Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Radiomics-based detection of lateralization in primary aldosteronism: a multicenter study

Marta Araujo-Castro

martaazul.2a@hotmail.com

Hospital Universitario Ramón y Cajal

Miguel Ángel Gómez-Bermejo

Hospital Universitario Ramón y Cajal

Eva Moya

Hospital Universitario Infanta Leonor

Angel Rebollo Román

Hospital Universitario Reina Sofía

Rosario Fernández

Hospital Universitario Central de Asturias

Edelmiro Menendez Torre

Hospital Universitario Central de Asturias

Cristina Robles Lazaro

Complejo Hospitalario de Salamanca

Jose María Recio

Complejo Hospitalario de Salamanca

Emilia Gomez Hoyos

Hospital Clínico Universitario de Valladolid

Iago Millan Entenza

Hospital Universitario Ramón y Cajal

Alvaro Eduardo Flandes Iparraguirre

Hospital Universitario Ramón y Cajal

Jesús Fraguas Martín

Hospital Universitario Ramón y Cajal

Juan Trujillo García

Hospital Universitario Ramón y Cajal

Pastorin Salis Rodrigo

Hospital Universitario Infanta Leonor

Daniel Prieto Morán

Hospital Universitario Ramón y Cajal

Article

Keywords: primary aldosteronism, radiomics, adrenal venous sampling, aldosterone, adrenal gland

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11240494/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Purpose: To investigate whether adrenal CT radiomics can distinguish unilateral from bilateral primary aldosteronism (PA), focusing particularly on inter-lateral asymmetry of radiomic features.

Methods: Patients were selected from the multicenter Spanish SPAIN-ALDO registry and included if adrenal computed tomography (CT) imaging and definitive subtype classification by adrenal venous sampling were available. Portal-phase CT scans from 33 patients (66 adrenal glands; 22 unilateral and 11 bilateral PA) were centrally analyzed. Adrenal glands were segmented independently by two radiologists. Radiomic features underwent reproducibility filtering (ICC ≥0.75), redundancy reduction, and bootstrap-based stability selection. Patient-level lateralization was primarily assessed using inter-lateral asymmetry indices, with repeated stratified cross-validation and balanced logistic regression.

Results: Two reproducible radiomic features—GLSZM small zone emphasis and GLSZM gray-level non-uniformity—showed consistent discriminatory signals for unilateral versus bilateral PA. Radiomics based on inter-lateral asymmetry achieved an AUC of 0.681, increasing to 0.733 when systolic blood pressure was added to the model. At the individual-gland level, Compactness2 and maximum 3D diameter were selected. Radiomics alone achieved an AUC of 0.660 for identifying the aldosterone-producing gland, with no improvement from adding clinical variables. Morphological features showed excellent interobserver reproducibility.

Conclusions: In this multicenter exploratory cohort, reproducible adrenal CT radiomic features demonstrated a potential signal for PA subtype classification. Inter-lateral radiomic asymmetry combined with systolic blood pressure showed the highest point-estimate performance for distinguishing unilateral from bilateral PA, while radiomics alone performed similarly for identifying the aldosterone-producing gland. These findings warrant validation in larger independent cohorts**.**

1. Introduction

Primary aldosteronism (PA) is the most common form of endocrine hypertension, accounting for approximately 5–10% of all hypertensive patients and in up to 20–30% of patients with resistant hypertension [1]. Beyond its effects on blood pressure (BP), PA is associated with an increased risk of cardiovascular, renal, and metabolic complications, independently of BP levels [2–5]. Accurate subtype differentiation is therefore essential, as unilateral and bilateral forms of PA differ substantially in their pathophysiology, prognosis, and therapeutic management [1, 6].

The two principal subtypes of PA are unilateral aldosterone excess—most frequently caused by aldosterone-producing adenomas (APAs) or, less commonly, unilateral adrenal hyperplasia—and bilateral adrenal hyperplasia [7]. Identification of unilateral disease is of paramount importance because adrenalectomy may achieve biochemical cure and substantially improve cardiovascular outcomes, whereas patients with bilateral disease require long-term treatment with mineralocorticoid receptor

antagonists (MRAs) or, in the near future, with aldosterone synthase inhibitors [6, 8]. Consequently, reliable subtype diagnosis represents a critical step in the clinical management of PA [1].

Current international guidelines recommend adrenal computed tomography (CT) as the initial imaging modality in all patients with confirmed PA, primarily to exclude adrenocortical carcinoma and identify structural adrenal abnormalities [1, 7]. However, conventional radiological assessment has limited accuracy for subtype classification. Several studies have demonstrated a substantial discordance between CT findings and adrenal venous sampling (AVS), the current gold standard for lateralization [9, 10]. Despite being considered the reference standard for subtype differentiation, AVS remains an invasive, technically demanding, and costly procedure that is available only in specialized centers [11]. Success rates vary considerably according to local expertise, and the procedure may not be feasible in all patients [12–14]. Consequently, there is growing interest in the development of non-invasive tools capable of improving subtype prediction and complementing existing diagnostic algorithms.

Radiomics has emerged as a promising approach for extracting quantitative information from medical images beyond what can be appreciated by the human eye. By converting radiological images into high-dimensional data, radiomics enables the characterization of lesion shape, intensity, texture, and spatial heterogeneity, potentially reflecting underlying molecular, histopathological, and functional properties. In oncology, radiomics has demonstrated value in tumor classification, prognostic stratification, and treatment prediction [15, 16]. More recently, these methods have been extended to adrenal disorders, where they offer the possibility of identifying imaging biomarkers associated with hormone secretion and disease subtype [17–19].

The application of radiomics to PA remains relatively limited and has primarily focused on differentiating APAs from non-functioning adrenal adenomas or improving subtype prediction when combined with clinical and biochemical variables [20, 21]. Keng *et al.* developed a clinical-radiomic nomogram based on unenhanced CT images in patients with unilateral adrenal adenomas and reported that integrating radiomic features with clinical variables improved identification of APAs [20]. Subsequently, Li *et al*. showed that CT-based radiomic signatures could distinguish APAs from non-functioning adrenal adenomas in patients with essential hypertension and explored prediction of postoperative biochemical outcomes [21]. More recently, machine-learning (ML) approaches integrating radiomic features with clinical and biochemical characteristics have been proposed for PA subtype classification. In a retrospective cohort of 158 patients who underwent AVS, combined models incorporating CT radiomics outperformed conventional imaging assessment for distinguishing unilateral from bilateral disease and for identifying the dominant adrenal gland in unilateral PA. However, the diagnostic performance of radiomics alone remained moderate, emphasizing the complexity of adrenal steroidogenic disorders and the need for further validation [22]. The largest radiomics study to date, based on the prospective German Conn Registry, analyzed 269 patients with PA and investigated whether CT-derived radiomic features could predict AVS lateralization. Although the addition of clinical parameters improved model performance, radiomics alone showed only modest discrimination between unilateral and bilateral

aldosterone excess, with overall areas under the receiver operating characteristic curve (AUC) close to 0.56 [23].

ML models combining radiomic and clinical variables have shown encouraging diagnostic performance, suggesting that quantitative imaging may identify subtle abnormalities that are not evident on routine visual assessment. Nevertheless, several important knowledge gaps remain. Existing studies have generally been limited by their single-center design, relatively small sample size, heterogeneous methodologies for image acquisition and segmentation, and the frequent incorporation of biochemical variables into predictive models. Moreover, few investigations have specifically addressed whether adrenal radiomic phenotypes differ intrinsically between unilateral and bilateral PA and whether these differences can provide insight into the biological heterogeneity of aldosterone-producing adrenal disease.

Therefore, the present multicenter Spanish study aimed to characterize the radiomic features of the adrenal glands in patients with unilateral and bilateral PA and to evaluate whether quantitative imaging biomarkers may contribute to the non-invasive differentiation of PA subtypes. By leveraging data from multiple expert centers, we sought to improve our understanding of adrenal radiomic phenotypes and explore their potential role in the diagnostic pathway of PA.

2. Methods

2.1. Study population

Patients were selected from the SPAIN-ALDO registry, a multicenter Spanish electronic database collecting clinical, biochemical, imaging, procedural, and treatment data from patients diagnosed with and/or followed for PA at six referral centers in Spain. The source cohort included 92 patients with available adrenal CT studies.

Patients were progressively excluded if their CT studies did not meet the technical imaging criteria (120 kV and slice thickness ≤ 3 mm), if no reference standard for lateralization was available (selective AVS and/or biochemical cure following adrenalectomy), or if no CT study acquired during the portal venous phase was available. The portal venous phase was considered the standard phase for radiomic assessment of the adrenal parenchyma, given that attenuation and texture features vary substantially across contrast-enhanced phases [24].

After these exclusions, 39 patients had adrenal CT studies available for centralized evaluation at the reference center (Hospital Universitario Ramón y Cajal, Madrid). The study was approved by the Ethics Committee of Hospital Universitario Ramón y Cajal (Madrid) [approval date: November 10, 2020, code: ACTA 401]

2.2. PA Diagnosis and AVS Lateralization

PA was diagnosed according to Endocrine Society clinical practice guidelines applicable during the inclusion period [1, 6]. At the time of diagnosis, plasma aldosterone concentration (PAC), direct renin concentration (DRC), and/or plasma renin activity (PRA) were measured in all participants. Subtype classification used successful AVS and/or biochemical cure after adrenalectomy, the latter classified according to Primary Aldosteronism Surgical Outcome (PASO) criteria when follow-up data were available [25]. AVS was performed according to a predefined protocol (without ACTH stimulation). Unilateral PA was defined by lateralized aldosterone secretion on successful bilateral AVS, with a lateralization index (LI) ≥ 4. Patients with complete biochemical cure were also classified as unilateral PA (correction of hypokalemia (if present preoperatively) and normalization of the aldosterone-to-renin ratio) [25]. Bilateral PA was defined by successful bilateral AVS showing no evidence of lateralized aldosterone secretion (LI < 4), consistent with bilateral aldosterone hypersecretion.

2.3. Radiomics study

Contrast-enhanced abdominal CT scans acquired during the portal venous phase were analyzed from the participating centers. Given the inherent variability between acquisition systems, technical homogeneity criteria were applied as a filtering approach rather than statistical harmonization between scanners (e.g., ComBat), in order to reduce scanner-related bias in feature extraction. Inclusion criteria required a fixed tube voltage of 120 kV and a slice thickness of ≤ 3 mm, to minimize partial-volume artifacts that could distort spatial and fine-texture features. It should be noted that this 3 mm criterion served exclusively as the upper inclusion threshold rather than the effective thickness of the cohort. In the final 33 patients, the actual slice thickness was 1.43 ± 0.35 mm (median 1.50 mm; range 0.80–3.00 mm), with 91% of cases acquired at ≤ 1.5 mm. This sub-millimetric/millimetric effective resolution substantially mitigates concerns regarding spatial resolution relative to the small dimensions of the adrenal glands and minimizes partial-volume artifacts.

Segmentation

Three-dimensional volumetric segmentation of the regions of interest (ROIs) encompassing both adrenal glands (right and left) was manually performed by three radiologists with varying levels of experience (M.A.G.B., 7 years of experience in abdominal radiology; J.F.M., 2 years; A.F.I., 2 years). The radiologists were alternately paired so that each adrenal gland was independently assessed by two different radiologists. Segmentation was performed using the open-source software LIFEx 7.0. Interobserver reproducibility of radiomic features was quantified from the two independent segmentations of each gland.

Spatial and Intensity Standardization

Following the recommendations of the Image Biomarker Standardisation Initiative (IBSI), anisotropic spatial resampling was performed (X = 0.8 mm, Y = 0.8 mm, Z = 1.5 mm; non-isotropic), together with relative intensity normalization (μ + 3σ). Six intensity features constrained by the reconstruction window

(0–400 HU) were excluded because they introduced truncation bias that could not be corrected by normalization.

Database Construction: Four data representations were constructed: (1) inter-lateral asymmetry indices, calculated as the right-to-left ratio for morphological variables and as the right-minus-left difference for the remaining features; (2) both adrenal glands treated as independent variables, with each selected feature entered separately for the right and left glands together with the clinical variables; (3) the right adrenal gland alone; and (4) the left adrenal gland alone. Because inter-lateral asymmetry directly represents the concept of lateralization, the confirmatory analysis for unilateral-versus-bilateral discrimination was restricted to this representation; the remaining three were retained only as exploratory comparators and are not reported as confirmatory results.

Feature Selection: Feature selection proceeded in three sequential stages, applied before any consideration of predictive performance for the first stage. (1) Interobserver reproducibility filtering: the intraclass correlation coefficient, ICC(3,1) (two-way mixed-effects, consistency, single measures), was calculated between two independent reader groups —Group 1 (radiologist A.F.I., with support from M.A.G.B.) and Group 2 (radiologist J.F.M., with support from M.A.G.B.); see reader composition and experience in Section 2.3— for each radiomic feature at the individual-gland level (i.e., comparing the two groups' values for the same gland) in a 39-patient reliability subsample (66 adrenal glands corresponding to the final confirmatory cohort); the six intensity features constrained by the reconstruction window were excluded beforehand; only features with ICC(3,1) ≥ 0.75 were retained. For the individual-gland model, this per-gland ICC directly reflects the reproducibility of the features actually entered into the model. For the lateralization model, which uses the right-minus-left difference of each feature, per-gland ICC was used as a proxy: reproducibility of the difference itself was not separately measured and was not assumed to be numerically identical to that of the underlying per-gland feature. (2) Correlation filtering: among the reproducible features, pairs with Pearson correlation > 0.90 were considered redundant, and this filter was recalculated within each training partition of the step below to avoid information leakage. (3) Stability ranking: a nested cross-validation framework was applied to the reproducible, non-redundant feature set. An outer loop (5 folds × 5 repetitions) partitioned the data, while an inner bootstrap resampling loop (50 iterations), applied exclusively to the training set of each outer fold, generated a stability ranking based on ANOVA (SelectKBest); this loop is independent from, and precedes, the 5 × 20 repeated cross-validation used to assess final model performance (Section 2.4). Final radiomic variables were selected according to their frequency of occurrence among the most discriminative features across the resulting rankings, without using the test set at any stage. Two features were selected for the lateralization model (inter-lateral difference in GLSZM small zone emphasis and in GLSZM gray-level non-uniformity) and two for the individual-gland model (morphological Compactness2 and maximum 3D diameter). The final models were additionally evaluated combined with systolic blood pressure and aldosterone levels, given their established pathophysiological relevance in this condition.

Prevention of Information Leakage

Feature scaling (z-score standardization using StandardScaler) was fitted exclusively on the training data within each partition and subsequently applied to the corresponding test set, never in the reverse direction. For the individual-adrenal-level analysis (66 adrenal glands), validation was performed using patient-level grouped partitioning (StratifiedGroupKFold), ensuring that both adrenal glands from the same individual always remained within the same subset (training or test), thereby preventing leakage of shared clinical information between paired glands.

2.4. Model and Statistical Validation

Balanced-class-weight logistic regression was used as the final modeling approach. Model performance was assessed using repeated stratified cross-validation (5 folds × 20 repetitions = 100 independent evaluations) for each of the four patient-level data representations. Performance metrics, including area under the receiver operating characteristic curve (AUC), Matthews correlation coefficient (MCC), accuracy, sensitivity, and specificity, were summarized as the mean and the P2.5–P97.5 percentile range calculated directly from the distribution of the 100 evaluations. This approach was used instead of the standard error of the mean, which may underestimate the uncertainty of generalization by failing to capture variability across cross-validation partitions, an issue that is particularly relevant in small cohorts such as the present study. As described in Section 2.3, interobserver ICC(3,1) was used as a mandatory prerequisite for feature entry (ICC(3,1) ≥ 0.75), calculated before any performance-based selection. Agreement was interpreted as poor (< 0.50), moderate (0.50–0.75), good (0.75–0.90), or excellent (> 0.90) [26]. ICC(3,1) was additionally calculated for basic morphological descriptors and, exploratorily, across the full feature set, to characterize the overall distribution of reproducibility in this cohort beyond the features entering the final models.

3. Results

3.1. Study population

Of the 92 patients with initially available adrenal CT studies, 53 were excluded because they did not meet the technical imaging criteria, lacked a reference standard for lateralization, or did not have a CT study acquired during the portal venous phase **(see** Section 2.1**)**, leaving 39 patients with imaging and clinical data available for centralized evaluation. Of these, an additional six patients (15.4%) were excluded because complete valid clinical data were unavailable, resulting in a final cohort of 33 patients (66 adrenal glands), 22 patients with unilateral PA and 11 with bilateral PA **(**Tables 1 and 2**).**

Table 1
Reasons for exclusion during the imaging screening phase (92 → 39 patients).

Reason for exclusion	n
Non-portal venous phase	19
Technical criteria not met (kVp ≠ 120 or slice thickness > 3 mm)	12
Unknown lateralization status (no AVS or biochemical cure after surgery)	12
Other (inaccessible archive, empty folder, no CT study available, or MRI only)	10
Total excluded	**53**

AVS: adrenal venous sampling

Table 2
Characteristics of the final cohort (n = 33).

Variable	Total (n = 33)	Unilateral (n = 22)	Bilateral (n = 11)
Age (years), mean ± SD	52.2 ± 13.6	53.9 ± 11.5	49.1 ± 17.1
SBP (mmHg), mean ± SD	158.9 ± 27.4	164.0 ± 31.2	148.6 ± 13.6
DBP (mmHg), mean ± SD	93.2 ± 12.3	94.6 ± 12.6	90.4 ± 11.6
Creatinine (mg/dL), mean ± SD	0.9 ± 0.2	0.8 ± 0.2	0.9 ± 0.2
Sodium (mmol/L), mean ± SD	142.4 ± 3.2	142.5 ± 3.3	142.0 ± 2.9
Potassium (mEq/L), mean ± SD	3.6 ± 0.6	3.5 ± 0.7	3.7 ± 0.3
Aldosterone (PAC; ng/dL), mean ± SD	44.0 ± 37.6	41.7 ± 39.5	48.5 ± 35.0
Renin (DRC; mIU/L), mean ± SD	21.2 ± 33.9	20.2 ± 31.6	23.0 ± 39.6
ARR (ng/dL/ mIU/L), mean ± SD	24.9 ± 33.0	21.5 ± 24.9	31.6 ± 45.9
Number of antihypertensive medications, mean ± SD	3.0 ± 1.2	3.0 ± 1.3	3.1 ± 0.9
Male sex, n (%)	14/33 (42%)	9/22 (41%)	5/11 (45%)
Diabetes, n (%)	6/33 (18%)	3/22 (14%)	3/11 (27%)
Microalbuminuria, n (%)	5/33 (15%)	4/22 (18%)	1/11 (9%)
Hypokalemia, n (%)	20/33 (61%)	15/22 (68%)	5/11 (45%)

3.2. Predictive performance for unilateral versus bilateral discrimination

Two features passed the reproducibility filter and were consistently top-ranked in the stability analysis of inter-lateral asymmetry indices: the right-minus-left difference in GLSZM small zone emphasis and in GLSZM gray-level non-uniformity. Using these two reproducible features, the radiomics-only model achieved an AUC of 0.681 (MCC 0.312). Adding systolic BP increased the point-estimate AUC to 0.733 (MCC 0.354), whereas further adding aldosterone returned performance to the radiomics-alone level (AUC 0.681; MCC 0.270) **(**Table 3, Fig. 1**).** As in the previous, non-reproducibility-filtered iteration of this analysis, systolic BP improved point-estimate performance while aldosterone did not, a pattern consistent with an increased risk of overfitting from additional variables in a small cohort. The right adrenal gland, both-glands, and left-gland representations were not re-evaluated with this reproducibility-filtered feature set and are not reported as confirmatory results **(see** Section 2.3**).**

Table 3
Predictive performance for unilateral versus bilateral discrimination, using inter-lateral asymmetry indices restricted to reproducibility-filtered features (ICC(3,1) ≥ 0.75): right-minus-left difference in GLSZM small zone emphasis and GLSZM gray-level non-uniformity.

Configuration	AUC (P2.5-P97.5)	MCC (P2.5-P97.5)	Accuracy (P2.5-P97.5)	Sensitivity, Unilateral (P2.5-P97.5)	Specificity, Bilateral (P2.5-P97.5)
Radiomics	0.681 [0.224, 1.000]	0.312 [-0.400, 1.000]	0.632 [0.286, 1.000]	0.590 [0.200, 1.000]	0.717 [0.000, 1.000]
+ SBP	**0.733 [0.336, 1.000]**	**0.354 [-0.360, 1.000]**	**0.668 [0.429, 1.000]**	**0.653 [0.250, 1.000]**	**0.687 [0.000, 1.000]**
+ SBP + Aldosterone	0.681 [0.300, 1.000]	0.270 [-0.360, 0.750]	0.634 [0.333, 0.857]	0.634 [0.250, 1.000]	0.625 [0.000, 1.000]

Abbreviations: SBP: systolic blood pressure. An exploratory comparison against models based on both glands, right gland alone, and left gland alone, using features selected without reproducibility pre-filtering, was performed in an earlier iteration of this analysis and is not reported here as a confirmatory result (see Section 2.3).

3.3. Predictive performance for identifying the aldosterone-producing adrenal gland (individual-gland level)

Two reproducible morphological features were selected for this analysis: Compactness2 and maximum 3D diameter. The radiomics-only model achieved an AUC of 0.660 (MCC 0.296) for identifying the aldosterone-producing gland. Adding systolic BP did not change the point-estimate AUC (0.660; MCC 0.238), and further adding aldosterone reduced performance (AUC 0.624; MCC 0.124). As in the patient-level analysis, the incorporation of clinical variables did not improve performance over the radiomics-only model. This is consistent with the nature of the clinical variables included, which were measured at the patient level and were therefore identical for both adrenal glands within the same individual;

consequently, they cannot provide information capable of differentiating which of the two glands is the aldosterone-producing gland in patients with unilateral disease. Validation was performed using patient-level grouped partitioning to prevent leakage of shared clinical information between paired adrenal glands from the same individual **(**Table 4 **and** Fig. 2**).**

Table 4
Predictive performance for identifying the aldosterone-producing adrenal gland at the individual-gland level, using two reproducibility-filtered morphological features (ICC(3,1) ≥ 0.75): Compactness2 and maximum 3D diameter.

Configuration	AUC (P2.5-P97.5)	MCC (P2.5-P97.5)	Accuracy (P2.5-P97.5)	Sensitivity, Producing gland (P2.5-P97.5)	Specificity, Non-producing gland (P2.5-P97.5)
Radiomics	**0.660 [0.323, 0.906]**	**0.296 [-0.185, 0.645]**	0.620 [0.417, 0.786]	0.567 [0.292, 0.778]	0.741 [0.250, 1.000]
+ SBP	0.660 [0.322, 0.905]	0.238 [-0.246, 0.694]	0.612 [0.417, 0.823]	0.605 [0.279, 0.952]	0.642 [0.119, 1.000]
+ SBP + Aldosterone	0.624 [0.279, 0.851]	0.124 [-0.316, 0.579]	0.568 [0.345, 0.769]	0.581 [0.246, 0.867]	0.549 [0.000, 1.000]

Abbreviations: SBP: systolic blood pressure

3.4. Interobserver reproducibility of radiomic features

Basic morphological descriptors showed good-to-excellent interobserver agreement (ICC(3,1) 0.850–0.957), including maximum 3D diameter (0.957), volume (0.929), approximate volume (0.925), voxel count (0.925), spherical diameter (0.898), and surface area (0.850) (Table 5). By construction, all four radiomic features entering the confirmatory models met the reproducibility threshold: GLSZM small zone emphasis and gray-level non-uniformity (lateralization model), and morphological Compactness2 and maximum 3D diameter (individual-gland model) **(**Table 5**).** For context, in an earlier iteration of this analysis in which features were selected by predictive performance alone, the resulting signature (GLCM sum/joint variance, intensity interquartile range, GLSZM high gray-level zone emphasis, surface-to-volume ratio) showed only moderate-to-poor reproducibility (ICC(3,1) 0.332–0.608), motivating the reproducibility-first selection strategy adopted here. Exploratory ranking of the full feature set identified several GLSZM and GLRLM descriptors with excellent ICC (0.970–0.998), whereas intensity-histogram features clustered at the bottom of the distribution (ICC − 0.010 to 0.174), with 95% confidence intervals that included zero **(**Table 6**).**

Table 5
Interobserver ICC(3,1) for basic morphological descriptors and for the radiomic features entering the confirmatory predictive models.

Feature	ICC(3,1)	Interpretation
Morphological descriptors		
Maximum 3D diameter	0.957	Excellent
Volume	0.929	Excellent
Approximate volume	0.925	Excellent
Voxel count	0.925	Excellent
Spherical diameter	0.898	Good
Surface area	0.850	Good
Features in the confirmatory models (per-gland ICC)*		
Lateralization model		
GLSZM small zone emphasis	0.998	Excellent
GLSZM gray-level non-uniformity	0.959	Excellent
Individual-gland model		
Morphological Compactness2	0.763	Good
Maximum 3D diameter	0.957	Excellent
Features from a prior, non-reproducibility- filtered iteration (not used in the confirmatory models)		
GLSZM high gray-level zone emphasis	0.608	Moderate
Morphological surface-to-volume ratio	0.576	Moderate
Intensity interquartile range	0.473	Poor
GLCM sum variance	0.397	Poor
GLCM joint variance	0.332	Poor

ICC(3,1): two-way mixed-effects, single-measure, consistency coefficient. Interpretation after Koo and Li [36]: poor < 0.50; moderate 0.50–0.75; good 0.75–0.90; excellent > 0.90. GLSZM, gray-level size zone matrix; GLCM, gray-level co-occurrence matrix; *ICC values for the lateralization-model features were measured at the individual-gland level (agreement between readers on the same gland), not on the derived right-minus-left difference used in the model; see Section 2.3 and Discussion.

Table 6
Radiomic features with the highest and lowest interobserver ICC(3,1) in the full extracted set (exploratory ranking).

Feature	ICC(3,1)	95% CI	Interpretation
Highest ICC			
GLSZM small-zone emphasis (IBSI:5QRC)	0.998	[1.00, 1.00]	Excellent
GLRLM short-run emphasis (IBSI:220V)	0.997	[0.99, 1.00]	Excellent
GLSZM zone percentage (IBSI:P30P)	0.996	[0.99, 1.00]	Excellent
GLSZM normalized zone-size non-uniformity (IBSI:VB3A)	0.996	[0.99, 1.00]	Excellent
GLSZM small-zone high gray-level emphasis (IBSI:HW1V)	0.995	[0.99, 1.00]	Excellent
GLRLM long-run high gray-level emphasis (IBSI:3KUM)	0.995	[0.99, 1.00]	Excellent
GLRLM long-run emphasis (IBSI:W4KF)	0.994	[0.99, 1.00]	Excellent
GLRLM run percentage (IBSI:9ZK5)	0.993	[0.99, 1.00]	Excellent
Intensity-based energy (IBSI:N8CA)	0.988	[0.98, 0.99]	Excellent
GLSZM large-zone high gray-level emphasis (IBSI:J17V)	0.983	[0.97, 0.99]	Excellent
GLCM inverse difference moment (IBSI:WF0Z)	0.979	[0.97, 0.99]	Excellent
Intensity-based energy (HU)	0.977	[0.96, 0.99]	Excellent
GLRLM run-length variance (IBSI:SXLW)	0.977	[0.96, 0.99]	Excellent
GLSZM large-zone emphasis (IBSI:48P8)	0.973	[0.96, 0.98]	Excellent
90th intensity percentile (IBSI:8DWT)	0.970	[0.95, 0.98]	Excellent
Lowest ICC			
GLSZM low gray-level zone emphasis (IBSI:XMSY)	0.174	[− 0.07, 0.40]	Poor
Intensity-histogram interquartile range (IBSI:WR0O)	0.172	[− 0.07, 0.40]	Poor
Histogram quartile coefficient of dispersion (IBSI:SLWD)	0.168	[− 0.08, 0.39]	Poor
GLRLM gray-level variance (IBSI:8CE5)	0.165	[− 0.08, 0.39]	Poor
Intensity-histogram kurtosis (IBSI:C3I7)	0.153	[− 0.09, 0.38]	Poor

Feature	ICC(3,1)	95% CI	Interpretation
Intensity-histogram variance (IBSI:CH89)	0.141	[− 0.10, 0.37]	Poor
Intensity-histogram standard deviation	0.140	[− 0.10, 0.37]	Poor
GLRLM low gray-level run emphasis (IBSI:V3SW)	0.139	[− 0.11, 0.37]	Poor
Minimum histogram-gradient gray level (IBSI:RHQZ)	0.133	[− 0.11, 0.36]	Poor
GLRLM short-run low gray-level emphasis (IBSI:HTZT)	0.118	[− 0.13, 0.35]	Poor
Histogram coefficient of variation (IBSI:CWYJ)	0.111	[− 0.13, 0.34]	Poor
GLSZM gray-level variance (IBSI:BYLV)	0.078	[− 0.17, 0.31]	Poor
Intensity-histogram mode (IBSI:AMMC)	0.047	[− 0.20, 0.28]	Poor
Intensity kurtosis (IBSI:IPH6)	0.002	[− 0.24, 0.24]	Poor
Histogram minimum gray level (IBSI:1PR8)	−0.010	[− 0.25, 0.23]	Poor

4. Discussion

In this study, radiomic features were required to meet an interobserver reproducibility threshold (ICC(3,1) ≥ 0.75) before any performance-based selection was applied. This reverses the order followed in an earlier iteration of this analysis, in which features were selected by predictive performance alone and only subsequently found to have moderate-to-poor reproducibility (ICC 0.332–0.608; Table 5). Under the reproducibility-first pipeline, the two features selected for lateralization —the inter-lateral difference in GLSZM small zone emphasis and gray-level non-uniformity— were derived from underlying per-gland features that showed excellent interobserver agreement (ICC 0.998 and 0.959, respectively), comparable to that of the simplest morphological descriptors (volume, diameter). Reproducibility was measured at the individual-gland level rather than on the derived right-minus-left difference itself; while a highly reproducible per-gland feature provides a reasonable basis for expecting a reasonably reproducible difference, this was not directly confirmed and should be verified in future work. Combined with systolic BP, this reproducible signature achieved an AUC of 0.733 (MCC 0.354) for discriminating unilateral from bilateral PA in internal cross-validation. Although point estimates are more modest than those obtained with the earlier, non-reproducibility-filtered feature set (AUC 0.826), they are, in our view, considerably more defensible: the predictive signal no longer relies on features whose values are substantially reader-

dependent. This finding supports, as a hypothesis, an approach integrating reproducible imaging-derived information with routinely available clinical variables; it does not establish clinical validity, and external validation remains necessary.

The individual-gland model similarly relied on reproducible features throughout: Compactness2 (ICC 0.763) and maximum 3D diameter (ICC 0.957), both morphological descriptors with good-to-excellent agreement. This finding illustrates that reproducibility and predictive information were not necessarily in conflict in this cohort once texture features derived from fine boundary detail were excluded a priori; rather, the earlier non-reproducible signature (GLCM variance, intensity interquartile range) appears to have captured boundary-dependent information that a stricter, reproducibility-first search was able to avoid while retaining discriminative features from a different part of the feature space (shape and gray-level distribution descriptors).

The potential value of explicit asymmetry metrics is supported by radiomics studies in bilateral organs, in which quantitative comparison of imaging features between paired structures has provided complementary discriminatory information beyond independent analysis of each side [27, 28]. In an earlier, non-reproducibility-filtered iteration of this analysis, the asymmetry-index representation also outperformed models entering both glands independently or either gland alone, although percentile intervals overlapped substantially; that comparison was not repeated with the reproducibility-filtered feature set and is not reported as a confirmatory result here. Explicit comparison between the two adrenal glands remains conceptually consistent with the pathophysiological definition of lateralization, and a workflow specifically designed to calculate inter-lateral asymmetry indices — as opposed to entering all features from both glands independently — should be prespecified for evaluation in an external validation cohort.

Systolic BP was associated with higher point-estimate AUCs, whereas aldosterone was not. The addition of systolic BP was associated with a higher point-estimate AUC across all four patient-level data representations. This finding is biologically plausible, as hypertension severity has been described as a clinical correlate of PA lateralization, with unilateral forms generally associated with more severe hypertension than bilateral disease [29, 30]. This interpretation is further supported by data from the Spanish SPAIN-ALDO registry, in which a combination of clinical, biochemical, and radiological features was highly predictive of unilateral PA. Specifically, among patients without relevant comorbidities, the presence of hypokalemia, systolic BP > 160 mmHg, PAC > 40 ng/dL, and a unilateral adrenal lesion > 1 cm with a normal contralateral adrenal gland was associated with a 98.5% probability of unilateral PA. These findings reinforce the concept that unilateral PA is frequently associated with a more severe clinical and biochemical phenotype and that integrating markers of disease severity with imaging findings may improve subtype prediction. In contrast, the additional inclusion of PAC in our current study did not improve performance and returned it to the radiomics-alone level, both for lateralization and, more markedly, for individual-gland identification (AUC 0.660 to 0.624). With a sample size of only 33 patients, increasing the number of predictors without a proportional gain in information increases the risk of overfitting, which may explain this pattern. From a practical perspective, these findings cautiously

suggest that systolic BP—a routinely available variable—may represent a more parsimonious complement to radiomic features than the additional inclusion of aldosterone in this cohort. Potassium was in fact evaluated as an additional clinical predictor in this study, without showing a relevant contribution; the ARR and saline infusion test results, however, were not available as alternative clinical predictors, and these variables were more informative in other series [31–33].

In an earlier, non-reproducibility-filtered iteration of this analysis, models restricted to the right adrenal gland alone showed markedly poorer performance (AUC 0.300–0.508) than those based on the left gland alone (AUC 0.728–0.791). This comparison was not repeated with the reproducibility-filtered feature set reported here and therefore does not bear directly on the confirmatory results in Tables 3 and 4; it is nonetheless noted because it may inform future work using single-gland representations. A plausible hypothesis, not confirmed statistically in this cohort, is the greater difficulty of manually delineating the right adrenal gland because of its anatomical proximity to the liver; in exploratory analyses, the right gland showed lower interpatient variability than the left (paired Wilcoxon test, statistically significant) and slightly greater morphological compactness, although the latter difference did not reach statistical significance ($p = 0.09$).

At the individual-gland level, the pattern is reversed. For identification of the aldosterone-producing adrenal gland, the radiomics-only model achieved the best performance (AUC 0.660, matched by the addition of systolic BP), with no improvement following the addition of clinical variables. This contrast with the patient-level analysis is consistent with the nature of these clinical variables: because they are measured at the patient level and are therefore identical for both adrenal glands within the same individual, they cannot provide information that distinguishes which gland is responsible for aldosterone excess in a patient with unilateral disease. This finding has a practical implication: as expected, systemic clinical variables do not provide gland-specific localization information. Importantly, this does not imply that radiomics alone is sufficient to guide adrenalectomy. In patients with bilateral disease, both adrenal glands were classified as aldosterone-producing based on the pathophysiological definition of diffuse bilateral hyperplasia, which should be considered when interpreting the sensitivity and specificity of this analysis [1].

Several limitations should be considered when interpreting these findings. First, the sample size was small (33 patients, 66 adrenal glands), as reflected by the wide intervals observed across all evaluated configurations. Several of these intervals, particularly those for the Matthews correlation coefficient, included the null value; therefore, the findings should be regarded as hypothesis-generating rather than definitive confirmatory evidence. Second, all results were derived from internal cross-validation within the same cohort, and no independent external cohort was available for validation. Third, although the final radiomic features were fixed before the confirmatory analysis, they were identified through exploratory analyses performed within the same cohort, reducing but not completely eliminating the risk of selection bias. Fourth, no comparison with conventional visual CT interpretation was available, and the model has not undergone validation according to TRIPOD recommendations [34]. Fifth, although the confirmatory features met a strict reproducibility threshold ($ICC(3,1) \geq 0.75$), ICC was not reported

separately for the right and left glands, and the reproducibility filter was derived from a 39-patient subsample rather than the full cohort. Finally, the proposed explanation for the poorer performance of the right-adrenal-gland model in the earlier, non-reproducibility-filtered iteration of this analysis remains exploratory and cannot be statistically confirmed given the available sample size; it was not re-examined with the reproducibility-filtered feature set.

5. Conclusions

In this multicenter exploratory study, CT-based radiomics showed a potential signal for PA classification when radiomic features were required, a priori, to meet an interobserver reproducibility threshold (ICC(3,1) ≥ 0.75) before any performance-based selection. This requirement excluded texture features that had shown good predictive performance but poor reproducibility in an earlier iteration of this analysis and instead retained features with good-to-excellent interobserver agreement (ICC 0.763–0.998): the inter-lateral difference in GLSZM small zone emphasis and gray-level non-uniformity for lateralization, and morphological Compactness2 and maximum 3D diameter for individual-gland identification. Explicit inter-lateral asymmetry, combined with systolic BP, yielded the highest point-estimate performance for distinguishing unilateral from bilateral disease (AUC 0.733); adding aldosterone provided no further benefit. At the individual-gland level, radiomics alone were sufficient (AUC 0.660), and no clinical variable improved on it. Because feature selection now prioritizes reproducibility, any residual laterality signal is less likely to be driven by reader-dependent boundary detail than in the earlier iteration of this analysis, although this has not been directly tested against an external reader panel. Given the small sample size and internal validation, these findings remain hypothesis-generating and require confirmation in larger, independent cohorts before clinical application.

Declarations

Conflicts of Interest: The authors declare no conflict of interest. Industry was not involved in any part of the study, including the design, data collection, analysis, or interpretation of the results.

Funding: This manuscript was funded by Instituto de Salud Carlos III (ISCIII), PI22/01525, co-funded by the European Union.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of the Hospital Universitario Ramón y Cajal, Madrid, Spain (approval date: 10th November 2020, code: ACTA 401)

Informed Consent Statement: The requirement for informed consent was waived by the Ethics Committee of the Hospital Universitario Ramón y Cajal, Madrid, Spain due to the retrospective nature of the study. Written informed consent was obtained from patients who continued follow-up or were prospectively included in the study.

Author Contributions: M.A.-C., D. P. M and M.A.G-B conceived and designed the study. M.A.-C., D. P. M and M. A.G-B contributed to patient selection and clinical data collection from the SPAIN-ALDO registry. D. P. M and M.A.G-B performed the adrenal CT image acquisition and/or radiological assessment, including independent adrenal gland segmentation. M.A.G.B., J.F.M. and A.F.I., performed the radiomics extraction and data preprocessing. D. P. M and M.A.G-B performed the statistical analysis, including reproducibility assessment, feature selection, stability analysis, and predictive modelling. M.A.-C., D. P. M and M.A.G-B interpreted the results. M.A.-C., D. P. M and M.A.G-B drafted the manuscript. All authors contributed to the critical revision of the manuscript for important intellectual content, approved the final version, and agreed to be accountable for all aspects of the work.

Data availability statement: The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Abstract count: 240: Number of keywords: 5; Word count: 4184; Number of tables: 6; Number of figures: 2; Number of references: 36

References

1. Adler, G. K. et al. Primary Aldosteronism: An Endocrine Society Clinical Practice Guideline. *J. Clin. Endocrinol. Metab.* **110**, 2453–2495. https://doi.org/10.1210/CLINEM/DGAF284 (2025).
2. Milliez, P. et al. Evidence for an increased rate of cardiovascular events in patients with primary aldosteronism. *J. Am. Coll. Cardiol.* **45**, 1243–1248. https://doi.org/10.1016/j.jacc.2005.01.015 (2005).
3. Monticone, S. et al. Cardiovascular events and target organ damage in primary aldosteronism compared with essential hypertension: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol.* **6**, 41–50. https://doi.org/10.1016/S2213-8587(17)30319-4 (2018).
4. Hundemer, G. L. et al. Renal outcomes in medically and surgically treated primary aldosteronism. *Hypertension* **72**, 658–666. https://doi.org/10.1161/HYPERTENSIONAHA.118.11568 (2018).
5. Reincke, M. et al. Observational study mortality in treated primary aldosteronism: The German conn's registry. *Hypertension* **60**, 618–624. https://doi.org/10.1161/HYPERTENSIONAHA.112.197111 (2012).
6. Araujo-Castro, M. et al. Practical consensus for the treatment and follow-up of primary aldosteronism: a multidisciplinary consensus document. *Endocrine* **85**, 532–544 (2024).
7. Funder, J. W. et al. The management of primary aldosteronism: Case detection, diagnosis, and treatment: An endocrine society clinical practice guideline. *J. Clin. Endocrinol. Metab.* **101**, 1889–1916. https://doi.org/10.1210/jc.2015-4061 (2016).
8. Menino, J., Vega, A., Goi, J. & Araujo-Castro, M. Baxdrostat for the treatment of primary aldosteronism: current evidence. Endocrine 2026;91. https://doi.org/10.1007/S12020-026-04570-2
9. Kempers, M. J. E. et al. Systematic review: Diagnostic procedures to differentiate unilateral from bilateral adrenal abnormality in primary aldosteronism. **151**. Ann Intern Med; (2009).

https://doi.org/10.7326/0003-4819-151-5-200909010-00007

10. Umakoshi, H. et al. Significance of Computed Tomography and Serum Potassium in Predicting Subtype Diagnosis of Primary Aldosteronism. *J. Clin. Endocrinol. Metab.* **103**, 900–908. https://doi.org/10.1210/JC.2017-01774 (2018).
11. Araujo-Castro, M. et al. Screening and diagnosis of primary aldosteronism. Consensus document of all the Spanish Societies involved in the management of primary aldosteronism. *Endocrine* **85**, 99–121. https://doi.org/10.1007/S12020-024-03751-1 (2024).
12. Rossi, G. P. et al. An expert consensus statement on use of adrenal vein sampling for the subtyping of primary aldosteronism. *Hypertension* **63**, 151–160. https://doi.org/10.1161/HYPERTENSIONAHA.113.02097 (2014).
13. Vonend, O. et al. Adrenal venous sampling: Evaluation of the german conn's registry. *Hypertension* **57**, 990–995. https://doi.org/10.1161/HYPERTENSIONAHA.110.168484 (2011).
14. Araujo-Castro, M. et al. Adrenal venous sampling in primary aldosteronism: Experience of a Spanish multicentric study (Results from the SPAIN-ALDO Register). *Endocrine* **78**, 363–372. https://doi.org/10.1007/S12020-022-03122-8 (2022).
15. Aerts, H. J. W. L. et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **5** https://doi.org/10.1038/NCOMMS5006 (2014).
16. Avanzo, M., Stancanello, J. & El Naqa, I. Beyond imaging: The promise of radiomics. *Phys. Med.* **38**, 122–139. https://doi.org/10.1016/j.ejmp.2017.05.071 (2017).
17. Broomand Lomer, N., Ahmadzadeh, A. M. & Azinfar, A. Preoperative differentiation of adrenal pheochromocytoma and lipid-poor adenoma using CT-based radiomics: a systematic review and meta-analysis. *Acta Radiol.* https://doi.org/10.1177/02841851261462504 (2026).
18. Mucha, M., Roszkowska, Z., Ambroziak, U., Bobrowicz, M. & Roszczyk, R. The application of artificial intelligence in adrenal imaging: current state of knowledge, challenges, and future directions. *Endocr. Connect.* https://doi.org/10.1530/EC-26-0267 (2026).
19. Ferro, M. et al. Artificial intelligence and radiomics applications in adrenal lesions: a systematic review. *Ther. Adv. Urol.* 17. https://doi.org/10.1177/17562872251352553 (2025).
20. He, K. et al. A Clinical-Radiomic Nomogram Based on Unenhanced Computed Tomography for Predicting the Risk of Aldosterone-Producing Adenoma. Front Oncol 2021;11. https://doi.org/10.3389/FONC.2021.634879/PDF
21. Chen, P. T. et al. Radiomics utilization to differentiate nonfunctional adenoma in essential hypertension and functional adenoma in primary aldosteronism. *Sci. Rep.* 12. https://doi.org/10.1038/S41598-022-12835-9 (2022).
22. Chen, P. T. et al. Machine Learning Model with Computed Tomography Radiomics and Clinicobiochemical Characteristics Predict the Subtypes of Patients with Primary Aldosteronism. *Acad. Radiol.* **31**, 1818–1827. https://doi.org/10.1016/j.acra.2023.10.015 (2024).
23. Mansour, N. et al. Integration of clinical parameters and CT-based radiomics improves machine learning assisted subtyping of primary hyperaldosteronism. *Front. Endocrinol. (Lausanne)*. **14**,

1244342. https://doi.org/10.3389/fendo.2023.1244342 (2023).

24. O'Shea, A., Kilcoyne, A., McDermott, E., O'Grady, M. & McDermott, S. Can radiomic feature analysis differentiate adrenal metastases from lipid-poor adenomas on single-phase contrast-enhanced CT abdomen? *Clin. Radiol.* **77**, e711–e718. https://doi.org/10.1016/j.crad.2022.06.015 (2022).
25. Williams, T. A. et al. Outcomes after adrenalectomy for unilateral primary aldosteronism: an international consensus on outcome measures and analysis of remission rates in an international cohort. *Lancet Diabetes Endocrinol.* **5**, 689–699. https://doi.org/https://dx.doi.org/10.1016/S2213-8587(17)30135-3 (2017).
26. Koo, T. K. & Li, M. Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **15**, 155–163. https://doi.org/10.1016/J.JCM.2016.02.012 (2016).
27. Turcu, A. F. & Auchus, R. Approach to the Patient with Primary Aldosteronism: Utility and Limitations of Adrenal Vein Sampling. *J. Clin. Endocrinol. Metab.* **106**, 1195–1208. https://doi.org/10.1210/CLINEM/DGAA952 (2021).
28. Jones, M. A., Sadeghipour, N., Chen, X., Islam, W. & Zheng, B. A multi-stage fusion framework to classify breast lesions using deep learning and radiomics features computed from four-view mammograms. *Med. Phys.* **50**, 7670–7683. https://doi.org/10.1002/MP.16419 (2023).
29. Rossitto, G. et al. Subtyping of primary aldosteronism in the AVIS-2 study: Assessment of selectivity and lateralization. *J. Clin. Endocrinol. Metab.* **105**, 2042–2052. https://doi.org/10.1210/clinem/dgz017 (2020).
30. Wannachalee, T. et al. Three Discrete Patterns of Primary Aldosteronism Lateralization in Response to Cosyntropin During Adrenal Vein Sampling. *J. Clin. Endocrinol. Metab.* **104**, 5867–5876. https://doi.org/10.1210/JC.2019-01182 (2019).
31. Küpers, E. M., Amar, L., Raynaud, A., Plouin, P. F. & Steichen, O. A clinical prediction score to diagnose unilateral primary aldosteronism. *J. Clin. Endocrinol. Metab.* **97**, 3530–3537. https://doi.org/10.1210/jc.2012-1917 (2012).
32. Kološová, B. et al. Validation of Existing Clinical Prediction Tools for Primary Aldosteronism Subtyping. *Diagnostics (Basel Switzerland)*. 12. https://doi.org/10.3390/DIAGNOSTICS12112806 (2022).
33. Hashimura, H. et al. Saline suppression test parameters may predict bilateral subtypes of primary aldosteronism. *Clin. Endocrinol. (Oxf)*. **89**, 308–313. https://doi.org/10.1111/CEN.13757 (2018).
34. Moons, K. G. M., Altman, D. G., Reitsma, J. B. & Collins, G. S. New Guideline for the Reporting of Studies Developing, Validating, or Updating a Multivariable Clinical Prediction Model: The TRIPOD Statement. *Adv. Anat. Pathol.* **22**, 303–305. https://doi.org/10.1097/PAP.0000000000000072 (2015).

Figures

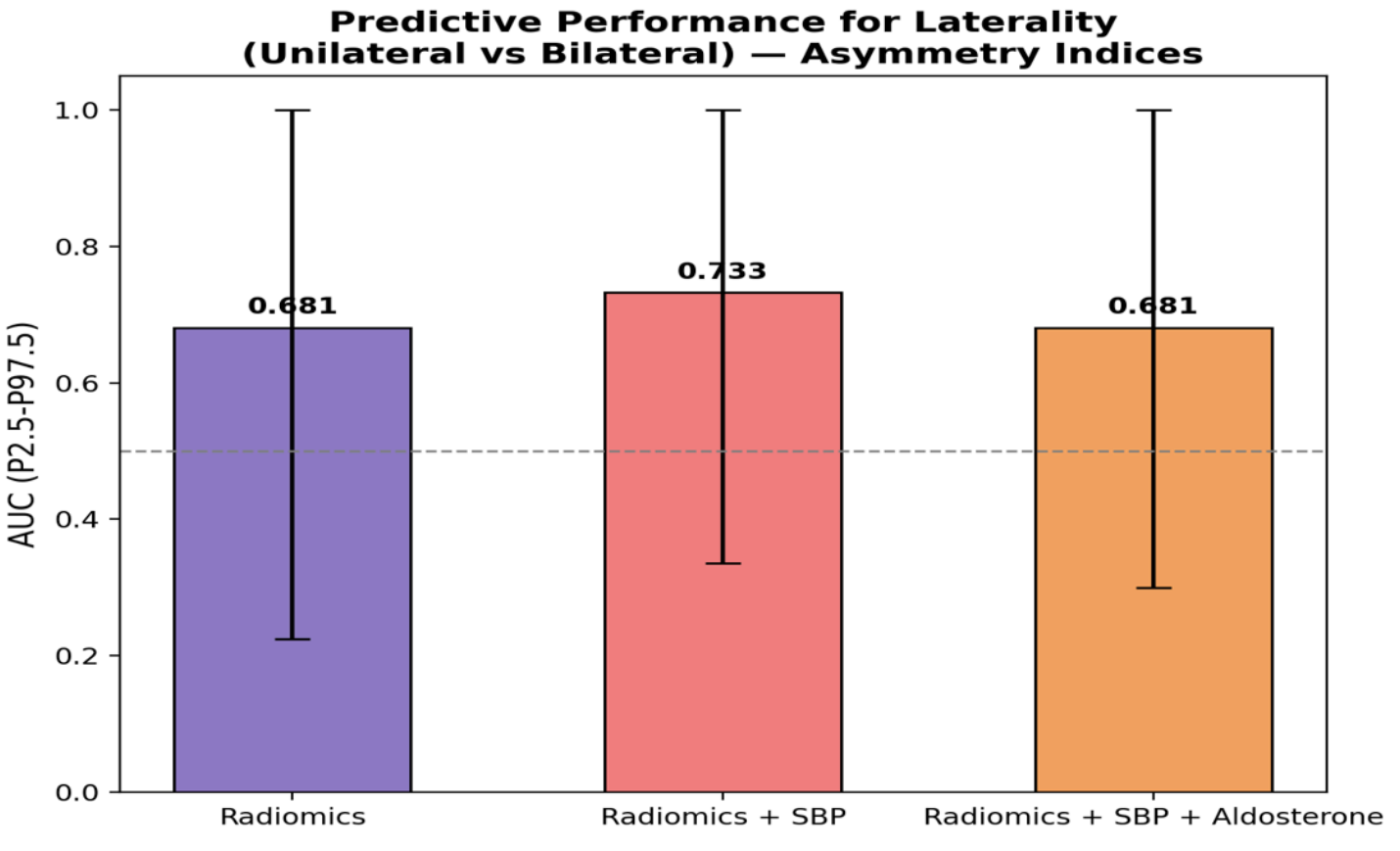


Figure 1

Predictive performance (mean AUC, bar chart) for discriminating unilateral from bilateral primary aldosteronism at the patient level, using inter-lateral asymmetry indices (two reproducibility-filtered features: GLSZM small zone emphasis and gray-level non-uniformity) across three clinical-variable configurations. Error bars represent the P2.5-P97.5 percentile range across 100 repeated stratified cross-validation partitions. n=33 patients (22 unilateral, 11 bilateral).

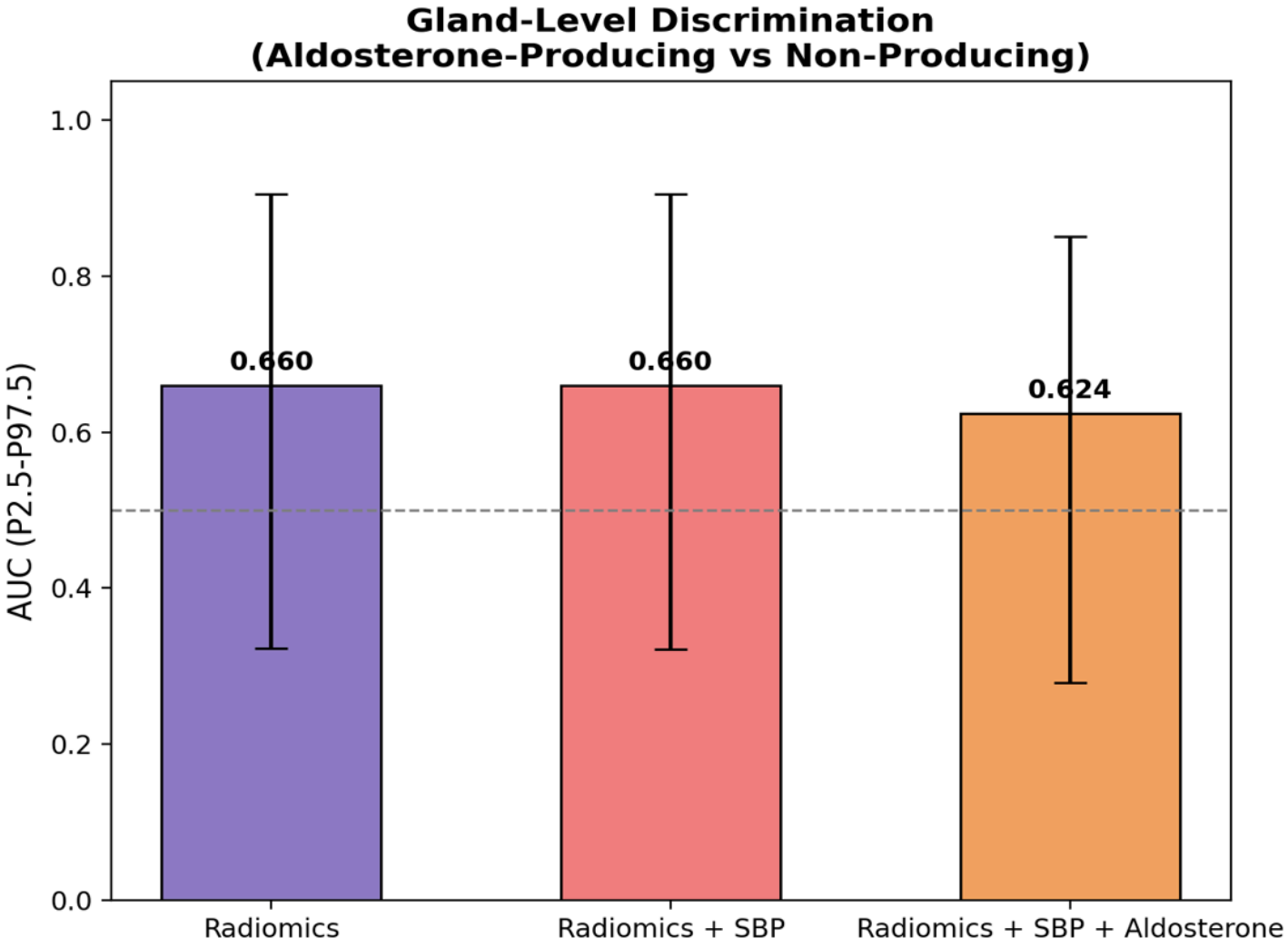


Figure 2

Predictive performance (mean AUC, bar chart) for identifying the aldosterone-producing adrenal gland at the individual-gland level, using two reproducibility-filtered morphological features (Compactness2 and maximum 3D diameter) across three clinical-variable configurations. Error bars represent the P2.5-P97.5 percentile range across 100 patient-grouped cross-validation partitions (not ROC curves). n=66 glands (33 patients).