European Journal of Endocrinology, 2025, **193**, 204–215
https://doi.org/10.1093/ejendo/lvaf145
Advance access publication 29 July 2025
Original Research

Machine learning-based classification of adrenal tumors using clinical, hormonal, and body composition data

Seung Shin Park,[1,2,†] Jongsung Noh,[3,†] Jinhee Kim,[4,†] Taesung Kim,[4] Hae Jin Seo,[3] Chang Ho Ahn,[5,6] Jaegul Choo,[4] Man Ho Choi,[3,‡,*] and Jung Hee Kim[1,2,‡,*]

[1]Department of Internal Medicine, Seoul National University College of Medicine, Seoul 03080, Korea
[2]Department of Internal Medicine, Seoul National University Hospital, Seoul 03080, Korea
[3]Center for Advanced Biomolecular Recognition, Korea Institute of Science and Technology, Seoul 02792, Korea
[4]Graduate School of Artificial Intelligence, Korea Advanced Institute of Science and Technology, Daejeon 34141, Korea
[5]Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam 13620, Korea
[6]Lunit, Seoul 06241, Korea

*Corresponding authors: Department of Internal Medicine, Seoul National University College of Medicine, 101 Daehak-ro, Seoul 03080, Korea. Email: jhee1@snu.ac.kr (J.H.K.); Center for Advanced Biomolecular Recognition, Korea Institute of Science and Technology, 5 Hwarang-ro 14-gil, Seoul 02792, Korea. Email: mh_choi@kist.re.kr (M.H.C.)

†. S.S.P., J.N., and J.K. contributed equally.
‡. M.H.C. and J.H.K. contributed equally.

Abstract

Objective: Accurate diagnosis of adrenal tumors, including mild autonomous cortisol secretion (MACS), adrenal Cushing's syndrome (ACS), primary aldosteronism (PA), pheochromocytoma (PCC), and nonfunctioning adrenal adenomas (NFAs), is crucial but challenging. We aimed to develop a machine learning (ML)-based single-step diagnostic method for differentiating adrenal tumors by integrating clinical data, serum adrenal hormone profiles (SAPs), and body composition data.

Methods: A total of 641 patients with adrenal tumors (MACS = 141, ACS = 64, PA = 265, PCC = 78, and NFA = 93), excluding adrenal metastases and adrenocortical carcinoma, were enrolled from Seoul National University Hospital. Patients were randomly divided into training and test cohorts at a 4:1 ratio. The ML models were developed to differentiate adrenal tumors using 32 clinical data points, 49 SAP markers, and 15 body composition data points.

Results: The best-performing ML model for differentiating all 5 adrenal tumors achieved a balanced accuracy of 0.78, sensitivity of 0.77, specificity of 0.93, and area under the curve (AUC) of 0.89. To distinguish MACS, ACS, PA, and PCC from NFA, the accuracies were 0.85, 0.94, 0.78, and 0.86, with AUCs of 0.96, 0.99, 0.90, and 0.94, respectively. The ML model differentiating between NFA and the other functioning adrenal tumors exhibited an accuracy of 0.75 and an AUC of 0.79. The SAP features were identified as the most critical for differentiation, whereas body composition data contributed only minimally.

Conclusions: The ML model demonstrates high diagnostic accuracy in differentiating adrenal tumor subtypes by integrating clinical data, body composition, and SAP, potentially reducing the need for invasive procedures and aiding clinical decision-making.

Keywords: adrenal gland, hyperaldosteronism, Cushing's syndrome, pheochromocytoma

Significance

The increasing use of computed tomography has led to a rise in adrenal tumor detection, necessitating more efficient diagnostic methods beyond current time-consuming and costly dynamic hormone tests. This study introduces a novel machine learning model that integrates clinical data, serum adrenal hormone profiles, and body composition data to differentiate adrenal tumors—including nonfunctioning adenomas, primary aldosteronism, pheochromocytoma, adrenal Cushing's syndrome, and mild autonomous cortisol secretion. By streamlining diagnosis and optimizing resource utilization, this approach may improve clinical decision-making. If validated in diverse settings, this machine learning-based tool has the potential to enhance accessibility to efficient and cost-effective adrenal tumor management.

Introduction

The incidence of adrenal incidentalomas has increased with the widespread use of advanced imaging techniques, which is estimated to be 1%-10% of abdominal imaging.[1] While most adrenal incidentalomas are benign and nonfunctioning, a significant subset can be hormonally active or malignant, necessitating accurate diagnosis and management.[2,3] The differential diagnosis of adrenal tumors encompasses nonfunctioning adrenal adenomas (NFAs), mild autonomous cortisol secretion (MACS), adrenal Cushing's syndrome (ACS),

Received: February 3, 2025. **Revised:** July 8, 2025. **Editorial Decision:** July 15, 2025
© The Author(s) 2025. Published by Oxford University Press on behalf of European Society of Endocrinology. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Downloaded from https://academic.oup.com/ejendo/article/193/2/204/8217437 by guest on 29 August 2025

primary aldosteronism (PA), and pheochromocytoma (PCC). Distinguishing adrenal tumors is crucial for determining appropriate treatment strategies.

Current diagnostic approaches to differentiate functioning adrenal tumors rely on a sequential combination of dynamic hormone tests.[2] Diagnosing requires repeated testing of aldosterone-to-renin ratios, followed by confirmatory testing, such as saline infusion tests.[4,5] To discriminate ACS from NFAs, two or more of the three tests of cortisol after a 1 mg overnight dexamethasone suppression test (1 mg ODST), late-night cortisol, and 24-h urine cortisol need to be satisfied.[2,6] These multiple tests are time-consuming and often require repeated visits or hospitalization.

Certain clinical indicators may also aid in diagnosis. For instance, hypokalemia and hypertension often raise suspicion of PA or Cushing's syndrome,[7,8] while uncontrolled hypertension, diabetes mellitus (DM), and dyslipidemia are more common in ACS or MACS.[2] Larger adrenal tumors are often seen in PCC.[9] Certain body composition characteristics are associated with specific adrenal diseases. A low body mass index (BMI) is frequently observed in PCC due to catecholamine-induced lipolysis,[10,11] whereas high visceral fat and low muscle mass may be characteristics of ACS.[12]

Steroid profiling by mass spectrometry allows comprehensive measurement of adrenal steroids and metabolites.[13-16] These profiles have proven useful in differentiating adrenal cortical carcinoma from adrenal adenomas and ACS from MACS.[13,15-18] In addition, a recent development involves the simultaneous profiling of adrenal steroids, catecholamines, and metanephrines to characterize metabolic features of adrenal tumors.[19] Thus, incorporating adrenal hormone profiling offers a more nuanced understanding of adrenal diseases. However, interpretation of these profiles can be complex because of the intricate interrelationships between adrenal hormones.

Advances in artificial intelligence, particularly machine learning (ML), offer promising improvements in diagnostic accuracy across various medical fields.[20] The strength of ML lies in its ability to integrate and analyze complex data with interrelated variables. In endocrinology, ML algorithms have been employed to diagnose, predict prognosis, and assist in treating various endocrine disorders, such as neuroendocrine disorders, thyroid disease, DM, and obesity.[21] In the context of adrenal diseases, several attempts have been made to differentiate subtypes of PA,[22] Cushing's syndrome,[23] and hypertension,[24] as well as to distinguish adrenal cortical carcinoma from adrenal cortical adenoma.[14,18] However, the application of ML algorithms in the simultaneous differentiation of various adrenal tumors remains unexplored.

Therefore, this study aimed to develop and validate ML models to differentiate adrenal tumors: NFA, MACS, ACS, PA, and PCC. We hypothesized that ML algorithms that incorporate clinical data, comprehensive serum adrenal hormone profiles (SAPs) by simultaneous measurement of steroids, catecholamines, and metanephrines, and body composition measures would contribute to a single-step, integrated diagnostic algorithm for adrenal tumors.

Material and methods

Study participants

A total of 641 patients with tumors, except adrenal cortical carcinoma, were consecutively enrolled at Seoul National University Hospital between 2018 and 2022: 93 NFA, 141 MACS, 64 ACS, 265 PA, and 78 PCC. As our cohort included only patients with primary adrenal tumors, those with adrenal metastases were not enrolled. Before 2021, patients who had donated samples to the biobank and consented to secondary use were enrolled. From 2021 onward, only patients who consented to participate in this study were enrolled. This study was approved by the International Review Board of Seoul National University Hospital (IRB Nos. 1801-010-911 and H-2105-008-1215) and registered at ClinicalTrials.gov (NCT04948970). All procedures were conducted in accordance with the ethical standards of the Declaration of Helsinki.

Adrenal tumor classification

Adrenal Cushing's syndrome was defined as cortisol > 1.8 μg/dL after the 1 mg ODST with clinical features of Cushing's syndrome (moon face, buffalo hump, central obesity, etc).[2] When available, additional biochemical parameters such as 24-h urinary free cortisol, midnight serum cortisol, and plasma ACTH levels were also considered in the diagnostic evaluation. Mild autonomous cortisol secretion was defined as cortisol > 1.8 μg/dL after 1 mg ODST without Cushingoid features.[2] Primary aldosteronism was defined as plasma aldosterone concentration (PAC) ≥ 6 ng/dL in the saline loading test in the seated position.[4] Pheochromocytoma was defined as cases with elevated serum or 24-h urine metanephrine or normetanephrine (NMN) levels that were pathologically confirmed as PCC after mass excision.[25] Nonfunctioning adrenal adenoma was defined as a benign adrenal mass, excluding ACS, MACS, PA, and PCC.

Plasma aldosterone concentration was measured using the SPAC-S Aldosterone Kit (TFB Inc.), and plasma renin activity (PRA) was measured using the PRA RIA Kit (TFB Inc.). Serum cortisol levels were measured using a radioimmunoassay kit (CIS Bio International, France). Serum metanephrine and NMN levels were measured at our hospital using an in-house liquid chromatography-tandem mass spectrometry (LC-MS/MS). These hormonal data were used only for diagnosis and excluded from ML model input features.

In this study, we aimed to distinguish between 7 different disease combinations. The following combinations were used: (1) All vs All (differentiating between NFA, MACS, ACS, PA, and PCC); (2) NFA vs others (NFA vs [MACS, ACS, PA, and PCC]); (3) NFA vs MACS; (4) NFA vs ACS; (5) NFA vs PA; (6) NFA vs PCC; and (7) NFA vs MACS vs ACS. The ML models were developed for each disease combination, and the prediction outcomes of each ML model were obtained.

Clinical data

A total of 32 clinical features were assessed in this study. Demographic data such as age, sex, BMI, and medical history, including menstruation status, hypertension, DM, dyslipidemia, cardiovascular disease (CVD), cerebrovascular accident, cancer, and previous fracture history, were included as binary variables. Radiological information from the abdominal computed tomography (CT) scans (mass size and site) was also obtained. Additionally, laboratory values including white blood cell count, hemoglobin, hematocrit, platelet count, neutrophil percentage, lymphocyte, eosinophil, absolute neutrophil count, aspartate aminotransferase, alanine aminotransferase, total cholesterol, total protein, albumin, blood urea nitrogen,

Downloaded from https://academic.oup.com/ejendo/article/193/2/204/8217437 by guest on 29 August 2025

creatinine, uric acid, sodium, potassium, and chloride levels were measured.

Serum cortisol levels after 1 mg ODST, serum or 24-h urine metanephrine, and NMN, PAC, and PRA were excluded from ML input features.

Body composition data

Body composition was evaluated using the body composition analysis DeepCatch software (https://medicalip.com/DeepCatch; DeepCatch, version 1.0.0.0; MEDICALIP Co., Ltd., Seoul, Korea).[26] Contrast-enhanced abdominal CT scans from the portal venous phase at the time of diagnosis were used to automatically detect the third lumbar vertebral body by measuring skeletal muscle area, visceral fat area, subcutaneous fat area (SFA), skeletal muscle volume, visceral fat volume, and subcutaneous fat volume. Additionally, the mean CT attenuation (Hounsfield units [HU]) of skeletal muscle, visceral fat, and subcutaneous fat was measured. Each volume and area parameter was adjusted for BMI, resulting in 15 derived variables utilized in this study.

Quantitative profiling of serum adrenal hormones

Simultaneous profiling of adrenal steroids, catecholamines, and metanephrines with a small serum sample volume (400 µL) was performed using a validated LC-MS assay.[19] This assay quantified 49 adrenal steroid analytes, including glucocorticoids, mineralocorticoids, and androgen precursors. The detailed steroid profiling process is described in Methods S1.

Developing a ML model for differentiating adrenal tumors

A schematic of the ML process for processing, feature selection, and model evaluation is shown in Figure 1. The dataset, data preprocessing, methodology development, and final model training and testing for ML model development to differentiate adrenal tumors are described in Figure 1 and Methods S1.

We allocated 80% (512 patients) for training and 20% (129 patients) for testing, maintaining class balance. The training data were further split into 80% (409 patients) for training

Figure 1. Schematic flow of machine learning process, including data processing, feature selection, and model evaluation. The process includes 4 main stages: model selection, outlier detection, feature selection, and resampling strategy. The dataset ($n = 641$) was split into a training set ($n = 512$) and a test set ($n = 129$) with class-balanced randomization. Input features included 32 clinical variables, 49 SAPs (LC-MS), and 15 CT-based body composition measures. The training data underwent preprocessing, model selection (from 8 classifiers), outlier handling, feature selection (Boruta or CFS), and resampling (SMOTETomek, BorderlineSMOTE, or ClusterCentroids). The final models, based on the top 3 classifiers (Easy Ensemble, Gradient Boosting, Random Forest), were trained on the full training cohort and evaluated on the test cohort across 28 classification tasks. ACS, adrenal Cushing's syndrome; CFS, correlation-based feature selection; CT, computed tomography; IQR, interquartile range; LC-MS, liquid chromatography-mass spectrometry; MACS, mild autonomous cortisol secretion; ML, machine learning; NFA, nonfunctioning adrenal adenoma; PA, primary aldosteronism; PCC, pheochromocytoma. This figure was created using BioRender.com.

Downloaded from https://academic.oup.com/ejendo/article/193/2/204/8217437 by guest on 29 August 2025

and 20% (103 patients) for validation, maintaining class balance.

The ML model development involved 4 steps: model selection, outlier detection, feature selection, and resampling optimization, evaluated 100 random iteration with balanced accuracy as the primary metric.

1. Model selection: We first evaluated 8 different classifiers for distinguishing all types of adrenal tumors (All vs All) and selected the 3 best-performing classifiers: Easy Ensemble Classifier (EE), Random Forest Classifier (RF), and Gradient Boosting Classifier (GB). For each model, we optimized the outlier detection, feature selection, and resampling algorithms (Figure S1).
2. Outlier detection: Outliers were handled using 2 methods—$3 \times 1.5 \times$ interquartile range (IQR) and $1.5 \times$ IQR. The $3 \times 1.5 \times$ IQR method outperformed the alternatives and was applied in subsequent analyses (Figure S1A).
3. Feature selection: 3 methods—default (using all variables), Boruta, and correlation-based feature selection (CFS)—were tested. Features appearing more than 50 times in 100 random iterations were retained, with Boruta yielding the best results (Figure S1B). Boruta identifies important features by comparing them to random noise features via a random forest.
4. Resampling optimization: To mitigate class imbalance, SMOTETomek (a combination of over- and under-sampling), Borderline-SMOTE (over-sampling), and ClusterCentroids (under-sampling) were evaluated. The optimal strategy was determined for each classifier: no resampling for EE, SMOTETomek for GB, and Borderline-SMOTE for RF (Figure S1C).

Using the optimized algorithms, we conducted classification analyses for 7 disease combinations: All-All, NFA-Others, NFA-PA, NFA-PCC, NFA-ACS, NFA-MACS, and NFA-MACS-ACS. Each combination used multiple parameters, clinical data, body composition, and SAP as features, resulting in 28 tasks for each. Each task was performed using the 3 best-performing classifiers (EE, RF, and GB). For each task, the model that demonstrated the highest balanced accuracy in the validation set was selected, trained on the entire training cohort, and validated in the test cohort.

After optimization, the final model was trained on the full training set of 512 patients and evaluated on the independent test set of 129 patients. Performance metrics included balanced accuracy, sensitivity, specificity, precision, area under the receiver operating characteristicscurve (AUC), *F*1 score, and kappa coefficient.

Statistical analysis

The clinical characteristics were compared among the 5 groups. For continuous variables, analysis of variance (ANOVA) was used and presented as mean ± standard deviation (SD). For categorical variables, the χ^2 test was used, and the results are presented as numbers (percentages). A receiver operating characteristic (ROC) curve analysis was conducted to evaluate the diagnostic performance of the developed ML model. The balanced accuracy, precision, sensitivity, specificity, AUC, and *F*1 score were calculated as performance matrices. SHapley Additive exPlanations (SHAP) values were calculated to understand the contributions of the features in our ML model.[27] Statistical significance was set at $P < .05$. All statistical analyses were performed using the R software (version 4.1; R Foundation for Statistical Computing, Vienna, Austria).

Results

Baseline characteristics of study subjects

The baseline characteristics of the study participants ($N = 641$) are shown in Table 1. The mean age was 54.6 years, with MACS patients being the oldest (58.7 years) and those with PCC the youngest (49.7 years). Sex distribution also varied significantly among the groups ($P < .001$), with ACS having the highest proportion of females (87.5%) and NFA having the lowest (43.0%). Pheochromocytoma patients had the lowest BMI (23.1 ± 3.6 kg/m²), while PA patients presented the highest BMI (26.1 ± 4.3 kg/m²). Hypertension was most prevalent in patients with PA (93.6%), while the prevalence of dyslipidemia ranged from 16.7% in PCC to 45.3% in ACS ($P = .001$). Tumor size ranged from 1.6 cm in PA to 3.7 cm in PCC, with the largest observed in PCC. The SAPs of study subjects were shown in Table S1.

Performance metrics of ML algorithm

The performance metrics of the optimal classifier for each task in the test cohort were evaluated (Figure 2; Figure S2). In the All-All combination task, the EE classifier using multiple parameters showed the best performance, with a balanced accuracy of 0.79, AUC of 0.89, sensitivity of 0.77, and specificity of 0.93. To distinguish NFAs from other conditions (NFA-others), the GB classifier with multiple parameters exhibited the best performance, with a balanced accuracy of 0.75, AUC of 0.79, and sensitivity of 0.93. Interestingly, for the NFA-PA and NFA-PCC classifications, using SAP profiles alone showed better accuracy than using multiple parameters. Specifically, for NFA-PA, the RF model with SAP achieved a balanced accuracy of 0.78 and an AUC of 0.86. Similarly, for NFA-PCC, the RF model with SAP showed a balanced accuracy of 0.78 and an AUC of 0.78.

Among all binary classifications, the models distinguishing NFA from ACS showed the highest performance. The EE classifier with multiple parameters demonstrated exceptional results with a balanced accuracy of 0.92, AUC of 0.99, sensitivity of 1.00, and specificity of 0.84. This was the best-performing model among the 28 adrenal tumor-discriminating tasks. For NFA-ACS, among the single parameter sets, clinical data and SAP both achieved balanced accuracies of 0.87, with AUCs of 0.96 and 0.98, respectively. In the NFA-MACS classification, SAP alone performed best among single parameter sets with a balanced accuracy of 0.80 and an AUC of 0.94. For NFA-MACS, the RF classifier with multiple parameters achieved a balanced accuracy of 0.85, an AUC of 0.96, a sensitivity of 0.96, and a specificity of 0.74. For the classification of NFA-MACS-ACS, the GB classifier with multiple parameters demonstrated superior performance with an accuracy of 0.81 and an AUC of 0.94, outperforming the use of single parameter sets. The ROC curves for 5 binary tasks out of the 7 classification tasks are shown in Figure S3.

Selected features for discriminating adrenal tumors

Serum adrenal hormone profiles contributed >70% of the selected features across tasks, whereas body composition data

Downloaded from https://academic.oup.com/ejendo/article/193/2/204/8217437 by guest on 29 August 2025

Table 1. Baseline characteristics of study subjects.

	NFA ($N = 93$)	PCC ($N = 78$)	PA ($N = 265$)	ACS ($N = 64$)	MACS ($N = 141$)	Total ($N = 641$)	*P*-value
Age (years)	56.6 ± 11.6	49.7 ± 15.2	51.9 ± 11.7	50.1 ± 13.9	58.7 ± 10.5	53.6 ± 12.6	<.001
Females, *N* (%)	40 (43.0)	36 (46.2)	137 (51.7)	56 (87.5)	77 (54.6)	346 (54.0)	<.001
BMI (kg/m^2)	25.6 ± 3.4	23.1 ± 3.6	26.1 ± 4.3	25.1 ± 3.7	25.9 ± 3.8	25.5 ± 4.0	<.001
HTN, *N* (%)	48 (51.6)	46 (59.0)	248 (93.6)	36 (56.2)	71 (50.4)	449 (70.0)	<.001
DM, *N* (%)	16 (17.2)	24 (30.8)	47 (17.7)	13 (20.3)	36 (25.5)	136 (21.2)	.070
Dyslipidemia, *N* (%)	25 (26.9)	13 (16.7)	64 (24.2)	29 (45.3)	50 (35.5)	181 (28.2)	.001
Mass size (cm)	2.0 ± 2.1	3.7 ± 3.0	1.6 ± 1.1	2.6 ± 1.2	2.1 ± 0.8	2.1 ± 1.7	<.001
ODST (nmol/L)	33.8 ± 9.7	74.9 ± 66.5	66.4 ± 46.8	440.7 ± 194.5	125.4 ± 83.8	113.1 ± 139.6	33.8 ± 9.7
PAC (nmol/L)	0.48 ± 0.22	0.68 ± 0.60	1.01 ± 0.89	0.48 ± 0.32	0.50 ± 0.31	0.73 ± 0.68	<.001
PRA (ng/mL/hr)	12.2 ± 92.3	17.7 ± 84.3	0.7 ± 2.3	1.4 ± 1.8	2.6 ± 3.5	4.9 ± 46.1	.023
Metanephrine (pmol/L)	1879.8 ± 2091.4	17 945.4 ± 28 283.9	1405.4 ± 615.6	987.5 ± 469.5	1691.1 ± 2460.8	3508.0 ± 11 285.4	<.001
Normetanephrine (pmol/L)	6387.5 ± 13 161.8	41 490.8 ± 35 167.7	3305.7 ± 1380.7	2327.5 ± 1012.9	4336.2 ± 7948.2	8528.4 ± 18 452.8	<.001

Continuous variables are presented as mean ± SD and were compared using ANOVA. Categorical variables are presented as number (percentage) and were compared using the χ^2 test. *P*-values indicate the overall difference among the groups.
Abbreviations: ACS, adrenal Cushing's syndrome; BMI, body mass index; DM, diabetes mellitus; HTN, hypertension; MACS, mild autonomous cortisol secretion; NFA, nonfunctioning adenoma; ODST, overnight dexamethasone suppression test; PA, primary aldosteronism; PAC, plasma aldosterone concentration; PCC, pheochromocytoma; PRA, plasma renin activity.

were rarely used. The selected important features are listed in Table 2 and illustrated in Figure 3. The All-All model with multiple parameters included 49 features: BMI, aldosterone (Aldo), corticosterone (B), 11-deoxycortisol (11-deoxyF), tetrahydro-11-deoxycortisol (THS), 20α-dihydrocortisol (20α-DHF), 18-hydroxycortisol (18-OHF), 21-deoxycortisol (21-deoxyF), cortisol (F), cortisone (E), 6β-hydroxycortisol (6β-OHF), tetrahydrocortisol (THF), tetrahydrocortisone (THE), 11-hydroxyandrostenedione (11-OHA4), dehydroepiandrosterone sulfate (DHEA-S), norepinephrine (NEP), NMN, maximal diameter, hematocrit, neutrophils, lymphocytes, eosinophils, serum protein, albumin, and potassium.

Figure 4 illustrates the SHAP values of the 20 most important features for differentiating between adrenal tumors. In the Other tasks, among the SAP, low levels of 18-OHF, Aldo, THE, and 21-deoxyF, as well as elevated 11-OHA4, were suggestive of NFA. Among the clinical data, an elevated neutrophil count was indicative of functional tumors belonging to the "others" category. In the NFA-PA task, high levels of 18-OHF and Aldo, as well as low levels of potassium, 11-OHA4, and NMN, were suggestive of PA. In the NFA-PCC task, high levels of NEP, NMN, and 20α-DHF; low Aldo/20α-DHF; large maximal tumor diameter; and low BMI were suggestive of PCC. In the NFA-MACS model, high levels of THE/DHEA-S, THE, and other cortisol metabolism-related features such as Aldo and E, as well as low levels of NMN, were suggestive of MACS. For the NFA-ACS classification, higher levels of cortisol metabolites such as THE, 6β-OHF, and 6β-OHF/DHEA-S were suggestive of ACS. Among the clinical features, higher neutrophil and lower lymphocyte counts were suggestive of ACS.

Discussion

We developed ML models to differentiate 5 adrenal tumors (NFA, PA, PCC, ACS, and MACS) and validated them in an independent test cohort ($n = 129$). Seven disease combinations were examined: All-All, NFA-others, NFA-PA, NFA-PCC, NFA-ACS, NFA-MACS, and NFA-MACS-ACS. The ML model utilized multiple parameters, including clinical, SAP, and body composition data. We evaluated the performance of each ML model for these tasks and identified the key features for each classification. The EE classifier using multiple parameters showed the best performance for the All-All combination, which was differentiated among the 5 diseases. Serum adrenal hormone profile has emerged as the most frequently utilized data, while body composition is rarely used to distinguish diseases.

To the best of our knowledge, this is the first ML model that differentiates all functional adrenal tumors using clinical data, SAP, and body composition. Our model to simultaneously differentiate all 5 types of adrenal tumors showed comparable performance with a balanced accuracy of 0.79 and an AUC of 0.89. This suggests the potential utility of this model in assisting the differentiation of adrenal tumor subtypes, which could support appropriate clinical decisions. The GB classifier, using multiple parameters, showed excellent performance in discriminating NFAs from other diseases with a high sensitivity of 0.93. This is meaningful given that NFAs are the most common adrenal incidentalomas and do not require further evaluation.

Among the tasks of differentiating NFAs from each adrenal tumor, distinguishing NFA from ACS showed the highest

Downloaded from https://academic.oup.com/ejendo/article/193/2/204/8217437 by guest on 29 August 2025

Disease	Type	Classifier	Performance Metrics					
			Balanced accuracy	Precision	Sensitivity	Specificity	AUC	F1
All-All	Multiple parameter	EE	0.79	0.77	0.77	0.93	0.89	0.77
	Clinical data	EE	0.58	0.57	0.63	0.90	0.82	0.59
	Body composition	RF	0.23	0.27	0.22	0.80	0.52	0.22
	SAP	RF	0.71	0.69	0.68	0.90	0.89	0.68
NFA vs others	Multiple parameter	GB	0.75	0.93	0.93	0.58	0.79	0.93
	Clinical data	EE	0.61	0.90	0.59	0.63	0.64	0.71
	Body composition	GB	0.48	0.85	0.75	0.21	0.44	0.80
	SAP	EE	0.68	0.91	0.78	0.58	0.73	0.84
NFA vs PA	Multiple parameter	RF	0.76	0.89	0.77	0.74	0.89	0.83
	Clinical data	RF	0.75	0.88	0.81	0.68	0.91	0.84
	Body composition	GB	0.51	0.74	0.66	0.37	0.46	0.70
	SAP	RF	0.78	0.91	0.77	0.79	0.86	0.84
NFA vs PCC	Multiple parameter	GB	0.75	0.67	0.88	0.63	0.91	0.76
	Clinical data	EE	0.66	0.75	0.60	0.58	0.78	0.67
	Body composition	GB	0.48	0.44	0.44	0.53	0.35	0.44
	SAP	RF	0.78	0.88	0.70	0.68	0.90	0.78
NFA vs ACS	Multiple parameter	EE	0.92	0.81	1.00	0.84	0.99	0.90
	Clinical data	EE	0.87	0.72	1.00	0.74	0.96	0.84
	Body composition	RF	0.44	0.33	0.31	0.58	0.35	0.32
	SAP	EE	0.87	0.72	1.00	0.74	0.98	0.84
NFA vs MACS	Multiple parameter	RF	0.85	0.84	0.96	0.74	0.96	0.90
	Clinical data	EE	0.52	0.62	0.57	0.47	0.59	0.59
	Body composition	GB	0.42	0.53	0.57	0.26	0.44	0.55
	SAP	GB	0.80	0.79	0.96	0.63	0.94	0.87
NFA vs MACS vs ACS	Multiple parameter	GB	0.81	0.82	0.82	0.87	0.94	0.82
	Clinical data	EE	0.69	0.67	0.65	0.83	0.77	0.65
	Body composition	EE	0.33	0.33	0.30	0.67	0.51	0.30
	SAP	GB	0.69	0.67	0.65	0.83	0.77	0.65

Figure 2. Test cohort performance metrics for the optimal classifier trained using multiple and clinical data, body composition, and serum steroid profiles. Balanced accuracy, precision, sensitivity, specificity, AUC, and *F*1 score are reported for each classification task across different input types: clinical data, body composition measures, SAPs, and their combination (multiple parameter). The best-performing classifier for each setting is presented among EE, GB, and RF. Classification tasks include All-All and specific pairwise disease differentiations (eg, NFA vs MACS, NFA vs ACS). ACS, adrenal Cushing's syndrome; AUC, area under the curve; EE, Easy Ensemble Classifier; GB, Gradient Boosting Classifier; MACS, mild autonomous cortisol secretion; NFA, nonfunctioning adenoma; PA, primary aldosteronism; PCC, pheochromocytoma; RF, Random Forest Classifier; SAP, serum adrenal hormone profile.

Table 2. Features selected by Boruta methods for each disease comparison among clinical data, body composition, and serum steroid profiles.

	Comparison	ML model	Feature number	Features
Multiple parameter	All vs All	EE	49	BMI, Aldo, B, 11-deoxyF, THS, 20α-DHF, 18-OHF, 21-deoxyF, F, E, 6β-OHF, THF, THE, 11-OHA4, DHEA-S, NEP, NMN, F/NEP, F/NMN, B/NEP, B/NMN, F/B, 11-deoxyF/17-OHP4, 21-deoxyF/17-OHP4, F/11-deoxyF, 6β-OHF/F, THE/E, 11-OHA4/A4, T/A4, DHEA-S/P5-S, THS/11-deoxyF, THS/DHEA-S, 21-deoxyF/DHEA-S, 6β-OHF/DHEA-S, THE/DHEA-S, B/20α-DHF, 18-OHF/20α-DHF, Aldo/20α-DHF, NEP/11-deoxyF, NMN/11-deoxyF, Max_diameter, HCT, Neutrophil_%, Lymphocyte, Eosinophil, ANC, TP, Albumin, K
	NFA vs Others	GB	13	Aldo, B, 11-deoxyF, 20α-DHF, 18-OHF, 21-deoxyF, F, E, 6β-OHF, THF, THE, 11-OHA4, Neutrophil_%
	NFA vs PA	RF	20	Aldo, B, 11-deoxyF, 20α-DHF, 18-OHF, 21-deoxyF, F, E, 6β-OHF, THF, THE, 11-OHA4, NMN, F/NEP, F/NMN, B/NEP, 18-OHF/20α-DHF, Aldo/20α-DHF, NEP/11-deoxyF, K

(*continued*)

Downloaded from https://academic.oup.com/ejendo/article/193/2/204/8217437 by guest on 29 August 2025

Table 2. Continued

	Comparison	ML model	Feature number	Features
	NFA vs PCC	GB	15	BMI, 20α-DHF, F, E, 6β-OHF, THF, NEP, NMN, F/NEP, F/11-deoxyF, 18-OHF/20α-DHF, Aldo/20α-DHF, Max_diameter, Neutrophil_%, Lymphocyte
	NFA vs ACS	EE	27	B, THS, 20α-DHF, 21-deoxyF, F, E, 6β-OHF, THF, THE, 11-OHA4, NMN, F/NEP, F/NMN, 11-OHA4/A4, T/A4, DHEA-S/P5-S, THS/11-deoxyF, THS/DHEA-S, 21-deoxyF/DHEA-S, 6β-OHF/DHEA-S, THE/DHEA-S, 18-OHF/20α-DHF, Aldo/20α-DHF, Max_diameter, Neutrophil_%, Lymphocyte, Cr
	NFA vs MACS	RF	17	Aldo, 20α-DHF, 21-deoxyF, F, E, 6β-OHF, THF, THE, 11-OHA4, NMN, THE/E, 11-OHA4/A4, 21-deoxyF/DHEA-S, 6β-OHF/DHEA-S, THE/DHEA-S, Max_diameter, Neutrophil_%
	NFA vs MACS vs ACS	GB	34	B, THS, 20α-DHF, 21-deoxyF, F, E, 6β-OHF, THF, THE, 11-OHA4, NMN, F/NEP, F/NMN, 11-deoxyF/17-OHP4, 6β-OHF/F, THE/E, 11-OHA4/A4, T/A4, DHEA-S/P5-S, THS/11-deoxyF, THS/DHEA-S, 21-deoxyF/DHEA-S, 6β-OHF/DHEA-S, THE/DHEA-S, 18-OHF/20α-DHF, Aldo/20α-DHF, Max_diameter, Neutrophil_%, Lymphocyte, Eosinophil, ANC, TP, Albumin, K
Clinical data	All vs All	EE	14	Age, BMI, Max_diameter, WBC, HCT, Neutrophil_%, Lymphocyte, Eosinophil, ANC, TP, Albumin, Cr, Na, K
	NFA vs Others	EE	7	Neutrophil_%, Lymphocyte, ANC, ALT, Uric acid, K
	NFA vs PA	RF	9	Age, HTN, Hb, HCT, AST, ALT, Chol, TP, K
	NFA vs PCC	EE	5	Age, BMI, Max_diameter, Neutrophil_%, Lymphocyte
	NFA vs ACS	EE	9	Max_diameter, Neutrophil_%, Lymphocyte, Eosinophil, ANC, TP, Cr, Uric acid
	NFA vs MACS	EE	6	Max_diameter, Neutrophil_%, Lymphocyte, ANC, Uric acid
	NFA vs MACS vs ACS	EE	16	Age, BMI, HTN, Max_diameter, WBC, HCT, Neutrophil_%, Lymphocyte, Eosinophil, ANC, Chol, TP, Albumin, Cr, Na, K
Body composition	All vs All	RF	2	SFA, SMA/BMI
	NFA vs Others	GB	—	—
	NFA vs PA	GB	—	—
	NFA vs PCC	GB	—	—
	NFA vs ACS	RF	1	SFA/BMI
	NFA vs MACS	GB	—	—
	NFA vs MACS vs ACS	GB	—	—
Serum adrenal hormone profiles	All vs All	RF	41	Aldo, B, 11-deoxyF, THS, 20α-DHF, 18-OHF, 21-deoxyF, F, E, 6β-OHF, THF, THE, 11-OHA4, DHEA-S, P5-S, NEP, NMN, F/NEP, F/NMN, B/NEP, B/NMN, F/B, F/E, 11-deoxyF/17-OHP4, 21-deoxyF/17-OHP4, F/11-deoxyF, 6β-OHF/F, THE/E, 11-OHA4/A4, T/A4, DHEA-S/P5-S, THS/11-deoxyF, THS/DHEA-S, 21-deoxyF/DHEA-S, 6β-OHF/DHEA-S, THE/DHEA-S, B/20α-DHF, 18-OHF/20α-DHF, Aldo/20α-DHF, NEP/11-deoxyF, NMN/11-deoxyF
	NFA vs Others	EE	11	Aldo, B, 20α-DHF, 18-OHF, 21-deoxyF, F, E, 6β-OHF, THF, THE, 11-OHA4
	NFA vs PA	RF	17	Aldo, B, 11-deoxyF, 20α-DHF, 18-OHF, 21-deoxyF, F, E, 6β-OHF, THF, THE, 11-OHA4, NMN, F/NEP, 18-OHF/20α-DHF, Aldo/20α-DHF, NEP/11-deoxyF
	NFA vs PCC	RF	11	20α-DHF, F, E, 6β-OHF, NEP, NMN, F/NEP, F/11-deoxyF, 18-OHF/20α-DHF, Aldo/20α-DHF, NEP/11-deoxyF
	NFA vs ACS	EE	21	B, THS, 20α-DHF, 21-deoxyF, F, E, 6β-OHF, THF, THE, NMN, F/NEP, F/NMN, T/A4, DHEA-S/P5-S, THS/11-deoxyF, THS/DHEA-S, 21-deoxyF/DHEA-S, 6β-OHF/DHEA-S, THE/DHEA-S, 18-OHF/20α-DHF, Aldo/20α-DHF
	NFA vs MACS	GB	13	Aldo, 20α-DHF, 21-deoxyF, F, E, THF, THE, NMN, THE/E, 11-OHA4/A4, 21-deoxyF/DHEA-S, 6β-OHF/DHEA-S, THE/DHEA-S
	NFA vs MACS vs ACS	GB	27	B, THS, 20α-DHF, 21-deoxyF, F, E, 6β-OHF, THF, THE, 11-OHA4, NMN, F/NEP, F/NMN, F/E, 11-deoxyF/17-OHP4, 6β-OHF/F, THE/E, 11-OHA4/A4, T/A4, DHEA-S/P5-S, THS/11-deoxyF, THS/DHEA-S, 21-deoxyF/DHEA-S, 6β-OHF/DHEA-S, THE/DHEA-S, 18-OHF/20α-DHF, Aldo/20α-DHF

Abbreviations: EE, Easy Ensemble Classifier; GB, Gradient Boosting Classifier; RF, Random Forest Classifier; NFA, nonfunctioning adenoma; PA, primary aldosteronism; PCC; pheochromocytoma; ACS, adrenal Cushing's syndrome; MACS, mild autonomous cortisol secretion; 11-OHA4, 11-hydroxyandrostenedione; 11-deoxyF, 11-deoxycortisol; 17-OHP4, 17α-hydroxyprogesterone; 18-OHF, 18-hydroxycortisol; 20α-DHF, 20α-dihydrocortisol; 21-deoxyF, 21-deoxycortisol; 6β-OHF, 6β-hydroxycortisol; A4, androstenedione; ALT, alanine transaminase; alloTHF, allo-tetrahydrocortisol; ANC, absolute neutrophil count; AST, aspartate aminotransferase; Aldo, aldosterone; B, corticosterone; BMI, body mass index; BUN, blood urea nitrogen; CAD, coronary artery disease; CVA, cerebrovascular attack; Chol, total cholesterol; Cl, chloride; Cortisol after 1 mg ODST, cortisol after 1 mg overnight dexamethasone suppression test; Cr, creatinine; DHEA-S, dehydroepiandrosterone sulfate; DM, diabetes mellitus; E, cortisone; F, cortisol; Fx, fracture; HCT, hematocrit; HTN, hypertension; Hb, hemoglobin; K, potassium; MN_S, metanephrine; Mass site, mass site on CT scan; Max_diameter, maximal size of mass on CT scan; NEP, norepinephrine; NMN, normetanephrine; NMN_S, normetanephrine; Na, sodium; Neutrophil_%, segmented neutrophil percentage; P5-S, pregnenolone sulfate; PAC, plasma aldosterone concentration; PLT, platelet; PRA, plasma renin activity; SFA, subcutaneous fat area; SFV, subcutaneous fat volume; SF_HU, subcutaneous fat Hounsfield unit; SMA, skeletal muscle area; SMV, skeletal muscle volume; SM_HU, skeletal muscle Hounsfield unit; T, testosterone; THE, tetrahydrocortisone; THF, tetrahydrocortisol; THS, tetrahydro-11-deoxycortisol; TP, total protein; VFA, visceral fat area; VFV, visceral fat volume; VF_HU, visceral fat Hounsfield unit; WBC, white blood cell.

performance, with the EE classifier achieving a balanced accuracy of 0.92 and an AUC of 0.99. In the NFA-PA and NFA-PCC tasks, the use of SAP alone showed better performance than multiple or other single-parameter sets. This highlights the potential of SAP as a powerful diagnostic tool for differentiating between adrenal tumors. Our models revealed that SAP alone accounted for over 70% of the selected features across all disease-differentiating tasks.

Downloaded from https://academic.oup.com/ejendo/article/193/2/204/8217437 by guest on 29 August 2025

Figure 3. Selected features by Boruta for the classification of All-All, NFA-others, NFA-PA, NFA-ACS, NFA-PCC, NFA-MACS, and NFA-MACS-ACS. Each radial plot represents features selected by the Boruta algorithm using one of four input types: multiple parameters (top left), clinical data (top right), body composition (bottom left), and SAPs (bottom right). From the outermost ring inward, each concentric layer corresponds to a specific classification task: All-All, NFA vs Others, NFA vs PA, NFA vs ACS, NFA vs PCC, NFA vs MACS, and NFA vs MACS vs ACS. Colors indicate the feature category: blue for clinical data, green for serum adrenal hormones, red for serum hormone ratios, and orange for body composition. Only features selected in more than 50 out of 100 Boruta runs are shown. 11-deoxyF, 11-deoxycortisol; 11-deoxyF/17-OHP4, 11-deoxycortisol/17-alpha-hydoxy progesterone; 11-OHA4, 11-hydroxyandrostenedione; 11-OHA4/A4, 11-hydroxyandrostenedione/androstenedione; 17-OHP4, 17-alpha-hydoxy progesterone; 18-OHF, 18-hydroxycortisol; 18-OHF/20α-DHF, 18-hydroxycortisol/20-alpha-dihydrocortisol; 20α-DHF, 20-alpha-dihydrocortisol; 20α-DHF/F, 20-alpha-dihydrocortisol/cortisol; 21-deoxyF, 21-deoxycortisol; 21-deoxyF/17-OHP4, 21-deoxycortisol/17-alpha-hydoxy progesterone; 21-deoxyF/DHEA-S, 21-deoxycortisol/dehydroepiandrosterone sulfate; 6β-OHF, 6-beta-hydroxycortisol; 6β-OHF/DHEA-S, 6-beta-hydroxycortisol/dehydroepiandrosterone sulfate; 6β-OHF/F, 6-beta-hydroxycortisol/cortisol; A4, androstenedione; ACS, adrenal Cushing's syndrome; Aldo, aldosterone; Aldo/20α-DHF, aldosterone/20-alpha-dihydrocortisol; alloTHF, allo-tetrahydrocortisol; alloTHF/F, allo-tetrahydrocortisol/cortisol; ALT, alanine transaminase; ANC, absolute neutrophil count; AST, aspartate aminotransferase; B, corticosterone; B/20α-DHF, corticosterone/20-alpha-dihydrocortisol; B/NEP, corticosterone/norepinephrine; B/NMN, corticosterone/normetanephrine; BMI, body mass index; BUN, blood urea nitrogen; CAD, coronary artery disease; Chol, total cholesterol; Cr, creatinine; CVA, cerebrovascular attack; DHEA-S, dehydroepiandrosterone sulfate; DHEA-S/P5-S, dehydroepiandrosterone sulfate/pregnenolone sulfate; DM, diabetes mellitus; E, cortisone; F, cortisol; F/11-deoxyF, cortisol/11-deoxycortisol; F/21-deoxyF, cortisol/21-deoxycortisol; F/B, cortisol/corticosterone; F/E, cortisol/cortisone; F/NEP, cortisol/norepinephrine; F/NMN, cortisol/normetanephrine; Fx, fracture; Hb, hemoglobin; HCT, hematocrit; HTN, hypertension; MACS, mild autonomous cortisol secretion; MN_S, metanephrine; NEP, norepinephrine; NEP/11-deoxyF, norepinephrine/11-deoxycortisol; Neutrophil_%, segmented neutrophil percentage; NFA, nonfunctioning adenoma; NMN, normetanephrine; NMN/11-deoxyF, normetanephrine/11-deoxycortisol; NMN_S, normetanephrine; P5-S, pregnenolone sulfate; PA, primary aldosteronism; PAC, plasma aldosterone concentration; PCC, pheochromocytoma; PLT, platelet; PRA, plasma renin activity; SAP, serum adrenal hormone profiles; SF_HU, subcutaneous fat Hounsfield unit; SFA, subcutaneous fat area; SFV, subcutaneous fat volume; SM_HU, skeletal muscle Hounsfield unit; SMA, skeletal muscle area; SMV, skeletal muscle volume; T, testosterone; T/A4, testosterone/androstenedione; THE, tetrahydrocortisone; THE/DHEA-S, tetrahydrocortisone/dehydroepiandrosterone sulfate; THE/E, tetrahydrocortisone/cortisone; THF, tetrahydrocortisol; THF/F, tetrahydrocortisol/cortisol; THS, tetrahydro-11-deoxycortisol; THS/11-deoxyF, tetrahydro-11-deoxycortisol/11-deoxycortisol; THS/DHEA-S, tetrahydro-11-deoxycortisol/dehydroepiandrosterone sulfate; TP, total protein; VF_HU, visceral fat Hounsfield unit; VFA, visceral fat area; VFV, visceral fat volume; WBC, whole blood cell.

Downloaded from https://academic.oup.com/ejendo/article/193/2/204/8217437 by guest on 29 August 2025

Figure 4. Beeswarm plots illustrating feature importance across classification tasks. SHAP value distributions are shown for (A) All-All, (B) NFA vs Others, (C) NFA vs PA, (D) NFA vs PCC, (E) NFA vs ACS, (F) NFA vs MACS, and (G) NFA vs MACS vs ACS. Each plot visualizes the contribution of individual features to the model output. The *x*-axis represents the SHAP value, indicating the impact of each feature on the prediction, while the *y*-axis lists the top-ranked features. Color denotes the actual feature value, with red representing high and blue representing low values. 11-OHA4, 11-hydroxyandrostenedione; 20α-DHF, 20-alpha-dihydrocortisol; 21-deoxyF, 21-deoxycortisol; 6β-OHF, 6-beta-hydroxycortisol; A4, androstenedione; Aldo, aldosterone; B, corticosterone; BMI, body mass index; Cr, creatinine; DHEA-S, dehydroepiandrosterone sulfate; E, cortisone; F, cortisol; K, potassium; Lymphocyte, lymphocyte; Max_diameter, maximum diameter; NEP, norepinephrine; Neutrophil, neutrophil; NMN, normetanephrine; P5-S, pregnenolone sulfate; THE, tetrahydrocortisone; THF, tetrahydrocortisol; THS, tetrahydro-11-deoxycortisol.

Downloaded from https://academic.oup.com/ejendo/article/193/2/204/8217437 by guest on 29 August 2025

This emphasizes the importance of comprehensive hormonal profiling for the diagnosis of adrenal tumors and suggests its value in routine clinical practice. The observed trends included increased glucocorticoids and their metabolites in ACS and MACS and decreased androgen metabolites (eg, DHEA-S), which is consistent with previous studies.[13,15-17] For the NFA-PA combination task, in addition to Aldo levels, steroids associated with cortisol metabolism (eg, F, 18-OHF, and 21-deoxyF) were identified as key features, suggesting a role for cortisol metabolism in distinguishing PA from NFA. This aligns with the findings that autonomous cortisol secretion may often accompany PA.[28] In the NFA-PCC analysis, catecholamine-related features, including NEP, NMN, F/NEP, and NEP/11-deoxyF, were highly influential.

The use of advanced feature selection methods, particularly the Boruta algorithm, enables the identification of the most relevant features for each classification task. This approach not only improves model performance but also provides insights into the key biological and clinical factors that distinguish different adrenal tumors. Clinical data such as neutrophil count, lymphocyte count, and maximal tumor diameter were identified as important features. The neutrophil/lymphocyte ratio could aid in differentiating Cushing's syndrome because it reflects the immunomodulatory effects of excess cortisol, which is consistent with findings that high neutrophil and low lymphocyte counts are key indicators of NFA-ACS differentiation.[29] Maximal tumor diameter was a relevant feature differentiating NFA-ACS, NFA-PCC, and NFA-MACS but not NFA-PA. This finding supports that PA does not significantly differ in size from NFA.[9,30] For NFA-PA differentiation, potassium level was the most critical factor, with lower potassium levels indicating PA, a finding aligned with clinical practice.[4] Notably, comorbidities and complications such as hypertension, DM, dyslipidemia, and CVD were unexpectedly underutilized in the models. This suggests that these conditions are not specific to adrenal tumors and, therefore, have limited value in adrenal tumor differentiation.

In our study, body composition features showed limited performance in adrenal tumor differentiation. Specific body composition patterns in various adrenal tumors, such as low fat and low muscle mass in PCC patients, high visceral fat in ACS patients, and high fat in PA patients, have been identified.[10,11,31] However, in our study, these features were not significantly utilized in ML models. When used alone, the SFA/BMI ratio was a feature of ACS differentiation. However, their performance has been limited. This discrepancy with previous studies indicates that, although body composition changes exist, they may lack specificity for ML-based tumor classification.

This study has several advantages. First, it is the first to develop an ML model for adrenal tumor classification using clinical data, SAP, and body composition data, employing a rigorous ensemble ML approach for enhanced predictive capabilities. Our iterative model optimization, including outlier detection, feature selection, and class imbalance, demonstrates a commitment to developing reliable and clinically applicable ML models. Second, we simultaneously measured serum steroids, NEP, and normetanephrine from a small volume of the same serum sample, an approach that has not been successfully achieved before. Third, our large sample size of 641 patients from retrospective and prospective cohorts, with performance evaluated in both the training and test groups, enhanced the robustness of the study. Finally, this study identified the key features that contribute to the distinct characteristics of the 5 adrenal tumors.

However, this study had several limitations. First, the sensitivity of the All-All combination, which simultaneously differentiated the 5 diseases, was relatively modest at 0.77. Therefore, our model cannot replace clinical judgment or other diagnostic tools. Second, as this was a single-center study, the patients were divided into training and test groups, and the ML models lacked external validation. Third, a substantial portion of the samples was obtained from previously donated specimens, which were primarily collected from hospitalized patients, leading to a class imbalance. Additionally, as this study did not include patients with adrenal metastases or adrenocortical carcinoma, the applicability of our findings to adrenal tumors of malignant etiology may be limited. Fourth, despite enrolling 641 patients, the division into 5 disease groups resulted in small sample sizes per group. The lower proportion of nonfunctioning adenomas may also limit the model's generalizability. Fifth, although body composition was included as a parameter, its use in the final model is limited. Sixth, due to the combined retrospective and prospective design, some clinical data—such as family history and precontrast HU of adrenal tumors—were inconsistently available. Moreover, although LC-MS–based steroid profiling offers detailed hormonal analysis, its limited availability may restrict the model's immediate applicability. Finally, because body composition assessment and steroid profiling are not routinely performed in current clinical workflows, the integration of this model into standard practice may be limited. However, in certain circumstances—such as when dynamic hormone testing or 24-h urine collection is impractical—this ML model, which relies on CT-derived body composition and steroid profiling via a single blood draw, could offer a practical diagnostic alternative.

Conclusion

An ML model was developed to differentiate adrenal diseases using clinical, SAP, and body composition data. The ML model will assist in distinguishing adrenal diseases in clinical settings in which dynamic hormone tests are difficult to perform. However, large-scale external validation of the model is required.

Acknowledgments

We thank the patients who participated in this study. Biospecimens and data used in this study were provided by the Biobank of Seoul National University Hospital, a member of the Korea Biobank Network.

Supplementary material

Supplementary material is available at *European Journal of Endocrinology online*.

Authors' contributions

Seung Shin Park (Conceptualization [equal], Formal analysis [equal], Methodology [equal], Writing—original draft [lead], Writing—review & editing [equal]), Jongsung Noh (Data curation [equal], Formal analysis [equal], Methodology [equal], Writing—review & editing [equal]), Jinhee Kim

Downloaded from https://academic.oup.com/ejendo/article/193/2/204/8217437 by guest on 29 August 2025

(Data curation [equal], Formal analysis [equal], Methodology [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal]), Taesung Kim (Data curation [equal], Formal analysis [equal], Writing—review & editing [equal]), Hae Jin Seo (Data curation [equal], Formal analysis [equal]), Chang Ho Ahn (Conceptualization [equal], Funding acquisition [equal], Writing—review & editing [equal]), Jaegul Choo (Data curation [equal], Formal analysis [equal], Writing—review & editing [equal]), Man Ho Choi (Conceptualization [equal], Formal analysis [equal], Funding acquisition [equal], Methodology [equal], Supervision [equal], Writing—original draft [equal], Writing—review & editing [equal]), and Jung Hee Kim (Conceptualization [equal], Funding acquisition [equal], Resources [equal], Supervision [lead], Writing—original draft [equal], Writing—review & editing [equal]).

Funding

This study was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare of the Republic of Korea (project no. HI21C0032). This research was supported by a grant of Patient-Centered Clinical Research Coordinating Center (PACEN) funded by the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2024-00397426). This study was supported by Korea Institute of Science and Technology Institutional Program (project no. 2E31156).

Conflict of interest: None declared.

Data availability

The datasets generated during and analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Ebbehoj A, Li D, Kaur RJ, *et al.* Epidemiology of adrenal tumours in Olmsted County, Minnesota, USA: a population-based cohort study. *Lancet Diabetes Endocrinol*. 2020;8(11):894-902. https://doi.org/10.1016/S2213-8587(20)30314-4
2. Fassnacht M, Tsagarakis S, Terzolo M, *et al.* European Society of Endocrinology clinical practice guidelines on the management of adrenal incidentalomas, in collaboration with the European Network for the Study of Adrenal Tumors. *Eur J Endocrinol*. 2023;189(1):G1-G42. https://doi.org/10.1093/ejendo/lvad066
3. Park SS, Kim JH. Recent updates on the management of adrenal incidentalomas. *Endocrinol Metab (Seoul)*. 2023;38(4):373-380. https://doi.org/10.3803/EnM.2023.1779
4. Funder JW, Carey RM, Mantero F, *et al.* The management of primary aldosteronism: case detection, diagnosis, and treatment: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*. 2016;101(5):1889-1916. https://doi.org/10.1210/jc.2015-4061
5. Young WF J. Diagnosis and treatment of primary aldosteronism: practical clinical perspectives. *J Intern Med*. 2019;285(2):126-148. https://doi.org/10.1111/joim.12831
6. Reincke M, Fleseriu M. Cushing syndrome: a review. *JAMA*. 2023;330(2):170-181. https://doi.org/10.1001/jama.2023.11305
7. Burrello J, Monticone S, Losano I, *et al.* Prevalence of hypokalemia and primary aldosteronism in 5100 patients referred to a tertiary hypertension unit. *Hypertension*. 2020;75(4):1025-1033. https://doi.org/10.1161/HYPERTENSIONAHA.119.14063
8. Torpy DJ, Mullen N, Ilias I, Nieman LK. Association of hypertension and hypokalemia with Cushing's syndrome caused by ectopic ACTH secretion: a series of 58 cases. *Ann N Y Acad Sci*. 2002;970(1):134-144. https://doi.org/10.1111/j.1749-6632.2002.tb04419.x
9. Kim J, Bae KH, Choi YK, *et al.* Clinical characteristics for 348 patients with adrenal incidentaloma. *Endocrinol Metab (Seoul)*. 2013;28(1):20-25. https://doi.org/10.3803/EnM.2013.28.1.20
10. Ko Y, Jeong H, Khang S, Lee J, Kim KW, Kim BJ. Change of computed tomography-based body composition after adrenalectomy in patients with pheochromocytoma. *Cancers (Basel)*. 2022;14(8):1967. https://doi.org/10.3390/cancers14081967
11. Lee SH, Kwak MK, Ahn SH, *et al.* Change of skeletal muscle mass in patients with pheochromocytoma. *J Bone Miner Metab*. 2019;37(4):694-702. https://doi.org/10.1007/s00774-018-0959-3
12. Garrapa GG, Pantanetti P, Arnaldi G, Mantero F, Faloia E. Body composition and metabolic features in women with adrenal incidentaloma or Cushing's syndrome. *J Clin Endocrinol Metab*. 2001;86(11):5301-5306. https://doi.org/10.1210/jcem.86.11.8059
13. Maser-Gluth C, Reincke M, Allolio B, Schulze E. Metabolism of glucocorticoids and mineralocorticoids in patients with adrenal incidentalomas. *Eur J Clin Invest*. 2000;30(s3):83-86. https://doi.org/10.1046/j.1365-2362.2000.0300s3083.x
14. Arlt W, Biehl M, Taylor AE, *et al.* Urine steroid metabolomics as a biomarker tool for detecting malignancy in adrenal tumors. *J Clin Endocrinol Metab*. 2011;96(12):3775-3784. https://doi.org/10.1210/jc.2011-1565
15. Masjkur J, Gruber M, Peitzsch M, *et al.* Plasma steroid profiles in subclinical compared with overt adrenal Cushing syndrome. *J Clin Endocrinol Metab*. 2019;104(10):4331-4340. https://doi.org/10.1210/jc.2018-02349
16. Berke K, Constantinescu G, Masjkur J, *et al.* Plasma steroid profiling in patients with adrenal incidentaloma. *J Clin Endocrinol Metab*. 2022;107(3):e1181-e1192. https://doi.org/10.1210/clinem/dgab751
17. Prete A, Bancos I. Mild autonomous cortisol secretion: pathophysiology, comorbidities and management approaches. *Nat Rev Endocrinol*. 2024;20(8):460-473. https://doi.org/10.1038/s41574-024-00984-y
18. Bancos I, Taylor AE, Chortis V, *et al.* Urine steroid metabolomics for the differential diagnosis of adrenal incidentalomas in the EURINE-ACT study: a prospective test validation study. *Lancet Diabetes Endocrinol*. 2020;8(9):773-781. https://doi.org/10.1016/S2213-8587(20)30218-7
19. Noh J, Lee C, Kim JH, Myung SW, Choi MH. LC-MS based simultaneous profiling of adrenal hormones of steroids, catecholamines, and metanephrines. *J Lipid Res*. 2023;64(11):100453. https://doi.org/10.1016/j.jlr.2023.100453
20. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. *Lancet*. 2020;395(10236):1579-1586. https://doi.org/10.1016/S0140-6736(20)30226-9
21. Assié G, Allassonnière S. Artificial intelligence in endocrinology: on track toward great opportunities. *J Clin Endocrinol Metab*. 2024;109(6):e1462-e1467. https://doi.org/10.1210/clinem/dgae154
22. Eisenhofer G, Durán C, Cannistraci CV, *et al.* Use of steroid profiling combined with machine learning for identification and subtype classification in primary aldosteronism. *JAMA Netw Open*. 2020;3(9):e2016209. https://doi.org/10.1001/jamanetworkopen.2020.16209
23. Braun LT, Osswald A, Zopp S, *et al.* Delineating endogenous Cushing's syndrome by GC-MS urinary steroid metabotyping. *EBioMedicine*. 2024;99:104907. https://doi.org/10.1016/j.ebiom.2023.104907
24. Reel PS, Reel S, van Kralingen JC, *et al.* Machine learning for classification of hypertension subtypes using multi-omics: a multicentre, retrospective, data-driven study. *EBioMedicine*. 2022;84:104276. https://doi.org/10.1016/j.ebiom.2022.104276
25. Tanabe A, Naruse M. Recent advances in the management of pheochromocytoma and paraganglioma. *Hypertens Res*. 2020;43(11):1141-1151. https://doi.org/10.1038/s41440-020-0531-0
26. Lee YS, Hong N, Witanto JN, *et al.* Deep neural network for automatic volumetric segmentation of whole-body CT images for body

Downloaded from https://academic.oup.com/ejendo/article/193/2/204/8217437 by guest on 29 August 2025

composition assessment. *Clin Nutr*. 2021;40(8):5038-5046. https://doi.org/10.1016/j.clnu.2021.06.025
27. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, California: Curran Associates; 2017:4768-4777.
28. Araujo-Castro M, Paja Fano M, Pla Peris B, *et al*. Autonomous cortisol secretion in patients with primary aldosteronism: prevalence and implications on cardiometabolic profile and on surgical outcomes. *Endocr Connect*. 2023;12(9):e230043. 10.1530/EC-23-0043.
29. Detomas M, Altieri B, Chifu I, *et al*. Subtype-specific pattern of white blood cell differential in endogenous hypercortisolism. *Eur J Endocrinol*. 2022;187(3):439-449. https://doi.org/10.1530/EJE-22-0211
30. Stavropoulos K, Imprialos KP, Katsiki N, *et al*. Primary aldosteronism in patients with adrenal incidentaloma: is screening appropriate for everyone? *J Clin Hypertens (Greenwich)*. 2018;20(5):942-948. https://doi.org/10.1111/jch.13291
31. Park SS, Ahn CH, Kim SW, Yoon JW, Kim JH. Subtype-specific body composition and metabolic risk in patients with primary aldosteronism. *J Clin Endocrinol Metab*. 2024;109(2):e788-e798. https://doi.org/10.1210/clinem/dgad520

Downloaded from https://academic.oup.com/ejendo/article/193/2/204/8217437 by guest on 29 August 2025