



Prediction of high proliferative index in pituitary macroadenomas using MRI-based radiomics and machine learning

Lorenzo Ugga¹ · Renato Cuocolo¹ · Domenico Solari² · Elia Guadagno³ · Alessandra D'Amico¹ · Teresa Somma² · Paolo Cappabianca² · Maria Laura del Basso de Caro³ · Luigi Maria Cavallo² · Arturo Brunetti¹

Received: 14 June 2019 / Accepted: 15 July 2019 / Published online: 2 August 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Purpose Pituitary adenomas are among the most frequent intracranial tumors. They may exhibit clinically aggressive behavior, with recurrent disease and resistance to multimodal therapy. The ki-67 labeling index represents a proliferative marker which correlates with pituitary adenoma aggressiveness. Aim of our study was 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.

Methods A total of 89 patients who underwent an endoscopic endonasal procedure for pituitary adenoma removal with available ki-67 labeling index were included. From T2w MR images, 1128 quantitative imaging features were extracted. To select the most informative features, different supervised feature selection methods were employed. Subsequently, a *k*-nearest neighbors (*k*-NN) classifier was employed to predict macroadenoma high or low proliferation index. Algorithm validation was performed with a train-test approach.

Results Of the 12 subsets derived from feature selection, the best performing one was constituted by the 4 highest correlating parameters at Pearson's test. These all showed very good (ICC ≥ 0.85) inter-observer reproducibility. The overall accuracy of the *k*-NN in the test group was of 91.67% (33/36) of correctly classified patients.

Conclusions 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.

Keywords Machine learning · Magnetic resonance imaging · Pituitary adenoma

Introduction

Pituitary adenomas are among the most frequent intracranial tumors, with a 1/1500 prevalence [1]. 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 and 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 neuro-oncological 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

LU and RC contributed equally to this work.

✉ Renato Cuocolo
renato.cuocolo@unina.it

¹ Department of Advanced Biomedical Sciences, University of Naples “Federico II”, via Sergio Pansini 5, 80131 Naples, Italy

² Department of Neurosciences, Reproductive and Odontostomatological Sciences, Division of Neurosurgery, University of Naples “Federico II”, Naples, Italy

³ Department of Advanced Biomedical Sciences, Pathology Section, University of Naples “Federico II”, Naples, Italy

to tumors with rapid growth, radiological invasion, and high Ki-67 proliferation index [2–4].

Considering the above, early identification of any radiological feature defining such behavior can be crucial in order to allow timely diagnosis and treatment. In this regard, radiomics, consisting of conversion of images into mineable data and subsequent analysis for decision support, has been emerging [5]. In particular, texture analysis is a post-processing technique for quantitative parameter extraction from pixel gray level heterogeneity. More recently, texture analysis-derived features have been used in association with data mining machine learning algorithms, aiding in the interpretation of a 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 [6]. 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 [7–14].

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.

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.

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 as follows: 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).

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 intrasuprasellar post-surgical tumor cavities. They were run according to techniques already described in previous publications [15–19].

Pathological data

Specimens were obtained as formalin-fixed tissue. For the evaluation of the proliferation index ki-67 (labeling index (LI)), “hotspot” areas were chosen at low magnification and an average of the values on 5 adjacent fields (at least 500 neoplastic cells) was calculated (Fig. 1).

MRI acquisition data

Imaging studies were performed either on a 1.5-Tesla scanner (Gyrosan 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, 180 × 180 mm; matrix, 288 × 288; thk, 3 mm, at 1.5 T; TR/TE, 3000/98 ms FOV, 200 × 200 mm; matrix, 384 × 384; thk, 3 mm, at 3 T).

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 freely available segmentation software (ITKSnap v3.6.0) [20] (Fig. 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 in an open-source Python radiomics software (Pyradiomics v2.1.2) [21]. 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 2 × 2 × 2 mm resolution. To optimize extraction time, making it less memory intensive, images were automatically cropped based on ROI masks keeping 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

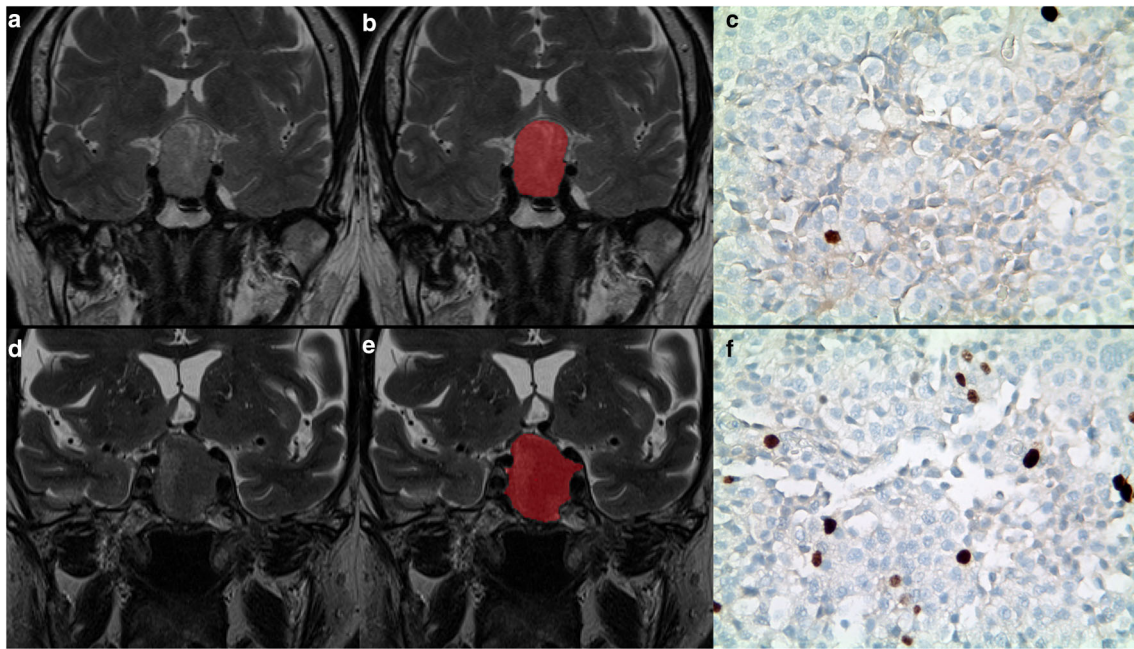


Fig. 1 Segmentation examples on coronal T2-weighted images in two patients with low (upper row, **a**) and high (lower row, **d**) proliferation index pituitary adenomas, showing hand-drawn ROI placement (**b** and **e**,

respectively). Pictures **c** and **f** show corresponding immunohistochemical evaluation ($\times 40$ magnification) of ki-67 cell labeling index, respectively, 1 and 6%

developers, we decided to use a fixed bin width that allowed us to obtain an ideal bin count between 16 and 128 [22]. 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 $\times 100$) was applied.

Various filters were also used to generate derived images for additional texture feature extraction. In particular, wavelet decomposition yields 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 changes 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 [23]. Gray level size zone matrix (GLSZM) features quantify gray level zones, defined as the number of connected voxels that share the same intensity value [24]. The gray level run length matrix (GLRLM) evaluates gray level runs, which are the length of consecutive pixels with the same gray level [25]. Neighboring gray-tone difference matrix (NGTDM) features

assess differences between pixel values and neighbor average gray value within a set distance [26]. 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 [27].

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 [28]. This process and subsequent steps in our machine learning pipeline were performed on a freely available data mining software (Weka v3.8) [29]. To balance data dimensionality reduction and loss of information, we tested eight-feature selection method belonging to three different families as follows:

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 [30];
2. Feature ranking methods using a direct Pearson's correlation and a ReliefF Evaluator, a filter-method approach sensitive to feature interactions [31];
3. Subset feature search selection method, which evaluates attribute subset worth considering both feature predictive ability and redundancy [31].

Absolute agreement intraclass 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 [32]. 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 Fig. 2.

Results

Subjects and pathological data

According to the 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 into 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 1).

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), while in 30 cases (13 functioning and 17 non-functioning lesions), it resulted equal to or greater than 3% (high proliferative index) [33].

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.

Image analysis

A total of 1128 features were extracted from each patient, represented on the correlation cluster map shown in Fig. 3.

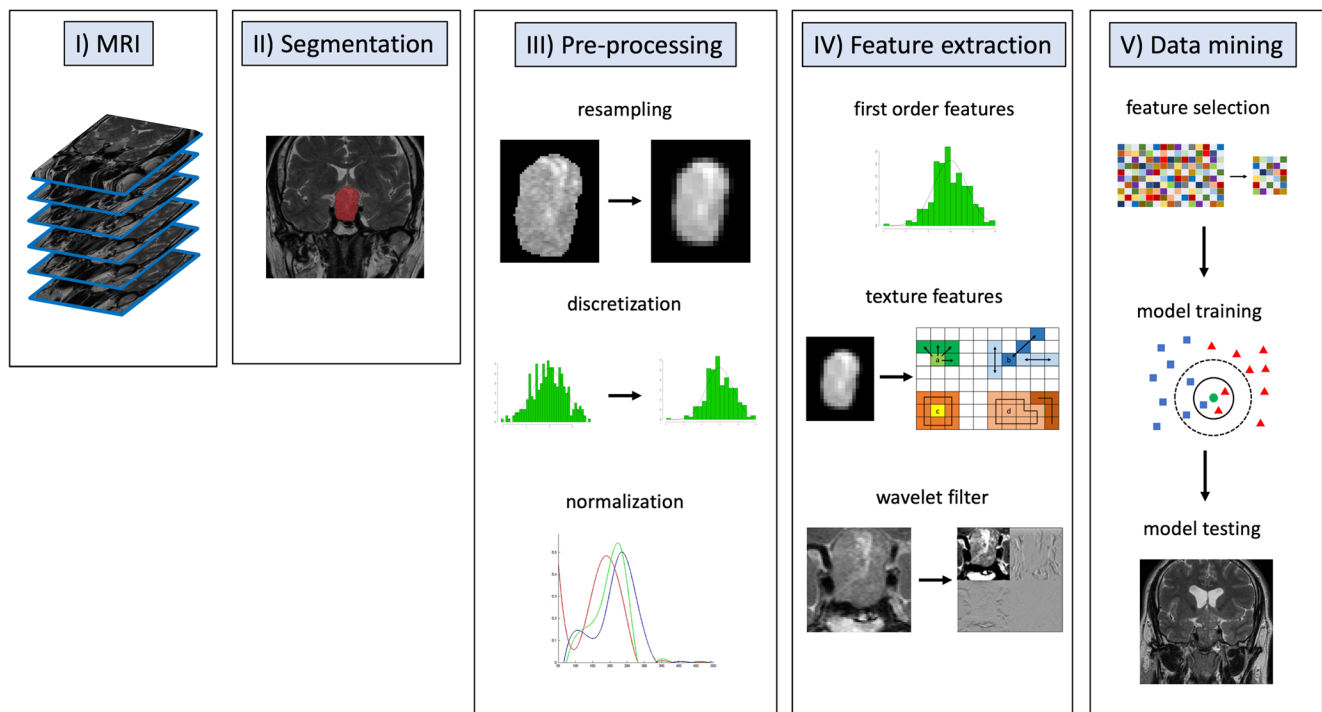


Fig. 2 Radiomic workflow pipeline. Manual segmentation of the lesions was performed on coronal T2-weighted images using a bidimensional polygonal ROI. Image pre-processing included resampling of all images and corresponding ROI masks to a $2 \times 2 \times 2$ mm resolution, discretization, and normalization 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 the macroadenoma proliferation index class with a train-test approach

Table 1 Patient population clinical data

Characteristic	Low ki-67 LI	High ki-67 LI
No. of patients	59	30
Age (mean) (year)	54.8	47.6
Sex		
M (no.) (%)	35 (59.3%)	16 (53.3%)
F (no.) (%)	24 (40.6%)	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, labeling index; F, functioning macroadenomas; NF, non-functioning macroadenomas

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 the following: 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 [34]. Figures 4 and 5 present univariate and pairwise feature distribution for this data subset in our population.

The k -NN overall accuracy was 91.67% (33/36) of correctly classified patients. Other evaluation metrics are reported in

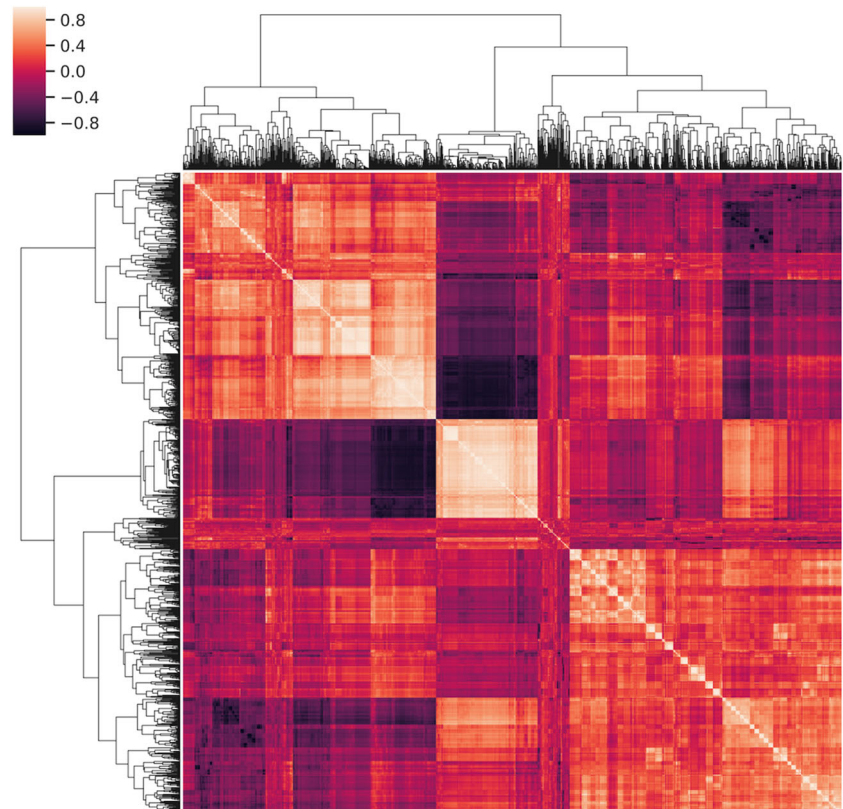
Fig. 3 Feature correlation matrix represented as a hierarchically clustered heatmap

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 as follows:

- 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 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.

Discussion

In the previous edition of the WHO Classification of Tumors of the Pituitary Gland (2004), pituitary neuroendocrine tumors were divided into a typical adenoma, atypical adenoma, and carcinoma. Atypical adenomas were identified by an elevated

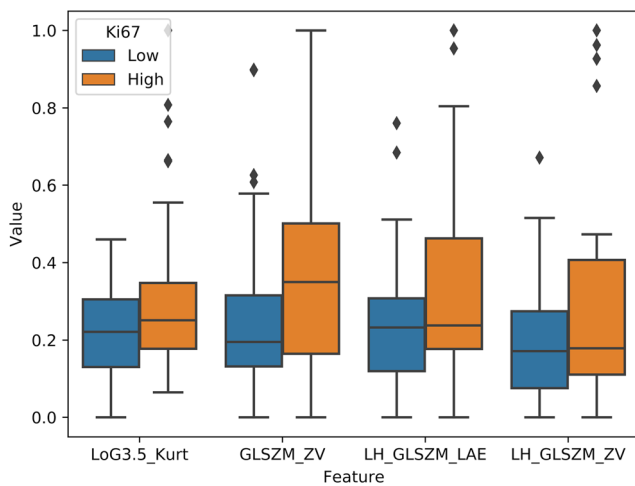


Fig. 4 Box plot of the distribution for each feature of the selected subset in relation to the proliferation index class

mitotic index and ki-67-labeling index $\geq 3\%$, suggestive of aggressive clinical behavior. Using this classification, atypical adenoma incidence was relatively variable and lacking valid and reliable prognostic correlations [35]. Therefore, Trouillas and colleagues proposed a new clinicopathological classification using MRI invasiveness signs (cavernous and/or sphenoid sinus involvement), immunocytochemistry and proliferative markers, and 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) [2]. 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 [36, 37]. However, it does not provide any hint to correlate these elements to define pituitary adenomas subtypes [38].

In this setting, artificial intelligence could contribute to better allocate individual cases in relation to aggressiveness. Machine learning applications have proved promising in recently published papers, providing insights into how predictive modeling can improve patient perioperative management. Hollon and colleagues reported an accuracy of 87% in the prediction of early postoperative outcomes in a retrospective cohort of 400 consecutive pituitary adenoma patients [39]. 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 [40]. Furthermore, using radiomics, MRI proved accurate in predicting non-functioning pituitary adenomas subtypes [41] and cavernous sinus invasion [42].

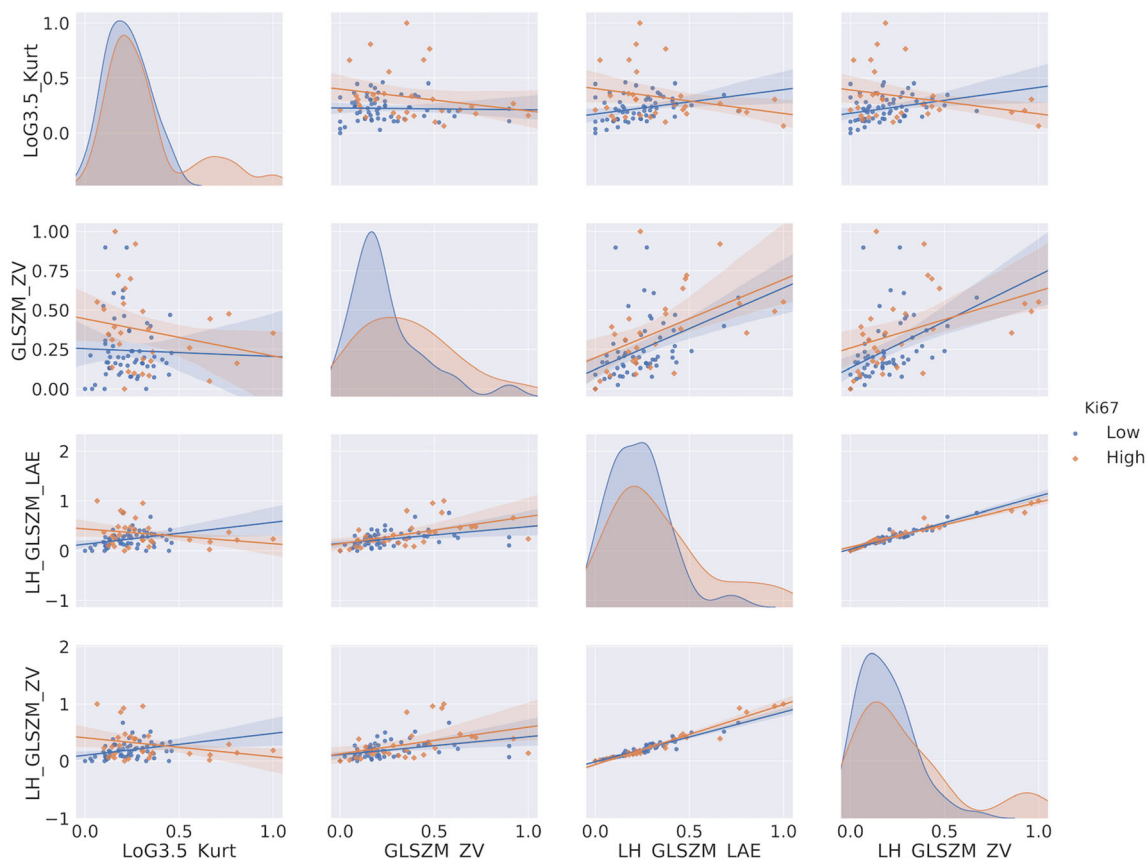


Fig. 5 Pairwise bivariate distribution with regression lines for the selected feature subset in relation to the proliferation index class

Table 2 *k*-NN accuracy metrics weighted average and by class

Class	Sensitivity	Specificity	Precision	MCC	<i>F</i> 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, labeling index; *WA*, weighted average; *MCC*, Matthews correlation coefficient; *ROC*, receiver operator curve; *PRC*, precision-recall curve

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 [43]. 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 the optimal feature number choice. It depends on both sample size and algorithm employed [44]. 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 [45]. 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 a decision process, limiting software assumption correctness assessment and subsequent wide-scale applicability. Furthermore, data required for training such as networks is exponentially larger and computational time is also increased compared with 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, the 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 [46] and on collagen content [47], useful for predicting tumor consistency. Regarding the last, a recent study by Zeynalova [48] 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 [49, 50].

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.

Table 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, labeling index

Conclusion

We analyzed pituitary adenoma proliferative index class pre-operative 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.

Funding No funding was received for this study.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in the studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Requirement for informed consent was waived by the local IRB.

References

- Daly AF, Rixhon M, Adam C, Dempegioti A, Tichomirowa MA, Beckers A (2006) High prevalence of pituitary adenomas: a cross-sectional study in the province of Liège, Belgium. *J Clin Endocrinol Metab* 91:4769–4775. <https://doi.org/10.1210/jc.2006-1668>
- 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. <https://doi.org/10.1007/s00401-013-1084-y>
- Paek K-I, Kim S-H, Song S-H, Choi SW, Koh HS, Youm JY, Kim Y (2005) Clinical significance of Ki-67 labeling index in pituitary macroadenoma. *J Korean Med Sci* 20:489. <https://doi.org/10.3346/jkms.2005.20.3.489>
- Chatzellis E, Alexandraki KI, Androulakis II, Kaltsas G (2015) Aggressive pituitary tumors. *Neuroendocrinology* 101:87–104. <https://doi.org/10.1159/000371806>
- Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. *Radiology* 278:563–577. <https://doi.org/10.1148/radiol.2015151169>
- Senders JT, Zaki MM, Karhade AV, Chang B, Gormley WB, Broekman ML, Smith TR, Amaout O (2018) An introduction and overview of machine learning in neurosurgical care. *Acta Neurochir* 160:29–38. <https://doi.org/10.1007/s00701-017-3385-8>
- Stanzione A, Cuocolo R, Cocozza S, Romeo V, Persico F, Fusco F, Longo N, Brunetti A, Imbriaco M (2019) Detection of extraprostatic extension of cancer on biparametric MRI combining texture analysis and machine learning: preliminary results. *Acad Radiol*:1–7. <https://doi.org/10.1016/j.acra.2018.12.025>
- Romeo V, Maurea S, Cuocolo R, Petretta M, Mainenti PP, Verde F, Coppola M, Dell'Aversana S, Brunetti A (2018) Characterization of adrenal lesions on unenhanced MRI using texture analysis: a machine-learning approach. *J Magn Reson Imaging* 48:198–204. <https://doi.org/10.1002/jmri.25954>
- Kadir T, Gleeson F (2018) Lung cancer prediction using machine learning and advanced imaging techniques. *Transl Lung Cancer Res* 7:304–312. <https://doi.org/10.21037/tlcr.2018.05.15>
- Fan M, Cheng H, Zhang P, Gao X, Zhang J, Shao G, Li L (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. <https://doi.org/10.1002/jmri.25921>
- Imbriaco M, Cuocolo R (2018) Does texture analysis of MR images of breast tumors help predict response to treatment? *Radiology* 286:421–423. <https://doi.org/10.1148/radiol.2017172454>
- Liu S, Zheng H, Zhang Y, Chen L, Guan W, Guan Y, Ge Y, He J, Zhou Z (2018) Whole-volume apparent diffusion coefficient-based entropy parameters for assessment of gastric cancer aggressiveness. *J Magn Reson Imaging* 47:168–175. <https://doi.org/10.1002/jmri.25752>
- Cuocolo R, Stanzione A, Ponsiglione A, Romeo V, Verde F, Creta M, la Rocca R, Longo N, Pace L, Imbriaco M (2019) Clinically significant prostate cancer detection on MRI: a radiomic shape features study. *Eur J Radiol* 116:144–149. <https://doi.org/10.1016/j.ejrad.2019.05.006>
- Romeo V, Ricciardi C, Cuocolo R, Stanzione A, Verde F, Sarno L, Improta G, Mainenti PP, D'Armiento M, Brunetti A, Maurea S (2019) Machine learning analysis of MRI-derived texture features to predict placenta accreta spectrum in patients with placenta previa. *Magn Reson Imaging*. <https://doi.org/10.1016/j.mri.2019.05.017>
- Cappabianca P, Cavallo LM, Solari D, Stagno V, Esposito F, de Angelis M (2014) Endoscopic endonasal surgery for pituitary adenomas. *World Neurosurg* 82:S3–S11. <https://doi.org/10.1016/j.wneu.2014.07.019>
- de Notaris M, Solari D, Cavallo LM, D'Enza AI, Enseñat J, Berenguer J, Ferrer E, Prats-Galino A, Cappabianca P (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. <https://doi.org/10.3171/2011.11.JNS111162>
- 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*
- Solari D, Cavallo LM, Cappabianca P (2014) Surgical approach to pituitary tumors. *Handb Clin Neurol* 124:291–301. <https://doi.org/10.1016/B978-0-444-59602-4.00019-8>
- Cappabianca P, Cavallo LM, Esposito F, de Divitiis O, Messina A, de Divitiis E (2008) Extended endoscopic endonasal approach to the midline skull base: the evolving role of transphenoidal surgery. In: Pickard JD, Akalan N, Di Rocco C et al (eds) *Advances and technical standards in neurosurgery*. Springer Vienna, Vienna, pp 151–199
- Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G (2006) User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 31:1116–1128. <https://doi.org/10.1016/j.neuroimage.2006.01.015>
- van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin JC, Pieper S, Aerts HJWL (2017) Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 77:e104–e107. <https://doi.org/10.1158/0008-5472.CAN-17-0339>
- Leijenaar RTH, Nalbantov G, Carvalho S, van Elmpt WJC, Troost EGC, Boellaard R, Aerts HJWL, Gillies RJ, Lambin P (2015) The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis. *Sci Rep* 5:11075. <https://doi.org/10.1038/srep11075>
- Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. *IEEE Trans Syst Man Cybern SMC-3*: 610–621. <https://doi.org/10.1109/TSMC.1973.4309314>

24. THIBAUT G, FERTIL B, NAVARRO C, PEREIRA S, CAU P, LEVY N, SEQUEIRA J, MARI JL (2013) Shape and texture indexes application to cell nuclei classification. *Int J Pattern Recognit Artif Intell* 27:1357002. <https://doi.org/10.1142/S0218001413570024>
25. Galloway MM (1975) Texture analysis using gray level run lengths. *Comput Graph Image Process* 4:172–179. [https://doi.org/10.1016/S0146-664X\(75\)80008-6](https://doi.org/10.1016/S0146-664X(75)80008-6)
26. Amadasun M, King R (1989) Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybern* 19:1264–1274. <https://doi.org/10.1109/21.44046>
27. Sun C, Wee WG (1983) Neighboring gray level dependence matrix for texture classification. *Comput Vision, Graph Image Process* 23: 341–352. [https://doi.org/10.1016/0734-189X\(83\)90032-4](https://doi.org/10.1016/0734-189X(83)90032-4)
28. Cai J, Luo J, Wang S, Yang S (2018) Feature selection in machine learning: a new perspective. *Neurocomputing* 300:70–79. <https://doi.org/10.1016/j.neucom.2017.11.077>
29. Eibe F, Hall MA, Witten IH (2016) The WEKA Workbench. Online Appendix for “Data Mining: Practical Machine Learning Tools and Techniques,” Fourth Ed. Morgan Kaufmann
30. 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. <https://doi.org/10.1002/jmri.22095>
31. 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. [https://doi.org/10.1016/0165-022X\(95\)00026-N](https://doi.org/10.1016/0165-022X(95)00026-N)
32. Aha DW, Kibler D, Albert MK (1991) Instance-based learning algorithms. *Mach Learn* 6:37–66. <https://doi.org/10.1023/A:1022689900470>
33. Thapar K, Kovacs K, Scheithauer BW, Stefanescu L, Horvath E, Peter J, P, Murray D, Laws ER Jr (1996) Proliferative activity and invasiveness among pituitary adenomas and carcinomas: an analysis using the MIB-1 antibody. *Neurosurgery* 38:99–107. <https://doi.org/10.1097/00006123-199601000-00024>
34. Brennan P, Silman A (1992) Statistical methods for assessing observer variability in clinical measures. *BMJ* 304:1491–1494. <https://doi.org/10.1136/bmj.304.6840.1491>
35. 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. <https://doi.org/10.1007/s10143-016-0740-9>
36. Vasiljevic A, Jouanneau E, Trouillas J, Raverot G (2016) Clinicopathological prognostic and theranostic markers in pituitary tumors. *Minerva Endocrinol* 41:377–389
37. Mete O, Lopes MB (2017) Overview of the 2017 WHO classification of pituitary tumors. *Endocr Pathol* 28:228–243. <https://doi.org/10.1007/s12022-017-9498-z>
38. Inoshita N, Nishioka H (2018) The 2017 WHO classification of pituitary adenoma: overview and comments. *Brain Tumor Pathol* 35:51–56. <https://doi.org/10.1007/s10014-018-0314-3>
39. Hollon TC, Parikh A, Pandian B, Tarpeh J, Orringer DA, Barkan AL, McKean EL, Sullivan SE (2018) A machine learning approach to predict early outcomes after pituitary adenoma surgery. *Neurosurg Focus* 45:1–9. <https://doi.org/10.3171/2018.8.FOCUS18268>
40. 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: <https://doi.org/10.3171/2018.8.FOCUS18243>
41. Zhang S, Song G, Zang Y, Jia J, Wang C, Li C, Tian J, Dong D, Zhang Y (2018) Non-invasive radiomics approach potentially predicts non-functioning pituitary adenomas subtypes before surgery. *Eur Radiol* 28:3692–3701. <https://doi.org/10.1007/s00330-017-5180-6>
42. 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. <https://doi.org/10.1007/s00330-018-5725-3>
43. Tamrazi B, Pekmezci M, Aboian M, Tihan T, Glastonbury CM (2017) Apparent diffusion coefficient and pituitary macroadenomas: pre-operative assessment of tumor atypia. *Pituitary* 20:195–200. <https://doi.org/10.1007/s11102-016-0759-5>
44. Hua J, Xiong Z, Lowey J, Suh E, Dougherty ER (2005) Optimal number of features as a function of sample size for various classification rules. *Bioinformatics* 21:1509–1515. <https://doi.org/10.1093/bioinformatics/bti171>
45. Zaharchuk G, Gong E, Wintermark M, Rubin D, Langlotz CP (2018) Deep learning in neuroradiology. *Am J Neuroradiol* 39: 1776–1784. <https://doi.org/10.3171/ajnr.A5543>
46. Sanei Taheri M, Kimia F, Mehrnahad M, Saligheh Rad H, Haghghatkhah H, Moradi A, Kazerooni AF, Alviri M, Absalan A (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. <https://doi.org/10.1177/1971400918809825>
47. 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: 12898–12905
48. Zeynalova A, Kocak B, Durmaz ES, Comunoglu N, Ozcan K, Ozcan G, Turk O, Tanriover N, Kocer N, Kizilkilic O, Islak C (2019) Preoperative evaluation of tumour consistency in pituitary macroadenomas: a machine learning-based histogram analysis on conventional T2-weighted MRI. *Neuroradiology* 61:767–774. <https://doi.org/10.1007/s00234-019-02211-2>
49. Heck A, Emblem KE, Casar-Borota O, Bollerslev J, Ringstad G (2016) Quantitative analyses of T2-weighted MRI as a potential marker for response to somatostatin analogs in newly diagnosed acromegaly. *Endocrine* 52:333–343. <https://doi.org/10.1007/s12020-015-0766-8>
50. Kocak B, Durmaz ES, Kadioglu P, Polat Korkmaz O, Comunoglu N, Tanriover N, Kocer N, Islak C, Kizilkilic O (2019) Predicting response to somatostatin analogues in acromegaly: machine learning-based high-dimensional quantitative texture analysis on T2-weighted MRI. *Eur Radiol* 29:2731–2739. <https://doi.org/10.1007/s00330-018-5876-2>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.