

## SUPPLEMENTAL DATA

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

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

(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

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

Table S3. Characteristics of Training *versus* Internal Validation cohort

Table S4. Characteristics of Developmental *versus* Validation cohort

Table S5. Diagnostic performance of machine learning based models

Table S6. Diagnostic performance of the models at training and validation

Table S7. Score development and validation: optimization of sensitivity and specificity

Table S8. Correlation of the PACT score with hormonal measurements at confirmatory testing

Table S9. Distribution of patients with PA according to the score

Table S10. Diagnostic performance in patients with unilateral PA

Figure S1. Diagnostic modelling

Figure S2. Predictor importance for machine learning based models

Figure S3. Flow chart for the management of patients with PA

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

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

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**

| <b>Variable (ref. PA confirmed)</b> | <b>OR (CI 95%)</b> | <b>P-value</b>   |
|-------------------------------------|--------------------|------------------|
| Female sex, n (%)                   | 0.42 (0.28-0.62)   | <b>&lt;0.001</b> |
| Systolic BP (mmHg)                  | 1.00 (0.99-1.01)   | 0.566            |
| Antihypertensive medication (DDD)   | 1.21 (1.07-1.36)   | <b>0.002</b>     |
| PRA at screening (ng/mL/h)          | 0.07 (0.03-0.19)   | <b>&lt;0.001</b> |
| Aldosterone at screening (ng/dL)    | 1.08 (1.06-1.11)   | <b>&lt;0.001</b> |
| Lowest Potassium (mEq/L)            | 0.15 (0.09-0.23)   | <b>&lt;0.001</b> |
| Organ damage, n (%)                 | 2.64 (1.74-4.01)   | <b>&lt;0.001</b> |
| 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)                         | <b>0.019</b>     |
| 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)                         | <b>&lt;0.001</b> |
| Duration of HTN (months)                   | 64 [21; 131]                 | 48 [11; 138]                       | <b>0.006</b>     |
| Systolic BP (mmHg)                         | 155 ± 20.3                   | 150 ± 19.6                         | <b>&lt;0.001</b> |
| 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]                  | <b>&lt;0.001</b> |
| BMI (Kg/sqm)                               | 25.7 ± 4.28                  | 27.0 ± 5.09                        | <b>&lt;0.001</b> |
| 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]                   | <b>&lt;0.001</b> |
| Lowest Potassium (mEq/L)                   | 3.8 ± 0.62                   | 3.5 ± 0.51                         | <b>&lt;0.001</b> |
| eGFR (mL/min)                              | 91 ± 17.0                    | 87 ± 19.9                          | <b>0.001</b>     |
| Diabetes, n (%)                            | 44 (6.3)                     | 36 (11.0)                          | <b>0.010</b>     |
| Organ damage, n (%)                        | 404 (58.0)                   | 129 (39.3)                         | <b>&lt;0.001</b> |
| CV events, n (%)                           | 81 (11.6)                    | 39 (11.9)                          | 0.907            |
| Confirmatory testing                       |                              |                                    |                  |
| Saline infusion test, n (%)                | 560 (80.5)                   | 319 (97.3)                         | <b>&lt;0.001</b> |
| Captopril Challenge test, n (%)            | 136 (19.5)                   | 9 (2.7)                            |                  |
| PRA pre-confirmatory test (ng/mL/h)        | 0.20 [0.10; 0.40]            | N.A.                               | N.A.             |
| DRC pre-confirmatory test (mU/L)           | N.A.                         | 3.4 [2.0; 8.3]                     | N.A.             |
| Aldosterone pre-confirmatory test (ng/dL)  | 22.7 [16.5; 31.4]            | 12.0 [7.5; 18.0]                   | <b>&lt;0.001</b> |
| PRA post-confirmatory test (ng/mL/h)       | 0.15 [0.10; 0.30]            | N.A.                               | N.A.             |
| DRC post-confirmatory test (mU/L)          | N.A.                         | 2.0 [2.0; 4.3]                     | N.A.             |
| Aldosterone post-confirmatory test (ng/dL) | 7.5 [3.5; 15.1]              | 5.6 [3.5; 11.4]                    | 0.068            |

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  $\pm$  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           | <b>Training cohort</b> (N = 522)   | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 79.7 |
|                     | PA confirmed                       | 272                 | 50                      | Sensitivity (%) | 84.5 |
|                     | PA <b>not confirmed</b>            | 56                  | 144                     | Specificity (%) | 72.0 |
|                     | <b>Validation cohort</b> (N = 174) | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 77.6 |
|                     | PA confirmed                       | 82                  | 17                      | Sensitivity (%) | 82.8 |
|                     | PA <b>not confirmed</b>            | 22                  | 53                      | Specificity (%) | 70.7 |
|                     | <b>Combined cohort</b> (N = 696)   | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 79.2 |
|                     | PA confirmed                       | 354                 | 67                      | Sensitivity (%) | 84.1 |
|                     | PA <b>not confirmed</b>            | 78                  | 197                     | Specificity (%) | 71.6 |
| RF Model            | <b>Training cohort</b> (N = 522)   | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 82.8 |
|                     | PA confirmed                       | 286                 | 36                      | Sensitivity (%) | 88.8 |
|                     | PA <b>not confirmed</b>            | 54                  | 146                     | Specificity (%) | 73.0 |
|                     | <b>Validation cohort</b> (N = 174) | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 79.9 |
|                     | PA confirmed                       | 84                  | 15                      | Sensitivity (%) | 84.8 |
|                     | PA <b>not confirmed</b>            | 20                  | 55                      | Specificity (%) | 73.3 |
|                     | <b>Combined cohort</b> (N = 696)   | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 82.0 |
|                     | PA confirmed                       | 370                 | 51                      | Sensitivity (%) | 87.9 |
|                     | PA <b>not confirmed</b>            | 74                  | 201                     | Specificity (%) | 73.1 |
| Linear SVM          | <b>Training cohort</b> (N = 522)   | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 80.7 |
|                     | PA confirmed                       | 272                 | 50                      | Sensitivity (%) | 84.5 |
|                     | PA <b>not confirmed</b>            | 51                  | 149                     | Specificity (%) | 74.5 |
|                     | <b>Validation cohort</b> (N = 174) | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 78.2 |
|                     | PA confirmed                       | 82                  | 17                      | Sensitivity (%) | 82.8 |
|                     | PA <b>not confirmed</b>            | 21                  | 54                      | Specificity (%) | 72.0 |
|                     | <b>Combined cohort</b> (N = 696)   | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 80.0 |
|                     | PA confirmed                       | 354                 | 67                      | Sensitivity (%) | 84.1 |
|                     | PA <b>not confirmed</b>            | 72                  | 203                     | Specificity (%) | 73.8 |
| Gaussian SVM        | <b>Training cohort</b> (N = 522)   | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 83.9 |
|                     | PA confirmed                       | 284                 | 38                      | Sensitivity (%) | 88.2 |
|                     | PA <b>not confirmed</b>            | 46                  | 154                     | Specificity (%) | 77.0 |
|                     | <b>Validation cohort</b> (N = 174) | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 74.7 |
|                     | PA confirmed                       | 81                  | 18                      | Sensitivity (%) | 81.8 |
|                     | PA <b>not confirmed</b>            | 26                  | 49                      | Specificity (%) | 65.3 |
|                     | <b>Combined cohort</b> (N = 696)   | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 81.6 |
|                     | PA confirmed                       | 365                 | 56                      | Sensitivity (%) | 86.7 |
|                     | PA <b>not confirmed</b>            | 72                  | 203                     | Specificity (%) | 73.8 |

The table shows real and predicted diagnosis (PA confirmed vs. **not confirmed**), 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. Diagnostic performance of the models at training and validation**

| Performance                         |                     | N   | Cut-off   | Sens (%) | Spec (%) | PPV (%) | NPV (%) | Acc (%) |
|-------------------------------------|---------------------|-----|-----------|----------|----------|---------|---------|---------|
| <b>LDA Model</b>                    | Training            | 522 | N.A.      | 84.5     | 72.0     | 82.9    | 74.2    | 79.7    |
|                                     | Internal Validation | 174 |           | 82.8     | 70.7     | 78.8    | 75.7    | 77.6    |
|                                     | External Validation | 328 |           | 65.9     | 92.3     | 90.5    | 70.8    | 78.4    |
| <b>RF Model</b>                     | Training            | 522 | N.A.      | 88.8     | 73.0     | 84.1    | 80.2    | 82.8    |
|                                     | Internal Validation | 174 |           | 84.8     | 73.3     | 80.8    | 78.6    | 79.9    |
|                                     | External Validation | 328 |           | 77.5     | 67.7     | 72.8    | 72.9    | 72.9    |
| <b>Linear SVM</b>                   | Training            | 522 | N.A.      | 84.5     | 74.5     | 84.2    | 74.9    | 80.7    |
|                                     | Internal Validation | 174 |           | 82.8     | 72.0     | 79.6    | 76.1    | 78.2    |
|                                     | External Validation | 328 |           | 63.6     | 94.8     | 93.2    | 70.0    | 78.4    |
| <b>Gaussian SVM</b>                 | Training            | 522 | N.A.      | 88.2     | 77.0     | 86.1    | 80.2    | 83.9    |
|                                     | Internal Validation | 174 |           | 81.8     | 65.3     | 75.7    | 73.1    | 74.7    |
|                                     | External Validation | 328 |           | 68.2     | 90.3     | 88.7    | 71.8    | 78.7    |
| <b>PACT Score (Best Accuracy)</b>   |                     | 522 | $\geq 8$  | 92.2     | 71.0     | 83.7    | 85.0    | 84.1    |
|                                     | Internal Validation | 174 |           | 91.9     | 73.3     | 82.0    | 87.3    | 83.9    |
|                                     | External Validation | 328 |           | 78.6     | 83.9     | 84.5    | 77.8    | 81.1    |
| <b>PACT Score (Max Sensitivity)</b> |                     | 522 | $\geq 5$  | 100.0    | 23.5     | 67.8    | 100.0   | 70.7    |
|                                     | Internal Validation | 174 |           | 100.0    | 21.3     | 62.7    | 100.0   | 66.1    |
|                                     | External Validation | 328 |           | 100.0    | 28.4     | 60.9    | 100.0   | 66.2    |
| <b>PACT Score (Max Specificity)</b> |                     | 522 | $\geq 13$ | 24.8     | 100.0    | 100.0   | 45.2    | 53.6    |
|                                     | Internal Validation | 174 |           | 21.2     | 100.0    | 100.0   | 49.0    | 55.2    |
|                                     | External Validation | 328 |           | 14.5     | 100.0    | 100.0   | 51.2    | 54.9    |

Sensitivity (Sens), specificity (Spec), positive/negative predictive value (PPV/ NPV), and accuracy (Acc) for proposed models (each indicator is derived considering a confirmed diagnosis of PA as referral category). LDA, Linear Discriminant Analysis; RF, Random Forest; SVM, Support Vector Machine; PACT, Primary Aldosteronism Confirmatory Testing Score. Internal and external validations are provided on Torino and Munich cohorts, respectively.



**Table S7. Score development and validation**

| PACT Score Accuracy                 |                                    | Predicted Diagnosis |                         | Performance     |       |
|-------------------------------------|------------------------------------|---------------------|-------------------------|-----------------|-------|
| Real Diagnosis (Cut-off $\geq 5$ )  | <b>Training cohort (N = 522)</b>   | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 70.7  |
|                                     | PA confirmed                       | 322                 | 0                       | Sensitivity (%) | 100.0 |
|                                     | PA <b>not confirmed</b>            | 153                 | 47                      | Specificity (%) | 23.5  |
|                                     | <b>Validation cohort (N = 174)</b> | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 66.1  |
|                                     | PA confirmed                       | 99                  | 0                       | Sensitivity (%) | 100.0 |
|                                     | PA <b>not confirmed</b>            | 59                  | 16                      | Specificity (%) | 21.3  |
|                                     | <b>Combined cohort (N = 696)</b>   | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 69.5  |
|                                     | PA confirmed                       | 421                 | 0                       | Sensitivity (%) | 100.0 |
|                                     | PA <b>not confirmed</b>            | 212                 | 63                      | Specificity (%) | 22.9  |
| Real Diagnosis (Cut-off $\geq 8$ )  | <b>Training cohort (N = 522)</b>   | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 84.1  |
|                                     | PA confirmed                       | 297                 | 25                      | Sensitivity (%) | 92.2  |
|                                     | PA <b>not confirmed</b>            | 58                  | 142                     | Specificity (%) | 71.0  |
|                                     | <b>Validation cohort (N = 174)</b> | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 83.9  |
|                                     | PA confirmed                       | 91                  | 8                       | Sensitivity (%) | 91.9  |
|                                     | PA <b>not confirmed</b>            | 20                  | 55                      | Specificity (%) | 73.3  |
|                                     | <b>Combined cohort (N = 696)</b>   | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 84.1  |
|                                     | PA confirmed                       | 388                 | 33                      | Sensitivity (%) | 92.2  |
|                                     | PA <b>not confirmed</b>            | 78                  | 197                     | Specificity (%) | 71.6  |
| Real Diagnosis (Cut-off $\geq 13$ ) | <b>Training cohort (N = 522)</b>   | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 53.6  |
|                                     | PA confirmed                       | 80                  | 242                     | Sensitivity (%) | 24.8  |
|                                     | PA <b>not confirmed</b>            | 0                   | 200                     | Specificity (%) | 100.0 |
|                                     | <b>Validation cohort (N = 174)</b> | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 55.2  |
|                                     | PA confirmed                       | 21                  | 78                      | Sensitivity (%) | 21.2  |
|                                     | PA <b>not confirmed</b>            | 0                   | 75                      | Specificity (%) | 100.0 |
|                                     | <b>Combined cohort (N = 696)</b>   | PA confirmed        | PA <b>not confirmed</b> | Accuracy (%)    | 54.0  |
|                                     | PA confirmed                       | 101                 | 320                     | Sensitivity (%) | 24.0  |
|                                     | PA <b>not confirmed</b>            | 0                   | 275                     | Specificity (%) | 100.0 |

The table shows real and predicted diagnosis (PA confirmed vs. **not confirmed**), 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 S8. Correlation of the PACT score with hormonal measurements at confirmatory testing**

| Variable   | Pearson's R | P-value |
|--|-------------|---------|
| Aldosterone pre-confirmatory test (ng/dL)            | 0.479       | <0.001  |
| ARR (PRA) pre-confirmatory test ([ng/dL]/[ng/mL/h])  | 0.247       | <0.001  |
| ARR (DRC) pre-confirmatory test ([ng/dL]/[mU/L])     | 0.415       | <0.001  |
| Aldosterone post-confirmatory test (ng/dL)           | 0.443       | <0.001  |
| ARR (PRA) post-confirmatory test ([ng/dL]/[ng/mL/h]) | 0.251       | <0.001  |
| ARR (DRC) post-confirmatory test ([ng/dL]/[mU/L])    | 0.240       | <0.001  |

Correlations between the PACT score and hormonal measurements at confirmatory testing (aldosterone and ARR pre- and post-test). The ARR (aldosterone-to-renin ratio) is calculated by dividing aldosterone for PRA or DRC. Pearson's R and P-value are reported for each correlation.

**Table S9. Distribution of patients with PA according to the score**

| Score Points | Total (n) | PA not confirmed |       | PA confirmed |       |
|--------------|-----------|------------------|-------|--------------|-------|
|              |           | (n)              | (%)   | (n)          | (%)   |
| 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.  |

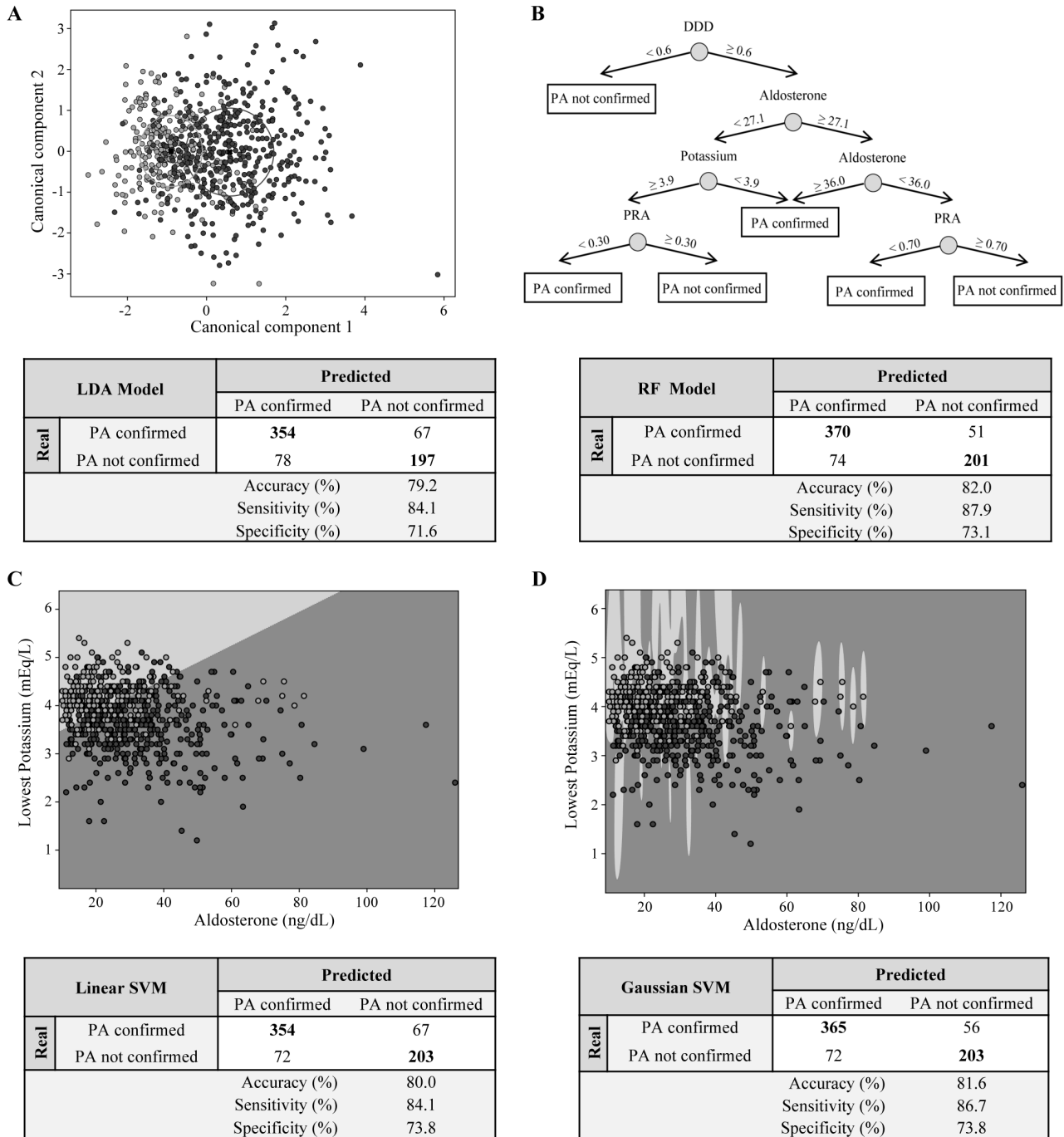
Number (n) and proportion (%) of patients stratified for diagnosis (PA confirmed vs. not confirmed) is shown according to the score in the developmental cohort of Torino (n=696). N.A., Not Applicable.

**Table S10. Diagnostic performance in patients with unilateral PA**

| Performance                        | Accuracy (%) | Missed UPA Patients (n) |
|------------------------------------|--------------|-------------------------|
| LDA Model                          | 95.0         | 11                      |
| RF Model                           | 97.7         | 5                       |
| Linear SVM                         | 95.0         | 11                      |
| Gaussian SVM                       | 96.8         | 7                       |
| PACT Score                         | 98.2         | 4                       |
| Flow chart – 1 <sup>st</sup> model | 100.0        | 0                       |
| Flow chart – 2 <sup>nd</sup> model | 100.0        | 0                       |

The table shows the accuracy of all the proposed models in patients with a diagnosis of unilateral PA (UPA; n=222) and the number of patients with UPA which would be missed by each different approach. See Figure 2 and Figure S3A for flow chart 1<sup>st</sup> and 2<sup>nd</sup> model, respectively. LDA, Linear Discriminant Analysis; RF, Random Forest; SVM, Support Vector Machine; PACT, Primary Aldosteronism Confirmatory Testing Score.

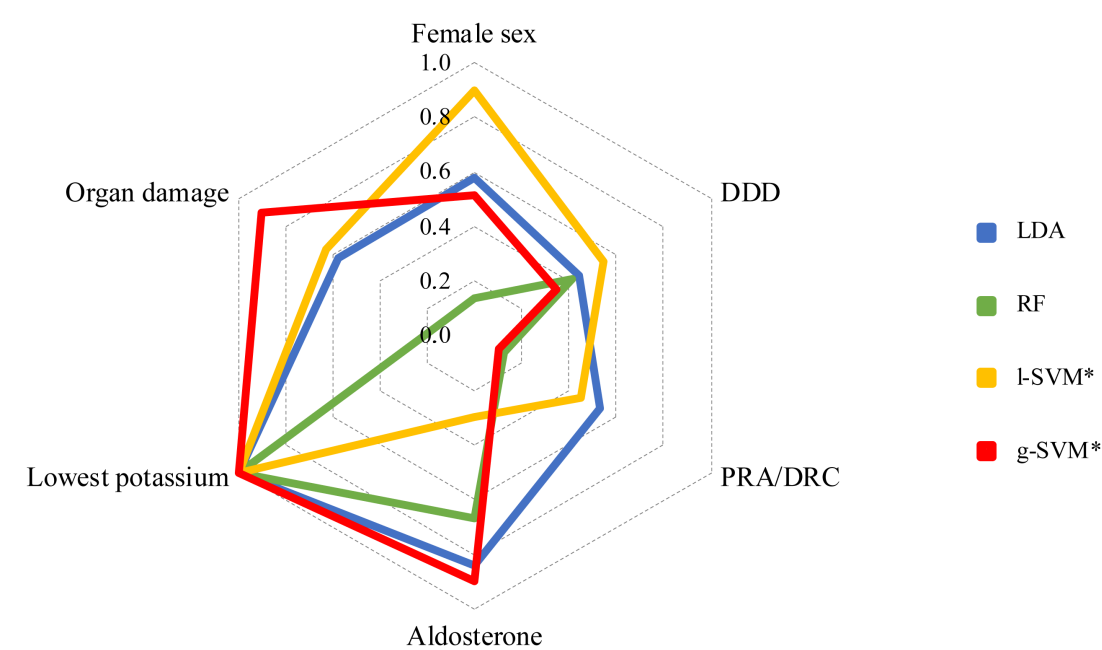
**Figure S1. Diagnostic modelling**



Machine learning based models to discriminate patients with a confirmed diagnosis of PA (n=421) from patients with **not confirmed** PA (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; **not confirmed** PA, 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  $\pm$  SD. (B)

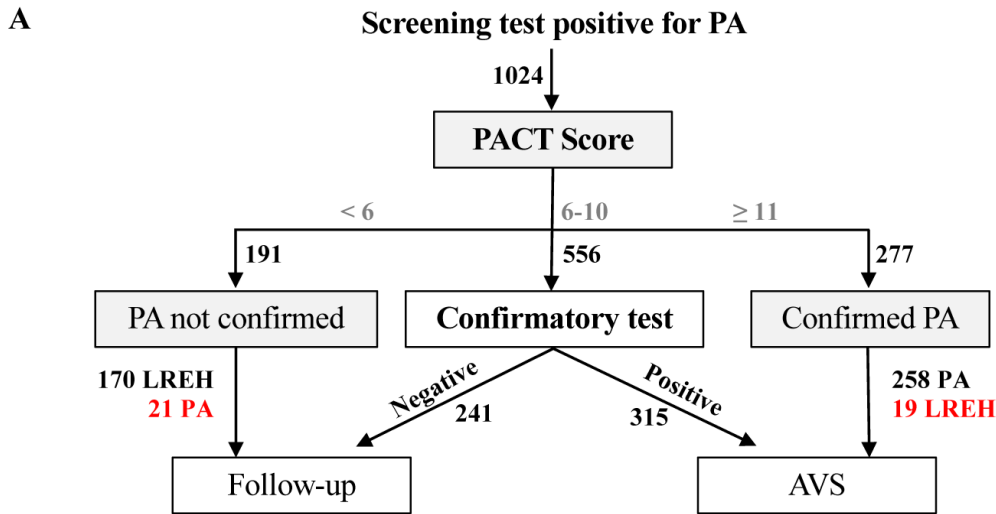
The first classification tree of the forest is shown for the prediction of PA confirmed vs. **not confirmed**. (C, D) Graphs showing the performance of SVM models (Support Vector Machine, Linear and Gaussian). Axes report the two best support vector classifiers: aldosterone at screening on x-axis and lowest recorded potassium levels on y-axis. Each patient is indicated by a point and diagnosis are reported by colour (confirmed PA, dark grey; **not confirmed** PA, grey). Model prediction areas are indicated by colours, as appropriated.

**Figure S2. Predictor importance for machine learning based models**



The graph shows predictor importances for the machine learning based models (coefficients are normalized between 0 and 1). \*For l-SVM and g-SVM, an estimate of predictor importances was computed by recursive feature elimination method. DDD, Defined Daily Dose; PRA, Plasma Renin Activity, DRC, Direct Renin Concentration; LDA, Linear Discriminant Analysis; RF, Random Forest; l-SVM and g-SVM, linear and gaussian Support Vector Machine.

**Figure S3.** Flow chart for the management of patients with PA



**B**

| Diagnosis  |                  | Predicted    |                  | Sensitivity (%) | 96.5 |
|--|------------------|--------------|------------------|-----------------|------|
|  |                  | PA confirmed | PA not confirmed | Specificity (%) | 95.6 |
| Real   | PA confirmed     | 573          | 21               | PPV (%)         | 96.8 |
|  | PA not confirmed | 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)** Management of patients with PA 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.