UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

SCUOLA DI MEDICINA E CHIRURGIA

Dipartimento di Scienze Biomediche Avanzate

Direttore Prof. Alberto Cuocolo

DOTTORATO DI RICERCA IN SCIENZE BIOMORFOLOGICHE E CHIRURGICHE

XXXIII CICLO

Coordinatore Prof. Alberto Cuocolo

TESI DI DOTTORATO

Radiomic data mining and machine learning on preoperative pituitary adenoma MRI

RELATORE

Ch.mo Prof. Arturo Brunetti

CANDIDATO

Dott. Lorenzo Ugga

Contents

1. Introduction

1.1 Pituitary adenomas

Pituitary adenomas are among the most frequent intracranial tumors, accounting for the majority of sellar/suprasellar masses in adults with a prevalence ranging from 1 in 865 persons to 1 in 2688 persons [1–5].

Conversely, pituitary carcinomas are rare neoplasms, representing less than 0.5% of the pituitary symptomatic tumors. They are indistinguishable from adenomas on imaging and defined only by the presence of central nervous system or systemic metastases [6].

Pituitary adenomas are classified by size into microadenomas (<10 mm), macroadenomas (≥10 mm), and giant adenomas (≥40 mm).

Although most pituitary adenomas arise within the sella turcica, ectopic adenomas have been described in different sites including sphenoid sinus, nasopharynx, third ventricle, and suprasellar cistern tumors [7].

About two-thirds of pituitary adenomas may secrete excess hormones. Clinical case-finding studies showed that 32% to 66% of tumors were secreting prolactin, 14% to 54% were clinically nonfunctioning adenomas, 8% to 16% were secreting growth hormone, 2% to 6% were secreting adrenocorticotropic hormone, and less than 1% were secreting thyrotropin persons [1–5].

In most clinically nonfunctioning adenomas, stain for gonadotropins or their subunits is positive. However, rarely they may stain for growth hormone, adrenocorticotropic hormone, or prolactin without hypersecretion of these hormones, and are defined “silent” somatotrope, corticotrope, or lactotrope tumors, respectively [8].

The pathogenesis of most pituitary adenomas remains unknown. Cells with multipotent progenitor have been identified in the adult pituitary gland and may play a key role in tumorigenesis. In the presence of alterations in the normal microenvironment, uncoordinated proliferation of these pituitary stem cells may lead to pituitary adenoma formation [9].

Trans-sphenoidal surgical resection is the initial treatment for all tumors except prolactinomas, for which dopamine agonists administration is recommended as first-line treatment. Irradiation is reserved for patients who do not achieve adequate reduction in tumor size or hormone levels after surgery and/or medical therapy, and it is now always performed with stereotaxic technique [10, 11].

1.2 State-of-the-art imaging of pituitary adenomas

MRI is the preferred imaging modality for detecting pituitary adenomas. The standard protocol includes a sagittal spin-echo T1-weighted sequence, a coronal turbo spin-echo T2-weighted sequence and a coronal spin-echo T1-weighted

sequence. Contrast administration is recommended when an intrasellar lesion is suspected on the sequences without contrast, followed by the acquisition of coronal T1-weighted images. It is preferable to use a lower dose of contrast medium because the excessive enhancement of the glandular parenchyma could hide small intraglandular lesions.
In the case of a macroadenoma with suprasellar extension, coronal and sagittal post-contrast T1-weighted sequences and 3D gradient echo T1-weighted acquisition are recommended. In the investigation of suspected microadenoma, in particular in patients with Cushing disease, dynamic contrast-enhanced MR in the coronal plane using a power injector has to be performed. Gradient echo T2*- and susceptibility-weighted imaging can be useful in visualizing intralesional hemorrhagic components, while diffusion-weighted imaging (DWI) and apparent diffusion coefficient (ADC) maps have been proposed for tumor consistency assessment with variable accuracy [12].
Computed tomography imaging is complementary to MR imaging for the depiction of bony changes, including remodeling or frank bone destruction.

1.3 Radiomics

Radiomics represents the conversion of digital medical images (computed tomography, magnetic resonance, or positron emission tomography images) into mineable high-dimensional data. This process is motivated by the concept that biomedical images contain information that reflects underlying pathophysiology and that these relationships can be revealed via quantitative image analyses [13].
Radiomics has been recognized to have a great potential offering different applications especially in the oncologic field, from tissue characterization to treatment response prediction or prognosis. For this reason, it plays an essential role in so-called “personalized medicine” for which the availability of robust and validated biomarkers is essential.
The radiomic workflow include several steps:

- acquiring the images;
- identifying the volumes of interest;
- segmenting the volumes;
- extracting and qualifying descriptive features from the volume
- using these to populate a database;
- mining these data to develop classifier models either alone or in combination with additional information.

Two types of features can be extracted in radiomics, semantic and agnostic ones. Semantic features are commonly used by radiologists to describe lesions and can be somehow quantized. On the other hand, agnostic features are mathematically extracted quantitative descriptors, and can be divided into first-, second-, or higher order statistical outputs. First-order statistics describe the distribution of grey level values of individual voxels (histogram-based

methods). Second-order statistics are general defined “texture” features and describe interrelationships between voxels with similar (or dissimilar) contrast values.

1.4 Artificial intelligence

Machine learning (ML) is a subset of Artificial Intelligence (AI), characterized by the ability of software to learn from data and make predictions without explicit prior programming. It is based on algorithms which can be divided based on learning type. In supervised learning, labelled data is available for the training process thus creating a ground truth. In unsupervised learning, the software finds hidden structure in data automatically without prior categorization of the training set. In reinforcement learning, positive and negative reinforcement contribute to the improvement of the predictive model [14]. In other words, it learns from the consequences of its interactions with a dynamic environment. ML has potential applications in multiple fields of medicine, though it has been most frequently applied to imaging. In this setting, it has shown promising performance in cancer patients for lesion detection, characterization and staging.[6–10] ML analysis has also been used in non-oncologic pathology, with promising results [15, 16]. Another interesting aim of ML is represented by the automated segmentation of medical images. Though manual segmentation represents the gold standard, AI tools might have higher accuracy and reproducibility [17].

Although AI has been widely used in imaging, it has other applications as it enables fast and accurate analysis of large amounts of data. This ability can be employed to create predictive models both for classification and regression.

1.5 Structure of the Manuscript and Objective/Contribution

The manuscript is organized as follows:

- Chapter 2 presents the first contribution of the thesis, which consists in evaluating the functional status of pituitary macroadenomas using a machine learning analysis. In particular, a machine learning approach has been applied in order to classify pituitary adenomas into functional and non-functional starting from preoperative MRI features. Multinomial logistic regression and K nearest neighbor classifiers achieved accuracies beyond 92.0% and an Area Under the Receiving Characteristic Operator Curve till 98.4%. The novelty of this paper lies in proving the ability of the combination of radiomics and machine learning to pre-operatively predict adenoma function, previously only possible with blood tests or histopathological analysis. The study included in this chapter has been presented at MEDICON 2019 and published as a conference paper on IFMBE Proceedings [18].
- Chapter 3 presents the second contribution of the thesis, to assess the accuracy of machine learning analysis of texture-derived parameters from pituitary adenomas preoperative MRI for the prediction of ki-67 proliferation

index class. The ki-67 labeling index represents a proliferative marker which correlates with pituitary adenoma aggressiveness. The higher this index, the higher the clinically aggressive behavior, with recurrent disease and resistance to multimodal therapy. From T2-weighted MR images, quantitative features were extracted. Subsequently, different supervised feature selection methods were employed and a k-nearest neighbors classifier was used to predict macroadenoma high or low proliferation index with a train-test approach. The overall accuracy of the k-NN in the test group was of 91.67% of correctly classified patients. In this study machine learning analysis of texture-derived parameters from preoperative T2 MRI has proven to be effective for the prediction of pituitary macroadenomas ki-67 proliferation index class. This might aid the surgical strategy making a more accurate preoperative lesion classification and allow for a more focused and cost-effective follow-up and long-term management. This chapter was published as a journal article on Neuroradiology [19].

- Chapter 4 presents the third contribution of the thesis, the crucial importance of preoperatively investigating pituitary adenoma consistency. Indeed, tumor consistency can influence the ease of lesion removal during surgery, especially when using a transsphenoidal approach. Unfortunately, it is not assessable on standard qualitative MRI. In this study, radiomic texture features were extracted from original and filtered T2-weighted MR images. After hyperparameter tuning via stratified 5-fold cross-validation, the Extra Trees classifier obtained an accuracy of 93%, demonstrating that preoperative T2-weighted MRI texture analysis and ML could predict pituitary macroadenoma consistency. This chapter was published as a journal article on Neuroradiology [20].
- Chapter 5 presents a narrative review which provides an overview of the main concepts in ML and to analyze its applications in the imaging of pituitary adenomas. Pre-treatment assessment and neurosurgical outcome prediction were the potential ML applications using magnetic resonance imaging. Regarding pre-treatment assessment, ML methods were used to have information about tumor consistency, predict cavernous sinus invasion and high proliferative index, discriminate null cell adenomas, which respond to neo-adjuvant radiotherapy from other subtypes, predict somatostatin analogues response and visual pathway injury. Regarding neurosurgical outcome prediction, the following applications were discussed: Gross total resection prediction, evaluation of Cushing disease recurrence after transsphenoidal surgery and prediction of cerebrospinal fluid fistula's formation after surgery. Although clinical applicability requires more replicability, generalizability and validation, results are promising, and ML software can be a potential power to facilitate better clinical decision making in pituitary tumor patients. This chapter was published as a journal article on Artificial Intelligence in Medical Imaging [21].

Finally, chapter 6 concludes the thesis by summarizing the contributions and providing some perspectives.

1.6 List of Publications and Award

Journal articles

- Ugga L, Cuocolo R, Solari D, Guadagno E, D'Amico A, Somma T, Cappabianca P, Del Basso de Caro ML, Cavallo LM, Brunetti A. Prediction of high proliferative index in pituitary macroadenomas using MRI-based radiomics and machine learning. Neuroradiology. 2019 Dec;61(12):1365-1373. doi: 10.1007/s00234-019-02266-1. Epub 2019 Aug 2. PMID: 31375883 [19];
- Cuocolo R, Ugga L, Solari D, Corvino S, D'Amico A, Russo D, Cappabianca P, Cavallo LM, Elefante A. Prediction of pituitary adenoma surgical consistency: radiomic data mining and machine learning on T2-weighted MRI. Neuroradiology. 2020 Dec;62(12):1649-1656. doi: 10.1007/s00234-020-02502-z. Epub 2020 Jul 23. PMID: 32705290 [20];
- Guerriero E, Ugga L, Cuocolo R. Artificial intelligence and pituitary adenomas: a review. Artificial Intelligence in Medical Imaging. 2020 Aug;1(2): 70-77. doi: 10.35711/aimi.v1.i2.70 [21].

Conference papers

- Ricciardi C, Cuocolo R, Cesarelli G, Ugga L, Improta G, Solari D, Romeo V, Guadagno E, Cavallo LM, Cesarelli M. Distinguishing Functional from Non-functional Pituitary Macroadenomas with a Machine Learning Analysis. In: Henriques J., Neves N., de Carvalho P. (eds) XV Mediterranean Conference on Medical and Biological Engineering and Computing - MEDICON 2019. IFMBE Proceedings, vol 76. Springer, Cham. doi: 10.1007/978-3-030-31635-8_221 [18].

Award

- Federico Bartolazzi 2019 Grant of the Italian Society of Neurosurgery for the research project "*Preoperative identification of pituitary macroadenomas with high proliferative index on magnetic resonance: radiomic study with machine learning approach".*

2. Pituitary adenoma functional status

2.1 Introduction

Pituitary adenomas are rare intracranial tumors, presenting with a prevalence of 1/1500 in the general population. In most cases, they are benign lesions, whose clinical manifestations are related to mass effect signs - depending on tumor size and/or to hormone hypersecretion syndromes. On the other hand, low dimension intrasellar adenomas may be clinically silent and their diagnosis often comes as an incidental finding on MR scans [22–25].

Radiomics, consisting in the conversion of images into mineable data and subsequent analysis for decision support, is an emerging field allowing tumor classification [13]. In particular, texture analysis is a postprocessing technique for quantitative parameter extraction from pixel grey level heterogeneity. It consists of statistical analysis based on both simple intensity value distribution histograms and more complex gray level distribution matrix analyses which also retain information on spatial distribution of voxel intensities [26].

In this setting, machine learning can be applied in order to predict the outcome of patients and help clinicians in decision-making [27–30]. There is a wide range of applications of machine learning in different areas of medicine, from cardiology to radiology [14, 31]. In particular, studies applying machine learning on texture analysis according to the "radiomic process" were described by Kumar et al. [32]; Zacharaki et al. classified brain tumor type and grade using MRI texture and shape through Linear Discriminant Analysis with Fisher's discriminant rule, k-nearest neighbour (KNN), nonlinear Support Vector Machine (SVM) and employing leave one out crossvalidation [33]; Juntu et al. differentiated benign from malignant soft-tissue tumours in T1-MRI images testing three classifiers (neural networks, decision tree and SVM) [34]; Romeo et al. characterized adrenal lesions on unenhanced MRI images [35]; finally, Stanzione and colleagues have recently demonstrated the potential of this approach in prostate cancer local staging [36].

Moreover, recent studies investigated the relevance of first and second order histogram features obtained by diffusion-weighted imaging magnetic resonance in differentiating functional from non-functional pituitary macroadenoma through a classic statistical analysis [37].

Therefore, the aim of this study is to apply machine learning algorithms on parameters obtained by texture analysis on MRI images in order to distinguish functional from non-functional pituitary macroadenomas.

2.2 Materials and methods

2.2.1 Subjects

We retrospectively reviewed data of 50 patients, who received so-called standard endoscopic endonasal approach for the removal of a pituitary adenoma, between January 2013 and December 2017, at the Division of Neurosurgery of the University of Naples 'Federico II' in Italy. All of them underwent preoperative MRI at our Institution prior to the surgical procedure. Demographic data, preoperative assessment - i.e., endocrinological and visual status and presenting signs - tumor features, prior Distinguishing Functional from Non-functional Pituitary 1823 treatments, surgical results and complications, were retrieved from our electronic database (Filemaker Pro 11 - File Maker Inc, Santa Clara, CA, USA).

2.2.2 MRI Acquisition and Texture Analysis

All exams were acquired on a 1.5-Tesla scanner (Gyroscan Intera, Philips, Eindhoven, The Netherlands). The imaging protocol always included a coronal T2-weighted Turbo Spin Echo sequence (TR/TE: 2600/89 ms, FOV: 180 x 180 mm; matrix: 288 x 288; thk: 3 mm) used for the following radiomic feature extraction. First of all, lesions were detected by an expert neuroradiologist who then performed their manual contouring by means of a bidimensional polygonal ROI after selection of the slice where it showed maximum extension (Fig. 2.1). Further editing with a brush tool was performed, when needed. This process was carried on using a freely available segmentation software (ITKSnap v3.6.0) [38].

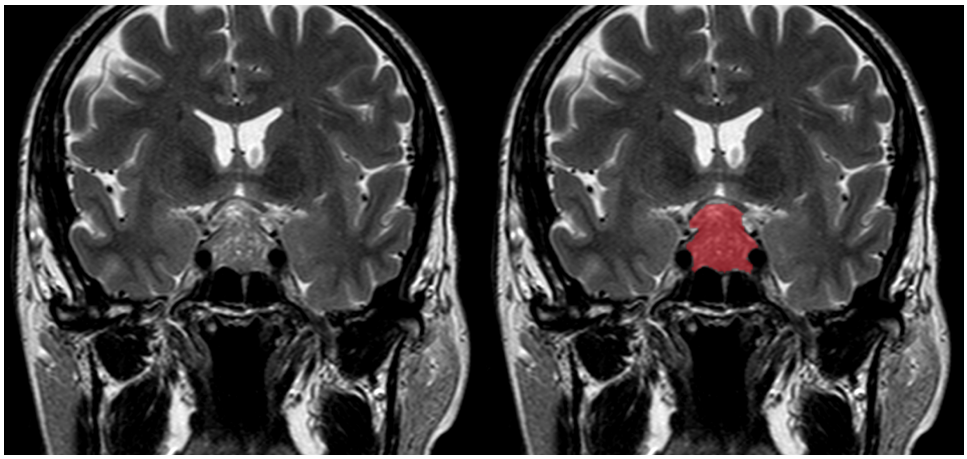

Fig. 2.1. Coronal T2-weighted MRI exam showing the maximum extension slice of a functioning pituitary macroadenoma (A). Image B depicts the result of the image annotation of the region of interest to be employed for subsequent texture feature extraction

Image pre-processing and feature extraction were performed on an open-source Python radiomics software (Pyradiomics v2.1.2) [39]. The first step consisted of image gray level normalization with a scale of 100). This step was mandatory since T2-weighted images are not quantitative and intensity values are not absolute in contrast to T2 maps.

The latter were not available as only routine clinical scans were selected for the analysis, also in order to guarantee reproducibility of the results in the clinical setting. Subsequently, all volumes and corresponding lesion masks were resampled to a 2 x 2 x 2 mm voxel resolution. The next pre-processing step consisted of intensity value discretization. For this task, a fixed bin width approach was chosen, obtaining an ideal bin count between 16 and 128, as suggested in previous studies [40]. The use of wavelet decomposition, yielding all possible combinations of High and Low pass filtering in the x, y and z dimensions, and edge enhancement Laplacian of Gaussian (LoG) filters, emphasizing gray level change at different texture coarseness, allowed additional feature extraction from the derived images. shape and first order statistics, we also obtained higher order class parameters. In detail, these were derived from the symmetrical Gray Level Co-occurrence Matrix (GLCM), Gray Level Size Zone Matrix (GLSZM), Gray Level Run Length Matrix (GLRLM, Neighboring Gray Tone Difference Matrix (NGTDM) and Gray Level Dependence Matrix (GLDM).

2.2.3 Tool

Knime analytics platform (v. 3.7.1) was chosen to conduct this machine learning study, as it is a well-known open-source platform implementing a wide range of machine learning algorithms and integrated with Weka, Python and other software; moreover, it was already employed in literature for other studies [41][42]. The algorithms used in this paper are briefly presented in the next section.

2.2.4 Algorithms and Evaluation Metrics

J48 is the Java implementation of a C4.5 decision tree [43], which consists of the evolution of the ID3 algorithm. It is an easy structure made up of leaves, representing classes, and nodes, representing test phases over an attribute. Multinomial Logistic Regression (MLR) with ridge estimator is applied through the “Logistic” node of Weka that follows the implementation of Le Cessie, and van Houwelingen [44, 45].

K Nearest Neighbour (KNN) is an easy instance-based classifier that assigns a label basing its choice on the dominance of a class in the nearest neighbours [46]. For all these algorithms, “smote” (Synthetic Minority Over-sampling Technique) was applied [47]. Smote generates artificial data by extrapolating between a real object of a given class and one of its nearest neighbours (of the same class). Boosting was implemented for J48, it converts weak learners into strong learners that predict with higher accuracy; it selects only the parameters that can improve the predictive ability of algorithms during the training phase, making the complexity in terms of dimension decrease and improving execution time [48]. The evaluation metrics employed in this study are:

- Accuracy: correct classifications over the total;
- Error: misclassifications over the total;

- Recall: the ratio of positives correctly classified;
- Precision: the ratio of positives correctly predicted in the positive class;
- Sensitivity: capacity to detect true positives;
- Specificity: capacity to detect true negatives.

Moreover, Area Under the Curve Receiving Characteristic Operator (AUCROC) was computed for each algorithm and for both bagging and boosting groups.

2.3 Results

Of the included lesions, 25 were functioning adenomas (5 Adreno-cortico-tropic Hormone, 8 Growth Hormone, 5 Growth Hormone/Prolactin, 6 Prolactin and 1 Thyroid-Stimulating Hormone secreting) and 25 non-functioning. A total of 1128 features was extracted from each patient.

Due to the small number of patients, smote technique was applied to make the number of records rise from 50 to 100. Then, a procedure of feature selection was applied to reduce the number of features extracted by the images: the matrix of correlation was computed among all variables and a threshold of correlation of 0.4 was chosen: all the variables with a correlation higher than the threshold were excluded because they did not add information to the classifiers. It allowed us to reduce the number of features from more than one thousand to 28. As the number of patients was not so high, leave one out was applied for all the implemented algorithms. J48, MLR and KNN were implemented together with the boosting node of Knime. Results are summarized in Table 2.1 while Table 2.2 shows the features used to build the models.

Table 2.1. Scores for each algorithm.

	Accuracy [%]	**Error [%]**	**Recall [%]**	**Precision [%]**	**Sensitivity [%]**	**Specificity [%]**	**AUCROC [%]**
Boosting J48	83.0	17.0	82.0	83.7	82.0	84.0	89.8
MLR	97.0	3.0	96.0	98.0	96.0	98.0	98.4
KNN	90.0	10.0	90.0	90.0	90.0	90.0	90.0

Table 2.2. Features used to build the predictive models.

log-sigma-2-0-mm-3D_gldm_LargeDependenceLowGrayLevelEmphasis	wavelet-LHH_firstorder_Skewness
log-sigma-2-5-mm-3D_glcm_InverseVariance	wavelet-LHH_firstorder_Entropy

log-sigma-2-5-mm-3D_firstorder_Range	wavelet-LLH_glcm_ClusterShade
log-sigma-3-0-mm-3D_gldm_LargeDependenceLowGrayLevelEmphasis	wavelet-LLH_glcm_Contrast
log-sigma-3-5-mm-3D_gldm_LargeDependenceLowGrayLevelEmphasis	wavelet-LLH_firstorder_Kurtosis
log-sigma-3-5-mm-3D_gldm_LargeDependenceHighGrayLevelEmphasis	wavelet-HLH_glcm_Id
log-sigma-3-5-mm-3D_firstorder_Maximum	wavelet-HLH_firstorder_Skewness
wavelet-HLL_glcm_ClusterProminence	wavelet-HLH_firstorder_Median
wavelet-HLL_firstorder_Skewness	wavelet-HHH_firstorder_Skewness
wavelet-HLL_firstorder_Maximum	wavelet-LLL_glcm_ClusterShade
wavelet-LHL_glcm_Correlation	wavelet-LLL_firstorder_Kurtosis
wavelet-LHL_firstorder_Skewness	original_glcm_Correlation
wavelet-LHL_firstorder_Mean	original_firstorder_Skewness
wavelet-LHH_gldm_DependenceVariance	original_firstorder_Kurtosis

MLR obtained the highest accuracy, recall, precision, sensitivity, specificity and AUCROC among the three implemented algorithms. Despite getting the lowest accuracy (83.0%), J48 reached an AUCROC comparable to the KNN's one.

2.4 Discussion

First, the MRI acquisition of 50 patients was performed at the department of Advanced Biomedical Sciences of the University Hospital "Federico II" of Naples. Furthermore, a texture analysis was conducted to extract more than one thousand quantitative features from the MRI images. The machine learning analysis was finally performed in order to carry out some evaluation metrics as regards the algorithms.

Mentioning other studies that employed radiomics and machine learning, Romeo et al. [35] characterized adrenal lesions with a diagnostic accuracy of 80%, while Juntu et al. [34] distinguished benign from malignant tumors with an accuracy of 93%; Zacharaki et al. [33] obtained 85% of accuracy classifying type and grade of brain tumors. Although a direct comparison with other studies would not be completely fair (due to the use of different datasets), this study shows greater capacity to correctly make classifications (functional and non-functional pituitary macroadenomas), exploiting features extracted through texture analysis. A comparison may be done with the study of Sanei et al. [37] who distinguished functional from non-functional pituitary macroadenomas with lower scores than those obtained through a machine learning analysis.

The functional status of pituitary lesions has a significant influence on the clinical manifestations of disease: the correct diagnosis and management is crucial for the selection of the correct therapeutic strategy and therefore cure this

multifaceted disease. Although a previous study has shown the promise of Apparent Diffusion Coefficient values of pituitary lesions in this assessment [38], Diffusion Weighted Imaging is not routinely performed in the imaging of the sellar region. It is known that this area is potentially more prone to artefacts on echo-planar imaging and this technique is time consuming. For these reasons, an approach that obtains similar results while employing routine MRI sequences has more potential for its application in the current clinical setting.

Of course, this study is affected by some limitations: the dataset was augmented with artificial data in order to improve its size allowing us to perform the analysis. Major dataset could be studied, and machine learning analysis could be performed to reach 100% accuracy in this classification. Nevertheless, machine learning proved to be the best way to distinguish functional from non-functional pituitary macroadenomas using texture analysis on MRI images.

2.5 Conclusion

This paper proved that the combination of radiomics and machine learning can be used to predict tumoral behavior pre-operatively while only blood tests or histopathological analysis were known as providers of this information.

3. Pituitary adenoma proliferative index

3.1 Introduction

Pituitary adenomas are among the most frequent intracranial tumors, with a 1/1500 prevalence [49]. They are mostly benign and typically presents with hormone hypersecretion syndromes and/or mass effect signs. Small intrasellar tumors can be clinically silent and diagnosed only as incidental MRI findings. Nonetheless, some entities among the different subtypes of adenoma show more aggressive and unpredictable behavior: these - such as sparsely granulated somatotroph adenomas, Crooke cell adenomas, and plurihormonal Pit-1-positive adenomas - tend to disclose local invasiveness, high recurrence risk and rarely present features of highly aggressive cancer. Surgery is considered the first line of treatment in most cases but in this latter group, multimodality management therapy - as per neurooncological guidelines - is mandatory. The possibility of predicting pituitary tumor behavior at the preoperative stage cannot be yet considered reliable as no valid factor has been identified, and though it remains the critical groundwork of the pathologists.

The 2017 WHO classification revealed that former categories of typical adenomas, atypical adenomas, and pituitary carcinomas have no relevance from a clinical standpoint, introducing the definition of "high-risk" adenomas in reference to tumors with rapid growth, radiological invasion, and high Ki-67 proliferation index [50–52].

Considering the above, early identification of any radiological feature defining such behavior can be crucial in order allow timely diagnosis and treatment. In this regard, radiomics, consisting in the conversion of images into mineable data and subsequent analysis for decision support, has been emerging [13]. In particular, texture analysis is a postprocessing technique for quantitative parameter extraction from pixel grey level heterogeneity. More recently, texture analysis-derived features have been used in association with data mining machine learning algorithms, aiding in the interpretation of the large amount of information produced. Machine learning is the branch of artificial intelligence dealing with computer algorithms capable of learning and improving in accuracy by analyzing datasets, without prior explicit programming [53]. It leads to the creation of automated predictive models to solve classification problems. The usefulness of the radiomics approach is being assessed in different fields of radiology [15, 35, 36, 54–58].

The aim of our study was to assess the accuracy of machine learning analysis of texture-derived parameters from preoperative MRI of pituitary macroadenomas for the prediction of ki-67 proliferation index class.

3.2 Materials and methods

This study was approved by the local institutional review board, which waived the necessity for informed consent due to the retrospective nature of the study.

3.2.1 Subjects

We retrospectively reviewed data of 108 consecutive patients, who underwent endoscopic endonasal procedures for pituitary adenoma removal between January 2013 and December 2017, at the University of Naples "Federico II" Neurosurgery Unit.

Only patients with available ki-67 labeling index in the histopathological report were included (n = 106). Exclusion criteria were: any previous treatment for pituitary adenoma (radiation or medical therapy) (n = 9), extensively necrotic or hemorrhagic lesions (n = 6), significant artifacts on the images used for the analysis (n = 2).

Demographic data, preoperative assessment, tumor features and histopathological characteristics were retrieved from our electronic database (Filemaker Pro 11 - File Maker Inc, Santa Clara, CA, USA).

3.2.2 Surgical approach

Endonasal surgical procedures were performed using a rigid 0-degree endoscope, 18 cm in length and 4 mm in diameter (Karl Storz Endoscopy, Tuttlingen, Germany), as the sole visualizing tool. The use of 30°-45° angled endoscopes was reserved to explore large intra-suprasellar post-surgical tumor cavities. They were run according to techniques already described in previous publications [25, 59–62].

3.2.3 Pathological data

Specimens were obtained as formalin-fixed tissue. For the evaluation of the proliferation index ki-67 (Labeling Index, L.I.), "hot spot" areas were chosen at low magnification and an average of the values on 5 adjacent fields (at least 500 neoplastic cells) was calculated (Fig. 3.1).

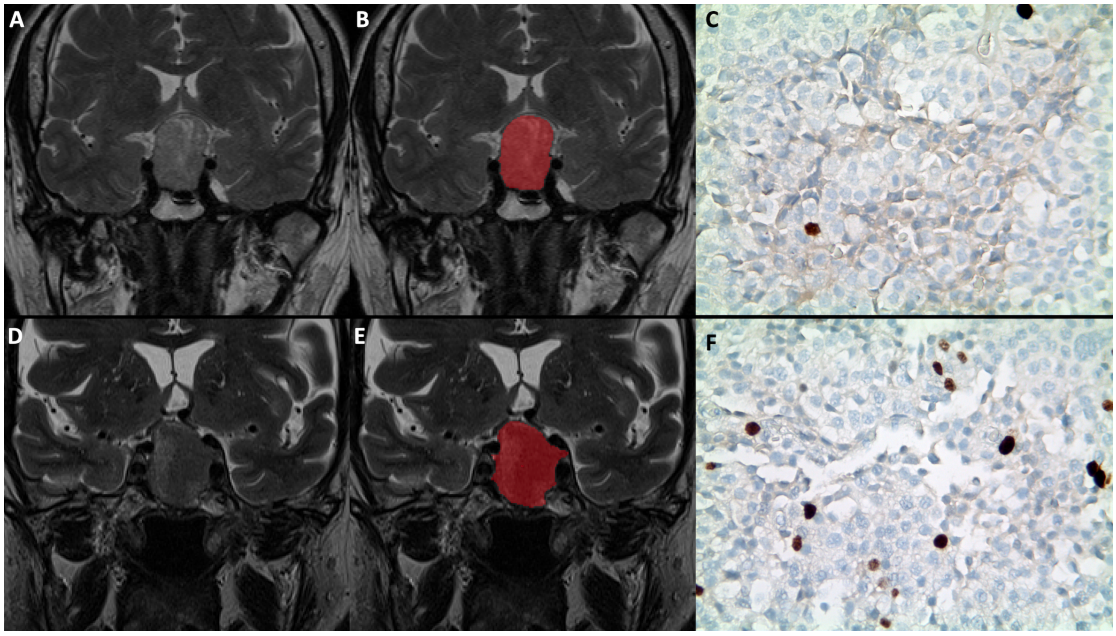

Fig. 3.1. Segmentation examples on coronal T2-weighted images in two patients with low (upper row, A) and high (lower row, D) proliferation index pituitary adenomas, showing hand-drawn ROI placement (B and E, respectively). Pictures C and F show corresponding immunohistochemical evaluation (40x magnification) of ki-67 cell labeling index, respectively, 1 and 6%.

3.2.4 MRI acquisition data

Imaging studies were performed either on a 1.5-Tesla scanner (Gyroscan Intera, Philips, Eindhoven, The Netherlands), or 3T MR scanner (Magnetom Trio, Siemens Medical Solutions, Erlangen, Germany). Both protocols included a coronal T2-w TSE sequence (TR/TE: 2600/89 ms, FOV: 180x180 mm; matrix: 288x288; thk: 3 mm, at 1.5 T; TR/TE: 3000/98 ms FOV: 200x200 mm; matrix: 384x384; thk: 3 mm, at 3 T).

3.2.5 Image analysis

Tumors were identified by an expert neuroradiologist who proceeded to manual segmentation using a bidimensional polygonal ROI on the slice of lesion maximum extension with further editing with a brush tool, when needed. This process was carried on using a freely available segmentation software (ITKSnap v3.6.0) [38] (Fig. 3.1). A second segmentation on a sample of 35 patients was performed by another neuroradiologist, blinded to the prior annotations and clinical data, to assess inter-operator reproducibility of the selected features.

Image pre-processing and feature extraction were performed on an open-source Python radiomics software (Pyradiomics v2.1.2) [39]. In particular, as two MRI scanners by different vendors and with distinct acquisition parameters were employed, image voxels were first normalized by subtracting the mean intensity and dividing by the standard deviation with an expected resulting range ~ [-3, 3], a mean of 0 and standard deviation of 1 in the normalized

image. Then, a scale of 100 was applied, resulting in an intensity range mainly within [-300, 300]. This would have been required even using a single scanner as non-quantitative sequences were evaluated (e.g., T2-w images and not T2 maps). All images and corresponding ROI masks were also resampled to a 2x2x2 mm resolution. To optimize extraction time, making it less memory intensive, images were automatically cropped based on ROI masks keeping a 10-voxel padding in order to consent subsequent filter application.

As for image discretization, the employed software allows both for a fixed bin count and bin width. As suggested by the developers, we decided to use a fixed bin width that allowed us to obtain an ideal bin count between 16 and 128 [40]. To do so, a preliminary extraction of first order parameters was performed in all patients, calculating using the gray-level range an optimal bin width of 3. To avoid negative gray values after normalization, potentially affecting first order feature extraction, an array shift of 300 (3 standard deviations x 100) was applied.

Various filters were also used to generate derived images for additional texture feature extraction. In particular, wavelet decomposition yielding all possible combinations of High and Low pass filtering in the x and y dimensions; edge enhancement Laplacian of Gaussian (LoG) filter emphasizing gray level change at different texture coarseness. Due to the applied resampling resolution and lesion size, we chose four LoG sigma values ranging from 2.0 (maximum fineness) to 3.5 mm (maximum coarseness) with 0.5 mm increments.

Regarding feature extraction, to bidimensional shape and first order statistics, commonly referred to as histogram analysis, we added higher order class parameters. In detail, the symmetrical Gray Level Co-occurrence Matrix (GLCM) characterizes image texture by calculating how often pairs of voxels with specific intensity levels and spatial relationship occur in an image, extracting statistical measures from the deriving matrix [63]. Gray Level Size Zone Matrix (GLSZM) features quantify gray level zones, defined as the number of connected voxels that share the same intensity value [64]. The Gray Level Run Length Matrix (GLRLM) evaluates gray level runs, which are the length of consecutive pixels with the same gray level [65]. Neighboring Gray Tone Difference Matrix (NGTDM) features assess differences between pixel values and neighbor average gray value within a set distance [66]. Finally, employing Gray Level Dependence Matrix (GLDM) derived parameters, the dependency of connected voxels, expressed as their number within a set distance, to the center voxel can be determined [67].

3.2.6 Statistical analysis

Feature scaling was performed through normalization. Subsequently, in order to remove irrelevant and redundant data, which can reduce computation time, improve learning accuracy, and facilitate a better understanding for the learning model or data, different supervised feature selection methods were employed [68]. This process and subsequent steps in our machine learning pipeline were performed on a freely available data mining software (Weka v3.8) [69]. To balance

data dimensionality reduction and loss of information, we tested eight feature selection method belonging to three different families:

1. Embedded classifier methods, both ranking attribute worth with the OneR (1R) algorithm and C4.5 decision tree, and selecting the best feature subset for C4.5 decision tree and Random Tree algorithms [34];
2. Feature ranking methods using a direct Pearson's correlation and a ReliefF Evaluator, a filter-method approach sensitive to feature interactions [70];
3. Subset feature search selection method, which evaluates attribute subset worth considering both feature predictive ability and redundancy [70].

Absolute agreement intra-class correlation coefficient (ICC) values were calculated for the selected texture features using SPSS version 17.

A *k*-nearest neighbors (*k*-NN) classifier was employed to predict macroadenoma proliferation index class. It represents a non-parametric, lazy learning algorithm, not making any assumption on the underlying data distribution and not using training data points to do generalization, deferring computation until classification [45]. We utilized a $k = 3$, linear nearest neighbor search with Euclidean distance function and no distance weighting.

Algorithm validation was performed with a train-test approach, randomly splitting the patients in training (60%) and test (40%) groups. Feature selection and model training were performed on the first while the latter was employed for classifier validation and calculation of accuracy metrics.

Our radiomics workflow pipeline is illustrated in Figure 3.2.

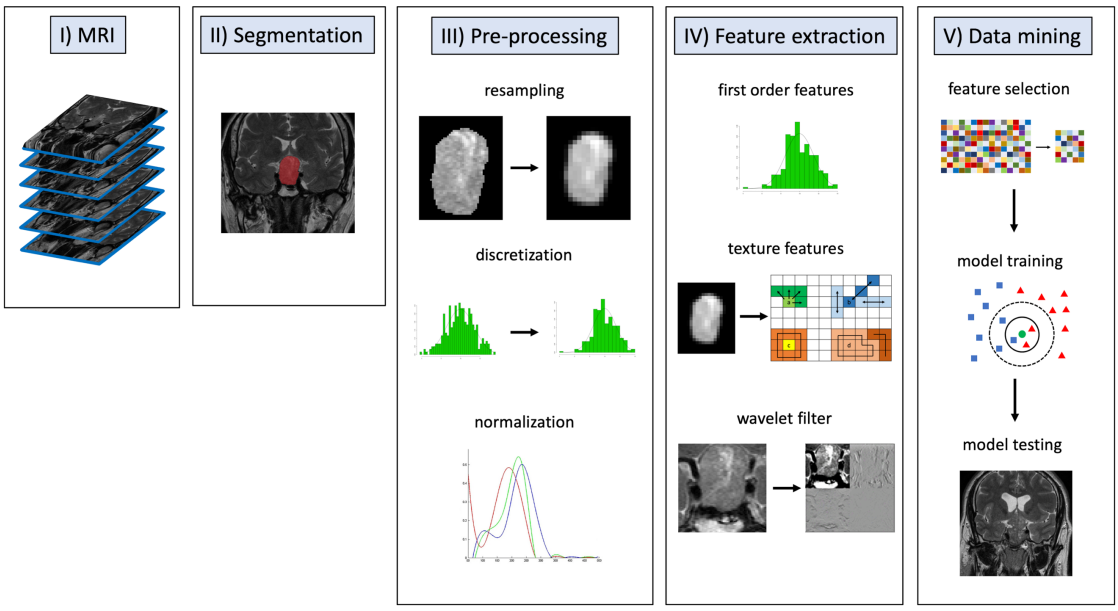

Fig. 3.2. Radiomic workflow pipeline. Manual segmentation of the lesions was performed on coronal T2-weighted images using a bidimensional polygonal ROI. Image pre-processing included resampling of all images and

corresponding ROI masks to a 2x2x2 mm resolution, discretization and normalization of voxel gray levels. Feature extraction (histogram and texture analysis) was performed on both native and filtered images. Subsequently, in order to remove irrelevant and redundant data, different supervised feature selection methods were employed. A *k*-NN classifier was employed to predict macroadenoma proliferation index class with a train-test approach.

3.3 Results

3.3.1 Subjects and pathological data

According to inclusion and exclusion criteria, the final population consisted of 89 patients (51 males and 38 females; mean age 52,17±1 years; range, 16 – 80). As stated above, they were divided in training (n = 53) and test (n = 36) groups.

Of the included lesions, 25 were functioning adenomas (5 ACTH, 8 GH, 5 GH/PRL, 6 PRL and 1 TSH secreting) and 64 non-functioning. Concerning tumor location, 13 were purely intrasellar, 40 presented suprasellar infradiaphragmatic extension, while 36 involved the supradiaphragmatic space (Table 3.1).

Table 3.1. Patient population clinical data.

Characteristic	**Low ki-67 LI**	**High ki-67 LI**
No. of patients	59	30
Age (mean) (yr)	54,8	47,6
Sex M (No.) (%) F (No.) (%)	35 (59.3%) 24 (40.6%)	16 (53.3%) 14 (46.7%)
Tumor type	F (12) NF (47)	F (13) NF (17)
MRI availability (No.)		
3 T	14	7
1.5 T	45	23

LI = labelling index; F = functioning macroadenomas; NF = non-functioning macroadenomas.

Surgical treatment was performed via “standard” endoscopic endonasal approach in 79 cases, with a transtuberculum transplanum “extended” approach in the remaining. Overall gross total removal was obtained in 74 cases with a subtotal removal in the other 15 patients.

In regard to pathological features, we identified 59 lesions (12 functioning and 47 non-functioning tumors) whose ki-67 was lesser than 3% (low proliferation index), whilst in 30 cases (13 functioning and 17 non-functioning lesions) it resulted equal to or greater than 3% (high proliferative index)[71].

The training group included 20 high proliferation and 33 low proliferation index patients, while the test one included respectively 10 and 26 patients for each class.

3.3.2 Image analysis

A total of 1128 features was extracted from each patient, represented on the correlation clustermap shown in Figure 3.3.

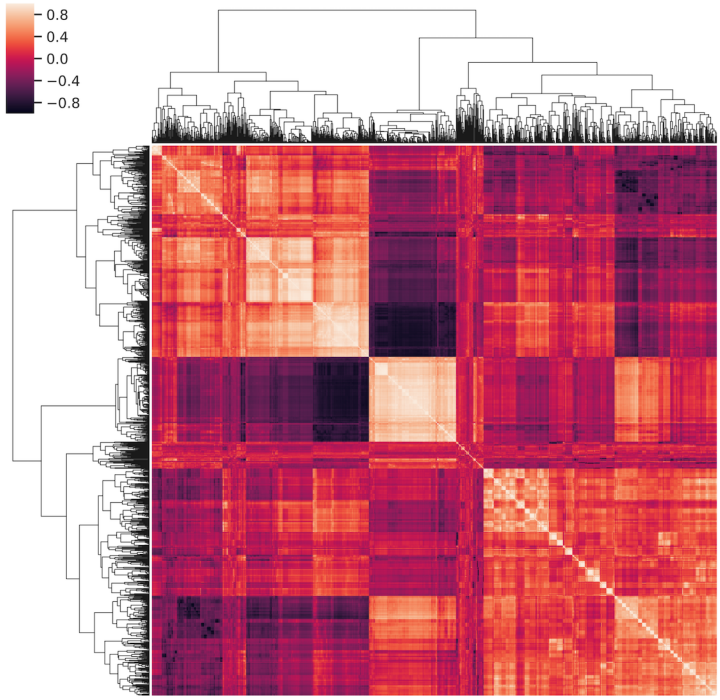

Fig. 3.3. Feature correlation matrix represented as a hierarchically-clustered heatmap.

Of the 12 subsets derived from feature selection (feature number range = 1-6), the best performing one was constituted by the 4 highest-correlating parameters at Pearson's test. They were: kurtosis derived from the filtered LoG (sigma = 3.5), Zone Variance from both the GLSZM of the original image and after low-high pass wavelet transform, and Large Area Emphasis from the GLSZM after low-high pass wavelet transform. Their ICC values were respectively 0.87, 0.97, 0.93 and 0.95, indicating very good reproducibility [72]. Figures 3.4 and 3.5 present univariate and pairwise feature distribution for this data subset in our population.

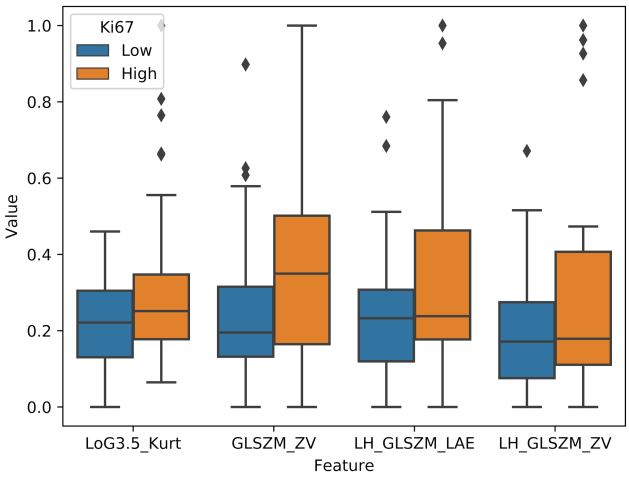

Fig. 3.4. Box plot of the distribution for each feature of the selected subset in relation to the proliferation index class.

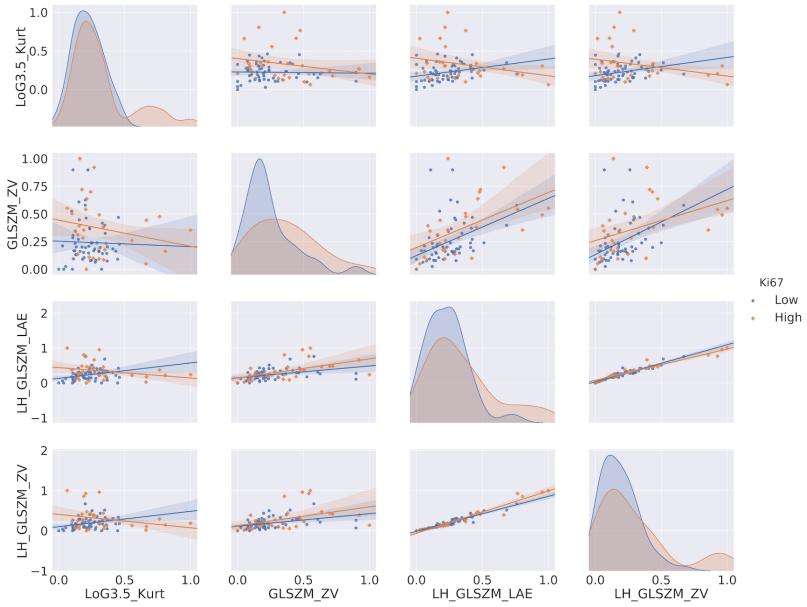

Fig. 3.5. Pairwise bivariate distribution with regression lines for the selected feature subset in relation to the proliferation index class.

The *k*-NN overall accuracy was 91.67% (33/36) of correctly classified patients. Other evaluation metrics are reported in table 2, derived from the confusion matrix in table 3. Among these, some accuracy measures more commonly used in machine learning but less in other fields included:

- F-score, the harmonic average of the precision (also Positive Predictive Value) and recall (also Sensitivity) ranging from 0 to 1 (perfect accuracy), which in our case was 0.92;

- Matthews Correlation Coefficient, a measure of the quality of binary classifications in machine learning (+1 representing a perfect prediction, 0 an average random prediction and -1 and inverse prediction), was 0.78;
- the area under the Precision-Recall Curve representing an alternative to the area under the Receiver Operator Characteristics curve that is considered more informative for imbalanced classes. A high area under the curve, 0.88 in our case, represents both high recall and high precision.

Table 3.2. *k*-NN accuracy metrics weighted average and by class.

Class	Sensitivity	Specificity	Precision	MCC	F-score	ROC Area	PRC Area
Low ki-67 LI	0,96	0,80	0,93	0.79	0.94	0,87	0,91
High ki-67 LI	0,80	0,96	0,89	0.79	0.84	0,87	0,79
WA	0,92	0,86	0,92	0.79	0.92	0,87	0,88

LI = labelling index; WA = weighted average; MCC = Matthews Correlation Coefficient; ROC = Receiver Operator Curve; PRC = Precision Recall Curve.

Table 3.3. Confusion matrix for the test group.

		Actual class	
		Low ki-67 LI	High ki-67 LI
Predicted class	Low ki-67 LI	25	1
	High ki-67 LI	2	8

LI = labelling index.

3.4 Discussion

In the previous edition of the WHO Classification of Tumors of the Pituitary Gland (2004), pituitary neuroendocrine tumors were divided into typical adenoma, atypical adenoma, and carcinoma. Atypical adenomas were identified by an elevated mitotic index and ki-67-labeling index ≥ 3%, suggestive of aggressive clinical behavior. Using this classification, atypical adenomas incidence was relatively variable and lacking valid and reliable prognostic correlations [73]. Therefore, Trouillas and colleagues proposed a new clinicopathological classification using MRI invasiveness signs (cavernous and/or sphenoid sinus involvement), immunocytochemistry and proliferative markers, labeling adenomas according to size, type and grade (grade 1a: non-invasive, 1b: non-invasive and proliferative, 2a: invasive,

2b: invasive and proliferative, and 3: metastatic) [50]. Along this line, the fourth WHO classification edition (2017) emphasizes pituitary adenoma histopathological aspects and molecular genetics, abandons the term "atypical" and strongly recommends an accurate assessment of tumor subtyping, proliferative potential and clinical parameters such as tumor invasion for consideration of aggressive adenomas [74, 75]. However, it does not provide any hint to correlate these elements to define pituitary adenomas subtypes [76].

In this setting, artificial intelligence could contribute to better allocate individual cases in relation to aggressiveness. Machine learning applications have proved promising in recent published papers, providing insights into how predictive modeling can improve patient perioperative management. Hollon and colleagues reported an accuracy of 87% in prediction of early postoperative outcomes in a retrospective cohort of 400 consecutive pituitary adenoma patients [77]. At the same time, Staartjes et al. demonstrated the usefulness of deep learning to preoperatively predict gross-total resection likelihood in 140 patients, reporting a 91% accuracy [78]. Furthermore, using radiomics, MRI proved accurate in predicting non-functioning pituitary adenomas subtypes [79] and cavernous sinus invasion [80].

To our knowledge, this is the first study investigating proliferative index prediction using a radiomics analysis, possibly affecting surgical approach and postoperative management**.** Interestingly, MRI had already proved promising in proliferative index prediction, using diffusion weighted imaging. Indeed, a strong correlation of ADC values and ki-67 has been recently reported [81]. Our study demonstrates that data mining from non-diffusion-weighted sequences can provide similar results.

A frequent limitation of radiomic machine learning studies is applicability across different sites, varying scanner vendors and field intensity. We chose to analyze images acquired on both 1.5 and 3 T scanners by different vendors, suggesting that our results could more easily be confirmed using exams from different Institutions/equipment. Obviously, this approach requires accurate pre-processing of images to reduce scanner-induced variability.

Another recurring issue in radiomics and machine learning applications is optimal feature number choice. It depends on both sample size and algorithm employed [82]. As texture analysis often yields very large datasets, data dimensionality reduction methods are necessary to select optimal subsets for the proposed classification problem. As error distribution usually cannot be known prior to classification, it is best to test a wider range of feature-set sizes derived from different selection methods. In our case, when feature selection output was represented by a ranking, a range of 4-6 parameters was used to create distinct sets. Among these, the 4-feature set ranked by direct correlation with high Ki-67 expression proved most effective in combination with *k*-NN.

Recently, a growing interest has been shown for deep learning applications in medical imaging [83]. Their complex technical aspects are surely more fascinating than simpler machine learning algorithms such as the *k*-NN we employed. On the other hand, deep learning presents its own sets of issues, such as the "black box" nature of its feature extraction

and selection as well as decision process, limiting software assumption correctness assessment and subsequent wide scale applicability. Furthermore, data required for training such networks is exponentially larger and computational time is also increased compared to post-processing pipelines such as the one we used. For these reasons, it would be more correct to start by using simpler, less resource-intensive machine learning methods, reserving more complex approaches in case satisfactory results are not obtainable by other means.

This study has some limitations which have to be pointed out. A further study on a more numerous population is necessary to further validate and possibly expand these results. Only T2 weighted images were used as contrast enhanced sequences were not taken into consideration due to the presence of both Gradient Echo and Spin Echo sequences; DWI was not performed for all lesions. However, obtaining valuable data without contrast agent administration could represent an added value. Finally, while very good, feature reproducibility was only tested after their selection on a subset of patients.

Obviously, proliferative index is only one of many aspects to be taken into consideration. However, with the proposed approach, from a T2-weighted sequence it might be possible to obtain data concerning size, invasiveness and proliferative index, as well as information on secretory activity [37] and on collagen content [84], useful for predicting tumor consistency. Regarding the last, a recent study by Zeynalova [85] employed a Neural Network on data extracted from 55 pituitary adenoma patients to assess, with good results (72.5% accuracy). It is interesting to note that only T2 images were employed, as done in our study. T2-weighted MR images also proved effective in predicting response to somatostatin analogues in patients with acromegaly and GH-secreting pituitary macroadenoma using a radiomic machine learning approach [86, 87].

For this reason, a possible future application of artificial intelligence in the study of macroadenomas could derive from consideration, in addition to data originating from advanced image analysis (intensity, texture, shape and wavelet), of clinical data, immunohistotype, proliferative indices and invasiveness parameters (cavernous sinuses, sphenoid sinus), also evaluating other omics such as proteomics and genomics, to improve lesion classification and disease treatment choice.

3.5 Conclusion

We analyzed pituitary adenoma proliferative index class preoperative prediction, based on T2-weighted MR imaging. Our findings might aid the surgical strategy making a more accurate preoperative lesion classification and allow for a more focused and cost-effective follow up and long-term management.

4. Pituitary adenoma consistency

4.1 Introduction

Pituitary adenomas are frequent tumors of the pituitary gland. Although most pituitary macroadenomas have a soft consistency, some are rather fibrous and therefore more challenging to remove by transsphenoidal adenomectomy. Indeed, tumor consistency has been reported as one of the principal determinants of transsphenoidal surgery success rate [88]. For this reason, the ability to preoperatively assess adenoma consistency could improve surgical planning and reduce complication rate and risk of residual tumor presence [89].

Radiomics, consisting of conversion of images into mineable data and subsequent analysis for decision support, has been gaining attention in recent years [13]. In particular, texture analysis is a post-processing technique allowing for quantitative description of pixel gray level heterogeneity. More recently, texture analysis-derived features have been used in association with data mining and machine learning algorithms, aiding in the interpretation of a large amount of information produced. Machine learning (ML) is the branch of artificial intelligence including algorithms capable of modeling themselves and improving in accuracy by analyzing datasets, without prior explicit programming [53]. It leads to the creation of predictive models that are able, among other tasks, to solve classification problems. The usefulness of the radiomic approach is being assessed in different fields of radiology [56, 58, 90–92].

Our aim was to assess the accuracy of a ML model trained on radiomic data mined from MRI exams to predict pituitary macroadenoma surgical consistency prior to an endoscopic endonasal procedure.

4.2 Methods and materials

4.2.1 Patient population

This retrospective study was conducted in accordance with the 1964 Helsinki Declaration and its later amendments. The local Institutional Review Board gave its approval and waived the need for informed consent. We reviewed all patients referred to our Institution for endoscopic endonasal pituitary adenoma removal (January 2013-December 2017). Those with history of previous treatment for pituitary adenoma (radiation or medical therapy) at the time of MRI, lesions smaller than 10 mm, extensively necrotic or hemorrhagic areas or significant artifacts on the images used for the analysis were excluded.

4.2.2.Consistency assessment

All patients were operated on by two neurosurgeons with over 10 years of experience in a third level referral center in the field of pituitary surgery [93]. Tumor consistency, classified as soft or fibrous, was assessed in blinded doublecheck by the two surgeons according to the lesions' inner surgical features. In detail, adenomas easily removable with conventional maneuvers of curettage and suction were defined as soft; more resistant ones, difficult to remove and thus requiring more complex maneuvers, such as extracapsular dissection, were classified as fibrous [59, 94–96].

4.2.3 Image acquisition

All patients underwent MRI exams either on a 1.5 (Gyroscan Intera, Philips, Eindhoven, The Netherlands) or 3 Tesla MR scanner (Magnetom Trio, Siemens Medical Solutions, Erlangen, Germany). The imaging protocol always included a coronal T2-weighted (T2-w) Turbo Spin Echo sequence whose detailed parameters are reported in table 4.1.

Table 4.1. Coronal T2-weighted Turbo Spin Echo sequence parameters for 1.5 and 3 Tesla scanners.

	TR	**TE**	**FOV**	**matrix**	**thk**	**ETL**	**Slice gap**	**Acquisition time**
1.5 Tesla	2600 ms	89 ms	180x180 mm	288x288	3 mm	17	no gap	2 min 17 s
3 Tesla	3000 ms	98 ms	200x200 mm	384x384	3 mm	18	no gap	3 min 22 s

TR: Repetition Time; TE: Echo Time; FOV: Field of View; thk: slice thickness; ETL: echo train length.

4.2.4 Handcrafted radiomics

Adenomas were manually annotated by a neuroradiologist (8 years of experience) by placing a 2D polygonal region of interest (ROI) on the coronal slice of maximum lesion extension on a freely available segmentation software (ITKSnap v3.8.0) (Figure 4.1). Two other readers (both > 5 years' experience) also performed lesion segmentation on all patients, blinded to the first neuroradiologist's ROI placement, to perform radiomic feature stability testing.

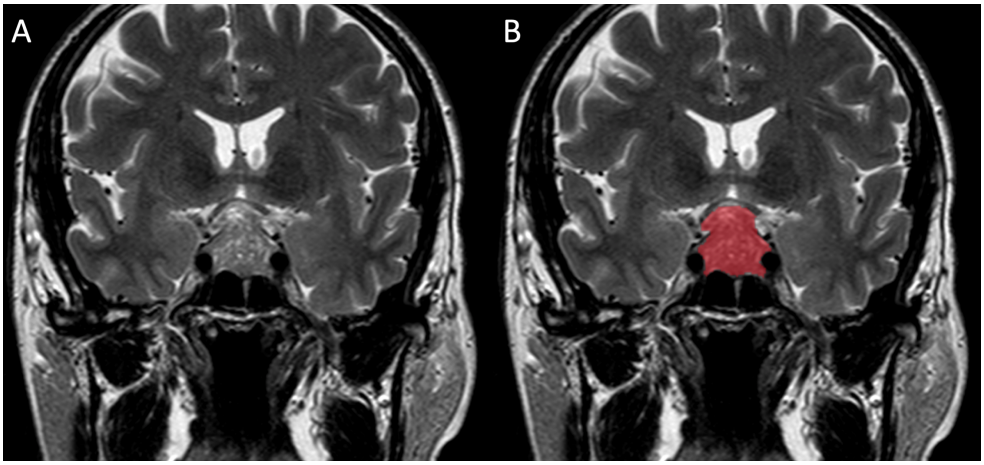

Fig. 4.1. Pituitary macroadenoma segmentation example on coronal T2-weighted (A), showing hand-drawn ROI placement (B).

A freely available, well-established and open-source Python software platform was used for image pre-processing and radiomic parameter extraction (Pyradiomics, v2.2.0). First of all, images and ROIs were resampled to a 2x2x2 mm isotropic voxel, as required for further pre-processing (i.e., correct use of image filters). All voxel intensity values were also normalized by subtracting the mean intensity and dividing by the standard deviation and discretized by using a fixed bin width (=3). Filtered images were also employed for feature extraction in addition to the pre-processed original T2-w ones. In particular, a Laplacian of Gaussian filter, with sigma values ranging from 2.0 (most fine texture) to 3.5 (most coarse) in 0.5 increments, and all available combinations of wavelet decomposition high and low-pass filtering in the x, y and z dimensions. While 2D ROIs were drawn, we still chose to employ a 3-dimensional wavelet decomposition as after resampling the software will detect an ROI z-axis value >1. This is not an issue for the analysis as we excluded shape features and the following feature selection steps will remove all redundant parameters that could have been extracted from similar wavelet decomposition-derived images.

4.2.5 Data Mining and Machine Learning

Initial assessment and processing of the extracted data was performed on Python in particular using the numpy, pandas and scikit-learn packages. First of all, the intraclass correlation coefficient (ICC) was calculated for each parameter as extracted using ROIs from the three readers. A two-way, absolute agreement and single rater ICC was employed and only features with values ≥0.75 were considered stable. Non-informative, low variance (variance ≤0.1) features were also excluded from the dataset. Then, a pairwise correlation matrix was calculated for these in order to remove all features with an intercorrelation ≥0.8. As we expect an unbalanced dataset due to the relative rarity of fibrous adenomas compared to soft ones, the Synthetic Minority Oversampling Technique (SMOTE) was employed [42]. Then, 80% of the data was used for hyperparameter tuning via stratified 5-fold cross-validation while a 20% hold-out set was

employed for its testing on unseen data. In detail, the following steps were exclusively performed on the first set. A normalization scaler was calculated to remove biases due to feature scale and was later applied to the hold-out test set. Finally, recursive feature elimination (RFE), employing a logistic regression algorithm and stratified 5-fold cross-validation, was used to select the better performing feature subset.

The resulting data was used to train an ensemble learning meta-algorithm, the Extra Trees Classifier (ET). These often demonstrate good performance on radiomic medical image data [43]. Its performance for consistency prediction was finally assessed on the test set.

Accuracy metrics were obtained using the scikit-learn package and further analyzed on the R software (R for Unix/Linux, version 3.4.4, the R Foundation for Statistical Computing, 2014). In particular, DeLong's test (pROC package) was used to obtain 95% confidence intervals (95%CI) of the area under the Receiver Operating Characteristic curve (AUC) and the confusion matrix function (caret package) those of the classifier's accuracy and compare its performance to the no information rate.

The described radiomics workflow pipeline is illustrated in Figure 4.2.

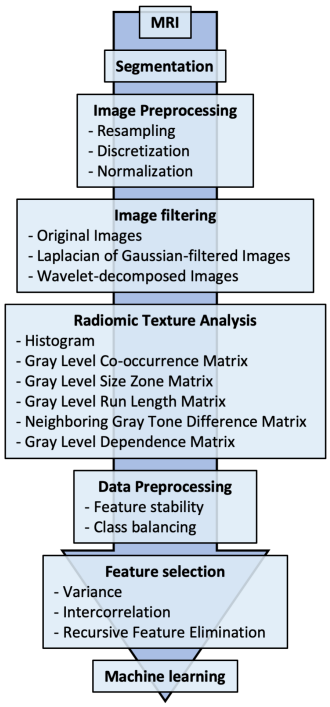

Fig. 4.2. Radiomics workflow pipeline.

4.3 Results

According to selection's criteria, 89 patients were included in this study; 51 were males and 38 females, with mean age 52.17 ± 14.64 years (range 16-80). Average lesion size was 25 ± 8 mm (range 8-46 mm). The pituitary lesions were

classified as soft in 68 patients and fibrous in the remaining 21. In detail, 19 soft (8 ACTH, 7 GH, 4 PRL) and 6 fibrous (2 ACTH, 2 PRL, 1 GH, 1 TSH) were functioning (25/89, 28% in total). In none of the cases there was discordance among the neurosurgeons in lesion classification. Patient population clinical data are presented in Table 4.2.

Table 4.2. Patient population clinical data.

	Tumor consistency		
	Total (n = 89)	**Soft (n = 68)**	**Fibrous (n = 21)**
Age (mean) (year)	52.2 ± 14.6	53.2 ± 15.5	54.6 ± 14.6
Sex			
Males (n) (%)	51 (57%)	39 (57%)	12 (57%)
Females (n) (%)	38 (43%)	29 (43%)	9 (43%)
Tumor type			
Functioning (n) (%)	25 (28%)	19 (28%)	6 (21%)
Non-functioning (n) (%)	64 (72%)	49 (72%)	15 (78%)

A total of 1118 texture features were extracted, including first and higher order texture features from the original and filtered images. The correlation cluster map of the extracted features is shown in Figure 4.3. Their detailed description is available in the online Pyradiomics documentation (https://pyradiomics.readthedocs.io/en/latest/features.html). After feature stability analysis, 741 were retained for the subsequent steps. Of these, 4 had low variance while 625 were highly-intercorrelated. RFE then identified a 14-feature subset as most accurate (Figure 4.4), whose features are listed below:

- Original firstorder Energy
- Original firstorder Kurtosis
- Original firstorder Skewness
- Log-sigma-3-0-mm-3D firstorder Skewness
- Log-sigma-3-0-mm-3D glcm Imc1
- Log-sigma-3-0-mm-3D glszm SizeZoneNonUniformityNormalized
- Log-sigma-3-5-mm-3D firstorder Minimum
- Wavelet-LHL firstorder InterquartileRange
- Wavelet-LHL gldm DependenceVariance
- Wavelet-LHH firstorder 10Percentile
- Wavelet-LHH firstorder Maximum

- Wavelet-HLH firstorder Maximum
- Wavelet-HLH firstorder Mean
- Wavelet-HLH firstorder Minimum.

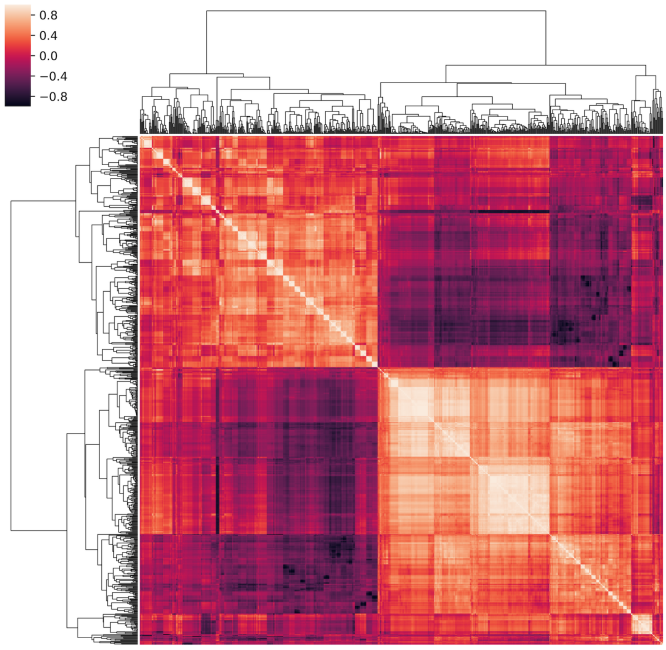

Fig. 4.3. Hierarchically clustered heatmap of the feature correlation matrix. Features with an intercorrelation above the selected threshold (≥ 0.8) were removed from the dataset.

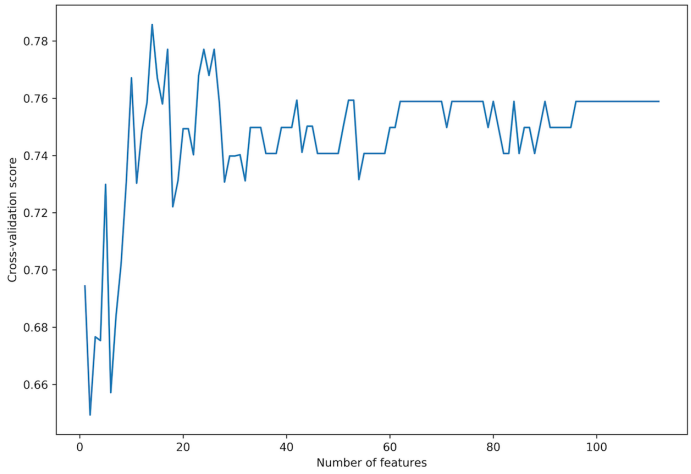

Fig. 4.4. Plot of the feature selection process by recursive feature elimination. The x-axis contains the total number of features, from which one is removed at each iteration. The y-axis contains the average cross-validation score for each feature total.

The ET model obtained an overall accuracy, in terms of correctly classified lesions, of 86% (±10%) in the training set cross-validation. The classifier tuned parameters are reported in the supplementary materials. In the test set, the accuracy was of 93% (95%CI = 77-99%), sensitivity of 100% and specificity of 87%. The AUC was of 0.99 (95%CI = 0.97-1.00) (Figure 4.5), equal to the area under the Precision-Recall curve (0.99), often used in binary ML classifications (Fig AUC). The classifier was significantly better ($p = 8e^{-6}$) than the no information rate. The confusion matrix and detailed accuracy metrics are shown in Tables 4.3 and 4.4.

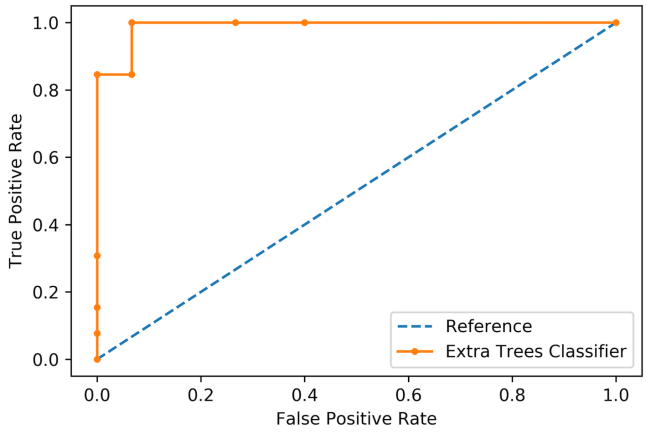

Fig. 4.5. Receiver Operating Characteristics curve of the Extra Trees Classifier accuracy.

Table 4.3. Confusion matrix for the test group.

		Predicted class	
		Soft	Fibrous
Actual class	Soft	13	2
	Fibrous	0	13

Table 4.4. Extra Trees Classifier accuracy metrics

Class	**Recall**	**Precision**	**F-score**	**AUC**	**AUPRC**

Soft	0.87	1.00	0.93	0.99	0.99
Fibrous	1.00	0.87	0.93	0.99	0.99
WAvg	0.94	0.93	0.93	0.99	0.99

WAvg = weighted average, AUC = Area under the Receiver Operating Characteristic curve; AUPRC = Area under the Precision-Recall Curve.

4.4 Discussion

Preoperative assessment of pituitary macroadenoma consistency is useful for planning surgical approach and reducing residuals and recurrence's rate. For this reason, several studies have investigated the correlation between preoperative MRI features and tumor hardness. In particular, there are conflicting studies on the value of the relative signal on T2-weighted MRI and the macroadenoma consistency, with some works demonstrating a positive correlation between low signal and hardness [97–100] and other concluding that relative signal intensity values do not correlate [101–104]. Indeed besides collagen amount, which mainly correlates with the hardness, other factors such as intratumoral hematoma, amyloid, iron, calcification or protein-rich fluid may affect the T2 signal intensity [105].

Diffusion weighted imaging ability to predict tumor consistency also showed divergent results, both indicating a significant correlation [88, 106] and not [101, 107]. Furthermore, the lower spatial resolution and the presence of susceptibility artefacts in the sellar region related to bone and sinus pneumatization limit the use of this technique.

Finally, in two studies by Romano et al. and Yamamoto et al. contrast-enhanced MRI showed a strong correlation for tumor consistency [89, 108]. Perfusion imaging parameters have also been investigated as possible biomarkers of pituitary macroadenoma consistency, but no added value was found compared to precontrast T1-weighted images [109].

A more interesting advanced technique in this setting is represented by MR elastography. Pituitary adenoma stiffness was found to correlate with their consistency and, if it became widely available, could offer additional data to mine with a radiomic approach [110, 111].

Regarding texture analysis there are only three studies exploring this issue, to the best of our knowledge. In the first, Rui et al. explored the value of MRI texture analysis in assessing pituitary macroadenoma consistency, obtaining good accuracy values [112]. However, this study was conducted using contrast-enhanced 3D-SPACE images and without a ML approach. Fan and colleagues explored this issue in acromegalic patients using ML for radiomic feature selection prior to building a nomogram obtaining an AUC of 0.81 [113]. Zeynalova et al also performed an analysis on ML preoperative evaluation of pituitary macroadenoma consistency [85]. Their study presented some similarities with our own. They also used PyRadiomics for feature extraction from bidimensional ROIs, although they utilized different

sigma settings (2, 4 and 6 mm) for the LoG filter and obtained a total of 162 parameters. The lower number of features is probably due to their exclusive focus on first order histogram-derived ones. These are more reproducible but convey less information on tissue texture compared to higher order parameters. As their in-plane resolution was higher (0.5x0.8 mm) they were able to use a 1x1 mm resampling size compared to our 2x2 mm. It is interesting to note the use of a very narrow bin width value of 0.06, as the number of bins should not exceed 128, following the developer recommendations. They also performed a feature robustness assessment with our same ICC threshold, while their inter-correlation threshold was lower (0.7 vs 0.8). After data dimensionality reduction, they identified 6 informative features using the Weka data mining platform and a wrapper-based selector. In our study the entire analysis was conducted using the scikit-learn Python package. Some other major differences are represented by the use of cross-validation, without further assessment on a separate test set. Their reported accuracy is 72.5%, with an AUC of 0.71. Therefore, our algorithm presents a clearly superior performance. This could be in part explained by their use of a Multilayer Perceptron neural network, which may not be the best suited algorithm for a small dataset obtained from 55 patients. Finally, Zenyalova and colleagues also used collagen amount within the tumor on histopathological examination for their reference standard. As consistency information is mainly useful for surgical strategy planning, we believe intraoperative consistency assessment represents a more practical and useful reference standard as the final recipient of the information should be a neurosurgeon. As the two neurosurgeons involved in our study never had disagreements, we also found this assessment to be reproducible.

By analyzing our confusion matrix, it can be seen that the mistakes made by the classifier were 2 cases of soft lesions identified as fibrous. Given the clinical setting of our investigation, this kind of error is somewhat more acceptable than a false negative, as it would be more auspicable to sometimes overestimate the difficulty of a surgery rather than the opposite.

In our study, we chose to employ an ET ML algorithm. This belongs to the decision tree ensemble methods, in particular constituted by a large number of highly randomized decision trees which are fitted on data sub-samples. Each of these outputs a prediction, and a majority vote determines the final outcome. Ensemble learning is based on the assumption that a decision by committee made by a large number of weak classifiers will perform better than a single algorithm. A sufficient diversification of the random trees included in the ET is guaranteed by random sampling, with replacement, of patients from the training dataset (bootstrap aggregation or bagging) and of their available features (n = 3 in our case). This in turn ensures low correlation of each tree, improving the ET's overall performance [43]. As the dataset lesion classes were imbalanced, SMOTE was employed. This is a known solution to address this issue and has demonstrated its value in the setting of medical imaging radiomic ML analysis [42, 114–116].

We have chosen a handcrafted radiomics approach rather than a fully automated deep learning one as this gave us better control on the initial data analysis and following ML model construction. Both approaches have been object of discussion in current literature as they possess peculiar merits and limitations. It is our belief that a handcrafted analysis is more appropriate for relatively smaller datasets as it allows greater involvement of radiologists and better understanding of the whole pipeline. Only when extremely large datasets will become available in medical imaging, as in other fields, the less time-consuming completely neural network-based approach will be a practical necessity. Until then, the value of greater involvement of the radiologist and finer quality control of patient or lesion data outweigh the larger amount of time needed to extract medical imaging radiomics. Furthermore, medicine and especially treatments are evolving in the direction of precision, patient-tailored therapies. Contrary to the current desire in radiology to aggregate as many patients as possible to train ML algorithms, this determines a need to work with ever smaller patient subgroups within each pathological entity. Therefore, a future with space for both deep learning software to apply on large populations and engineered approaches for more specific tasks can be envisioned.

Our study has some limitations which have to be acknowledged. As is often the case for ML, future studies on larger populations is necessary to confirm and possibly expand our results. The need for oversampling given the unbalanced nature of the classes further highlight this necessity but was expected given epidemiological data. Only T2-weighted images were used, without investigating the added value of other sequences. However, obtaining valuable data without contrast agent administration and streamlining the pipeline to incorporate a single MRI sequence could also represent an added value. Furthermore, considering previous works, T2-weighted MRI alone proved effective to provide data concerning proliferative index [19], secretory activity [37] and response prediction to somatostatin analogues in patients with acromegaly and GH secreting pituitary macroadenoma [86, 87].

4.5 Conclusion

The ML model trained on radiomic data extracted from T2-weighted MRI demonstrated a high accuracy in the classification of soft and fibrous pituitary macroadenomas. Therefore, this tool could prove valuable in the presurgical planning of these patients if further developed and validated on larger datasets.

5. Artificial intelligence and pituitary adenomas: a review

5.1 Introduction

Pituitary adenomas are benign tumors accounting for 15%-20% of all intracranial neoplasms, with an incidence of 80–90 cases per 100000 population [1, 117]. Microadenomas are defined as tumors < 10 mm in maximum diameter, whereas larger adenomas are considered macroadenomas. Their peak age of presentation is between the fourth and seventh decades. Almost two-thirds of pituitary adenomas are hormone-secreting, prolactin most commonly, followed by growth hormone, corticotropin and thyrotropin, and cause typical hypersecretion syndromes. Non-functioning, small intrasellar tumors can be clinically silent and diagnosed only as incidental magnetic resonance findings, while bulky pituitary macroadenomas typically present with mass effect signs, such as headache, visual disturbances, and hypopituitarism [118, 119]. The 2017 World Health Organization (WHO) classification adopted pituitary adenohypophyseal cell lineage as the main principle guiding the classification of adenomas. According to this principle we distinguish the acidophilic lineage (in which the involved transcription factor is PIT1), the corticotrope lineage (TPIT transcription factor), and the gonadotroph lineage (SF1 transcription factor). Null-cell adenomas (NCAs) are now defined as tumors that have no immunohistochemical evidence of cell-type-specific differentiation considering both pituitary hormones and transcription factors. Furthermore, in the new WHO classification the term “atypical adenoma” has been abandoned and replaced by “high risk adenoma”, in reference to tumors with high proliferation index and tendency to invasion. In particular, emphasis is placed on the evaluation of tumor proliferation (mitotic count and Ki-67 index), tumor invasion, and on special adenomas variants for which clinical behavior has been shown to be more aggressive due to their intrinsic histological features: lactotroph adenoma in men, sparsely granulated somatotroph adenoma, the silent corticotroph adenoma, the Crooke’s cell adenoma and the plurihormonal PIT1-positive adenoma [76]. Magnetic resonance imaging (MRI) is the investigation of choice for a complete evaluation of pituitary adenomas [120]. Various parameters regarding the extent, consistency, and contrast enhancement can be analyzed in order to help neurosurgeons in planning an appropriate surgical approach and long-term follow-up [107].

Attempting to predict invasion (cavernous and/or sphenoid sinus involvement) based on imaging is an important challenge. The Knosp classification is one of the more commonly used systems to determine the likelihood of cavernous sinus invasion by pituitary macroadenomas, but the highest accuracy of this grading system is observed in extreme cases of overt invasion or non-invasion, while sensitivity and specificity are low in intermediate cases [121, 122].

Tumor consistency in pituitary macroadenomas has been known to be one of the main factors that determine the success rate of the transsphenoidal approach. The role of MRI in predicting the consistency of pituitary macroadenomas is

controversial. Several studies suggested that relative signal intensity or signal intensity ratio on T2-weighted MRI correlates with the tumor consistency, while some others concluded that they have no predictive value [97, 103, 105]. A similar controversy has been reported in several studies which investigated the usefulness of diffusion-weighted imaging in tumor consistency prediction [101, 106, 123].

Considering the above, it is still difficult to achieve an early identification of clinical and radiological features suggestive of an aggressive behavior, characterized by rapid growth, local invasion, and high ki-67 proliferation index. In this setting, artificial intelligence (AI) has proved promising in recently published papers. Machine learning (ML) is a subfield of AI that employs algorithms to allow computers to learn directly from the data and subsequently perform predictions without explicit prior programming. The potential impact of ML on medicine, and particularly medical imaging, is relative to its ability to analyze large datasets including gray level textural features that humans do not consciously assess. Unlike classical rule-based algorithms, machine learning can take advantage of increased exposure to new data and learn over time [124]. ML techniques can be further divided into supervised, unsupervised learning and reinforcement learning [13, 14, 125, 126]. In supervised learning there is a ground truth which is directly used to guide the algorithm training process. The goal of the resulting model is usually to learn a general rule that maps inputs to outputs and is applicable to new, unseen cases. In unsupervised learning there is no preliminary labeling and therefore its goal is to cluster the given inputs based solely on the underlying data structure. Finally, reinforcement learning consists of a computer program performing an assigned task in a dynamic environment and consequently receiving feedback as a positive or negative reinforcement. To improve algorithm's performance, these approaches can be combined, some examples are semi-supervised, self-supervised and multi-instance learning.

Deep learning (DL) is an ML approach employing networks inspired by brain's structure, with a large number of simple interconnected units performing complicated tasks. The DL algorithms most applied to medical imaging are convolutional neural networks. Lower level information inputs, derived from imaging data transformed in feature vectors, form connections to the next level or "layer" of neurons. Each neuron in this second layer can combine the inputs from lower level neurons to form a newer, more complex output. As the number of intermediate or hidden layers increases, the final output from the highest layer becomes richer and more complex.

ML tasks are not limited to tumor property prediction but include many possible applications in other medical imaging and daily workflow fields, such as image acquisition, segmentation, image quality analytics, automated dose estimation and radiology reporting [83, 127, 128]. Despite the high number of recent ML successes, there are still many limitations in its clinical use [15, 16, 129, 130]. First of all, an obstacle to AI adoption in the clinical setting is identifiable in its limited interpretability, especially true for DL. Clinicians are consequently reluctant to trust and to adopt something whose decision process is not fully understood. Secondly, ML research has to deal with issues due to the nature of the

health domain, including the lack of large amounts of data, necessary during the training phase, the need for algorithm frequent updating and potential model overfitting.

This review aims to give an overview of the current applications of ML methods in pituitary adenomas evaluation.

5.2 Pretreatment assessment

Tumor consistency is one of the main factors that determine the success-rate of transsphenoidal adenomectomy. For this reason, pre-operative information about tumor consistency would help neurosurgeons in planning the most appropriate surgical approach. Zeynalova *et al* [85] demonstrated the utility of ML-based histogram analysis (from 55 pituitary adenoma patients) to predict tumor consistency and compared it with a conventional signal intensity ratio (SIR) evaluation. Histogram features were extracted from coronal T2-weighted original, filtered and transformed MRI images by manual segmentation. The high dimensionality of the histogram texture features was reduced with reproducibility analysis, collinearity analysis and wrapper-based feature selection. They employed the artificial neural network (ANN) as ML classifier. The reference-standard was consensual evaluations of neurosurgeons and pathologists. For histogram analysis, the ANN correctly classified 72.5% of pituitary macroadenomas with an area under the receiver operating characteristic (ROC) curve (AUC) value of 0.710. As for SIR evaluation, accuracy and AUC values were 74.5% and 0.551, respectively. Considering AUC values, ML-based histogram analysis performed better than SIR evaluation.

Fan *et al* [113] demonstrated how a radiomics model can assist neurosurgeons in predicting tumor consistency in patients with acromegaly before surgery and facilitate the determination of an appropriate therapeutic approach. 158 patients (training group $n = 100$, validation group $n = 58$) were included in this retrospective study, while 30 were enrolled in a prospective multi-center study for model validation. The consistency of the tumor was classified as soft or firm according to the neurosurgeon's evaluation. All patients underwent MRI examination which included T1-, T2- and contrast-enhanced T1-weighted sequences in the coronal plane, used for feature extraction. The radiomics features were collected based on the regions of interest drawn by an expert neuroradiologist and verified by a second expert. Total 1561 quantitative features were collected for every sequence. The radiomics features were determined using the elastic net feature selection algorithm, and the radiomics signature was constructed. Next, a radiomics model was developed using the radiomics signature and clinical characteristics, which were further screened according to the Akaike information criterion. Then, 30 patients with acromegaly from three hospitals were enrolled for multicenter validation of the model. The prediction accuracy was then evaluated through ROC analyses and associated classification measures. The radiomics model constructed in this study showed an AUC of 0.83 and 0.81 in the primary and validation cohorts, respectively. In conclusion, this model was convenient to use and could accurately predict the tumor consistency in a multicenter prospective validation before surgery.

The pre-operative prediction of cavernous sinus (CS) invasion by pituitary adenomas (Knosp grade 2-3) can help neurosurgeons in planning the surgical approach, follow-up, and long-term management. Niu *et al* [80] used a radiomics method to predict CS invasion, enrolling 194 patients with Knosp grade 2-3 (training set $n = 97$; test set $n = 97$) and extracting 2553 quantitative imaging texture features from contrast-enhanced T1- and T2-weighted MR images. A linear support vector machine (SVM) was used to fit the predictive model, then a nomogram was constructed incorporating radiomics signature and clinico-radiological risk factors. Radiomics model yielded an AUC of 0.852 and 0.826 for the training and test set, respectively. The nomogram yielded an AUC of 0.899 in the training test and 0.871 in the test set.

According to the 2017 WHO classification, "high risk" pituitary adenomas are tumors with rapid growth, radiological invasion, and high Ki-67 proliferation index. MRI had already proved promising in proliferative index prediction, using diffusion-weighted imaging. Indeed, Tamrazi *et al* [81] performed a retrospective review of diffusion imaging and immunohistochemical characteristics of 17 with pituitary macroadenomas and demonstrated an inverse relationship between apparent diffusion coefficient values and Ki-67. In this context, machine learning can be effective for the early identification of "high risk" adenomas and could allow making a more accurate pre-operative assessment and long-term follow-up. Regarding the last, a recent study by Ugga *et al* [19] employed ML analysis of texture-derived parameters from pre-operative coronal T2-weighted MR images. A total of 89 patients that underwent endoscopic endonasal procedure were included. Pituitary adenomas were classified in high versus low Ki-67 proliferation index according to pathological data. Total 1128 features were extracted, and different supervised feature selection methods were employed to select the most informative features. A k-nearest neighbors (k-NN) classifier was used to predict the proliferative index, then algorithm validation was performed with a train-test approach. The accuracy of k-NN in the test group was 91.67% of correctly classified patients.

Non-functioning pituitary adenomas are a huge group of adenomas and can be divided in NCAs, oncocytomas and gonadotrophic adenomas. Patients with NCAs are more likely to respond to neo-adjuvant radiotherapy, so radiomics could play a role in discriminating preoperatively NCAs from other subtypes. Zhang *et al* [79] enrolled 112 patients (training set $n = 75$; test set $n = 37$) with non-functioning pituitary adenomas who underwent MR examination. In their retrospective study a SVM trained a predictive model that was validated using a ROC analysis on an independent test set. Then, a nomogram was constructed incorporating clinical characteristics and the radiomics signature for a more individualized predictive model. T1-weighted image features yielded an AUC value of 0.83 and 0.80 for the training and test sets, respectively. The nomogram incorporating sex and the T1 radiomics signature yielded good calibration in the training and test sets (concordance index of 0.854 and 0.857, respectively).

Somatostatin analogues (SAs) response prediction is an essential information in acromegalic patient medical treatment in the presence of GH-secreting pituitary adenomas. Indeed, this medical treatment can improve the surgical outcome, but it is burdened by high costs. Heck *et al* [86] showed how quantitative analysis of T2-weighted MR images could predict response to SAs in patients with acromegaly. However, they verified that conventional visual T2 intensity assessment achieved similar results. This retrospective cohort study included 58 newly diagnosed patients. Parameters from the T2 histogram analyses (T2 intensity ratio and T2 homogeneity ratio) were correlated to visually assessed T2 intensity (hypo-, iso- or hyperintense), baseline characteristics, response to SA treatment, and histological granulation pattern (anti-Cam5.2). T2 intensity ratio was lowest in the hypointense tumors and highest in the hyperintense tumors. T2 intensity at baseline correlated with reduction in GH (r = -0.67) and IGF-1 (r = -0.36) after primary SA treatment (n = 34). The T2 homogeneity ratio correlated with adenoma size reduction (r = -0.45). Sparsely granulated adenomas, which are typically associated to resistance to SAs, had a higher T2 intensity than densely or intermediately granulated adenomas. In conclusion, using T2 histogram analyses the authors found that high T2 intensity correlated with more aggressive adenoma subtypes, larger adenoma size, lower GH and IGF-1 production, and blunted response to an octreotide test dose at baseline. Moreover, a better biochemical response to SA therapy was observed in adenomas with low T2 intensity. In their retrospective study, Kocak *et al* [87] demonstrated the potential role of ML-based high-dimensional quantitative texture analysis (qTA) in predicting SAs response in acromegalic patients with a GH-secreting pituitary adenoma. They showed how ML performs better than relative signal intensity (rSI) evaluation or immunohistochemical granulation pattern evaluation. Coronal T2-weighted images of 47 patients (24 SA responsive and 23 SA resistant patients) were used for qTA and quantitative and qualitative rSI evaluation, while the immunohistochemical evaluation was based on the granulation pattern of the adenomas. ML classifiers were k-NN and C4.5 algorithm. The reference standard was the biochemical response status (6 months post-therapy). Predictive performance of qTA was compared with that of the quantitative and qualitative rSI and immunohistochemical evaluation. For the qTA, k-NN correctly classified 85.1% macroadenomas with an AUC of 0.847. The accuracy and AUC ranges of the other methods were lower, equal to 57.4/70.2% and 0.575/0.704, respectively.

Pituitary tumor growth can lead to compression of the anterior visual pathways, leading to visual impairment, which is the most common and earliest symptom in this pathology. In their retrospective study Lilja *et al* [131] demonstrated that diffusion Tensor imaging (DTI) and a prediction model may be an additional diagnostic tool that provides objective data about visual pathway injury, guiding treatment decisions. Total 23 patients with pituitary adenomas and 20 healthy patients underwent a complete neuro-ophthalmological examination and an MRI study, which included 3D T1-weighted and DTI sequences. A prediction model using logistic regression was constructed to test the capability of DTI parameters to correctly classify a subject as a patient (before surgery) or a control. Total 12 features quantifying mean

DTI parameters from the optic tract regions were included. Based on the axial diffusivity and fractional anisotropy, the prediction model could separate patients from controls with high sensitivity. The prediction model correctly classified all patients with visual field defects (sensitivity = 1.0), 9 of 12 patients without visual field defects (sensitivity = 0.75), and 17 of 20 controls (specificity = 0.85).

5.3 Neurosurgical outcome prediction

Gross total resection (GTR) is the main surgical goal in transsphenoidal surgery for most pituitary adenomas. Predictive analytics for GTR may help in surgical decision-making, especially in intermediate cases (Knops grade 2-3A). In their retrospective study, Staartjes *et al* [78] investigated the potential value of deep neural network for predicting GTR in comparison with the Knops classification and logistic regression. They enrolled a total of 140 patients who underwent endoscopic transsphenoidal surgery and trained a deep neural network to predict GTR from 16 preoperatively available neuro-radiological and procedural variables. Their DL model (AUC = 0.96; accuracy = 91%; sensitivity = 94%; specificity = 89%) outperformed both the Knosp classification (AUC = 0.87; accuracy = 81%; sensitivity = 92%; specificity = 70%) and logistic regression (AUC = 0.86; accuracy = 82%; sensitivity = 81%; specificity = 83%).

In their retrospective study, Liu *et al* [132] aimed to develop machine learning-based predictive models to evaluate Cushing disease recurrence after initial transsphenoidal surgery and to investigate their performance. Seventeen radiomic features including tumor volume computed from pre-operative MRI (contrast-enhanced T1-weighted MRI) and other pre/post-operative clinical variables were evaluated. Five supervised ML algorithms, including decision tree, gradient boosting decision tree, random forest (RF), adaptive boosting, and extreme gradient boost and 2 conventional models (Logistic regression, Naïve Bayes) were applied. Models were evaluated based on their AUC. The study demonstrated that ML-based predictive models for neurosurgical outcomes performed well, better than some conventional models such as logistic regression. Using 17 variables, several ML-based predictive models for recurrence were developed, and most of them (4/5) maintained high performance, with AUCs ranging from 0.694 to 0.781 which were much higher than that of conventional statistics. The best performance (AUC = 0.781) was obtained introducing 8 variables to RF algorithm, which was much better than that of logistic regression (AUC = 0.684) and that of using only postoperative morning serum cortisol (AUC = 0.635). According to the feature selection algorithms, the top predictors were age, postoperative serum cortisol, and postoperative ACTH.

Cerebrospinal fluid (CSF) fistulas remain a major complication of transnasal transsphenoidal surgery for pituitary adenoma. Staartjes *et al* [133] developed a neural network–based model with the aim of classifying pituitary surgeries in having high versus low-risk of CSF leak. From a prospective registry, 154 patients who underwent endoscopic transnasal transsphenoidal surgery for pituitary adenoma were identified and underwent an MRI study. Moreover, risk

factors for intraoperative CSF leaks were identified using conventional statistical methods. Selected features included both imaging features from inter-carotid distances and other clinical pre/post-operative variables. The authors built a predictive model for intraoperative CSF leaks based on a deep multilayer perceptron with 5 hidden layers. The deep neural network–based prediction model could identify patients at high risk for intraoperative CSF leak. It correctly classified 88% of patients in the test set, with an AUC of 0.84. Sensitivity and specificity were high, of 83% and 89% respectively. The positive predictive value was 71%, negative predictive value was 94%, and F1 score was 0.77.

5.4 Conclusion

We reviewed a set of articles related to ML applications in pituitary adenomas. These studies showed that ML has a certain potential to improve the diagnostic performance of MRI in pre-treatment assessment and neurosurgical outcome prediction. In current studies there is not a standardized procedure, ML methodologies vary a lot, different types of classifiers are applied and only a few models are validated on an external set. The major limits of these studies are the replicability and generalizability. Publicly available datasets are needed, and clinical applicability still requires more robust validation across different sites, scanner vendors and field intensity. However, the research in the years is growing rapidly and ML software can be a potential power to facilitate better clinical decision making in pituitary tumor patients.

6. Conclusions and Perspectives

The presented studies demonstrate the promising role of radiomics in the preoperative evaluation of patients with pituitary macroadenomas. In fact, using radiomic features extracted from T2-weighted images, and analyzing them with a machine learning approach, we obtained high accuracy values in predicting functional status, proliferative index and tumor consistency. Although further studies are needed to validate the reported results, these preliminary studies lay the foundations for the practical use of radiomics in a clinical setting, which could influence patient management toward precision medicine.

Bibliography

1. Fernandez A, Karavitaki N, Wass JAH (2010) Prevalence of pituitary adenomas: a community-based, cross-sectional study in Banbury (Oxfordshire, UK). Clin Endocrinol (Oxf) 72:377–382 . doi: 10.1111/j.1365-2265.2009.03667.x
2. Raappana A, Koivukangas J, Ebeling T, Pirilä T (2010) Incidence of Pituitary Adenomas in Northern Finland in 1992–2007. J Clin Endocrinol Metab 95:4268–4275 . doi: 10.1210/jc.2010-0537
3. Gruppetta M, Mercieca C, Vassallo J (2013) Prevalence and incidence of pituitary adenomas: a population based study in Malta. Pituitary 16:545–553 . doi: 10.1007/s11102-012-0454-0
4. Tjörnstrand A, Gunnarsson K, Evert M, et al (2014) The incidence rate of pituitary adenomas in western Sweden for the period 2001–2011. Eur J Endocrinol 171:519–526 . doi: 10.1530/EJE-14-0144
5. Agustsson TT, Baldvinsdottir T, Jonasson JG, et al (2015) The epidemiology of pituitary adenomas in Iceland, 1955–2012: a nationwide population-based study. Eur J Endocrinol 173:655–664 . doi: 10.1530/EJE-15-0189
6. Ragel BT, Couldwell WT (2004) Pituitary carcinoma: a review of the literature. Neurosurg Focus 16:1–9 . doi: 10.3171/foc.2004.16.4.8
7. Cattin F (2016) MRI Technique and Radiological Anatomy of the Pituitary Gland. In: MRI of the Pituitary Gland. Springer International Publishing, Cham, pp 1–11
8. Cooper O, Melmed S (2012) Subclinical hyperfunctioning pituitary adenomas: The silent tumors. Best Pract Res Clin Endocrinol Metab 26:447–460 . doi: 10.1016/j.beem.2012.01.002
9. Melmed S (2011) Pathogenesis of pituitary tumors. Nat Rev Endocrinol 7:257–266 . doi: 10.1038/nrendo.2011.40
10. Loeffler JS, Shih HA (2011) Radiation Therapy in the Management of Pituitary Adenomas. J Clin Endocrinol Metab 96:1992–2003 . doi: 10.1210/jc.2011-0251
11. Ding D, Starke RM, Sheehan JP (2014) Treatment paradigms for pituitary adenomas: defining the roles of radiosurgery and radiation therapy. J Neurooncol 117:445–457 . doi: 10.1007/s11060-013-1262-8
12. Go JL, Rajamohan AG (2017) Imaging of the Sella and Parasellar Region. Radiol Clin North Am 55:83–101 . doi: 10.1016/j.rcl.2016.09.002
13. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: Images Are More than Pictures, They Are Data. Radiology 278:563–577 . doi: 10.1148/radiol.2015151169
14. Choy G, Khalilzadeh O, Michalski M, et al (2018) Current applications and future impact of machine learning in radiology. Radiology
15. Romeo V, Ricciardi C, Cuocolo R, et al (2019) Machine learning analysis of MRI-derived texture features to

predict placenta accreta spectrum in patients with placenta previa. Magn Reson Imaging. doi: 10.1016/j.mri.2019.05.017

16. Islam MM, Nasrin T, Walther BA, et al (2019) Prediction of sepsis patients using machine learning approach: A meta-analysis. Comput Methods Programs Biomed 170:1–9 . doi: 10.1016/j.cmpb.2018.12.027
17. Deniz CM, Xiang S, Hallyburton RS, et al (2018) Segmentation of the Proximal Femur from MR Images using Deep Convolutional Neural Networks. Sci Rep 8:16485 . doi: 10.1038/s41598-018-34817-6
18. Ricciardi C, Cuocolo R, Cesarelli G, et al (2020) Distinguishing Functional from Non-functional Pituitary Macroadenomas with a Machine Learning Analysis. pp 1822–1829
19. Ugga L, Cuocolo R, Solari D, et al (2019) Prediction of high proliferative index in pituitary macroadenomas using MRI-based radiomics and machine learning. Neuroradiology 61:1365–1373 . doi: 10.1007/s00234-019-02266-1
20. Cuocolo R, Ugga L, Solari D, et al (2020) Prediction of pituitary adenoma surgical consistency: radiomic data mining and machine learning on T2-weighted MRI. Neuroradiology. doi: 10.1007/s00234-020-02502-z
21. Guerriero E, Ugga L, Cuocolo R (2020) Artificial intelligence and pituitary adenomas: A review. Artif Intell Med Imaging 1:70–77 . doi: 10.35711/aimi.v1.i2.70
22. Gittleman H, Ostrom QT, Farah PD, et al (2014) Descriptive epidemiology of pituitary tumors in the United States, 2004-2009: Clinical article. J Neurosurg. doi: 10.3171/2014.5.JNS131819
23. Ostrom QT, Gittleman H, Liao P, et al (2014) CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2007-2011. Neuro Oncol 16:iv1–iv63 . doi: 10.1093/neuonc/nou223
24. Wilson PJ, Omay SB, Kacker A, et al (2018) Endonasal endoscopic pituitary surgery in the elderly. J Neurosurg. doi: 10.3171/2016.11.JNS162286
25. Solari D, Cavallo LM, Cappabianca P (2014) Surgical approach to pituitary tumors. In: Handbook of Clinical Neurology. pp 291–301
26. Lubner MG, Smith AD, Sandrasegaran K, et al (2017) CT texture analysis: Definitions, applications, biologic correlates, and challenges. Radiographics
27. Malik MM, Abdallah S, Ala'raj M (2018) Data mining and predictive analytics applications for the delivery of healthcare services: a systematic literature review. Ann. Oper. Res.
28. Santini S, Pescape A, Valente AS, et al (2017) Using fuzzy logic for improving clinical daily-care of β-thalassemia patients. In: IEEE International Conference on Fuzzy Systems
29. Improta G, Russo MA, Triassi M, et al (2018) Use of the AHP methodology in system dynamics: Modelling

and simulation for health technology assessments to determine the correct prosthesis choice for hernia diseases. Math Biosci 299:19–27 . doi: 10.1016/j.mbs.2018.03.004

30. Romano M, Bifulco P, Improta G, et al (2013) Symbolic dynamics in cardiotocographic monitoring. In: 2013 E-Health and Bioengineering Conference, EHB 2013

31. Johnson KW, Torres Soto J, Glicksberg BS, et al (2018) Artificial Intelligence in Cardiology. J Am Coll Cardiol 71:2668–2679 . doi: 10.1016/j.jacc.2018.03.521

32. Kumar V, Gu Y, Basu S, et al (2012) Radiomics: the process and the challenges. Magn Reson Imaging 30:1234–1248 . doi: 10.1016/j.mri.2012.06.010

33. Zacharaki EI, Wang S, Chawla S, et al (2009) Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. Magn Reson Med. doi: 10.1002/mrm.22147

34. Juntu J, Sijbers J, De Backer S, et al (2010) Machine learning study of several classifiers trained with texture analysis features to differentiate benign from malignant soft-tissue tumors in T1-MRI images. J Magn Reson Imaging 31:680–689 . doi: 10.1002/jmri.22095

35. Romeo V, Maurea S, Cuocolo R, et al (2018) Characterization of Adrenal Lesions on Unenhanced MRI Using Texture Analysis: A Machine-Learning Approach. J Magn Reson Imaging 48:198–204 . doi: 10.1002/jmri.25954

36. Stanzione A, Cuocolo R, Cocozza S, et al (2019) Detection of Extraprostatic Extension of Cancer on Biparametric MRI Combining Texture Analysis and Machine Learning: Preliminary Results. Acad Radiol 1–7 . doi: 10.1016/j.acra.2018.12.025

37. Sanei Taheri M, Kimia F, Mehrnahad M, et al (2019) Accuracy of diffusion-weighted imaging-magnetic resonance in differentiating functional from non-functional pituitary macro-adenoma and classification of tumor consistency. Neuroradiol J 32:74–85 . doi: 10.1177/1971400918809825

38. Yushkevich PA, Piven J, Hazlett HC, et al (2006) User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. Neuroimage 31:1116–1128 . doi: 10.1016/j.neuroimage.2006.01.015

39. van Griethuysen JJM, Fedorov A, Parmar C, et al (2017) Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res 77:e104–e107 . doi: 10.1158/0008-5472.CAN-17-0339

40. Leijenaar RTH, Nalbantov G, Carvalho S, et al (2015) The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. Sci Rep 5:11075 . doi: 10.1038/srep11075

41. Mannarino T, Assante R, Ricciardi C, et al (2019) Head-to-head comparison of diagnostic accuracy of stress-

only myocardial perfusion imaging with conventional and cadmium-zinc telluride single-photon emission computed tomography in women with suspected coronary artery disease. J Nucl Cardiol. doi: 10.1007/s12350-019-01789-7

42. Romeo V, Ricciardi C, Cuocolo R, et al (2019) Machine learning analysis of MRI-derived texture features to predict placenta accreta spectrum in patients with placenta previa. Magn Reson Imaging 64:71–76 . doi: 10.1016/j.mri.2019.05.017

43. Breiman L (2001) Random forests. Mach Learn. doi: 10.1023/A:1010933404324

44. Cessie S Le, Houwelingen JC Van (1992) Ridge Estimators in Logistic Regression. Appl Stat 41:191 . doi: 10.2307/2347628

45. Aha DW, Kibler D, Albert MK (1991) Instance-based learning algorithms. Mach Learn. doi: 10.1023/A:1022689900470

46. Keller JM, Gray MR, Givens JA (1985) A fuzzy K-nearest neighbor algorithm. IEEE Trans Syst Man Cybern SMC-15:580–585 . doi: 10.1109/TSMC.1985.6313426

47. Chawla N V., Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res. doi: 10.1613/jair.953

48. Freund Y, Schapire RE (1997) A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. J Comput Syst Sci 55:119–139 . doi: 10.1006/jcss.1997.1504

49. Daly AF, Rixhon M, Adam C, et al (2006) High Prevalence of Pituitary Adenomas: A Cross-Sectional Study in the Province of Liège, Belgium. J Clin Endocrinol Metab 91:4769–4775 . doi: 10.1210/jc.2006-1668

50. Trouillas J, Roy P, Sturm N, et al (2013) A new prognostic clinicopathological classification of pituitary adenomas: A multicentric case-control study of 410 patients with 8 years post-operative follow-up. Acta Neuropathol 126:123–135 . doi: 10.1007/s00401-013-1084-y

51. Paek K-I, Kim S-H, Song S-H, et al (2005) Clinical Significance of Ki-67 Labeling Index in Pituitary Macroadenoma. J Korean Med Sci 20:489 . doi: 10.3346/jkms.2005.20.3.489

52. Chatzellis E, Alexandraki KI, Androulakis II, Kaltsas G (2015) Aggressive Pituitary Tumors. Neuroendocrinology 101:87–104 . doi: 10.1159/000371806

53. Senders JT, Zaki MM, Karhade A V., et al (2018) An introduction and overview of machine learning in neurosurgical care. Acta Neurochir (Wien) 160:29–38 . doi: 10.1007/s00701-017-3385-8

54. Kadir T, Gleeson F (2018) Lung cancer prediction using machine learning and advanced imaging techniques. Transl Lung Cancer Res 7:304–312 . doi: 10.21037/tlcr.2018.05.15

55. Fan M, Cheng H, Zhang P, et al (2018) DCE-MRI texture analysis with tumor subregion partitioning for

predicting Ki-67 status of estrogen receptor-positive breast cancers. J Magn Reson Imaging 48:237–247 . doi: 10.1002/jmri.25921

56. Imbriaco M, Cuocolo R (2018) Does Texture Analysis of MR Images of Breast Tumors Help Predict Response to Treatment? Radiology 286:421–423 . doi: 10.1148/radiol.2017172454

57. Liu S, Zheng H, Zhang Y, et al (2018) Whole-volume apparent diffusion coefficient-based entropy parameters for assessment of gastric cancer aggressiveness. J Magn Reson Imaging 47:168–175 . doi: 10.1002/jmri.25752

58. Cuocolo R, Stanzione A, Ponsiglione A, et al (2019) Clinically significant prostate cancer detection on MRI: A radiomic shape features study. Eur J Radiol 116:144–149 . doi: 10.1016/j.ejrad.2019.05.006

59. Cappabianca P, Cavallo LM, Solari D, et al (2014) Endoscopic Endonasal Surgery for Pituitary Adenomas. World Neurosurg 82:S3–S11 . doi: 10.1016/j.wneu.2014.07.019

60. de Notaris M, Solari D, Cavallo LM, et al (2012) The "suprasellar notch," or the tuberculum sellae as seen from below: definition, features, and clinical implications from an endoscopic endonasal perspective. J Neurosurg 116:622–629 . doi: 10.3171/2011.11.JNS111162

61. Kassam A, Snyderman CH, Mintz A, et al (2005) Expanded endonasal approach: the rostrocaudal axis. Part I. Crista galli to the sella turcica. Neurosurg Focus. doi: 190103 [pii]

62. Cappabianca P, Cavallo LM, Esposito F, et al (2008) Extended endoscopic endonasal approach to the midline skull base: the evolving role of transsphenoidal surgery. In: Pickard JD, Akalan N, Di Rocco C, et al (eds) Advances and Technical Standards in Neurosurgery. Springer Vienna, Vienna, pp 151–199

63. Haralick RM, Shanmugam K, Dinstein I (1973) Textural Features for Image Classification. IEEE Trans Syst Man Cybern SMC-3:610–621 . doi: 10.1109/TSMC.1973.4309314

64. THIBAULT G, FERTIL B, NAVARRO C, et al (2013) Shape and texture indexes application to cell nuclei classification. Int J Pattern Recognit Artif Intell 27:1357002 . doi: 10.1142/S0218001413570024

65. Galloway MM (1975) Texture analysis using gray level run lengths. Comput Graph Image Process 4:172–179 . doi: 10.1016/S0146-664X(75)80008-6

66. Amadasun M, King R (1989) Textural features corresponding to textural properties. IEEE Trans Syst Man Cybern 19:1264–1274 . doi: 10.1109/21.44046

67. Sun C, Wee WG (1983) Neighboring gray level dependence matrix for texture classification. Comput Vision, Graph Image Process 23:341–352 . doi: 10.1016/0734-189X(83)90032-4

68. Cai J, Luo J, Wang S, Yang S (2018) Feature selection in machine learning: A new perspective. Neurocomputing 300:70–79 . doi: 10.1016/j.neucom.2017.11.077

69. Eibe F, Hall MA, Witten IH (2016) The WEKA Workbench. Online Appendix for "Data Mining: Practical

Machine Learning Tools and Techniques," Fourth Edi. Morgan Kaufmann

70. Siva Kumar N (1996) An affinity method for the purification of mannose 6-phosphate receptor proteins (MPR 215) from rat tissues and goat liver. J Biochem Biophys Methods 31:181–184 . doi: 10.1016/0165-022X(95)00026-N

71. Thapar K, Kovacs K, Scheithauer BW, et al (1996) Proliferative Activity and Invasiveness among Pituitary Adenomas and Carcinomas: An Analysis Using the MIB-1 Antibody. Neurosurgery 38:99–107 . doi: 10.1097/00006123-199601000-00024

72. Brennan P, Silman A (1992) Statistical methods for assessing observer variability in clinical measures. BMJ 304:1491–1494 . doi: 10.1136/bmj.304.6840.1491

73. Del Basso De Caro M, Solari D, Pagliuca F, et al (2017) Atypical pituitary adenomas: clinical characteristics and role of ki-67 and p53 in prognostic and therapeutic evaluation. A series of 50 patients. Neurosurg Rev 40:105–114 . doi: 10.1007/s10143-016-0740-9

74. Vasiljevic A, Jouanneau E, Trouillas J, Raverot G (2016) Clinicopathological prognostic and theranostic markers in pituitary tumors. Minerva Endocrinol 41:377–89

75. Mete O, Lopes MB (2017) Overview of the 2017 WHO Classification of Pituitary Tumors. Endocr Pathol 28:228–243 . doi: 10.1007/s12022-017-9498-z

76. Inoshita N, Nishioka H (2018) The 2017 WHO classification of pituitary adenoma: overview and comments. Brain Tumor Pathol 35:51–56 . doi: 10.1007/s10014-018-0314-3

77. Hollon TC, Parikh A, Pandian B, et al (2018) A machine learning approach to predict early outcomes after pituitary adenoma surgery. Neurosurg Focus 45:1–9 . doi: 10.3171/2018.8.FOCUS18268.

78. Staartjes VE, Serra C, Muscas G, et al (2018) Utility of deep neural networks in predicting gross-total resection after transsphenoidal surgery for pituitary adenoma: a pilot study. 45:1–7 . doi: 10.3171/2018.8.FOCUS18243.

79. Zhang S, Song G, Zang Y, et al (2018) Non-invasive radiomics approach potentially predicts non-functioning pituitary adenomas subtypes before surgery. Eur Radiol 28:3692–3701 . doi: 10.1007/s00330-017-5180-6

80. Niu J, Tian J, Zhou W, et al (2018) Preoperative prediction of cavernous sinus invasion by pituitary adenomas using a radiomics method based on magnetic resonance images. Eur Radiol 29:1625–1634 . doi: 10.1007/s00330-018-5725-3

81. Tamrazi B, Pekmezci M, Aboian M, et al (2017) Apparent diffusion coefficient and pituitary macroadenomas: pre-operative assessment of tumor atypia. Pituitary 20:195–200 . doi: 10.1007/s11102-016-0759-5

82. Hua J, Xiong Z, Lowey J, et al (2005) Optimal number of features as a function of sample size for various classification rules. Bioinformatics 21:1509–1515 . doi: 10.1093/bioinformatics/bti171

83. Zaharchuk G, Gong E, Wintermark M, et al (2018) Deep Learning in Neuroradiology. Am J Neuroradiol 39:1776–1784 . doi: 10.3174/ajnr.A5543

84. Wei L, Lin S-A, Fan K, et al (2015) Relationship between pituitary adenoma texture and collagen content revealed by comparative study of MRI and pathology analysis. Int J Clin Exp Med 8:12898–905

85. Zeynalova A, Kocak B, Durmaz ES, et al (2019) Preoperative evaluation of tumour consistency in pituitary macroadenomas: a machine learning-based histogram analysis on conventional T2-weighted MRI. Neuroradiology 61:767–774 . doi: 10.1007/s00234-019-02211-2

86. Heck A, Emblem KE, Casar-Borota O, et al (2016) Quantitative analyses of T2-weighted MRI as a potential marker for response to somatostatin analogs in newly diagnosed acromegaly. Endocrine 52:333–343 . doi: 10.1007/s12020-015-0766-8

87. Kocak B, Durmaz ES, Kadioglu P, et al (2019) Predicting response to somatostatin analogues in acromegaly: machine learning-based high-dimensional quantitative texture analysis on T2-weighted MRI. Eur Radiol 29:2731–2739 . doi: 10.1007/s00330-018-5876-2

88. Alimohamadi M, Sanjari R, Mortazavi A, et al (2014) Predictive value of diffusion-weighted MRI for tumor consistency and resection rate of nonfunctional pituitary macroadenomas. Acta Neurochir (Wien) 156:2245–2252 . doi: 10.1007/s00701-014-2259-6

89. Romano A, Coppola V, Lombardi M, et al (2017) Predictive role of dynamic contrast enhanced T1-weighted MR sequences in pre-surgical evaluation of macroadenomas consistency. Pituitary 20:201–209 . doi: 10.1007/s11102-016-0760-z

90. Cuocolo R, Cipullo MB, Stanzione A, et al (2019) Machine learning applications in prostate cancer magnetic resonance imaging. Eur Radiol Exp 3:35 . doi: 10.1186/s41747-019-0109-2

91. Mokrane F-Z, Lu L, Vavasseur A, et al (2020) Radiomics machine-learning signature for diagnosis of hepatocellular carcinoma in cirrhotic patients with indeterminate liver nodules. Eur Radiol 30:558–570 . doi: 10.1007/s00330-019-06347-w

92. Lotan E, Jain R, Razavian N, et al (2019) State of the Art: Machine Learning Applications in Glioma Imaging. Am J Roentgenol 212:26–37 . doi: 10.2214/AJR.18.20218

93. Solari D, Zenga F, Angileri FF, et al (2019) A Survey on Pituitary Surgery in Italy. World Neurosurg. doi: 10.1016/j.wneu.2018.11.186

94. Cappabianca P, Cavallo LM, de Divitiis O, et al (2008) Endoscopic pituitary surgery. Pituitary 11:385–390 . doi: 10.1007/s11102-008-0087-5

95. Solari D, Pivonello R, Caggiano C, et al (2019) Pituitary Adenomas: What Are the Key Features? What Are the

Current Treatments? Where Is the Future Taking Us? World Neurosurg 127:695–709 . doi: 10.1016/j.wneu.2019.03.049

96. Di Maio S, Cavallo LM, Esposito F, et al (2011) Extended endoscopic endonasal approach for selected pituitary adenomas: early experience. J Neurosurg 114:345–353 . doi: 10.3171/2010.9.JNS10262

97. Snow RB, Johnson CE, Morgello S, et al (1990) Is Magnetic Resonance Imaging Useful in Guiding the Operative Approach to Large Pituitary Tumors? Neurosurgery 26:801–803 . doi: 10.1227/00006123-199005000-00011

98. Iuchi T, Saeki N, Tanaka M, et al (1998) MRI Prediction of Fibrous Pituitary Adenomas. Acta Neurochir (Wien) 140:779–786 . doi: 10.1007/s007010050179

99. NAGANUMA H, SATOH E, NUKUI H (2002) Technical Considerations of Transsphenoidal Removal of Fibrous Pituitary Adenomas and Evaluation of Collagen Content and Subtype in the Adenomas. Neurol Med Chir (Tokyo) 42:202–213 . doi: 10.2176/nmc.42.202

100. Smith K, Leever J, Chamoun R (2015) Prediction of Consistency of Pituitary Adenomas by Magnetic Resonance Imaging. J Neurol Surg Part B Skull Base 76:340–343 . doi: 10.1055/s-0035-1549005

101. Suzuki C, Maeda M, Hori K, et al (2007) Apparent diffusion coefficient of pituitary macroadenoma evaluated with line-scan diffusion-weighted imaging. J Neuroradiol 34:228–235 . doi: 10.1016/j.neurad.2007.06.007

102. CHAKRABORTTY S, OI S, YAMAGUCHI M, et al (1993) Growth Hormone producing Pituitary Adenomas: MR Characteristics and Pre- and Postoperative Evaluation. Neurol Med Chir (Tokyo) 33:81–85 . doi: 10.2176/nmc.33.81

103. Bahuleyan B, Raghuram L, Rajshekhar V, Chacko AG (2006) To assess the ability of MRI to predict consistency of pituitary macroadenomas. Br J Neurosurg 20:324–326 . doi: 10.1080/02688690601000717

104. Thotakura AK, Patibandla MR, Panigrahi MK, Mahadevan A (2017) Is it really possible to predict the consistency of a pituitary adenoma preoperatively? Neurochirurgie 63:453–457 . doi: 10.1016/j.neuchi.2017.06.003

105. Hagiwara A, Inoue Y, Wakasa K, et al (2003) Comparison of Growth Hormone–producing and Non–Growth Hormone–producing Pituitary Adenomas: Imaging Characteristics and Pathologic Correlation. Radiology 228:533–538 . doi: 10.1148/radiol.2282020695

106. Pierallini A, Caramia F, Falcone C, et al (2006) Pituitary Macroadenomas: Preoperative Evaluation of Consistency with Diffusion-weighted MR Imaging—Initial Experience. Radiology 239:223–231 . doi: 10.1148/radiol.2383042204

107. Boxerman JL, Rogg JM, Donahue JE, et al (2010) Preoperative MRI Evaluation of Pituitary Macroadenoma:

Imaging Features Predictive of Successful Transsphenoidal Surgery. Am J Roentgenol 195:720–728 . doi: 10.2214/AJR.09.4128

108. Yamamoto J, Kakeda S, Shimajiri S, et al (2014) Tumor Consistency of Pituitary Macroadenomas: Predictive Analysis on the Basis of Imaging Features with Contrast-Enhanced 3D FIESTA at 3T. Am J Neuroradiol 35:297–303 . doi: 10.3174/ajnr.A3667

109. Ma Z, He W, Zhao Y, et al (2016) Predictive value of PWI for blood supply and T1-spin echo MRI for consistency of pituitary adenoma. Neuroradiology 58:51–57 . doi: 10.1007/s00234-015-1591-8

110. Sakai N, Takehara Y, Yamashita S, et al (2016) Shear stiffness of 4 common intracranial tumors measured using mr elastography: Comparison with intraoperative consistency grading. Am J Neuroradiol. doi: 10.3174/ajnr.A4832

111. Hughes JD, Fattahi N, Van Gompel J, et al (2016) Magnetic resonance elastography detects tumoral consistency in pituitary macroadenomas. Pituitary. doi: 10.1007/s11102-016-0706-5

112. Rui W, Wu Y, Ma Z, et al (2019) MR textural analysis on contrast enhanced 3D-SPACE images in assessment of consistency of pituitary macroadenoma. Eur J Radiol 110:219–224 . doi: 10.1016/j.ejrad.2018.12.002

113. Fan Y, Hua M, Mou A, et al (2019) Preoperative Noninvasive Radiomics Approach Predicts Tumor Consistency in Patients With Acromegaly: Development and Multicenter Prospective Validation. Front Endocrinol (Lausanne) 10: . doi: 10.3389/fendo.2019.00403

114. Antonelli M, Johnston EW, Dikaios N, et al (2019) Machine learning classifiers can predict Gleason pattern 4 prostate cancer with greater accuracy than experienced radiologists. Eur Radiol 29:4754–4764 . doi: 10.1007/s00330-019-06244-2

115. Fehr D, Veeraraghavan H, Wibmer A, et al (2015) Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. Proc Natl Acad Sci 112:E6265–E6273 . doi: 10.1073/pnas.1505935112

116. Zhang C, Bi J, Soda P (2017) Feature selection and resampling in class imbalance learning: Which comes first? An empirical study in the biological domain. In: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, pp 933–938

117. McDowell BD, Wallace RB, Carnahan RM, et al (2011) Demographic differences in incidence for pituitary adenoma. Pituitary 14:23–30 . doi: 10.1007/s11102-010-0253-4

118. Chen L, White WL, Spetzler RF, Xu B (2011) A prospective study of nonfunctioning pituitary adenomas: presentation, management, and clinical outcome. J Neurooncol 102:129–138 . doi: 10.1007/s11060-010-0302-x

119. Asa SL, Ezzat S (2009) The Pathogenesis of Pituitary Tumors. Annu Rev Pathol Mech Dis 4:97–126 . doi:

10.1146/annurev.pathol.4.110807.092259

120. Gupta K, Sahni S, Saggar K, Vashisht G (2018) Evaluation of clinical and magnetic resonance imaging profile of pituitary macroadenoma: A prospective study. J Nat Sci Biol Med 9:34 . doi: 10.4103/jnsbm.JNSBM_111_17

121. Knosp E, Steiner E, Kitz K, Matula C (1993) Pituitary Adenomas with Invasion of the Cavernous Sinus Space. Neurosurgery 33:610–618 . doi: 10.1227/00006123-199310000-00008

122. Micko ASG, Wöhrer A, Wolfsberger S, Knosp E (2015) Invasion of the cavernous sinus space in pituitary adenomas: endoscopic verification and its correlation with an MRI-based classification. J Neurosurg 122:803–811 . doi: 10.3171/2014.12.JNS141083

123. Mohamed FF, Abouhashem S (2013) Diagnostic value of apparent diffusion coefficient (ADC) in assessment of pituitary macroadenoma consistency. Egypt J Radiol Nucl Med 44:617–624 . doi: 10.1016/j.ejrnm.2013.05.012

124. Erickson BJ, Korfiatis P, Akkus Z, Kline TL (2017) Machine Learning for Medical Imaging. RadioGraphics 37:505–515 . doi: 10.1148/rg.2017160130

125. Wang S, Summers RM (2012) Machine learning and radiology. Med Image Anal 16:933–951 . doi: 10.1016/j.media.2012.02.005

126. Schmidt J, Marques MRG, Botti S, Marques MAL (2019) Recent advances and applications of machine learning in solid-state materials science. npj Comput Mater 5:83 . doi: 10.1038/s41524-019-0221-0

127. Mazurowski MA, Buda M, Saha A, Bashir MR (2019) Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. J Magn Reson Imaging 49:939–954 . doi: 10.1002/jmri.26534

128. McBee MP, Awan OA, Colucci AT, et al (2018) Deep Learning in Radiology. Acad Radiol 25:1472–1480 . doi: 10.1016/j.acra.2018.02.018

129. Ciompi F, Chung K, van Riel SJ, et al (2017) Towards automatic pulmonary nodule management in lung cancer screening with deep learning. Sci Rep 7:46479 . doi: 10.1038/srep46479

130. Cuocolo R, Perillo T, De Rosa E, et al (2019) Current applications of big data and machine learning in cardiology. J Geriatr Cardiol 16:601–607 . doi: 10.11909/j.issn.1671-5411.2019.08.002

131. Lilja Y, Gustafsson O, Ljungberg M, et al (2017) Visual pathway impairment by pituitary adenomas: quantitative diagnostics by diffusion tensor imaging. J Neurosurg 127:569–579 . doi: 10.3171/2016.8.JNS161290

132. Liu Y, Liu X, Hong X, et al (2019) Prediction of Recurrence after Transsphenoidal Surgery for Cushing's Disease: The Use of Machine Learning Algorithms. Neuroendocrinology 108:201–210 . doi:

10.1159/000496753

133. Staartjes VE, Zattra CM, Akeret K, et al (2020) Neural network–based identification of patients at high risk for intraoperative cerebrospinal fluid leaks in endoscopic pituitary surgery. J Neurosurg 133:329–335 . doi: 10.3171/2019.4.JNS19477