

# A Supervised Machine learning-powered tool: intraoperative CSF leak predictor in endoscopic transsphenoidal surgery for pituitary adenomas.

Leonardo Tariciotti<sup>1</sup>, Giorgio Fiore<sup>2</sup>, Giorgio Carrabba<sup>3</sup>, Giulio Bertani<sup>3</sup>, Emanuele Ferrante<sup>4</sup>, Valerio Caccavella<sup>5</sup>, Pier Paolo Mattogno<sup>5</sup>, Giovanna Mantovani<sup>3</sup>, and Marco Locatelli<sup>3</sup>

<sup>1</sup>University of Milan

<sup>2</sup>Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico.

<sup>3</sup>Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico

<sup>4</sup>Fondazione IRCCS Cà Granda Ospedale Maggiore

<sup>5</sup>Fondazione Policlinico Agostino Gemelli IRCCS

November 4, 2020

## Abstract

Background: Despite advances in endoscopic transnasal transsphenoidal surgery (E-TNS) for pituitary adenomas (PAs), cerebrospinal fluid (CSF) leakage remains a life-threatening complication as it predisposes to meningitis and tension pneumocephalus. The purpose of the current study is to develop an accurate supervised ML model able to predict the risk of intraoperative CSF leakage by comparing different machine learning (ML) methods. Methods: A cohort of patients consecutively treated via E-TNS for PAs was selected. Clinical, radiological and endocrinological preoperative data were reviewed and elaborated through a features selecting algorithm. A customized pipeline of several ML models was programmed and trained in parallel; the best five models were included for further analyses. Selected risk factors were then used for training and hyperparameters optimization. Results: Intraoperative CSF leak occurred in 54 (22,6%) of 238 patients. Best risk's predictors were: non secreting status, older age, x-, y- and z-axes diameters, ICD and R ratio. The random forest (RF) classifier outperformed other models, with an AUC of 0,84, high sensitivity (87%) and specificity (82%). Positive predictive value and negative predictive value were 69% and 93% respectively. F1 score was 0,87. Conclusion: A supervised machine learning prediction model able to identify patients at higher risk of intraoperative CSF leakage was trained and internally validated. The random forest classifier showed the best performance across all models selected by the authors. RF models might predict surgical outcomes in heterogeneous multimorbid and fragile populations outperforming classical statistical analyses and other machine learning models.

## Abstract:

**Objectives:** The purpose of the current study is to develop an accurate supervised ML model able to predict the risk of intraoperative CSF leakage by comparing different machine learning methods.

**Design:** Retrospective diagnostic-prognostic analysis.

**Setting :** Endoscopic transsphenoidal surgery for pituitary adenomas (PAs).

**Participants:** A cohort of patients consecutively treated via E-TNS for PAs was selected. Clinical, radiological and endocrinological preoperative data were reviewed and elaborated through a feature selecting algorithm. A customized pipeline of several ML models was programmed and trained in parallel for CSF

leakage prediction; the best five models were included for further analyses. Selected risk factors were then used for training and hyperparameters optimization.

**Main outcome measures :** The main outcome metrics were: accuracy, sensitivity, specificity, PPV, NPV and F1-score of each ML model.

**Results:** Intraoperative CSF leak occurred in 54 (22,6%) of 238 patients. Best risk's predictors were: non secreting status, older age, x-, y- and z-axes diameters, ICD and R ratio. The random forest (RF) classifier outperformed other models, with an AUC of 0,84, high sensitivity (87%) and specificity (82%). Positive predictive value and negative predictive value were 69% and 93% respectively. F1 score was 0,87.

**Conclusion :** A supervised machine learning prediction model able to identify patients at higher risk of intraoperative CSF leakage was trained and internally validated. The random forest classifier showed the best performance across all models selected by the authors. RF models might predict surgical outcomes in heterogeneous multimorbid and fragile populations outperforming classical statistical analyses and other machine learning models.

### Keywords :

Machine learning;

Supervised machine learning;

Pituitary adenoma;

Intraoperative CSF leakage;

CSF fistula;

### Key points:

1. Machine learning application in neurosurgery can provide additional help to physicians and patients, optimizing risk-adjusted presurgical counselling, planning of tailored interventional therapies and integrating quantitative risk assessment in both legal and quality control perspectives.
2. Pituitary surgery has reached highly effective optimization and complications occurrence is rare; however, CSF leakages still predispose to life threatening complications such as meningitis and tension pneumocephalus.
3. A random forest classifier predicted best the occurrence or not of intraoperative CSF leakage, outperforming other machine learning architectures like SVM classifier, artificial neural network, logistic multivariate regression and naïve-Bayes classifier.
4. Patients flagged as at high risk of CSF leakage according to machine learning classifier might receive additional attention in identifying even occult intraoperative leakages and sellar floor repair to prevent post-operative life threatening complications.
5. The accurate prediction of intraoperative CSF poses an advancement to the state of the art of current surgical practice.

### Introduction:

Pituitary Adenomas (PAs) account for 16% of all newly diagnosed primary CNS tumors, the third most common central nervous system tumors after meningiomas and gliomas. PAs are more frequent in women under the age of 50, while in later life their incidence becomes higher in males(1).

Surgical resection is usually the first-line treatment for large or functioning tumors except prolactinomas. The modern surgical approach includes two main TS techniques: microscopic and endoscopic approaches. The literature does not uniformly agree on the superiority of one technique over the other but there is an incontrovertible and consolidated tendency to prefer the endoscopic approach, being considered less invasive and more effective(2).

Despite continuous advances, three major complications are still associated with TS endoscopic surgery and are represented by hypopituitarism, diabetes insipidus, and CSF rhinorrhea(3). In a national study(4), the incidence of the latter was reported between 1.5% and 4.2%. Given the high risk of postoperative meningitis, it has to be considered a life-threatening condition.

The population affected by PAs results made up of multimorbid young adults or frail elderly patients – exposed to higher risk of perioperative complications – who would benefit most from a tailored multimodal treatment and effective preoperative risk assessment.

Since machine learning (ML) has already demonstrated its reliability in improving neurosurgical care, particularly by increasing the efficiency and precision of preoperative planning and surgical outcome prediction(5), the purpose of the current study is to investigate whether a comprehensive supervised ML model trained and internally validated on clinical, radiological and endocrinological preoperative data can predict the risk of intraoperative CSF leakage.

## Materials and Methods

### Participants:

We retrospectively reviewed a cohort of patients consecutively subjected to endonasal endoscopic transsphenoidal surgery (E-TNS) for pituitary adenomas (PA), performed by the same skull base team in the interval between January 2014 and January 2020. All patients were treated by a dedicated staff according to the same recommendations and protocols. PAs treated via transcranial approaches were excluded.

We included patients with at least 95% data collection (demographics, endocrinological, historical, radiological and intraoperative occurrence of CSF leakage), reliable information on extension of resection and post-operative CSF leakage occurrence, KI67 status at pathology report, early postoperative (within 72 hours) and 3-months postoperative volumetric contrast-enhanced MRI. Mild CSF leakages were repaired with synthetic dural patches and fibrin sealants combined with mucoperiosteal flap, while moderate to severe CSF leakages required multilayer autologous fascia lata grafts, often in association with tissue sealants and mucoperiosteal flap. All patients received nasal packing for the first 24 hours and those suffering from intraoperative CSF leak had subsequent CSF drainage through seriated lumbar punctures or external spinal shunt according to leak severity.

### Variables of interest and outcome:

All medical charts were reviewed by two investigators (L.T. and G.F.) and cross-matching of all tabular data was performed afterwards. Radiological (x-axis: coronal ; y-axis: craniocaudal ; z axis: anteroposterior diameters ) and resection measurements (volumetric EOR assessment on axial post contrast T1 scans) were performed by senior neuroradiologists. Osteo-dural invasiveness was stated as preoperative radiological bone invasion at T2-weighted, contrast-enhanced-T1 and CT scans(6). “R ratio” was defined as the ratio between horizontal tumor diameter and inter-carotid distance at the horizontal intracavernous segment (ICD). All PAs were graded according to Knosp(7) and Hardy(8) classifications by the senior author. All postoperative information (KI67, extension of resection and postoperative CSF leak occurrence) were discarded after exploratory analysis and not included in model building.

### Statistical analysis

Quantitative data were tested for normality. Normal variables were reported as mean  $\pm$  standard deviation (SD) and compared by Student t-test, skewed variables were reported as median and interquartile range (IQR) and tested by Wilcoxon test. Categorical variables were reported as absolute “counts (percentages)” and intergroup comparison was computed using Fisher’s exact test or Chi-square test. Missing data (<5%) were relocated using predictive mean matching. For all traditional hypotheses, p values < 0.05 were considered statistically significant.

A multivariate logistic regression model was fitted for independent predictors of intraoperative CSF leak. Sex, age, PA secretion status, prior EE-TNS surgery, R ratio, ICD, Knosp grades, Hardy grades, volume,

diameters in 3 axes and osteo-dural invasiveness were extracted as independent variables in the dataset, which was split preserving similar class distribution into a training (70%) and a hold-out-test sets (30%). Features selection was computed with the BORUTA algorithm(9) (parameter “importance” threshold: 70%) based on a random forest classifier model.

Class imbalance was corrected oversampling the minority class of the training set (SMOTE-NC: Synthetic minority over-sampling technique for nominal and continuous features(10)).

Based on an exploratory accuracy evaluation on the training set performed with a customized pipeline for model comparison (TPOT(11)), we selected 5 supervised-machine learning models across different subgroups of ML methods (Bayesian models, generalized linear models, marginal classifiers, decision trees generators and deep learning architectures) showing F1-score > 0.60 on training set without parameters tuning: naïve-bayes classifier (NB), multivariate logit regression (MLS), support-vector machine classifier (SVM-C), random forest classifier (RF) and artificial neural network (ANN) were included. We trained 5 optimized models on the training set, tuned hyperparameters with 10-fold cross validation using GridSearch(12,13) and measured their performance on the hold-out test set (data never approached by the model before) for final validation. Extensive performance metrics were analyzed and models were ranked based on overall area under the curve (AUC). Interval of confidence was reported for AUC imputed on both training and testing set. “Dropout” and “Early stop” methods were implemented to minimize overfitting in the ANN. Additional materials on the rationale of supervised-machine learning models implemented in the current study and structure of the ANN are provided in the **Supplementary File**. .

All statistical analyses were computed in Python language (v. 3.7.5.), with Anaconda? graphical user interface (GUI; <http://www.anaconda.com>) implementing scikit learn, pandas, numpy, seaborn, statsmodel libraries and built-in-Keras TensorFlow framework(14) (Google Brain Team?, Google LLC; <http://tensorflow.org>). The integrated pipeline of the current study is summarized in **Figure 1**.

## Results

### Population:

238 patients consecutively operated between 2014 and 2020 were included. Demographics characteristics of the population are reported in **Table 1** . Intraoperative CSF leakage occurred in 54 patients (22,6 %). Postoperative CSF leak occurred in 5 patients (2,1 %). All of them had experienced intraoperative CSF fistula and no cases of meningitis or tension pneumocephalus occurred. PAs with Ki67>3% at pathology examination were more represented in patients who suffered from intraoperative CSF leakage (48,1%vs 10,3%, p = 0.001). At the 3-month radiological follow-up, GTR was achieved in 120 patients (50,2 %).

### Classical inferential analysis of CSF leak predictors(**Table 2**):

Independent predictors of intraoperative CSF leak in the multivariate logistic regression analysis were: non secreting status (OR:9,7, p=0.00001), osteo-dural invasiveness (OR:0,34, p=0.01), higher age (OR:1,03, p=0.04) and reduced ICD (OR:0,88, p=0.0087). Overall, the multivariate logistic regression model showed a moderate predicting power (AUC: 0,60. Sensitivity: 55%. Specificity: 63%).

### Supervised machine learning predictive models

The Boruta algorithm selected the most relevant features according to model performance metrics: non secreting status, age, x-axis, y-axis, z-axis, ICD and R ratio. These were included in any further analysis. The dimensionality reduction of the implemented data corresponded to shortened training time, simplification of imputations and improvement in classification performance(15). Optimized models were ranked according to AUC in **Table 3**. Selected features importance for the optimal model were reported for interpretability (**Figure 2** ; a correlation plot is provided for comparison)

The random forest classifier demonstrated to be the most accurate ML model showing high training parameters (AUC: 0.88. Accuracy: 87%, Sensitivity: 95%. Specificity: 80%. F1-score: 0.88) and overall high discriminative capacity in predicting intraoperative CSF leak occurrence in the hold-out test set for internal

validation (AUC:0.84. Accuracy: 84%. Sensitivity: 87%. Specificity: 82%. PPV: 69%. NPV: 93%. F-1 score: 0.87).

## Discussion:

The ability to predict outcomes of specific surgical treatments is becoming mandatory to provide the highest standards of surgical care. It helps propose the treatment that fits best and allocate specific resources to patients most likely to benefit from them. On the personalized level, routine use of predictive technology could be of value in presurgical counseling, giving accurate prognostic information to both physicians and patients, powering medical decision, mitigating expectations and disappointment given by the incomprehension of patients and care-givers and reducing legal issues. On the systemic medical level, timely refraining from unnecessary testing or treatment would result in decreased complication rates and healthcare costs.

In the literature, the association between CSF leakage and preoperative risk factors has already been investigated, without univocal results. A recent review(16) recognized that only suprasellar intraventricular extension was consistently associated with CSF leakage. Other preoperative features considered associated with increased risk of CSF leak were: lower age, higher BMI and ACTH secretion. However, the preoperative characteristics of PAs patients, above all their incredible heterogeneity, make the use of traditional statistical models less reliable compared to other CNS diseases. Meta-analyses are lacking for the same reason.

We provided a comparative analysis of performance for classical statistical methods *vs* ML, and also between different ML technologies, demonstrating the feasibility to develop an internally validate prediction model based on supervised machine learning architecture.

According to our initial exploratory analysis, less invasive PAs (Knosp grade 2 or lower) were more represented among cases suffering from intraoperative CSF leakage (31,5% vs 14,7%) compared to more invasive PAs (Knosp grade 3: leak 5,6 % vs no-leak 18,5%). Parasellar extension, therefore, could play a major role into planning resection goal and determining intraoperative CSF leaks, since the latter were more frequently associated with a radical approach intention. A shorter ICD was more represented in patients who experienced intraoperative CSF leak (19,5+-4,13 vs 21,6+-3,84) as reduced ICD might facilitate cavernous sinus invasion, require additional tissue manipulation and overall laterally reduce surgical corridor during endoscopic approach, resulting in higher chance of traction on surrounding tissue and damage of the arachnoid.

Of interest, sellar osteodural invasiveness was related to lower chance of intraoperative CSF leak (59,2% vs 35,2%). A more prominent expansion downward the sphenoid sinus might have determined a less suprasellar growth, decreasing the risk of arachnoid tearing and leak.

Despite only moderate accuracy, the multivariate regression identified non-secreting status, osteodural invasiveness, older age and ICD as risk factors. In our experience, an increased age seems to be related to increased odds, in discordance with previous studies. It could be due to a predominant prevalence of macroadenomas (n=190, 82,8%) in our cohort, most likely occurring in older patients with higher risk of CSF leakage. Intraoperative (22.69%) and postoperative (2,1%) CSF leakage rate in our population is in line with previous literature, determining a reliable milestone for generalizability of the model(17). Being ACTH- and GH- secreting PAs a minor proportion of the population investigated, it is not surprising the current study did not show a relevance for these variables. In fact, intraoperative CSF leak occurrence could be higher in younger adults suffering from Cushing's disease, since TNS surgery plays a curative role in this invalidating condition and aggressive strategies are usually recommended. On the contrary, GH-secreting microadenomas in older adults could receive more conservative treatments as adjuvant medical therapies are shown to be suitable for disease control. Non secreting status – associated with major extrasellar invasion and firm consistency with not well defined frequency – might be addressed as major risk factor (highest odds in our study). In our traditional analysis, non-secreting status showed increased odds of CSF leak occurrence independently from diameter measurements, implying additional factors might work in favor of augmented risk of leakage.

These patients might require additional attention when no macroscopic leakages are identified during surgery:

in fact, patients selected by the predictive tool as at higher risk of CSF leakage might benefit of careful exploration of the surgical corridor for identifying even occult low-flow leakages occasionally left undetected during endoscopic resection. These patients might require preventive sellar floor repair even in the absence of CSF leakage after scrupulous inspection, especially in frail multimorbid patients who would suffer most from postoperative complications.

The implementation of a feature selecting algorithm (BORUTA) help identify the most performing predictors in our population: **“non-secreting status”, “higher age”, “x-axis”, “y-axis”, “z-axis”, “ICD” and “R ratio”**. Compared to previous classical analyses, the Knosp classification was outperformed by R ratio. Moreover, in addition to the confirmed endocrinological and demographic predictors (higher age and non-secreting status), dimensional tumor measurements also resulted highly predictive of CSF leakage occurrence.

Our supervised ML model successfully passed internal validation test, with particular reference to random forest classifier which showed high discriminative capacity in predicting intraoperative CSF leak occurrence (**Table 3**).

To the best of our knowledge, this is the first study to compare different ML models and their performances on CSF leakage occurrence prediction in EE-TNS surgery. Unlike previous studies, where different models were arbitrary picked by the investigators without arguing the reasons (about 87,5% of all studies, according to a recent systematic review(18)), our workflow included a preliminary parallel analysis of a consistent number of different models among which we selected a subgroup for further analyses based on F-1 score before hyperparameters tuning.

Startjees et colleagues already investigated the reliability of a predictive ML model for intraoperative CSF leakages(19): they reported high accuracy on a small monocentric population with a single ANN-based model. In our analysis, RF outperformed every other tested model, including ANN, in accordance with previous evidence where tabular data were used as inputs[36]. RFs - in fact - can train on small datasets and deal with missing data, while ANNs require larger datasets and features normalization. In addition, the application of ANN on small population results less prone to generalizability because of overfitting and biases in the sample characteristics might play a major role. It is also noteworthy that ANNs, as well as other deep learning technologies, work on implicit relationships between input and output features. This so called “black-box” process prevents explicit workflow of the analysis from being extracted. On the contrary, RFs can be manipulated with several approaches to extract the most important input features for further clinical discussion and implementation(22).

Despite high-quality data training permits ML models to follow complex non-linear interactions and compute accurate predictions, the interpretation of such results should remain speculative and experimental. A limitation of our study, in fact, is the monocentric design which could poses undersurface biases if patients treated in other institutions and by different surgeons are tested with our tool. Therefore, an external validation must investigate the generalizability of the model.

With prospective population inclusion and external data validation we will be able in the near future to generalize this ML-powered tool in prevision of a deployment in the clinical practice.

## Conclusions:

We believe that machine learning can improve the current planning and perioperative management of PAs. In this study, we provided a pipeline for training and validating different supervised machine learning models. Our random forest classifier (RF) predicted intraoperative CSF leak occurrence with an accuracy of 87% in the training set and 84% in hold-out test (Sensitivity 87%. Specificity 82%). We encourage other institutions to join our mission and share their surgical experience to develop a tool able to assist daily neurosurgical practice.

## References:

1. Ostrom QT, Cioffi G, Gittleman H, Patil N, Waite K, Kruchko C, et al. CBTRUS Statistical Report:

- Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012-2016. *Neuro Oncol.* 2019;
2. Tabaee A, Anand VK, Barron Y, Hiltzik DDH, Brown SM, Kacker A, et al. Endoscopic pituitary surgery: A systematic review and meta-analysis: Clinical article. *J Neurosurg.* 2009;111(3):545–54.
3. Nishioka H, Haraoka J, Ikeda Y. Risk factors of cerebrospinal fluid rhinorrhea following transsphenoidal surgery. *Acta Neurochir (Wien).* 2005;147(11):1163–6.
4. Ivan C, Ann R, Craig B, Debi P. Topic Review Complications of Transsphenoidal Surgery : Results of a National Survey , Review of the Literature , and Personal Experience Abstract The primary objectives of this report were , first , to determine the number and incidence of complications . 1997;(February):225–37.
5. Senders JT, Zaki MM, Karhade A V., Chang B, Gormley WB, Broekman ML, et al. An introduction and overview of machine learning in neurosurgical care. *Acta Neurochir (Wien).* 2018;160(1):29–38.
6. C B Luo, M M Teng, S S Chen, J F Lirng, F C Chang, W Y Guo, et al. Imaging of Invasiveness of Pituitary Adenomas - PubMed. *Kaohsiung J Med Sci [Internet].* 2000 [cited 2020 May 10];16(1):26–31. Available from: <https://pubmed.ncbi.nlm.nih.gov/10741013/>
7. Micko ASG, Wohrer A, Wolfsberger S, Knosp E. Invasion of the cavernous sinus space in pituitary adenomas: Endoscopic verification and its correlation with an MRI-based classification. *J Neurosurg.* 2015;
8. Hardy J, Vezina JL. Transsphenoidal neurosurgery of intracranial neoplasm. *Adv Neurol.* 1976;
9. Kursu MB, Jankowski A, Rudnicki WR. Boruta - A system for feature selection. *Fundam Informaticae.* 2010;
10. Chawla N V., Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res.* 2002;
11. Olson RS, Moore JH. TPOT: A Tree-Based Pipeline Optimization Tool for Automating Machine Learning. In 2019.
12. Brownlee J. A Gentle Introduction to k-fold Cross-Validation. *machinelearningmastery.com.* 2019.
13. Ghawi R, Pfeffer J. Efficient Hyperparameter Tuning with Grid Search for Text Categorization using kNN Approach with BM25 Similarity. *Open Comput Sci.* 2019;
14. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: A system for large-scale machine learning. In: *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2016.* 2016.
15. Hall MA, Holmes G. Benchmarking Attribute Selection Techniques for Discrete Class Data Mining. *IEEE Trans Knowl Data Eng.* 2003;
16. Lobatto DJ, de Vries F, Zamanipoor Najafabadi AH, Pereira AM, Peul WC, Vliet Vlieland TPM, et al. Preoperative risk factors for postoperative complications in endoscopic pituitary surgery: a systematic review. *Pituitary.* 2018;21(1):84–97.
17. Strickland BA, Lucas J, Harris B, Kulubya E, Bakhsheshian J, Liu C, et al. Identification and repair of intraoperative cerebrospinal fluid leaks in endonasal transsphenoidal pituitary surgery: Surgical experience in a series of 1002 patients. *J Neurosurg.* 2018;
18. Qiao N. A systematic review on machine learning in sellar region diseases: Quality and reporting items. *Endocr Connect.* 2019;
19. Staartjes VE, Zattra CM, Akeret K, Maldaner N, Muscas G, Bas van Niftrik CH, et al. Neural network–based identification of patients at high risk for intraoperative cerebrospinal fluid leaks in endoscopic pituitary surgery. *J Neurosurg.* 2019;

20. Nawar S, Mouazen AM. Comparison between random forests, artificial neural networks and gradient boosted machines methods of on-line Vis-NIR spectroscopy measurements of soil total nitrogen and total carbon. *Sensors* (Switzerland). 2017;
21. Senders JT, Staples P, Mehrtash A, Cote DJ, Taphoorn MJB, Reardon DA, et al. An Online Calculator for the Prediction of Survival in Glioblastoma Patients Using Classical Statistics and Machine Learning. *Clin Neurosurg*. 2020;
22. Banerjee M, Ding Y, Noone AM. Identifying representative trees from ensembles. *Stat Med*. 2012;

## Figures:

**Figure 1:** Study design and Machine learning model building pipeline. Here are reported all phases of the current study: 1) Patients selection: inclusion and exclusion criteria definition; 2) Data extraction: medical chart review and radiological measurements; 3) Data preprocessing: dataset construction and variables definition; 4) Data splitting: definition of a training set (70% of overall data) and an hold-out test set (30%) for final internal validation; 5) Features selection: the features selecting BORUTA algorithm discarded not relevant variables in improving model accuracy and pointed out a minor proportion of preoperative data as implicated in intraoperative CSF leakage occurrence; 6) Minory class imbalance oversampling: as CSF leakage occurred in a small proportion (22,6%) of patients, performance evaluation of ML models would be negatively influenced by such a disproportion between outcome classes (occurrence of CSF intraoperative leakage or not). The SMOTE-NC algorithm permits comparison by oversampling minority class ; 7) Model selection on training set: a customized pipeline based on TPOT was coded and best models according to F-1 score were picked; 8) Model optimization: 10-fold CV was run for hyperparameter optimization; 9) Model performance report on hold-out test set: best five optimized models were tested on data never trained on before (hold-out) and their performance was reported.

**Figure 2:** Left: Features importance plot as computed by the features selecting algorithm BORUTA Right: Correlation plot provided for comparison shows classical statistical inference by Pearson correlation test.

## Additional materials:

### Supplementary file:

**Table 1:** Machine learning dictionary and overview

**File 1:** Models parameters

### Hosted file

CO\_Tables.pdf available at <https://authorea.com/users/373120/articles/490905-a-supervised-machine-learning-powered-tool-intraoperative-csf-leak-predictor-in-endoscopic-transsphenoidal-surgery-for-pituitary-adenomas>

