SUPPLEMENTAL DATA

Development and validation of a prediction score to avoid confirmatory testing in patients with suspected primary aldosteronism.

Jacopo Burrello MD[1*], Martina Amongero MS[2*], Fabrizio Buffolo MD[1], Elisa Sconfienza MD[1], Vittorio Forestiero MD[1], Alessio Burrello MS[3], Christian Adolf MD[4], Laura Handgriff MD[4], Martin Reincke MD[4], Franco Veglio MD[1], Tracy Ann Williams PhD[1,4], Silvia Monticone MD, PhD[1#], Paolo Mulatero MD[1#].

(1) Division of Internal Medicine and Hypertension, Department of Medical Sciences, University of Torino, Italy. (2) Department of Mathematical Sciences G. L. Lagrange, Polytechnic University of Torino, Italy. (3) Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi" (DEI), University of Bologna, Italy. (4) Medizinische Klinik und Poliklinik IV, Klinikum der Universität, Ludwig-Maximilians-Universität München, Munich, Germany.

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

Summary

Table S1. Patient Characteristics of Study Cohort: Univariate Regression Analysis

Variable (ref. PA confirmed)	OR (CI 95%)	*P*-value
Age at diagnosis (years)	0.99 (0.98-1.01)	0.202
Female sex, n (%)	0.34 (0.24-0.46)	**<0.001**
Duration of HTN (months)	1.01 (1.00-1.01)	0.089
Systolic BP (mmHg)	1.01 (1.01-1.02)	**0.003**
Diastolic BP (mmHg)	1.01 (1.00-1.03)	0.052
Antihypertensive medication (DDD)	1.40 (1.27-1.54)	**<0.001**
BMI (Kg/sqm)	1.02 (0.99-1.06)	0.204
PRA at screening (ng/mL/h)	0.28 (0.14-0.57)	**<0.001**
Aldosterone at screening (ng/dL)	1.06 (1.04-1.08)	**<0.001**
Lowest Potassium (mEq/L)	0.13 (0.09-0.19)	**<0.001**
eGFR (mL/min)	1.00 (0.99-1.01)	0.666
Diabetes, n (%)	1.60 (0.82-3.12)	0.166
Organ damage, n (%)	3.13 (2.28-4.29)	**<0.001**
CV events, n (%)	2.16 (1.27-3.67)	**0.004**

Odds ratio (OR) and the 95% confidence interval (CI) were evaluated by univariate logistic regression analysis for each variable. An OR greater than 1 indicates an increased likelihood of confirmed PA, and an OR less than 1 a decreased likelihood. HTN, Hypertension; BP, Blood Pressure; DDD, Defined Daily Dose (average maintenance dose per day for a drug used for its main indication in adults); PRA, Plasma Renin Activity; eGFR, estimated Glomerular Filtration Rate; CV, Cardiovascular. Organ damage is defined as presence of left ventricular hypertrophy at echocardiography and/or microalbuminuria.

Table S2. Patient Characteristics of Study Cohort: Multivariate Regression Analysis

Variable (ref. PA confirmed)	OR (CI 95%)	*P*-value
Female sex, n (%)	0.42 (0.28-0.62)	**<0.001**
Systolic BP (mmHg)	1.00 (0.99-1.01)	0.566
Antihypertensive medication (DDD)	1.21 (1.07-1.36)	**0.002**
PRA at screening (ng/mL/h)	0.07 (0.03-0.19)	**<0.001**
Aldosterone at screening (ng/dL)	1.08 (1.06-1.11)	**<0.001**
Lowest Potassium (mEq/L)	0.15 (0.09-0.23)	**<0.001**
Organ damage, n (%)	2.64 (1.74-4.01)	**<0.001**
CV events, n (%)	1.40 (0.72-2.72)	0.315

Odds ratio (OR) and the 95% confidence interval (CI) were evaluated by multivariate logistic regression analysis for variables associated to a confirmed PA diagnosis in the univariate model. An OR greater than 1 indicates an increased likelihood of confirmed PA, and an OR less than 1 a decreased likelihood. BP, Blood Pressure; DDD, Defined Daily Dose (average maintenance dose per day for a drug used for its main indication in adults); PRA, Plasma Renin Activity; CV, Cardiovascular. Organ damage is defined as presence of left ventricular hypertrophy at echocardiography and/or microalbuminuria.

Table S3. Characteristics of Training versus Internal Validation cohort

Variable	Combined Cohort (n=696)	Training Cohort (n=522)	Validation Cohort (n=174)	*P*-value
Confirmed PA, n (%)	421 (60.5)	322 (61.7)	99 (56.9)	0.263
Subtyping, UPA (%)	133 (19.1)	98 (18.8)	35 (20.1)	0.313
Age at diagnosis (years)	50 ± 9.9	50 ± 10.2	50 ± 9.3	0.770
Female sex, n (%)	318 (45.7)	239 (45.8)	79 (45.4)	0.930
Duration of HTN (months)	64 [21; 131]	59 [21; 128]	75 [23; 134]	0.300
Systolic BP (mmHg)	155 ± 20.3	155 ± 20.4	155 ± 20.2	0.889
Diastolic BP (mmHg)	95 ± 11.0	95 ± 11.1	94 ± 10.8	0.531
Antihypertensive medication (DDD)	2.15 [1.00; 4.00]	2.00 [1.00; 3.69]	2.33 [1.00; 4.00]	0.765
BMI (Kg/sqm)	25.7 ± 4.28	25.9 ± 4.23	25.4 ± 4.45	0.224
PRA at screening (ng/mL/h)	0.30 [0.15; 0.40]	0.22 [0.15; 0.40]	0.30 [0.20; 0.45]	0.086
Aldosterone at screening (ng/dL)	25.6 [18.7; 35.5]	25.8 [18.8; 35.5]	24.3 [18.5; 35.1]	0.791
Lowest Potassium (mEq/L)	3.8 ± 0.62	3.8 ± 0.62	3.8 ± 0.61	0.414
eGFR (mL/min)	91 ± 17.0	91 ± 17.2	91 ± 16.6	0.914
Diabetes, n (%)	44 (6.3)	33 (6.3)	11 (6.3)	1.000
Organ damage, n (%)	404 (58.0)	298 (57.1)	106 (60.9)	0.375
CV events, n (%)	81 (11.6)	62 (11.9)	19 (10.9)	0.733

Characteristics of patients included in the developmental cohort: patients from the combined cohort (n=696) were randomly assigned to training (n=522), or validation cohort (n=174). HTN, Hypertension; BP, Blood Pressure; DDD, Defined Daily Dose (average maintenance dose per day for a drug used for its main indication in adults); PRA, Plasma Renin Activity; eGFR, estimated Glomerular Filtration Rate; CV, Cardiovascular. Organ damage is defined as presence of left ventricular hypertrophy at echocardiography and/or microalbuminuria. 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 S4. Characteristics of Developmental versus Validation cohort

Variable	Developmental Cohort (n=696)	External Validation Cohort (n=328)	*P*-value
Confirmed PA, n (%)	421 (60.5)	173 (52.7)	**0.019**
Subtyping, UPA (%)	133 (19.1)	89 (27.1)	0.299
Age at diagnosis (years)	50 ± 9.9	50 ± 13.5	0.467
Female sex, n (%)	318 (45.7)	192 (58.5)	**<0.001**
Duration of HTN (months)	64 [21; 131]	48 [11; 138]	**0.006**
Systolic BP (mmHg)	155 ± 20.3	150 ± 19.6	**<0.001**
Diastolic BP (mmHg)	95 ± 11.0	93 ± 12.4	0.136
Antihypertensive medication (DDD)	2.15 [1.00; 4.00]	1.00 [0.00; 2.50]	**<0.001**
BMI (Kg/sqm)	25.7 ± 4.28	27.0 ± 5.09	**<0.001**
PRA at screening (ng/mL/h)	0.30 [0.15; 0.40]	N.A.	N.A.
DRC at screening (mU/L)	N.A.	2.7 [2.0; 5.6]	N.A.
Aldosterone at screening (ng/dL)	25.6 [18.7; 35.5]	12.8 [8.2; 20.0]	**<0.001**
Lowest Potassium (mEq/L)	3.8 ± 0.62	3.5 ± 0.51	**<0.001**
eGFR (mL/min)	91 ± 17.0	87 ± 19.9	**0.001**
Diabetes, n (%)	44 (6.3)	36 (11.0)	**0.010**
Organ damage, n (%)	404 (58.0)	129 (39.3)	**<0.001**
CV events, n (%)	81 (11.6)	39 (11.9)	0.907

Characteristics of patients included in the analysis: patients from the developmental cohort from Torino (n=696) were compared to patients from the external validation cohort from Munich (n=328). HTN, Hypertension; BP, Blood Pressure; DDD, Defined Daily Dose (average maintenance dose per day for a drug used for its main indication in adults); PRA, Plasma Renin Activity; eGFR, estimated Glomerular Filtration Rate; CV, Cardiovascular. Organ damage is defined as presence of left ventricular hypertrophy at echocardiography and/or microalbuminuria. 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 S5. Diagnostic performance of machine learning based models

	PACT Score Accuracy	Predicted Diagnosis		Performance	
LDA Model	**Training cohort** (N = 522)	PA confirmed	PA excluded	Accuracy (%)	79.7
	PA confirmed	272	50	Sensitivity (%)	84.5
	PA excluded	56	144	Specificity (%)	72.0
	Validation cohort (N = 174)	PA confirmed	PA excluded	Accuracy (%)	77.6
	PA confirmed	82	17	Sensitivity (%)	82.8
	PA excluded	22	53	Specificity (%)	70.7
	Combined cohort (N = 696)	PA confirmed	PA excluded	Accuracy (%)	79.2
	PA confirmed	354	67	Sensitivity (%)	84.1
	PA excluded	78	197	Specificity (%)	71.6
RF Model	**Training cohort** (N = 522)	PA confirmed	PA excluded	Accuracy (%)	82.8
	PA confirmed	286	36	Sensitivity (%)	88.8
	PA excluded	54	146	Specificity (%)	73.0
	Validation cohort (N = 174)	PA confirmed	PA excluded	Accuracy (%)	79.9
	PA confirmed	84	15	Sensitivity (%)	84.8
	PA excluded	20	55	Specificity (%)	73.3
	Combined cohort (N = 696)	PA confirmed	PA excluded	Accuracy (%)	82.0
	PA confirmed	370	51	Sensitivity (%)	87.9
	PA excluded	74	201	Specificity (%)	73.1
Linear SVM	**Training cohort** (N = 522)	PA confirmed	PA excluded	Accuracy (%)	80.7
	PA confirmed	272	50	Sensitivity (%)	84.5
	PA excluded	51	149	Specificity (%)	74.5
	Validation cohort (N = 174)	PA confirmed	PA excluded	Accuracy (%)	78.2
	PA confirmed	82	17	Sensitivity (%)	82.8
	PA excluded	21	54	Specificity (%)	72.0
	Combined cohort (N = 696)	PA confirmed	PA excluded	Accuracy (%)	80.0
	PA confirmed	354	67	Sensitivity (%)	84.1
	PA excluded	72	203	Specificity (%)	73.8
Gaussian SVM	**Training cohort** (N = 522)	PA confirmed	PA excluded	Accuracy (%)	83.9
	PA confirmed	284	38	Sensitivity (%)	88.2
	PA excluded	46	154	Specificity (%)	77.0
	Validation cohort (N = 174)	PA confirmed	PA excluded	Accuracy (%)	74.7
	PA confirmed	81	18	Sensitivity (%)	81.8
	PA excluded	26	49	Specificity (%)	65.3
	Combined cohort (N = 696)	PA confirmed	PA excluded	Accuracy (%)	81.6
	PA confirmed	365	56	Sensitivity (%)	86.7
	PA excluded	72	203	Specificity (%)	73.8

The table shows real and predicted diagnosis (PA confirmed vs. excluded), accuracy, sensitivity, specificity for the training cohort (n=522), the validation cohort (n=174), and the combined cohort from Torino (n=696). Diagnostic performance is shown for LDA (linear discriminant analysis), RF (random forest), linear and gaussian SVM (support vector machine) models.

Table S6. Score development and validation

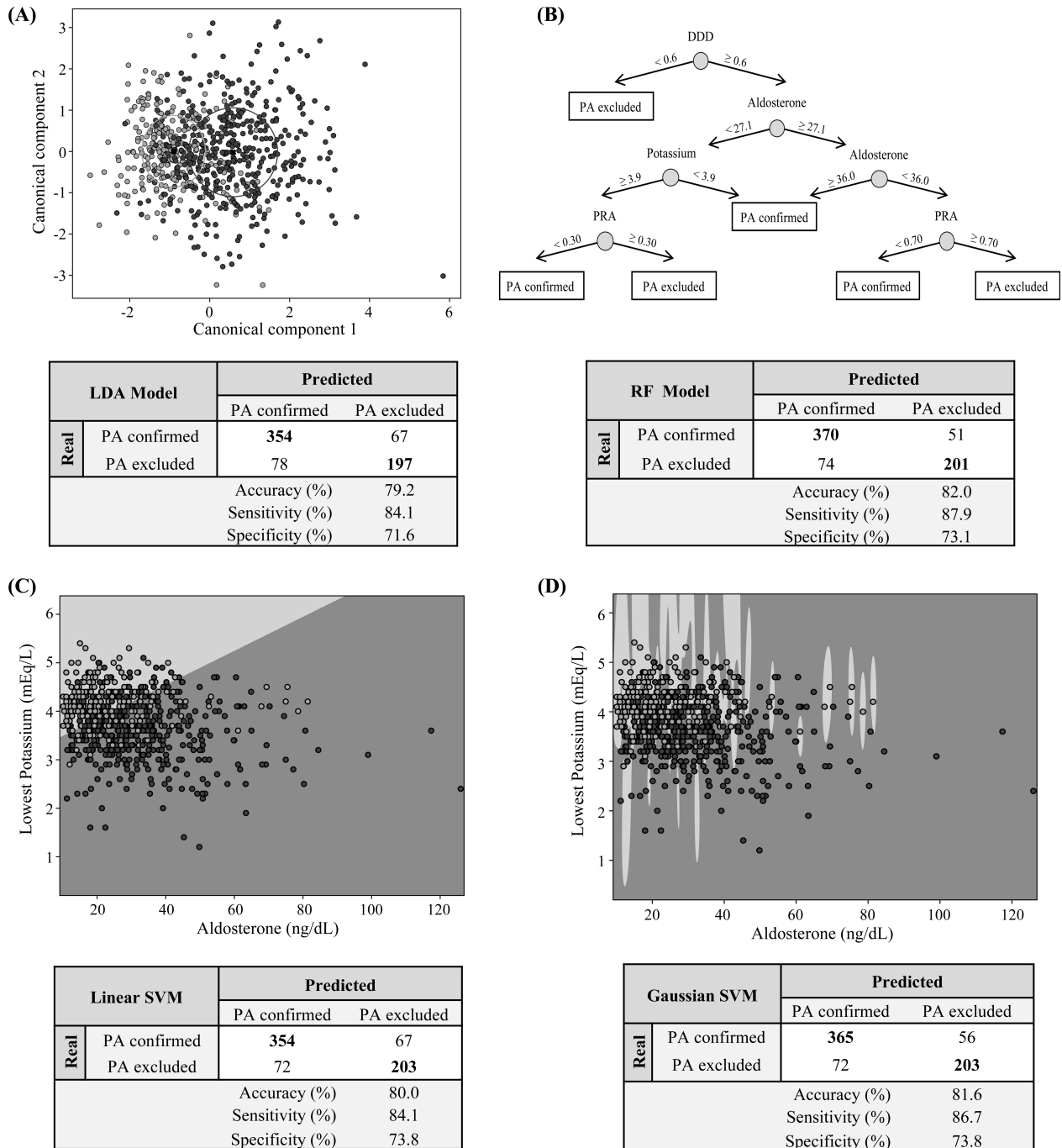| PACT Score Accuracy	Predicted Diagnosis		Performance	
Real Diagnosis (Cut-off ≥ 5)				
Training cohort (N = 522)	PA confirmed	PA excluded	Accuracy (%)	70.7
PA confirmed	322	0	Sensitivity (%)	100.0
PA excluded	153	47	Specificity (%)	23.5
Validation cohort (N = 174)	PA confirmed	PA excluded	Accuracy (%)	66.1
PA confirmed	99	0	Sensitivity (%)	100.0
PA excluded	59	16	Specificity (%)	21.3
Combined cohort (N = 696)	PA confirmed	PA excluded	Accuracy (%)	69.5
PA confirmed	421	0	Sensitivity (%)	100.0
PA excluded	212	63	Specificity (%)	22.9
Real Diagnosis (Cut-off ≥ 8)				
Training cohort (N = 522)	PA confirmed	PA excluded	Accuracy (%)	84.1
PA confirmed	297	25	Sensitivity (%)	92.2
PA excluded	58	142	Specificity (%)	71.0
Validation cohort (N = 174)	PA confirmed	PA excluded	Accuracy (%)	83.9
PA confirmed	91	8	Sensitivity (%)	91.9
PA excluded	20	55	Specificity (%)	73.3
Combined cohort (N = 696)	PA confirmed	PA excluded	Accuracy (%)	84.1
PA confirmed	388	33	Sensitivity (%)	92.2
PA excluded	78	197	Specificity (%)	71.6
Real Diagnosis (Cut-off ≥ 13)				
Training cohort (N = 522)	PA confirmed	PA excluded	Accuracy (%)	53.6
PA confirmed	80	242	Sensitivity (%)	24.8
PA excluded	0	200	Specificity (%)	100.0
Validation cohort (N = 174)	PA confirmed	PA excluded	Accuracy (%)	55.2
PA confirmed	21	78	Sensitivity (%)	21.2
PA excluded	0	75	Specificity (%)	100.0
Combined cohort (N = 696)	PA confirmed	PA excluded	Accuracy (%)	54.0
PA confirmed	101	320	Sensitivity (%)	24.0
PA excluded	0	275	Specificity (%)	100.0

The table shows real and predicted diagnosis (PA confirmed vs. excluded), accuracy, sensitivity, specificity for the training cohort (n=522), the validation cohort (n=174), and the combined cohort from Torino (n=696). Diagnostic performance is shown for the PACT (Primary Aldosteronism Confirmatory Testing) score. A cut-off of equal or greater than 5 identifies patients with a confirmed diagnosis of PA with the maximum sensitivity; a cut-off of equal or greater than 8 identifies patients with a confirmed diagnosis of PA with the higher accuracy; a cut-off of equal or greater than 13 identifies patients with a confirmed diagnosis of PA with the maximum specificity.

Table S7. Distribution of PA patients according to the score

Score Points	Total (n)	PA excluded (n)	PA excluded (%)	PA confirmed (n)	PA confirmed (%)
0.0-2.0	14	14	100.0	0	0.0
2.1-4.0	49	49	100.0	0	0.0
4.1-6.0	138	106	76.8	32	23.2
6.1-8.0	145	63	43.4	82	56.6
8.1-10.0	137	25	18.2	112	81.8
10.1-12.0	112	18	16.1	94	83.9
12.1-14.0	87	0	0.0	87	100.0
14.1-16.0	14	0	0.0	14	100.0
Total	696	275	N.A.	421	N.A.

Number (n) and proportion (%) of patients stratified for diagnosis (PA excluded *vs.* confirmed) is shown according to the score in the developmental cohort of Torino (n=696). N.A., Not Applicable.
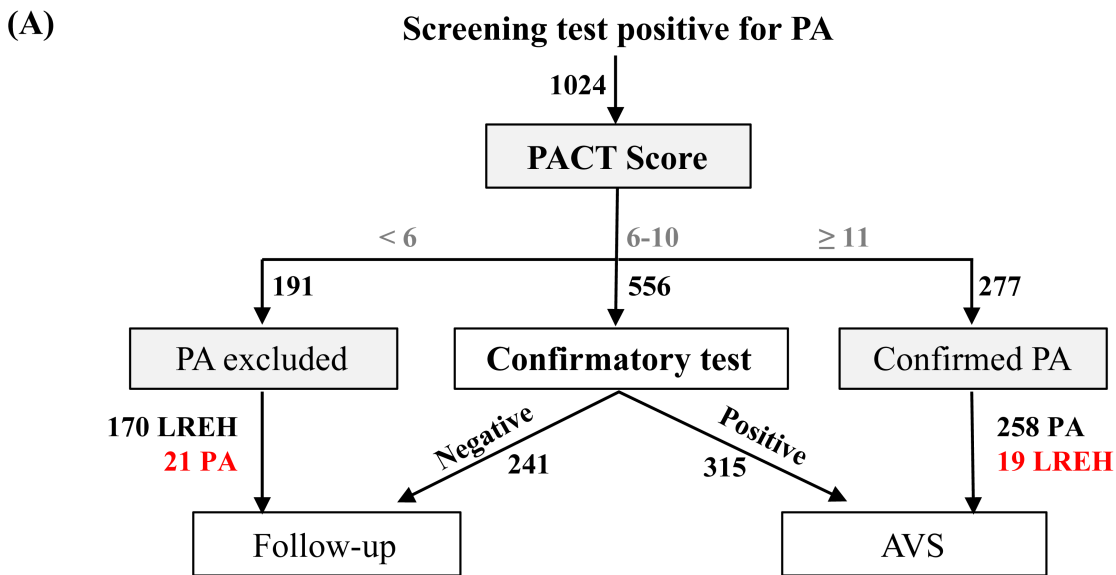
Figure S1. Diagnostic modelling

(A)

(B)

LDA Model		Predicted	
		PA confirmed	PA excluded
Real	PA confirmed	**354**	67
	PA excluded	78	**197**
	Accuracy (%)	79.2	
	Sensitivity (%)	84.1	
	Specificity (%)	71.6	

RF Model		Predicted	
		PA confirmed	PA excluded
Real	PA confirmed	**370**	51
	PA excluded	74	**201**
	Accuracy (%)	82.0	
	Sensitivity (%)	87.9	
	Specificity (%)	73.1	

(C)

(D)

Linear SVM		Predicted	
		PA confirmed	PA excluded
Real	PA confirmed	**354**	67
	PA excluded	72	**203**
	Accuracy (%)	80.0	
	Sensitivity (%)	84.1	
	Specificity (%)	73.8	

Gaussian SVM		Predicted	
		PA confirmed	PA excluded
Real	PA confirmed	**365**	56
	PA excluded	72	**203**
	Accuracy (%)	81.6	
	Sensitivity (%)	86.7	
	Specificity (%)	73.8	

Machine learning based models to discriminate patients with a confirmed diagnosis of PA (n=421) from patients with PA excluded (n=275). The models included the 6 variables with the highest prediction power. Confusion matrix shows real and predicted diagnosis, accuracy, sensitivity, and specificity for each model in the developmental cohort (n=696). Data on training and validation of the models are reported in Table S5. (**A**) Canonical plot representing diagnostic performance of LDA; each patient is indicated by a point and diagnosis are reported by colour (confirmed PA, black; PA excluded, grey). The axes (canonical component 1 and 2) are calculated by weighted linear combination of the 6 variables included in the model to maximize the separation between groups. The crosses indicate the means of (canonical 1; canonical 2) for patients with UPA or BPA, the ellipse included patients with a linear combination coefficient that falls within the mean ± SD. (**B**)

The first classification tree of the forest is shown for the prediction of PA confirmed *vs.* PA excluded. (**C**, **D**) Graphs showing the performance of SVM models (Support Vector Machine, Linear and Gaussian). Axes report the two best support vector classifiers: aldosterone at screening on x-axis and lowest recorded potassium levels on y-axis. Each patient is indicated by a point and diagnosis are reported by colour (confirmed PA, dark grey; PA excluded, grey). Model prediction areas are indicated by colours, as appropriated.

Figure S2. Flow chart for the management of PA patients

(A)

Screening test positive for PA

1024

PACT Score

< 6   6-10   ≥ 11

191   556   277

| PA excluded | Confirmatory test | Confirmed PA |

170 LREH
21 PA

Negative 241   Positive 315

258 PA
19 LREH

| Follow-up | | AVS |

(B)

Diagnosis		Predicted			
		PA confirmed	PA excluded	Sensitivity (%)	96.5
				Specificity (%)	95.6
Real	PA confirmed	573	21	PPV (%)	96.8
	PA excluded	19	411	NPV (%)	95.1
Accuracy 96.1% - Necessary confirmatory test – 45.7%					

Flow chart for the management of patients with a positive screening test (Developmental Cohort + External Validation Cohort; n=1,024). (**A**) PA patient management using the PACT score; the number of patients is indicated in bold; cut-offs are indicated in grey. Misclassified patients are reported in red. (**B**) Confusion matrix representing real and predicted subtype diagnosis, sensitivity, specificity, positive and negative predictive value (PPV; NPV). AVS, Adrenal Venous Sampling; PA, Primary Aldosteronism; LREH, Low Renin Essential Hypertensive patients (PA positive screening test with a negative confirmatory test); PACT, Primary Aldosteronism Confirmatory Testing Score.