BMC Medical Imaging

RESEARCH

Open Access

Development and validation of a model predicting adrenal lipid-poor adenoma based on the minimum attenuation value from non-contrast CT: a dual-center retrospective study

Hanlin Zhu[1,3†], Mengwei Wu[2,3†], Bo Feng[1], Haifeng Zhang[1], Chunfeng Hu[3], Tong Zhang[3] and Zhijiang Han[3*]

Abstract

Objective The early differentiation of adrenal lipid-poor adenomas from non-adenomas is a crucial step in reducing excessive examinations and treatments. This study seeks to construct an eXtreme Gradient Boosting (XGBoost) predictive model utilizing the minimum attenuation values (minAVs) from non-contrast CT (NCCT) scans to identify lipid-poor adenomas.

Materials and methods Retrospective analysis encompassed clinical data, minAVs, CT histogram (CTh), mean attenuation values (meanAVs), and lesion diameter from patients with pathologically or clinically confirmed adrenal lipid-poor adenomas across two medical institutions, juxtaposed with non-adenomas. Variable selection transpired in Institution A (training set), with XGBoost models established based on minAVs and CTh separately. Institution B (validation set) corroborated the diagnostic efficacy of the two models. Receiver operator characteristic (ROC) curve analysis, calibration curves, and Brier scores assessed the diagnostic performance and calibration of the models, with the Delong test gauging differences in the area under the curve (AUC) between models. SHapley Additive exPlanations (SHAP) values elucidated and visualized the models.

Results The training set comprised 136 adrenal lipid-poor adenomas and 126 non-adenomas, while the validation set included 46 and 40 instances, respectively. In the training set, there were substantial inter-group differences in minAVs, CTh, meanAVs, diameter, and body mass index (BMI) ($p < 0.05$ for all). The AUC for the minAV and CTh models were 0.912 (95% confidence interval [CI]: 0.866–0.957) and 0.916 (95% CI: 0.873–0.958), respectively. Both models exhibited good calibration, with Brier scores of 0.141 and 0.136. In the validation set, the AUCs were 0.871 (95% CI: 0.792–0.951) and 0.878 (95% CI: 0.794–0.962), with Brier scores of 0.156 and 0.165, respectively. The Delong test revealed no statistically significant differences in AUC between the models ($p > 0.05$ for both). SHAP value analysis for the minAV model suggested that minAVs had the highest absolute weight (AW) and negative contribution.

†Hanlin Zhu and Mengwei Wu contributed equally to this work.

*Correspondence:
Zhijiang Han
hzj1022@zju.edu.cn
Full list of author information is available at the end of the article

Conclusion The XGBoost predictive model based on minAVs demonstrates effective discrimination between adrenal lipid-poor adenomas and non-adenomas. The minAV variable is easily obtainable, and its diagnostic performance is comparable to that of the CTh model. This provides a basis for patient diagnosis and treatment plan selection.

Keywords Adrenal incidentaloma, Lipid-poor adenoma, Computed tomography, Minimum attenuation value, SHAP

Introduction

Adrenal incidentaloma refers to adrenal nodules with a diameter greater than 10.0 mm discovered incidentally during the assessment of non-adrenal diseases [1]. In adults, the prevalence of adrenal incidentalomas ranges from approximately 1.0% to 6.0% [2], with the highest incidence observed in the 50 to 60 age group, reaching around 7.0% [3]. Among adrenal incidentalomas, 75.0% are non-functional adenomas, 14.0% are functional adenomas, and non-adenomas, including metastatic tumors and pheochromocytomas, account for 11.0% [4]. The differentiation between functional and non-functional adrenal adenomas is primarily accomplished through specialized endocrine laboratory tests. At the same time, imaging examinations are mainly employed to distinguish between adenomas and non-adenomas and to monitor changes in tumor size [5].

Non-contrast CT (NCCT) is a commonly employed method for diagnosing adrenal adenomas. In the diagnosis of adrenal adenomas, a threshold of 10 Hounsfield units (HU) is widely accepted to classify adenomas into lipid-rich adenomas (≤10 HU) and lipid-poor adenomas (>10 HU), based on the abundant lipid content in the cytoplasm of lipid-rich adenomas. This standard is based on the widely accepted CT value classification method in the field. The lipid-rich adenomas are diagnostically more straightforward due to the abundance of lipid content in their cytoplasm. However, approximately 30.0% of adrenal adenomas are lipid-poor, posing challenges in their differentiation from non-adenomas using NCCT alone [6]. Additionally, for solid tumor components, a threshold of >20 HU increase in attenuation value on post-contrast CT is defined as significant enhancement, commonly used to distinguish enhancing lesions, such as metastatic tumors, from non-enhancing ones. Consequently, the early and accurate diagnosis of lipid-poor adenomas has emerged as a pivotal research focus. Bae et al. [7] pioneered the use of CT histogram (CTh) technology for diagnosing adrenal adenomas, achieving a sensitivity of 82.0% and a specificity of 100.0%, respectively. Nevertheless, the sensitivity of CTh in differentiating lipid-poor adenomas significantly decreases, highlighting the need for further research in this area. In the lipid-poor adenoma subgroup, CTh specificity remained constant, but sensitivity dramatically decreased, reported in the literature as only 46.0% to 53.0% [8, 9]. Moreover, CTh calculation necessitates specific software for assessment, adding complexity to clinical applications [10]. In 2016, our research team [11] discovered that the NCCT minimum attenuation value (minAV) can predict adrenal adenomas. In 2022, dual-institution data were incorporated to validate the value of minAVs in diagnosing lipid-poor adenomas, achieving sensitivity ranging from 58.6% to 69.1% at specificities of 95.8% to 98.6% [12]. While minAVs can enhance sensitivity in diagnosing lipid-poor adenomas, the diagnosis of adrenal lesions in clinical practice is typically intricate, with a single variable insufficient to capture comprehensive lesion information. Therefore, this study intends to construct a multivariate eXtreme Gradient Boosting (XGBoost) predictive model based on minAVs, contrasting it with the CTh model. The objective is to elucidate the efficacy of the minAV model in discriminating lipid-poor adenomas from non-adenomas and to harness SHapley Additive exPlanations (SHAP) [13] values for the interpretation and visualization of the XGBoost model.

Materials and methods

This study was approved by the Ethics Committee of the First People's Hospital affiliated with Westlake University and Affiliated Quzhou Hospital of Wenzhou Medical University (Quzhou People's Hospital) (Approval numbers: IIT-20220811–0128-01, QZPYH-ER-2023–007), and patient informed consent was waived. Furthermore, it adhered to the ethical standards set forth by the World Medical Association's Declaration of Helsinki.

Study population

The study retrospectively analyzed patients with adrenal incidentalomas who underwent NCCT examinations and possessed surgical and puncture-related pathological outcomes or met reference standards through clinical confirmation from January 2017 to December 2023 at Medical Institution A (Hangzhou First People's Hospital affiliated with Westlake University Medical School) and B (Zhejiang Quzhou People's Hospital).

The specific criteria for clinical confirmation are as follows:

① A clinical diagnosis of adrenal metastatic tumors is established when patients have a history of extra-adrenal malignant tumors, and new adrenal lesions are identified on imaging or there is an increase in the volume of

pre-existing lesions (diameter growth rate > 5 mm/year) [14]. For solid tumor components, while an increase in attenuation value of > 20 HU on post-contrast CT is commonly used to define significant enhancement, in this study, we adopted a threshold of > 10 HU difference in attenuation values before and after CT enhancement, as per the methodology of Nagayama et al. [15]. This lower threshold helps in more sensitively identifying potential solid components, thereby facilitating a more accurate assessment of adrenal incidentalomas. ② All other non-metastatic lesions require pathological confirmation. ③ The distinction between lipid-rich and lipid-poor adenomas is made based on the pre-contrast CT attenuation value, with a threshold of 10 HU used to classify adenomas into lipid-rich (≤ 10 HU) and lipid-poor (> 10 HU) categories.

The inclusion criteria for patients with adrenal incidentalomas are as follows: ① Age ≥ 18 years; ② All patients with adrenal incidentalomas who have undergone NCCT examinations of the abdomen or chest, with comprehensive imaging that includes adrenal lesions; ③ In cases where patients have bilateral adrenal lesions, only the largest one in diameter is selected for inclusion to mitigate clustering effects. The exclusion criteria encompass: ① Average NCCT value of adrenal lesions ≤ 10 HU; ② Maximum diameter of the tumor < 10.0 mm; ③ Adrenal cysts or tumors with a solid component < 50.0%; ④ CT scan slice thickness greater than 5.0 mm; ⑤ Unqualified CT image quality, presence of foreign bodies, or motion artifacts, among others.

According to the inclusion and exclusion criteria, a total of 348 lesions were enrolled across the two medical institutions (Fig. 1), comprising 185 cases in females and 163 cases in males, with ages ranging from 18 to 90 years and a mean age of 53.4 ± 14.8 years. Among the incorporated adrenal incidentalomas, there were 182 lipid-poor adenomas, 96 metastatic tumors (65 clinically confirmed), 41 pheochromocytomas, 11 lymphomas, 9 ganglioneuromas, 5 cortical carcinomas, and 4 schwannomas.

Image techniques

All CT images at Medical Institution A were acquired using the Light-Speed 16-slice Spiral CT system (General Electric Healthcare, Wisconsin, USA), Revolution 64-slice Spiral CT system, and SOMATOM Definition Flash 128-slice Spiral CT system (Siemens, Munich, Germany). The imaging parameters for NCCT were as follows: 120 kV, automatic mAs modulation, and image reconstruction with slice thicknesses of 5.00 mm and 3.75 mm.

Medical Institution B utilized the Optima 16-slice Spiral CT system (General Electric Healthcare, Wisconsin, USA) and uCT510 32-slice Spiral CT system (United Imaging, Shanghai, China) for obtaining all NCCT images. The imaging parameters for NCCT were consistent, featuring 120 kV, automatic mAs modulation, and image reconstruction with slice thicknesses of 5.00 mm and 3.75 mm.

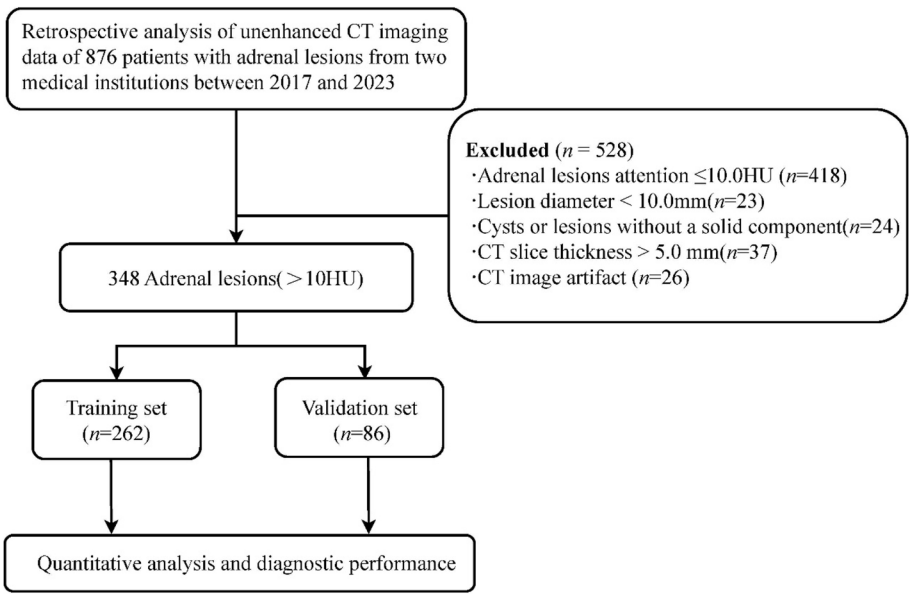

Fig. 1 Flow diagram of patients investigated in this study

Image analysis

CT quantitative data were acquired by radiologists A (with 9 years of experience) and B (with 7 years of experience) using the PACS system (RadinfoSystems, Zhejiang, China) workstation and uAI software (version V1.4, United Imaging, Shanghai, China), under the premise of being unaware of pathological results. Simultaneously, clinical baseline information such as gender, age, and BMI was recorded. Four weeks later, radiologist B randomly selected 100 lesions to measure again, assessing the repeatability of the quantitative data.

For minAV measurement, a 4-point method was employed, involving the measurement of the maximum transverse and longitudinal diameters on the plane of the lesion's largest axis to obtain the intersection. The midpoint between the intersection and the tumor edge was designated as the measurement point, with a region of interest (ROI) range of 19.0 ~ 24.0 mm^2. The measurement method for CTh and meanAVs involved selecting an ROI on the plane of the lesion's largest axis that encompassed at least 2/3 of the lesion, and uAI software was utilized to extract CTh data from the ROI. Care was taken to avoid areas of necrosis, bleeding, and calcification when selecting each ROI.

Model construction and statistical analysis

In the data from Institution A (training set), univariate analysis was adopted to screen variables, retaining those with p values < 0.05. Subsequently, two XGBoost diagnostic models were built based on minAVs and CTh, respectively. The diagnostic performance of both models was compared in Institution B (validation set). SHAP values were harnessed to interpret and visualize the minAV model.

The Shapiro–Wilk test evaluated the normality of continuous variables. Normally distributed data were presented using mean ± standard deviation, with inter-group differences assessed using the Student's t-test. Non-normally distributed data were represented using medians and quartiles, and inter-group differences were evaluated using the Wilcoxon rank-sum test. Categorical variables were expressed as proportions (percentages), with between-group comparisons conducted using the chi-square test. The inter-rater correlation coefficient (ICC) assayed the repeatability of minAVs and CTh. The diagnostic performance of the minAV and CTh models was confirmed by calculating the area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, and accuracy. Calibration curves were introduced to examine model calibration, whereas the Brier score quantified calibration degree, with scores below 0.1 reflecting excellent calibration, 0.1 to 0.2 considered good, 0.2 to 0.3 regarded as fair, and scores above 0.3 denoting poor calibration. Comparison between the two ROC curves employed the Delong test. A p value < 0.05 was considered statistically significant. All analyses were carried out with the assistance of Python software (version 3.7.1, Python Software Foundation, Delaware, USA) and R software (version 4.0.1; R Foundation for Statistical Computing, Vienna, Austria).

Results

General characteristics of lipid-adenoma and non-adenoma

The training set encompassed 262 adrenal lesions (Table 1), comprising 136 lipid-poor adenomas (51.9%) and 126 non-adenomas (48.1%). Gender distribution was equal, with males and females each accounting for 50.0%, and the median age was 55 years. The validation set encompassed 86 adrenal lesions, with 46 lipid-poor adenomas (53.5%) and 40 non-adenomas (46.5%). In this set, there were 32 males (37.2%) and 54 females (62.8%), with a median age of 50 years.

Univariate analysis of lipid-poor adenoma and non-adenoma

In the training set, adrenal lipid-poor adenomas exhibited lower meanAVs, minAVs, and diameters compared to non-adenomas (medians: 26.8 HU vs 34.1 HU, -1.6 HU vs 10.6 HU, 22.8 mm vs 31.1 mm, p values all < 0.001). Additionally, the negative pixel percentage and BMI were higher in the lipid-poor adenoma group vis-à-vis the non-adenoma group (medians: 3.3% vs 0.0%, 23.8 kg/m^2 vs 21.6 kg/m^2, p values all < 0.001). Nevertheless, there existed no statistically significant disparities in terms of age, gender, or lesion location (Table 1).

Repeatability analysis of quantitative CT metrics

The ICC for minAVs demonstrated a slightly superior level compared to that of CTh. Precisely, the recorded values were 0.90 (95% CI: 0.83, 0.94) and 0.86 (95% CI: 0.75, 0.91), respectively (p < 0.05 for both).

Comparison of the diagnostic performance of the two models

In the training set, the AUC values for minAV and CTh models were comparable (0.912 vs. 0.916). The Delong test unveiled no remarkable difference between groups (Z = -0.372, p = 0.712). Sensitivity for minAV and CTh models was 83.1% and 81.6%, respectively, while specificity was 84.1% and 86.5%, and accuracy was 83.6% and 84.0% (Table 2 and Fig. 2A). Both the minAV and CTh models exhibited good calibration, with Brier scores of 0.141 and 0.136, respectively (Fig. 2B). In the external validation set, the AUC values for the minAV and CTh models were similarly comparable (0.871

Table 1 Baseline clinical imaging characteristics of patients

Variables	Training set			Validation set		
	Lipid-poor adenoma (*n* = 136)	Non-adenoma (*n* = 126)	*p*	Lipid-poor adenoma (*n* = 46)	Non-adenoma (*n* = 40)	*p*
Age(year), Med (Q1,Q3) [b]	53.0 (44.0, 62.0)	56.0 (46.0, 67.0)	0.190	51.5 (43.2, 61.0)	47.5 (36.5, 59.5)	0.268
Gender, n (%)[c]			0.536			0.106
Female	71 (52.2)	60 (47.6)		33 (71.7)	21 (52.5)	
Male	65 (47.8)	66 (52.4)		13 (28.3)	19 (47.5)	
Location, n (%)[c]			0.726			0.022
Right	49 (36.0)	49 (38.9)		13 (28.3)	22 (55.0)	
Left	87 (64.0)	77 (61.1)		33 (71.7)	18 (45.0)	
BMI(kg/m^2), Med (Q1,Q3)[b]	23.8 (22.1, 26.2)	21.6 (21.5, 24.8)	**< 0.001**	23.8 (21.8, 24.5)	21.6 (21.2, 25.1)	0.135
Diameter(mm), Med (Q1,Q3)[b]	22.8 (16.7, 28.0)	31.1 (22.3, 40.0)	**< 0.001**	24.8 (19.4, 29.0)	35.5 (25.7, 43.5)	**< 0.001**
MeanAV (HU), Med (Q1,Q3)[b]	26.8 (20.2, 33.7)	34.1 (29.4, 37.9)	**< 0.001**	24.6 (21.6, 31.6)	33.1 (30.0, 37.5)	**< 0.001**
CTh (%), Med (Q1,Q3)[b]	3.3 (0.5, 10.0)	0.0 (0.0, 0.5)	**< 0.001**	1.2 (0.5, 3.4)	0.0 (0.0, 0.3)	**< 0.001**
MinAV (HU), Mean ± SD[a]	-0.3 ± 13.8	10.0 ± 11.9	**< 0.001**	0.0 ± 12.3	9.9 ± 10.2	**< 0.001**
Spacing (mm), n (%)[c]			1.000			0.427
3.75	72 (52.9)	67 (53.2)		18 (39.1)	20 (50.0)	
5.00	64 (47.1)	59 (46.8)		28 (60.9)	20 (50.0)	

Med Median, *Q1* First Quartile, *Q3* Third Quartile SD, standard deviation, *MeanAV* Mean attenuation value, *MinAV* Minimum attenuation value, CTh (%), Percentage negative pixels, *BMI* Body Mass Index, *HU* Hounsfield unit

[a] Student's *t* test

[b] Wilcoxon rank-sum test

[c] chi-square test

Table 2 Comparison of the diagnostic performance of the models

	AUC	Sensitivity	Specificity	Accuracy	PPV	NPV
Training set						
MinAV model	0.912 (0.866–0.957)	83.1% (73.5%-93.4%)	88.1% (75.4%-95.2%)	85.1% (80.9%-88.9%)	87.9% (79.5%-94.6%)	82.7% (75.8%-92.3%)
CTh model	0.916 (0.873–0.958)	83.8% (75.0%-92.6%)	88.1% (77.8%-95.2%)	85.9% (81.7%-89.7%)	88.2% (81.0%-94.4%)	83.6% (77.2%-90.7%)
Validation set						
MinAV model	0.871 (0.792–0.951)	84.8% (69.6%-95.7%)	85.0% (70.0%-95.0%)	83.7% (76.7%-91.9%)	86.4% (77.1%-95.3%)	82.2% (71.7%-93.8%)
CTh model	0.878 (0.794–0.962)	84.8% (73.9%-95.7%)	85.0% (72.5%-95.0%)	84.9% (76.7%-91.9%)	86.7% (77.8%-95.1%)	83.7% (73.9%-94.4%)

AUC area under the curve, *CI* confidence interval, *MinAV* minimum attenuation value, *CTh* CT histogram, *PPV* Positive Predictive Value, *NPV* Negative Predictive Value, The parentheses indicate the 95% confidence interval

vs. 0.878) (Fig. 2C). The Delong test showed no notable difference between the two models (Z = -0.243, $p = 0.813$). Sensitivity, specificity, and accuracy for the minAV model were 82.6%, 71.7%, and 82.5%, respectively, and for the CTh model, they were 87.5%, 82.6%, and 79.1%. Both models demonstrated good calibration, with Brier scores of 0.156 and 0.165, respectively (Fig. 2D).

Subgroup analysis

Based on the First Quartile and Third Quartile of meanAVs in adrenal lesions, three subgroups were created (Table 3). The AUC values for the minAV model ranged from 0.718 to 0.863 across the subgroups, while the AUC values for the CTh model ranged from 0.700 to 0.838. In the meanAV 10–25 HU subgroup and the > 35 HU subgroup, the minAV model outperformed the CTh

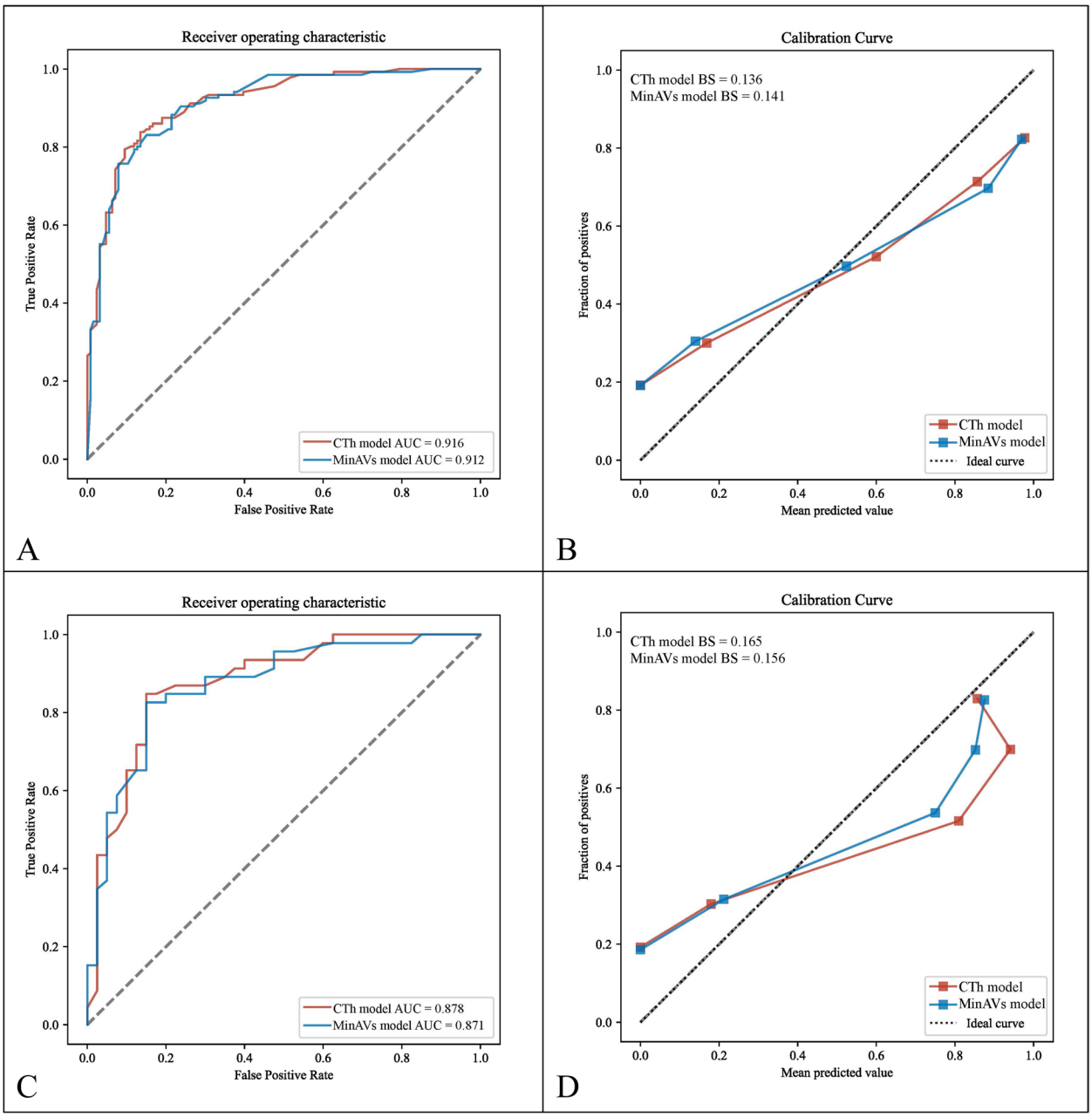

Fig. 2 ROC analysis and calibration curves of minAV and CTh XGBoost models Figures **A** and **B** represent the training set, while Figures **C** and **D** correspond to the validation set. Note: AUC stands for the area under the curve, and BS represents the Brier score

model in terms of AUC (0.863 vs. 0.784 and 0.718 vs. 0.700, respectively). Within the subgroup with meanAVs greater than 25 HU to 35 HU, the former exhibited a slightly lower AUC compared to the latter (0.808 vs. 0.838).

SHAP value analysis of the minAV model

The analysis of SHAP values for the minAV model unraveled that the variables, listed in descending order as per their weights, were lesion minAVs, diameter, BMI, and meanAVs. The absolute average SHAP values for these variables stood at 0.241, 0.154, 0.148, and 0.04, respectively. MinAV, diameter, and meanAV were considered negative contributions, whereas BMI was regarded as a positive contribution (Fig. 3). Two cases were selected from the external validation set for individualized predictions using SHAP values (Fig. 4).

Table 3 Subgroup analysis of diagnostic performance for the minAV model and CTh model

	MinAV model			CTh model		
	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
S1	0.863 (0.725–0.988)	97.5% (93.6%-100%)	75.0% (50.0%-100.0%)	0.784 (0.635–0.906)	90.1% (83.1%-96.3%)	66.7% (37.5%-90.9%)
S2	0.808 (0.744–0.866)	79.1% (68.9%-87.8%)	82.5% (74.4%-89.8%)	0.838 (0.776–0.893)	82.1% (72.1%-90.4%)	85.6% (77.8%-92.0%)
S3	0.718 (0.618–0.809)	55.9% (39.3%-72.7%)	87.7% (79.2%-95.7%)	0.700 (0.607–0.792)	47.1% (30.3%-64.3%)	93.0% (85.2%-98.4%)

S Subgroup, S1 > 10HU, ≤ 25HU (81 adenomas, 12 non-adenomas), S2 > 25HU, ≤ 35HU (67 adenomas, 97 non-adenomas), S3, > 35HU (34 adenomas, 57 non-adenomas); AUC, area under the curve; the parentheses indicate the 95% confidence interval

Discussion

This research conducted a comparative analysis involving 182 adrenal lipid-poor adenomas and 166 non-adenomas from two medical institutions. Two multivariate XGBoost models were developed based on minAVs and CTh. Our study results indicate that there was no substantial difference in the predictive performance of the two models for lipid-poor adenomas. In the SHAP value analysis, minAVs exhibited the highest weight in the model, and its simplicity of acquisition makes it a crucial indicator for diagnosing lipid-poor adenomas. Thus, this study highly recommends the use of the minAV model in clinical practice.

While MRI combined with chemical shift (CS) sequences has been widely recognized as an effective method for characterizing adrenal lesions (AL), its diagnostic accuracy is inherently dependent on the abundance of intracellular lipids within adenoma cells [16]. As emphasized by Stanzione et al. [16], when adrenal tumors contain insufficient cytoplasmic lipids, signal loss or heterogeneous signal loss on CS out-of-phase images may not be observed, encompassing lesions such as lipid-poor adenomas and non-adenomas. Consequently, differentiating between lipid-poor adenomas and non-adenomas remains a formidable challenge even with MRI. Our proposed XGBoost model based on minimum attenuation values (minAVs) from non-contrast CT provides an efficacious solution to this clinical conundrum, serving as a complementary approach to MRI in the evaluation of indeterminate adrenal lesions.

Despite the lower lipid content in adrenal lipid-poor adenomas, it remains higher than non-adenomas, making accurate detection a key aspect of distinguishing between the two [11]. In 2019, Zbyněk et al. [9] utilized CTh for the diagnosis of adrenal lipid-poor adenomas. CTh tallies the CT values of all pixels within the tumor ROI, displaying the distribution of pixel values in a histogram. It distinguishes adenomas from non-adenomas by calculating the percentage of negative pixels. Notwithstanding, lipid-poor adenomas have a lower percentage of negative pixels, and when using the widely accepted standard of a negative pixel percentage of 10.0%

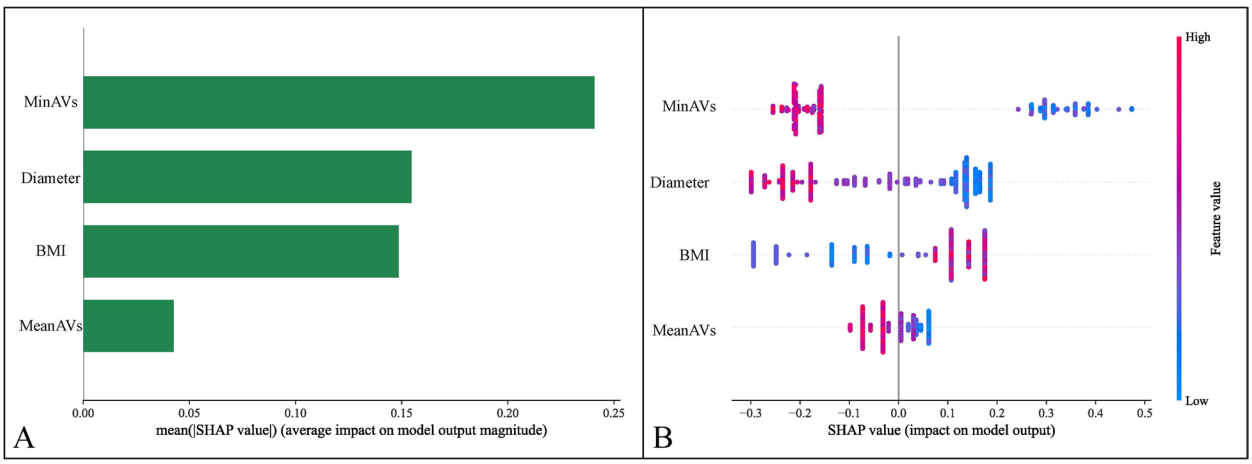

Fig. 3 SHAP value analysis of the minAV XGBoost model (**A**) Bar graph of absolute SHAP values, where the x-axis is arranged in descending order of variable weights, and the y-axis represents absolute SHAP values. **B** Scatter diagram of SHAP values, with each point representing a patient; the color indicates the magnitude of the variable value, with blue points on the right reflecting negative contributions and red points on the right indicating positive contributions

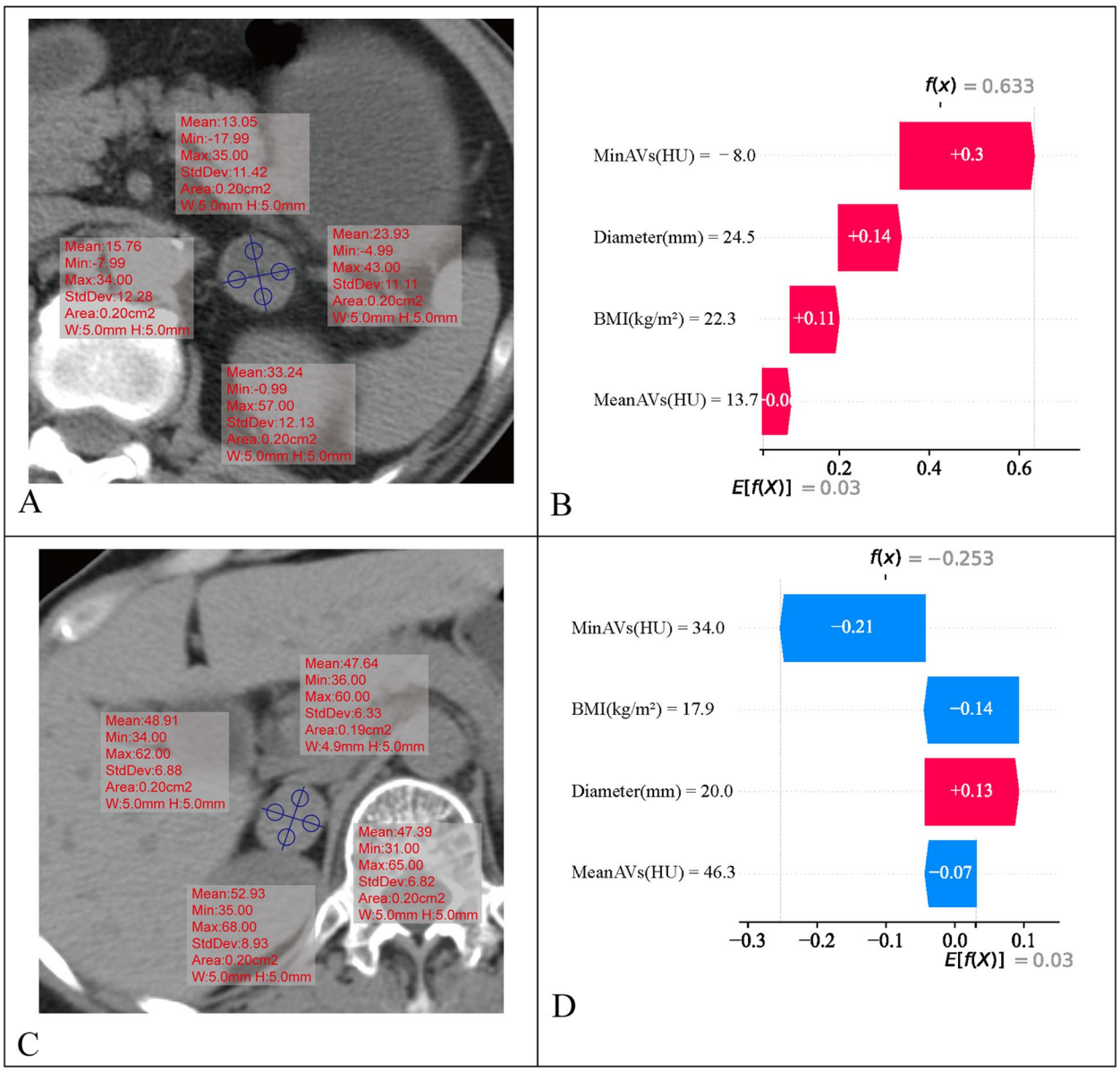

Fig. 4 Visualization of true positive and true negative cases using SHAP (**A**) Left adrenal lipid-poor adenoma in a patient, with various CT quantitative parameters obtained through the 4-point measurement method; (**B**) SHAP value waterfall plot, the minAV model predicted adrenal lipid-poor adenoma with f(x) = 0.633, and probability converted through the sigmoid function was 65.3%; (**C**) Right adrenal pheochromocytoma in a patient; (**D**) The minAV model predicted lipid-poor adenoma with f(x) = -0.253, probability converted through the sigmoid function was 44.4%

for diagnosing adenomas, the sensitivity is only 53.0%, with a specificity of 100.0%.In a retrospective single-institution study conducted by Wu et al. in 2020 [12], which incorporated 58 adrenal lipid-poor adenomas and 36 non-adenomas,NCCT data were analyzed using a Gaussian index reflecting the distribution of negative pixels in CTh after noise correction to forecast lipid-poor adenomas.The Gaussian index demonstrated an AUC of 0.77, with sensitivity and specificity of 89.7% and 55.6%, respectively. Although the Gaussian index can augment the sensitivity of diagnosing adrenal lipid-poor adenomas, specificity conspicuously decreases. Additionally, the Gaussian index requires complex calculations and equipment parameter calibration, limiting its clinical application. In our work, the minAV model predicted the sensitivity and specificity of adrenal lipid-poor adenomas in data from two medical institutions, both exceeding 80.0%. Within the subgroup with values greater than 35HU, the minAV model outperformed the CTh model in terms of AUC and sensitivity, suggesting that minAVs can

accurately detect the lipid components within the cytoplasm of lipid-poor adenomas [12]. Furthermore, this research harnessed SHAP values for the visualization of XGBoost models. SHAP values, as a post hoc explanation method, quantify the contribution of each variable to the model's output, addressing the "black box" problem in machine learning and enhancing the model's interpretability [13].

In the SHAP value analysis of XGBoost models, minAVs displayed the highest weight, with a value of 0.241, surpassing that of the lesion diameter (0.154), as well as the combined effect of BMI and meanAVs (0.188). MinAVs contributed negatively to the model, indicating that smaller minAVs increased the probability of diagnosing adrenal lipid-poor adenomas. However, when drawing a ROI that encompasses at least half to two-thirds of the lesion, operators find it challenging to avoid minor necrosis and cystic alterations [17], leading to a reduction in the diagnostic specificity of minAVs. In comparison, the 4-point measurement approach reflects the heterogeneity of different regions within the lesion while avoiding necrosis and cystic changes, thereby attenuating the false-positive rate of minAVs. The absolute SHAP value of adrenal lesion diameter in the model ranked second only to minAVs. Currently, most scholars support using 4.0 cm as a threshold to differentiate between benign and malignant adrenal lesions, but its specificity is only 34.0% to 61.0% [18]. Existing clinical guidelines do not recommend relying solely on the 4.0 cm cutoff value to decide whether to perform surgery but suggest determining the appropriate treatment based on comprehensive factors such as the CT AVs and growth rate of the lesion. Therefore, the multivariate predictive model developed in this research aligns with current clinical needs. BMI is the only clinical variable in the model, and some studies have corroborated a positive correlation between BMI and the AV of adrenal adenomas, although the underlying pathophysiology remains unclear [19]. The variation in BMI in our dataset may be attributed to the fact that some patients with metastatic tumors in the non-adenoma group were in the late stages of malignancy. MeanAVs and minAVs both reflect the CT attenuation of the lesion, and there is an association between the two. Thus, this study employed XGBoost models capable of handling correlated variables to mitigate the risk of overfitting. Additionally, the significantly higher SHAP value of minAVs compared to meanAVs further demonstrated the former's superiority in diagnosing adrenal lipid-poor adenomas.

Limitations of this study include: ① The retrospective design inevitably introduces selection bias and statistical bias; ② The investigation only explored the diagnostic performance of the models on 3.75 mm and 5 mm slice thickness,and whether they apply to other thicknesses requires further investigation.Despite the effectiveness of the models, only CT scans with slice thicknesses of 3.75 mm and 5 mm were analyzed in this study. Similar to radiomics studies, resampling the voxel size as an image preprocessing step may enhance the reproducibility of attenuation values across different slice thicknesses. Future studies could investigate the impact of such preprocessing on the diagnostic performance of the models, potentially improving their generalizability and accuracy. ③ The study did not examine imaging signs such as cystic changes, necrosis, and calcification, which may provide additional information for differentiating adrenal lipid-poor adenomas from non-adenomas.

Conclusion

The XGBoost model built based on minAVs, lesion diameter, BMI, and meanAVs efficaciously distinguishes between adrenal lipid-poor adenomas and non-adenomas. Its diagnostic performance is comparable to that of the CTh model. Moreover, the simplicity of variable acquisition in the minAV model makes it more practical for clinical application, providing a basis for personalized diagnosis and treatment planning for patients and avoiding unnecessary examinations.

Acknowledgements

The authors would like to Thank you Department of Radiology, Affiliated Hangzhou First People's Hospital, Westlake University School of Medicine, Help and support from chief physician Han Zhijiang. The authors would also like to thank all those who participated in this study.

Authors' contributions

Study design/planning: Hanlin Zhu,Mengwei Wu and Zhijiang Han. Study conduct: all authors. Data analysis: Hanlin Zhu,Mengwei Wu and Bo Feng. Writing paper: Hanlin Zhu and Mengwei Wu Revising paper: all authors.

Funding

This work was supported by Quzhou Science and Technology Plan Project, NO.2023K121 and by Hangzhou Biomedical and Health Industry Development Support Science and Technology Project, NO.2022WJC175 and by Hangzhou Medical and Health Science and Technology Program, NO. A20220386, NO. A20220121, NO.20211231Y058.The funding bodies had no role in the design of the study, collection, analysis, and interpretation of data, or writing of the manuscript.

Availability of data and materials

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

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the First People's Hospital affiliated with Westlake University and Affiliated Quzhou Hospital of Wenzhou Medical University (Quzhou People's Hospital) (Approval numbers: IIT-20220811–0128-01, QZPYH-ER-2023–007), and patient informed consent was waived. Furthermore, it adhered to the ethical standards set forth by the World Medical Association's Declaration of Helsinki.

Consent for publication

Not Applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Radiology, Hangzhou Ninth People's Hospital (Hangzhou Red Cross Hospital Qiantang Campus), Hangzhou 310012, China. [2]Department of Radiology, The Quzhou Affiliated Hospital of Wenzhou Medical University, (Quzhou People's Hospital), Quzhou 324003, China. [3]Department of Radiology, Affiliated Hangzhou First People's Hospital, Westlake University School of Medicine, Hangzhou 310006, China.

Received: 21 February 2024 Accepted: 5 August 2024
Published online: 12 August 2024

References

1. Mayo-Smith WW, Song JH, Boland GL, et al. Management of incidental adrenal masses: a white paper of the acr incidental findings committee[J]. J Am Coll Radiol. 2017;14(8):1038–44.
2. Fassnacht M, Dekkers OM, Else T, et al. European society of endocrinology clinical practice guidelines on the management of adrenocortical carcinoma in adults, in collaboration with the european network for the study of adrenal tumors[J]. Eur J Endocrinol. 2018;179(4):G1-g46.
3. Sherlock M, Scarsbrook A, Abbas A, et al. Adrenal incidentaloma[J]. Endocr Rev. 2020;41(6):775–820.
4. Barzon L, Sonino N, Fallo F, et al. Prevalence and natural history of adrenal incidentalomas[J]. Eur J Endocrinol. 2003;149(4):273–85.
5. Schieda N, Siegelman ES. Update on ct and mri of adrenal nodules[J]. AJR Am J Roentgenol. 2017;208(6):1206–17.
6. Chung R, Garratt J, Remer EM, et al. Adrenal neoplasms: Lessons from adrenal multidisciplinary tumor boards[J]. Radiographics. 2023;43(7):e220191.
7. Bae KT, Fuangtharnthip P, Prasad SR, et al. Adrenal masses: Ct characterization with histogram analysis method[J]. Radiology. 2003;228(3):735–42.
8. Kartik SJ, Fenella W, Sangeet G, et al. Comparison of ct histogram analysis and chemical shift mri in the characterization of indeterminate adrenal nodules[J]. Am J Roentgenol. 2006;187:1303.
9. Zbyněk T, Petr K, Paulína S, et al. Influence of slice thickness on result of ct histogram analysis in indeterminate adrenal masses[J]. Abdom Radiol. 2019;44:1461.
10. Tongdee R, Tongdee T, Jin M G, et al. Comparison of ct histogram analysis and mean attenuation methods in characterization of adrenal masses: A phantom study[C]. Chicago: Radiological Society of North America 2004 Scientific Assembly and Annual Meeting; 2004. https://archive.rsna.org/2004/4412135.html.
11. Zhu M, Qu J, Han Z. Evaluate the efficacy of minimum attenuation value in differentiation of adrenal adenomas from nonadenomas on unenhanced ct[J]. Clin Imaging. 2016;40(1):86–9.
12. Han Z, Wu M, Wei P, et al. Differential diagnostic value of plain ct scan in adrenal adenoma and non-adenoma: A two-center control study of mean attenuation value, min imum attenuation value, and ct histogram[J]. Front Endocrinol (Lausanne). 2022;13:1007870.
13. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable ai for trees[J]. Nat Mach Intell. 2020;2(1):56–67.
14. Corwin MT, Navarro SM, Malik DG, et al. Differences in growth rate on ct of adrenal adenomas and malignant adrenal nodules[J]. AJR Am J Roentgenol. 2019;213(3):632–6.
15. Nagayama Y, Inoue T, Oda S, et al. Unenhanced dual-layer spectral-detector ct for characterizing indeterminate adrenal lesions[J]. Radiology. 2021;301(2):369–78.
16. Stanzione A, Verde F, Galatola R, et al. Qualitative heterogeneous signal drop on chemical shift (cs) mr imaging: Correlative quantitative analysis between cs signal intensity index and contrast washout parameters using t1-weighted sequences[J]. Tomography. 2021;7(4):961–71.
17. Zhu H, Wu M, Wei P, et al. A modified method for ct radiomics region-of-interest segmentation in adrenal lipid-poor adenomas: A two-institution comparative study[J]. Front Oncol. 2023;13:1086039.
18. Iñiguez-Ariza NM, Kohlenberg JD, Delivanis DA, et al. Clinical, biochemical, and radiological characteristics of a single-center retrospective cohort of 705 large adrenal tumors[J]. Mayo Clin Proc Innov Qual Outcomes. 2018;2(1):30–9.
19. Nishie A, Asayama Y, Ishigami K, et al. Impact of body mass index on ct attenuation of adrenal adenoma[J]. Eur J Radiol. 2018;108:184–8.

Publisher's Note