frontiers | Frontiers in Surgery

TYPE Original Research
PUBLISHED 20 November 2025
DOI 10.3389/fsurg.2025.1685442

OPEN ACCESS

EDITED BY
Bin Zhou,
The Affiliated Hospital of Qingdao University, China

REVIEWED BY
Ziying Lin,
Xiamen University, China
Zhe Xi,
Sun Yat-sen University, China

*CORRESPONDENCE
Haining Zheng
✉ zhenghaining010@163.com
Chaoyang Wen
✉ wencypkuih@163.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 13 August 2025
ACCEPTED 15 September 2025
PUBLISHED 20 November 2025

CITATION
Zheng H, Gao M, Cui J, Zhang X, Li W, Ma X and Wen C (2025) Preoperative differentiation of retroperitoneal ganglioneuroma and schwannoma using an ultrasonography-based multivariable model and simplified score: development and single-center internal validation.
Front. Surg. 12:1685442.
doi: 10.3389/fsurg.2025.1685442

Preoperative differentiation of retroperitoneal ganglioneuroma and schwannoma using an ultrasonography-based multivariable model and simplified score: development and single-center internal validation

Haining Zheng[1*†], Meiying Gao[1†], Jin Cui[1†], Xiaoying Zhang[2], Wenjie Li[3], Xuemei Ma[1] and Chaoyang Wen[1*]

[1]Department of Ultrasound, Peking University International Hospital, Beijing, China, [2]Department of Pathology, Peking University International Hospital, Beijing, China, [3]Department of Retroperitoneal Tumor Surgery, Peking University International Hospital, Beijing, China

Objective: This study aims to develop and internally validate a multivariable logistic regression model and a simplified scoring system, based on standardized ultrasonographic features, for the preoperative differentiation of retroperitoneal ganglioneuroma (GN) from schwannoma (SW), and to evaluate their discrimination, calibration, and clinical utility.

Methods: We retrospectively included patients with retroperitoneal GN or SW confirmed by surgical pathology. Standardized ultrasonographic features were extracted, and candidate predictors were selected using least absolute shrinkage and selection operator (LASSO) regression, while retaining potential confounders (age, sex, lesion long diameter). A multivariable model was constructed, and a six-variable simplified score was derived. Discrimination [area under the curve (AUC)], calibration (intercept, slope, Brier score), and decision curve analysis (DCA) were evaluated using stratified fivefold cross-validation and bootstrap resampling ($B = 2{,}000$). Two task-oriented thresholds were predefined: R1 [rule-out, sensitivity (Se) ≥ 0.95] and S1 [standard diagnosis, specificity (Sp) ≥ 0.50].

Results: A total of 74 patients were included (GN, 25, 33.8%; SW, 49, 66.2%). After optimism correction, the multivariable model achieved an AUC of 0.930, and the simplified score achieved an AUC of 0.917. Independent predictors included pelvic extraperitoneal location (*loc_pelvic* = 1), absence of cystic/necrotic change, and lower SD/LD ratio. For R1, the model threshold of 0.149 yielded Se = 0.960, Sp = 0.837, and negative predictive value (NPV) = 0.976; the score threshold of 0.206 yielded Se = 1.000, Sp = 0.592, and NPV = 1.000. For S1, the model threshold of 0.426 yielded Se = 0.920 and Sp = 0.939, and the score threshold of 0.594 yielded Se = 0.760 and Sp = 0.918.

Conclusion: Both the multivariable model and the simplified score demonstrated excellent performance in differentiating GN from SW, suggesting potential value as rapid, interpretable tools for bedside use and in resource-limited settings. Their clinical utility should be confirmed through external validation and recalibration in multicenter, prospective cohorts and further enhanced through integration with multimodal imaging such as CT, MRI, and contrast-enhanced ultrasound (CEUS).

KEYWORDS

retroperitoneal tumor, ganglioneuroma, schwannoma, ultrasonographic features, predictive model, simplified scoring system, decision curve analysis

Introduction

Ganglioneuroma (GN) and schwannoma (SW) are relatively common benign neurogenic tumors of the retroperitoneum. Although both entities are histologically benign, they differ markedly in biological behavior, preferred surgical approaches, and postoperative surveillance strategies. Consequently, accurate preoperative discrimination between GN and SW is essential for individualized surgical planning and optimizing long-term patient outcomes (1–3). Evidence from both multicenter and single-center cohort studies has described the imaging appearances, clinical management, and prognoses of GN (1, 4), emphasizing the real-world value of reliable preoperative identification. In contrast to malignant retroperitoneal sarcomas, these two benign tumors warrant distinctly different management pathways and are associated with substantially different prognostic profiles (5, 6). Although both are benign, GN typically follows an indolent clinical course after complete excision with lower surveillance intensity, whereas SW may exhibit higher local recurrence risk requiring closer postoperative surveillance and, in selected cases, wider *en bloc* resection margins. These distinctions motivate an accurate preoperative differentiation to individualize surgical planning and follow-up.

In contemporary clinical practice, contrast-enhanced computed tomography (CT) and magnetic resonance imaging (MRI) remain the cornerstone modalities for evaluating the origin, anatomical extent, and relationships of retroperitoneal tumors to adjacent structures (7–9). Recent advances in radiomics—exemplified by the multicenter RADSARC-R study published in *Lancet Oncology* in 2023 (10)—have demonstrated that CT-derived radiomics signatures can classify histologic subtypes and grades of retroperitoneal sarcomas with high accuracy. Nonetheless, such approaches depend on high-quality cross-sectional imaging and advanced post-processing platforms, which limit their feasibility in emergency, bedside, or resource-limited settings.

Ultrasonography, in contrast, offers real-time imaging, wide availability, low cost, and freedom from ionizing radiation, making it an indispensable tool for the initial assessment and follow-up of retroperitoneal masses (11, 12). Prior studies have shown that ultrasonography can depict key morphologic and vascular characteristics of neurogenic tumors—including tumor shape, margin definition, internal echotexture, cystic or necrotic change, posterior acoustic enhancement, and vascularity—which can aid in differentiating between pathologic subtypes and distinguishing benign from malignant lesions (12–15). For palpable or superficial soft tissue masses, ultrasonography is consistently recommended as the first-line imaging modality for triage, while MRI is reserved for complex or suspicious lesions to enable detailed characterization and staging (11, 12). Notably, the 2023 European Society of Musculoskeletal Radiology (ESSR) consensus introduced standardized interpretive criteria for adult soft tissue tumors, and the 2022 Society of Radiologists in Ultrasound (SRU) consensus by Jacobson et al. established uniform terminology and key interpretive points for superficial soft tissue mass evaluation. In the present study, we adopted these terminology frameworks to standardize the ultrasonographic assessment of retroperitoneal neurogenic tumors, thereby improving the reproducibility and consistency of feature interpretation.

On this basis, we developed and internally validated a multivariable logistic regression model grounded in a unified dictionary of ultrasonographic features to distinguish GN from SW preoperatively. From this model, we derived a simplified scoring system designed to be both interpretable and readily applicable at the bedside. The potential clinical utility of this approach warrants further evaluation in larger, multicenter cohorts.

Methods

Study design and ethical approval

This was a single-center retrospective observational study conducted in compliance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) principles. The protocol was approved by the institutional ethics committee, with informed consent waived due to the retrospective use of anonymized imaging and pathology data.

Study population

The patients who underwent surgical resection of a retroperitoneal tumor between December 2016 and November 2024 were identified from the institutional database. The inclusion criteria were as follows:

1. Histopathological diagnosis of GN or SW
2. Preoperative ultrasonographic images of sufficient quality for feature assessment

3. Complete clinical and pathological data

The exclusion criteria were as follows:

1. Concomitant malignant tumors
2. Missing key imaging planes or poor image quality
3. Imaging-to-surgery interval >3 months

Ultrasonographic examination

All examinations were performed by radiologists with at least 5 years of abdominal ultrasonography experience, using a Philips iU Elite color Doppler system (1–5 MHz transducer). Standard transverse, longitudinal, and any additional planes required for lesion characterization were obtained and stored in the picture archiving and communication system (PACS). B-mode and color Doppler modes were applied, with scanning parameters adjusted to the patient's habitus. Only cases with diagnostic-quality images, including transverse and longitudinal planes, were included; however, as a retrospective PACS review, we acknowledge that certain features (e.g., calcification) might not be captured if not present on the stored planes.

Features and data preprocessing

A standardized ultrasonographic feature dictionary was used to ensure consistency. The recorded variables included the following:

- Location: pelvic extraperitoneal (presacral/iliac fossa; yes/no). For clarity, the retroperitoneum inherently includes the pelvic portion; here, "pelvic extraperitoneal" was used as a shorthand for presacral/iliac fossa location.
- Shape: regular vs. irregular
- Margin: well-defined vs. ill-defined
- Internal echotexture: hypoechoic vs. non-hypoechoic
- Cystic/necrotic change: present vs. absent
- Calcification: present vs. absent
- Posterior acoustic enhancement: present vs. absent
- Tumor vessel encasement sign: present vs. absent
- Vascularity: present vs. absent on color Doppler
- Sex: male/female
- Age: years
- Tumor diameters: long diameter (LD), short diameter (SD), and SD/LD ratio

Two radiologists independently evaluated images, and disagreements were resolved by consensus. All features were coded as binary or continuous variables as appropriate. Data completeness was verified (no missing data). Both radiologists were blinded to histopathological results during feature assessment.

Model development

Feature selection was performed using least absolute shrinkage and selection operator (LASSO) logistic regression, retaining potential confounders (age, sex, LD). The selected variables were entered into a multivariable logistic regression to obtain adjusted odds ratios.

A simplified scoring system was derived from the final binary predictors, assigning one point per predictor (total score 0–6). A sensitivity analysis compared equal-weight and coefficient-weighted scoring for diagnostic performance and calibration.

Model evaluation and validation

Model performance was assessed in terms of the following:

- Discrimination: receiver operating characteristic (ROC) analysis with area under the curve (AUC) calculation
- Calibration: intercept, slope, and Brier score estimation; stratified calibration with 10 equal-frequency bins
- Uniform shrinkage: adjustment of coefficients using the optimism-corrected calibration slope (16)
- Clinical utility: decision curve analysis (DCA) to evaluate net benefit
- Internal validation: stratified fivefold out-of-fold (OOF) cross-validation for the simplified score
- Additional validation: nested cross-validation and bootstrap (.632+) resampling to assess model robustness

Threshold strategies

Two task-oriented thresholds were predefined:

- R1 (rule-out): sensitivity $\geq$0.95, maximizing specificity under this constraint
- S1 (standard diagnosis): specificity $\geq$0.50, maximizing the Youden index under this constraint

Statistical analysis

Unless otherwise specified, values were reported to three decimal places; $P < 0.001$ were reported as "<0.001." Se, Sp, positive predictive value (PPV), and negative predictive value (NPV) were calculated with 95% confidence intervals using the Wilson method. Normality of continuous variables was tested; appropriate parametric or non-parametric tests were applied. Categorical variables were analyzed using the χ^2 test or Fisher's exact test. All analyses were performed in Python (v3.9) with standard statistical and plotting libraries.

Results

Patient characteristics

A total of 74 patients were included: 25 (33.8%) with GN and 49 (66.2%) with SW. Baseline demographic and ultrasonographic characteristics are summarized in Table 1. A study flow

TABLE 1 Baseline characteristics of patients with GN vs. SW.

Variable	SW	GN	*P*-value
Sex	33 (67.3%) 16 (32.7%)	15 (60.0%) 10 (40.0%)	0.610
Age (years)	46.00 ± 15.56 [50.00 (33.00, 59.00)]	32.92 ± 15.01 [27.00 (23.00, 43.00)]	<0.001
Symptoms	5 (10.2%) 5 (10.2%) 21 (42.9%) 6 (12.2%) 12 (24.5%)	2 (8.0%) 0 (0.0%) 20 (80.0%) 1 (4.0%) 2 (8.0%)	0.035
Number of lesions	43 (87.8%) 6 (12.2%)	24 (96.0%) 1 (4.0%)	0.411
Location	25 (51.0%) 5 (10.2%) 19 (38.8%)	1 (4.0%) 1 (4.0%) 23 (92.0%)	<0.001
LD (cm)	7.72 ± 3.66 [7.60 (5.30, 9.60)]	12.21 ± 5.41 [11.40 (7.70, 15.30)]	<0.001
SD (cm)	5.37 ± 2.58 [4.90 (3.90, 6.40)]	5.34 ± 2.64 [4.80 (3.50, 6.40)]	0.749
SD/LD ratio	0.72 ± 0.16 [0.73 (0.58, 0.85)]	0.46 ± 0.17 [0.42 (0.38, 0.54)]	<0.001
Shape	10 (20.4%) 39 (79.6%)	18 (72.0%) 7 (28.0%)	<0.001
Margin	3 (6.1%) 46 (93.9%)	13 (52.0%) 12 (48.0%)	<0.001
Internal echo	48 (98.0%) 0 (0.0%) 1 (2.0%)	19 (76.0%) 6 (24.0%) 0 (0.0%)	0.001
Cystic/necrosis	24 (49.0%) 25 (51.0%)	23 (92.0%) 2 (8.0%)	<0.001
Calcification	43 (87.8%) 6 (12.2%)	19 (76.0%) 6 (24.0%)	0.317
Posterior echo	21 (42.9%) 27 (55.1%) 1 (2.0%)	4 (16.0%) 21 (84.0%) 0 (0.0%)	0.046
Blood flow	16 (32.7%) 8 (16.3%) 1 (2.0%) 24 (49.0%)	5 (20.0%) 3 (12.0%) 2 (8.0%) 15 (60.0%)	0.396
Vessel encasement	46 (93.9%) 3 (6.1%)	12 (48.0%) 13 (52.0%)	<0.001

GN, ganglioneuroma; SW, schwannoma; LD, long diameter; SD, short diameter; SD/LD, short-to-long diameter ratio. *P*-values were calculated using *t*-test or Mann–Whitney *U*-test for continuous variables, and chi-square test or Fisher's exact test for categorical variables as appropriate. Location was dichotomized as pelvic extraperitoneal (presacral/iliac fossa) vs. non-pelvic for modeling. Symptoms were categorized as follows: (1) asymptomatic/incidental finding; (2) abdominal pain/discomfort; (3) low back pain/discomfort; (4) lower-limb pain/discomfort; (5) palpable mass without discomfort.

description has been added, reporting numbers potentially eligible, excluded with reasons, and included (see Supplementary eFigure S2).

Univariate analysis

All odds ratios (ORs) were calculated with GN as the outcome event. Binary features significantly associated with a higher likelihood of GN included ill-defined margin (OR = 16.6, $P < 0.001$), tumor vessel encasement sign (OR = 16.6, $P < 0.001$), and irregular shape (OR = 10.0, $P < 0.001$). Features favoring SW (OR < 1 for GN) included pelvic extraperitoneal location (OR = 0.04, $P < 0.001$), cystic/necrotic change (OR = 0.083, $P < 0.001$), and posterior acoustic enhancement (OR = 0.254, $P = 0.006$).

Among continuous variables, the median LD was significantly larger in GN [Hodges–Lehmann (HL) difference $\approx +3.8$ cm, $P \approx 3.5 \times 10^{-4}$], whereas the SD/LD ratio was significantly lower in GN (HL difference ≈ -0.269, $P \approx 1.9 \times 10^{-7}$). Details are shown in Figure 1A (binary predictors) and Figure 1B (continuous predictors), as well as Supplementary eTables S1 and S2.

Multivariable analysis

LASSO regression selected three major predictors (location, cystic/necrotic change, SD/LD), which were entered into multivariable logistic regression together with potential confounders (LD, age, sex).

The final model included six variables:

- Pelvic extraperitoneal location (OR = 0.067, $P = 0.029$; favoring SW)
- Absence of cystic/necrotic change (OR = 0.023, $P = 0.008$; favoring GN)
- SD/LD per +1 unit (OR = 0.00067, $P = 0.017$; lower ratio favoring GN)
- LD (borderline effect, OR = 1.375, $P = 0.055$)
- Age and sex (non-significant)

Details are provided in Supplementary eTable S3A; scaled and shrunken effects are summarized in Supplementary eTables S3B and C, respectively.

Model performance

The multivariable model demonstrated strong discrimination: apparent AUC = 0.967 (95% CI, 0.926–0.995), optimism-corrected AUC = 0.930. The simplified score (fivefold cross-validation) yielded an AUC = 0.917 (95% CI, 0.846–0.968) (Figures 2A,B; Tables 2 and 3).

Calibration

For the multivariable model, the apparent slope was 1.000 with a Brier score of 0.068; the optimism-corrected slope was 0.442 with a Brier score of 0.101. The simplified score, using fivefold out-of-fold calibration, had an intercept of −0.003, a slope of 0.483, and a Brier score of 0.100 (Figure 3; Table 4). The optimism-corrected calibration slope of 0.442 indicates that the apparent predictions were over-extreme; substantial shrinkage and two-step recalibration (adjusting the intercept and then the slope) are recommended prior to external application.

Additional performance metrics:

- Fivefold out-of-fold calibration (reduced six-variable model): intercept = −0.003, slope = 0.483, Brier score = 0.100, AUC = 0.904.

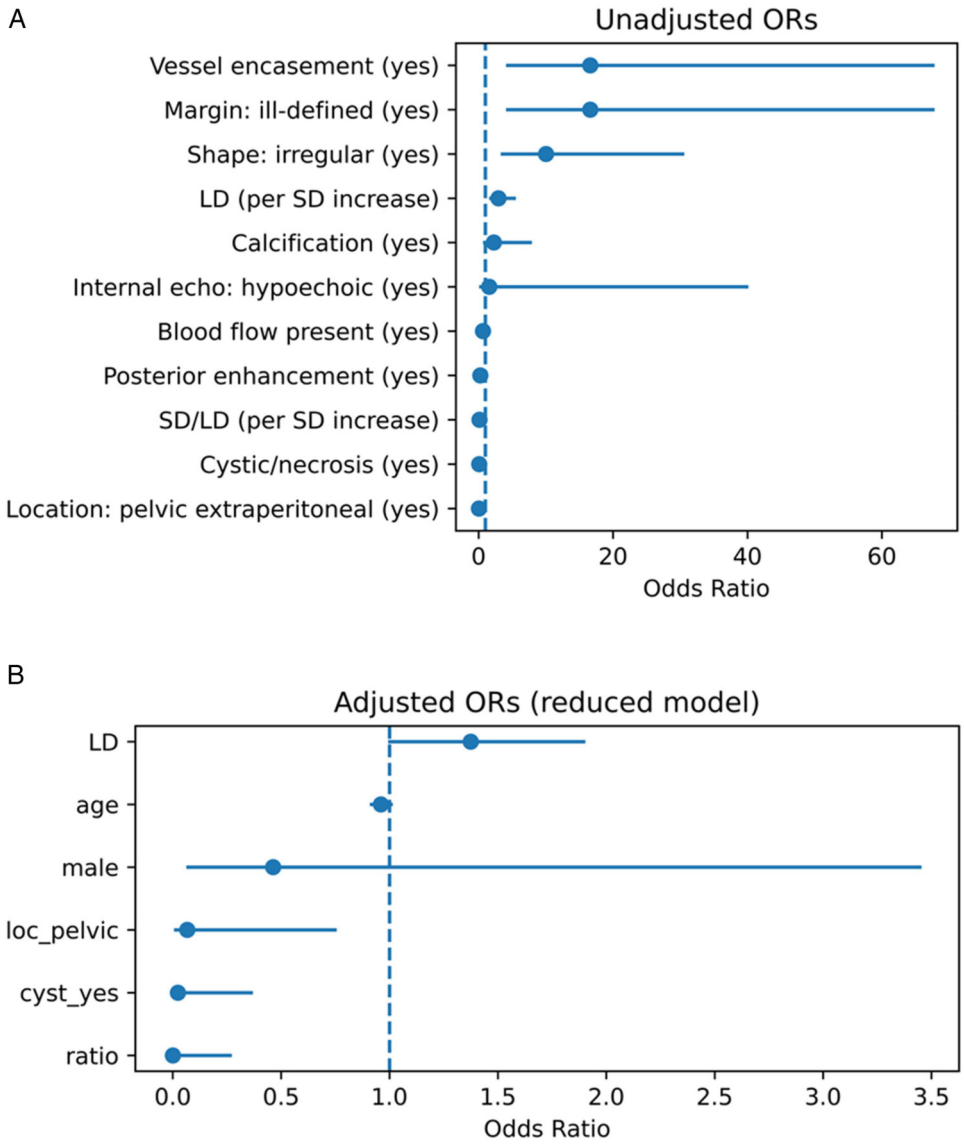

FIGURE 1
(A) Univariate associations between ultrasound features and the diagnosis of GN vs. SW. Forest plot showing unadjusted odds ratios (ORs) and 95% confidence intervals (CIs) from univariate analyses. Binary variables are shown as ORs; continuous variables are shown per unit increase or Hodges–Lehmann median difference. GN, ganglioneuroma; SW, schwannoma; LD, long diameter; SD, short diameter; SD/LD, short-to-long diameter ratio. (B) Adjusted multivariable associations between ultrasound features and the diagnosis of GN vs. SW. Forest plot showing adjusted odds ratios (ORs) and 95% confidence intervals (CIs) from the reduced multivariable logistic regression model. Variables included were selected by LASSO and adjusted for potential confounders (age, sex, long diameter). GN, ganglioneuroma; SW, schwannoma; LD, long diameter; SD, short diameter; SD/LD, short-to-long diameter ratio.

- L1-regularized logistic regression with nested cross-validation: outer AUC = 0.956 ± 0.030, Brier score = 0.088 ± 0.050.
- Bootstrap (.632+): AUC = 0.948, Brier score = 0.113.

Precision–recall analysis

Average precision (AP) was 0.937 for the multivariable model and 0.915 for the simplified score (Supplementary eFigure S1 and eTable S4). Notably, Supplementary eFigure S1 illustrates the PR curve for the multivariable model and shows the simplified score, highlighting both tools' robustness in imbalanced data contexts. Stratified calibration and extended calibration metrics are presented in Supplementary eTables S5A and B.

Decision curve analysis

Across the threshold probability range of 0.20–0.80, the multivariable model consistently provided higher net benefit than the treat-all and treat-none strategies (Figure 4),

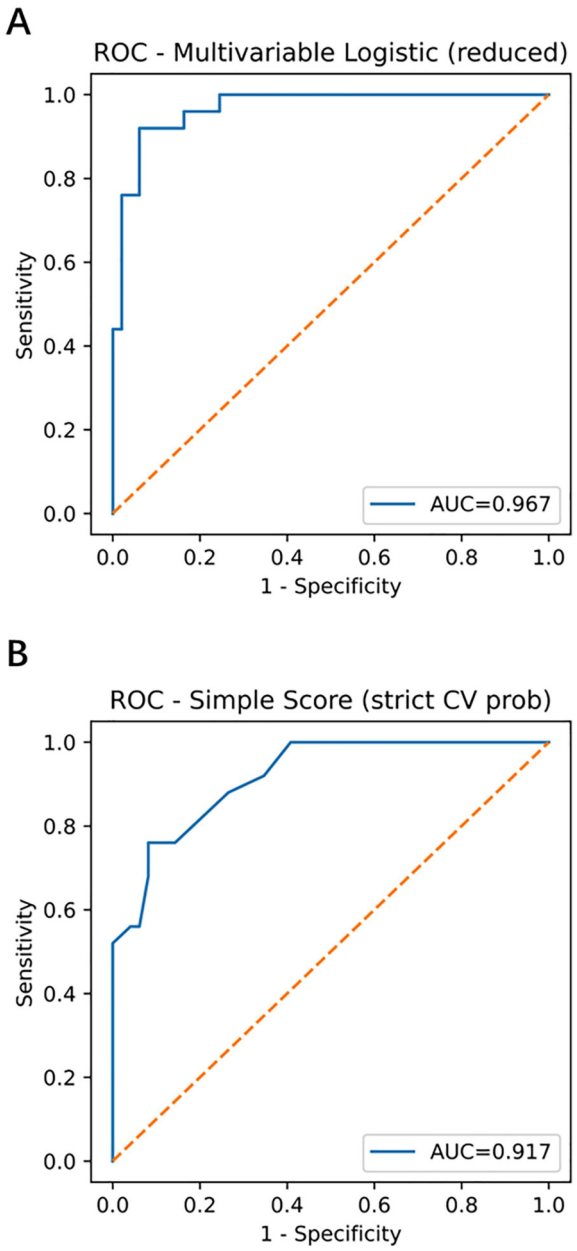

FIGURE 2
Receiver operating characteristic (ROC) curves of the multivariable model and simplified score. **(A)** Multivariable model with apparent AUC = 0.967 (95% CI, 0.926–0.995) and optimism-corrected AUC = 0.930. **(B)** Simplified score with stratified fivefold cross-validation AUC = 0.917 (95% CI, 0.846–0.968). ROC, receiver operating characteristic; AUC, area under the curve; CI, confidence interval; GN, ganglioneuroma; SW, schwannoma.

supporting its potential clinical utility in preoperative decision-making.

Threshold-based diagnostic performance

For clinical application, two predefined thresholds were evaluated:

- R1 (rule-out): high sensitivity (≥0.95) for excluding GN
- S1 (standard diagnosis): specificity ≥ 0.50 for confirming GN

TABLE 2 Discrimination of the multivariable model (AUC, 95% CI).

Model	AUC	95% CI (lower)	95% CI (upper)
Multivariable model	0.967	0.926	0.995

AUC, area under the curve; CI, confidence interval.

TABLE 3 Performance of the simplified score (stratified fivefold CV AUC, 95% CI).

Model	AUC	95% CI (lower)	95% CI (upper)
Simple score (strict fivefold CV prob)	0.917	0.846	0.968

AUC, area under the curve; CI, confidence interval; CV, cross-validation.

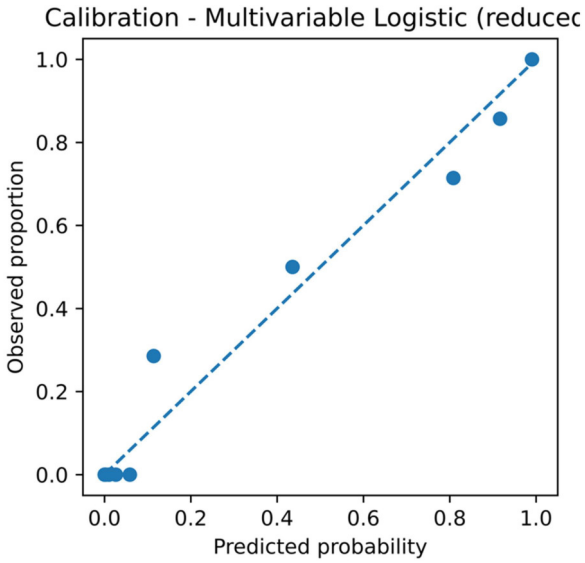

FIGURE 3
Calibration plot of the multivariable logistic regression model. The plot shows the relationship between predicted probabilities and observed proportions. Apparent slope = 1.000 and Brier score = 0.068; optimism-corrected slope = 0.442 and Brier score = 0.101, based on 2,000 bootstrap resamples. CI, confidence interval; GN, ganglioneuroma; SW, schwannoma.

Model R1/S1 thresholds: 0.149/0.426
Score R1/S1 thresholds: 0.206/0.594
Model performance (Table 5):

- R1: Se = 0.960, Sp = 0.837, PPV = 0.750, NPV = 0.976
- S1: Se = 0.920, Sp = 0.939, PPV = 0.885, NPV = 0.958

Score performance (Table 6):

- R1: Se = 1.000, Sp = 0.592, PPV = 0.556, NPV = 1.000
- S1: Se = 0.760, Sp = 0.918, PPV = 0.826, NPV = 0.882

Confusion matrices are provided in Supplementary eTable S6, illustrating the trade-off between sensitivity and specificity for clinical decision-making. These results align with the model's intended clinical roles described in the Discussion, where R1 supports high-sensitivity exclusion and S1 enables more

TABLE 4 Calibration metrics: intercept, slope, and Brier score.

Phase	Intercept	Slope	Brier score
Apparent	0.000	1.000	0.068
Optimism-corrected	0.040	0.442	0.101

CI, confidence interval. Calibration metrics were estimated using 2,000 bootstrap resamples.

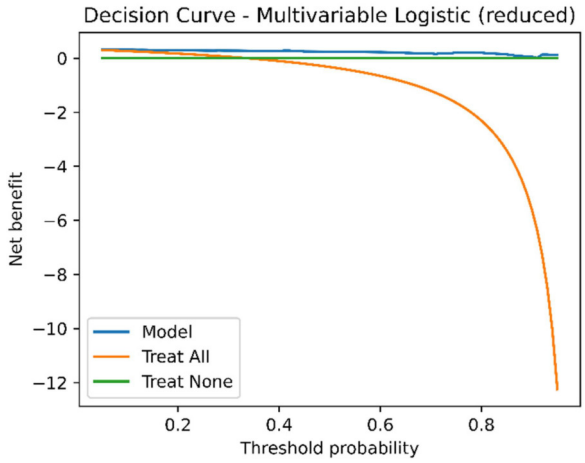

FIGURE 4
Decision curve analysis (DCA) of the multivariable model. Net benefit is plotted against threshold probability. The multivariable model shows a higher net benefit than both treat-all and treat-none strategies across threshold probabilities from 0.2 to 0.8. DCA, decision curve analysis; GN, ganglioneuroma; SW, schwannoma.

confident confirmation. Calculator parameters and an example are summarized in Supplementary eTables S11 and S12.

Extended score analysis

- Equal-weight vs. weighted scoring: AUC = 0.931 vs. 0.940; minimal difference, equal-weight retained for bedside use (Supplementary eTable S7).
- Score-to-probability mapping: For example, a score of 3 corresponds to an estimated probability of 0.206, 4 points ≈ 0.708, and 5 points ≈ 0.957. This mapping enables direct translation of a patient's score into an interpretable probability of GN, facilitating risk communication and clinical decision-making, and allowing individualized probability estimates (see Supplementary eTable S8), as further discussed in the section on clinical applicability.

Discussion

Key findings and underlying biological mechanisms

In this study, we developed and internally validated a multivariable logistic regression model and a simplified scoring system based on standardized ultrasonographic features to differentiate retroperitoneal GN from SW. Both tools demonstrated excellent discrimination, calibration, and clinical net benefit (apparent AUC = 0.967; optimism-corrected AUC = 0.930; simplified score cross-validated AUC = 0.917; Figure 2 and Supplementary eTable S9), with calibration slopes and Brier scores indicating robust model performance. Independent predictors in the final model included pelvic extraperitoneal location (*loc_pelvic* = 1; OR = 0.067, $P = 0.029$; favoring SW), absence of cystic/necrotic change (OR = 0.023, $P = 0.008$; favoring GN), and lower SD/LD ratio (OR = 0.00067 per +1 unit, $P = 0.017$; lower ratios favoring GN). LD showed a borderline effect ($P = 0.055$), whereas age and sex were not significant (Supplementary eTables S3A,C).

These findings have clear imaging–pathologic correlations:

1. Histologic/degenerative mechanisms

SW often contains Antoni B regions with marked degenerative changes (cystic degeneration, hemorrhage, myxoid change), producing hypoechoic cystic areas with possible posterior enhancement on B-mode ultrasonography; GN has a denser, more mature stroma, with a lower incidence of cystic change (8, 9, 17).

2. Geometric ratio characteristics

GN typically extends longitudinally along the sympathetic chain or nerve axis, producing a longer LD and narrower SD (lower SD/LD ratio); SW grows in an expansile manner, resulting in a

TABLE 5 Task-oriented thresholds (model): R1 (Se ≥ 0.95) and S1 (Sp ≥ 0.50) (for preoperative assessment, exploratory).

Strategy	Threshold	Sensitivity	Specificity	PPV	NPV
S1: Sp ≥ 0.50	0.426	0.920	0.939	0.885	0.958
R1: rule-out (Se ≥ 0.95)	0.149	0.960	0.837	0.750	0.976

PPV, positive predictive value; NPV, negative predictive value; Se, sensitivity; Sp, specificity.

TABLE 6 Task-oriented thresholds (simplified score): R1 (Se ≥ 0.95) and S1 (Sp ≥ 0.50) (for preoperative assessment, exploratory).

Strategy	Threshold	Sensitivity	Specificity	PPV	NPV
S1: Sp ≥ 0.50	0.594	0.760	0.918	0.826	0.882
R1: rule-out (Se ≥ 0.95)	0.206	1.000	0.592	0.556	1.000

PPV, positive predictive value; NPV, negative predictive value; Se, sensitivity; Sp, specificity.

more rounded or ovoid transverse section (2, 3, 18, 19). Moreover, pelvic retroperitoneal SWs likely originate from pelvic plexus nerves and tend to expand centrifugally within confined spaces, which is consistent with their more rounded morphology and with our finding that pelvic retroperitoneal location independently predicts SW.

Compared with most prior studies that relied on descriptive features or MRI-based models (AUC typically 0.85–0.90), our approach—integrating a unified ultrasonographic feature dictionary and the SD/LD ratio—achieved higher discrimination while using an accessible, non-ionizing modality. These structure–function correspondences likely underpin the stability of our model and strengthen its interpretability. In contrast to CT/MRI radiomics pipelines that require high-end imaging quality and dedicated post-processing, ultrasonography offers portability, timeliness, and bedside availability; our standardized ultrasonographic feature dictionary leverages these strengths while maintaining interpretability.

Clinical translation of threshold strategies

We defined two task-oriented thresholds (Tables 5 and 6, Supplementary eTable S10):

- R1 (rule-out, Se $\geq$ 0.95): model threshold = 0.149 (Se = 0.960, NPV = 0.976); score threshold = 0.206 (Se = 1.000, NPV = 1.000), enabling zero false negatives in-sample.
- S1 (standard diagnosis, Sp $\geq$ 0.50): model threshold = 0.426 (Se = 0.920, Sp = 0.939, PPV = 0.885, NPV = 0.958); score threshold = 0.594 (Se = 0.760, Sp = 0.918).

In rapid triage within ultrasound departments, R1 could identify low-probability cases suitable for follow-up or de-escalation, reducing unnecessary advanced imaging, whereas S1 could flag moderate-to-high-probability cases for further MRI characterization or preoperative planning. Decision curve analysis confirmed higher net benefit than treat-all or treat-none strategies across clinically relevant ranges (20, 21). These roles align directly with our results, where R1 achieved maximal sensitivity and S1 balanced specificity with predictive value. However, these findings remain exploratory and require prospective multicenter validation. These thresholds require prospective multicenter validation before widespread adoption.

Complementarity of B-mode and contrast-enhanced ultrasound

B-mode ultrasonography and contrast-enhanced ultrasound (CEUS) often yield concordant macroscopic findings (e.g., SW more frequently shows cystic change and heterogeneous enhancement), but CEUS uniquely assesses microvascular perfusion. CEUS–pathology correlation studies in retroperitoneal SW have demonstrated early-phase moderate-to-high enhancement patterns consistent with Antoni A/B vascular characteristics (15).

Following EFSUMB non-hepatic CEUS guidelines (22), integrating CEUS into our R1/S1 framework could refine risk stratification: CEUS might enhance specificity in equivocal or atypical cases, particularly in hypervascular SW subtypes, while R1/S1 thresholds provide structured decision-making for broader application. Future work should explore multimodal integration of CEUS parameters with standardized ultrasonographic features.

Molecular and embryologic perspectives

From a molecular and developmental standpoint:

- **SOX10**: A neural crest differentiation marker, diffusely positive in SW, reflecting stable Schwannian differentiation and correlating with its frequent degenerative changes (23).
- **GD2**: Highly expressed in immature neuroblastic tumors but low or absent in mature GN, suggesting potential roles in molecular imaging or targeted contrast agents for morphologically ambiguous retroperitoneal neurogenic tumors (24).
- **Embryologic pathway**: GN distribution follows neural crest migration patterns, often extending "*en bloc*" along the sympathetic chain, which explains its lower SD/LD ratio and tendency to abut but not invade major vessels (25).

Linking these molecular profiles to ultrasonographic features may pave the way for multi-omic predictive models that combine imaging phenotypes with tissue biomarkers. Integration of these molecular profiles with ultrasonographic phenotypes could further improve model interpretability and precision imaging.

5 Limitations and future directions

This single-center retrospective study had a limited sample size [$N = 74$; GN = 25; events per variable (EPV) ≈ 4], which may introduce selection and spectrum bias. Some predictors, such as LD and sex, had wide confidence intervals, highlighting the need for larger datasets to confirm effect stability. Despite LASSO selection, bootstrap optimism correction, and uniform shrinkage, the limited sample size ($N = 74$; GN = 25; EPV ≈ 4) implies residual risk of overfitting and statistical imprecision, warranting cautious interpretation and external validation. Generalisability may be limited by single-center case-mix, operator experience, and device/vendor differences; multicenter prospective validation across diverse settings is warranted.

To mitigate overfitting, we applied bootstrap optimism correction (AUC reduced from 0.967 to 0.930; Brier score increased from 0.068 to 0.101; calibration slope = 0.442) and uniform shrinkage (Figure 3, Supplementary eTables S9 and S3C). Before external application, we recommend two-step recalibration (adjust intercept a and then slope b) (16) and recalculation of PPV/NPV based on local prevalence. We provide executable Excel calculators (Supplementary Data Sheets

S1 and S2) and comma-separated values (CSV) recalibration examples (Supplementary Dataset S1) to support reproducibility.

Future studies should recruit multicenter, prospective cohorts including non-surgical cases, leverage semi-automated or AI-assisted ultrasonographic feature extraction, and evaluate the integration of CEUS and molecular data. These strategies may enhance model generalizability and promote clinical adoption.

This study developed and internally validated a multivariable logistic regression model (AUC = 0.930) and a simplified scoring system (AUC = 0.917) based on standardized ultrasonographic features to differentiate retroperitoneal GN from SW. Both tools demonstrated excellent discrimination, calibration, and clinical net benefit. The three core predictors—absence of cystic/necrotic change, lower SD/LD ratio, and pelvic extraperitoneal location—have clear imaging–pathologic underpinnings. These findings suggest that the proposed model and score could serve as rapid, interpretable tools for bedside and resource-limited settings. External validation in multicenter, prospective cohorts, ideally incorporating multimodal imaging features, is warranted to enhance diagnostic performance in complex or morphologically overlapping cases.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding authors.

Ethics statement

The study was approved by the Institutional Review Board of Peking University International Hospital (approval number: 2025-KY-0070-01). Given the retrospective design and the use of existing imaging and pathology data, the requirement for informed consent was waived by the ethics committee.

Author contributions

HZ: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. MG: Writing – original draft. JC: Data curation, Validation, Writing – review & editing. XZ: Data curation, Investigation, Writing – review & editing. WL: Validation, Writing – review & editing. XM: Data curation, Writing – original draft. CW: Methodology, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was financially supported by Peking University International Hospital Research Funds, China (Nos. YN2018QN01 and YN2025ZD01).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence, and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fsurg.2025.1685442/full#supplementary-material

References

1. Zhang QW, Song T, Yang PP, Hao Q. Retroperitoneum ganglioneuroma: imaging features and surgical outcomes of 35 cases at a Chinese institution. *BMC Med Imaging*. (2021) 21:114. doi: 10.1186/s12880-021-00643-y

2. Lonergan GJ, Schwab CM, Suarez ES, Carlson CL. Neuroblastoma, ganglioneuroblastoma, and ganglioneuroma: radiologic-pathologic correlation. *Radiographics*. (2002) 22(4):911–34. doi: 10.1148/radiographics.22.4.g02jl15911

3. Rha SE, Byun JY, Jung SE, Chun HJ, Lee HG, Lee JM. Neurogenic tumors in the abdomen: tumor types and imaging characteristics. *Radiographics*. (2003) 23(1):29–43. doi: 10.1148/rg.231025050

4. Xiao J, Zhao Z, Li B, Zhang T. Primary retroperitoneal ganglioneuroma: a retrospective cohort study of 32 patients. *Front Surg*. (2021) 8:642451. doi: 10.3389/fsurg.2021.642451

5. Behranwala KA, A'Hern R, Thomas JM. Surgical management of primary retroperitoneal sarcoma. *Br J Surg*. (2004) 91(5):547–57. doi: 10.1002/bjs.4533

6. Lewis JJ, Leung D, Woodruff JM, Brennan MF. Retroperitoneal soft-tissue sarcoma: analysis of 500 patients. *Ann Surg*. (1998) 228(3):355–65. doi: 10.1097/00000658-199809000-00008

7. Nwawka OK, Adriaensen M, Andreisek G, Drakonaki EE, Lee KS, Lutz AM, et al. Imaging of peripheral nerves: AJR expert panel narrative review. *AJR Am J Roentgenol*. (2025) 224:e2431064. doi: 10.2214/AJR.24.31064

8. Beaman FD, Kransdorf MJ, Menke DM. Schwannoma: radiologic-pathologic correlation. *Radiographics*. (2004) 24(5):1477–81. doi: 10.1148/rg.245045001

9. Pilavaki M, Chourmouzi D, Kiziridou A, Skordalaki A, Zarampoukas T, Drevelengas A. Imaging of peripheral nerve sheath tumors with pathologic correlation: pictorial review. *Eur J Radiol*. (2004) 52(3):229–39. doi: 10.1016/j.ejrad.2003.12.001

10. Arthur A, Orton MR, Emsley R, Vit S, Kelly-Morland C, Strauss D, et al. A CT-based radiomics classification model for the prediction of histological type and tumour grade in retroperitoneal sarcoma (RADSARC-R): a retrospective multicohort analysis. *Lancet Oncol*. (2023) 24(11):1277–86. doi: 10.1016/S1470-2045(23)00462-X

11. Noebauer-Huhmann IM, Vanhoenacker FM, Vilanova JC, Tagliafico AS, Weber MA, Lalam RK, et al. Soft tissue tumor imaging in adults: European Society of Musculoskeletal Radiology-Guidelines 2023—overview, and primary local imaging: how and where? *Eur Radiol*. (2024) 34(7):4427–37. doi: 10.1007/s00330-023-10425-5

12. Jacobson JA, Middleton WD, Allison SJ, Dahiya N, Lee KS, Levine BD, et al. Ultrasonography of superficial soft-tissue masses: society of radiologists in ultrasound consensus conference statement. *Radiology*. (2022) 304(1):18–30. doi: 10.1148/radiol.211101

13. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. (2015) 162(1):55–63. doi: 10.7326/M14-0697

14. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. TRIPOD: explanation and elaboration. *Ann Intern Med*. (2015) 162(1):W1–W73. doi: 10.7326/M14-0698

15. Safai Zadeh E, Görg C, Prosch H, Görg M, Trenker C, Westhoff CC, et al. The value of contrast-enhanced ultrasound in differentiating benign from malignant retroperitoneal masses. *Eur J Radiol*. (2024) 178:111596. doi: 10.1016/j.ejrad.2024.111596

16. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. 2nd ed. Cham: Springer (2019). doi: 10.1007/978-3-030-16399-0

17. Rodriguez FJ, Folpe AL, Giannini C, Perry A. Pathology of peripheral nerve sheath tumors: diagnostic overview and update on selected diagnostic problems. *Acta Neuropathol*. (2012) 123(3):295–319. doi: 10.1007/s00401-012-0954-z

18. Pilavaki M, Chourmouzi D, Kiziridou A, Skordalaki A, Zarampoukas T, Drevelengas A. Imaging of peripheral nerve sheath tumors with pathologic correlation: pictorial review. *Eur J Radiol*. (2004) 52(3):229–39. doi: 10.1016/j.ejrad.2003.12.001

19. Carone L, Messana G, Vanoli A, Pugliese L, Gallotti A, Preda L. Correlation between imaging and histology in benign solitary retroperitoneal nerve sheath tumors: a pictorial review. *Insights Imaging*. (2024) 15:132. doi: 10.1186/s13244-024-01709-5

20. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. (2006) 26(6):565–74. doi: 10.1177/0272989X06295361

21. Vickers AJ, van Calster B, Steyerberg EW. Decision curve analysis in the evaluation of radiology research. *Eur Radiol*. (2022) 32(9):5787–9. doi: 10.1007/s00330-022-08685-8

22. Piscaglia F, Nolsøe C, Dietrich CF, Cosgrove DO, Gilja OH, Bachmann Nielsen M, et al. The EFSUMB guidelines and recommendations on the clinical practice of contrast enhanced ultrasound (CEUS): update 2011 on non-hepatic applications. *Ultraschall Med*. (2012) 33(1):33–59. doi: 10.1055/s-0031-1281676

23. Sy AL, Hoang MP. SOX10. *J Clin Pathol*. (2023) 76(10):649–53. doi: 10.1136/jcp-2023-208924

24. Wei X, Li S, Wang Y. Expression of GD2 and GD3 in peripheral neuroblastic tumors. *Indian J Pathol Microbiol*. (2025) 68(1):17–22. doi: 10.4103/ijpm.ijpm_618_23

25. Le Douarin NM, Kalcheim C. *The Neural Crest*. 2nd ed. Cambridge: Cambridge University Press (1999). doi: 10.1017/CBO9780511897948