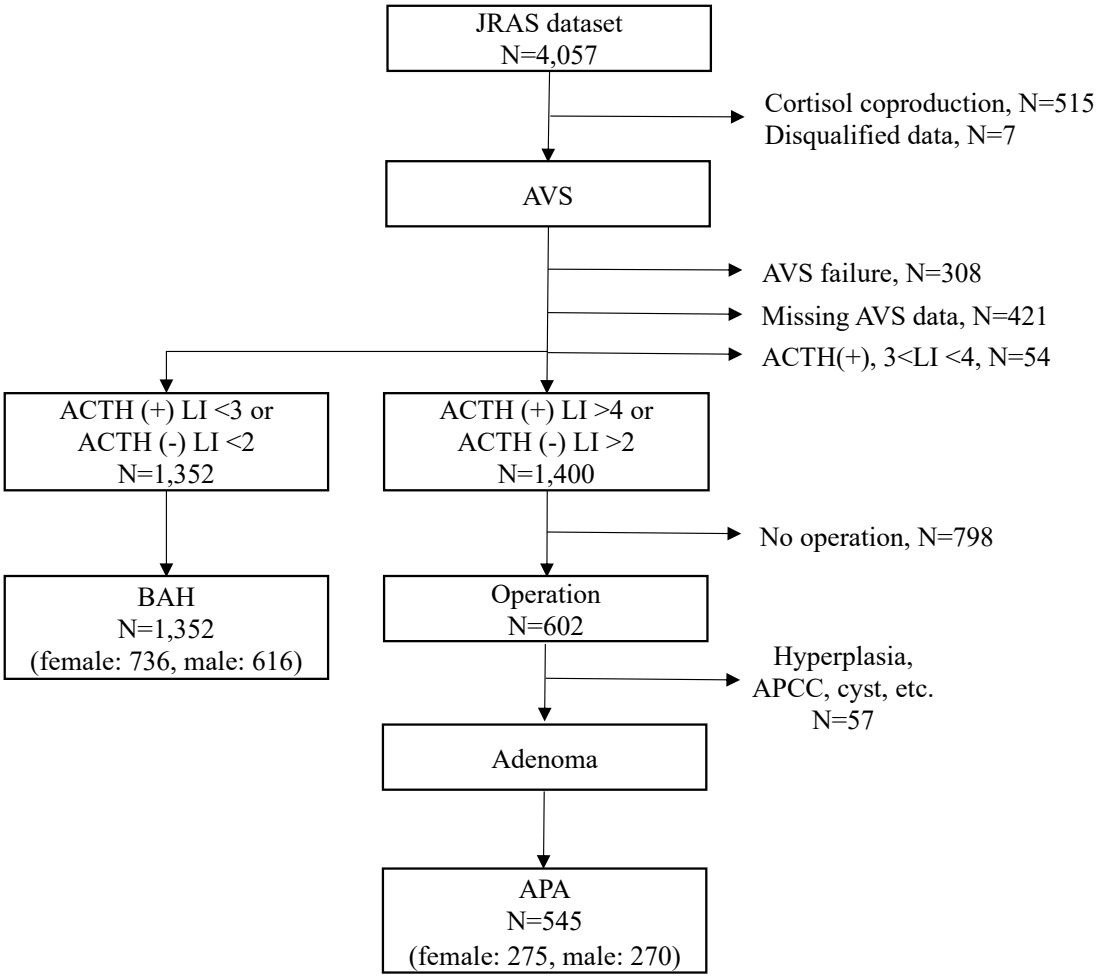
Supplementary Figure

Table of Contents

Figure S1. Definition of APA and BAH in JRAS dataset.

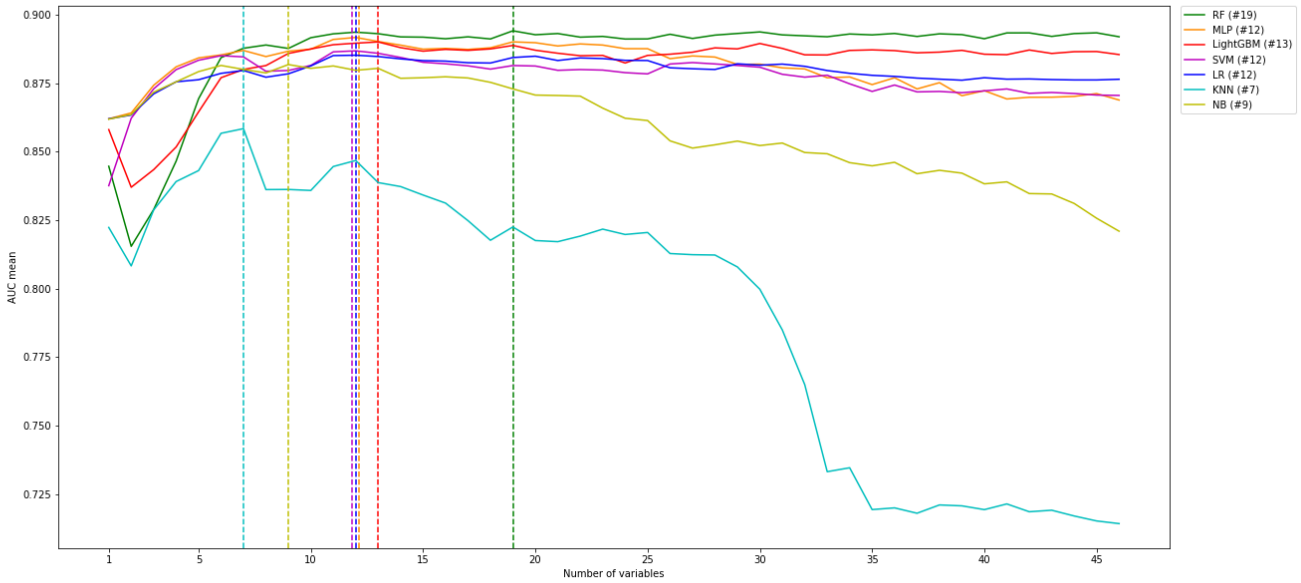


545 of APA patients (female: 275, male: 270) and 1,352 of BAH patients (female: 736, male: 616) were extracted in JRAS dataset.

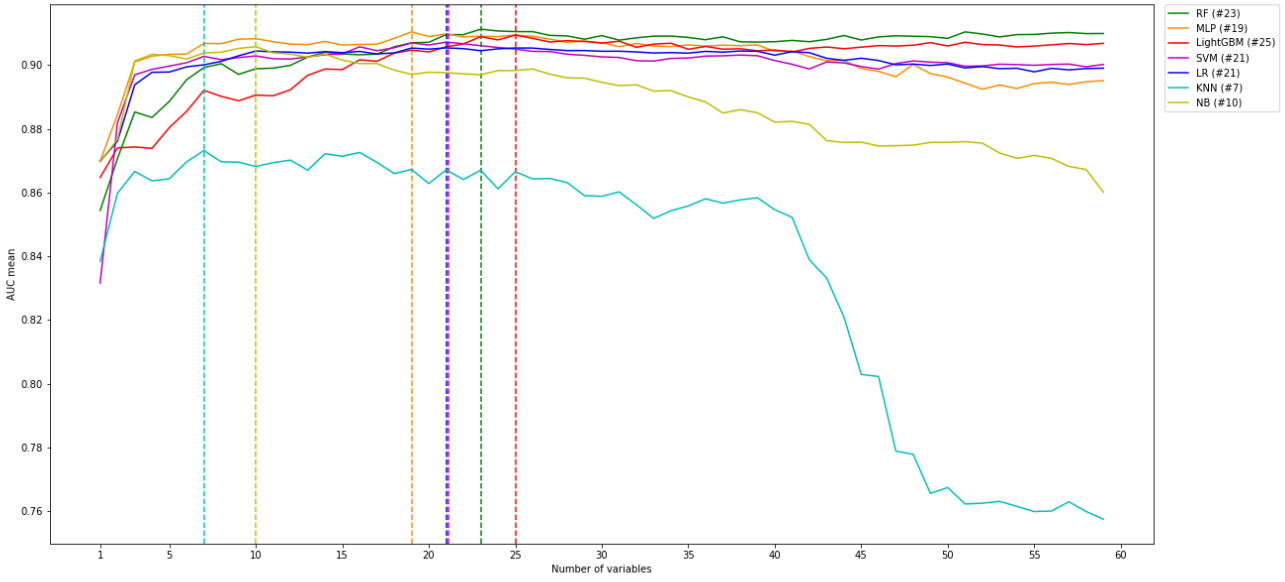
Abbreviation: ACTH, adrenocorticotropic hormone; APA, aldosterone-producing adenoma; APCC, aldosterone-producing cell cluster; AVS, adrenal venous sampling; BAH, bilateral adrenal hyperplasia; JRAS, Japan Rare/Intractable Adrenal Diseases Study; LI, lateralized index.

Figure S2. The relationship between the number of variables and the mean AUC of seven MLAs in two datasets.

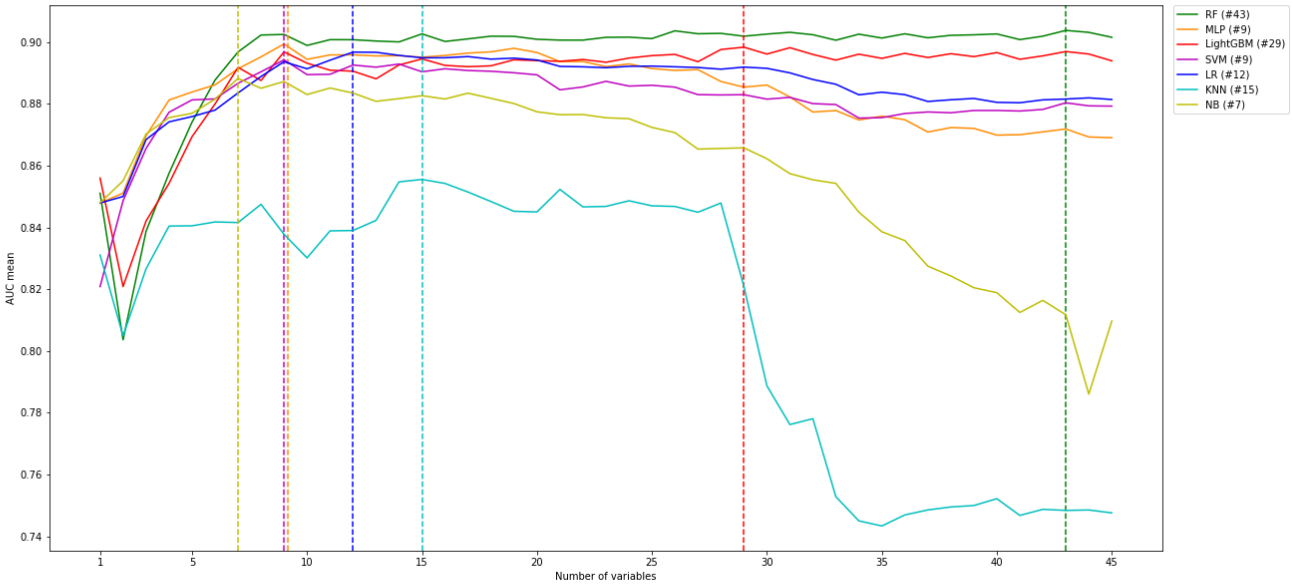
(A) all patients in the screening dataset



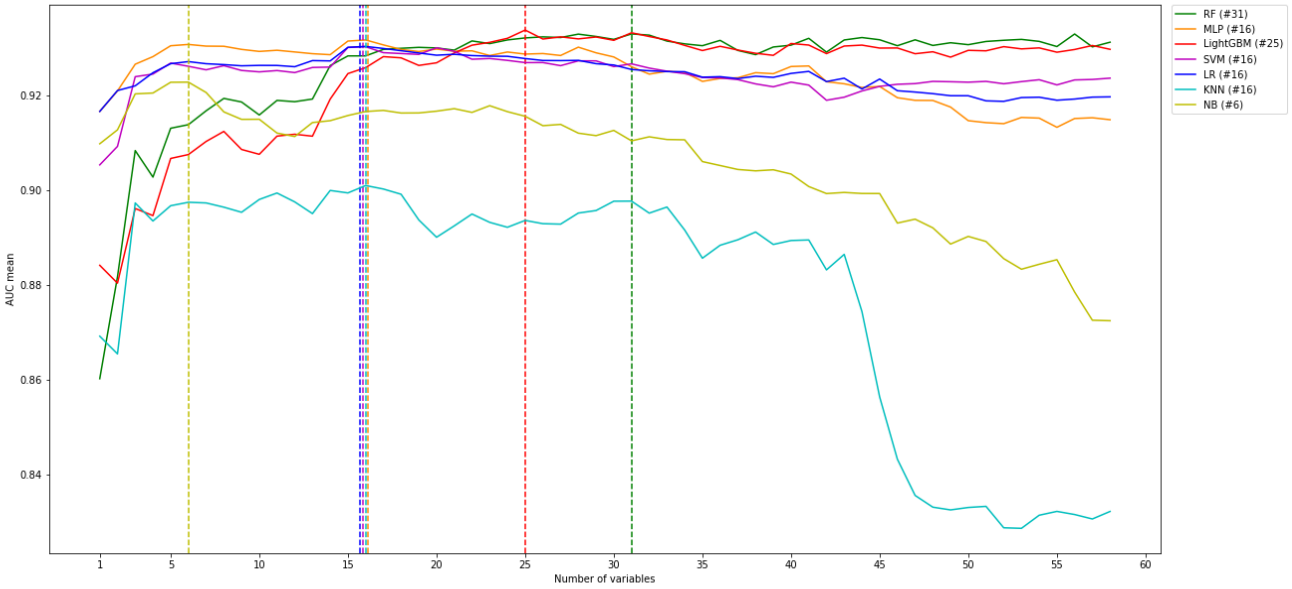
(B) all patients in the confirmatory test dataset



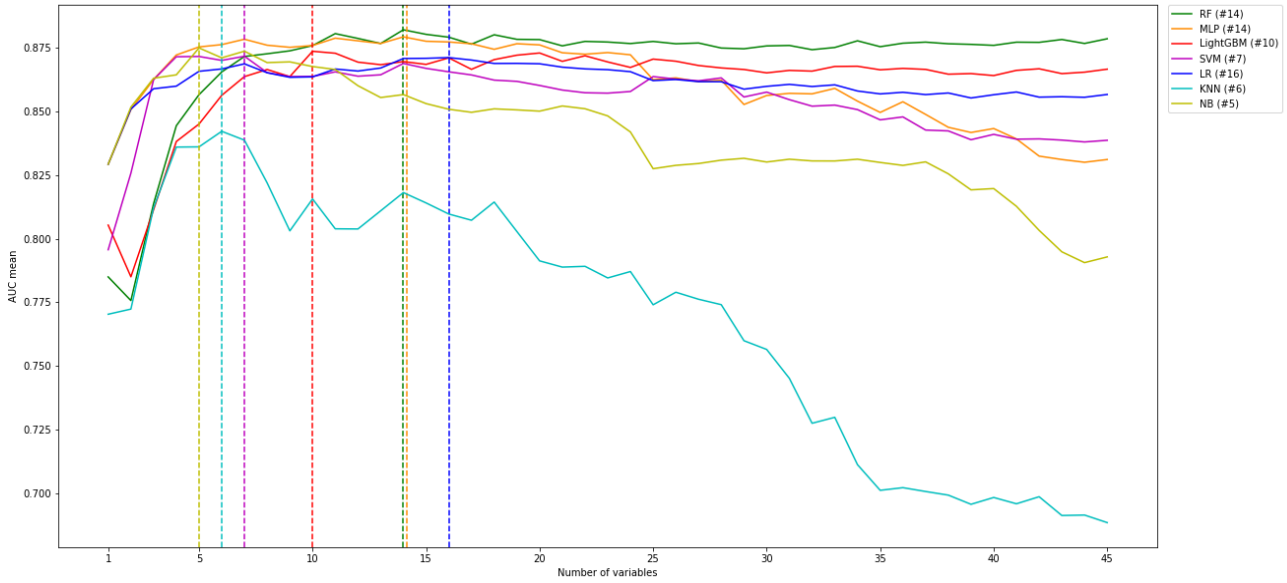
(C) female patients in the screening dataset



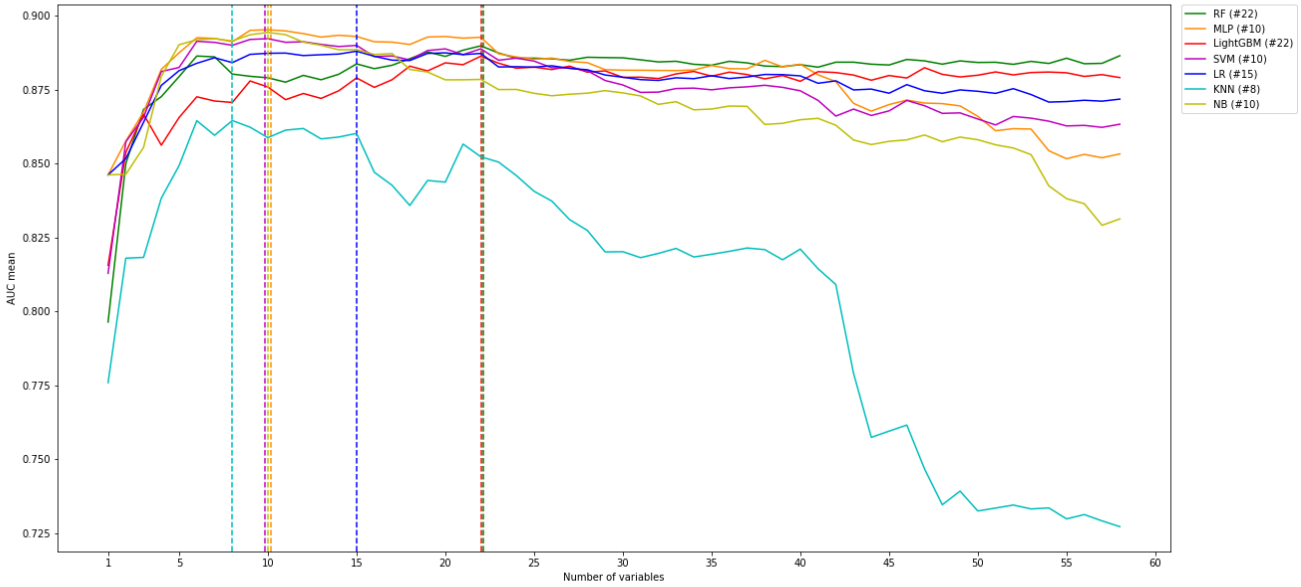
(D) female patients in the confirmatory test dataset



(E) male patients in the screening dataset



(F) male patients in the confirmatory test dataset



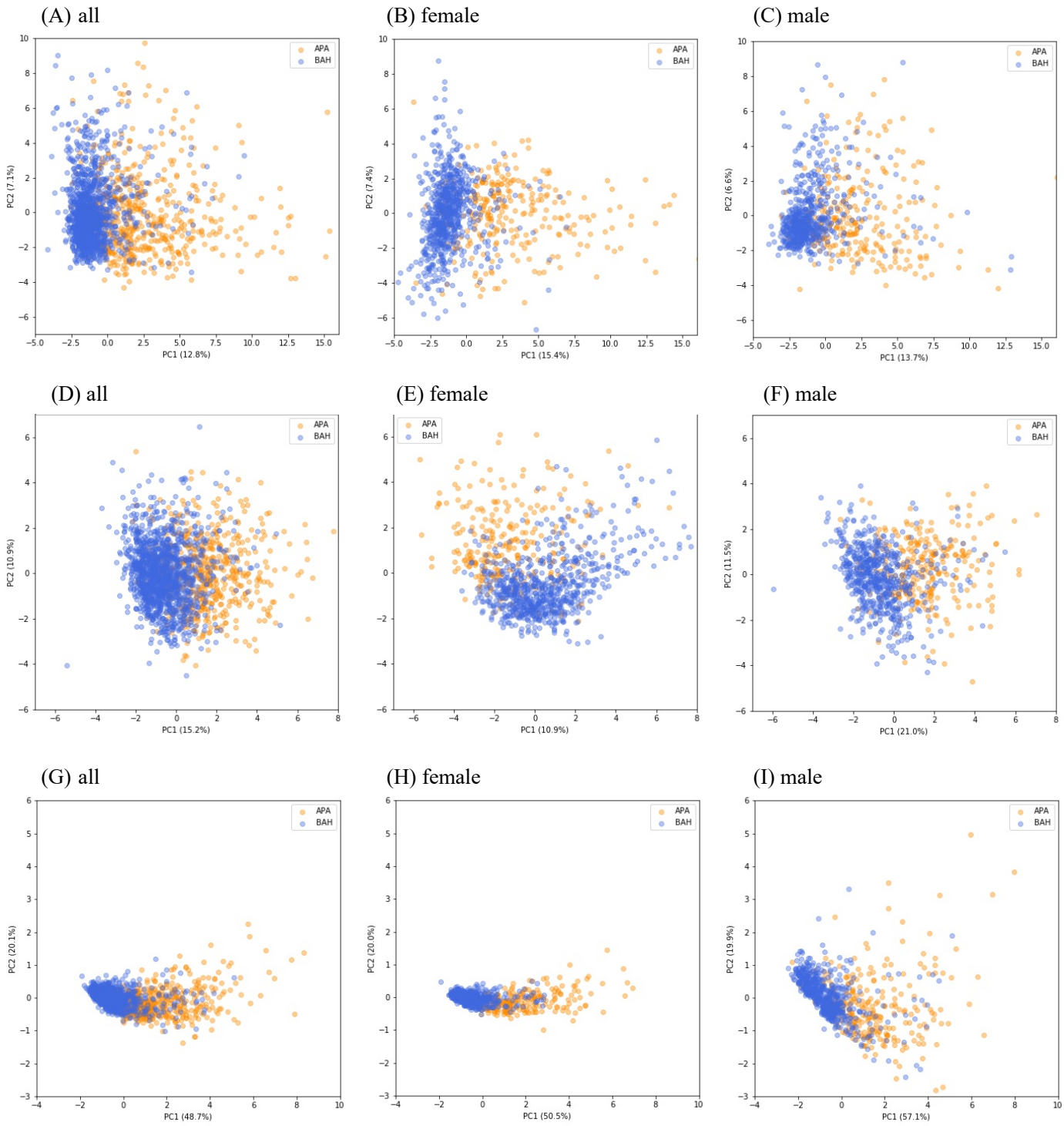
The relationship between the number of variables and the mean AUC of seven MLAs in two datasets.

In the screening dataset, RF, MLP, LightGBM, SVM, LR, KNN, and NB had the best average AUC using (A) 19, 12, 13, 12, 12, 7, and 9 variables (female; (C) 43, 9, 29, 9, 12, 15, and 7 variables, male; (E) 14, 14, 10, 7, 16, 6, and 5 variables), respectively.

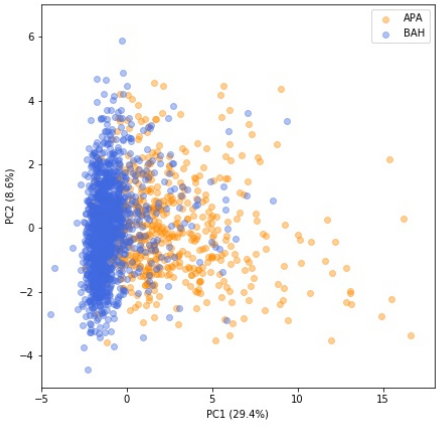
In the confirmatory test dataset, RF, MLP, LightGBM, SVM, LR, KNN, and NB had the best average AUC using (B) 23, 19, 25, 21, 21, 7, and 10 variables (female; (D) 31, 16, 25, 16, 16, 16, and 6 variables, male; (F) 22, 10, 22, 10, 15, 8, and 10 variables), respectively.

Abbreviation: AUC, area under the curve; KNN, k-nearest neighbor algorithm; LightGBM, Light Gradient Boosting Machine; LR, Logistic Regression; MLA, machine-learning algorithm; MLP, multiple-layer perceptron; NB, Naïve Bayes; RF, Random Forest; SVM, Support Vector Machine.

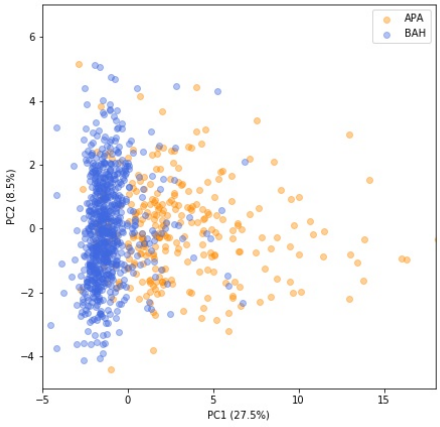
Figure S3. PCA plots of APA and BAH by total patients, female, and male.



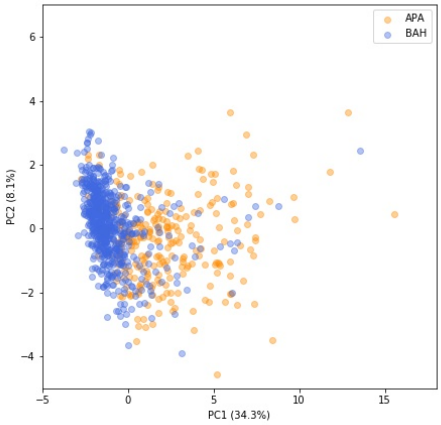
(J) all



(K) female



(L) male



PCA plots for all 59 variables in Confirmatory test data after completion with MissForest (A. overall, B. female, C. male). PCA plots of variables in the screening datasets used in the ELM. Variables were used in descending order of importance ranking in Table 2A. (D. all 1-19th; E. female 1-43rd; F. male 1-16th). PCA plots using only Top 5 variables in the screening datasets (G. all; H. female; I. male). PCA plots of variables from the confirmatory dataset used in the ELM. Variables were used in descending order of importance ranking in Table 2B (J. all 1-25th; K. female 1-31st; L. male 1-22nd).

Abbreviation: APA, aldosterone-producing adenoma; BAH, bilateral adrenal hyperplasia; ELM, ensemble learning model; PCA, principal component analysis.