



# Multivariable Diagnostic Prediction Model to Detect Hormone Secretion Profile From T2W MRI Radiomics with Artificial Neural Networks in Pituitary Adenomas

## Hipofiz Makroadenomlarında T2-MRG Tabanlı Radyomiks ve Sinir Ağları ile Hormon Salgılama Profiline Tahmini

© Begumhan BAYSAL, © Mehmet Bilgin ESER, © Mahmut Bilal DOĞAN, © Muhammet Arif KURSUN

Istanbul Goztepe Prof. Dr. Suleyman Yalcin City Hospital, Clinic of Radiology, Istanbul, Turkey

### ABSTRACT

**Objective:** This study aims to develop neural networks to detect hormone secretion profiles in the pituitary adenomas based on T2 weighted magnetic resonance imaging (MRI) radiomics.

**Methods:** This retrospective model-development study included a cohort of patients with pituitary adenomas (n=130) from January 2015 to January 2020 in one tertiary center. The mean age was 46.49±13.69 years, and 76/130 (58.46%) were women. Three observers segmented lesions on coronal T2 weighted MRI, and an interrater agreement was evaluated using the Dice coefficient. Predictors were determined as radiomics features (n=851). Feature selection was based on intraclass correlation coefficient, coefficient variance, variance inflation factor, and LASSO regression analysis. Outcomes were identified as 7 hormone secretion profiles [non-functioning pituitary adenoma, growth hormone-secreting adenomas, prolactinomas, adrenocorticotrophic hormone-secreting adenomas, pluri-hormonal secreting adenomas (PHA), follicle-stimulating hormone and luteinizing hormone-secreting adenomas, and thyroid-stimulating hormone adenomas]. A multivariable diagnostic prediction model was developed with artificial neural networks (ANN) for 7 outcomes. ANN performance was presented as an area under the receiver operating characteristic curve (AUC) and accepted as successful if the AUC was >0.85 and p-value was <0.01.

**Results:** The performance of the ANN distinguishing prolactinomas from other adenomas was validated (AUC=0.95, p<0.001, sensitivity: 91%, and specificity: 98%). The model distinguishing PHA had the lowest AUC (AUC=0.74 and p<0.001). The AUC values for the other five ANN were >0.85 and p values were <0.001.

**Conclusions:** This study was successful in training neural networks that could differentiate the hormone secretion profile of pituitary adenomas.

**Keywords:** Pituitary adenoma, magnetic resonance imaging, machine learning, artificial intelligence, radiomics

### ÖZ

**Amaç:** Bu çalışma, T2 ağırlıklı manyetik rezonans görüntüleme (MRG) radyomikslerine dayalı olarak hipofiz adenomlarında hormon salgılama profillerini tespit etmek için sinir ağları geliştirmeyi amaçlamaktadır.

**Yöntemler:** Bu retrospektif model geliştirme çalışması, üçüncü basamak bir merkezde Ocak 2015 ile Ocak 2020 arasında hipofiz adenomu olan hastalardan oluşan bir kohortu içermektedir (n=130). Ortalama hasta yaşı 46,49±13,69 yıldır ve 76/130'u (%58,46) kadındır. Üç gözlemci, koronal T2 ağırlıklı MRG'de lezyonları segmente etti ve Dice katsayısı kullanılarak gözlemciler arası uyum değerlendirildi. Prediktörler radyomiks parametreleri olarak belirlendi (n=851). Parametre seçimi, sınıf içi korelasyon katsayısına, katsayı varyansına, varyans inflasyon faktörüne ve LASSO regresyon analizine dayanmaktadır. Sonuçlar yedi farklı hormon salgılama profili olarak tanımlandı [non-fonksiyone hipofiz adenomu, büyüme hormonu salgılayan adenomlar, prolaktinomalar, adrenokortikotropik hormon salgılayan adenomlar, pluri-hormonal adenomlar (PLSA), folikül uyarıcı hormon ve luteinize edici hormon salgılayan adenomlar ile tiroid uyarıcı hormon salgılayan adenomlar]. Yedi hormon için yapay sinir ağları (YSA) ile çok değişkenli bir tanısal tahmin modeli geliştirildi. YSA performansı, alıcı işletim karakteristik eğrisinin altındaki alan (AUC) olarak sunuldu ve AUC >0,85 ve p değeri <0,01 başarılı kabul edildi.

**Bulgular:** YSA, AUC=0,95, p<0,001, duyarlılık: %91, özgüllük: %98 değerleri ile prolaktinomaları diğer adenomlardan ayırdı. PLSA için AUC=0,74 ve p<0,001'di. Diğer beş YSA için ise AUC değerleri >0,85 ve p<0,001 idi.

**Sonuçlar:** Bu çalışma, hipofiz adenomlarının hormon salgılama profilini ayırt edebilen sinir ağlarının eğitiminde başarılı olmuştur.

**Anahtar kelimeler:** Hipofiz adenomu, manyetik rezonans görüntüleme, makine öğrenimi, yapay zeka, radyomiks

**Presented in:** The article was presented at the 59<sup>th</sup> ASNR Annual Meeting in 2021.

**Address for Correspondence:** B. Baysal, Istanbul Goztepe Prof. Dr. Suleyman Yalcin City Hospital, Clinic of Radiology, Istanbul, Turkey

**E-mail:** baysalbegumhan@yahoo.com **ORCID ID:** orcid.org/0000-0003-0470-1683

**Received:** 03 December 2021

**Accepted:** 25 January 2022

**Online First:** 03 February 2022

**Cite as:** Baysal B, Eser MB, Dogan MB, Kursun MA. Multivariable Diagnostic Prediction Model to Detect Hormone Secretion Profile From T2W MRI Radiomics with Artificial Neural Networks in Pituitary Adenomas. Medeni Med J 2022;37:36-43

## INTRODUCTION

Pituitary adenoma is the second most common primary central nervous system tumor and constitutes approximately 14% of intracranial masses<sup>1-3</sup>. Pituitary adenomas are classified according to their size as microadenoma (<1 cm), macroadenoma (≥1 cm) and giant adenoma (>4 cm). Additionally, 54-62% of pituitary adenomas are active hormone secreting tumors [growth hormone-secreting adenomas (GHSA), prolactinomas, adrenocorticotrophic hormone-secreting adenomas (ACTHSA), pluri-hormonal adenomas (PHA), follicle-stimulating hormone and luteinizing hormone-secreting adenomas (FSH&LHSA), and thyroid-stimulating hormone adenomas (TSHA)], and 38-46% of them are non-functioning<sup>4-6</sup>. However, it is possible to determine the hormone secretion profile by using plasma hormone concentrations. Currently, due to increasing use of radiological imaging, many pituitary adenomas are detected incidentally<sup>7</sup>. For these tumors, it may be possible to estimate the hormone secretion profile at the time of imaging by exploiting the heterogeneity<sup>8</sup>.

Radiomics is a quantitative approach that extracts many image features from medical images and allows the development of diagnostic tools<sup>8</sup>. The success of this approach in determining tumor subtypes has been studied and confirmed in some other tumors<sup>9,10</sup>. In addition, in a limited number of recent studies, a model based on radiomics features was developed to predict tumor consistency in patients with pituitary adenoma<sup>11-15</sup>.

In prior radiomics studies, the stability of the radiomics feature was only evaluated at the level of interrater agreement with the intraclass correlation coefficient<sup>11</sup>. Therefore, this approach may be inadequate to detect stability of radiomics feature. However, the recent statement offered that stable features also should have high precision and accuracy<sup>16</sup>. Therefore, creating diagnostic models based on stable radiomics features may positively affect reproducibility, precision, and accuracy.

This study aims to develop neural networks to detect hormone secretion profiles in the pituitary adenomas based on T2 weighted magnetic resonance imaging (MRI) radiomics.

## MATERIALS and METHODS

### Ethical Considerations

This retrospective model-development study was done after it was approved by the Istanbul Medeniyet University Goztepe Prof. Dr. Suleyman Yalcin Training and

Research Hospital Clinical Research Ethics Committee (decision no: 2020/0304, date: 05.18.2020), and written informed consent was waived. The STARD 2015 statement was followed to document the study, and white papers and statements of multiple societies were followed<sup>17-21</sup>. This study was scored (18/36) with a radiomics quality score<sup>17</sup>.

### Study Population and Data Collection

This model-development study was carried out in a single tertiary-care center. From the patients documented between January 2015 to January 2020, 130 patients who met the inclusion criteria were included in the study. Inclusion criteria were determined as compliance with the following criteria: 1. The MRI, including T2W sequences, of the patient must be present. 2. Image quality should be sufficient to allow segmentation. Patients diagnosed in our center, but whose imaging was performed in another center were excluded. The MRI protocol is described in Table 1.

### Predictors: Analysis of the T2W Images

Three radiologists with 8 years, 3 years, and 1 year of experience performed segmentation using 3D Slicer software, version 4.10.2 (<https://www.slicer.org>). Segmentation was done volumetrically on T2W images. The 851 radiomics feature, which is the predictor of this study, was extracted with the PyRadiomics (version 2.2.0). All the features (shape, first order, and high order) in this module were selected. Resampling was done, normalization was enabled, and wavelet-based filters were activated (Figure 1).

### Outcomes

Outcomes were identified as 7 hormone secretion profiles [non-functioning pituitary adenoma (n=19), GHSA (n=21), prolactinomas (n=64), ACTHSA (n=6), PHA (n=6), FSA&LHSA (n=8), and TSHA (n=6)].

### Features Stability Analyses: Interobserver Agreement Evaluation and Coefficient of Variation Analysis

Segmentations and radiomic features were separately assessed for interobserver agreement. For segmentation, the Dice similarity coefficient was used to measure interobserver reliability, while intraclass correlation coefficient (ICC - 3,k), two-way random effects model, and absolute agreement were used for radiomics features<sup>22</sup>. Features with an ICC>0.75 were included in the coefficient of variation (CoV) analysis, with those presenting >15% variances being eliminated<sup>16</sup>. The predictor features that passed the CoV analysis were subjected to Spearman's

Table 1. Magnetic resonance imaging protocol for pituitary gland used in the study.			
	TIW	T2W-FSE	Dynamic-GRE
Plane	Sagittal + coronal	Sagittal + coronal	Coronal
Fat suppression	-	-	-
Repetition time (ms)	380	4,334	365
Echo time (ms)	Min full*	102	Min full*
Flip angle (°)	160	160	160
Slice thickness (mm)	3	2	3
Field of view (mm)	160x160	160x160	160x160
Bandwidth (Hz)	20.83	23.82	23.82
Matrix (mm x mm)	288x288	288x288	288x160
Scan time (ms)	139	178	320
Time after contrast injection (sec)	-	-	0, 30, 45, 75, 120, 180, 240, 300

TIW: T1 weighted, T2W: T2 weighted, FSE: Fast spin echo, GRE: Gradient echo. This study was conducted with two different magnetic resonance imaging devices [1.5 Tesla GE Optima MR450w and MR360 (General Electric, Chicago, Illinois, USA)]. Also, three different gadolinium contrast agents were used. \*Indicates 4-16 ms.

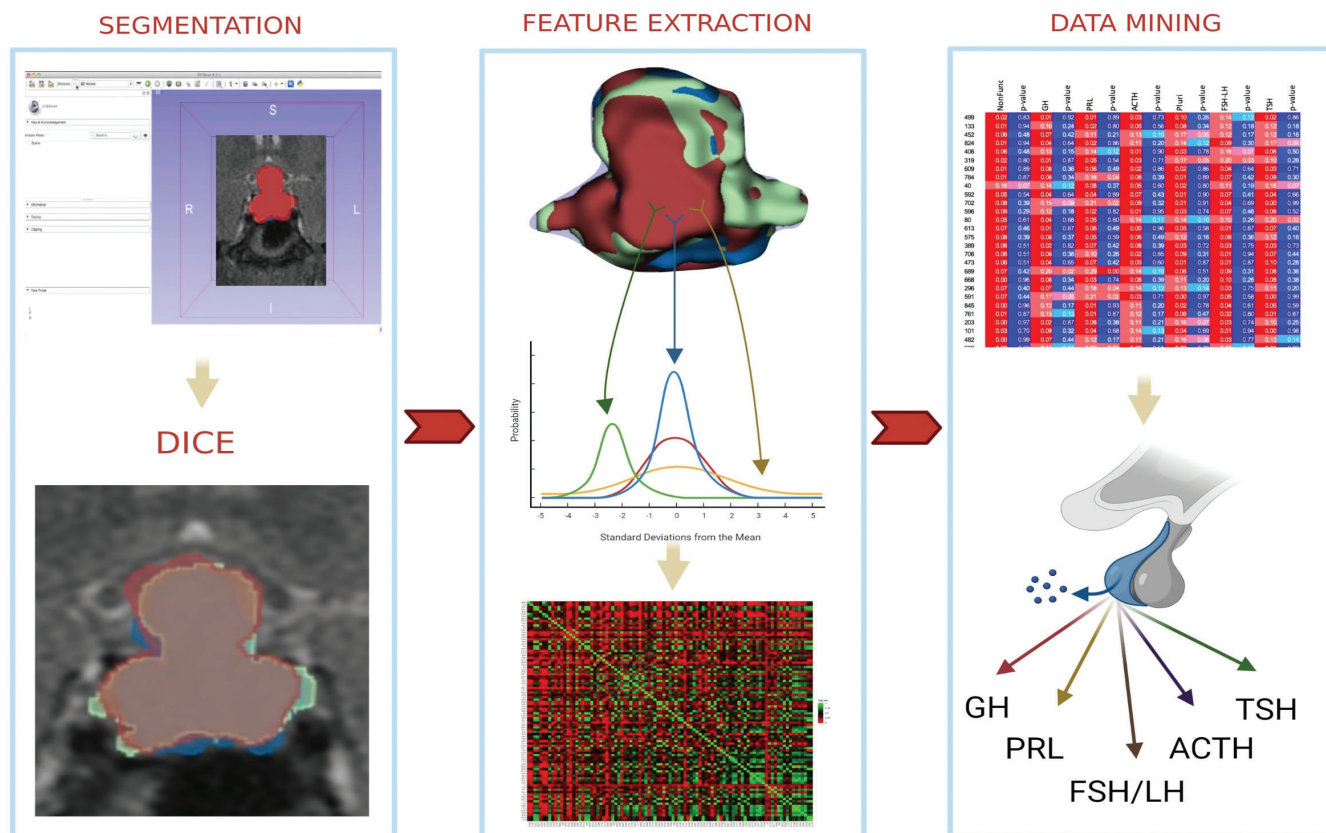


Figure 1. Pipeline of the study.

GH: Growth hormone, ACTH: Adrenocorticotropic hormone, TSH: Thyroid-stimulating hormone, FSH/LH: Follicle-stimulating hormone/ luteinizing hormone, PRL: Prolactin

correlation (SC) analysis, and correlation matrixes were performed for variance inflation factor (VIF) analysis.

**Features Selection Analyses: Collinearity-multicollinearity Evaluation and Least Absolute Shrinkage and Selection Operator Regression**

VIF analyses were performed to reduce the collinearity-multicollinearity using the formula  $1/1-R^2$ . If the VIF was above 10, the feature was eliminated<sup>23</sup>. The features with smaller CoV were preserved in this elimination process. Further, validated imaging biomarkers were evaluated using SC analysis between features and outcomes ( $p < 0.01$ ).

Features were selected with the least absolute shrinkage and selection operator (LASSO) with L1 normalization. Random sampling and 5-fold cross-validation were used for seeding LASSO.

**Structuring Artificial Neural Networks**

For training, networks of multilayer perceptron and radial basis function were selected. The software appointed the number of layers, the number of neurons, error function, hidden activation, and output activation in these models. The software used random number generator for sampling 70% of the patients as train, 15% as a test, and 15% as a validation (hold-out) set for each training session of neurons. These subgroups were in a similar distribution in terms of predictors and outcomes. Hyperparameter tuning was made with the “early stopping” algorithm. The “Early stopping” algorithm trains the neural networks with the “training” set and performs hyperparameter tuning with the “test” set at the end of each epoch. Neuron training continues as long

as the error rate decreases in both groups. The training is terminated when the error rate starts to increase in the “test” set. Finally, neuron performance is measured with the validation (hold-out) set.

**Statistical Analysis**

Statistical analyses and neural network development were performed using the TIBCO Statistica version 13.0.5 (TIBCO Software, Palo Alto, CA). Neural network results with the highest diagnostic accuracy are presented with area under the receiver operating characteristic curve (AUC) with 95% confidence intervals. In receiver operating curve analysis, if AUC was  $> 0.85$  and p-value was  $< 0.01$ , then it was considered a validated classifier neural network<sup>16</sup>.

**RESULTS**

**Patient’s Characteristics**

This study included 130 consecutive patients with pituitary adenoma. The mean age was  $46.49 \pm 13.69$  years, and 76/130 (58.46%) were women. All patients were Caucasians. A full summary of clinicopathologic characteristics of the patients is presented in Table 2.

**Model Development and Specification**

The interobserver median Dice coefficient values for segmentations were 0.84 [interquartile range (IQR): 0.06] between observers 1 and 2; 0.84 (IQR: 0.17) between observers 1 and 3; and 0.79 (IQR: 0.20) between observers 2 and 3.

The 204 features were eliminated by using ICC ( $< 0.75$ ). By using CoV analysis ( $> 0.15$ ), 552 features were eliminated. Finally, another 44 features were eliminated

**Table 2. Characteristics of the participants.**

Characteristics	All patients (n=130)
Age (mean ± SD, year)	46.49±13.69
Sex (female/male; n, %)	76/54 (58.5%/41.5%)
Median tumor size cm <sup>3</sup> (IQR)	0.92 (2.41)
<b>Hormonal subtypes</b>	
Non-functioning pituitary adenomas, (n, %)	19, 14.6%
Growth hormone-secreting adenomas, (n, %)	21, 16.2%
Prolactin secreting adenomas, (n, %)	64, 49.2%
Adrenocorticotrophic hormone-secreting adenomas, (n, %)	6, 4.6%
Pluri-hormonal adenomas, (n, %)	6, 4.6%
Follicle-stimulating hormone and luteinizing hormone-secreting adenomas, (n, %)	8, 6.2%
Thyroid-stimulating hormone secreting adenomas (n, %)	6, 4.6%
SD: Standard deviation, IQR: Interquartile range	

by using VIF analysis due to collinearity. Most of the radiomics features were found to be unstable (n=800, 94%).

Stable predictors (n=51) and all outcomes were used for correlation analysis, and correlation matrixes were created to evaluate the unadjusted relation between each candidate predictor and outcomes (Figure 2). In this analysis, all SC coefficients were below 0.30, with p<0.01 for only five predictors (Figure 3). Finally, LASSO

regression was used for regularization, and the most relevant predictors were selected for neural network training.

### Diagnostic Prediction Model Results

The performance of the ANN distinguishing prolactinomas from other adenomas was validated (AUC=0.95, p<0.001, sensitivity: 91%, and specificity: 98%). The model distinguishing PHA had the lowest AUC (AUC=0.74 and p<0.001). Results of seven neural networks are presented in Table 3.

### DISCUSSION

The most obvious result of this study was that prolactinomas, which were found in about half of the included patients, were predicted with high accuracy based on the heterogeneity in the T2W MRI images. However, the model distinguishing PHA had the lowest AUC. Difficulty in distinguishing these tumors with more than one cell group suggests that the results are not random and related to tissue heterogeneity.

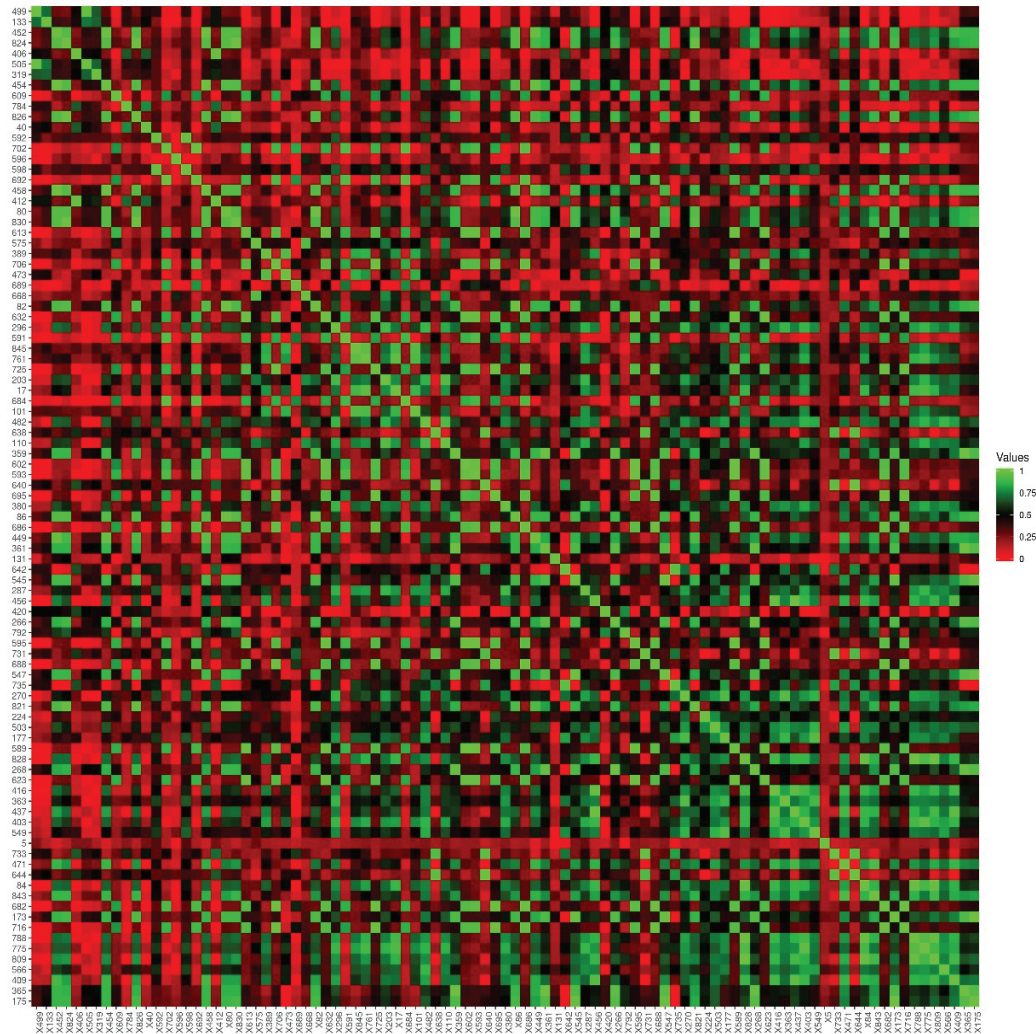
There are limited studies in the literature on the classification of pituitary adenomas from MRI images<sup>11-15</sup>. The four of these studies investigated surgical consistency after surgical excision of adenomas<sup>11,13-15</sup>. In the study, which included 89 macroadenomas, Cuocolo et al.<sup>11</sup> predicted 28 patients' outcomes in the test group, and only two soft tumors were misclassified as fibrous tumors. However, all fibrous tumors were correctly classified. Fan et al.<sup>14</sup> reported that adding clinical data such as age, sex, hormone levels to the model improved the model's accuracy. These results meant that patients who might require re-surgery were identified by imaging the early phase of the disease. This information can make the surgeon confident for surgical planning and reduce residuals and recurrence rates. A second benefit is that the patient can be informed that the tumor is consistent and may need re-surgery in the future. In another study, Peng et al.<sup>12</sup> used T1W, contrast-enhanced T1W, and T2W MRI images and three different machine learning algorithms, and they predicted three different immunohistochemical classes of pituitary adenomas preoperatively. They observed that T2W radiomics based model's accuracy was the highest. The best classifier was the support vector machine. Considering these results, we did our study with T2W radiomics features and pre-trained neural networks.

Currently, radiomics studies are facing a reproducibility crisis. Therefore, the European Society of Radiology (ESR) has recently presented the statement for imaging biomarkers stability such as radiomics<sup>16</sup>.

	NonFunc	p-value	GH	p-value	PRL	p-value	ACTH	p-value	Pluri	p-value	FSH-LH	p-value	TSH	p-value
499	0.02	0.83	0.01	0.92	0.01	0.89	0.03	0.73	0.10	0.26	0.14	0.12	0.02	0.86
133	0.01	0.94	0.10	0.24	0.02	0.80	0.05	0.56	0.08	0.34	0.12	0.18	0.12	0.18
452	0.06	0.48	0.07	0.42	0.11	0.21	0.13	0.15	0.17	0.05	0.12	0.17	0.12	0.16
824	0.01	0.94	0.04	0.64	0.02	0.86	0.11	0.20	0.14	0.12	0.09	0.30	0.17	0.06
406	0.06	0.48	0.13	0.15	0.14	0.12	0.01	0.90	0.03	0.78	0.16	0.07	0.06	0.50
319	0.02	0.80	0.01	0.87	0.05	0.54	0.03	0.71	0.17	0.05	0.20	0.03	0.10	0.26
609	0.01	0.89	0.08	0.38	0.06	0.49	0.02	0.86	0.02	0.86	0.04	0.84	0.03	0.71
784	0.01	0.87	0.08	0.34	0.18	0.04	0.08	0.39	0.01	0.89	0.07	0.42	0.09	0.30
40	0.16	0.07	0.14	0.12	0.08	0.37	0.05	0.60	0.02	0.80	0.11	0.19	0.16	0.07
592	0.05	0.54	0.04	0.64	0.04	0.69	0.07	0.43	0.01	0.90	0.07	0.41	0.04	0.66
702	0.08	0.39	0.15	0.09	0.21	0.02	0.09	0.32	0.01	0.91	0.04	0.69	0.00	0.99
596	0.09	0.29	0.12	0.18	0.02	0.82	0.01	0.95	0.03	0.74	0.07	0.46	0.06	0.52
80	0.05	0.61	0.04	0.68	0.05	0.60	0.14	0.11	0.14	0.10	0.10	0.26	0.20	0.02
613	0.07	0.46	0.01	0.87	0.06	0.49	0.00	0.96	0.05	0.58	0.01	0.87	0.07	0.40
575	0.08	0.39	0.08	0.37	0.05	0.59	0.06	0.49	0.12	0.16	0.08	0.36	0.12	0.16
389	0.06	0.51	0.02	0.82	0.07	0.42	0.08	0.39	0.03	0.72	0.03	0.75	0.03	0.73
706	0.08	0.51	0.08	0.38	0.10	0.26	0.02	0.85	0.09	0.31	0.01	0.94	0.07	0.44
473	0.06	0.51	0.04	0.65	0.07	0.42	0.05	0.60	0.01	0.87	0.01	0.87	0.10	0.28
689	0.07	0.42	0.20	0.02	0.29	0.00	0.14	0.10	0.06	0.51	0.09	0.31	0.08	0.36
668	0.00	0.96	0.08	0.34	0.03	0.74	0.08	0.39	0.11	0.20	0.10	0.26	0.08	0.38
296	0.07	0.40	0.07	0.44	0.18	0.04	0.14	0.12	0.13	0.14	0.03	0.75	0.11	0.20
591	0.07	0.44	0.17	0.05	0.21	0.02	0.03	0.71	0.00	0.97	0.05	0.58	0.00	0.99
845	0.00	0.96	0.12	0.17	0.01	0.93	0.11	0.20	0.02	0.78	0.04	0.61	0.05	0.59
761	0.01	0.87	0.13	0.13	0.01	0.87	0.12	0.17	0.06	0.47	0.02	0.80	0.01	0.87
203	0.00	0.97	0.02	0.87	0.08	0.38	0.11	0.21	0.16	0.07	0.03	0.74	0.10	0.25
101	0.03	0.70	0.09	0.32	0.04	0.68	0.14	0.13	0.04	0.69	0.01	0.94	0.00	0.98
482	0.00	0.99	0.07	0.44	0.12	0.17	0.11	0.21	0.16	0.08	0.03	0.77	0.13	0.14
638	0.02	0.83	0.14	0.11	0.22	0.01	0.06	0.51	0.02	0.78	0.13	0.15	0.01	0.93
110	0.00	0.99	0.01	0.89	0.07	0.45	0.11	0.20	0.16	0.07	0.02	0.85	0.11	0.20
359	0.00	0.99	0.09	0.30	0.07	0.42	0.11	0.20	0.15	0.09	0.09	0.33	0.15	0.09
380	0.09	0.33	0.03	0.76	0.17	0.06	0.12	0.17	0.07	0.41	0.05	0.56	0.05	0.58
131	0.05	0.56	0.11	0.21	0.04	0.68	0.05	0.55	0.01	0.94	0.02	0.84	0.09	0.33
642	0.04	0.69	0.11	0.21	0.14	0.12	0.05	0.55	0.08	0.39	0.11	0.22	0.06	0.49
545	0.01	0.95	0.07	0.44	0.04	0.64	0.07	0.42	0.14	0.12	0.08	0.36	0.13	0.15
287	0.02	0.79	0.02	0.85	0.06	0.47	0.13	0.15	0.09	0.32	0.07	0.44	0.08	0.39
456	0.06	0.51	0.09	0.32	0.12	0.18	0.11	0.20	0.11	0.22	0.12	0.19	0.05	0.54
420	0.04	0.62	0.13	0.15	0.01	0.92	0.01	0.95	0.06	0.52	0.28	0.00	0.06	0.50
792	0.03	0.74	0.01	0.91	0.07	0.45	0.07	0.45	0.13	0.15	0.20	0.02	0.09	0.30
731	0.07	0.43	0.15	0.09	0.29	0.00	0.09	0.32	0.03	0.76	0.15	0.08	0.01	0.93
270	0.01	0.93	0.09	0.33	0.13	0.16	0.11	0.23	0.16	0.08	0.02	0.86	0.11	0.21
224	0.06	0.52	0.02	0.83	0.10	0.28	0.10	0.27	0.13	0.16	0.09	0.32	0.02	0.79
503	0.00	0.98	0.12	0.16	0.12	0.16	0.11	0.19	0.14	0.12	0.08	0.36	0.08	0.34
177	0.01	0.91	0.09	0.32	0.10	0.27	0.08	0.36	0.13	0.13	0.00	1.00	0.12	0.18
828	0.03	0.74	0.03	0.78	0.00	1.00	0.14	0.11	0.11	0.22	0.07	0.42	0.07	0.40
363	0.06	0.52	0.16	0.07	0.21	0.02	0.12	0.18	0.11	0.21	0.00	0.99	0.11	0.21
403	0.06	0.50	0.06	0.51	0.16	0.07	0.13	0.13	0.15	0.08	0.03	0.75	0.08	0.39
549	0.00	0.97	0.16	0.08	0.15	0.08	0.07	0.40	0.13	0.15	0.00	0.99	0.11	0.19
5	0.24	0.01	0.05	0.54	0.24	0.01	0.06	0.52	0.04	0.68	0.10	0.27	0.06	0.52
471	0.05	0.58	0.13	0.14	0.14	0.10	0.12	0.18	0.15	0.09	0.11	0.22	0.11	0.20
843	0.04	0.64	0.07	0.42	0.02	0.83	0.10	0.24	0.14	0.11	0.13	0.15	0.11	0.21
566	0.03	0.76	0.09	0.29	0.11	0.22	0.06	0.50	0.12	0.17	0.02	0.79	0.07	0.45

**Figure 2.** Correlation matrix between predictors and outcomes.

GH: Growth, hormone, PRL: Prolactin, ACTH: Adrenocorticotrophic hormone, FSH/LH: Follicle-stimulating hormone/luteinizing hormone, TSH: Thyroid-stimulating hormone



**Figure 3.** Heatmap of the predictors. Each predictor coded with a variable number and an available list of variables in a supplemental file. With this Spearman rank correlation analysis, this heatmap created the high collinear variables eliminated by VIF analyses.

VIF: Varince inflation factor

Cuocolo et al.<sup>11</sup> and Zeynalova et al.<sup>13</sup> evaluated the reproducibility of radiomics features by using ICC and included the features with ICC>0.75 and ICC>0.90, respectively. Peng et al.<sup>12</sup>, Fan et al.<sup>14</sup>, and Rui et al.<sup>15</sup> did not evaluate the reproducibility of radiomics features. In this study, we followed the ESR statement to evaluate the feature’s stability. Therefore, we eliminated high variance features by using CoV and high collinear features by using VIF analysis<sup>16</sup>. Although Cuocolo et al.<sup>11</sup> did not accept variance and collinearity as a criterion of stability, they also eliminated these features similar to our study.

The incidence of incidental adenoma is increasing due to the increasing frequency of imaging<sup>7</sup>. Detecting

these lesions’ secretion profiles and consistency at the time of imaging can be beneficial for accelerating patient management. Due to several studies on tumor stiffness and consistency, we focused on the secretion profile in this study<sup>11,13-15</sup>. We hypothesized that the cells that determined the secretion profile could be detected by quantitative analysis in this study and we thought that estimating PHA with the lowest accuracy while estimating prolactinoma with the highest accuracy confirmed this hypothesis. Because each of the pluri-hormonal tumors has different amounts of different secretory types of cells, this condition restricts imaging profiling whereas imaging profiling in a tumor containing a single type of cell, such as a prolactinoma, is succesful.

**Table 3. Neural networks performance results.**

Outcomes	Non-functioning pituitary adenoma	Growth hormone secreted adenoma	Prolactinoma	ACTH secreted adenoma	Pluri-hormonal adenoma	FSH-LH secreted adenoma	TSH secreted adenoma
Training algorithm	BFGS 0	BFGS 0	BFGS 68	BFGS 0	BFGS 0	BFGS 13	BFGS 0
Error function	Entropy	Entropy	Entropy	Entropy	SOS	Entropy	Entropy
Hidden activation	Tahn	Tahn	Tahn	Tahn	Tahn	Tahn	Tahn
Output activation	Softmax	Softmax	Softmax	Softmax	Exponential	Softmax	Softmax
AUC*	0.87	0.89	0.95	0.94	0.74	0.96	0.95
CI 95% lower	0.80	0.83	0.91	0.89	0.57	0.92	0.94
CI 95% upper	0.94	0.95	0.99	0.99	0.91	0.99	0.99
Acc (%)	94	92	95	97	97	96	95
Sen (%)	74	67	83	83	50	50	100
Spec (%)	97	97	98	98	99	99	95
PPV (%)	82	82	98	63	75	80	50
NPV (%)	96	94	92	99	98	97	100

\*p<0.001. AUC: Area under the receiver operating characteristic curve, CI: Confidence interval, ACTH: Adrenocorticotrophic hormone, FSH-LH: Follicle-stimulating hormone and luteinizing hormone, TSH: Thyroid-stimulating hormone, Sen: Sensitivity, Spec: Specificity, PPV: Positive predictive value, NPV: Negative predictive value, BFGS: Broyden-Fletcher-Goldfarb-Shanno algorithm, Acc: Accuracy

This study had several limitations. First, prolactinomas were found in half of the patients, and this neural network trained balanced distribution; however other networks have not. Second, the ground truth was hormone plasma levels because our patient population was consisted of patients admitting to the outpatient clinic of endocrinology. Third, the study was single-centered. However, radiomics features were subjected to rigorous stability analyses to increase reproducibility and precision, and the internal validation methods were used in training neural networks to increase accuracy.

**CONCLUSIONS**

Soon, this study and previous studies will become parts of a complex web and accumulate, allowing us to obtain much more quantitative data on patients than current. Until then, we need to increase our quantitative data and closely test our imaging biomarkers’ reproducibility, precision, and accuracy. This study shows that the ANN distinguishes with 95% accuracy whether a pituitary adenoma is a prolactinoma.

**Ethics**

**Ethics Committee Approval:** This retrospective model-development study was done after it was approved by the Istanbul Medeniyet University Goztepe Prof. Dr. Suleyman Yalcin Training and Research

Hospital Clinical Research Ethics Committee (decision no: 2020/0304, date: 05.18.2020).

**Informed Consent:** Written informed consent was waived.

**Peer-review:** Externally and internally peer-reviewed.

**Author Contributions**

Surgical and Medical Practices: B.B., M.B.E., M.B.D., M.A.K., Concept: B.B., M.B.E., M.B.D., M.A.K., Design: B.B., M.B.E., M.B.D., Data Collection and/or Processing: B.B., M.B.D., M.A.K., Analysis and/or Interpretation: B.B., M.B.E., Literature Search: B.B., M.B.E., M.B.D., Writing: B.B., M.B.E., M.B.D., M.A.K.

**Conflict of Interest:** The authors have no conflict of interest to declare.

**Financial Disclosure:** The authors declared that this study has received no financial support.

**REFERENCES**

1. Dolecek TA, Propp JM, Stroup NE, Kruchko C. CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2005-2009. *Neuro Oncol.* 2012;14(Suppl 5):1-49.
2. Ezzat S, Asa SL, Couldwell WT, et al. The prevalence of pituitary adenomas: A systematic review. *Cancer.* 2004;101:613-9.

## B. Baysal et al. Predicting Hormone Secretion of Macroadenomas from Radiomics

3. Guo S, Wang Z, Kang X, Xin W, Li X. A Meta-Analysis of Endoscopic vs. Microscopic Transsphenoidal Surgery for Non-functioning and Functioning Pituitary Adenomas: Comparisons of Efficacy and Safety. *Front Neurol.* 2021;12:614382.
4. Agustsson TT, Baldvinsdottir T, Jonasson JG, et al. The epidemiology of pituitary adenomas in Iceland, 1955-2012: a nationwide population-based study. *Eur J Endocrinol.* 2015;173:655-64.
5. Raappana A, Koivukangas J, Ebeling T, Pirilä T. Incidence of pituitary adenomas in Northern Finland in 1992-2007. *J Clin Endocrinol Metab.* 2010;95:4268-75.
6. Tjörnstrand A, Gunnarsson K, Evert M, et al. The incidence rate of pituitary adenomas in western Sweden for the period 2001-2011. *Eur J Endocrinol.* 2014;171:519-26.
7. Sivakumar W, Chamoun R, Nguyen V, Couldwell WT. Incidental pituitary adenomas. *Neurosurg Focus.* 2011;31:18.
8. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* 2012;48:441-6.
9. Lu CF, Hsu FT, Hsieh KL, et al. Machine Learning-Based Radiomics for Molecular Subtyping of Gliomas. *Clin Cancer Res.* 2018;24:4429-36.
10. Huang YQ, Liang CH, He L, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol.* 2016;34:2157-64.
11. Cuocolo R, Ugga L, Solari D, et al. Prediction of pituitary adenoma surgical consistency: radiomic data mining and machine learning on T2-weighted MRI. *Neuroradiology.* 2020;62:1649-56.
12. Peng A, Dai H, Duan H, et al. A machine learning model to precisely immunohistochemically classify pituitary adenoma subtypes with radiomics based on preoperative magnetic resonance imaging. *Eur J Radiol.* 2020;125:108892.
13. Zeynalova A, Kocak B, Durmaz ES, et al. Preoperative evaluation of tumour consistency in pituitary macroadenomas: a machine learning-based histogram analysis on conventional T2-weighted MRI. *Neuroradiology.* 2019;61:767-74.
14. Fan Y, Hua M, Mou A, et al. Preoperative noninvasive radiomics approach predicts tumor consistency in patients with acromegaly: Development and multicenter prospective validation. *Front Endocrinol (Lausanne).* 2019;10:403.
15. Rui W, Wu Y, Ma Z, et al. MR textural analysis on contrast enhanced 3D-SPACE images in assessment of consistency of pituitary macroadenoma. *Eur J Radiol.* 2019;110:219-24.
16. European Society of Radiology (ESR). ESR Statement on the Validation of Imaging Biomarkers. *Insights Imaging.* 2020;11:76.
17. Park JE, Kim D, Kim HS, et al. Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and TRIPOD statement. *Eur Radiol.* 2020;30:523-36.
18. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *Radiology.* 2015;277:826-32.
19. Jaremko JL, Azar M, Bromwich R, et al. Canadian Association of Radiologists White Paper on Ethical and Legal Issues Related to Artificial Intelligence in Radiology. *Can Assoc Radiol J.* 2019;70:107-18.
20. Geis JR, Brady AP, Wu CC, et al. Ethics of artificial intelligence in radiology: Summary of the joint European and North American multisociety statement. *Radiology.* 2019;293:436-40.
21. European Society of Radiology (ESR). What the radiologist should know about artificial intelligence - an ESR white paper. *Insights Imaging.* 2019;10:44.
22. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med.* 2016;15:155-63.
23. Kim JH. Multicollinearity and misleading statistical results. *Korean J Anesthesiol.* 2019;72:558-69.