

Prediction of hyperaldosteronism subtypes when adrenal vein sampling is unilaterally successful

*Original*

Prediction of hyperaldosteronism subtypes when adrenal vein sampling is unilaterally successful / Burrello, Jacopo; Burrello, Alessio; Pieroni, Jacopo; Sconfienza, Elisa; Forestiero, Vittorio; Amongero, Martina; Rossato, Denis; Veglio, Franco; Williams, Tracy A; Monticone, Silvia; Mulatero, Paolo. - In: EUROPEAN JOURNAL OF ENDOCRINOLOGY. - ISSN 0804-4643. - 183:6(2020), pp. 657-667. [10.1530/EJE-20-0656]

*Availability:*

This version is available at: 11583/2978528 since: 2023-05-16T11:04:40Z

*Publisher:*

Oxford University press

*Published*

DOI:10.1530/EJE-20-0656

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

Oxford University Press preprint/submitted version

This article has been accepted for publication in EUROPEAN JOURNAL OF ENDOCRINOLOGY Published by Oxford University Press

(Article begins on next page)

1 **Prediction of hyperaldosteronism subtypes when adrenal vein sampling is**  
2 **unilaterally successful**

3 Jacopo Burrello<sup>1\*</sup>, Alessio Burrello<sup>2\*</sup>, Jacopo Pieroni<sup>1</sup>, Elisa Sconfienza<sup>1</sup>, Vittorio Forestiero<sup>1</sup>,  
4 Martina Amongero<sup>3</sup>, Denis Rossato<sup>4</sup>, Franco Veglio<sup>1</sup>, Tracy A. Williams<sup>1,5</sup>, Silvia Monticone<sup>1#</sup>,  
5 Paolo Mulatero<sup>1#</sup>.

6  
7 (1) Division of Internal Medicine and Hypertension, Department of Medical Sciences, University of  
8 Torino, Italy. (2) Department of Electrical, Electronic and Information Engineering "Guglielmo  
9 Marconi" (DEI), University of Bologna, Italy. (3) Department of Mathematical Sciences G. L.  
10 Lagrange, Polytechnic University of Torino, Italy. (4) AOU Città della Salute e della Scienza -  
11 Service of Radiology, Torino, Italy. (5) Medizinische Klinik und Poliklinik IV, Klinikum der  
12 Universität, Ludwig-Maximilians-Universität München, Munich, Germany.

13 \* Contributed equally and should be considered as joint first authors.

14 # Contributed equally and should be considered as joint last authors.

15  
16 Corresponding author: Prof. Paolo Mulatero - Division of Internal Medicine and Hypertension Unit,  
17 Department of Medical Sciences, University of Torino, Città della Salute e della Scienza di Torino,  
18 Via Genova 3, 10126 Torino, Italy. Telephone/Fax number: 0039.011.633.6959 / 0039.011.633.6931.  
19 E-mail: paolo.mulatero@unito.it

20 Manuscript word count: 3,440 excluding tables, references and figure captions.

21 Abstract word count: 250.

22 Number of Tables: 3; Number of Figures: 4; 1 Supplementary file.

23 **Short title:** Subtype diagnosis in primary aldosteronism

24 **Keywords:** aldosterone; primary aldosteronism; contralateral suppression; adrenal venous sampling;  
25 machine learning.

26 **ABSTRACT**

27 **Objective** – Adrenal venous sampling (AVS) is the gold standard to discriminate patients with  
28 unilateral primary aldosteronism (UPA) from bilateral disease (BPA). AVS is technically-demanding  
29 and in cases of unsuccessful cannulation of adrenal veins, the results may not always be interpreted.  
30 The aim of our study was to develop diagnostic models to distinguish UPA from BPA, in cases of  
31 unilateral successful AVS and the presence of contralateral suppression of aldosterone secretion.

32 **Design** – Retrospective evaluation of 158 patients referred to a tertiary hypertension unit who  
33 underwent AVS. We randomly assigned 110 patients to a training cohort and 48 patients to a  
34 validation cohort to develop and test the diagnostic models.

35 **Methods** – Supervised machine learning algorithms and regression models were used to develop and  
36 validate two prediction models and a simple 19-point score system to stratify patients according to  
37 their subtype diagnosis.

38 **Results** – Aldosterone levels at screening and after confirmatory testing, lowest potassium, ipsilateral  
39 and contralateral imaging findings at CT scanning, and contralateral ratio at AVS, were associated  
40 with a diagnosis of UPA and were included in the diagnostic models. Machine learning algorithms  
41 correctly classified the majority of patients both at training and validation (accuracy 82.9-95.7%).  
42 The score system displayed a sensitivity/specificity of 95.2/96.9%, with an AUC of 0.971. A flow-  
43 chart integrating our score correctly managed all patients except 3 (98.1% accuracy), avoiding the  
44 potential repetition of 77.2% of AVS procedures.

45 **Conclusions** – Our score could be integrated in clinical practice and guide surgical decision-making  
46 in patients with unilateral successful AVS and contralateral suppression.

47 **ABBREVIATION LIST:** A/C, Aldosterone-to-Cortisol ratio; ARR, Aldosterone-to-Renin Ratio;  
48 AVS, Adrenal Venous Sampling; BPA, Bilateral Primary Aldosteronism; CLR, Contralateral ratio;  
49 CT, Computed Tomography; DDD, Defined Daily Dose; IVC, Inferior Vena Cava; LAV, Left  
50 Adrenal Vein; LDA, Linear Discriminant Analysis; LI, Lateralization Index; PA, Primary  
51 Aldosteronism; PRA, Plasma Renin Activity; RAV, Right Adrenal Vein; RF, Random Forest; SI,  
52 Selectivity Index; UPA, Unilateral Primary Aldosteronism.

53

## 54 **INTRODUCTION**

55 Primary aldosteronism (PA) is a common secondary cause of arterial hypertension<sup>1,2</sup>, associated with  
56 an increased cardiovascular risk compared with patients with essential hypertension<sup>3,4</sup>. This condition  
57 may be determined either by unilateral or bilateral hypersecretion of aldosterone by the adrenal  
58 glands, justifying the two major subtypes of PA, aldosterone producing adenoma, and bilateral  
59 adrenal hyperplasia.

60 Patients with a diagnosis of unilateral PA (UPA) derive a clinical benefit after surgical adrenalectomy  
61 in more than 80% of cases<sup>5-7</sup>. Medical therapy with mineralocorticoid receptor antagonists is  
62 recommended for patients with bilateral disease<sup>7</sup>. The correct treatment leads to a significant  
63 improvement of long-term outcomes, reverting the cardiovascular risk excess displayed by these  
64 patients<sup>8</sup>. For these reasons, an early diagnosis of PA and subtype differentiation is fundamental.  
65 Guidelines recommends adrenal venous sampling (AVS) as the gold standard to discriminate UPA  
66 from bilateral PA (BPA)<sup>1-3</sup>.

67 However, AVS may not be interpretable when cannulation is unsuccessful for one or both adrenal  
68 veins. The cannulation of the left adrenal vein (LAV) is relatively straight forward, because it merges  
69 with the inferior phrenic vein to form a common vessel draining into the left renal vein. Conversely,  
70 the cannulation of the right adrenal vein (RAV), which is shorter and smaller, directly drains into the  
71 inferior vena cava (IVC) at an acute angle and may thus be challenging<sup>9,10</sup>. The correct cannulation

72 of RAV and LAV is assessed by the selectivity index (SI), which is calculated as the cortisol ratio  
73 between adrenal veins and IVC. The ratio between the aldosterone-to-cortisol (A/C) ratio in the  
74 dominant adrenal vein and the A/C ratio in the non-dominant vein is defined as lateralization index  
75 (LI), and displays high sensitivity and specificity to distinguish UPA from bilateral forms<sup>3,10,11</sup>.

76 Several studies have described the diagnostic performance of the contralateral ratio (CLR, defined as  
77 A/C ratio in the non-dominant adrenal vein divided by A/C ratio in the IVC). Contralateral  
78 suppression is defined in the presence of an A/C ratio in the non-dominant adrenal vein lower than  
79 the A/C ratio in IVC, and a CLR lower than 1 is used in some centers to define lateralization when  
80 only one adrenal vein is successfully cannulated<sup>12,13</sup>. Patients with UPA should display a contralateral  
81 normal adrenal gland with suppressed aldosterone secretion from the *zona glomerulosa*<sup>14,15</sup>.  
82 Nevertheless, the CLR cannot predict UPA by itself, because about one third of patients with BPA  
83 also display contralateral suppression<sup>11</sup>, and for the occurrence of apparent bilateral aldosterone  
84 suppression, defined as adrenal A/C ratio bilaterally lower than that in IVC<sup>16</sup>. Thresholds to interpret  
85 CLR and assess lateralization of aldosterone secretion have been not yet defined.

86 Therefore, the aim of our study was to develop and validate a diagnostic model able to discriminate  
87 UPA from BPA, when the AVS is unilaterally successful and the adrenal vein A/C ratio is lower than  
88 the A/C ratio in the IVC (so called contralateral suppression).

## 90 **METHODS**

### 91 ***Study cohort and data extraction***

92 We retrospectively evaluated 158 patients who underwent AVS in the Hypertension Unit of Torino  
93 (Italy), between 2008 and 2019. All patients gave informed written consent to the study according to  
94 Helsinki declaration. The study was approved by the local ethical committee of the hospital “Città  
95 della Salute e della Scienza” of Torino. PA was diagnosed according with the Endocrine Society  
96 Guideline<sup>1</sup>. Patients were screened using the aldosterone-to-plasma renin activity (PRA) -ratio

97 (ARR); if possible, all interfering drugs were withdrawn for at least 4 weeks prior to analysis (6 weeks  
98 for mineralocorticoid receptor antagonist and diuretics). The cut-off for a positive screening test was  
99 an ARR greater than  $30 \text{ ng/dL/ng} \cdot \text{mL}^{-1} \cdot \text{h}^{-1}$  together with an aldosterone concentration greater than  
100  $10 \text{ ng/dL}$ . Patients with a positive screening test underwent confirmatory/exclusion testing by either  
101 intravenous saline loading test or a captopril challenge test<sup>1</sup>. Confirmed patients with PA underwent  
102 multi-slice computed tomography (CT) scanning and AVS. An adrenal nodule was reported in the  
103 presence of a mass equal or greater than 8 mm. AVS was performed under basal conditions or after  
104 continuous cosyntropin (ACTH) infusion (76 vs. 82, respectively). The catheterization of adrenal  
105 veins was sequential in both protocols. In case of cosyntropin stimulation, the infusion was performed  
106 at  $50 \text{ ug/h}$ , starting 30 minutes before the procedure. AVS procedures were performed by a single  
107 experienced interventional radiologist. The radiologist firstly cannulates the right adrenal vein and  
108 subsequently the left adrenal vein (within 15 minutes). ACTH stimulation was necessary for patients  
109 who need preparation with steroids before the procedure for a history of contrast allergy (or other  
110 allergy;  $n=13$ ), or when the procedure was not performed in the early morning ( $n=69$ )<sup>10</sup>. Adrenal vein  
111 cannulation was considered successful in the presence of a SI equal or greater than 3 under basal  
112 conditions, and of 5 under cosyntropin infusion. The diagnosis of UPA was established when the LI  
113 was at least 4<sup>10</sup>.

114 Patients were included in the study if: 1) the AVS was bilaterally successful; 2) in presence of  
115 contralateral suppression, that is when at least one adrenal vein A/C ratio was lower than the A/C  
116 ratio in the IVC ( $\text{CLR} < 1$ ). Hypercortisolism was excluded in all patients. Enrolled patients were  
117 randomly assigned to a training cohort ( $N = 110$ ) or to a validation cohort ( $N = 48$ ), with a ratio 70:30.  
118 Contralateral and ipsilateral adrenal glands were defined on the basis of AVS findings. In case of  
119 UPA, the contralateral adrenal was the gland on the side opposite to the lateralization. In case of BPA  
120 diagnosis, the contralateral adrenal was defined as the gland showing an A/C ratio lower than in the  
121 IVC. Ipsilateral and contralateral adrenal imaging at CT scanning were defined accordingly. To build

122 our model, aldosterone and cortisol measurements from the IVC and from the adrenal vein with A/C  
123 ratio lower than in the IVC were used. When UPA is predicted, the ipsilateral adrenal gland should  
124 be removed. For all patients, antihypertensive medication was quantified as Defined Daily Dose  
125 (DDD), which is the average maintenance dose per day for a drug used for its main indication in  
126 adults.

### 127 ***Machine learning and statistical analysis***

128 IBM SPSS Statistics 22 (*IBM Corp., Armonk, New York, USA*) and Python 3.5 (library, scikit-learn)  
129 were used for statistics. The study followed the TRIPOD statement for transparent reporting of a  
130 multivariable prediction model for individual prognosis or diagnosis.

131 Sample size was calculated on the estimated variability for median value of CLR. With a mean effect  
132 size (Cohen's d coefficient) of 1.24, a minimum power ( $1 - \beta$  error probability) of 95%, and a  
133 significance level ( $\alpha$  error) of 0.05, the estimated total sample size was 44. Sample size calculation  
134 was carried out with GPower 3.1.

135 Kolmogorov–Smirnov test was used to assess variable distribution. Normally and non-normally  
136 distributed variables were reported as mean  $\pm$  standard deviation, or median and interquartile range,  
137 and analyzed by T-student, or Mann-Whitney's test, respectively. Categorical variables were reported  
138 as absolute number and percentage distribution and analyzed by Chi-square test. Logistic regression  
139 analyses were used for univariate and multivariate models; an odds ratio (OR) greater than 1 was  
140 associated with an increased likelihood of diagnosis of UPA. Supervised machine learning techniques  
141 were used to develop the diagnostic models and the clinical score, as previously described<sup>17</sup>. Briefly,  
142 linear discriminant analysis (LDA) and random forest (RF) algorithms were applied to develop two  
143 different diagnostic models able to discriminate UPA vs. BPA using the best discriminant patients'  
144 parameters, previously selected by univariate and multivariate regression analysis. The models were  
145 developed after correction of dataset imbalance using a SMOTE algorithm (Synthetic Minority  
146 Oversampling Technique). Briefly, SMOTE imputes new patient data (defined as “neighbour”), along

147 lines that link real patients from the training dataset in the virtual space created by patient parameters,  
148 in order to balance the number of UPA and BPA patients at model training. A number of neighbours  
149 equal to 3 (K=3), corresponded to the higher accuracy and was used in all analyses.

150 The LDA computes a set of coefficients for linear combination of each variable to classify patients  
151 according with their diagnosis. The predicted diagnosis is derived from the following equation: UPA  
152 =  $LDAcoeff_1 * Variable_1 + LDAcoeff_2 * Variable_2 + \dots + LDAcoeff_n * Variable_n > 9.7425$ . The RF  
153 algorithm is composed by 20 classification trees with a maximum number of 8 splits. The predicted  
154 diagnosis resulted from the outcome of each classification tree. If at least 11 of 20 trees of the RF  
155 predicted UPA, then the patient is classified accordingly. Machine learning models were tested by a  
156 10-fold validation algorithm: (i) the cohort was randomly divided into 10 groups; (ii) the model was  
157 trained within the first 9 groups and validated with the remaining group; (iii) the process was repeated  
158 10 times, rotating the validation group at each round. The accuracy at validation resulted from the  
159 mean of the accuracy obtained at each round. A 19-point score based on the same discriminant  
160 parameters was finally developed in the training cohort and tested by 10-fold cross validation (with  
161 bootstrapping) and in both validation and combined cohorts. Points for each reference category and  
162 relative cut-offs were automatically derived to achieve the best accuracy. The analysis of ROC curve  
163 was used to assess the area under the curve (AUC) and derive the Youden Index ( $J = \text{sensitivity} +$   
164  $\text{specificity} - 1$ ). Accuracy was defined as the ratio between “true positive” plus “true negative” divided  
165 by the total number of patients included in each cohort. An online tool was developed to calculate the  
166 score and define the diagnosis (available at [https://github.com/ABurrello/CLR-](https://github.com/ABurrello/CLR-score/raw/master/CLR%20Score%20Calculator.xlsm)  
167 [score/raw/master/CLR%20Score%20Calculator.xlsm](https://github.com/ABurrello/CLR-score/raw/master/CLR%20Score%20Calculator.xlsm)). All data supporting the findings of this study  
168 are available from the corresponding author upon reasonable request.

169

## 170 **RESULTS**

### 171 *Patient characteristics*

172 One hundred and fifty-eight patients underwent AVS for PA subtyping and were included in the  
173 analysis. The mean age at diagnosis of PA was  $50 \pm 10$ , mean BP was 164/98, with a duration of  
174 hypertension of 67 [23; 130] months. Seventy-six patients underwent AVS under basal conditions,  
175 and 82 after ACTH stimulation. No differences were observed between patients who underwent AVS  
176 with or without ACTH except for SI higher under ACTH stimulation ( $P = 0.002$ ) (Supplementary  
177 Table S1). For this reason, patients were grouped together for all subsequent analyses, irrespective of  
178 the protocol used for AVS.

179 Clinical and biochemical characteristics of patients after subtype diagnosis are reported in Table 1;  
180 126 patients were diagnosed as UPA, and 32 as BPA. The high prevalence of unilateral forms is  
181 expected, because only patients with contralateral suppression at AVS were enrolled in this study.  
182 Patients with a diagnosis of UPA were younger ( $49 \pm 10.3$  vs.  $55 \pm 7.1$ ;  $P = 0.001$ ), frequently females  
183 (42.1% vs. 21.9%;  $P = 0.036$ ), with a higher DDD (4 [2.5; 6] vs. 2.9 [1.4; 4.3]  $P = 0.008$ ). The  
184 prevalence of microalbuminuria, left ventricular hypertrophy, and prior cardiovascular events was  
185 not different between groups.

186 UPA patients displayed lower potassium levels ( $3.1 \pm 0.6$  vs.  $3.9 \pm 0.4$  mEq/L;  $P < 0.001$ ) and higher  
187 aldosterone, both at screening (38.2 [25.9; 49.8] vs. 30.2 [21.1; 41] ng/dL;  $P = 0.044$ ) and after  
188 confirmatory testing (21.9 [13.9; 35.5] vs. 11.4 [7.7; 19.9] ng/dL;  $P < 0.001$ ). At CT scanning,  
189 ipsilateral imaging was normal in 1.6% of patients with UPA, compared to 43.8% of BPA patients;  
190 conversely, contralateral imaging was normal in 84.9% of UPA and 53.1% of BPA patients ( $P <$   
191  $0.001$  for both comparisons). After AVS, contralateral A/C and CLR were both lower in patients with  
192 UPA (4.1 [1.5; 8.3] vs. 7.9 [3.2; 12] and 0.3 [0.2; 0.5] vs. 0.7 [0.5; 0.8], respectively).

193 Univariate logistic regression analysis confirmed the association with the diagnosis of UPA for all  
194 these parameters (Supplementary Table S2). In particular, female sex (OR 2.59), age at diagnosis  
195 (OR 0.94), DDD (OR 1.43), lowest potassium (0.07), aldosterone at screening (OR 1.01) and after  
196 confirmatory testing (1.01), a normal ipsilateral and contralateral imaging (OR 0.02 and OR 4.97,

197 respectively), contralateral A/C (OR 0.91), and CLR (OR 0.01). Six of these 10 variables were  
198 selected considering their discriminative performance and expert knowledge<sup>18</sup>, and were introduced  
199 in the multivariate model (Table 2), which confirmed their independent association with UPA  
200 diagnosis (female sex, age at diagnosis, DDD, and contralateral A/C were no longer associated with  
201 UPA when introduced in the multivariate analysis; data not shown).

202

### 203 *Linear discriminant analysis and random forest models*

204 Patients included in the analysis were randomized into two subgroups: the training cohort (N = 110)  
205 and the validation cohort (N = 48). No differences were found for all the evaluated variables  
206 (Supplementary Table S3).

207 Supervised machine learning algorithms were used to develop two diagnostic models: an LDA model  
208 and a RF classification algorithm. The models were developed in the training cohort and subsequently  
209 tested by both internal validation (10-fold cross validation applied to the same training cohort) and  
210 external validation (on the randomly selected validation cohort). Supplementary Table S4 reports the  
211 diagnostic performance at training, internal and external validation, and on the combined cohort. The  
212 linear combination of the 6 above-selected variables is represented in the canonical plot (Figure 1A);  
213 each patient is represented by a point, and the clear separation according to their subtype diagnosis  
214 indicates the high diagnostic performance of the LDA model (accuracy 92.3%, 88.9%, 84.6%,  
215 respectively at training, internal, and external validation). In the combined cohort, 112 of 126 patients  
216 with UPA and 30 of 32 with BPA, were correctly classified, resulting in a sensitivity and specificity  
217 of 88.8% and 93.4%, respectively (Figure 1B). Internal and external validation confirmed the high  
218 predictive performance, with a minimum overfitting effect (3.4-7.7%).

219 The RF classification algorithm (the first tree of the RF is reported in Figure 2A) correctly  
220 discriminated 145 of 158 patients (accuracy 91.7%) from the combined cohort, with a sensitivity and  
221 specificity of 92.1% and 90.3%, respectively (Figure 2B). Also in this case, the accuracy at external

222 and internal validation by 10-fold cross validation was high (accuracy 82.9% and 89.6%, respectively;  
223 overfitting effect 6.0-12.7%; see Supplementary Table S4). Ipsilateral imaging and CLR were the  
224 strongest predictors in the LDA model, whereas lowest potassium levels and aldosterone post-  
225 confirmatory test in the RF (Figures 1C and 2C). Estimate predictor importance is reported in  
226 Supplementary Table S5 for both LDA and RF models.

227

### 228 ***Prediction score and management of PA patients***

229 The 6 variables used in the LDA and RF models were used to develop the CLR score, a 19-points  
230 scoring system to discriminate patients with a diagnosis of UPA vs. BPA. The CLR score was  
231 developed in the training cohort and then tested by 10-fold cross validation, and in the validation and  
232 combined cohorts. Figure 3A show how each parameter was categorized and points assigned. The  
233 analysis of the ROC curve confirmed the reliable performance of CLR score, with an AUC of 0.971  
234 (95% CI 0.947-0.994; Figure 3B). In the training cohort, a CLR score greater than 11 displayed the  
235 higher accuracy (96.4%), with the correct classification of 84 of 88 UPA patients (sensitivity 95.5%),  
236 whereas a score equal or lower than 11 classified all the BPA patients (specificity 100%). By 10-fold  
237 cross validation, we confirmed a reliable accuracy (96.3%). The diagnostic performance of the CLR  
238 score at external validation remained very high, with an accuracy, sensitivity, and specificity of  
239 93.7%, 94.7% and 90.0%, respectively, higher than the machine learning algorithms (confusion  
240 matrix for training, validation, and combined cohorts are reported in Supplementary Tables S4 and  
241 S6). A cut-off of greater than 8 optimized sensitivity, correctly classified 86 of 88, and 37 of 38  
242 patients with UPA, respectively at training and at validation, with a sensitivity of 97.8% and 97.4%.  
243 A cut-off lower than 13 maximized specificity, and correctly identified all patients with a diagnosis  
244 of BPA, both at training and at validation (specificity 100%).  
245 Using 11 as cut-off, patients which underwent AVS under basal conditions were correctly  
246 discriminated in 94.7% of cases, whereas after stimulation with ACTH in 96.3% of cases.

247 Figure 4A shows the stratification of patients with a diagnosis of UPA or BPA for the CLR score.  
248 The scoring system was directly correlated with the proportion of patients with a diagnosis of UPA.  
249 In addition, all patients with a score greater than 13 had UPA (N = 88), whereas all patients with a  
250 score equal or lower than 3 had BPA (N = 6; Supplementary Table S7).

251 Overall, 102 of the 126 patients with UPA underwent adrenalectomy. According to the PASO  
252 criteria<sup>5</sup>, at follow-up of 6-12 months after surgery, patients with UPA displayed complete clinical  
253 and biochemical success in 50% and 96.1% of cases, respectively (Supplementary Table S8). All the  
254 6 patients with UPA misclassified as BPA displayed partial or absent clinical outcome (4 with a  
255 complete and 2 with a partial biochemical outcome). Forty three percent (41 of 96) of correctly  
256 predicted patients with UPA displayed a partial clinical outcome and 4% absent clinical outcome;  
257 biochemical outcome was complete in 94 of 96 correctly classified patients and partial in the  
258 remaining 2 patients. No statistically significant differences were observed in the clinical and  
259 biochemical outcomes for patients who underwent unstimulated *vs.* ACTH stimulated AVS  
260 (Supplementary Table 8).

261 The CLR score was finally integrated into a flow chart for the management of patients with PA with  
262 a unilateral successful AVS and contralateral suppression (CLR < 1; Figure 4B and 4C). Patients with  
263 a score equal to or greater than 13 were classified as “probable UPA” with indication to surgical  
264 intervention. Notably, all patients with UPA identified in this way (N = 104) correctly underwent  
265 unilateral adrenalectomy. Patients with a score equal to or lower than 8 were classified as “probable  
266 BPA” and treated with MRA (N = 18), resulting in 15 patients with BPA correctly managed and 3  
267 UPA patients, who missed the possibility to be operated. All the other patients (N = 36), with a score  
268 ranging between 8.5 and 12.5 should repeat AVS. The application of the prediction score in our  
269 clinical context would result in the correct management of 155 of 158 patients (accuracy 98.1%),  
270 allowing a subtype diagnosis with the repetition of only 22.8% of AVS procedures (sensitivity 97.6%  
271 and specificity 100.0%).

272

273 **DISCUSSION**

274 We propose two prediction models based on supervised machine learning algorithms, and an easily  
275 applicable scoring system, to discriminate UPA from BPA. An online free-downloadable tool is  
276 provided and allows the application of a flow-chart for the management of patients undergoing  
277 unilateral successful AVS in the presence of contralateral suppression.

278 Although the rate of bilaterally successful cannulation during AVS may be higher than 90% when  
279 performed by an experienced radiologist<sup>19</sup>, several centers do not have a comparable reliability, with  
280 a successful rate ranging between 8.0 and 30.5%, especially where a low number of AVS are  
281 performed per year<sup>20-22</sup>. Given its course and anatomical variability<sup>23</sup>, the main reason for  
282 unsuccessful AVS is the failed cannulation of RAV because of difficulties in identification of the  
283 vessel. Therefore, over the last few years, some studies have tried to develop indices to predict  
284 subtype diagnosis from incomplete AVS data<sup>24-28</sup>.

285 Kline et al. demonstrated a high performance of the CLR in the identification of UPA; sensitivity and  
286 specificity were 90% and 94%, respectively, using a CLR < 1.4 as cut-off<sup>24</sup>. However, the authors  
287 did not validate their approach in a second cohort of patients and observed a high accuracy only in  
288 patients who underwent cosyntropin-stimulated AVS, whereas in unstimulated procedures the  
289 accuracy was lower. Many patients with bilateral disease may be misclassified using a CLR cut-off  
290 of 1.4 that includes subjects without contralateral suppression of aldosterone secretion. On the other  
291 hand, Lin et al. proposed very low cut-offs (CLR < 0.07 for left and 0.08 for right lateralisation) in a  
292 cohort of 160 patients who underwent AVS in basal conditions. Specificity was 100%, but sensitivity  
293 was unacceptable for clinical application, ranging between 27.5 and 40%, thus missing the majority  
294 of patients with a diagnosis of UPA<sup>28</sup>. In a subsequent study, Durivage et al. applied multinomial  
295 regression modelling in patients with unsuccessful right adrenal vein cannulation. The authors  
296 correctly identified 75.1% and 64.1% of UPA patients in cosyntropin-stimulated and unstimulated

297 AVS, respectively<sup>26</sup>. In another study performed in 62 AVS procedures<sup>25</sup>, a CLR equal or lower than  
298 0.5 displayed a specificity and positive predictive value of 100% but missed about half of patients  
299 with UPA (sensitivity 58%). Strajina et al. tested the same cut-off in a cohort of 150 patients; the  
300 specificity was 100%, but 20% of patients with UPA were classified as BPA and missed the  
301 possibility to be cured by adrenalectomy<sup>27</sup>.

302

303 In our cohort, 79.7% of patients were diagnosed as UPA, thus being representative of the prevalence  
304 of unilateral disease in patients with contralateral suppression<sup>24-28</sup>. With our machine learning-based  
305 models and with the CLR score, we obtained a higher diagnostic performance compared to previous  
306 methods. We achieved a 97.6% sensitivity and 100% specificity with the proposed flow-chart for  
307 patient management. Combining the cut-offs which optimized sensitivity and specificity in our flow-  
308 chart, all patients predicted as UPA by the CLR score were correctly diagnosed (positive predictive  
309 value 100%), whereas only 3 patients with a diagnosis of UPA were misclassified as BPA and would  
310 have missed the possibility of surgical treatment but would have a long-term pharmacological  
311 treatment. The cut-off with the maximal specificity of the CLR score displayed a sensitivity higher  
312 than 80% both at training and validation, outperforming previous scores. The accuracy was similarly  
313 very high, independent of the AVS protocol (with or without cosyntropin stimulation).

314 In our study, 20.3% of patients had BPA and displayed a CLR lower than 1, thus confirming a high  
315 prevalence of contralateral suppression in patients with bilateral disease<sup>11</sup>. Using a CLR equal or  
316 lower than 0.5<sup>25,27</sup>, 8 of 32 patients with BPA (25%) still received a diagnosis of UPA and would be  
317 operated inappropriately. Indeed, a large multicenter study demonstrated that contralateral  
318 suppression was not associated with a better clinical and biochemical outcome after surgery<sup>13</sup>,  
319 whereas the studies evaluating the diagnostic performance of CLR did not assess post-surgical  
320 outcomes<sup>24,25,27,28</sup>. With our flow-chart, only 3 UPA patients were classified as BPA would have  
321 missed the possibility to be surgically treated. When assessed with the PASO criteria<sup>5</sup>, these 3 patients

322 displayed partial clinical success, without normalization of aldosterone levels in 2 cases (partial  
323 biochemical success).

324

325 Contralateral suppression alone is not considered a reliable parameter to recommend  
326 adrenalectomy<sup>10</sup>. Similarly, CT scanning and/or clinical and biochemical parameters do not allow to  
327 predict subtype diagnosis in patients with PA<sup>7,29</sup>.

328 Combining CLR with CT scanning findings and potassium and aldosterone levels, we considerably  
329 improved the diagnostic performance described for the CLR alone<sup>24-28</sup>. To evaluate the net benefit of  
330 our clinical score, compared to CLR alone, we assessed the respective diagnostic performance by  
331 ROC curve analysis. The AUC for CLR was 0.836 (CI 95% 0.759-0.913), compared with 0.971 for  
332 the clinical score (+16.1%). A CLR < 0.45 displayed an accuracy of 84.8% in the discrimination of  
333 patients with UPA from those with BPA. In the combined cohort, the score, LDA and RF models  
334 reached an accuracy of 95.5%, 90.0%, and 91.7% with a net benefit ranging between 6.1% and 12.6%.  
335 Our score **could be** of interest for centers with a low rate of bilateral adrenal vein cannulation, since  
336 it correctly classifies 98.1% of patients that display contralateral suppression and avoid the repetition  
337 of AVS in 77.2% of the cases. This approach **would** allow to recommend adrenalectomy (100%  
338 positive predictive value and specificity), without the repetition of the AVS in case of unsuccessful  
339 right adrenal vein cannulation.

340

341 There are some limitations to this study. This was a retrospective analysis of patients referred to a  
342 single center: **we validated our scoring system in a separate randomly selected population, but the**  
343 **algorithm should be validated in independent prospective cohorts from other centers, to assess his**  
344 **generalizability and exclude selection bias**. The strengths of our study include the high accuracy of  
345 the proposed prediction models, integrating machine learning algorithms and a scoring-system,  
346 allowing the correct management of patients with unilateral successful AVS and contralateral

347 suppression. Moreover, we developed an online tool to calculate the CLR score automatically,  
348 allowing the easy application of our flow-chart in clinical practice.

349 In conclusion, AVS remains a technically demanding and invasive procedure but still essential to  
350 distinguish UPA from BPA. Our findings suggest that integrating clinical and biochemical  
351 parameters, CT scanning imaging, and CLR, partial AVS data could be used to define the subtype  
352 diagnosis, facilitating surgical decision-making in case of missed cannulation of one adrenal vein.

353

#### 354 **Declarations**

355 **Conflict(s) of Interest/Disclosure(s):** the authors declare no conflict of interest that could be  
356 perceived as prejudicing the impartiality of the research reported.

357 **Source(s) of Funding:** This research did not receive any specific grant from any funding agency in  
358 the public, commercial or not-for-profit sector.

359 **REFERENCES**

- 360 1. Funder JW, Carey RM, Mantero F, Murad MH, Reincke M, Shibata H, Stowasser M, Young WF  
361 Jr. The Management of Primary Aldosteronism: Case Detection, Diagnosis, and Treatment: An  
362 ES Clinical Practice Guideline. *Journal of Clinical Endocrinology and Metabolism* 2016 **101**  
363 1889-1916.
- 364 2. Mulatero P, Monticone S, Deinum J, Amar L, Prejbisz A, Zennaro MC, Beuschlein F, Rossi GP,  
365 Nishikawa T, Morganti A, et al. Genetics, prevalence, screening and confirmation of primary  
366 aldosteronism: a position statement and consensus of the Working Group on Endocrine  
367 Hypertension of The European Society of Hypertension. *J Hypertens* 2020 [Epub ahead of print].
- 368 3. Mulatero P, Sechi LA, Williams TA, Lenders JWM, Reincke M, Satoh F, Januszewicz A, Naruse  
369 M, Doumas M, Veglio F, et al. Subtype diagnosis, treatment, complications and outcomes of  
370 primary aldosteronism and future direction of research: a position statement and consensus of  
371 the Working Group on Endocrine Hypertension of the European Society of Hypertension. *J*  
372 *Hypertens* 2020 [Epub ahead of print].
- 373 4. Monticone S, D'Ascenzo F, Moretti C, Williams TA, Veglio F, Gaita F, Mulatero P.  
374 Cardiovascular events and target organ damage in primary aldosteronism compared with  
375 essential hypertension: a systematic review and meta-analysis. *The Lancet Diabetes &*  
376 *Endocrinology* 2018 **6** 41-50.
- 377 5. Williams TA, Lenders JWM, Mulatero P, Burrello J, Rottenkolber M, Adolf C, Satoh F, Amar  
378 L, Quinkler M, Deinum J, et al. Outcomes after adrenalectomy for unilateral primary  
379 aldosteronism: an international consensus on outcome measures and analysis of remission rates  
380 in an international cohort. *The Lancet Diabetes & Endocrinology* 2017 **5** 689-699.
- 381 6. Williams TA, Burrello J, Sechi LA, Fardella CE, Matrozoza J, Adolf C, Baudrand R, Bernardi  
382 S, Beuschlein F, Catena C, et al. Computed Tomography and Adrenal Venous Sampling in the  
383 Diagnosis of Unilateral Primary Aldosteronism. *Hypertension* 2018 **72** 641-649.

- 384 7. Takeda M, Yamamoto K, Akasaka H, Rakugi H, Naruse M, Takeda Y, Kurihara I, Itoh H,  
385 Umakoshi H, Tsuiki M, et al., JPAS Study Group. Clinical characteristics and postoperative  
386 outcomes of primary aldosteronism in the elderly. *Journal of Clinical Endocrinology and*  
387 *Metabolism* 2018 **103** 3620-3629.
- 388 8. Hundemer GL, Curhan GC, Yozamp N, Wang M, Vaidya A. Cardiometabolic outcomes and  
389 mortality in medically treated primary aldosteronism: a retrospective cohort study. *The Lancet*  
390 *Diabetes & Endocrinology* 2018 **6** 51-59.
- 391 9. Daunt N. Adrenal vein sampling: how to make it quick, easy, and successful. *Radiographics*  
392 2005 **25** S143-S158.
- 393 10. Monticone S, Viola A, Rossato D, Veglio F, Reincke M, Gomez-Sanchez C, Mulatero P. Adrenal  
394 vein sampling in primary aldosteronism: towards a standardised protocol. *The Lancet Diabetes*  
395 *& Endocrinology* 2015 **3** 296-303.
- 396 11. Young WF, Stanson AW, Thompson GB, Grant CS, Farley DR, van Heerden JA. Role for  
397 adrenal venous sampling in primary aldosteronism. *Surgery* 2004 **136** 1227-1235.
- 398 12. Wolley MJ, Gordon RD, Ahmed AH, Stowasser M. Does contralateral suppression at adrenal  
399 venous sampling predict outcome following unilateral adrenalectomy for primary aldosteronism?  
400 A retrospective study. *Journal of Clinical Endocrinology and Metabolism* 2015 **100** 1477-1484.
- 401 13. Monticone S, Satoh F, Viola A, Fischer E, Vonend O, Bernini G, Lucatello B, Quinkler M,  
402 Ronconi V, Morimoto R, et al. Aldosterone suppression on contralateral adrenal during adrenal  
403 vein sampling does not predict blood pressure response after adrenalectomy. *Journal of Clinical*  
404 *Endocrinology and Metabolism* 2014 **99** 4158-4166.
- 405 14. Espiner EA, Ross DG, Yandle TG, Richards AM, Hunt PJ. Predicting surgically remedial  
406 primary aldosteronism: role of adrenal scanning, posture testing, and adrenal vein sampling.  
407 *Journal of Clinical Endocrinology and Metabolism* 2003 **88** 3637-3644.

- 408 15. Auchus RJ, Wians FH Jr, Anderson ME, Dolmatch BL, Trimmer CK, Josephs SC, Chan D,  
409 Toomay S, Nwariaku FE. What we still do not know about adrenal vein sampling for primary  
410 aldosteronism. *Hormone and Metabolic Research* 2010 **42** 411-415.
- 411 16. Shibayama Y, Wada N, Naruse M, Kurihara I, Ito H, Yoneda T, Takeda Y, Umakoshi H, Tsuiki  
412 M, Ichijo T, et al. The Occurrence of Apparent Bilateral Aldosterone Suppression in Adrenal  
413 Vein Sampling for Primary Aldosteronism. *Journal of the Endocrine Society* 2018 **2** 398-407.
- 414 17. Burrello J, Burrello A, Stowasser M, Nishikawa T, Quinkler M, Prejbisz A, Lenders JWM, Satoh  
415 F, Mulatero P, Reincke M, et al. The Primary Aldosteronism Surgical Outcome Score for the  
416 Prediction of Clinical Outcomes After Adrenalectomy for Unilateral Primary Aldosteronism.  
417 *Annals of Surgery* 2019 [Epub ahead of print].
- 418 18. Roe KD, Jawa V, Zhang X, Chute CG, Epstein JA, Matelsky J, Shpitser I, Taylor CO. Feature  
419 engineering with clinical expert knowledge: A case study assessment of machine learning model  
420 complexity and performance. *PLoS One* 2020 **15** e0231300.
- 421 19. Young WF, Stanson AW. What are the keys to successful adrenal venous sampling (AVS) in  
422 patients with primary aldosteronism? *Clinical Endocrinology* 2009 **70** 14-17.
- 423 20. Vonend O, Ockenfels N, Gao X, Allolio B, Lang K, Mai K, Quack I, Saleh A, Degenhart C,  
424 Seufert J, et al. Adrenal venous sampling: evaluation of the German Conn's registry.  
425 *Hypertension* 2011 **57** 990-995.
- 426 21. Elliott P, Holmes DT. Adrenal vein sampling: substantial need for technical improvement at  
427 regional referral centres. *Clinical Biochemistry* 2013 **46** 1399-1404.
- 428 22. Harsha A, Trerotola SO. Technical aspects of adrenal vein sampling. *Journal of Vascular and*  
429 *Interventional Radiology* 2015 **26** 239.
- 430 23. Matsuura T, Takase K, Ota H, Yamada T, Sato A, Satoh F, Takahashi S. Radiologic anatomy of  
431 the right adrenal vein: preliminary experience with MDCT. *AJR American journal of*  
432 *roentgenology* 2008 **191** 402-408.

- 433 24. Kline GA, Chin A, So B, Harvey A, Pasiaka JL. Defining contralateral adrenal suppression in  
434 primary aldosteronism: implications for diagnosis and outcome. *Clinical Endocrinology* 2015 **83**  
435 20-27.
- 436 25. Pasternak JD, Epelboym I, Seiser N, Wingo M, Herman M, Cowan V, Gosnell JE, Shen WT,  
437 Kerlan RK Jr, Lee JA, et al. Diagnostic utility of data from adrenal venous sampling for primary  
438 aldosteronism despite failed cannulation of the right adrenal vein. *Surgery* 2016 **159** 267-273.
- 439 26. Durivage C, Blanchette R, Soulez G, Chagnon M, Gilbert P, Giroux MF, Bourdeau I, Oliva VL,  
440 Lacroix A, Therasse E. Adrenal venous sampling in primary aldosteronism: multinomial  
441 regression modeling to detect aldosterone secretion lateralization when right adrenal sampling is  
442 missing. *Journal of Hypertension* 2017 **35** 362-368.
- 443 27. Strajina V, Al-Hilli Z, Andrews JC, Bancos I, Thompson GB, Farley DR, Lyden ML, Dy BM,  
444 Young WF, McKenzie TJ. Primary aldosteronism: making sense of partial data sets from failed  
445 adrenal venous sampling-suppression of adrenal aldosterone production can be used in clinical  
446 decision making. *Surgery* 2018 **163** 801-806.
- 447 28. Lin L, Zhou L, Guo Y, Liu Z, Chen T, Liu Z, Wang K, Li J, Zhu Y, Ren Y. Can incomplete  
448 adrenal venous sampling data be used in predicting the subtype of primary aldosteronism?  
449 *Annales Endocrinology* 2019 **80** 301-307.
- 450 29. K Kempers MJ, Lenders JW, van Outheusden L, van der Wilt GJ, Schultze Kool LJ, Hermus  
451 AR, Deinum J. Systematic review: diagnostic procedures to differentiate unilateral from bilateral  
452 adrenal abnormality in primary aldosteronism. *Annals of Internal Medicine* 2009 **151** 329-337.

453 **FIGURE LEGENDS**

454 *Legend to Figure 1 – Diagnostic Modelling: Linear Discriminant Analysis.* The LDA model  
455 included the 6 variables with the highest prediction power for subtype diagnosis in the training  
456 cohort (N = 110). Contralateral adrenal gland corresponds to the adrenal with the A/C ratio lower  
457 than the A/C ratio in the inferior vena cava, while ipsilateral gland corresponds to the opposite side.  
458 When unilateral PA (UPA) is predicted, the ipsilateral adrenal gland should be removed. **Panel A**,  
459 canonical plot representing diagnostic performance of LDA; each patient is indicated by a point and  
460 subtype diagnosis are reported by colour (UPA, black; BPA, bilateral PA, grey). The axes (canonical  
461 component 1 and 2) are calculated by weighted linear combination of the 6 variables included in the  
462 model to maximize the separation between groups. The crosses indicate the means of (canonical 1;  
463 canonical 2) for patients with UPA or BPA, the ellipse included patients with a linear combination  
464 coefficient that falls within the mean  $\pm$  SD. **Panel B**, confusion matrix reporting real and predicted  
465 diagnosis, accuracy, sensitivity, specificity in the combined cohort, and 10-fold cross validation.  
466 **Panel C**, histogram representing normalized LDA coefficients for each variable included in the  
467 model (see also Supplementary Table S5).

468  
469 *Legend to Figure 2 – Diagnostic Modelling: Random Forest.* The RF algorithm included the 6  
470 variables with the highest classification power for subtype diagnosis in the training cohort (N =  
471 110). Contralateral adrenal gland corresponds to the adrenal with the A/C ratio lower than the A/C  
472 ratio in the inferior vena cava, while ipsilateral gland corresponds to the opposite side. When  
473 unilateral PA (UPA) is predicted, the ipsilateral adrenal gland should be removed. **Panel A**, the first  
474 classification tree of the forest is shown for the prediction of UPA vs. BPA (bilateral PA). **Panel B**,  
475 confusion matrix reporting real and predicted diagnosis, accuracy, sensitivity, specificity in the  
476 combined cohort, and 10-fold cross validation. **Panel C**, histogram representing normalized  
477 predictive coefficients for each variable included in the model (see also Supplementary Table S5).

478

479 *Legend to Figure 3 – Score development.* Univariate/multivariate regression analyses and  
480 coefficients from the LDA and RF models were used to assign points to each variable according to  
481 stratification level. The score was developed in the training cohort (N = 110) and tested on the  
482 validation cohort (N = 48). Contralateral adrenal gland corresponds to the adrenal with the A/C ratio  
483 lower than the A/C ratio in the inferior vena cava, while ipsilateral gland corresponds to the opposite  
484 side. When unilateral PA (UPA) is predicted, the ipsilateral adrenal gland should be removed. **Panel**  
485 **A**, table showing included variables and scoring-point system. **Panel B**, receiver operating  
486 characteristics (ROC) curve to assess AUC (area under the curve) and the best cut-off for the score  
487 in the combined cohort (N = 158).

488

489 *Legend to Figure 4 – Score performance and management of PA patients.* Flow chart for the  
490 management of patients with PA, unilateral successful AVS, and CLR < 1. Contralateral adrenal  
491 gland corresponds to the adrenal with the A/C ratio lower than the A/C ratio in the inferior vena  
492 cava, while ipsilateral gland corresponds to the opposite side. When unilateral PA (UPA) is  
493 predicted, the ipsilateral adrenal gland should be removed. **Panel A**, histogram showing the  
494 proportion of patients (y-axis, %) for each subtype diagnosis (UPA, black; BPA, bilateral PA, grey),  
495 stratified by score points (x-axis) on the combined cohort. The total number of patients (N) for each  
496 AVS score level and their proportion (%) are reported in Table S5. **Panel B**, PA patient management  
497 using our score; the number of patients is indicated in bold; cut-offs and misclassified patients are  
498 indicated in grey. **Panel C**, confusion matrix representing real and predicted subtype diagnosis,  
499 sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). CLR,  
500 Contralateral Ratio; AVS, Adrenal Venous Sampling; MRA, Mineral Receptor Antagonist.

501

502 **TABLE LEGENDS**

503 *Legend to Table 1* – Characteristics of patients included in the analysis stratified for diagnosis:  
504 unilateral PA (UPA; N=126) vs. bilateral PA (BPA; N=32). Antihypertensive medication is  
505 expressed as Defined Daily Dose (DDD). DDD is the average maintenance dose per day for a drug  
506 used for its main indication in adults. HTN, Hypertension; BP, Blood Pressure; PRA, Plasma Renin  
507 Activity; LVH, Left Ventricular Hypertrophy; Echo, Echocardiography; CV, Cardiovascular; CT,  
508 Computed Tomography; SI, Selectivity Index; A/C, Aldosterone to Cortisol ratio. Normally and  
509 non-normally distributed variables were reported as mean  $\pm$  standard deviation or median  
510 [interquartile range], respectively. Categorical variables were reported as absolute number (n) and  
511 proportion (%).

512  
513 *Legend to Table 2* – Logistic regression analysis was performed to assess the odds ratio (OR),  $\beta$   
514 estimate and the 95% confidence interval (CI) for each variable. Univariate and multivariate analysis  
515 are shown as indicated. An OR greater than 1 indicates an increased likelihood of unilateral PA  
516 (UPA), and an OR less than 1 a decreased likelihood (i.e. an OR of 1.01 for aldosterone means an  
517 increase of 1% of the likelihood of UPA for each 1 ng/dL increase of aldosterone concentration).  
518 Aldosterone at screening, lowest potassium, aldosterone post-confirmatory test, and contralateral  
519 ratio were treated as continuous variables; ipsilateral and contralateral imaging (defined as normal  
520 in absence of nodules, or with thickening  $<$  4 mm) were treated as categorical variables.

**Table 1. Patient Characteristics of Study Cohort**

Variable	UPA (N = 126)	BPA (N = 32)	P-value
Female sex, n (%)	53 (42.1)	7 (21.9)	<b>0.036</b>
Age at diagnosis (years)	49 ± 10.3	55 ± 7.1	<b>0.001</b>
Duration of HTN (months)	66 [22; 156]	75 [44; 124]	0.721
Systolic BP (mmHg)	164 ± 25.7	164 ± 20.3	0.934
Diastolic BP (mmHg)	97 ± 14.3	98 ± 12.7	0.824
Antihypertensive medication (DDD)	4.0 [2.5; 6.0]	2.9 [1.4; 4.3]	<b>0.008</b>
PRA at screening (ng/mL/h)	0.3 [0.20; 0.40]	0.20 [0.10; 0.22]	<b>0.014</b>
Aldosterone at screening (ng/dL)	38.2 [25.9; 49.8]	30.2 [21.1; 41.0]	<b>0.044</b>
Lowest potassium (mEq/L)	3.1 ± 0.6	3.9 ± 0.4	<b>&lt; 0.001</b>
Aldosterone post-confirmatory test (ng/dL)	21.9 [13.9; 35.5]	11.4 [7.7; 19.9]	<b>&lt; 0.001</b>
Microalbuminuria, n (%)	37 (29.6)	6 (19.0)	0.333
LVH at Echo, n (%)	70 (55.2)	24 (76.0)	0.058
CV events, n (%)	17 (13.6)	6 (18.5)	0.519
Largest nodule at CT scanning (diameter, mm)	13 [10; 18]	11 [9; 20]	0.433
Normal ipsilateral imaging, n (%)	2 (1.6)	14 (43.8)	<b>&lt;0.001</b>
Normal contralateral imaging, n (%)	107 (84.9)	17 (53.1)	<b>&lt;0.001</b>
Peripheral A/C ratio	15.9 [3.7; 26.2]	13.3 [7.2; 18.1]	0.145
Contralateral A/C ratio	4.1 [1.5; 8.3]	7.9 [3.2; 12.0]	<b>0.012</b>
Contralateral ratio	0.3 [0.2; 0.5]	0.7 [0.5; 0.8]	<b>&lt; 0.001</b>

*Legend to Table 1* – Characteristics of patients included in the analysis stratified for diagnosis:

unilateral PA (UPA; N=126) vs. bilateral PA (BPA; N=32). Antihypertensive medication is expressed as Defined Daily Dose (DDD). DDD is the average maintenance dose per day for a drug used for its main indication in adults. HTN, Hypertension; BP, Blood Pressure; PRA, Plasma Renin Activity; LVH, Left Ventricular Hypertrophy; Echo, Echocardiography; CV, Cardiovascular; CT, Computed Tomography; SI, Selectivity Index; A/C, Aldosterone to Cortisol ratio. Normally and non-normally distributed variables were reported as mean ± standard deviation or median [interquartile range], respectively. Categorical variables were reported as absolute number (n) and proportion (%).

**Table 2. Selected discriminant variables for a diagnosis of unilateral PA**

Variable (ref. BPA)	Univariate analysis			Multivariate Analysis		
	OR (CI 95%)	$\beta$	<i>P</i>	OR (CI 95%)	$\beta$	<i>P</i>
Aldosterone at screening (ng/dL)	1.01 (1.01 – 1.01)	0.003	<b>0.046</b>	1.01 (1.01-1.03)	0.001	<b>0.032</b>
Lowest potassium (mEq/L)	0.07 (0.03 – 0.20)	-2.639	<b>&lt; 0.001</b>	0.04 (0.01 – 0.26)	-3.139	<b>0.001</b>
Aldosterone post-confirmatory test (ng/dL)	1.01 (1.01 – 1.02)	0.006	<b>0.003</b>	1.01 (1.01 – 1.02)	0.004	<b>0.038</b>
Normal ipsilateral imaging	0.02 (0.01 – 0.10)	-3.876	<b>&lt; 0.001</b>	0.02 (0.01 – 0.19)	-4.120	<b>0.001</b>
Normal contralateral imaging	4.97 (2.13 – 11.61)	1.603	<b>&lt; 0.001</b>	5.35 (1.02 – 33.11)	1.678	<b>0.041</b>
Contralateral ratio	0.01 (0.01 – 0.03)	-5.710	<b>&lt; 0.001</b>	0.01 (0.01 – 0.08)	-6.532	<b>0.001</b>

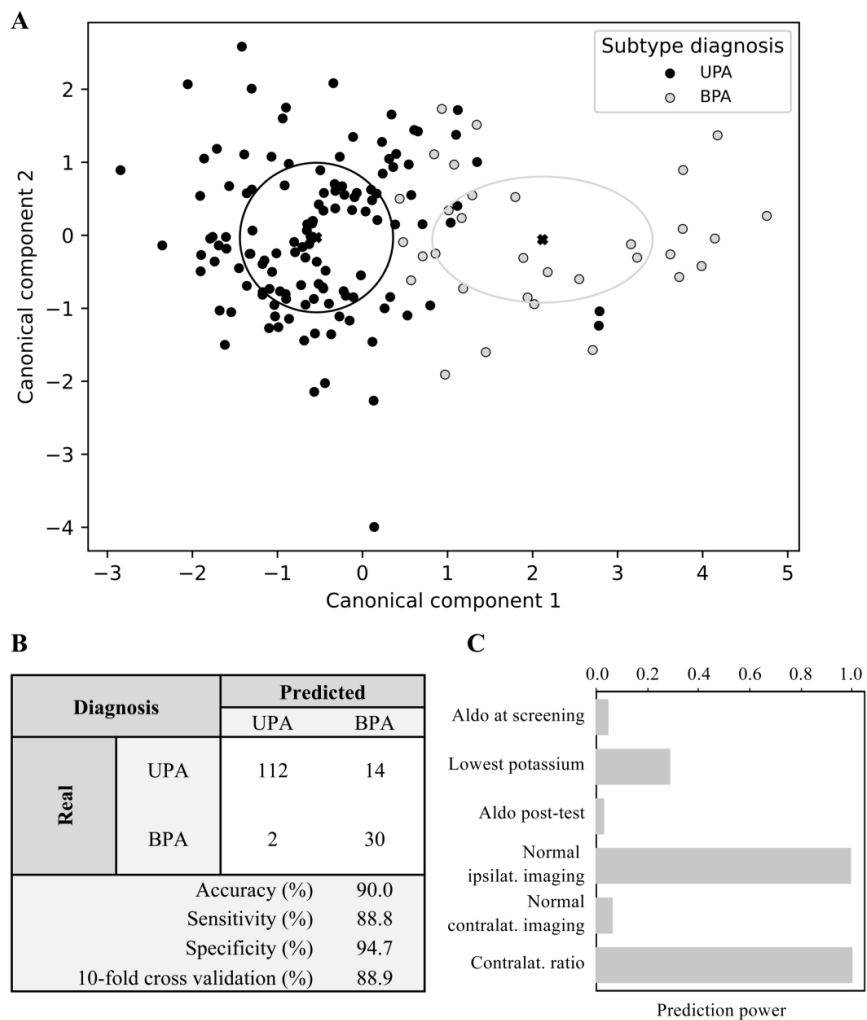


Figure 1. Diagnostic Modelling: Linear Discriminant Analysis.

126x141mm (600 x 600 DPI)

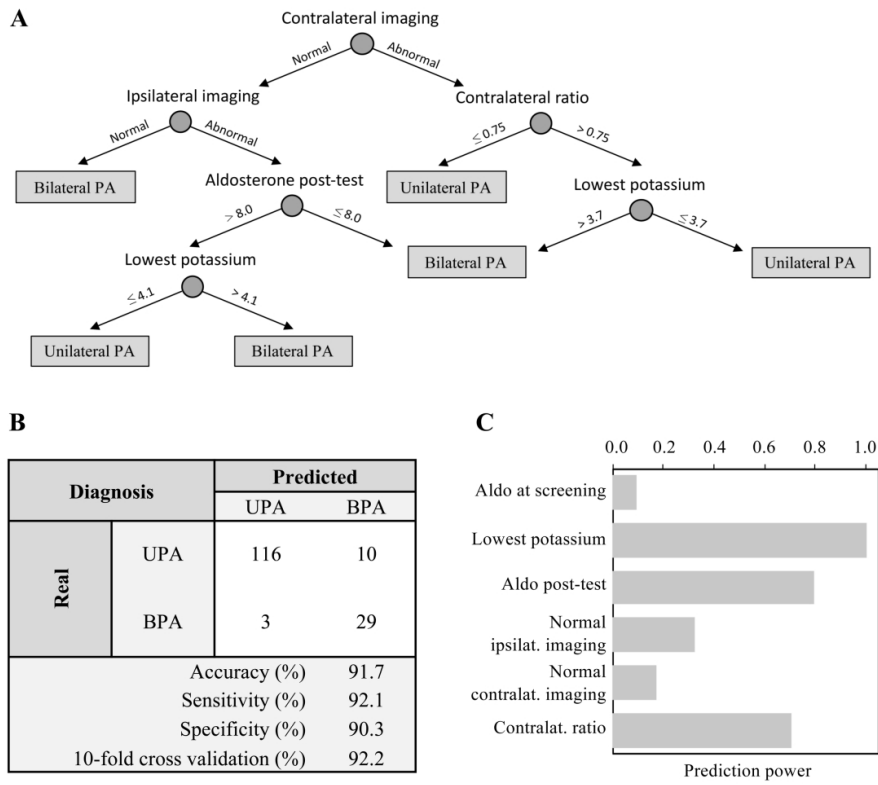


Figure 2. Diagnostic Modelling: Random Forest

127x110mm (600 x 600 DPI)

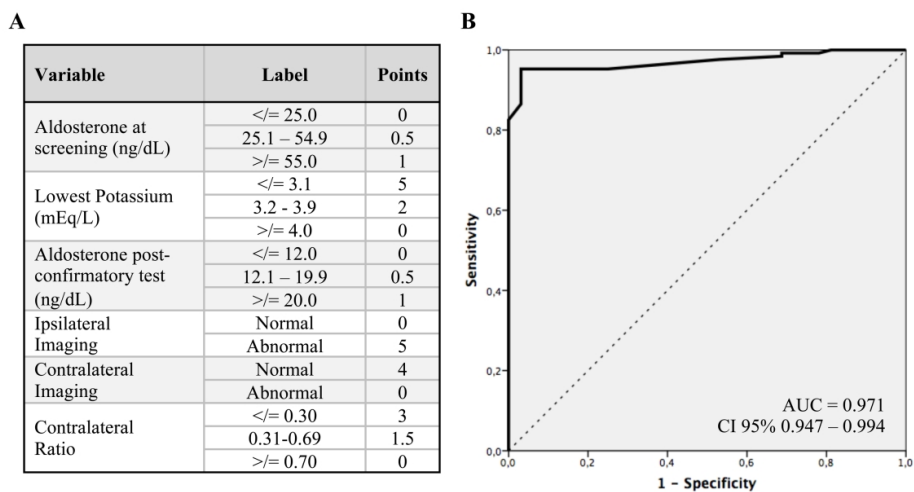


Figure 3. Score development

139x75mm (600 x 600 DPI)

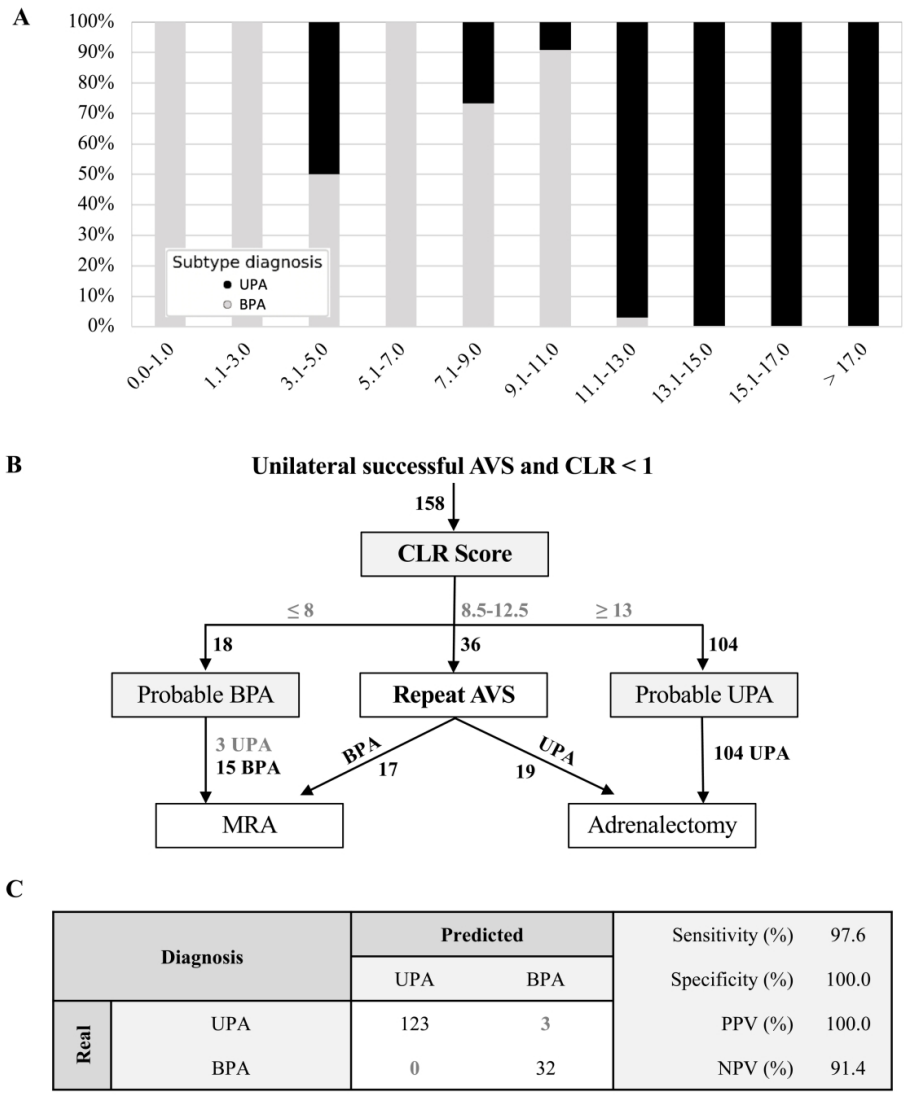


Figure 4. Score performance and management of PA patients

131x155mm (600 x 600 DPI)

**Supplementary Table 1. Patient Characteristics of Study Cohort**

Variable	Basal (N = 76)	ACTH (N = 82)	P-value
Diagnosis of UPA, n (%)	57 (75.0)	69 (84.1)	0.153
Female sex, n (%)	27 (35.5)	33 (40.2)	0.542
Age at diagnosis (years)	51 ± 9.3	49 ± 10.5	0.156
Duration of HTN (months)	69 [24; 176]	66 [22; 125]	0.342
Systolic BP (mmHg)	163 ± 25.2	164 ± 23.8	0.821
Diastolic BP (mmHg)	98 ± 13.0	97 ± 15.0	0.756
Antihypertensive medication (DDD)	4.0 [2.4; 5.7]	3.7 [2.2; 5.0]	0.671
PRA at screening (ng/mL/h)	0.20 [0.10; 0.30]	0.20 [0.15; 0.40]	0.656
Aldosterone at screening (ng/dL)	38.9 [25.8; 50.5]	34.8 [24.3; 45.3]	0.198
Lowest potassium (mEq/L)	3.3 ± 0.7	3.1 ± 0.7	0.085
Aldosterone post-confirmatory test (ng/dL)	19.6 [11.9; 32.7]	19.6 [10.8; 32.0]	0.738
Microalbuminuria, n (%)	23 (30.4)	21 (25.0)	0.541
LVH at Echo, n (%)	49 (65.1)	44 (53.7)	0.188
CV events, n (%)	10 (12.7)	13 (16.4)	0.549
Largest nodule at CT scanning (diameter, mm)	12 [10; 16]	14 [10; 20]	0.189
Normal ipsilateral imaging, n (%)	8 (10.5)	8 (9.8)	0.873
Normal contralateral imaging, n (%)	62 (81.6)	62 (75.6)	0.362
Contralateral SI	14.6 [7.5; 33.3]	23.0 [13.9; 48.6]	<b>0.002</b>
Peripheral A/C ratio	17.8 [3.4; 28.7]	14.4 [4.1; 19.7]	0.348
Contralateral A/C ratio	4.4 [1.2; 9.9]	4.9 [1.6; 8.6]	0.867
Contralateral ratio	0.3 [0.2; 0.6]	0.4 [0.3; 0.6]	0.274

Clinical characteristics of patients included in the analysis: patients were stratified according to the protocol of adrenal venous sampling (basal vs. ACTH). Antihypertensive medication is expressed as Defined Daily Dose (DDD). DDD is the average maintenance dose per day for a drug used for its main indication in adults. UPA, Unilateral Primary Aldosteronism; HTN, Hypertension; BP, Blood Pressure; PRA, Plasma Renin Activity; LVH, Left Ventricular Hypertrophy; Echo, Echocardiography; CV, Cardiovascular; CT, Computed Tomography; SI, Selectivity Index; A/C, Aldosterone to Cortisol ratio. Normally and non-normally distributed variables were reported as mean ± standard deviation or median [interquartile range], respectively. Categorical variables were reported as absolute number (n) and proportion (%).

**Supplementary Table 2. Patient Characteristics of Study Cohort**

Variable (ref. <b>BPA</b> )	OR (CI 95%)	$\beta$	P-value
Female sex, n (%)	2.59 (1.04 – 6.44)	0.943	<b>0.040</b>
Age at diagnosis (years)	0.94 (0.90 – 0.98)	-0.065	<b>0.006</b>
Duration of HTN (months)	1.01 (0.99 – 1.01)	0.002	0.236
Systolic BP (mmHg)	1.00 (0.98 – 1.02)	-0.001	0.933
Diastolic BP (mmHg)	1.00 (0.96 – 1.03)	-0.004	0.822
Antihypertensive medication (DDD)	1.43 (1.09 – 1.87)	0.357	<b>0.009</b>
PRA at screening (ng/mL/h)	4.40 (0.56 – 34.43)	1.480	0.159
Aldosterone at screening (ng/dL)	1.01 (1.01 – 1.01)	0.003	<b>0.046</b>
Lowest potassium (mEq/L)	0.07 (0.03 – 0.20)	-2.639	<b>&lt; 0.001</b>
Aldosterone post-confirmatory test (ng/dL)	1.01 (1.01 – 1.02)	0.006	<b>0.003</b>
Microalbuminuria, n (%)	1.79 (0.54 – 5.88)	0.582	0.337
LVH at Echo, n (%)	0.39 (0.14 – 1.05)	-0.942	0.063
CV events, n (%)	0.69 (0.23 – 2.13)	-0.368	0.521
Largest nodule at CT scanning (diameter, mm)	0.99 (0.92 – 1.06)	-0.015	0.683
Normal ipsilateral imaging, n (%)	0.02 (0.01 – 0.10)	-3.876	<b>&lt; 0.001</b>
Normal contralateral imaging, n (%)	4.97 (2.13 – 11.61)	1.603	<b>&lt; 0.001</b>
Peripheral A/C ratio	1.03 (1.00 – 1.06)	0.028	0.071
Contralateral A/C ratio	0.91 (0.85 – 0.97)	-0.093	<b>0.007</b>
Contralateral ratio	0.01 (0.01 – 0.03)	-5.710	<b>&lt; 0.001</b>

Univariate logistic regression analysis was performed to assess the odds ratio (OR),  $\beta$  estimate and the 95% confidence interval (CI) for each variable. An OR greater than 1 indicates an increased likelihood of unilateral primary aldosteronism (UPA), and an OR less than 1 a decreased likelihood (i.e. an OR of 1.01 for aldosterone means an increase of 1% of the likelihood of UPA for each 1 ng/dL increase of aldosterone concentration). Antihypertensive medication is expressed as Defined Daily Dose (DDD). DDD is the average maintenance dose per day for a drug used for its main indication in adults. HTN, Hypertension; BP, Blood Pressure; PRA, Plasma Renin Activity; LVH, Left Ventricular Hypertrophy; Echo, Echocardiography; CV, Cardiovascular; CT, Computed Tomography; SI, Selectivity Index; A/C, Aldosterone to Cortisol ratio. Ipsilateral and contralateral imaging were defined as “normal” in absence of nodules, or with thickening < 4 mm.

**Supplementary Table 3. Patient Characteristics of Study Cohort**

Variable	Combined Cohort (N = 158)	Training Cohort (N = 110)	Validation Cohort (N = 48)	P-value
ACTH protocol, n (%)	82 (51.9)	53 (48.2)	29 (60.4)	0.157
Diagnosis of UPA, n (%)	126 (79.7)	88 (80.0)	38 (79.2)	0.905
Female sex, n (%)	60 (38.0)	40 (36.4)	20 (41.7)	0.528
Age at diagnosis (years)	50 ± 10.0	50 ± 9.3	50 ± 11.5	0.949
Duration of HTN (months)	67 [23; 130]	66 [23; 128]	68 [24; 161]	0.825
Systolic BP (mmHg)	164 ± 24.5	163 ± 24.5	166 ± 24.6	0.534
Diastolic BP (mmHg)	98 ± 13.9	98 ± 13.8	97 ± 14.3	0.925
Antihypertensive medication (DDD)	3.8 [2.2; 5.0]	3.7 [2.4; 5.0]	4.0 [2.2; 5.5]	0.986
PRA at screening (ng/mL/h)	0.20 [0.12; 0.40]	0.29 [0.15; 0.40]	0.20 [0.10; 0.30]	0.166
Aldosterone at screening (ng/dL)	36.0 [25.5; 47.8]	35.9 [25.6; 47.3]	37.9 [25.0; 49.8]	0.867
Lowest potassium (mEq/L)	3.2 ± 0.7	3.2 ± 0.7	3.2 ± 0.7	0.658
Aldosterone post-confirmatory test (ng/dL)	19.6 [11.3; 32.0]	20.1 [13.1; 32.3]	16.4 [9.5; 33.0]	0.172
Microalbuminuria, n (%)	43 (27.5)	32 (29.1)	11 (21.9)	0.394
LVH at Echo, n (%)	94 (59.2)	65 (59.1)	29 (59.5)	0.963
CV events, n (%)	23 (14.6)	16 (14.4)	7 (15.0)	0.934
Largest nodule at CT scanning (diameter, mm)	13 [10; 19]	14 [10; 20]	12 [10; 18]	0.219
Normal ipsilateral imaging, n (%)	16 (10.1)	10 (9.1)	6 (12.5)	0.514
Normal contralateral imaging, n (%)	124 (78.5)	84 (76.4)	40 (83.3)	0.327
Contralateral SI	18.8 [10.2; 41.5]	18.8 [11.3; 47.5]	20.2 [7.9; 30.3]	0.463
Peripheral A/C ratio	17.8 [3.4; 28.7]	15.9 [3.7; 23.1]	14.3 [5.4; 26.0]	0.950
Contralateral A/C ratio	4.4 [1.2; 9.9]	4.4 [1.4; 9.1]	5.5 [1.8; 9.3]	0.456
Contralateral ratio	0.3 [0.2; 0.6]	0.3 [0.2; 0.6]	0.4 [0.2; 0.7]	0.393

Characteristics of patients included in the analysis: patients from the combined cohort (N = 158) were randomly assigned to training (N = 110), or validation cohort (N = 48). Antihypertensive medication is expressed as Defined Daily Dose (DDD). DDD is the average maintenance dose per day for a drug used for its main indication in adults. UPA, Unilateral Primary Aldosteronism; HTN, Hypertension; BP, Blood Pressure; PRA, Plasma Renin Activity; LVH, Left Ventricular Hypertrophy; Echo, Echocardiography; CV, Cardiovascular; CT, Computed Tomography; SI, Selectivity Index; A/C, Aldosterone to Cortisol ratio. Normally and non-normally distributed variables were reported as mean ± standard deviation or median [interquartile range], respectively. Categorical variables were reported as absolute number (n) and proportion (%).

**Supplementary Table 4. Diagnostic performance**

Score Performance	N	Cut-off	Sens (%)	Spec (%)	PPV (%)	NPV (%)	Acc (%)
<b>LDA Model (Training)</b>	110		90.6 (88.2-93.0)	99.4 (96.4-100.0)	99.8 (99.0-100.0)	72.6 (67.4-77.8)	92.3 (90.1-94.5)
Internal Validation	110	N.A.	88.9 (85.6-92.2)	88.9 (77.7-100.0)	97.0 (94.1-99.9)	66.8 (59.5-74.1)	88.9 (85.3-92.5)
External Validation	48		84.7 (82.7-86.6)	84.2 (73.1-95.3)	95.4 (92.2-98.6)	59.1 (55.1-63.1)	84.6 (82.1-87.1)
Combined cohort	158		88.8 (86.8-90.8)	94.7 (91.1-98.3)	98.5 (97.5-99.5)	68.3 (64.7-71.9)	90.0 (88.5-91.6)
<b>RF Model (Training)</b>	110		95.8 (94.1-97.5)	94.9 (91.7-98.1)	98.7 (97.9-99.5)	85.1 (80.0-90.2)	95.6 (94.2-97.0)
Internal Validation	110	N.A.	92.2 (89.1-95.3)	79.5 (66.5-92.6)	94.7 (91.6-97.8)	71.9 (63.6-80.2)	89.6 (86.1-93.1)
External Validation	48		83.6 (79.0-88.2)	80.0 (77.8-82.2)	94.1 (93.8-94.4)	56.4 (49.2-63.6)	82.9 (79.3-86.5)
Combined cohort	158		92.1 (90.2-94.0)	90.3 (88.1-92.5)	97.4 (96.8-98.0)	74.5 (69.8-79.2)	91.7 (90.1-93.3)
<b>CLR Score - Best Accuracy</b>	110		95.5 (88.9-98.2)	100.0 (85.1-100.0)	100.0 (96.0-100.0)	84.7 (65.7-93.3)	96.4 (88.1-98.6)
Internal Validation	110	> 11	95.3 (83.9-100.0)	100.0 (97.5-100.0)	100.0 (98.4-100.0)	88.3 (59.9-95.4)	96.3 (87.7-100.0)
External Validation	48		94.7 (82.7-99.1)	90.0 (59.6-99.5)	97.3 (88.6-99.9)	81.7 (47.6-96.7)	93.7 (77.9-99.2)
Combined Cohort	158		95.2 (90.0-97.8)	96.9 (84.3-99.8)	99.2 (95.8-99.9)	83.7 (68.2-92.0)	95.5 (88.8-98.2)
<b>CLR Score (Sensitivity Optimization)</b>	110		97.8 (92.1-99.6)	45.5 (26.9-65.3)	87.8 (83.4-92.0)	83.8 (46.0-97.6)	87.3 (79.1-92.7)
Internal Validation	110	> 8	96.9 (87.6-100.0)	51.7 (28.0-75.4)	88.8 (72.0-100.0)	75.0 (47.0-96.9)	87.3 (71.0-94.7)
External Validation	48		97.4 (86.5-99.9)	50.0 (23.7-76.3)	88.1 (81.2-94.1)	83.5 (31.6-99.5)	87.5 (73.4-95.0)
Combined Cohort	158		97.6 (93.2-99.4)	46.9 (30.9-63.6)	87.9 (84.2-91.5)	83.2 (53.6-96.4)	87.3 (80.6-92.1)
<b>CLR Score (Specificity Optimization)</b>	110		81.8 (72.5-88.5)	100.0 (85.1-100.0)	100.0 (95.1-100.0)	57.9 (43.6-68.5)	85.4 (75.0-90.8)
Internal Validation	110	≥ 13	79.2 (51.4-93.2)	100.0 (97.8-100.0)	100.0 (97.3-100.0)	54.5 (24.8-83.1)	82.7 (59.5-95.8)
External Validation	48		84.2 (69.6-92.6)	100.0 (72.2-100.0)	100.0 (90.5-100.0)	62.5 (38.5-78.1)	87.5 (70.1-94.1)
Combined Cohort	158		82.5 (75.0-88.2)	100.0 (89.3-100.0)	100.0 (96.5-100.0)	59.2 (47.6-68.3)	86.0 (77.9-90.6)

Sensitivity (Sens), specificity (Spec), positive/negative predictive value (PPV/ NPV), and accuracy (Acc) for proposed models (each indicator is derived considering unilateral PA as referral diagnosis and reported together with its 95% binomial confidence interval; confidence intervals are derived by normal approximation method and capped to 100%). Diagnostic indices are provided at training, at internal validation by 10-fold cross validation, at external validation, and in the combined cohort. A cut-off of greater than 11 corresponded to the higher model accuracy, as assessed by ROC curve analysis. The cut-offs of > 8 and ≥ 13 increased sensitivity or specificity, respectively, providing an overall model accuracy of at least 85% at validation. LDA, Linear Discriminant Analysis; RF, Random Forest.

**Supplementary Table 5. Estimate predictor importance for LDA and RF models**

<b>Variable</b>	<b>LDA</b>	<b>RF</b>
Aldosterone at screening (ng/dL)	0.047626	0.093263
Lowest potassium (mEq/L)	-0.287831	1.000000
Aldosterone post-confirmatory test (ng/dL)	0.033038	0.797178
Normal ipsilateral imaging, n (%)	-0.994998	0.327111
Normal contralateral imaging, n (%)	0.064665	0.172839
Contralateral ratio	-1.000000	0.707113

Normalized LDA and RF importance coefficients computed at the training of the models. A highest absolute value corresponds to the best predictor in each model.

**Supplementary Table 6. Score development and validation**

AVS Score Accuracy		Predicted Diagnosis		Performance		
Real Diagnosis (Cut-off > 11)	Training cohort (N = 110)		UPA	BPA	Accuracy (%)	96.4
		UPA	84	4	Sensitivity (%)	95.5
		BPA	0	22	Specificity (%)	100.0
	Validation cohort (N = 48)		UPA	BPA	Accuracy (%)	93.7
		UPA	36	2	Sensitivity (%)	94.7
		BPA	1	9	Specificity (%)	90.0
	Combined cohort (N = 158)		UPA	BPA	Accuracy (%)	95.5
		UPA	120	6	Sensitivity (%)	95.2
		BPA	1	31	Specificity (%)	96.9
Real Diagnosis (Cut-off > 8)	Training cohort (N = 110)		UPA	BPA	Accuracy (%)	87.3
		UPA	86	2	Sensitivity (%)	97.8
		BPA	12	10	Specificity (%)	45.5
	Validation cohort (N = 48)		UPA	BPA	Accuracy (%)	87.5
		UPA	37	1	Sensitivity (%)	97.4
		BPA	5	5	Specificity (%)	50.0
	Combined cohort (N = 158)		UPA	BPA	Accuracy (%)	87.3
		UPA	123	3	Sensitivity (%)	97.6
		BPA	17	15	Specificity (%)	46.9
Real Diagnosis (Cut-off >= 13)	Training cohort (N = 110)		UPA	BPA	Accuracy (%)	85.4
		UPA	72	16	Sensitivity (%)	81.8
		BPA	0	22	Specificity (%)	100.0
	Validation cohort (N = 48)		UPA	BPA	Accuracy (%)	87.5
		UPA	32	6	Sensitivity (%)	84.2
		BPA	0	10	Specificity (%)	100.0
	Combined cohort (N = 158)		UPA	BPA	Accuracy (%)	86.0
		UPA	104	22	Sensitivity (%)	82.5
		BPA	0	32	Specificity (%)	100.0

The table shows the real and predicted subtype diagnosis, accuracy sensitivity, specificity for the training cohort (N = 110), the validation cohort (N = 48), and the combined cohort (N = 158). A cut-off of greater than 11 identifies patients with unilateral primary aldosteronism (UPA) with the higher accuracy. A cut-off of greater than 8 identifies patients with a diagnosis of UPA with an optimized sensitivity; cut-off of greater than or equal to 13 identifies patients with a diagnosis of UPA with an optimized specificity. BPA, Bilateral Primary Aldosteronism.

**Supplementary Table 7. Distribution of PA patients according to the score**

Score points	Total (N)	UPA		BPA	
		(N)	(%)	(N)	(%)
0.0-1.0	1	0	0.0	1	100.0
1.1-3.0	5	0	0.0	5	100.0
3.1-5.0	2	1	50.0	1	50.0
5.1-7.0	3	0	0.0	3	100.0
7.1-9.0	15	4	26.7	11	73.3
9.1-11.0	11	1	9.1	10	90.9
11.1-13.0	33	32	97.0	1	3.0
13.1-15.0	21	21	100.0	0	0.0
15.1-17.0	34	34	100.0	0	0.0
> 17.0	33	33	100.0	0	0.0
<i>Total</i>	158	126	N.A.	32	N.A.

The number (N) and the proportion (%) of patients stratified for subtype diagnosis (unilateral PA *versus* bilateral PA) is shown according to the score in the combined cohort (N = 158).

**Supplementary Table 8. Sub-analysis according to patient outcome**

<b>Subgroup of patients</b>	<b>Clinical Outcome</b>	<b>Biochemical Outcome</b>
<i>UPA predicted as UPA (true prediction)</i>		
Complete, n (%)	51 (53.1)	94 (97.9)
Partial, n (%)	41 (42.7)	2 (2.1)
Absent, n (%)	4 (4.2)	0 (0.0)
<i>UPA predicted as BPA (false prediction)</i>		
Complete, n (%)	0 (0.0)	4 (66.7)
Partial, n (%)	5 (83.3)	2 (33.3)
Absent, n (%)	1 (16.7)	0 (0.0)
<i>Unstimulated AVS</i>		
Complete, n (%)	21 (42.9)	48 (98.0)
Partial, n (%)	27 (55.1)	1 (2.0)
Absent, n (%)	1 (2.0)	0 (0.0)
<i>ACTH-stimulated AVS</i>		
Complete, n (%)	30 (56.7)	50 (94.3)
Partial, n (%)	19 (35.8)	3 (5.7)
Absent, n (%)	4 (7.5)	0 (0.0)

Clinical and biochemical outcomes according to the PASO criteria were reported for patients with unilateral primary aldosteronism (UPA) after a follow-up of 6-12 months. Patients were stratified according to the CLR score predicted diagnosis (true vs. false predictions) or AVS protocol (ACTH-stimulated vs. unstimulated procedures).