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.

Development and Validation of a Routine Clinical Indicator-Based Predictive Model for Malignancy Risk Stratification in Thyroid Nodules

xiuping qiu
Shengli Clinical Medical College of Fujian Medical University
gang chen
Shengli Clinical Medical College of Fujian Medical University
mei tu
tm13859596628@163.com

Longyan First Affiliated Hospital of Fujian Medical University

Research Article

Keywords: Thyroid nodules, Papillary thyroid carcinoma, Predictive model, Thyroid function, Risk stratification

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10965948/v1

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

Additional Declarations: No competing interests reported.

Development and Validation of a Routine Clinical Indicator-Based Predictive Model for Malignancy Risk Stratification in Thyroid Nodules

Xiuping Qiu[1], MM, Gang Chen[1*], MD, Mei Tu[2*], MB

1 Department of Endocrinology, Shengli Clinical Medical College of Fujian Medical University, Fuzhou 350001, China.

2 Department of Endocrinology, Longyan First Affiliated Hospital of Fujian Medical University, Longyan 364000, Fujian Province, China

Corresponding author

Main:

Mei Tu

Department of Endocrinology,

Longyan First Affiliated Hospital of Fujian Medical University,

No. 105 of Jiuyi North Road, Xinluo District,

Longyan 364000,

Fujian Province, China.

Tel: 86-13859596628

Email: tm13859596628@163.com

Secondary:

Gang Chen

Department of Endocrinology,

Shengli Clinical Medical College of Fujian Medical University,

No. 134 of Dongjie, Gulou District,

Fuzhou 350001,

Fujian Province, China.

Tel: 86-13509337027

Email: chengangfj@163.com

Abstract

Objective: To develop and validate a predictive model based on routine clinical indicators for assessing the risk of malignancy in thyroid nodules, thereby providing a reference for the preoperative differentiation of papillary thyroid carcinoma.

Methods: A total of 1,754 patients with thyroid nodules who underwent thyroidectomy at Longyan First Hospital from January 2021 to December 2023 were retrospectively enrolled, including 754 patients with benign nodules and 1,000 patients with papillary thyroid carcinoma（PTC）. Data on thyroid function, thyroid autoantibodies, inflammatory markers, biochemical parameters, and nodule volume were collected. Variables were selected using LASSO regression and the Boruta algorithm, followed by

bidirectional stepwise selection based on the Akaike Information Criterion to construct the final multivariable logistic regression model.

Results: The final model included six predictors: nodule volume, FT3, FT4, TSH, serum total protein, and total bilirubin. The model demonstrated good discrimination, with AUC values of approximately 0.87 in the training cohort and 0.86 in the validation cohort, as well as favorable calibration and clinical net benefit.

Conclusion: This study developed a prediction model for differentiating papillary thyroid carcinoma from benign thyroid nodules based on nodule volume, FT3, FT4, TSH, total bilirubin, and serum total protein.

Trial registration: Clinical trial number: not applicable

Keywords:
Thyroid nodules; Papillary thyroid carcinoma; Predictive model; Thyroid function; Risk stratification

Introduction

Thyroid nodules are highly prevalent in the general population, with ultrasound-detected lesions identified in approximately 20–76% of adults [1,2]. Although the majority of these nodules are benign, approximately 5–15% harbor malignancy, predominantly papillary thyroid carcinoma [3]. Accurate differentiation between benign and malignant lesions is critical to avoid unnecessary invasive procedures while ensuring timely intervention for cancerous cases [4]. Current diagnostic standards rely heavily on ultrasonography and fine-needle aspiration biopsy (FNAB), yet FNAB

yields non-diagnostic or indeterminate results in 10–30% of cases, leading to diagnostic uncertainty and repeat procedures [4]. Emerging evidence suggests that serum biomarkers may complement imaging; for instance, elevated thyroid-stimulating hormone (TSH) levels are independently associated with an increased risk of thyroid malignancy [5]. Furthermore, systemic inflammatory markers such as the neutrophil-to-lymphocyte ratio (NLR), systemic immune-inflammation index (SII), and platelet-to-hemoglobin ratio (PHR) have been linked to tumor aggressiveness and poor prognostic features [6-8]. Similarly, thyroid peroxidase antibody(TPOAb), thyroglobulin antibody(TGAb), thyroid-stimulating hormone receptor antibody(TRAb), and dyslipidemia profiles Total cholesterol (TC), Triglycerides(TG), Low-density lipoprotein(LDL),High-Density Lipoprotein(HDL) reflect underlying autoimmune and metabolic states that may modulate cancer risk, although their individual predictive values remain controversial or context-dependent [9] [10–13]. Additionally, thyroid nodule volume has been correlated with malignancy risk and advanced disease stages [14,15]. Despite these associations, the collective predictive efficacy of integrating clinical, biochemical, hematological, and volumetric parameters remains inadequately characterized in a multivariable context. This study aims to construct a comprehensive predictive model by analyzing thyroid function indices (FT3, FT4, TSH), autoantibodies (TPOAb, TgAb, TRAb), lipid profiles (TC, TG, LDL-C, HDL, Serum Total Protein (STP),Total Bilirubin (TB)), systemic inflammatory markers (White Blood Cell count(WBC), Neutrophil count(NEU), Lymphocyte count(LYM), NLR, SII, PHR), and thyroid nodule volume (TNV), categorized into tertiles according

to measured nodule volume. We hypothesize that this integrated approach will demonstrate superior accuracy in stratifying malignancy risk compared to models relying on isolated parameters, thereby refining clinical decision-making.

Research Methods

This retrospective case-control study included 1,754 patients with thyroid nodules who underwent thyroidectomy at Longyan City First Hospital between January 2021 and December 2023. This cohort comprised 754 individuals with benign thyroid nodules and 1,000 cases of papillary thyroid carcinoma (PTC). Baseline data—including FT3,FT4,TSH, NEU ,MON, TPOAb,TGAb,TRAb,TNV, and other relevant clinical variables—were extracted from the initial fasting blood samples obtained upon patient admission. All inclusion and exclusion criteria were established based on previously published relevant literature. Inclusion criteria mandated: (1) surgical and pathological confirmation of either benign thyroid nodules or PTC; and (2) availability of complete clinical and pathological records for analysis. Exclusion criteria encompassed: (1) presence of acute inflammation or infection; (2) diagnosis of autoimmune disorders; (3) history of other malignant neoplasms; (4) prolonged use of immunomodulatory agents; (5) existence of endocrine abnormalities or thyroid dysfunction; (6) concurrent hepatobiliary diseases or abnormal dietary statuses, such as chronic fasting; and (7) incomplete clinical documentation. The diagnosis of PTC strictly adhered to the 2017 World Health Organization (WHO) classification standards for thyroid tumors, with all pathological findings independently verified by two senior pathologists [16] . All clinical data pertaining to enrolled participants were retrieved from

the hospital's electronic medical record system via a standardized extraction protocol.To ensure the fidelity and precision of the research dataset, a standardized protocol for data extraction was rigorously implemented.

Ethical authorization for this study was granted by the Ethics Review Board of Longyan City First Hospital (Approval No. LYREC2024-K043-01), with all procedures adhering strictly to the ethical principles outlined in the Declaration of Helsinki. Due to the retrospective observational nature of the study and the complete anonymization of patient medical records, the requirement for written informed consent was waived following approval by the institutional ethics committee and applicable regulatory guidelines. Strict confidentiality was maintained for all clinical data pertaining to enrolled participants, which were utilized exclusively for academic investigation. Throughout the research process, robust measures were enacted to preserve participant privacy; specifically, any identifiers capable of revealing individual subjects were excised to ensure data security and protect patient confidentiality.

Statistical Methods

Data Preprocessing and Missing Value Management

The normality of continuous variables was evaluated using the Shapiro-Wilk test, supplemented by an analysis of skewness and kurtosis coefficients, to guide the selection of appropriate descriptive statistics and modeling strategies. Categorical variables were summarized as frequencies and percentages. Regarding missing data, variables and subjects exhibiting a

missingness rate of 50% or higher were excluded prior to analysis. The dataset was first randomly split into training and validation cohorts. Multiple imputation was then performed within the training cohort, and the imputation model was applied to the validation cohort to avoid information leakage.Remaining missing values were addressed via the Multiple Imputation by Chained Equations (MICE) algorithm: predictive mean matching (PMM) was applied to continuous variables, logistic regression (logreg) to binary variables, and multinomial logistic regression (polyreg) to multicategorical variables. Model estimates were pooled across the five imputed datasets using Rubin's rules. To mitigate multicollinearity, identification IDs and composite derived indices (NLR, SII, PHR) were omitted from the imputation predictor matrix.

Descriptive Statistics and Inter-group Comparisons

Continuous variables following a normal distribution were expressed as mean ± standard deviation (Mean ± SD). Comparisons between two groups were conducted using independent samples t-tests, whereas one-way analysis of variance (ANOVA) was employed for comparisons involving three or more groups. Non-normally distributed continuous variables were reported as median and interquartile range [Median (IQR)]; the Mann-Whitney U test was utilized for two-group comparisons, while the Kruskal-Wallis H test was applied for multiple-group scenarios. Differences in categorical variables across groups were assessed using the Pearson χ^2 test; when theoretical frequencies fell below 5, Fisher's exact test with Monte Carlo simulation was substituted. All statistical tests were two-tailed, with statistical significance defined as $\alpha = 0.05$. For the construction of predictive models, the total cohort was randomly partitioned into a training set (70%)

and an internal validation set (30%), stratified by the binary outcome (thyroid malignancy). Baseline characteristics were compared between these subsets to verify balanced allocation.

Feature Engineering

Continuous variables exhibiting an absolute skewness coefficient exceeding 0.8 underwent Yeo-Johnson power transformation to enhance distributional symmetry. Categorical predictors were encoded into dummy variables using k-1 coding, with the initial category serving as the reference group. A dual framework for feature selection was established to identify robust predictive factors from the pool of candidate variables: (1) Least Absolute Shrinkage and Selection Operator (LASSO) regression with a binomial family and 10-fold cross-validation. Predictors retaining non-zero coefficients, filtered by the regularization parameter $\lambda.1se$ (within one standard error of the minimum criterion), were preserved; (2) The Boruta algorithm, grounded in random forest variable importance assessment, with a significance threshold of $p = 0.01$. To reduce model complexity and improve clinical applicability, the intersecting variables identified by LASSO and Boruta were further entered into a multivariable logistic regression model, followed by bidirectional stepwise selection based on the minimum AIC.To derive a robust set of confirmed "important" features, the procedure was executed over 99 iterations. Variables concurrently identified by both methodological approaches were intersected and subsequently incorporated into a multivariable logistic regression framework. Thyroid malignancy status served as the binary dependent variable, coded as 1 for malignant cases. To refine the predictor set and construct the final predictive model, bidirectional

stepwise selection guided by the Akaike Information Criterion (AIC) was employed. Model outputs are presented as odds ratios (ORs) accompanied by their respective 95% confidence intervals (CIs). Multicollinearity was assessed via variance inflation factors (VIFs), with

values below 5 deemed acceptable.

Model Performance Evaluation and Internal Validation

Build the model in the training set and perform internal validation in a held-out validation cohort. Model performance will be evaluated in terms of discrimination, calibration, and clinical utility.

Discrimination： Receiver operating characteristic (ROC) curves will be plotted, and the area under the curve (AUC) with its 95% confidence interval will be calculated using the bootstrap method with 1,000 resamples. The optimal cut-off value will be determined according to the Youden index. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score, and Brier score will also be reported.

Calibration： Model goodness-of-fit will be assessed using the Hosmer–Lemeshow test with five groups ($g = 5$). Calibration curves will be plotted. Calibration-in-the-large and calibration slope will be estimated using the bootstrap method with 1,000 resamples, with ideal values of 0 and 1, respectively.

Clinical utility： Decision curve analysis (DCA) will be performed to calculate the net benefit across threshold probabilities ranging from 0% to 100%. Given the case–control sampling design, the population prevalence

was prespecified in the DCA for both the training and internal validation cohorts to correct for oversampling bias, and bootstrap resampling with 500 iterations was used to improve curve stability, and bootstrap resampling with 500 iterations will be used to improve the stability of the curves.

Software and reporting : All statistical analyses will be conducted using R software, version 4.5, and relevant R packages,including mice, glmnet, boruta, pROC, rms, and rmda

Results

1. Baseline Characteristics

Table 1 delineates the comparison of clinical and biochemical baseline characteristics across the entire cohort, the training set (n = 1,228), and the internal validation set (n = 526). Univariate analyses demonstrated that baseline characteristics were well balanced between the training and internal validation cohorts. No statistically significant intergroup differences were detected for any evaluated variables (all $P > 0.05$), encompassing the primary outcome (with both subsets having an identical proportion of malignant cases of 57.0%, reflecting the stratified split rather than the population prevalence), age, sex distribution, thyroid function markers (TSH, FT3, FT4), thyroid autoantibodies, systemic peripheral inflammatory indices (WBC, NLR, etc.), and the three-tiered nodule volume stratification (TNV_Group). These results confirm that stratified random splitting was effectively implemented with negligible allocation bias, thereby supporting the comparability of the training and validation cohorts for subsequent model development and internal validation.

Table 1. Baseline characteristics of the training and validation sets

Characteristics[1]	Overall[2]	Training Set N = 1,228[2]	Validation Set N = 526[2]	Test statistic[3]	P value[4]
TEAM				0.000	>0.999
Benign	754(43.0%)	528(43.0%)	226(43.0%)		
Malignant	1,000(57.0%)	700(57.0%)	300(57.0%)		
Gender				1.40	0.236
1	1,438(82.0%)	1,016(82.7%)	422(80.2%)		
2	316(18.0%)	212(17.3%)	104(19.8%)		
Age				0.181	0.671
Median (Q1, Q3)	47.0 (37.0, 53.0)	46.0 (37.0, 53.0)	47.0 (37.0, 53.0)		
TCmmol/L				0.753	0.385
Median (Q1, Q3)	5.1 (4.4, 5.8)	5.1 (4.4, 5.7)	5.1 (4.4, 5.8)		
TGmmol/L				0.970	0.325
Median (Q1, Q3)	1.2 (0.9, 1.8)	1.2 (0.9, 1.8)	1.2 (0.8, 1.8)		
TyG				0.439	0.508
Median (Q1, Q3)	8.6 (8.2, 9.0)	8.6 (8.2, 9.0)	8.6 (8.2, 9.0)		
LDLmmol/L				2.21	0.137
Median (Q1, Q3)	3.2 (2.6, 3.7)	3.2 (2.6, 3.7)	3.2 (2.7, 3.8)		
GLUmmol/L				0.492	0.483
Median (Q1, Q3)	5.4 (5.0, 6.0)	5.4 (5.0, 6.0)	5.4 (5.0, 5.9)		
FT3pmol/L				2.16	0.142
Median (Q1, Q3)	5.1 (4.6, 5.6)	5.1 (4.6, 5.6)	5.1 (4.7, 5.6)		
FT4 pmol/L				0.159	0.690
Median (Q1, Q3)	12.7 (10.8, 16.0)	12.7 (10.9, 15.9)	12.7 (10.8, 16.1)		
TSHmIU/L				1.30	0.255
Median (Q1, Q3)	1.6 (1.0, 2.4)	1.6 (1.0, 2.5)	1.6 (1.0, 2.4)		
TPOAB IU/mL				0.547	0.460
Median (Q1, Q3)	7.1 (0.8, 22.4)	6.9 (0.8, 22.6)	7.1 (0.8, 21.5)		
TGAB IU/mL				0.860	0.354
Median (Q1, Q3)	4.3 (0.9, 15.4)	6.3 (0.9, 15.6)	2.2 (0.9, 15.2)		
TRAB IU/L				0.617	0.432
Median (Q1, Q3)	0.8 (0.5, 1.0)	0.8 (0.4, 1.0)	0.8 (0.5, 1.0)		
NEU×10⁹/L				0.010	0.921
Median (Q1, Q3)	4.4 (3.2, 6.8)	4.4 (3.3, 6.7)	4.3 (3.2, 7.2)		
WBC×10⁹/L				0.006	0.940
Median (Q1, Q3)	6.5 (5.3, 8.9)	6.5 (5.3, 8.8)	6.5 (5.2, 9.1)		
MON×10⁹/L				0.001	0.969
Median (Q1, Q3)	0.3 (0.2, 0.4)	0.3 (0.2, 0.4)	0.3 (0.2, 0.4)		
PLT×10⁹/L				0.001	0.980
Median (Q1, Q3)	257.5 (222.0, 300.0)	258.0 (222.5, 298.0)	255.0 (221.0, 304.0)		
STP g/L				0.063	0.802
Median (Q1, Q3)	72.2 (68.2, 75.5)	72.2 (68.4, 75.4)	71.9 (67.7, 75.8)		

Characteristics[1]	Overall[2]	Training Set N = 1,228[2]	Validation Set N = 526[2]	Test statistic[3]	P value[4]
ALB g/L				0.668	0.414
Mean ± SD	42.7 ± 3.6	42.7 ± 3.6	42.8 ± 3.7		
HDL mmol/L				0.012	0.914
Median (Q1, Q3)	1.3 (1.1, 1.6)	1.4 (1.1, 1.6)	1.3 (1.1, 1.6)		
TB umol/L				0.604	0.437
Median (Q1, Q3)	12.5 (9.8, 16.5)	12.4 (9.8, 16.2)	12.6 (9.8, 16.9)		
Hb g/L				1.78	0.182
Median (Q1, Q3)	133.0 (123.0, 142.0)	133.0 (123.0, 142.0)	135.0 (123.0, 144.0)		
LYM ×10⁹/L				0.182	0.670
Median (Q1, Q3)	1.4 (0.9, 2.0)	1.4 (0.9, 2.0)	1.4 (0.9, 2.0)		
NLR				0.522	0.470
Median (Q1, Q3)	3.1 (2.0, 6.4)	3.2 (2.0, 6.2)	2.9 (1.9, 6.7)		
SII				0.844	0.358
Median (Q1, Q3)	827.8 (495.1, 1,720.2)	838.4 (505.1, 1,676.5)	798.3 (463.4, 1,845.2)		
PHR				0.229	0.633
Median (Q1, Q3)	1.9 (1.6, 2.3)	1.9 (1.6, 2.3)	1.9 (1.6, 2.3)		
TNV_Group cm³				1.73	0.420
Q1 (Low)	593(33.8%)	405(33.0%)	188(35.7%)		
Q2 (Medium)	586(33.4%)	410(33.4%)	176(33.5%)		
Q3 (High)	575(32.8%)	413(33.6%)	162(30.8%)		

[1]Abbreviations: IQR, interquartile range; SD, standard deviation.

[2]Continuous variables presented as median (IQR) or mean (SD); categorical variables as n (%)

[3]Continuous variables were compared using Student's t-test or Mann-Whitney U test, as appropriate. Categorical variables were compared using Pearson's chi-square test or Fisher's exact test.

[4]Statistical significance set at P < 0.05 (two-tailed)

TC Total cholesterol;TG Triglycerides;**TyG** Triglyceride-glucose index;LDL Low-density lipoprotein;GLU glucose;FT3 Free Triiodothyronine;FT4 Free Thyroxine; TSH Thyroid-Stimulating Hormone ;TPOAb Thyroid Peroxidase Antibody ;TGAb: Thyroglobulin Antibody;TRAb Thyrotropin Receptor Antibody ;NEU Neutrophil count ; WBC: White Blood Cell count MON Monocyte count ;PLT: Platelet count ;STP: Serum Total Protein ;ALB: Albumin ;HDL High-Density Lipoprotein ; TB Total Bilirubin ;Hb: Hemoglobin ; LYM Lymphocyte count;NLR Neutrophil-to-Lymphocyte Ratio;SII: Systemic Immune-Inflammation Index;PHR: Platelet-to-Hemoglobin Ratio; TNV Thyroid nodule volume

Table 2 delineates the baseline demographic and clinical profiles of the 1,754 participants, categorized according to histopathological diagnosis into a benign nodule cohort (n = 754) and a malignant papillary thyroid

carcinoma (PTC) group (n = 1,000). Univariate statistical comparisons revealed significant heterogeneity in clinical characteristics between these two distinct populations. Specifically, patients diagnosed with PTC were notably younger than their counterparts with benign lesions, exhibiting a median age of 45.00 years compared to 49.00 years ($P < 0.001$). Furthermore, a pronounced disparity was observed regarding nodule volume distribution: the malignant cohort demonstrated a substantially higher prevalence of small nodules, with 45.6% falling within the lowest tertile (TNV Q1), versus only 18.2% in the benign group ($P < 0.001$). Conversely, large nodules (TNV Q3) were predominantly associated with benign pathology, comprising 62.5% of the benign cohort compared to merely 10.4% of the malignant cases. In terms of biochemical markers, the PTC group presented with elevated concentrations of TSH, FT4, and thyroid-specific autoantibodies, including TPOAb and TGAb, alongside altered peripheral.

Several systemic inflammatory markers, including WBC, NEU, and the NLR, exhibited highly significant disparities across groups (all $P < 0.001$). Conversely, no statistically meaningful variations were observed between cohorts concerning sex distribution ($P = 0.241$), TG($P = 0.418$), the triglyceride-glucose (TyG) index ($P = 0.115$), or HDL ($P = 0.394$).

Table 2. Baseline characteristics by Malignant and benign thyroid nodules

Characteristics[1]	Overall[2]	Benign N = 754[2]	Malignant N = 1,000[2]	Test statistic[3]	P value[4]
Gender				1.37	0.241
1	1,438(82.0%)	628(83.3%)	810(81.0%)		
2	316(18.0%)	126(16.7%)	190(19.0%)		
Age				39.2	<0.001
Median (Q1, Q3)	47.00 (37.00, 53.00)	49.00 (39.00, 56.00)	45.00 (36.00, 51.00)		
TC mmol/L				16.4	<0.001
Median (Q1, Q3)	5.08 (4.41, 5.75)	5.20 (4.55, 5.86)	4.96 (4.36, 5.69)		
TGmmol/L				0.657	0.418
Median (Q1, Q3)	1.22 (0.86, 1.79)	1.20 (0.87, 1.81)	1.23 (0.85, 1.78)		
TyG				2.48	0.115
Median (Q1, Q3)	8.57 (8.19, 9.03)	8.60 (8.20, 9.11)	8.55 (8.18, 8.99)		
LDLmmol/L				3.57	0.059
Median (Q1, Q3)	3.17 (2.63, 3.73)	3.25 (2.72, 3.73)	3.12 (2.59, 3.73)		
GLUmmol/L				9.41	0.002
Median (Q1, Q3)	5.40 (5.01, 5.95)	5.49 (5.04, 6.05)	5.35 (4.99, 5.89)		
FT3 pmol/L				90.6	<0.001
Median (Q1, Q3)	5.07 (4.64, 5.57)	5.25 (4.85, 5.72)	4.92 (4.51, 5.39)		
FT4 pmol/L				194	<0.001
Median (Q1, Q3)	12.73 (10.84, 15.97)	11.59 (10.41, 13.29)	14.63 (11.34, 17.20)		
TSHmIU/L				71.5	<0.001
Median (Q1, Q3)	1.57 (0.99, 2.42)	1.31 (0.80, 2.12)	1.77 (1.14, 2.63)		
TPOAB IU/mL				151	<0.001
Median (Q1, Q3)	7.07 (0.80, 22.36)	1.40 (0.50, 10.81)	13.55 (1.80, 28.60)		
TGAB IU/mL				287	<0.001
Median (Q1, Q3)	4.30 (0.90, 15.39)	0.90 (0.90, 4.60)	10.88 (0.90, 21.31)		
TRAB IU/L				72.1	<0.001
Median (Q1, Q3)	0.80 (0.46, 1.02)	0.80 (0.80, 1.11)	0.80 (0.35, 0.96)		
NEU×10⁹/L				127	<0.001
Median (Q1, Q3)	4.40 (3.23, 6.78)	3.95 (3.07, 5.12)	5.18 (3.40, 9.74)		
WBC×10⁹/L				218	<0.001
Median (Q1, Q3)	6.52 (5.29, 8.90)	5.97 (4.81, 7.08)	7.55 (5.72, 11.11)		
MON×10⁹/L				42.3	<0.001
Median (Q1, Q3)	0.32 (0.24, 0.42)	0.34 (0.28, 0.43)	0.31 (0.22, 0.41)		
PLT×10⁹/L				3.58	0.059
Median (Q1, Q3)	257.50 (222.00, 300.00)	253.00 (218.00, 298.00)	261.00 (224.00, 301.50)		
STP g/L				36.7	<0.001

Characteristics[1]	Overall[2]	Benign N = 754[2]	Malignant N = 1,000[2]	Test statistic[3]	P value[4]
Median (Q1, Q3)	72.20 (68.20, 75.50)	73.15 (69.30, 76.20)	71.30 (67.50, 74.90)		
ALB g/L				37.2	<0.001
Mean ± SD	42.71 ± 3.64	43.31 ± 3.52	42.25 ± 3.66		
HDL mmol/L				0.725	0.394
Median (Q1, Q3)	1.34 (1.13, 1.59)	1.35 (1.14, 1.58)	1.33 (1.13, 1.59)		
TB umol/L				103	<0.001
Median (Q1, Q3)	12.50 (9.80, 16.50)	11.10 (8.70, 14.60)	13.60 (10.80, 17.70)		
Hb g/L				10.3	0.001
Median (Q1, Q3)	133.00 (123.00, 142.00)	132.00 (121.00, 141.00)	134.00 (124.00, 143.00)		
LYM 10^{-9}/L				416	<0.001
Median (Q1, Q3)	1.44 (0.91, 1.95)	1.83 (1.47, 2.18)	1.03 (0.75, 1.57)		
NLR				532	<0.001
Median (Q1, Q3)	3.09 (1.97, 6.43)	2.11 (1.57, 2.87)	5.25 (2.88, 9.61)		
SII				464	<0.001
Median (Q1, Q3)	827.84 (495.08, 1,720.17)	537.64 (393.49, 811.94)	1,362.11 (731.62, 2,527.54)		
PHR				0.155	0.694
Median (Q1, Q3)	1.93 (1.63, 2.32)	1.94 (1.61, 2.34)	1.93 (1.64, 2.30)		
TNV_Group cm^3				529	<0.001
Q1 (Low)	593(33.8%)	137(18.2%)	456(45.6%)		
Q2 (Medium)	586(33.4%)	146(19.4%)	440(44.0%)		
Q3 (High)	575(32.8%)	471(62.5%)	104(10.4%)		

[1]Abbreviations: IQR, interquartile range; SD, standard deviation.

[2]Continuous variables presented as median (IQR) or mean (SD); categorical variables as n (%)

[3]Continuous variables were compared using Student's t-test or Mann-Whitney U test, as appropriate. Categorical variables were compared using Pearson's chi-square test or Fisher's exact test

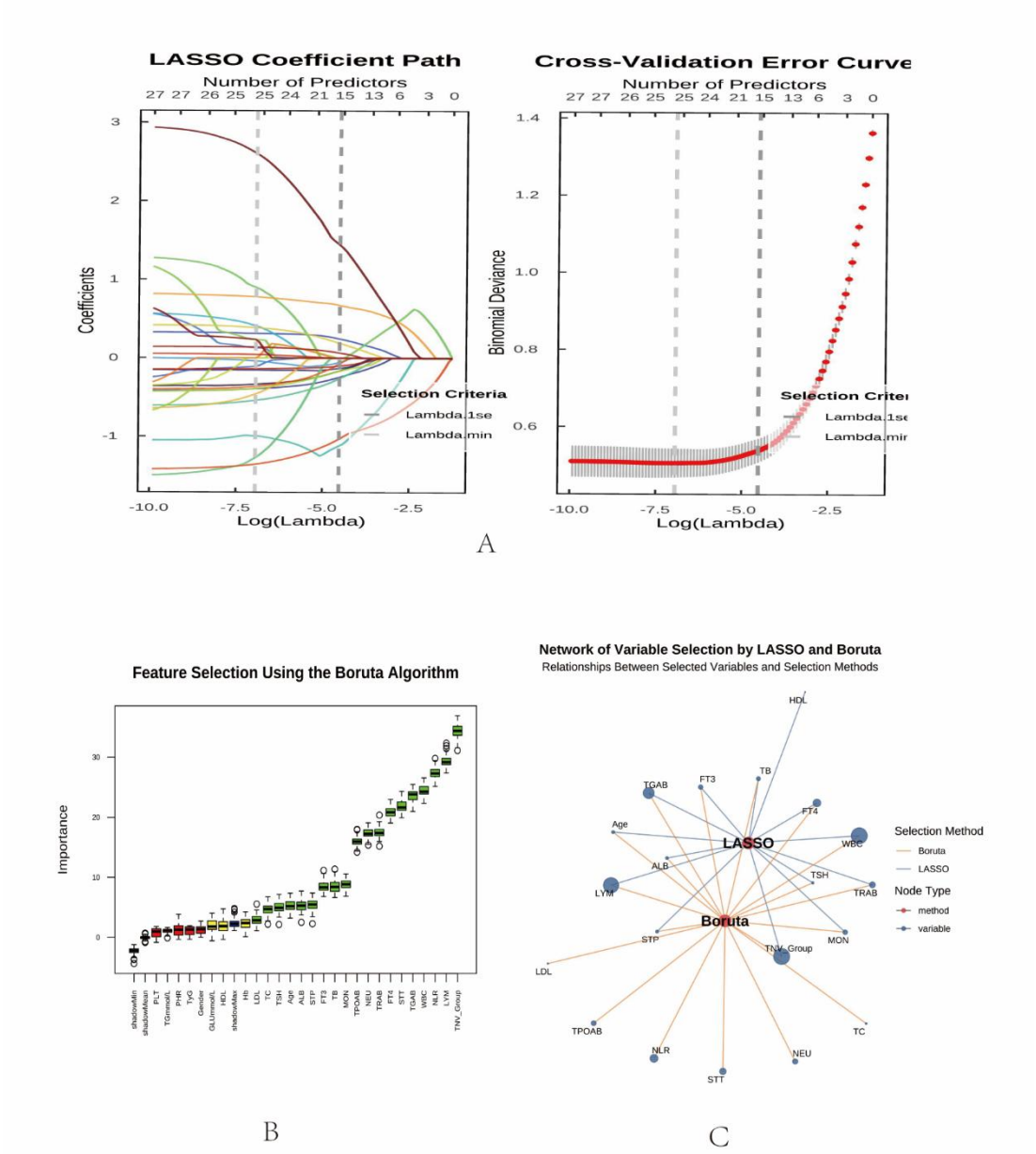
[4]Statistical significance set at $P < 0.05$ (two-tailed)

2.Feature Selection

To mitigate the potential for model overfitting and address multicollinearity concerns, we employed a dual-screening approach utilizing two distinct machine learning and statistical algorithms prior to conducting multivariate analysis. Initially, we applied Least Absolute Shrinkage and Selection Operator (LASSO) regression incorporating 10-fold cross-validation.

Adhering to the more robust one-standard-error (1-SE) rule to ensure generalizability, we identified an optimal penalty parameter of 0.01088. This process yielded a refined subset of 15 predictive features for subsequent modeling.

Concurrently, feature selection was comprehensively executed using the Boruta algorithm, which is grounded in random forest methodology. Following 99 iterative cycles, 19 features were confirmed as significant, whereas five irrelevant variables—specifically Gender, PHR, PLT, TG, and the TyG index—were excluded. Subsequently, we extracted the intersection of variables identified by both LASSO regression and the Boruta algorithm, yielding 13 robust core predictors: age, FT3, FT4, TSH, TGAb, TRAb, WBC, MON, STP, ALB, TB, and LYM, TNV. To further refine the model and enhance its clinical utility, these intersecting variables were integrated with clinical expertise into a multivariate logistic regression framework. Variable optimization and elimination were performed via bidirectional stepwise selection (direction = "both"), guided by the minimum Akaike Information Criterion (AIC). After bidirectional stepwise selection based on the AIC, six predictors were retained in the final malignancy risk prediction model: TNV_Group, FT3, FT4, TSH, STP, and TB. This specific variable combination not only achieved a balance between model fit, parsimony, and clinical interpretability. but also offers substantial practical guidance for clinical application. (Fig.1)



.

Figure 1. Variable selection using LASSO regression and the Boruta algorithm

A: LASSO coefficient path (left) and cross-validation deviance curve (right).

B: Variable importance ranking plot generated by the Boruta algorithm.

C: Network diagram showing variables screened by the LASSO regression and Boruta algorithm.

3. Multivariate logistic regression models were constructed.

Multivariate logistic regression analysis identified nodule volume, thyroid function markers, and associated biochemical parameters as independent predictors of malignancy risk in thyroid nodules. Notably, nodule volume was independently associated with lower odds of malignancy and showed a distinct dose–response relationship. Relative to the lowest volume tertile (Q1), the intermediate tertile (Q2) showed a non-significant trend toward reduced malignancy risk (OR = 0.74, 95% CI: 0.52–1.07, P = 0.1075). In contrast, the highest volume tertile (Q3) was associated with 93% lower odds of malignancy compared with the lowest tertile (Q1) (OR = 0.07, 95% CI: 0.05–0.10, P < 0.001). Furthermore, free thyroxine (FT4; OR = 1.27, 95% CI: 1.21–1.33, P < 0.001), thyroid-stimulating hormone (TSH; OR = 1.22, 95% CI: 1.09–1.37, P = 0.0008), and total bilirubin (TB; OR = 1.06, 95% CI: 1.03–1.09, P < 0.001) were established as independent risk factors for papillary thyroid carcinoma. Conversely, free triiodothyronine (FT3; OR = 0.55, 95% CI: 0.47–0.64, P < 0.001) and serum total protein (STP; OR = 0.95, 95% CI: 0.92–0.97, P = 0.0002) were independently associated with decreased odds of malignancy. (Fig2) Prior to model construction, multicollinearity diagnostics were conducted. The variance inflation factors (VIFs) for all included covariates ranged from 1.010 to 1.386, well below the conventional threshold of 5, thereby confirming the absence of multicollinearity bias among the study variables. (Table 3)

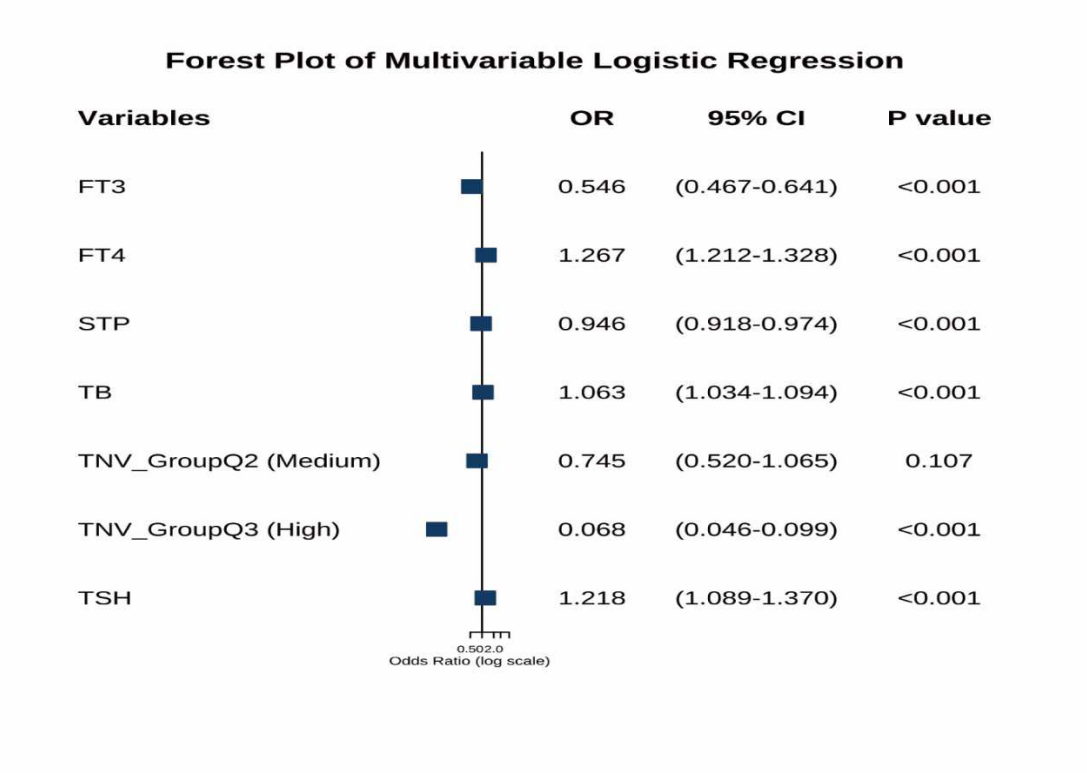


Fig 2 Forest plot of multivariate logistic regression showing independent predictors for thyroid nodule malignant risk. OR and 95% CI of each predictor are displayed.

Table 3. Multicollinearity diagnostics

Variable	VIF
TNV group Q2 (Medium) cm^3	1.345
TNV group Q3 (High) cm^3	1.386
FT3 pmol/L	1.268
FT4 pmol/L	1.327
TSH mIU/L	1.031
STP umol/L	1.010
TB g/L	1.019

Visual predictive tool: To improve clinical usability, a static nomogram was constructed in this study based on the model （Fig.3）

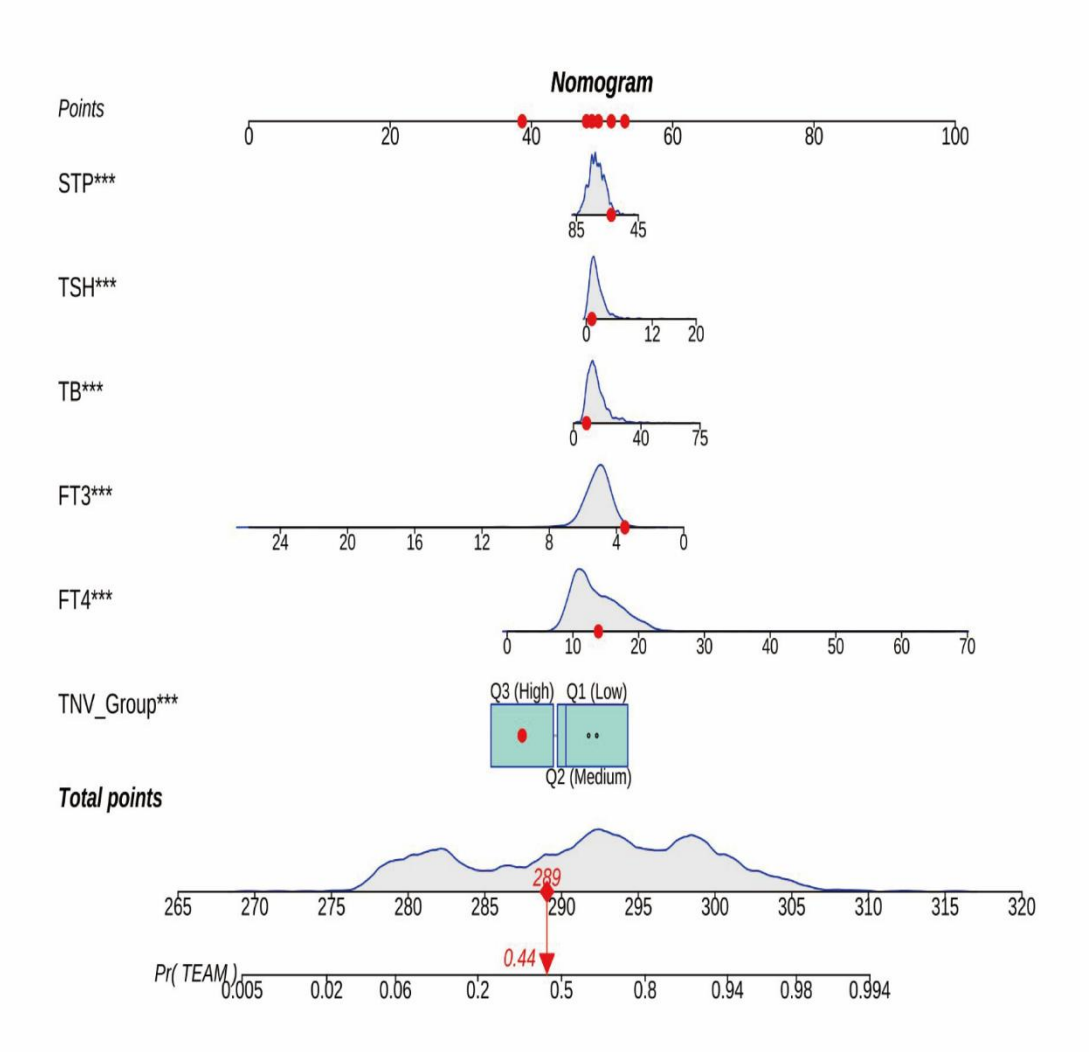


Fig3.Nomogram constructed based on the final prediction model for individualized prediction of papillary thyroid carcinoma risk in patients with thyroid nodules.

4. Model Performance Evaluation and Internal Validation

Discrimination

The multivariable prediction model developed in this study aimed to estimate the probability of PTC among surgically treated patients with thyroid nodules. The model demonstrated good predictive performance in

both the training and internal validation cohorts. As shown in Figure 4 and Table 4, the AUC was 0.870 (95% CI: 0.850–0.890) in the training cohort and 0.860 (95% CI: 0.828–0.891) in the internal validation cohort, suggesting favorable and consistent discriminative ability. Classification performance was further evaluated using the optimal cut-off value determined by the maximum Youden index. Within the training cohort, the model yielded an accuracy of 0.797, sensitivity of 0.825, specificity of 0.760, F1-score of 0.822, and a Kappa statistic of 0.585. In the validation set, performance metrics remained robust, with accuracy at 0.764, sensitivity at 0.804, specificity at 0.712, and an F1-score of 0.795. Furthermore, both positive predictive value (PPV) and negative predictive value (NPV) maintained elevated levels across both cohorts, ranging between 0.732 and 0.819. The Brier scores were 0.141 (95% CI: 0.128–0.153) in the training cohort and 0.149 (95% CI: 0.132–0.168) in the validation cohort, suggesting acceptable overall prediction accuracy.(Fig. 4; Table 4)

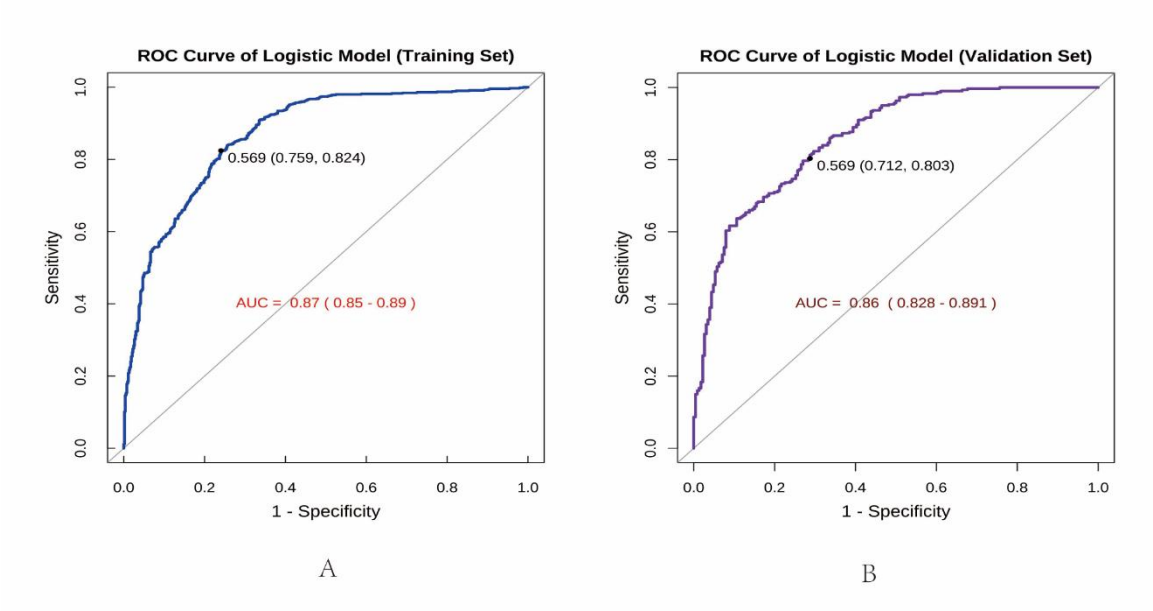


Figure 4. Receiver operating characteristic (ROC) curves for the training set (A) and validation （B）.

Table 4. Comparison of predictive performance between the training and validation sets

Dataset	Metric	Mean	CI_95
Training Set	AUC	0.87	(0.849 - 0.89)
Training Set	Accuracy	0.797	(0.774 - 0.819)
Training Set	Sensitivity	0.825	(0.795 - 0.853)
Training Set	Specificity	0.760	(0.723 - 0.798)
Training Set	PPV	0.819	(0.791 - 0.849)
Training Set	NPV	0.766	(0.729 - 0.802)
Training Set	F1	0.822	(0.799 - 0.845)
Training Set	Kappa	0.585	(0.537 - 0.631)
Training Set	Brier	0.141	(0.128 - 0.153)
Validation Set	AUC	0.860	(0.827 - 0.888)
Validation Set	Accuracy	0.764	(0.728 - 0.800)
Validation Set	Sensitivity	0.804	(0.755 - 0.848)
Validation Set	Specificity	0.712	(0.654 - 0.772)
Validation Set	PPV	0.788	(0.744 - 0.833)
Validation Set	NPV	0.732	(0.673 - 0.789)
Validation Set	F1	0.795	(0.760 - 0.830)
Validation Set	Kappa	0.517	(0.443 - 0.591)
Validation Set	Brier	0.149	(0.132 - 0.168)

Model Calibration

To evaluate model calibration, this study employed a dual-validation approach combining the Hosmer-Lemeshow (H-L) goodness-of-fit test with visual inspection of calibration curves. The H-L test indicated no statistically significant discrepancy between predicted malignancy probabilities and actual observed incidence rates in either the training cohort ($\chi^2 = 6.421$, $P = 0.093$) or the internal validation cohort ($\chi^2 = 7.085$, $P = 0.069$), as both P-values exceeded 0.05. Consequently, the null hypothesis of adequate fit was not rejected. Further quantitative assessment revealed that the calibration intercept and slope were close to the ideal values, indicating acceptable agreement between predicted and observed probabilities. In the training cohort, the calibration intercept was 0.000 and the slope was 1.000. In the internal validation cohort, the intercept was −0.101 and the slope was 1.000, suggesting generally acceptable calibration.(Fig. 5).

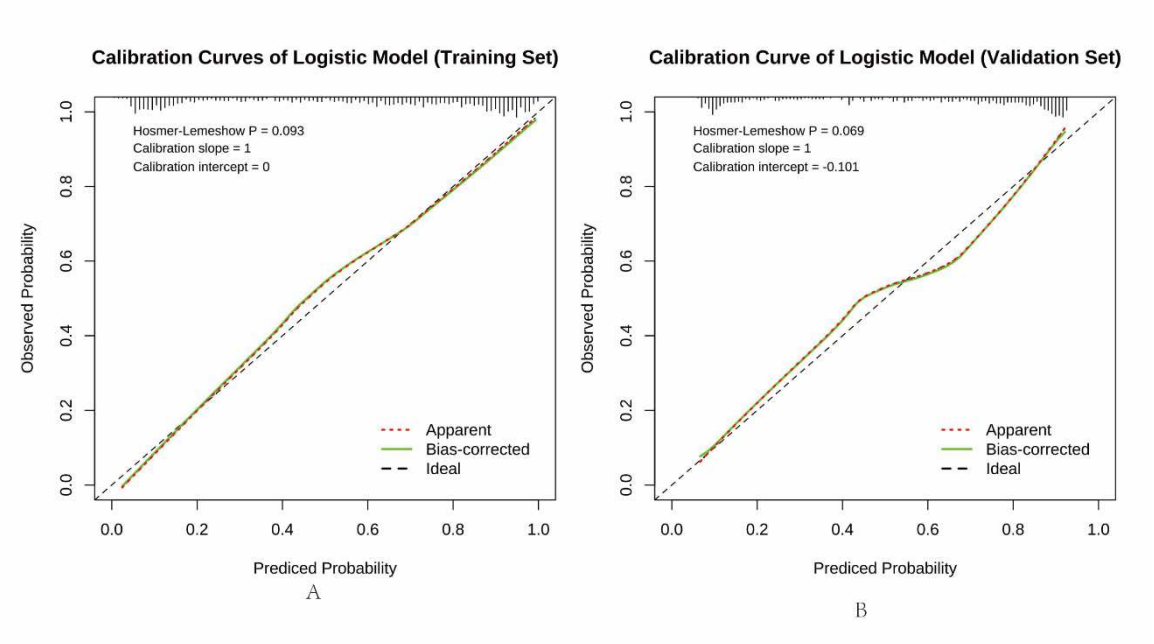


Figure 5. Calibration curves for the training set (A) and validation set (B).

Clinical Utility

To assess clinical applicability, decision curve analysis (DCA) was conducted to quantify the net benefit of the malignancy risk prediction model under realistic clinical conditions. The model showed a higher net benefit than the treat-all and treat-none strategies across a broad range of clinically relevant threshold probabilities. The internal validation cohort showed comparable findings, with the model providing greater net benefit over threshold probabilities ranging from 0.01 to 0.93. These findings suggest that the model delivers potential clinical value across a spectrum of decisions, from monitoring low-risk individuals to recommending biopsy or surgical intervention for high-risk patients, these findings support its potential clinical utility as an adjunctive tool, pending external validation. (Fig. 6).

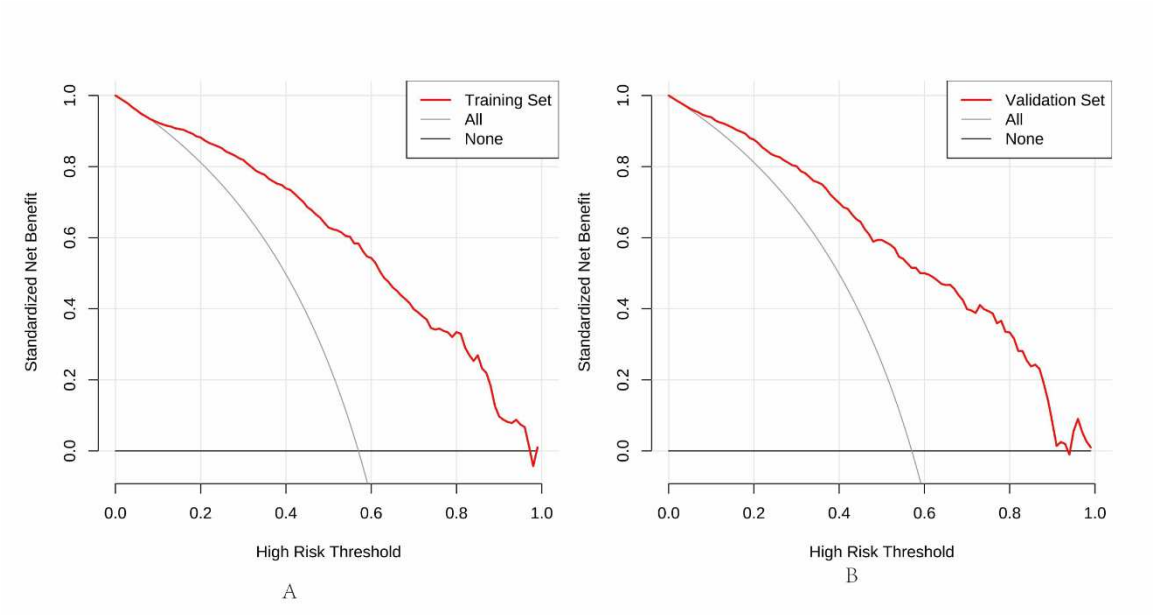


Figure 6. Decision curve analysis (DCA) for the training set (A) and validation set (B)

Discussion

Thyroid nodules are highly prevalent, with ultrasound screening detecting them in up to 67% of the population, creating a critical need for accurate risk stratification to avoid unnecessary invasive procedures [1]. Conventional systems like TIRADS rely on morphological features but demonstrate limited specificity, particularly for nodules with indeterminate cytology [17–19]. To address this gap, we developed a parsimonious prediction model incorporating six readily available parameters: nodule volume, FT3, FT4, TSH, TB, and STP. Internal validation demonstrated favorable discriminative performance, with an Area Under the Curve (AUC) of approximately 0.87 in the training cohort and 0.86 in the validation cohort, alongside favorable calibration and consistent positive net clinical benefit. A notable finding was the inverse association between nodule volume and malignant potential, where nodules in the highest volume tertile (Q3) demonstrated a 93% reduction in malignancy odds (OR = 0.07, $P < 0.001$). This finding challenges the traditional dogma that larger nodules harbor greater risk, as small-volume nodules (Q1) accounted for 45.6% of all malignant cases, whereas only 10.4% of malignancies were observed in large nodules (Q3). Smaller nodule volume was associated with a higher probability of malignancy, which may reflect the predominance of papillary thyroid microcarcinoma, the increased detection of small suspicious lesions by high-resolution ultrasonography, and the fact that larger nodules are more

often benign hyperplastic or adenomatous nodules. [20] . These findings suggest that, in this surgically confirmed case-control cohort, smaller nodule volume was associated with a higher likelihood of PTC. However, this association should not be interpreted as evidence that l larger nodules carry negligible clinical risk, because selection bias, tumor subtype distribution, and the absence of ultrasound risk stratification may influence this relationship. Distinct thyroid hormone profiles effectively differentiated benign from malignant nodules, revealing a pattern characterized by lower FT3 and higher FT4 levels in malignant cases. One possible explanation is that malignant thyroid tissue may alter deiodinase activity, potentially impairing T4-to-T3 conversion and contributing to higher FT4 and lower FT3 levels.

[21,22] . Furthermore, our model integrated TB and STP, linking systemic oxidative stress and nutritional status to thyroid tumorigenesis [23,24] . Elevated TB may reflect alterations in hepatic metabolism, oxidative stress responses, or systemic inflammatory status, although the precise biological link with thyroid carcinogenesis remains to be clarified,while lower STP may indicate poorer nutritional status or systemic inflammatory burden, which could be associated with tumor-related metabolic alterations,both of which may accelerate oncogenesis [25] . Since these markers are routine assays, their incorporation incurs no additional cost while enhancing the clinical applicability of the model. Our dual-variable selection approach combined LASSO regression (λ.1se = 0.01088) with the Boruta algorithm to mitigate overfitting and multicollinearity, yielding VIF values under 1.386 for all predictors. The model showed acceptable calibration, with non-significant Hosmer–Lemeshow tests, calibration slopes close to the ideal

value, and Brier scores of 0.141 in the training cohort and 0.149 in the validation cohort. Decision curve analysis demonstrated positive net clinical benefit across broad threshold ranges (0.08–0.97 in training; 0.01–0.93 in validation), supporting its potential clinical utility as an adjunctive risk-stratification tool. Nevertheless, limitations include the retrospective single-center design, lack of external validation, and potential batch effects from assay drift. In this case–control dataset, PPV and NPV ranged from 0.732 to 0.819; however, these values should be interpreted in light of the sampling design and may not directly reflect performance in a general clinical population.

Conclusion

This study developed a malignancy risk prediction model for thyroid nodules based on nodule volume, FT3, FT4, TSH, total bilirubin, and serum total protein. The results showed that the model had good discriminatory ability and clinical applicability, and could assist in preoperative risk stratification and clinical decision-making. Further multicenter prospective studies are needed to validate and optimize the model.

Ethics approval and consent to participate

All experimental procedures involving human tissue were conducted with ethical approval from the Ethics Committee of Longyan First Affiliated Hospital of Fujian Medical University (LYREC2024-K043-01).Due to the retrospective observational nature of the study and the complete anonymization of patient medical records, the requirement for written

informed consent was waived following approval by the institutional ethics committee and applicable regulatory guidelines. This study was conducted in accordance with the declaration of Helsinki.

Not applicable

Availability of data and materials

The datasets analysed during the current study are not publicly available due to patient privacy restrictions but are available from the corresponding author on reasonable request with approval from the institutional ethics committee

Conflict of interests

The authors declare that they have no competing interests.

Funding

Sponsored by Fujian Province Natural Science Foundation （Grant number：2025J011669）

Funding Support for the Longyan Medical Science and Technology Team for Oncology and Metabolic Diseases

Acknowledgments

We would like to acknowledge the hard and dedicated work of all the staff that implemented the intervention and evaluation components of the study.

Authors' contributions

Conception and design of the research: XQ, MT

Acquisition of data: XQ, MT

Analysis and interpretation of the data: GC

Statistical analysis: GC

Writing of the manuscript: XQ

Critical revision of the manuscript for intellectual content: XQ

All authors read and approved the final draft.

References

1. Mu C, Ming X, Tian Y, et al. Mapping global epidemiology of thyroid nodules among general population: A systematic review and meta-analysis. Front Oncol, 2022, 12: 1029926

2. Pan Q, Wang Y, Wang G. The Association Between Hyperhomocysteinemia and Thyroid Nodule Prevalence in an Adult Population. Metab Syndr Relat Disord, 2020, 18(8): 368 ~ 372

3. Grani G, Sponziello M, Filetti S, et al. Thyroid nodules: diagnosis and management. Nat Rev Endocrinol, 2024, 20(12): 715 ~ 728

4. Vural V, Calis H, Mayir B, et al. Predictors of Malignancy in Patients With a Repeated Diagnosis of Bethesda Category in Thyroid Nodules Who Underwent Surgery. Diagn Cytopathol, 2025, 53(7): 344 ~ 350

5. Tran T V T, Kitahara C M, Leenhardt L, et al. The effect of thyroid dysfunction on breast cancer risk: an updated meta-analysis. Endocr Relat Cancer, 2023, 30(1): e220155

6. Lee K H, Seok E Y, Kim E Y, et al. Different prognostic values of individual hematologic parameters in papillary thyroid cancer due to age-related changes in immunity. Ann Surg Treat Res, 2019, 96(2): 70 ~ 77

7. Wu Z, Jiang Y, Gong H, et al. Diagnostic value of preoperative systemic inflammatory markers and carcinoembryonic antigen in medullary thyroid carcinoma and the risk factors affecting its prognosis. Gland Surg, 2025, 14(1): 13 ~ 27

8. Shrestha B L, Kc A K, Rajbhandari P, et al. Does the Preoperative Neutrophil-to-lymphocyte Ratio and Platelet-to-lymphocyte Ratio Associate with Clinic-pathological Characteristics in Papillary Carcinoma of Thyroid. Kathmandu Univ Med J (KUMJ), 2021, 19(74): 225 ~ 229

9. Chen S, Wu F, Hai R, et al. Thyroid disease is associated with an increased risk of breast cancer: a systematic review and meta-analysis. Gland Surg, 2021, 10(1): 336 ~ 346

10. Viola N, Agate L, Caprio S, et al. Thyroid autoimmunity, thyroglobulin autoantibodies, and thyroid cancer prognosis. Endocr Relat Cancer, 2023, 30(7): e230042

11. Ge J, Lin B, Sun Y, et al. Preoperative thyroglobulin antibody concentrations and prediction of papillary thyroid carcinoma recurrence. Endocrine, 2025, 90(1): 237 ~ 247

12. Kim Y S, Yun J S, Kim H, et al. Acromegaly and the risk of cancer: a nationwide population-based cohort study in Korea. Eur J Endocrinol, 2025, 192(3): 220 ~ 227

13. Park J H, Choi M, Kim J H, et al. Metabolic Syndrome and the Risk of Thyroid Cancer: A Nationwide Population-Based Cohort Study. Thyroid, 2020, 30(10): 1496 ~ 1504

14. Huang Y C, Huang Z H, Cai H Z, et al.[Active surveillance for thyroid micro-malignant nodules]. Zhonghua Zhong Liu Za Zhi, 2022, 44(11): 1214 ~ 1220

15. Pappa T, Ahmadi S, Bikas A, et al.Thyroid Nodule Shape Independently Predicts Risk of Malignancy. J Clin Endocrinol Metab, 2022, 107(7): 1865 ~ 1870

16. Yang P, Gao F, Peng S, et al.Discriminating epigenetic landscapes: multi-omics characterization of benign thyroid nodules versus papillary thyroid carcinomas. Thyroid Res, 2025, 18(1): 49

17. Grimmichova T, Pacesova P, Hill M, et al.Thyroid Cancer Detection in a Routine Clinical Setting: Performance of ACR TI-RADS, FNAC, and Molecular Testing in Prospective Cohort Study. Biomedicines, 2022, 10(5): 954

18. He Y P, Xu H X, Zhao C K, et al. Cytologically indeterminate thyroid nodules: increased diagnostic performance with combination of US TI-RADS and a new scoring system. Sci Rep, 2017, 7(1): 6906

19. Seminati D, Capitoli G, Leni D, et al. Use of Diagnostic Criteria from ACR and EU-TIRADS Systems to Improve the Performance of Cytology in Thyroid Nodule Triage. Cancers (Basel), 2021, 13(21): 5439

20. Samà M T, Grosso E, Mele C, et al. Molecular characterisation and clinical correlation of papillary thyroid microcarcinoma. Endocrine, 2021, 71(1): 149 ~ 157

21. Sun J, Liu J, Wu T T, et al. Sensitivity to thyroid hormone indices are associated with papillary thyroid carcinoma in Chinese patients with thyroid nodules. BMC Endocr Disord, 2023, 23(1): 126

22. Xu H, Liu H, Hu X, et al. Reduced sensitivity to thyroid hormones is associated with differentiated thyroid cancer in the euthyroid thyroidectomy population. Front Endocrinol (Lausanne), 2025, 16: 1595002

23. Talpacci E, Morelli S, Leone V, et al.Predictors of response to lenvatinib in advanced differentiated thyroid cancer: focus on the CONUT score. Endocrine, 2025, 89(1): 143 ~ 150

24. Chen I W, Lin C W, Lin P T, et al.Non-thyroidal illness syndrome predicts poor prognosis in hepatocellular carcinoma patients treated with atezolizumab and bevacizumab. Cancer Immunol Immunother, 2026, 75(2): 59

25. Jia J L, Han J H, Pang R, et al. Predictors of poor prognosis in long-term survivors of differentiated thyroid cancer with psychiatric disorders. World J Psychiatry, 2025, 15(5): 103628