

2 **Table 1:** Demographic statistics of the population
3

		No Intraoperative CSF leakage				Intraoperative CSF leakage				P value
		Count	N %	Mean	SD	Count	N %	Mean	SD	
Sex	Female	99	53,8%			32	59,3%			P > 0.05
	Male	85	46,2%			22	40,7%			P > 0.05
Age (years)				55,48	15,55			59,42	12,91	P > 0.05
Prior TNS surgery		39	21,2%			10	18,5%			P > 0.05
Non-Functioning		135	73,4%			44	81,5%			P > 0.05
Prolactin		4	2,2%			1	1,9%			P > 0.05
GH		15	8,2%			2	3,7%			P > 0.05
ACTH		25	13,6%			7	13,0%			P > 0.05
Secreting status: Other		5	2,7%			0	0,0%			P > 0.05
Preoperative medical therapy		40	21,7%			10	18,5%			P > 0.05
Atypical PA (KI67>3%)*										
	>=3	19	10,3%			26	48,1%			p = 0.001
Y-axis (mm)				18,50	10,21			18,54	9,67	P > 0.05
Z-axis (mm)				15,95	7,46			15,56	7,83	P > 0.05
X-axis (mm)				18,01	8,65			16,93	8,21	P > 0.05
Total Volume (mm3)				4726,18	5833,07			4223,15	6453,75	P > 0.05
ICD (mm)				21,59	3,84			19,49	4,13	p = 0.001
R ratio				0,82	0,36			0,89	0,40	P > 0.05

Knosp grade	0	44	23,9%	14	25,9%	P > 0.05
	1	66	35,9%	16	29,6%	P > 0.05
	2	27	14,7%	17	31,5%	p = 0.005
	3A/3B	34	18,5%	3	5,6%	p = 0.021
	4	13	7,1%	4	7,4%	P > 0.05
Hardy grade (Sellar)	0	44	23,9%	15	27,8%	P > 0.05
	1	42	22,8%	9	16,7%	P > 0.05
	2	57	31,0%	20	37,0%	P > 0.05
	3	33	17,9%	9	16,7%	P > 0.05
	4	8	4,3%	1	1,9%	P > 0.05
Hardy grade (Suprasellar)	0	60	32,6%	16	29,6%	P > 0.05
	A	74	40,2%	23	42,6%	P > 0.05
	B	44	23,9%	10	18,5%	P > 0.05
	C	4	2,2%	3	5,6%	P > 0.05
	D	2	1,1%	2	3,7%	P > 0.05
Osteodural Invasiveness		109	59,2%	19	35,2%	P = 0.002
Gross total resection (GTR)*		93	50,3%	27	50,0%	P > 0.05
Post-operative CSF leakage*		0	0,00%	5	2,1 %	P = 0.003

4 Hereby all demographic characteristics of the population examined are reported, stratified per classes (occurrence of intraoperative CSF leakage or not)

5 (*) = Marked variables represent post-operative and follow-up characteristics. They were excluded from predictive analyses as for the aim of the study.

6 SD: standard derivation. CSF: cerebrospinal fluid. TNS: transnasal-surgery. GH: growth hormone. ACTH: adrenocorticotrophic hormone. PA: pituitary adenoma. Y-axis: craniocaudal diameter.

7 Z-axis: anteroposterior diameter. X-axis: coronal diameter. ICD: intercarotid distance.

Table 2: Multivariate logistic regression analysis

	OR	5% C.I.	95% C.I.	P-value
Sex	0,80	0,3669	1,7569	0,5827
Non Secreting	9,77	3,1708	30,0959	0,0001
Prolactin	25,59	0,7388	88,6930	0,0730
GH	0,84	0,0712	9,8567	0,8882
ACTH	2,69	0,4850	14,9334	0,2575
Other	0,01	0,0001	99,8979	0,9943
Reoperation	0,85	0,2987	2,4110	0,7580
Preoperative drug therapy	1,12	0,2697	4,6337	0,8778
Knosp grade (Parasellar)	1,17	0,7051	1,9491	0,5398
Hardy grade (Sella)	1,17	0,6871	2,0040	0,5581
Hardy grade (Suprasellar)	1,46	0,7155	2,9984	0,2964
Osteodural invasiveness	0,34	0,1473	0,7996	0,0132
Age	1,03	0,9995	1,0517	0,0449
Y-axis: craniocaudal max. diameter	0,95	0,8619	1,0414	0,2632
Z-axis: anteroposterior max. diameter	0,98	0,8616	1,1181	0,7787
X-axis: lateral max. diameter	0,94	0,8140	1,0752	0,3479
Volume	1,00	1,0000	1,0002	0,2038
ICD	0,88	0,8018	0,9685	0,0087
R ratio	0,97	0,1127	8,3384	0,9773

The table reports the traditional multivariate regression analysis performed on the independent variables as predictors of the occurrence of intraoperative CSF leakage. Overall accuracy of the multivariate logistic regression model was moderate (0.60).

OR: odd ratio. C.I.: confidence interval. CSF: cerebrospinal fluid. TNS: transnasal-surgery. GH: growth hormone. ACTH: adrenocorticotrophic hormone. PA: pituitary adenoma. Y-axis: craniocaudal diameter. Z-axis: anteroposterior diameter. X-axis: coronal diameter. ICD: intercarotid distance.

19 Table 3: Comparative machine learning models performance analysis.

20

21

Model	Metrics	Training Set	95% C.I.	Test Set (Validation)	95% C.I.	Ranks
Random Forest Classifier (RF)	AUC	0,880	0,860-0,900	0,840	0,82-0,86	#1
	Accuracy	87%		84%		
	Sensitivity	95%		87%		
	Specificity	80%		82%		
	PPV (Precision)	82%		69%		
	NPV	94%		93%		
	F-1 score	0,880		0,870		
SVM Classifier (SVM-C)	AUC	0,980	0,971-0,989	0,720	0,691-0,749	#2
	Accuracy	98%		60%		
	Sensitivity	98%		50%		
	Specificity	98%		98%		
	PPV (Precision)	90%		50%		
	NPV	90%		72%		
	F-1 score	0,900		0,740		
Artificial Neural Network (ANN)	AUC	0,820	0,769-0,876	0,690	0,661-0,719	#3
	Accuracy	80%		68%		
	Sensitivity	91%		69%		
	Specificity	90%		68%		
	PPV (Precision)	70%		33%		
	NPV	100%		91%		
	F-1 score	0,700		0,320		
Logistic Regression	AUC	0,701	0,655-0,747	0,600	0,568-0,632	#4

Classifier (LRC)	Accuracy	70%		60%		
	Sensitivity	66%		55%		
	Specificity	74%		63%		
	PPV (Precision)	72%		50%		
	NPV	69%		82%		
	F-1 score	0,690		0,400		
Naive Bayes Classifier (NB)	AUC	0,980	0,968-0,992	0,550	0,496-0,604	#5
	Accuracy	67%		59%		
	Sensitivity	97%		67%		
	Specificity	97%		50%		
	PPV (Precision)	96%		37%		
	NPV	97%		78%		
	F-1 score	0,710		0,610		

22 The selected machine learning models were ranked according to their AUC on hold-out test set. AUC: area under the curve. PPV: positive predictive value. NPV: negative predictive value.