a grayscale

175 normalization.

176

177 **Conventional features**

178 Conventional features were calculated using wavelet decomposition filters. Wavelet decomposition was

179 performed by applying either a low- or high-pass filter to the re-quantized MR images in each two-

180 dimensional image using the Coiflet1 mother wavelet. The radiomic features included histogram- and

texture-based features, i.e., gray level co-occurrence matrix (GLCM) [26], gray level run-length matrix (GLRLM) [27], gray level size-zone matrix (GLSZM) [28], and neighborhood gray-tone difference matrix (NGTDM) [29]. The 54 radiomic features are listed in the Supplementary information (Table S1). The 14 histogram- and 40 texture- based features were calculated from each of four wavelet-decomposed images per MR imaging sequence to obtain 432 radiomic features (54 radiomic features × 4 wavelet images × 2 MR imaging sequences) in total for each patient.

Topological features

Figures 2 and 3 show the calculation flows of topological features and b0 and b1 Betti maps from a T1 weighted MR images. The 41,472 topological features (54 radiomic features × 3 Betti maps × 128 threshold levels × 2 MR imaging sequences) were derived from 128 Betti maps for each patient. These Betti maps were produced from binary MR images by computing Betti numbers within kernel sizes. The kernel size and its shifting pixel were optimized for the ranges of 5 to 11 and 1 to 5, respectively [21]. The binary images were obtained by thresholding the re-quantized MR images with the values of 0 to 127. The b0 values were calculated by counting connected components according to the eight-connectivity within kernels. The b1 values were calculated by counting connected components in inversed binary images with the four-connectivity, with excluding the connected components that were attached to the edges of ROI as shown in Fig. 3. The ratio of b1 to b0 (b1/b0), which express the number of holes per connected component,

was also calculated as well. Betti numbers were calculated using in-house software developed in MATLAB 2017b.

Harmonization

A ComBat harmonization algorithm [30-32], which is effective in reducing effects induced by differences in acquisition protocol and scanner, was applied for the radiomic features. Therefore, we performed all the calculations with the ComBat harmonization algorithm in this study. Each value Y_{ijf} for a feature f of a patient j from a scanner i can be represented in a location and scale (L/S) model as

$$Y_{ijf} = \alpha_f + X\beta_f + \gamma_{if} + \delta_{if}\varepsilon_{ijf},$$

where α_f is the mean of the feature f, X is the design matrix for the covariates of interest, and β_f is the vector of regression coefficients for the feature f corresponding to X. The error terms ε_{ijf} can be assumed to follow a normal distribution with expected values of a mean of zero and variance of ${\delta_{if}}^2$. The γ_{if} and δ_{if} indicate the additive and multiplicative batch effects corresponding to the scanner i for the feature f, respectively. The harmonized value Y_{ijf}^{ComBat} is defined as

$$Y_{ijf}^{ComBat} = \frac{Y_{ijf} - \hat{\alpha}_f - X\hat{\beta}_f - \hat{\gamma}_{if}}{\hat{\delta}_{if}} + \hat{\alpha}_f + X\hat{\beta}_f,$$

where $\hat{\alpha}_f$, $\hat{\beta}_f$, $\hat{\gamma}_{if}$ and $\hat{\delta}_{if}$ are the estimators of the parameters α_f, β_f, γ_{if}, and δ_{if}. An empirical Bayes used for the estimation of the parameters $\hat{\gamma}_{if}$ and $\hat{\delta}_{if}$.

Construction of radiomic signatures

Two types of radiomic signatures were independently constructed from 432 conventional features and 41,472 topological features using an elastic net model in the "glmnet" package of R software version 4.0.1. The elastic net model includes the L1 and L2 norm penalties. In the elastic net model, the blending parameter α, which adjusts the effect of the least absolute shrinkage and selection operator penalty and ridge-regression penalties, was optimized by changing from 0.1 to 1. The radiomic signatures consisted of frequent features, which had the largest sum of absolute coefficients that were obtained based on Cox-net algorithm by the 4-fold cross-validation in 100 iterations [21, 33-35]. The Cox-net algorithm selected significant features with minimum cross-validation errors between reference PGC malignancy grades and predicted ones at each time, because of the random variables for splitting cases into 4 datasets for the 4-fold cross-validation test. Therefore, the frequent features were to be feasible for prediction of PGC malignancy grades. The number of features by each approach was optimized by changing the number of features from 2 to 6 to obtain the highest area under the receiver operating characteristic curve (AUC) in the training step. That was because we have a rule of the maximum number of features, which should be smaller than or equal to one-tenth of the number of cases (i.e., maximum number of features: 60/10=6) [36, 37].

Data augmentation

Figure 4 shows the illustration of the 4-fold cross-validation test with SMOTE. A whole data was divided into 4 sub-datasets of A, B, C, and D in this figure. The proposed approach was assessed using the 4-fold cross-validation test. The 29 or 30 training cases (3/4 folds) were increased to 112 or 120 cases to avoid an overfitting problem using SMOTE. SMOTE is an oversampling method that creates synthetic data based on feature space similarities between existing data [38]. The synthetic data vector, $\boldsymbol{x}_{new}$, is expressed as follows:

$$\boldsymbol{x}_{new} = \boldsymbol{x}_i + (\widehat{\boldsymbol{x}}_i - \boldsymbol{x}_i)\delta,$$

where $\boldsymbol{x}_i$ is the minority instance vector, $\widehat{\boldsymbol{x}}_i$ is one of the k-nearest neighbors for x_i, and $\delta \in [0, 1]$ is a random number.

Fourteen low/intermediate-grade cases and 15 high-grade cases (n=15) in 29 training cases were increased by four times to obtain 56 and 60 cases, respectively. To make balanced datasets with the same number of 112 training cases, four synthetic cases were randomly excluded from the 60 low/intermediate-grade and 60 high-grade cases.

Prediction of malignancy grade

Logistic classification (LC) was employed as a predictive model for PGC malignancy grade. LC can predict the probability of a binary event occurring by using a logistic sigmoid function [39].

Evaluation

The performance of our approach was evaluated using accuracy, the AUC, sensitivity, and specificity, and the topology-based approach was compared with the conventional (wavelet-based) approach. In addition, the robustness index (RI) was calculated to assess the robustness of models between the training and validation data [40]. The RI is defined as

$$\mathrm{RI} = \frac{\mathrm{AUC}_{\mathrm{train}} + \mathrm{AUC}_{\mathrm{valid}}}{1 + |\mathrm{AUC}_{\mathrm{train}} - \mathrm{AUC}_{\mathrm{valid}}|},$$

where $\mathrm{AUC}_{\mathrm{train}}$ and $\mathrm{AUC}_{\mathrm{valid}}$ represent the AUCs for the prediction of the PGC malignancy grade in the training and validation procedures, respectively.

Results

The parameters in the Betti numbers were a kernel size of 11 and a shifting pixel of 5. Five significant features, namely, b1_T1_GLCM_sum average_6, b0_T1_GLCM_sum average_26, b0_T2_NGTDM_coarseness_2, b0_T2_Histogram_Variance_55, and b0_T1_GLSZM_ Small zone emphasis (SZE)_52, were chosen for the proposed approach. The last number in each feature name represents the threshold value of the MR images for calculating Betti maps. The value α in the elastic net model was set at 1 for the topology-based approach and conventional approach to allow the accuracy to reach a maximum value. Figure 5 shows b0 and b1 Betti maps for 6 patients with high- and low/intermediate-grade PGC. The high-grade PGC seemed to have more holes within the PGC regions than the

low/intermediate-grade PGC. Figure 6 shows scatter plots of the five topological features for the 112 training cases selected using the elastic net.

Figure 7 shows scatter plots of the radiomic features obtained from the conventional and topology-based approaches. The features in this figure had the maximum absolute coefficients in the elastic net among the features with nonzero coefficients. The topological features of low/intermediate-grade and high-grade PGCs seemed to be more distinguishable than features of the conventional approach.

A summary of the results is presented in Table 2. The topology-based approach with SMOTE achieved an accuracy of 0.975±0.050, AUC of 0.997, a sensitivity of 0.95±0.100, specificity of 1.00, and RI of 1.99, whereas the conventional approach with SMOTE resulted in an accuracy of 0.694±0.072, AUC of 0.737, sensitivity of 0.65±0.100, specificity of 0.738±0.094, and RI of 1.40. On the other hand, the topology-based approach without SMOTE reached an accuracy of 0.975±0.050, AUC of 0.997, sensitivity of 0.95±0.100, specificity of 1.00, and RI of 1.99; while the conventional approach without SMOTE resulted in an accuracy of 0.692±0.083, AUC of 0.792, sensitivity of 0.60±0.163, specificity of 0.80±0.163, and RI of 1.55. Consequently, the topology-based approach may not need the data augmentation.

The topology-based approach in the hold-out validation test produced an accuracy of 0.917, AUC of 0.972, a sensitivity of 1.00, specificity of 0.833, and RI of 1.92. Furthermore, the topology-based approach in the independent test showed an accuracy of 1.00, a sensitivity of 1.00, and specificity of 1.00.

Discussion

In this study, the topology-based approach showed a better performance compared with the conventional approach. The cytological diagnostic accuracy of FNAC in the 39 patients with PGC employed in this study was 0.795, which may have included intra- and inter-observer variations. On the other hand, the proposed approach achieved an accuracy of 0.975, which may be feasible for stratifying patients with PGC into high-, low/intermediate-grade malignancies for treatment decision-making.

The Betti number indicates heterogeneity in the cancerous region. The topology-based approach without the LoG filter resulted in an accuracy of 0.884, AUC of 0.900, sensitivity of 0.85, specificity of 0.95, and an RI of 1.73 using the same parameter as Betti maps. Image noise can affect cancer heterogeneity, and heterogeneity may be evaluated accurately by suppressing the image noise. The LoG filter reasonably enhances the heterogeneity with noise suppression.

The b1 maps represent the number of holes in the cancer region, and the GLCM_sum average represents the relationship between the occurrences of pairs with lower and higher intensity values. The high-grade PGC had a number of holes with various heterogeneous sizes in the cancer region, while the low/intermediate-grade PGC had a large number of homogeneous small holes. The b0 maps mean the number of connected components, and the NGTDM represents the difference between the gray value and the average gray value of its neighbors. The low/intermediate-grade PGC had large connected components within the PGC regions. The histograms of high-grade PGC show the high variability compared with

low/intermediate-grade PGC. GLSZM represents the number of connected pixels that share the same gray level intensity, and SZN represents the variability of size zone volumes. Low/intermediate-grade PGC indicated a low value of GLSZM_SZN compared with high-grade PGC. High-grade PGC might be more heterogeneous than low/intermediate-grade PGC on MR images.

The deep learning (DL) method was already reported in our previous study [17], which a co-author [H.K.] reported using the same cases in this study. Therefore, we compared between the proposed approach and previous study. A previous study [17] investigated a non-invasive approach for predicting PGC malignancy grade based on radiomic biomarkers in preoperative MR images. They obtained the highest accuracy of 0.854 using a VGG-16-based DL model in the hold-out validation test. The PGC regions that showed hyperintensity on T1 and T2 weighted images were misclassified in their study. On the other hand, the proposed approach achieved the accuracy of 0.917 in the hold-out validation test, and correctly classified the case. The prediction accuracy of the proposed approach was improved so that we used the topological features which were independent of the image contrast. However, the proposed approach misclassified a case with a low grade PGC region that showed low hyperintensity on a T1 weighted image but mixed intensity on a T2 weighted image due to necrosis.

This study has three limitations. First, T1- and T2 MR images were used. Many types of MR images, such as fat suppression imaging (short TI inversion recovery, Dixon method), diffusion-weighted imaging, apparent diffusion coefficient imaging, and dynamic study data, could be useful for radiomic

research. For example, Martin *et al.* and Nemeth *et al.* reported a breast MRI radiomics study using multiparametric MR images [41, 42]. Most cases had no DWI and DCE-MRI images in our database. To ensure the number of cases, the case without the DWI and DCE-MRI images were excluded. As mentioned above, the DWI and DCE-MRI images can be useful for improvement of prediction accuracy if they can be used. Second, interobserver variability was included in the contouring cancer region. The third limitation was the number of cases used in this study. The accuracy of machine-learning techniques depends on the amount of training data. To avoid the overfitting problem due to the less sample size, the SMOTE technique was applied to our proposed approach. However, the proposed approach should be applied to various PGC cases in order to improve the accuracy and robustness of prediction model. In addition, we should apply the prediction model to independent large datasets from different institutions to improve its generalization. Therefore, it is necessary to increase the number of PGC cases to reliably predict the malignancy.

Conclusions

The predictive performance for the malignancy grade of PGC using a topology-based approach had higher accuracy than that of the conventional approach. This study indicated that the topology-based radiomic features could be used for the noninvasive prediction of the malignancy grade of PGCs.

Acknowledgements

This work was partially supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number JP17K15808. The authors would like to express their gratitude to all members of the Arimura Laboratory (http://www.shs.kyushu-u.ac.jp/~arimura) for their valuable comments for this study.

Author contribution

Ikushima: study conception and design, acquisition of data, analysis and interpretation of data, drafting of manuscript, critical revision; Arimura: study conception and design, acquisition of data, analysis and interpretation of data, critical revision; Yasumatsu: acquisition of data, analysis and interpretation of data, critical revision; Kamezawa: study conception and design, acquisition of data, analysis and interpretation of data, critical revision; Ninomiya: study conception and design, acquisition of data, analysis and interpretation of data, critical revision.

References

1. Chang JW, Hong HJ, Ban MJ, Shin YS, Kim WS, Koh YW, et al. (2015) Prognostic factors and treatment outcomes of parotid gland cancer: a 10-year single-center experience. Otolaryngol Head Neck Surg 153(6), 981-989.
2. El-Naggar AK, Chan JK, Grandis JR et al (2017) World Health Organization classification of tumours. Chapter 7, tumours of salivary glands. IARC press, Lyon, pp 160–202

3. Nishikawa S, Kawata R, Higashino M, Lee K, Terada T, Kurisu Y. et al. (2015) Assessing the histological type and grade of primary parotid carcinoma by fine-needle aspiration and frozen section. Auris Nasus Larynx 42(6):463-8.

4. Omura S, Kawata R, Higashino M, Nishikawa S, Terada T, Haginomori SI, Kurisu Y, Hirose Y. (2020) Challenges with preoperative diagnosis of low/intermediate-grade carcinoma of the parotid gland: single-center study of 112 patients. Eur Arch Otorhinolaryngol. Jul;277(7):2031-2039. doi: 10.1007/s00405-020-05871-6. Epub 2020 Mar 12. PMID: 32166415.

5. Leon B, John WE, Peter R, David S (2005) World Health Organization Classification of Tumours, Pathology and Genetics of Head and Neck Tumours, IARC Press, Lyon, pp 209-281.

6. Godballe C, Schultz JH, Krogdahl A, Møller-Grøntved A, Johansen J (2003) Parotid carcinoma: impact of clinical factors on prognosis in a histologically revised series. Laryngoscope 113(8):1411-7.

7. Lima RA, Tavares MR, Dias FL, Kligerman J, Nascimento MF, Barbosa MM. et al. Clinical prognostic factors in malignant parotid gland tumors. Otolaryngol Head Neck Surg. 2005;133(5):702-8.

8. Walvekar RR, Filho PAA, Seethala RR, Gooding WE, Heron DE, Johnson JT. et al. Clinicopathologic features as stronger prognostic factors than histology or grade in risk stratification of primary parotid malignancies. Head Neck. 2011;33(2):225-31.

9. Brierley J, Gospodarowicz MK, Wittekind C. (2017) Union for International Cancer Control (UICC) TNM classification of malignant tumours. Wiley Blackwell, Chichester, pp 241.

10. Eytan DF, Yin LX, Maleki Z, Koch WM, Tufano RP, Eisele DW. et al. (2018). Utility of preoperative fine needle aspiration in parotid lesions. Laryngoscope 128(2):398-402.

11. Seethala RR, LiVolsi VA, Baloch ZW (2005) Relative accuracy of fine-needle aspiration and frozen section in the diagnosis of lesions of the parotid gland. Head Neck 27(3), 217-223.

12. Zbären P, Guélat D, Loosli H, Stauffer E (2008) Parotid tumors: fine-needle aspiration and/or frozen section. Otolaryngol Head Neck Surg 139(6):811-5.

13. Mallon DH, Kostalas M, MacPherson FJ, Parmar A, Drysdale A, Chisholm E. et al (2013) The diagnostic value of fine needle aspiration in parotid lumps. Ann R Coll Surg Engl 95(4):258-62.

14. Zhou Z, Qian X, Hu J, Ma X, Zhou S, Dai Y, Zhu J (2021) CT-based peritumoral radiomics signatures for malignancy grading of clear cell renal cell carcinoma. Abdom Radiol, 46(6), 2690-2698.

15. Gong L, Xu M, Fang M, Zou J, Yang S, Yu X, Tian J (2020). Noninvasive prediction of high-grade prostate cancer via biparametric MRI radiomics. J Magn Reson Imaging 52(4), 1102-1109.

16. Li S, Liu J, Xiong Y, Pang P, Lei P, Zou H, Luo P (2021). A radiomics approach for automated diagnosis of ovarian neoplasm malignancy in computed tomography. Sci Rep 11(1), 1-9.

17. Kamezawa, H, Arimura, H, Yasumatsu, R, Ninomiya, K, Haseai S (2020) Preoperative and Non-Invasive Approach for Radiomic Biomarker-Based Prediction of Malignancy Grades in Patients with Parotid Gland Cancer in Magnetic Resonance Images. Medical Imaging and Information Sciences 37(4), 66-74.

18. Ho K, Lin H, Ann D K, Chu P G, Yen Y (2011) An overview of the rare parotid gland cancer. Head Neck Oncol 3(1), 1-7.

19. Lee Y. Y. P., Wong K. T., King A. D., Ahuja A. T. (2008). Imaging of salivary gland tumours. European Journal of Radiology, 66(3), 419-436.

20. Nakane K, Tsuchihashi Y, Matsuura N (2013) A simple mathematical model utilizing topological invariants for automatic detection of tumor areas in digital tissue images. In Diagnostic pathology (Vol. 8, No. 1, pp. 1-4). BioMed Central.

21. Ninomiya K, Arimura H (2020) Homological radiomics analysis for prognostic prediction in lung cancer patients. Phys Med 69, 90-100.

22. Adcock A, Rubin D, Carlsson G. (2014) Classification of hepatic lesions using the matching metric. Comput Vis Image Underst 121:36–42. https://doi.org/10.1016/j.cviu.2013.10.014.

23. Nakane K, Takiyama A, Mori S, Matsuura N. (2015) Homology-based method for detecting regions of interest in colonic digital images. Diagn Pathol 10:1–5. https://doi.org/10.1186/s13000-015-0244-x.

24. Le QC, Arimura H, Ninomiya K, Kodama T, Moriyama T. (2022) Can Persistent Homology Features Capture More Intrinsic Information about Tumors from ^{18}F-Fluorodeoxyglucose Positron Emission Tomography/Computed Tomography Images of Head and Neck Cancer Patients? Metabolites. 12(10):972. doi: 10.3390/metabo12100972.

25. Moons KG, Altman DG, Reitsma JB, et al. (2015) Transparent Reporting of a multivariable prediction

model for Individual Prognosis or Diagnosis (TRIPOD) : explanation and elaboration, Ann Intern Med, 162(1), W 1-73.

26. Haralick RM, Dinstein I, Shanmugam K (1973) Textural features for image classification. IEEE Trans Syst Man Cybern https://doi.org/10.1109/TSMC.1973.4309314.

27. Galloway MM. (1975) Texture analysis using gray level run lengths. Comput Graph Image Process 4:172–9. https://doi.org/10.1016/S0146-664X(75)80008-6.

28. Thibault G, Fertil B, Navarro C, Pereira S (2009) Texture indexes and gray level size zone matrix application to cell nuclei classification. Pattern Recognit Inf Process pp.140-145

29. Amadasun M, King R (1989) Textural features corresponding to textural properties. IEEE Trans Syst Man Cybern 19:1264–74. https://doi.org/10.1109/21.44046.

30. Johnson W. E., Li C. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics, 8(1), 118-127.

31. Fortin J. P, Parker D, Tunç B, Watanabe T, Elliott M. A, Ruparel K, Shinohara R. T (2017) Harmonization of multi-site diffusion tensor imaging data. Neuroimage, 161, 149-170.

32. Fortin J. P, Cullen N, Sheline Y. I, Taylor W. D, Aselcioglu I, Cook P. A, Shinohara R. T (2018) Harmonization of cortical thickness measurements across scanners and sites. Neuroimage, 167, 104-120.

33. Friedman J, Hastie T, Tibshirani R. (2010) Regularization paths for generalized linear models via

433 coordinate descent. J Stat Softw 33:1–22. https://doi.org/10.1016/j.expneurol.2008.01.011.

434 34. Simon N, Friedman J, Hastie T, Tibshirani R. (2011) Regularization paths for Cox's proportional

435 hazards model via coordinate descent. J Stat Softw ;39:1–13. https://doi.org/10.18637/jss.v039.i05.

436 35. Sun H, Lin W, Feng R, Li H. (2014) Network-regularized high-dimensional Cox regression for analysis

437 of genomic data. Stat Sin ;24:1433–59. https://doi.org/10.5705/ss.2012.317.

438 36. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. (1996) A simulation study of the number

439 of events per variable in logistic regression analysis. J Clin Epidemiol. 49:1373–1379.

440 37. Arimura H, Soufi M, Ninomiya K, Kamezawa H, Yamada M. (2018) Potentials of radiomics for cancer

441 diagnosis and treatment in comparison with computer-aided diagnosis. Radiol Phys Technol. 11:365–

442 374.

443 38. He H, Garcia EA (2009) Learning from Imbalanced Data. IEEE Trans Knowl Data Eng 21: (9), 1263-

444 1284

445 39. Harrel FE (2001) Regression Modeling Strategies With Applications to Linear Models, Logistic

446 Regression, and Survival Analysis. Springer, New York

447 40. Ninomiya K, Arimura H, Yoshitake T, Hirose T, Shioyama Y (2022) Synergistic combination of a

448 topologically invariant imaging signature and a biomarker for the accurate prediction of symptomatic

449 radiation pneumonitis before stereotactic ablative radiotherapy for lung cancer: A retrospective

450 analysis. PloS one, 17(1), e0263292.

41. Saint Martin M. J., Orlhac F, Akl P, Khalid F, Nioche C, Buvat I, Frouin F (2021) A radiomics pipeline dedicated to Breast MRI: validation on a multi-scanner phantom study. Magn Reson Mater Phy 34(3); 355-366.

42. Nemeth A, Chaudet P, Leporq B, Heudel P. E, Barabas F, Tredan O, Beuf O (2021) Multicontrast MRI-based radiomics for the prediction of pathological complete response to neoadjuvant chemotherapy in patients with early triple negative breast cancer. Magn Reson Mater Phy 34(6);833-844.

Figure captions

Figure 1 An overall scheme of the proposed approach. The ROI for a PGC region are shown by green lines. MR, magnetic resonance; T1WI, T1-weighted image; T2WI, T2-weighted image.

Figure 2 A Calculation flow of topological features. The ROI for a PGC region are shown by green lines.

Figure 3 A calculation flow of b0 and b1 Betti maps from a T1 weighted MR images. The ROI for a PGC region are shown by green lines.

Figure 4 An illustration of 4-fold cross-validation test with SMOTE. A whole data was divided into 4 sub-datasets of A, B, C, D in this figure. High: High-grade cases; Low/int: Low/intermediate-grade cases.

Figure 5 b0 and b1 Betti maps from T1 and T2 weighted MR images for 6 patients with high- and low/intermediate-grade PGC.

Figure 6 Scatter plots of topological features for 112 training cases selected by using the elastic net: (a) b1_T1_GLCM_SumAverage_26 and b0_T1_GLCM_SumAverage_6, (b) b1_T1_GLCM_SumAverage_26 and b0_T1_NGTDM_Coarseness_2, (c) b1_T1_GLCM_SumAverage_26 and b0_T2_Hist_Variance_55, and (d) b1_T1_GLCM_SumAverage_26 and b0_T1_GLSZM_SZE_52.

Figure 7 Scatter plots of radiomic features for 112 training cases obtained from conventional and proposed approaches: (a) topological features and (b)conventional features.

Table 1. Summary of patient characteristics.

Characteristics		No. of cases
Number		42
Sex	Male	27
	Female	15
Age (years)	Range	14 – 82
	Mean	56.0
Histological malignancy	Low	18 (43%)
	Intermediate	3 (7%)
	High	21 (50%)
Clinical stage	I	4 (9.5%)
	II	13 (30.9%)
	III	7 (16.7%)
	IV	18 (42.9%)

485 Table 2. Comparison of prediction performances in training and validation datasets for the 4-fold
486 cross-validation test

Method	Dataset	Accuracy	AUC	Sensitivity	Specificity	Robustness index
Topology w/ SMOTE	Train (n=29 or 30)	1.00	1.00	1.00	1.00	1.99
	Validation	0.975±0.050	0.997	0.95±0.100	1.00	
Conventional w/ SMOTE	Train (n=112)	0.853±0.019	0.919	0.862±0.111	0844±0.087	1.40
	Validation	0.694±0.072	0.737	0.650±0.100	0.738±0.094	
Topology w/o SMOTE	Train (n=29 or 30)	1.00	1.00	1.00	1.00	1.99
	validation	0.975±0.050	0.997	0.95±0.100	1.00	
Conventional w/o SMOTE	Train (n=112)	0.692±0.045	0.858	0.692±0.033	0.911±0.107	1.55
	Validation	0.692±0.083	0.792	0.60±0.163	0.80±0.163	

487

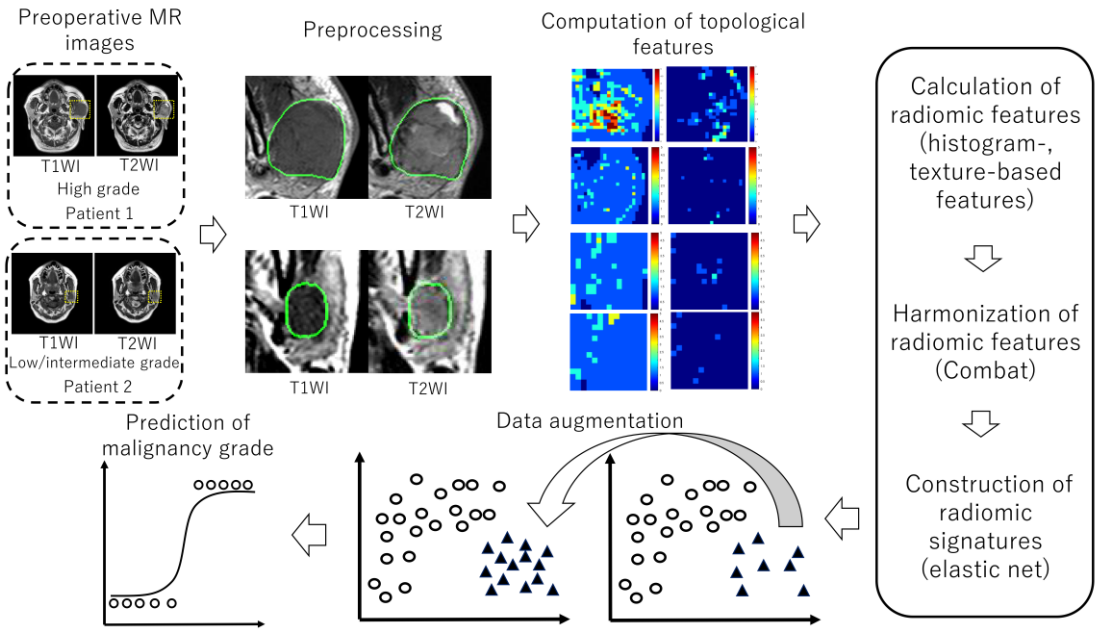
Preoperative MR images
T1WI
T2WI
High grade
Patient 1
T1WI
T2WI
Low/intermediate grade
Patient 2
Preprocessing
T1WI
T2WI
T1WI
T2WI
Computation of topological features
Calculation of radiomic features (histogram-, texture-based features)
Harmonization of radiomic features (Combat)
Construction of radiomic signatures (elastic net)
Data augmentation
Prediction of malignancy grade

Figure 1

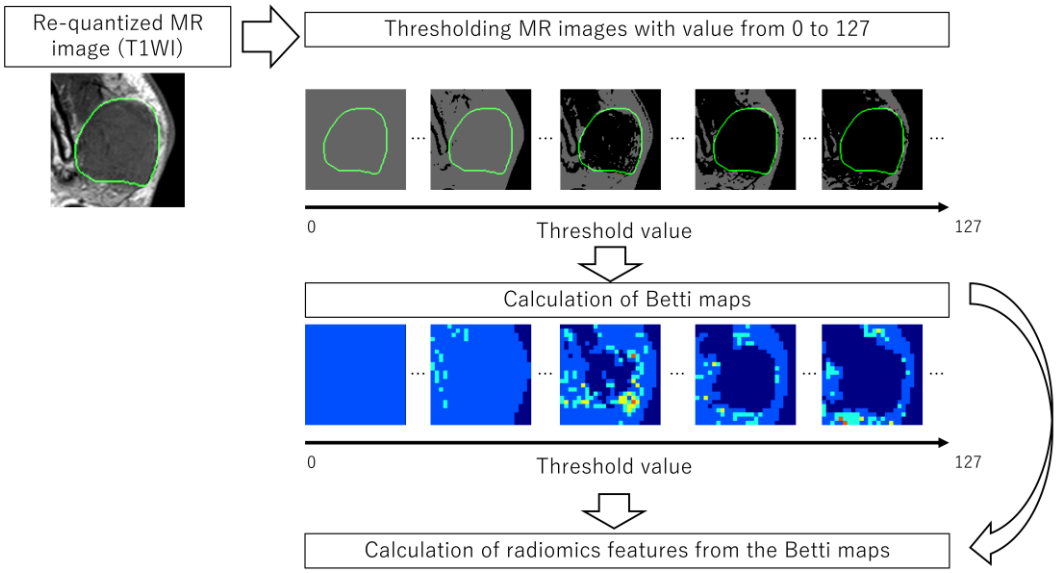
Re-quantized MR image (T1WI)
Thresholding MR images with value from 0 to 127
0
Threshold value
127
Calculation of Betti maps
0
Threshold value
127
Calculation of radiomics features from the Betti maps

Figure 2

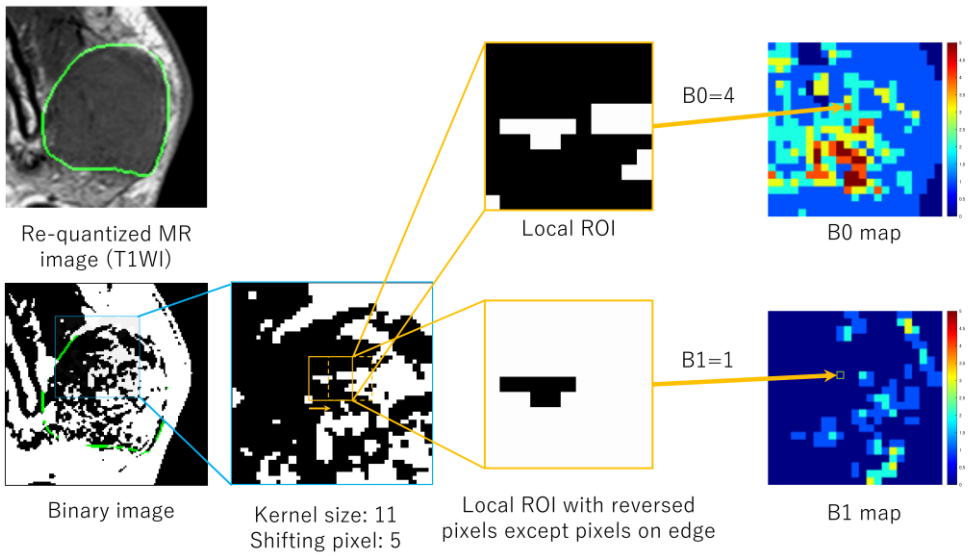
Re-quantized MR image (T1WI)
Binary image
Kernel size: 11
Shifting pixel: 5
Local ROI
Local ROI with reversed pixels except pixels on edge
B0=4
B1=1
B0 map
B1 map

Figure 3

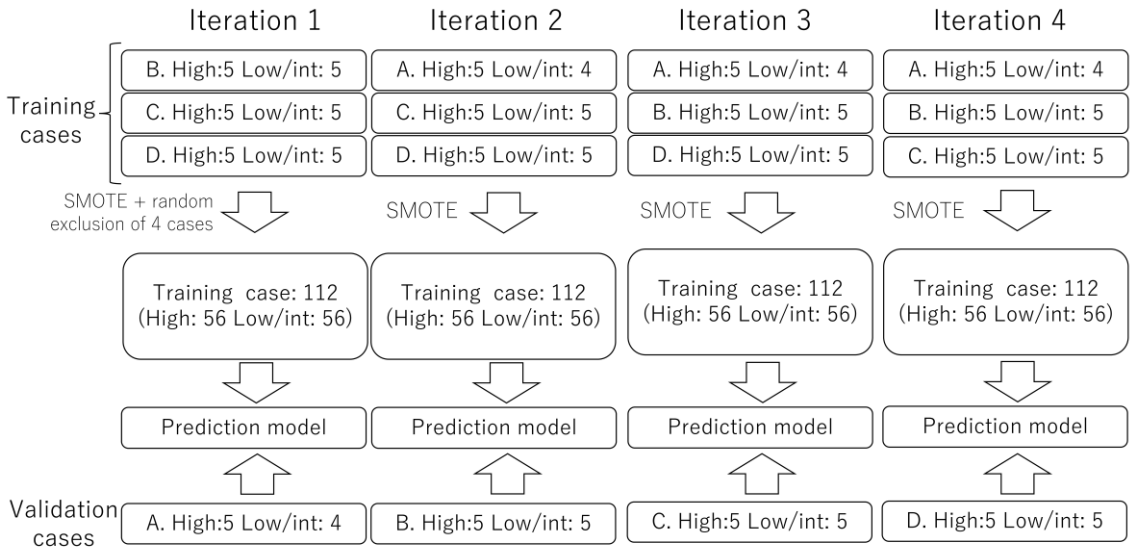
Iteration 1
Iteration 2
Iteration 3
Iteration 4
Training cases
B. High:5 Low/int: 5
C. High:5 Low/int: 5
D. High:5 Low/int: 5
A. High:5 Low/int: 4
C. High:5 Low/int: 5
D. High:5 Low/int: 5
A. High:5 Low/int: 4
B. High:5 Low/int: 5
D. High:5 Low/int: 5
A. High:5 Low/int: 4
B. High:5 Low/int: 5
C. High:5 Low/int: 5
SMOTE + random exclusion of 4 cases
SMOTE
SMOTE
SMOTE
Training case: 112 (High: 56 Low/int: 56)
Training case: 112 (High: 56 Low/int: 56)
Training case: 112 (High: 56 Low/int: 56)
Training case: 112 (High: 56 Low/int: 56)
Prediction model
Prediction model
Prediction model
Prediction model
Validation cases
A. High:5 Low/int: 4
B. High:5 Low/int: 5
C. High:5 Low/int: 5
D. High:5 Low/int: 5

Figure 4

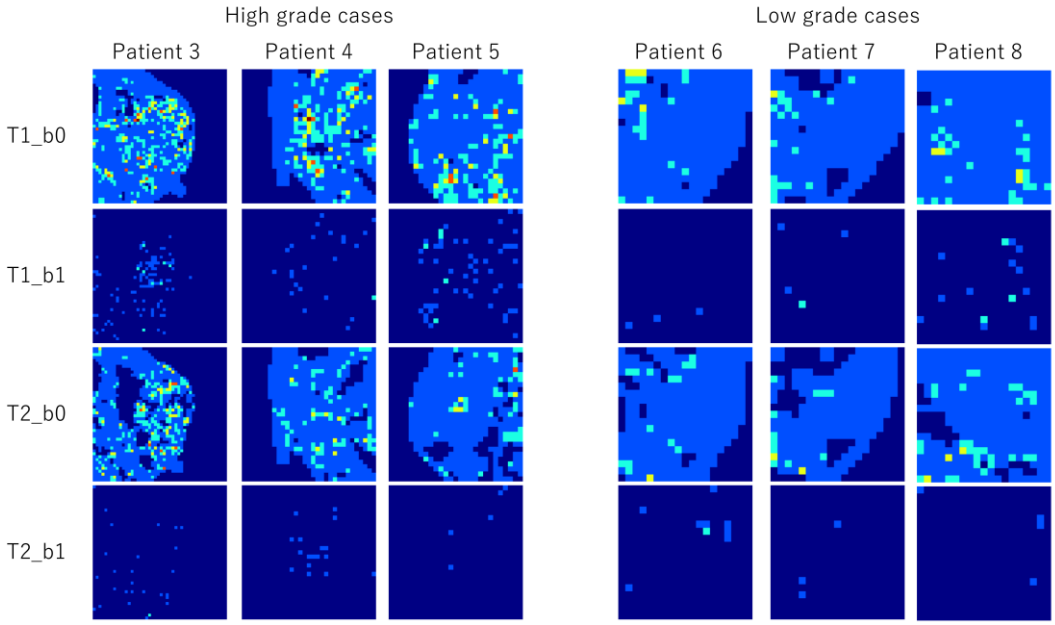
High grade cases
Low grade cases
Patient 3
Patient 4
Patient 5
Patient 6
Patient 7
Patient 8
T1_b0
T1_b1
T2_b0
T2_b1

Figure 5

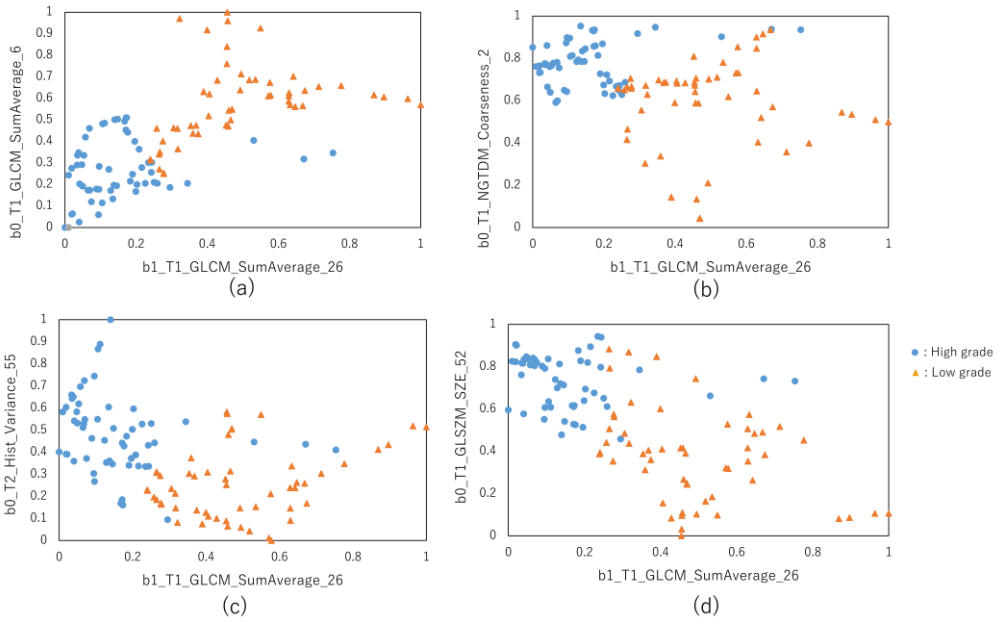
b0_T1_GLCM_SumAverage_6
b1_T1_GLCM_SumAverage_26
(a)
b0_T1_NGTDM_Coarseness_2
b1_T1_GLCM_SumAverage_26
(b)
b0_T2_Hist_Variance_55
b1_T1_GLCM_SumAverage_26
(c)
b0_T1_GLSZM_SZE_52
b1_T1_GLCM_SumAverage_26
(d)
: High grade
: Low grade

Figure 6

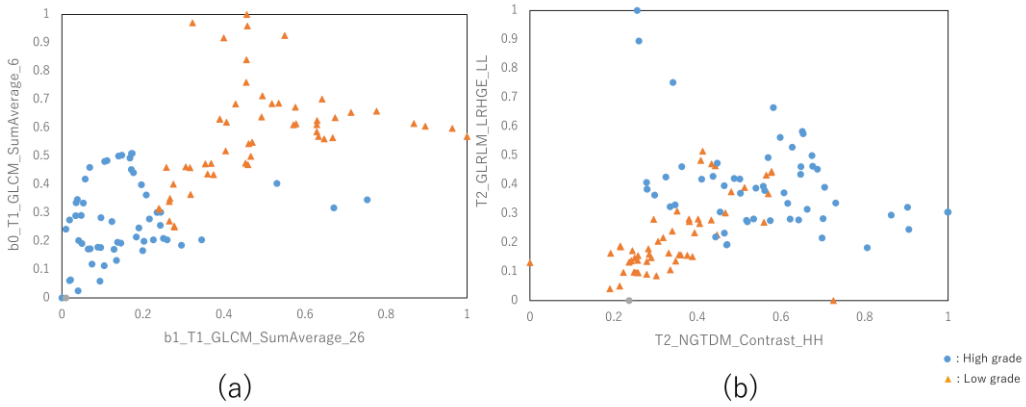
b0_T1_GLCM_SumAverage_6
b1_T1_GLCM_SumAverage_26
T2_GLRLM_LRHGE_LL
T2_NGTDM_Contrast_HH
: High grade
: Low grade
(a)
(b)

Figure 7