Research article

A machine learning model to precisely immunohistochemically classify pituitary adenoma subtypes with radiomics based on preoperative magnetic resonance imaging

AiJun Peng[a,1], HuMing Dai[b,1], HaiHan Duan[b], YaXing Chen[a], JianHan Huang[a], LiangXue Zhou[a,*], LiangYin Chen[b,c,**]

[a] Department of Neurosurgery, West China Hospital, Sichuan University, Chengdu, 610041, Sichuan Province, China
[b] College of Computer Science, Sichuan University, Chengdu, 610065, Sichuan Province, China
[c] The Institute for Industrial Internet Research, Sichuan University, Chengdu, Sichuan Province, China

ARTICLE INFO

ABSTRACT

Purpose: The type of pituitary adenoma (PA) cannot be clearly recognized with preoperative magnetic resonance imaging (MRI) but can be classified with immunohistochemical staining after surgery. In this study, a model to precisely immunohistochemically classify the PA subtypes by radiomic features based on preoperative MR images was developed.

Methods: Two hundred thirty-five pathologically diagnosed PAs, including t-box pituitary transcription factor (Tpit) family tumors (n = 55), pituitary transcription factor 1 (Pit-1) family tumors (n = 110), and steroidogenic factor 1 (SF-1) family tumors (n = 70), were retrospectively studied. T1-weighted, T2-weighted and contrast-enhanced T1-weighted images were obtained from all patients. Through imaging acquisition, feature extraction and radiomic data processing, 18 radiomic features were used to train support vector machine (SVM), k-nearest neighbors (KNN) and Naïve Bayes (NBs) models. Ten-fold cross-validation was applied to evaluate the performance of these models.

Results: The SVM model showed high performance (balanced accuracy 0.89, AUC 0.9549) whereas the KNN (balanced accuracy 0.83, AUC 0.9266) and NBs (balanced accuracy 0.80, AUC 0.9324) models displayed low performance based on the T2-weighted images. The performance of the T2-weighted images was better than that of the other two MR sequences. Additionally, significant sensitivity (P = 0.031) and specificity (P = 0.012) differences were observed when classifying the PA subtypes by T2-weighted images.

Conclusions: The SVM model was superior to the KNN and NBs models and can potentially precisely immunohistochemically classify PA subtypes with an MR-based radiomic analysis. The developed model exhibited good performance using T2-weighted images and might offer potential guidance to neurosurgeons in clinical decision-making before surgery.

1. Introduction

Pituitary adenomas (PAs) are considered common benign neuroendocrine neoplasms that arise from the anterior pituitary gland. These tumors are contained primarily in the sellar region though they may grow up to the suprasellar region, down to the clivus or extend to the parasellar region. The prevalence of PAs is 80–90 cases per 100,000 population every year, and PAs account for approximately 15 % of brain tumors [1–4].

PAs are classified into microadenoma (< 1 cm), macroadenoma (1 − 4 cm) and giant adenoma (> 4 cm). Additionally, they have been considered hormone hypersecretion adenomas and nonfunctioning PAs because of the hormone levels in circulation [5,6]. Recently, PAs have been classified based on the combination of tumor hormonal content and pituitary transcription factors [7–9]. Transcription factors, including t-box pituitary transcription factor (Tpit), pituitary

---

transcription factor 1 (Pit-1) and steroidogenic factor 1 (SF-1), can regulate hormonal activity and adenohypophyseal cell differentiation [8,9]. Corticotroph tumors that express adrenocorticotropic hormone arise from the Tpit lineage, while somatotroph tumors expressing growth hormone and lactotroph tumors expressing prolactin arise from Pit-1. Additionally, gonadotroph tumors expressing β-follicle-stimulating hormone, β-luteinizing hormone or α-subunits arise from SF-1 [10,11]. Therefore, immunohistochemical staining is still the main technological method for classifying PAs [11,12].

The treatments of PAs include transsphenoidal surgery, medication, radiotherapy and observation [13]. Different types of PAs are treated differently [13,14]. Most lactotroph tumor patients treated with medications have better outcomes [15]. However, not all patients with prolactin tumors have significantly high prolactin levels; therefore, it is essential for clinicians to choose the optimal treatment [16]. Additionally, silent corticotroph tumors and Crooke's cell adenomas are highly aggressive [11,12] and have a high risk for recurrence [10]. Therefore, these patients should be carefully followed clinically [17–19]. Thus, precisely classifying PAs can substantially benefit patients with normal or slightly abnormal hormone levels in circulation, especially patients with a high risk of recurrence. The correct PA classification can help patients avoid unnecessary pituitary surgery. To address this problem, we developed a classification model to precisely classify PA subtypes before surgery with radiomics using a machine learning (ML) approach.

Radiomics converts medical imaging data into large amounts of quantitative image features. At present, it has been widely used to extract all kinds of tumor features [20–22]. In this study, we attempted to develop a robust model based on ML to classify PA subtypes using magnetic resonance imaging (MRI) sequences (T1-weighted, T2-weighted, and contrast-enhanced T1-weighted images).

2. Materials and methods

2.1. Patient population

A total of 235 patients with histologically confirmed PA and with corresponding preoperative MR images from the Neurosurgery Department were retrospectively included in this single-center study between January 2016 and February 2019. The clinical characteristics, including sex, age, hormone-hypersecreting tumors, Knosp grade, and tumor diameter, were collected. Two experienced neuropathologists in the Pathology Department from the aforementioned hospital reviewed all diagnostic materials from the archived histopathological slides for all patients to obtain an accurate PA diagnosis.

The inclusion criteria were as follows: (1) all patients were diagnosed with PAs by histology, and PAs were clearly graded by immunohistochemical staining; (2) the quality of the MR images was good and without obvious artifacts; and (3) all of the MR images were obtained within one week before surgery. The exclusion criteria included the following: (1) patients who had undergone surgery or radiotherapy for PA; (2) PAs with no clear immunohistochemical staining; (3) poor MR image quality with obvious artifacts.

2.2. MRI acquisition

All patients underwent MRI of the sellar region with an MRI scanner (Siemens, 3.0 T, Trio, Germany) by the same coil, consisting of T1-weighted, T2-weighted and contrast-enhanced T1-weighted images. In addition, all sequence images were measured by a 2D-spin echo sequence. The specific parameters for the coronal MRI acquisition settings that were defined for these MRI modalities are provided in Table 1. We obtained all Digital Imaging and Communications in Medicine (DICOM) MR images from image archives and communication systems for further qualitative and quantitative analysis.

2.3. Tumor segmentation

All the MR images were loaded onto the open source software ITK-SNAP (version 3.8.0, www.itk-snap.org), and a three-dimensional region-of-interest (ROI) covered the whole tumor and was delineated on each slice of the MR images. This manual tumor segmentation was performed individually by one neurosurgeon and one neuroradiologist (with 14 and 13 years of experience in neuro-oncology diagnosis, respectively). Then, the results were reviewed by the corresponding author and another expert radiologist.

2.4. Radiomics feature extraction

The methodology to extract radiomics features was based on the segmentation results from the aforementioned section. Then, we used the Simple ITK software library (http://www.simpleitk.org/) to read each DICOM image slice of each MR sequence for every patient and integrated them into one three-dimensional original nearly raw raster data (NRRD) image. The same process was also performed for each image slice with an ROI mask, generating a three-dimensional labeled NRRD image. Then, we standardized the original image (f(x) = $\frac{x - \mu_x}{\sigma_x}$, where x and f(x) are the original and normalized intensities, respectively, and μ_x and σ_x are the mean and standard deviation of the image intensity values, respectively) and transformed it with a wavelet transform.

Afterward, we used PyRadiomics 1.2.0 (https://pyradiomics.readthedocs.io/) to extract radiomics features. Eighteen first-order statistical features were extracted from the original images, including 14 shape features, 22 gray level cooccurrence matrix (GLCM) features, 16 gray level run length matrix (GLRLM) features, 16 gray level size zone matrix (GLSZM) features, and 14 gray level dependence matrix (GLDM) features. Thus, a total of 100 features were extracted from the original images of each MR sequence. Additionally, 688 texture features of the same type were extracted from eight wavelet transform images. Therefore, 788 individual radiomics features were extracted from each MR sequence. The details of the feature descriptions can be found elsewhere (http://www.radiomics.io/pyradiomics.html).

2.5. Data preprocessing

Class balance is a key factor to reveal the actual performance of ML classifiers [23]. Moreover, since the incidence of the Tpit family tumors was approximately 10–15 % of all PAs [24], the numbers of Tpit family tumors were the smallest among these three types. Meanwhile, each subtype was partitioned into 10 subsets, 9 of which were randomly used to train the classifiers, and the remaining subset was used for validation. First, the training set was standardized. Then, these 297 training samples were also standardized with the standard scaler package (https://scikit-learn.org/stable/modules/preprocessing.html). Through standardization, the mean of the data was mapped to zero, and the standard deviation was mapped to 1. In addition, the standardized formula was f(x) = $\frac{x - \mu_x}{\sigma_x}$, where x and f(x) are the original and normalized samples, respectively, and μ_x and σ_x are the mean and standard deviation of the sample values, respectively. Furthermore, the standardized model in the training set was applied to the test set. Considering the data balance, the sample size of the SF-1 family tumors and Tpit family tumors in the training set was increased to 99, which was equal to the training set samples of Pit-1 family tumors, by adopting the datasets augmentation approach (SMOTE [25], https://pypi.org/project/imbalanced-learn/).

2.6. Dimensional reduction

Because high-dimensional radiomic features may contain redundant and irrelevant information, which may result in overfitting and

Table 1
MRI modalities of each sequence for pituitary adenoma examination FoV: Field of view, MRI: Magnetic resonance imaging, TE: echo time, TR: repetition time.

	TR (ms)	TE (ms)	Fov (mm)	Slice thickness (mm)	Slice	Voxel size (mm)
T1-weighted	600	8.1	200	2	16	0.8*0.6*2.0
T2-weighted	4000	93	220	2	16	0.8*0.6*2.0
Contrast enhanced T1-weighted	232	8.1	200	2	16	0.9*0.6*2.0

FoV: Field of view, MRI: Magnetic resonance imaging, TE: echo time, TR: repetition time.

decrease the performance of the classifiers, dimensional reduction was necessary [23]. To achieve the best dimensionality reduction effect and effectively avoid over-fitting, ten-fold cross-validation was carried out with reduced dimensionality of the training set and test set. Then, 1–54 dimensions were tested. In addition, the performance of 18 dimensions was better than that of any other dimensions. Therefore, the 788 individual radiomic features were reduced to 18 radiomic features.

2.7. Classification methods and model development

Three common ML methods were used separately: support vector machine (SVM, https://scikit-learn.org/stable/modules/svm.html, scikit-learn software package) with the linear kernel function, k-nearest neighbors (KNN, https://scikit-learn.org/stable/modules/neighbors.html, n neighbors = 3) and Naïve Bayes (NBs, https://scikit-learn.org/stable/modules/naive_bayes.html). Ten-fold cross-validation was applied to evaluate the performance of the classifiers with reduced features in classifying PA subtypes. The overall workflow of radiomics processing is shown in Fig. 1.

2.8. Statistical analysis

Statistical analysis was performed by using SPSS v.23.0 (Armonk, New York, United States). The continuous variables, which are expressed as the means ± standard deviations, and categorical variables of the PA subtypes were compared by one-way ANOVA. In addition, the sensitivity, specificity and accuracy were compared by a nonparametric test (Mann-Whitney U test). A two-sided P value less than 0.05 was considered statistically significant. The performance of each PA subtype in every MR sequence was measured by using confusion matrix-derived metrics, including sensitivity, specificity, accuracy and the area under the curve (AUC) value of the receiver operating characteristic (ROC) curve. Additionally, we used the macro-average ROC to evaluate the performance of the multiclassification classifiers.

3. Results

3.1. Clinical characteristics of each PA subtype

Pit-1 family tumors (59.09 %) and Tpit family tumors (58.18 %) occurred more frequently in female patients than in male patients; in contrast, SF-1 family tumors (58.57 %) occurred more frequently in male patients than in female patients. The mean age of patients was 42.87 years old (ranging from 14 to 75 years old) for those with Pit-1 family tumors, 58.73 years old (ranging from 23 to 75 years old) for those with Tpit family tumors, and 45.44 years old (ranging from 13 to 76 years) for those with SF-1 family tumors. Hormone-hypersecreting tumors were the most common in Pit-1 family tumors (74.55 %) and the least common in SF-1 family tumors (8.57 %). The tumor diameter in the SF-1family tumors was the largest compared with that in the other two types of family tumors (P < 0.001, Table 2). Grade 3–4 tumors occurred more frequently than grade 0–2 tumors in SF-1 family tumors, while the opposite trend was observed in Pit-1 family tumors and Tpit family tumors (P = 0.001, Table 2).

3.2. Hyper-parameter optimization

During the process of adjusting the n neighbors in the KNN model, ten-fold cross-validation was used to evaluate the value of k with its value ranging from 1 to 9. According to the correlation between k and the accuracy of the KNN model, when k equals 3, the performance was the best (Fig. 2). Additionally, according to the pre-experiments (data not shown), the performance of the linear model was much better than that of the Gaussian kernel. Additionally, Niu et al. [26] reported that the performance of the linear SVM model is better in the prediction of cavernous sinus invasion by PAs. In this study, the correlation between the dimension and the accuracy showed that there was no over-fitting phenomenon in the 18 dimensions, but the accuracy decreased while the trend tended to be stable after 18 dimensions (Fig. 3)

3.3. Confusion matrix for classifying PA subtypes in each MR sequence

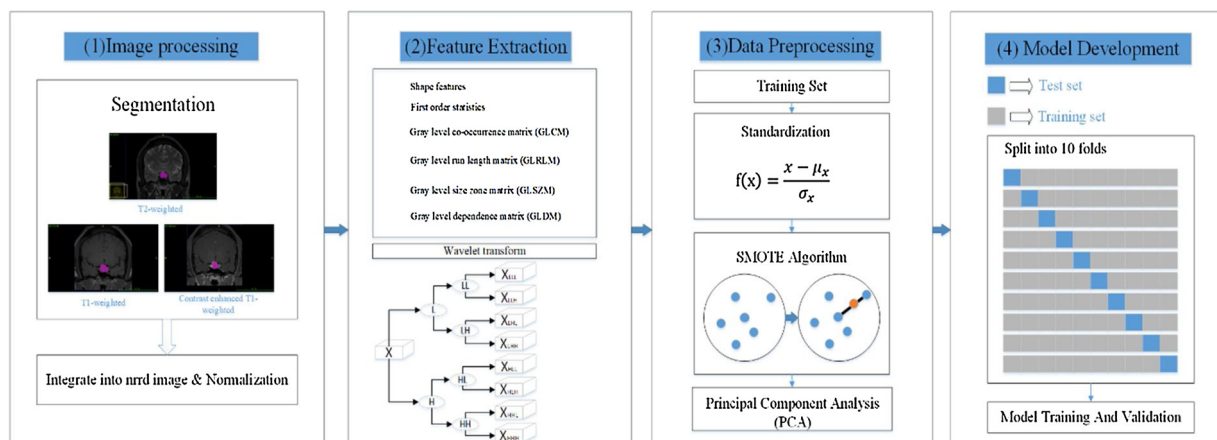Fig. 4 shows the mean confusion matrix results for accurately

**Fig. 1.** The overall workflow of radiomics processing.

Table 2
The basic characteristics for these three pituitary adenoma subtypes.

	Pit-1 family tumors	SF-1 family tumors	Tpit family tumors	P-value
Gender (%)				0.510
Male	45 (40.91 %)	41 (58.57 %)	23 (41.82 %)	
Female	65 (59.09 %)	29 (41.43 %)	32 (58.18 %)	
Age (years)	42.87 ± 12.39	50.13 ± 11.75	45.44 ± 14.19	0.527
Hormone hypersecreting tumors (%)				< 0.001
Yes	82 (74.55 %)	6 (8.57 %)	31 (56.36 %)	
No	28 (25.45 %)	64 (91.43 %)	24 (43.64 %)	
Knosp grade (%)				0.001
Grades 0–2	74 (67.27 %)	36 (51.43 %)	45 (81.82 %)	
Grade 3–4	36 (32.73 %)	34 (48.57 %)	10 (18.18 %)	
Tumor diameter (cm)	1.91 ± 0.89	2.84 ± 0.84	1.40 ± 1.17	< 0.001

Tpit, t-box pituitary transcription factor; Pit-1, pituitary transcription factor 1; SF-1, steroidogenic factor 1.
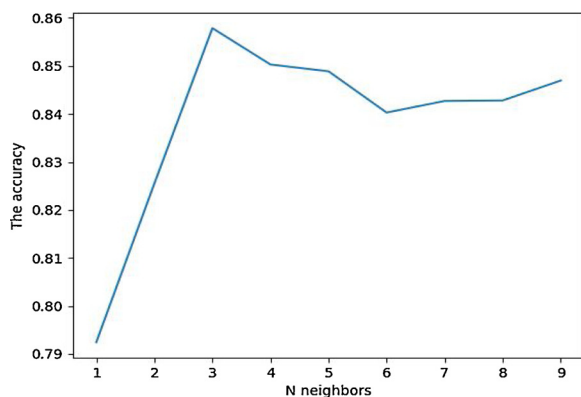
Fig. 2. Correlation between the neighbor of KNN and the accuracy.
This picture shows that the performance of the KNN model was better when k equaled 3.
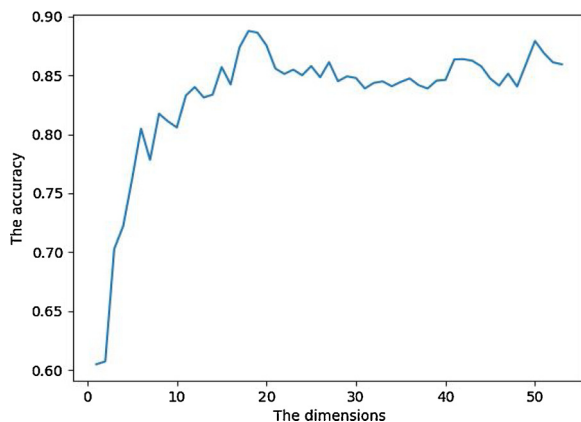
Fig. 3. Correlation between reduced dimensionality and the accuracy.
This picture shows that the performance of the classifier model was the best when the number of dimensions was 18. Additionally, there was no over-fitting phenomenon with 18 dimensions, but the accuracy decreased, while the trend tended to be stable after 18 dimensions.

classifying each PA subtype in T1-weighted, T2-weighted and contrast-enhanced T1-weighted images on three classifiers. These matrix results were constructed by ten-fold cross-validation in the test set of each MR sequence. According to the mean matrix results, the performance of the SVM model (Fig. 4A–C) was better than that of the KNN (Fig. 4D–F) and NBs (Fig. 4G–I) models in T1-weighted, T2-weighted and contrast-enhanced T1-weighted images. Additionally, based on the mean accuracy results, the SVM model, compared with the KNN and NBs models, was the best model for immunohistochemically classifying PA subtypes (Table 3). Additionally, the SVM model showed a good performance

(balanced accuracy 0.89), whereas the KNN (balanced accuracy 0.83) and NBs (balanced accuracy 0.80) models showed weak performance based on T2-weighted images (Table 3). The classification of each PA subtype with the SVM model in T2-weighted images (Fig. 4B) showed better performance than the model in T1-weighted (Fig. 4A) and contrast-enhanced T1-weighted (Fig. 4C) images. Moreover, the performance of each fold for the confusion matrix of the SVM model is shown in the Supplementary file (Figs. 1–3).

3.4. Results of the comparison of each MR sequence for classifying PA subtypes with the SVM model

The classification results of the PA subtypes for each MR sequence are given in Table 4. We obtained satisfactory classification results by using the SVM model with the mean sensitivity, specificity and accuracy for each MR sequence of the PA subtypes (Table 4). The performance of T2-weighted images was better than that of the other two MR sequences, and significant differences in sensitivity (P = 0.031) and specificity (P = 0.012) in classifying the PA subtypes in T2-weighted images were observed. However, the accuracy for classifying each PA subtype in T2-weighted images was similar (P = 0.401), which demonstrated that the SVM model was more robust in T2-weighted images compared with the other two MR images (Table 4).

### 3.5. Macro-average ROC curve for the evaluation of PA subtype classification

Fig. 3 shows one of the 10 macro-average ROC curves and the ROC curve of each PA subtype according to the SVM model in T1-weighted (Fig. 5A), T2-weighted (Fig. 5B) and contrast-enhanced T1-weighted (Fig. 5C) images, the KNN model in T1-weighted (Fig. 5D), T2-weighted (Fig. 5E), and contrast-enhanced T1-weighted (Fig. 5F) images and the NBs model in T1-weighted (Fig. 5G), T2-weighted (Fig. 5H), and contrast-enhanced T1-weighted (Fig. 5I) images. The mean AUC values obtained with the SVM model were 0.8762, 0.9549 and 0.8806; those from the KNN model were 0.8598, 0.9266 and 0.7947; and those from the NBs model were 0.8492, 0.9324 and 0.8309 in T1-weighted, T2-weighted and contrast-enhanced T1-weighted images, respectively. The performance of the macro-average ROC curve and the ROC curve obtained from these three models are shown in the Supplementary file (Figs. 4–12). Additionally, the AUC values obtained from the SVM model were higher than those obtained the other two models, as shown in the Supplementary file (Figs. 4–6). Therefore, the macro-average ROC curve for the evaluation of the multiclassifier showed that T2-weighted images performed better than the other two MR sequences.

3.6. Radiomics features for the distribution of PA subtypes

A total of 788 individual radiomics features were extracted from one patient. Surprisingly, the heat map with covariance analysis showed
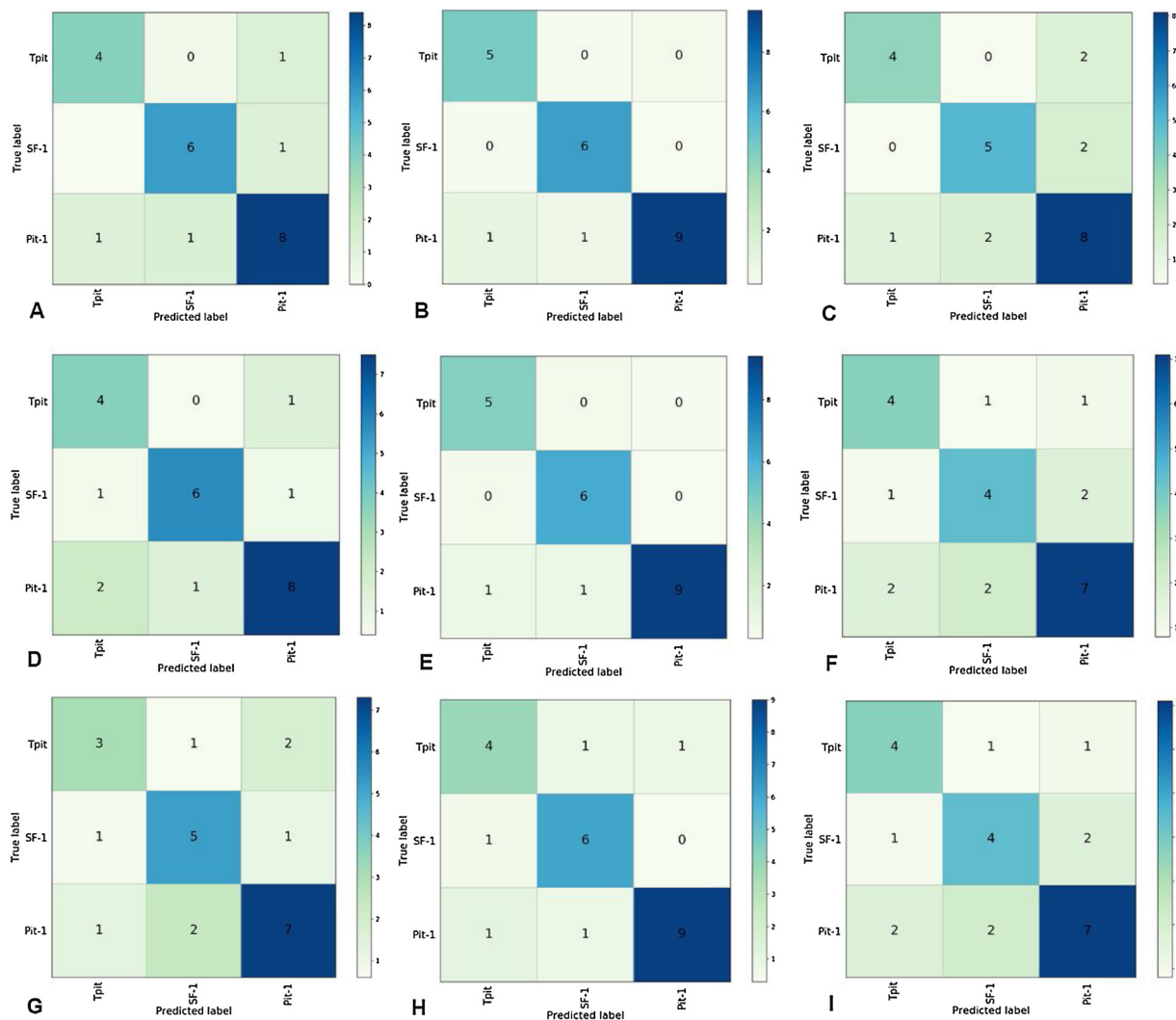
Fig. 4. Mean confusion matrix results for accurately classifying each PA subtype with three ML models with ten-fold cross-validation in MR sequences. The mean confusion matrix results for accurately classifying each PA subtype with the SVM model (A–C), the KNN model (D–F) and the NBs model (G–H) with ten-fold cross-validation in T1-weighted (A, D, G), T2-weighted (B, E, H), and contrast-enhanced T1-weighted (C, F, I) images.

Table 3
The mean balanced accuracy of these three classifiers for each MR sequence.

Classifiers	MR sequence		
	T1-weighted	T2-weighted	Contrast enhanced T1-weighted
SVM	0.77	0.89	0.71
KNN	0.72	0.83	0.65
NBs	0.66	0.80	0.65

outstanding performance for differentiating Tpit family tumors from Pit-1 family tumors and SF-1 family tumors with radiomics features in T2-weighted images (Fig. 6). Fig. 6 shows the sample distribution. In addition, different colors represent different numerical values, which were used to judge the approximate distribution of the sample. In Fig. 6, rows represent individual features, while columns represent individual patients. Additionally, columns 0–55, 56–125, and 126–235 in Fig. 6 represent patients with Tpit family tumors, patients with SF-1 family tumors, and patients with Pit-1 family tumors, respectively. According

Table 4
The classification results of the pituitary adenoma subtypes for each MR sequence.

Variables	MR sequence	Pit-1 family tumors	SF-1 family tumors	Tpit family tumors	P-value
Sensitivity	T1-weighted	0.76	0.83	0.75	0.134
	T2-weighted	0.81	0.93	0.86	0.031
	Contrast enhanced T1-weighted	0.74	0.71	0.69	0.703
Specificity	T1-weighted	0.80	0.90	0.93	0.221
	T2-weighted	0.82	0.89	0.85	0.012
	Contrast enhanced T1-weighted	0.75	0.88	0.91	0.515
Accuracy	T1-weighted	0.76	0.88	0.88	0.939
	T2-weighted	0.91	0.94	0.91	0.401
	Contrast enhanced T1-weighted	0.74	0.83	0.86	0.541

Tpit, t-box pituitary transcription factor; Pit-1, pituitary transcription factor 1; SF-1, steroidogenic factor 1.
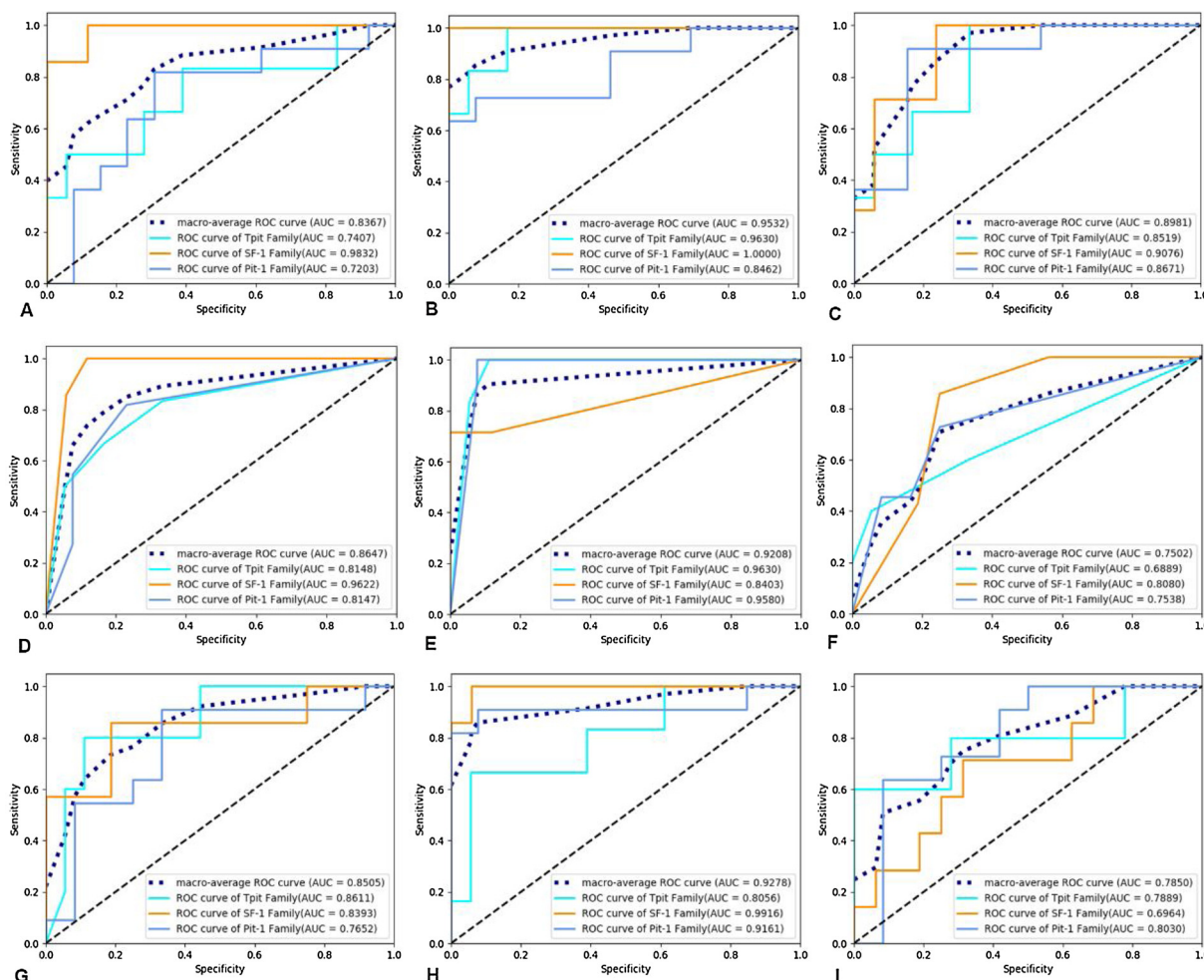
Fig. 5. One of the ten macro-average ROC curves and the ROC curve of each PA subtype in MR sequences.
The AUC value for the macro-average ROC curve of each PA subtype with the SVM model (A–E), the KNN model (D–F) and the NBs model (G–H) with ten-fold cross-validation in T1-weighted (A, D, G), T2-weighted (B, E, H), and contrast-enhanced T1-weighted (C, F, I) images.
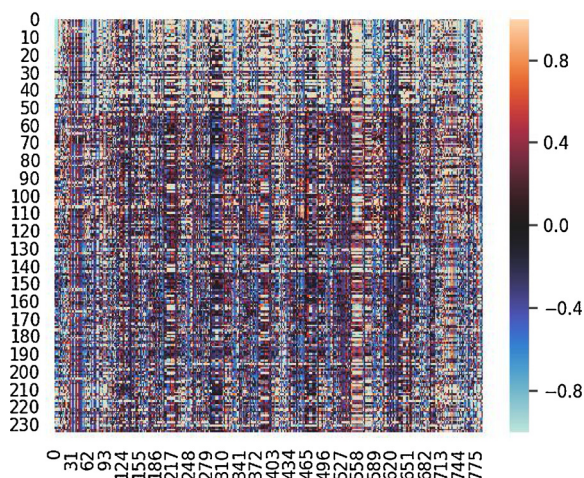
Fig. 6. Heat map for differentiating the three PAs subtypes with radiomics features in T2-weighted images.
This figure demonstrates the sample distribution, which could be judged through color distribution. Rows represent individual features, while columns represent individual patients. Additionally, columns 0–55, 56–125, and 126–235 represent each PA subtype.

to the first three principal components after the dimensional reduction of radiomics features, we also created a three-dimensional space plot that demonstrated the spatial distribution of these three PA subtypes by which they can be distinguished from each other on T2-weighted images (Fig. 7).

4. Discussion

In this study, we found that our SVM model was superior to the KNN and NBs models and has potential value in classifying PA subtypes with radiomics analysis based on preoperative MR images. We demonstrated that by using the ML approach, according to pituitary transcription factors, PA subtypes can be precisely classified with preoperative coronal MR images through feature extraction and quantitative analysis. The results from our study suggested that coronal T2-weighted MR sequences had better performance in classifying PA subtypes than T1-weighted and contrast-enhanced T1-weighted sequences.

Currently, using the immunohistochemical characteristics of tumor cells has been the main approach to accurately classify PAs. Moreover, the 2017 WHO classification of PAs emphasizes the role of nuclear transcription factors, including Tpit, Pit-1, and SF-1 in the classification of PAs [8,10]. Hormone hypersecretion PAs are easy to classify, but PAs that are hormone-negative before surgery are difficult to accurately classify. Additionally, due to the lack of reliable antibodies against Tpit in the market [10], Tpit family tumors cannot be accurately diagnosed. In addition, some of the nonfunctioning PAs, including silent
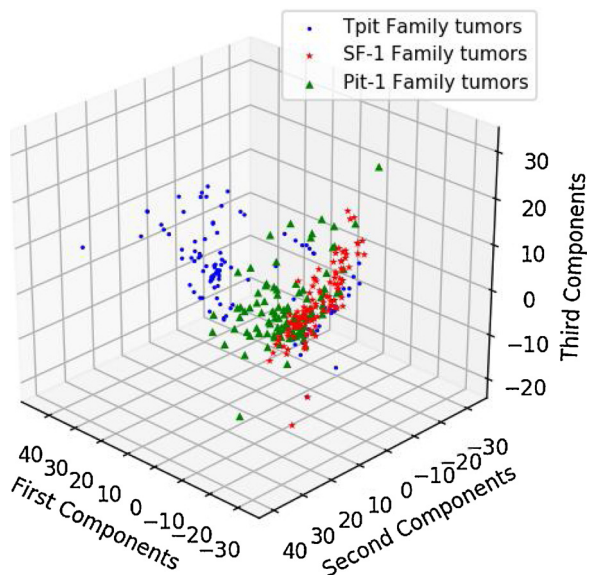
Fig. 7. A three-dimensional space plot demonstrates the spatial distribution of the three PA subtypes.

The x, y and z axes represent the first three principal components after the dimensional reduction of radiomics features in T2-weighted images. The blue solid points, orange asterisks and green triangles represent patients with Tpit family tumors, patients with SF-1 family tumors, and patients with Pit-1 family tumors, respectively.

corticotroph tumors and Crooke's cell adenomas, have a high risk of aggression and recurrence after surgery [27,28]. Moreover, medication is recommended for lactotroph tumors (a type of Pit-1 family tumor), few of which need surgery [13]. Therefore, developing an ML-based model to accurately classify PA subtypes by using MR images can provide patients and their families with useful suggestions and offer potential guidance for neurosurgeons in clinical decision-making before surgery.

In this study, Pit-1 and Tpit family tumors showed a slight predominance in females, whereas SF-1 family tumors showed a slight predominance in males that presented with older age (Table 2). Additionally, most of the Pit-1 and Tpit family tumors had hormone hypersecretion with smaller tumors and less invasion of the cavernous sinus than SF-1 family tumors (Table 2). Due to their nonfunctional SF-1, growing tumors with compression in the structures surrounding the sellar region often cause patients to experience symptoms including visual impairment, headache, and especially hypopituitarism. The results in our study were consistent with those in previous reports [10,29,30].

To our knowledge, only one study has been reported concerning nonfunctioning PA types [22]. In this study, we built an SVM model to accurately classify three PA subtypes with radiomics based on T1-weighted, T2-weighted and contrast-enhanced weighted images through ten-fold cross-validation, which may avoid sampling bias. We found that the accuracy was higher for SF-1 family tumors in T1- and T2-weighted images, whereas the accuracy was higher for Pit-1 family tumors in contrast-enhanced T1-weighted images. This finding is probably due to tumor characteristics, including size, texture consistency or shape; additionally, the contrast agent influencing the intensity of tumors may be another reason. Therefore, the potential mechanism needs to be further studied.

Data handling is an important process used to reveal the actual performance of ML classifiers [23]. In our study, the images and data were standardized for the following reasons: (1) to eliminate the dimensional relationship between variables to make the data comparable; and (2) to improve the convergence speed and reduce the calculation amount. Moreover, the high-dimensionality features of small data may

cause overfitting, and an imbalanced class can lead to misleading results [23]. In our study, PCA was used to select the sensitive components features that would make our model more reliable and robust. Furthermore, we found that the accuracy of classifiers with high-dimensionality features was slightly higher than that with dimensional feature reduction (data not shown). The probable explanation for this finding may be related to the relevant data and noise. In addition, another possible reason may be that single features and multiple fusion features play a common role in classification with high-dimensional features. However, dimensional reduction can reduce the interference caused by noise to the classification and improve the stability and robustness of the model. In this study, we found that the performance of the classifier model was the best, and no over-fitting phenomenon was observed when the radiomic features were reduced to approximately 2.5 % of the total radiomic features. Therefore, we selected ten-fold cross-validation to evaluate the performance of our models with dimensional radiomics feature reduction. Surprisingly, we achieved satisfactory results.

In our study, we used feature extraction from only MR image sequences before surgery to classify PA subtypes. Additionally, our SVM model was a multiclassifier model, while most previous studies [22,31,32] performed binary classification for nonfunctioning PAs and PA invasion of the cavernous sinus. Zeynalova et al. [33] reported that the T2-weighted MR sequence was better in predicting the consistency of pituitary macroadenomas with an ML-based approach. Additionally, Kocak et al. [34] showed that texture analysis on T2-weighted images by using the ML-based approach had better performance in predicting the response to somatostatin analogs (SA) in GH-secreting macro-PAs. Therefore, we speculate that the texture consistency of T2-weighted images may be better than that of T1-weighted and contrast-enhanced T1-weighted images, but this needs to be further explored in future research. Zhang et al. [22] reported that T1-weighted MR images before surgery performed better in discriminating nonfunctioning PA subtypes than contrast-enhanced T1-weighted images with a radiomics approach. However, they did not compare T1-weighted images with T2-weighted images. In this study, we used not only the accuracy, sensitivity, specificity, and AUC of each PA subtype but also the macro-average ROC curve to evaluate the performance of these three classifiers (Fig. 7). We found that these three models can classify PA subtypes on T2-weighted images, consistent with the aforementioned studies [33,34].

In our study, the SVM model demonstrated a good classification performance, whereas the KNN and NBs models demonstrated a lower classification performance based on T2-weighted images. Probable explanations for this finding may be as follows: (1) the radiomics feature datasets with continuous values may affect the classification of the probabilistic NBs model; and (2) the SVM classifies by finding a few SVMs to obtain a good hyperplane, while the KNN model finds the data most similar to the classified samples from all the data. Therefore, the SVM is a sparse model, and the KNN is a nonsparse model. Additionally, the SVM was robust to noise and overfitting [35], so the effect of the SVM model was better than that of the KNN model.

Additionally, through color distribution, which can be used to judge the approximate distribution of the samples, we found that high-dimensional features were correlated with each PA subtype and could be used to differentiate these three PA subtypes (Fig. 6). Moreover, the three-dimensional space plot demonstrated that these three PA subtypes can be differentiated from each other with the first three principal component features through PCA (Fig. 7). This finding illustrates that these three PA subtypes may process their own MR characteristics, but further research is needed.

Our study has some limitations. First, manually labeling tumors is a time-consuming and laborious task, especially when the dataset is large. This problem may be solved when our work on automatic tumor recognition is finished. Second, all MR images were collected with one type of MR machine from a single center. Different hospitals have

different MR machines that produce different MR images. Therefore, a future multicenter trial with different MR images is needed to validate our model. Third, due to the lower incidence of thyrotroph adenomas and plurihormonal adenomas, we did not include these patients; therefore, the possibility of selection bias cannot be ignored. Finally, in this study, we built models for classifying only three PA subtypes according to pituitary transcription factors, but we did not continue to classify growth hormone and lactotroph tumors that belong to Pit-1 family tumors. Therefore, further study needs to be explored to validate our model.

5. Conclusions

We developed a robust SVM model with radiomics features based on preoperative MR images to precisely immunohistochemically classify PA subtypes. The SVM model using radiomics features extracted from T2-weighted images had a better performance compared with that from T1-weighted and contrast-enhanced T1-weighted images. This model could offer useful suggestions to patients and their families and potential guidance for neurosurgeons in clinical decision-making before surgery.

Contributors

Peng AiJun, HuMing Dai, HaiHan Duan, YaXing Chen, and JianHan Huang were involved in the study design and data collection. The manuscript was written by Peng AiJun and HuMing Dai. Peng AiJun, HuMing Dai and Huang JianHan were responsible for the statistical analyses and interpretation of study results. Zhou LiangXue and LiangYin Chen served as principal investigator for the study, contributed data and participated in data interpretation as well as in the development of the manuscript. All authors have made substantial contributions to this study and approved submission to this journal.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ejrad.2020.108892.

References

[1] S. Melmed, Pathogenesis of pituitary tumors, nature reviews, Endocrinology 7 (5) (2011) 257–266.
[2] A. Raappana, J. Koivukangas, T. Ebeling, T. Pirila, Incidence of pituitary adenomas in Northern Finland in 1992–2007, J. Clin. Endocrinol. Metab. 95 (9) (2010) 4268–4275.
[3] A. Fernandez, N. Karavitaki, J.A. Wass, Prevalence of pituitary adenomas: a community-based, cross-sectional study in Banbury (Oxfordshire, UK), Clin. Endocrinol. 72 (3) (2010) 377–382.
[4] A.F. Daly, M. Rixhon, C. Adam, A. Dempegioti, M.A. Tichomirowa, A. Beckers, High prevalence of pituitary adenomas: a cross-sectional study in the province of Liege, Belgium, J. Clin. Endocrinol. Metab. 91 (12) (2006) 4769–4775.
[5] E. Manojlovic-Gacic, B.E. Engstrom, O. Casar-Borota, Histopathological classification of non-functioning pituitary neuroendocrine tumors, Pituitary 21 (2) (2018) 119–129.
[6] S.A. Paschou, A. Vryonidou, D.G. Goulis, Pituitary incidentalomas: a guide to assessment, treatment and follow-up, Maturitas 92 (2016) 143–149.
[7] M. Shibuya, Welcoming the new WHO classification of pituitary tumors 2017:
revolution in TTF-1-positive posterior pituitary tumors, Brain Tumor Pathol. 35 (2) (2018) 62–70.
[8] H. Nishioka, N. Inoshita, New WHO classification of pituitary adenomas (4th edition): assessment of pituitary transcription factors and the prognostic histological factors, Brain Tumor Pathol. 35 (2) (2018) 57–61.
[9] N. Inoshita, H. Nishioka, The 2017 WHO classification of pituitary adenoma: overview and comments, Brain Tumor Pathol. 35 (2) (2018) 51–56.
[10] M.B.S. Lopes, The 2017 World Health Organization classification of tumors of the pituitary gland: a summary, Acta Neuropathol. 134 (4) (2017) 521–535.
[11] E.R. Laws Jr., D.L. Penn, C.S. Repetti, Advances and controversies in the classification and grading of pituitary tumors, J. Endocrinol. Invest. 42 (2) (2019) 129–135.
[12] S.L. Asa, O. Mete, What's new in pituitary pathology? Histopathology 72 (1) (2018) 133–141.
[13] M.E. Molitch, Diagnosis and treatment of pituitary adenomas: a review, JAMA 317 (5) (2017) 516–524.
[14] Y. Greenman, N. Stern, Optimal management of non-functioning pituitary adenomas, Endocrine 50 (1) (2015) 51–55.
[15] C. Araujo, O. Marques, R. Almeida, M.J. Santos, Macroprolactinomas: longitudinal assessment of biochemical and imaging therapeutic responses, Endocrine 62 (2) (2018) 470–476.
[16] G.U. Mehta, R.R. Lonser, Management of hormone-secreting pituitary adenomas, Neurooncology 19 (6) (2017) 762–773.
[17] O. Mete, M.B. Lopes, Overview of the 2017 WHO classification of pituitary tumors, Endocr. Pathol. 28 (3) (2017) 228–243.
[18] O. Cooper, Silent corticotroph adenomas, Pituitary 18 (2) (2015) 225–231.
[19] A. Ben-Shlomo, O. Cooper, Silent corticotroph adenomas, Pituitary 21 (2) (2018) 183–193.
[20] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R.G. van Stiphout, P. Granton, C.M. Zegers, R. Gillies, R. Boellard, A. Dekker, H.J. Aerts, Radiomics: extracting more information from medical images using advanced feature analysis, Eur. J. Cancer (Oxford, England: 1990) 48 (4) (2012) 441–446.
[21] H.J. Aerts, E.R. Velazquez, R.T. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M.M. Rietbergen, C.R. Leemans, A. Dekker, J. Quackenbush, R.J. Gillies, P. Lambin, Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, Nat. Commun. 5 (2014) 4006.
[22] S. Zhang, G. Song, Y. Zang, J. Jia, C. Wang, C. Li, J. Tian, D. Dong, Y. Zhang, Non-invasive radiomics approach potentially predicts non-functioning pituitary adenomas subtypes before surgery, Eur. Radiol. 28 (9) (2018) 3692–3701.
[23] B. Kocak, E.S. Durmaz, E. Ates, O. Kilickesmez, Radiomics with artificial intelligence: a practical guide for beginners, Diagn. Interv. Radiol. (Ankara, Turkey) 25 (6) (2019) 485–495.
[24] J. Seltzer, C.E. Ashton, T.C. Scotton, D. Pangal, J.D. Carmichael, G. Zada, Gene and protein expression in pituitary corticotroph adenomas: a systematic review of the literature, Neurosurg. Focus 38 (2) (2015) E17.
[25] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321.
[26] J. Niu, S. Zhang, S. Ma, J. Diao, W. Zhou, J. Tian, Y. Zang, W. Jia, Preoperative prediction of cavernous sinus invasion by pituitary adenomas using a radiomics method based on magnetic resonance images, Eur. Radiol. 29 (3) (2019) 1625–1634.
[27] H. Nishioka, N. Inoshita, O. Mete, S.L. Asa, K. Hayashi, A. Takeshita, N. Fukuhara, M. Yamaguchi-Okada, Y. Takeuchi, S. Yamada, The complementary role of transcription factors in the accurate diagnosis of clinically nonfunctioning pituitary adenomas, Endocr. Pathol. 26 (4) (2015) 349–355.
[28] O. Mete, S.L. Asa, Clinicopathological correlations in pituitary adenomas, Brain Pathol. (Zurich, Switzerland) 22 (4) (2012) 443–453.
[29] G. Kontogeorgos, E. Thodou, The gonadotroph origin of null cell adenomas, Hormones (Athens, Greece) 15 (2) (2016) 243–247.
[30] C. Ramirez, S. Cheng, G. Vargas, S.L. Asa, S. Ezzat, B. Gonzalez, L. Cabrera, G. Guinto, M. Mercado, Expression of Ki-67, PTTG1, FGFR4, and SSTR 2, 3, and 5 in nonfunctioning pituitary adenomas: a high throughput TMA, immunohistochemical study, J. Clin. Endocrinol. Metab. 97 (5) (2012) 1745–1751.
[31] X. Chen, Y. Tong, Z. Shi, H. Chen, Z. Yang, Y. Wang, L. Chen, J. Yu, Noninvasive molecular diagnosis of craniopharyngioma with MRI-based radiomics approach, BMC Neurol. 19 (1) (2019) 6.
[32] J. Niu, S. Zhang, S. Ma, J. Diao, W. Zhou, J. Tian, Y. Zang, W. Jia, Preoperative prediction of cavernous sinus invasion by pituitary adenomas using a radiomics method based on magnetic resonance images, Eur. Radiol. 29 (3) (2019) 1625–1634.
[33] A. Zeynalova, B. Kocak, E.S. Durmaz, N. Comunoglu, K. Ozcan, G. Ozcan, O. Turk, N. Tanriover, N. Kocer, O. Kizilkilic, C. Islak, Preoperative evaluation of tumour consistency in pituitary macroadenomas: a machine learning-based histogram analysis on conventional T2-weighted MRI, Neuroradiology 61 (7) (2019) 767–774.
[34] B. Kocak, E.S. Durmaz, P. Kadioglu, O. Polat Korkmaz, N. Comunoglu, N. Tanriover, N. Kocer, C. Islak, O. Kizilkilic, Predicting response to somatostatin analogues in acromegaly: machine learning-based high-dimensional quantitative texture analysis on T2-weighted MRI, Eur. Radiol. 29 (6) (2019) 2731–2739.
[35] J.T. Senders, O. Arnaout, A.V. Karhade, H.H. Dasenbrock, W.B. Gormley, M.L. Broekman, T.R. Smith, Natural and Artificial Intelligence in Neurosurgery: A Systematic Review, Neurosurgery 83 (2) (2018) 181–192.