Discrimination between non-functioning pituitary adenomas and hypophysitis using machine learning methods based on magnetic resonance imaging-derived texture features

Serdar Sahin[1] · Gokcen Yildiz[2] · Seda Hanife Oguz[3] · Orkun Civan[2] · Ebru Cicek[4] · Emre Durcan[1] · Nil Comunoglu[5] · Hande Mefkure Ozkaya[1] · Aysim Buge Oz[5] · Figen Soylemezoglu[6] · Kader Karli Oguz[7] · Selçuk Dagdelen[3] · Tomris Erbas[3] · Osman Kizilkilic[2] · Pinar Kadioglu[1,8]

Abstract

Purpose Hypophysitis is a heterogeneous condition that includes inflammation of the pituitary gland and infundibulum, and it can cause symptoms related to mass effects and hormonal deficiencies. We aimed to evaluate the potential role of machine learning methods in differentiating hypophysitis from non-functioning pituitary adenomas.

Methods The radiomic parameters obtained from T1A-C images were used. Among the radiomic parameters, parameters capable of distinguishing between hypophysitis and non-functioning pituitary adenomas were selected. In order to avoid the effects of confounding factors and to improve the performance of the classifiers, parameters with high correlation with each other were eliminated. Machine learning algorithms were performed with the combination of gray-level run-length matrix-low gray level run emphasis, gray-level co-occurrence matrix-correlation, and gray-level co-occurrence entropy.

Results A total of 34 patients were included, 17 of whom had hypophysitis and 17 had non-functioning pituitary adenomas. Among the 38 radiomics parameters obtained from post-contrast T1-weighted images, 10 tissue features that could differentiate the lesions were selected. Machine learning algorithms were performed using three selected parameters; gray level run length matrix-low gray level run emphasis, gray-level co-occurrence matrix-correlation, and gray level co-occurrence entropy. Error matrices were calculated by using the machine learning algorithm and it was seen that support vector machines showed the best performance in distinguishing the two lesion types.

Conclusions Our analysis reported that support vector machines showed the best performance in distinguishing hypophysitis from non-functioning pituitary adenomas, emphasizing the importance of machine learning in differentiating the two lesions.

Keywords Machine learning · Hypophysitis · Non-functioning pituitary adenoma, support vector machines

✉ Pinar KadiogluMD
kadioglup@yahoo.com

1   Division of Endocrinology and Metabolism, Department of Internal Medicine, Cerrahpasa Faculty of Medicine, Istanbul University-Cerrahpasa, Istanbul, Turkey

2   Department of Radiology, Cerrahpasa Faculty of Medicine, Istanbul University-Cerrahpasa, Istanbul, Turkey

3   Division of Endocrinology and Metabolism, Department of Internal Medicine, Faculty of Medicine, Hacettepe University, Ankara, Turkey

4   Department of Internal Medicine, Cerrahpasa Faculty of

Medicine, Istanbul University- Cerrahpasa, Istanbul, Turkey

5   Department of Pathology, Cerrahpasa Faculty of Medicine, Istanbul University-Cerrahpasa, Istanbul, Turkey

6   Department of Pathology, Faculty of Medicine, Hacettepe University, Ankara, Turkey

7   Department of Radiology, Faculty of Medicine, Hacettepe University, Ankara, Turkey

8   Cerrahpasa Medical Faculty, Department of Internal Medicine, Division of Endocrinology and Metabolism, Istanbul University-Cerrahpasa, Kocamustafapasa Street No: 53, 34098 Fatih, Istanbul, Turkey

Introduction

Hypophysitis is a heterogeneous condition that includes inflammation of the pituitary gland and infundibulum, and it can cause symptoms related to mass effect and hormonal deficiencies [1]. The prevalence of hypophysitis ranges from 0.2 to 0.88% [2–4], and the annual incidence of hypophysitis is 1 case per 9 million individuals [4]. According to the anatomic region of inflammation, hypophysitis is classified as adenohypophysitis, infundibuloneurohypophysitis, and panhypophysitis [4]. According to histopathologic features, it is classified as lymphocytic (71.8%), granulomatous (18.6%), and rarely, xanthomatous (3.3%), immunoglobulin (Ig)-G4-related, necrotizing pituitary, and mixed forms [5]. Secondary hypophysitis can develop in the course of inflammatory, autoimmune, vascular, infectious, and neoplastic diseases, and with adverse effects of some drugs [5].

Pituitary adenomas are the most common intracranial neoplasms with a prevalence of 0.1%, their prevalence at autopsy is 15% [6, 7]. Approximately 35% of pituitary adenomas do not secrete hormones [8]. Non-functioning pituitary adenomas (NFPAs) can be detected in the examination of hypopituitarism, incidentally, or in the examination of neurologic symptoms due to mass effect such as visual disturbances and headache [9, 10]. There are also non-adenoma lesions such as hypophysitis, which cause symptoms similar to NFPA and do not always require surgery. Autoantibodies, imaging, and pituitary biopsy are used for differential diagnosis between hypophysitis and NFPA. However, there is currently no reliable autoantibody available [8].

Machine learning can help create a more reliable assisted diagnostic tool. Establishing the correct diagnosis in these patients may provide better clinical decision support. Machine learning has advantages over other predictive methods as it enables a predictive computer model to automatically learn the best predictive features found in the training data. Unlike using a human operator to manually identify these features, machine learning models can automatically identify the most robust predictive features and potentially generalize this knowledge to new patient groups.

In this study, we aimed to develop an automatic system that can distinguish between hypophysitis and NFPA, rather than a manual method performed by a human operator.

Materials and methods

The medical records of patients with hypophysitis and NFPAs who were followed by the Endocrinology and Metabolism outpatient clinic of Istanbul University-Cerrahpasa, Cerrahpasa Medical School, and Hacettepe University, Hacettepe Medical School were retrospectively reviewed.

This study was approved by the Local Ethics Committee of Cerrahpasa Faculty of Medicine.The study adhered to the Tenets of Helsinki.

Patient selection

Inclusion criteria were (a) NFPAs with histopathologic diagnosis; (b) clinical and radiologic or histopathologically established diagnosis of hypophysitis; (c) high-quality dynamic contrast-enhanced pituitary magnetic resonance imaging (MRI) before surgery or treatment. The exclusion criteria were as follows: (a) a history of another intracranial disease; (b) having undergone radiotherapy, or neurosurgery before MRI; (c) complicated lesion with hemorrhage in NFPAs; (d) presence of apoplexy.

MRI protocol

Pituitary MRI was available in all patients, including Gd-based contrast-enhanced gadopentetate dimeglumine (0.1 mmol/kg) T1-weighted (W) sequences and T2-weighted sequences. A 3.0 T MRI scanner (Ingenia; Philips Healthcare, Best, Netherlands) was used, acquiring axial, coronal, and sagittal data. Dynamic MRI scanwas conducted within 200 s following intravenous injection of the contrast agent. Examples of MRI of NFPA and hypophysitis are shown in Fig. 1.

MR texture analysis

The volume of interest (VOI) was manually segmented using the LifeX softwareby one senior and one junior radiologist in a three-dimensional (3D) fashion, slice by slice on
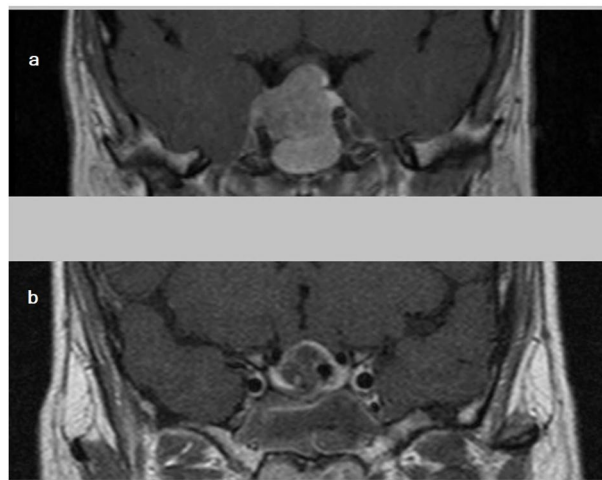

Fig. 1 Non-functioning pituitary adenoma (a)and hypophysitis (b) on post-Gd T1-W MRI images

coronal and sagittal MRI [11]. The entire tumor volume and hypophysitis involvement areas were included in the VOI. Any disagreement was corrected by the senior radiologist. Feature reproducibility was assessed by the intraclass correlation coefficient (ICC) with the cutoff value of 0.8 and the coefficient of variation (CV) with the cutoff value of 20%.

Texture features were extracted using the LifeX software on post-contrast T1-weighted images (post-Gd T1-W) and T2-W images. Voxels within the VOI outside the range $m \pm 3sd$ were rejected and not considered in MRI texture analysis. Spatial resampling was 0.5 mm (x), 0.5 (y), and 4 mm (z) for each image. Intensity rescaling was performed as a pretreatment step by choosing 256 as the number of grayscale levels. Texture features were extracted from each T2-W coronal, T1-W coronal, and T1-W sagittal MRI.

We extracted first-order features based on intensity-based histograms. Six histogram features were computed for both recurrent and non-recurrent tumor volumes: mean, variance, skewness, kurtosis, entropy, and energy. Second-order features were extracted: six features from grey-level co-occurrence matrix (GLCM), 11 features from the grey-level run-length matrix (GLRLM), 11 features from grey-level zone length matrix (GLZLM), and four features from neighborhood grey-level dependence matrix (NGLDM). A total of 38 features were computed for each image.

From these features, those that could differentiate the two lesion classes significantly ($p < 0.01$) were determined using the Chi-square test. Classifiers using these parameters were trained using different machine learning methods with Matlab software and their performance in classifying NFPA and hypophysitis lesions on test data was measured. The machine learning algorithms were used for model development: linear discriminat analysis, fine, medium and coarse decision trees, k-nearest neighbors, support vector machines (SVM), naive Bayes, ensemble classifiers. The learning and testing phase were performed using 10-fold internal and 10-fold external cross-validation. The performances of the methods used were compared using the Matthews correlation coefficient (MCC). An algorithm that correctly classified the two lesion types with high probability was selected using receiver operating characteristics (ROC) analysis and calculation of error matrices.

Results

In this study, a total of 34 patients, 17 (50%) of whom had NFPAs and 17 (50%) had hypophysitis, were evaluated. Of the 17 patients with hypophysitis, 14 (82%) had primary hypophysitis and three (18%) had secondary hypophysitis. One case of secondary hypophysitis was Erdheim-Chester disease, one was IgG4-related hypophysitis, and the other

was cancer immunotherapy-associated hypophysitis. Nine (52.9%) of the patients with hypophysitis were female and eight (47.1%) were male, and eight (47.9%) of the patients with NFPA were female and nine (52.9%) were male. There was no significant difference between the groups in terms of sex ($p = 0.732$). The age at the diagnosis of NFPA was found to be significantly higher, and the median age at diagnosis of hypophysitis was 27.0 [range 23.0–43.0] years, and the median age at diagnosis of NFPA was 50.0 [range, 45.0–56.0] years ($p < 0.001$).

Among the 38 radiomic parameters obtained from T1A-C images, 10 parameters were selected that could differentiate the lesions. These 10 tissue features ($p < 0.01$) are listed in Fig. 2 in order of importance. To avoid the effect of confounding factors and to increase the performance of the classifiers, the parameters with high correlation (correlation coefficient > 0.7) with each other were eliminated using the independent component analysis method. The differences between hypophysitis and NFPA were significant for gray-level run-length matrix-low gray level run emphasis (GLRLMLGLRE), gray-level co-occurrence matrix-correlation, and gray-level co-occurrence entropy. Machine learning algorithms were performed with the combination of these three selected texture features (Fig. 3).

Error matrices were calculated by using the machine learning algorithm's decision tree classifier, Bayesian classifier, K-nearest neighbor algorithm, and SVM using the Matlab software. It was seen that SVM (Sensitivity: 0.74, specificity: 0.97, PPV: 0.96, NPV: 0.78, accuracy: 0.85, MCC: 0.72) showed the best performance in distinguishing the two lesion types using ROC analysis (Fig. 4) and
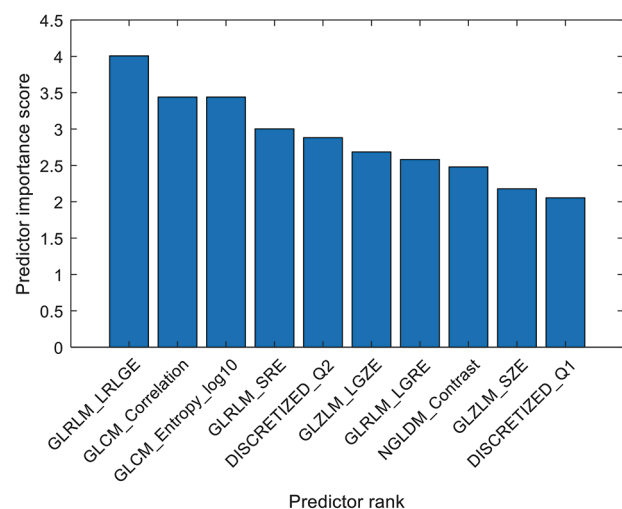

Fig. 2 Ten tissue features used to differentiate hypophysitis from non-functioning pituitary adenoma. Univariate chi-square test is used to calculated the importance score of each texture feature. The values in scores are the negative logs of the p-values
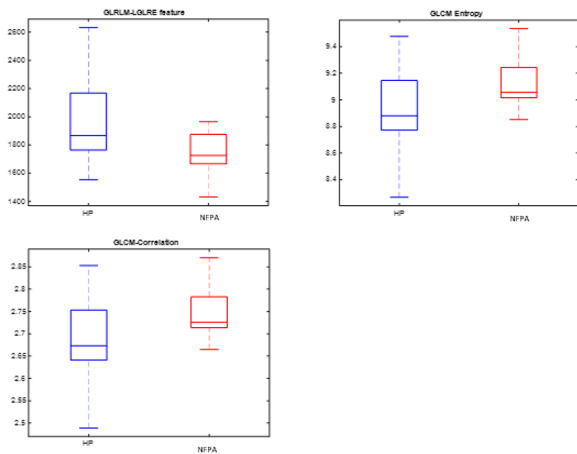
Fig. 3 Boxplots of three features with the highest predictor importance score. GLRLM_LGLRE feature has a significantly higher mean in HP cases whereas GLCM Entropy and correlation is higher in NFPA
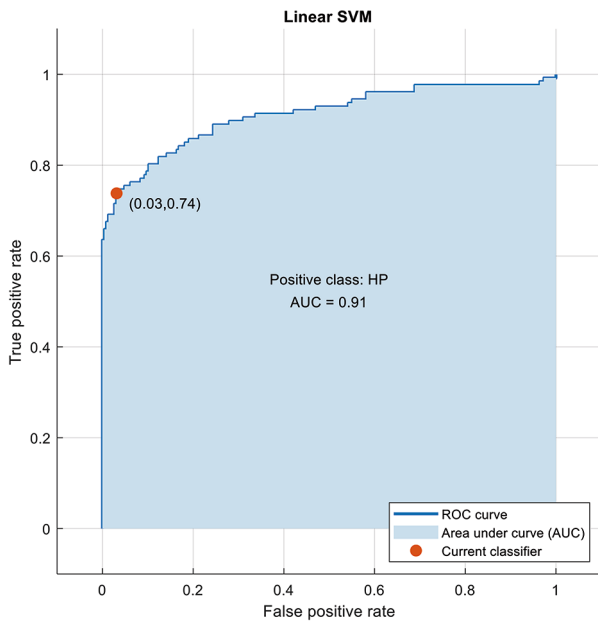

Fig. 4 The classification results of SVM with the three highest-ranked features. The linear SVM has the highest performance among all classifiers with AUC of 0.91. **HP**: Hypophysitis, **SVM**: Support vector machines, **AUC**: The area under the ROC curve
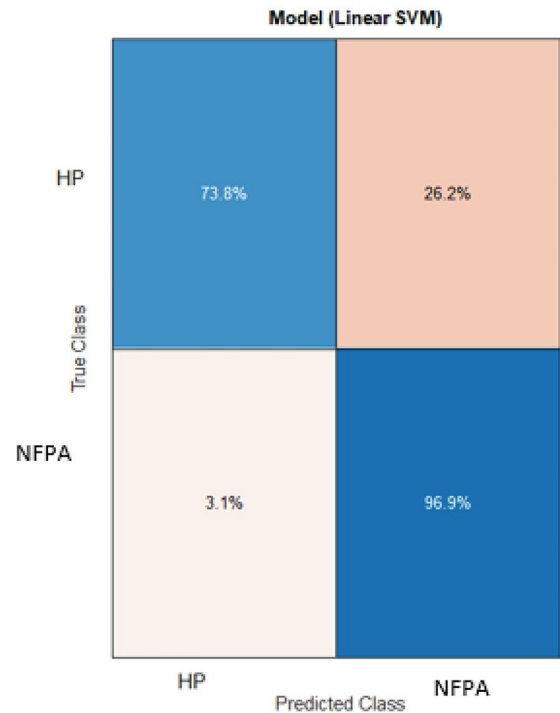

Fig. 5 The performance of the SVM model with confusion matrix. The algorithm differentiate hypophysitis from nonfunctioning adenoma with a sensitivity = 0.74, specificity = 0.97, precision = 0.96, negative predictive value = 0.78, false positive rate = 0.03, false discovery rate = 0.039, false negative rate = 0.04, accuracy = 0.85. Matthews Correlation Coefficient = 0.72. **HP**: Hypophysitis, **NFPA**: Non-functioning pituitary adenoma, **SVM**: Support vector machines

Discussion

This study reports that texture analysis-based machine learning tools show feasible performance in discriminating hypophysitis from NFPAs on post-Gd T1-W MRI. In this discrimination, three texture features, GLRLMLGLRE, gray-level co-occurrence matrix-correlation, and gray-level co-occurrence entropy, were significant. SVM successfully differentiated the two lesions using ROC analysis and confusion matrix criteria.

Patients with hypophysitis may be misdiagnosed as having NFPA and may even undergo surgery [12–14]. Although some radiologic and biochemical methods have been developed [8], a non-invasive method that clearly distinguishes hypophysitis from NFPA has not been identified at present. Gutenberg et al. reported a new radiologic score to distinguish autoimmune hypophysitis from NFPA [8]. Although the use of this score provides a significant improvement in the differential diagnosis of hypophysitis and NFPA, discrimination is still problematic [2]. For this reason, ML has recently gained an important place in the differential diagnosis of hypophysitis and NFPA. In our study, SVM was

confusion matrix criteria (Fig. 5). In contrast, K-nearest neighbor has the least performance in distinguishing the two lesion types (Sensitivity: 0.64, specificity: 0.52, PPV: 0.57, NPV: 0.60, accuracy: 0.58, MCC: 0.11). The results of the prediction performance of machine-learning classifiers are shown in Table 1.

Table 1 The prediction performances of machine-learning classifiers

Method	Sensitivity	Specificity	PPV	NPV	Accuracy	MCC
Decision Tree	0.76	0.58	0.65	0.71	0.67	0.35
Linear SVM	0.74	0.97	0.96	0.78	0.85	0.72
Kernel Naïve Bayes	0.70	0.52	0.60	0.64	0.61	0.23
KNN	0.64	0.52	0.57	0.60	0.58	0.11

PPV: Positive predictive values, **NPV**: Negative predictive values, **MCC**: Matthews Correlation Coefficient, **KNN**: K-nearest neighbors

able to successfully differentiate hypophysitis from NFPA. However, due to the rarity of hypophysitis, the number of our patients was small, but with larger studies to be done, the place of ML in the differentiation of hypophysitis and NFPA may be stronger.

ML is used successfully in diagnosis and predicting the outcomes of post-treatment of diseases in the sellar region. Kitajima et al. used artificial intelligence (AI), and they reported that their artificial neural network showed high performance in differentiating pituitary adenoma, craniopharyngioma, and Rathke's cleft cyst (area under the receiver operating characteristic curve, 0.990) [15]. In another study, Steiner et al. showed with AI that pituitary adenomas could be distinguished from normal pituitary tissue [16]. Ugga et al. evaluated the preoperative prediction of pituitary adenoma proliferative index class, and the authors demonstrated that ML is effective in predicting the pituitary macroadenomas ki-67 proliferation index class [17]. Zhang et al. reported that quantitative radiomic methods provided a preoperative prediction of NFPA subtypes, especially distinguishing between null cell adenomas and other subtypes, using T1 and CE-T1 images [18]. Cuocolo et al. evaluated the accuracy of tissue analysis combined with ML in the preoperative evaluation of pituitary macroadenoma consistency in patients undergoing endoscopic endonasal surgery. Trained on radiomic data extracted from T2-weighted MRI the ML model showed high accuracy in classifying soft and fibrous macroadenomas [19]. In our study, hypophysitis and NFPA were successfully differentiated using machine learning methods based on MRI-derived texture features.

In addition to being able to distinguish diseases, AI has also made significant advances in determining personalized treatments. Koçak et al. showed that the machine learning-based qTA of T2-weighted MRI performed better than the qualitative and quantitative relative signal intensity and immunohistochemical evaluations in predicting response to a somatostatin analog in patients with acromegaly [20]. In another study, Hollon et al. evaluated a large cohort of 400 patients with pituitary adenomas using machine learning and they were able to predict the early outcomes of pituitary adenoma surgery with 87% accuracy [21]. In our study, we could not evaluate the response to treatment, the parameters affecting the treatment, and the personalization of the treatment due to the small number of patients in the hypophysitis

group. However, larger-scale studies on this subject may contribute to the literature.

The main limitation of this study is the heterogeneity of the hypophysitis group. In addition, our study is retrospective in design and the number of patients is small.

In conclusion, we showed that machine learning tools had feasible performance in distinguishing hypophysitis from NFPA on post-Gd T1-W. With the development of this tool, patients with hypophysitis can be diagnosed using a noninvasive method and unnecessary surgeries can be avoided.

Declarations

References

1. Joshi MN, Whitelaw BC, Carroll PV (2018) Hypophysitis: diagnosis and treatment. Eur J Endocrinol 179(3):R151–R163. https://doi.org/10.1530/EJE-17-0009
2. Giuseppe B, Maria IM, Antonio B, Dario G, Katherine E, Antonio B, Annamaria DB (2016) Revisitation of autoimmune hypophysitis: knowledge and uncertainties on pathophysiological and clinical aspects. Pituitary 19(6):625–642. https://doi.org/10.1007/s11102-016-0736-z
3. Buxton N, Robertson I, Powell M, Chatterjee K (2001) Lymphocytic and granulocytic hypophysitis: A single centre experience. Br J Neurosurg 15(3):242–245. https://doi.org/10.1080/02688690120057664
4. Caturegli P, Newschaffer C, Olivi A et al (2005) Autoimmune hypophysitis. Endocr Rev 26(5):599–614. https://doi.org/10.1210/er.2004-0011
5. Gubbi S, Hannah-Shmouni F, Stratakis CA, Koch CA (2018) Primary hypophysitis and other autoimmune disorders of the sellar and suprasellar regions. Reviews in Endocrine and Metabolic Disorders 19(4):335–347. https://doi.org/10.1007/s11154-018-9480-1
6. Daly AF, Rixhon M, Adam C, Dempegioti A, Tichomirowa MA, Beckers A (2006) High prevalence of pituitary adenomas: A

cross-sectional study in the province of Liège, Belgium. J Clin Endocrinol Metab 91(12):4769–4775. https://doi.org/10.1210/jc.2006-1668

7. Ezzat S, Asa SL, Couldwell WT, Barr CE, Dodge WE, Vance ML, McCutcheon IE (2004) The prevalence of pituitary adenomas: A systematic review. Cancer 101(3):613–619. https://doi.org/10.1002/cncr.20412

8. Gutenberg A, Larsen J, Lupi I, Rohde V, Caturegli P (2009) A radiologic score to distinguish autoimmune hypophysitis from nonsecreting pituitary adenoma preoperatively. Am J Neuroradiol AJNR Am J Neuroradiol 30(9):1766–7172. https://doi.org/10.3174/ajnr.A1714

9. Ferrante E, Ferraroni M, Castrignanò T, Menicatti L, Anagni M, Reimondo G, Del Monte P, Bernasconi D, Loli P, Faustini-Fustini M, Borretta G, Terzolo M, Losa M, Morabito A, Spada A, Beck-Peccoz P, Lania AG (2006) Non-functioning pituitary adenoma database: A useful resource to improve the clinical management of pituitary tumors. Eur J Endocrinol 155(6):823–829. https://doi.org/10.1530/eje.1.02298

10. Dekkers OM, Pereira AM, Roelfsema F, Voormolen JHC, Neelis KJ, Schroijen MA, Smit JWA, Romijn JA (2006) Observation alone after transsphenoidal surgery for nonfunctioning pituitary macroadenoma. J Clin Endocrinol Metab 91(5):1796–1801. https://doi.org/10.1210/jc.2005-2552

11. Nioche C, Orlhac F, Boughdad S, Reuze S, Goya-Outi J, Robert C, Pellot-Barakat C, Soussan M, Frouin F, erique, Buvat I (2018) Lifex: A freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. Cancer Res 78(16):4786–4789. https://doi.org/10.1158/0008-5472.CAN-18-0125

12. Angelousi A, Cohen C, Sosa S, Danilowicz K, Papanastasiou L, Tsoli M, Pal A, Piaditis G, Grossman A, Kaltsas G (2018) Clinical, Endocrine and Imaging Characteristics of Patients with Primary Hypophysitis. Horm Metab Res 50(4):296–302. https://doi.org/10.1055/s-0044-101036

13. Korkmaz OP, Sahin S, Ozkaya HM, Apaydin T, Durmaz ES, Haliloglu O, Durcan E, Kadioglu P (2021) Primary hypophysitis: Experience of a Single Tertiary Center. Exp Clin Endocrinol Diabetes 129(1):14–21. https://doi.org/10.1055/a-0919-4388

14. Oguz SH, Soylemezoglu F, Sendur SN, Mut M, Oguz KK, Dagdelen S, Erbas T (2020) Clinical Characteristics, Management, and Treatment Outcomes of Primary Hypophysitis: A Monocentric Cohort. Horm Metab Res 52(4):220–227. https://doi.org/10.1055/a-1113-7777

15. Kitajima M, Hirai T, Katsuragawa S, Okuda T, Fukuoka H, Sasao A, Akter M, Awai K, Nakayama Y, Ikeda R, Yamashita Y, Yano S, Kuratsu J, Doi K (2009) Differentiation of Common Large Sellar-Suprasellar Masses. Effect of Artificial Neural Network on Radiologists' Diagnosis Performance. Acad Radiol 16(3):313 – 20. https://doi.org/10.1016/j.acra.2008.09.015

16. Steiner G, Mackenroth L, Geiger KD, Stelling A, Pinzer T, Uckermann O, Sablinskas V, Schackert G, Koch E, Kirsch M (2012) Label-free differentiation of human pituitary adenomas by FT-IR spectroscopic imaging. Anal Bioanal Chem 403(3):727–735. https://doi.org/10.1007/s00216-012-5824-y

17. Ugga L, Cuocolo R, Solari D, Guadagno E, D'Amico A, Somma T, Cappabianca P, Del Basso de Caro ML, Cavallo LM, Brunetti A (2019) Prediction of high proliferative index in pituitary macroadenomas using MRI-based radiomics and machine learning. Neuroradiology 61(12):1365–1373. https://doi.org/10.1007/s00234-019-02266-1

18. Zhang S, Song G, Zang Y, Jia J, Wang C, Li C, Tian J, Dong D, Zhang Y (2018) Non-invasive radiomics approach potentially predicts non-functioning pituitary adenomas subtypes before surgery. Eur Radiol 28(9):3692–3701. https://doi.org/10.1007/s00330-017-5180-6

19. Cuocolo R, Ugga L, Solari D, Corvino S, D'Amico A, Russo D, Cappabianca P, Cavallo LM, Elefante A (2020) Prediction of pituitary adenoma surgical consistency: radiomic data mining and machine learning on T2-weighted MRI. Neuroradiology 62(12):1649–1656. https://doi.org/10.1007/s00234-020-02502-z

20. Kocak B, Durmaz ES, Kadioglu P, Polat Korkmaz O, Comunoglu N, Tanriover N, Kocer N, Islak C, Kizilkilic O (2019) Predicting response to somatostatin analogues in acromegaly: machine learning-based high-dimensional quantitative texture analysis on T2-weighted MRI. Eur Radiol 29(6):2731–2739. https://doi.org/10.1007/s00330-018-5876-2

21. Hollon TC, Parikh A, Pandian B, Tarpeh J, Orringer DA, Barkan AL, McKean EL, Sullivan SE (2018) A machine learning approach to predict early outcomes after pituitary adenoma surgery. Neurosurg Focus 45(5):E8. https://doi.org/10.3171/2018.8.FOCUS18268