Am J Transl Res 2026;18(11):9910-9925
www.ajtr.org /ISSN:1943-8141/AJTR0174664

Original Article

Identification of independent risk factors for postoperative TSH target non-attainment for differentiated thyroid cancer and development of an XGBoost prediction model

Hanling Zeng, Li Shen

Department of Thyroid and Breast Surgery, The Fourth Affiliated Hospital of Nanjing Medical University, Nanjing 210031, Jiangsu, China

Received June 10, 2026; Accepted August 10, 2026; Epub November 15, 2026; Published November 30, 2026

Abstract: Aims: To identify independent risk factors fpr postoperative thyroid-stimulating hormone (TSH) target non-attainment in patients with differentiated thyroid cancer receiving endocrine therapy, and to develop and internally validate an interpretable XGBoost-based model for individualized risk prediction, thereby facilitating early risk stratification and optimization of postoperative endocrine management. Methods: We conducted a retrospective cohort study of consecutive patients with differentiated thyroid cancer who underwent surgery followed by postoperative endocrine therapy at the Fourth Affiliated Hospital of Nanjing Medical University between March and November 2025. Among 280 initially screened patients, 278 with complete follow-up data were included and classified according to postoperative TSH target attainment into a non-attainment group (n = 105) and an attainment group (n = 173). The primary endpoint was postoperative TSH target attainment during follow-up. Clinicopathological characteristics, biochemical parameters, treatment-related factors, and follow-up variables were analyzed to identify independent predictors using LASSO-assisted multivariable logistic regression. Subsequently, an individualized XGBoost model was developed and internally validated, with performance assessed using discrimination, calibration, decision-curve analysis, and subgroup and sensitivity analyses. Results: Among the 278 patients with differentiated thyroid cancer, postoperative TSH non-attainment occurred in 105 (37.8%). Preoperative thyroid function indices did not differ between groups. In contrast, after surgery and endocrine therapy, postoperative thyroglobulin (Tg) and TSH, but not FT3 or FT4, diverged significantly between groups, indicating that biochemical differences were primarily associated with disease- and treatment-related markers rather than overall thyroid hormone levels. LASSO-assisted multivariable analysis identified extrathyroidal invasion, lymph node metastasis, BRAF V600E mutation, elevated postoperative Tg levels, and intermediate/high recurrence risk as independent predictors of failure to achieve target TSH levels. In addition, endocrine therapy adherence demonstrated the strongest association, with poor adherence associated with the highest risk of TSH target non-attainment. Based on these variables, an interpretable XGBoost model was developed and achieved strong discrimination in both the training and validation cohorts (AUC, 0.893 and 0.870, respectively), accompanied by good calibration and consistent net clinical benefit in decision-curve analysis. Model performance remained robust across analyses stratified by sex, age, pathological characteristics, and sensitivity analyses. Conclusion: Postoperative thyroglobulin levels, recurrence risk stratification, extrathyroidal extension, lymph node metastasis, endocrine therapy adherence, and BRAF V600E mutation were identified as independent predictors of postoperative TSH target attainment in patients with differentiated thyroid cancer receiving endocrine therapy. The XGBoost model developed on the basis of these variables demonstrated strong predictive performance for identifying patients at risk of TSH target non-attainment and maintained good performance after internal validation.

Keywords: Differentiated thyroid cancer, TSH target attainment, postoperative endocrine therapy, independent predictors, XGBoost prediction model

Introduction

Differentiated thyroid cancer (DTC), including papillary and follicular thyroid carcinoma, accounts for the vast majority of thyroid cancers and has exhibited a consistent increase in incidence worldwide over recent decades [1-3]. Although the general prognosis of DTC is gener-

https://doi.org/10.62347/KVIN4874

ally favorable, postoperative management remains clinically significant because a considerable proportion of patients experience biochemical dysfunctions, structural recurrence, or require long-term surveillance and endocrine regulation. In this context, thyroid-stimulating hormone (TSH) suppression or levothyroxine dose adjustment has long been viewed as a cornerstone of postoperative management, based on the biological rationale that TSH may stimulate the proliferation of remaining thyroid tissue and thyroid cancer cells [4-6]. Nevertheless, postoperative TSH management is significantly more complex in clinical practice than a simple "the lower, the better" policy. Current American Thyroid Association (ATA) recommendations highlight that the extent and duration of TSH suppression should be tailored to the recurrence risk, response to treatment, and the balance between treatment-related harm against potential oncologic benefit [7]. This personalized treatment has become increasingly important, as the advantages of intensive TSH suppression appear to be more evident in selected high-risk populations, whereas unnecessary suppression can expose patients to cardiovascular, skeletal, and quality-of-life-related harms without apparent clinical benefit.

Although clinically important, achieving the desired postoperative TSH target remains challenging in clinical practice. The likelihood of achieving TSH targets may reflect a complex interplay among tumor aggressiveness, postoperative biochemical status, recurrence risk stratification, treatment intensity, and patient adherence [8, 9]. However, previous studies have primarily focused on recurrence, survival outcomes, or overall effectiveness of TSH suppression [10-12], rather than addressing the clinically relevant question of which patients are most likely to fail to achieve TSH targets after surgery. Consequently, evidence supporting the early identification of patients at risk of inadequate hormonal control after surgery remains limited. This represents a clinically significant gap, as failure to reach target TSH levels may compromise risk-adapted follow-up strategies, delay therapeutic optimization, and reflect increased complexity in postoperative management.

Meanwhile, machine-learning methods have increasingly been applied in thyroid cancer research due to their ability to capture nonlinear relationships and complex interactions that may not be adequately captured by conventional regression models. Recent studies have proposed that interpretable machine-learning frameworks, especially models integrating XGBoost with Shapley additive explanations (SHAP), can provide accurate and clinically meaningful prediction of thyroid cancer outcome while maintaining interpretability [13-15]. However, these strategies have rarely been applied to postoperative TSH target achievement, which represents a clinically relevant endocrine-management outcome with direct implications for personalized care.

Therefore, this study aimed to identify independent risk factors for postoperative TSH target non-achievement in patients with DTC undergoing endocrine therapy and to develop an interpretable XGBoost-based prediction model for individualized risk assessment. In a retrospective cohort of 278 patients, we assessed clinicopathological features, perioperative thyroid function indicators, molecular features, and treatment adherence, and further assessed model discrimination, calibration, clinical utility, subgroup performance, and sensitivity. This study intended to offer a clinically applicable tool to facilitate risk-adaptive endocrine management in DTC by focusing on postoperative TSH target attainment rather than conventional oncologic endpoints.

Methods

Case design

This retrospective single-center cohort study included consecutive patients with differentiated thyroid cancer (DTC) who underwent thyroid surgery followed by postoperative levothyroxine therapy at the Fourth Affiliated Hospital of Nanjing Medical University between March and November 2025. Clinical, pathological, treatment-related, and follow-up data were retrospectively retrieved from the hospital electronic medical record system and postoperative follow-up database. A total of 280 patients were screened for eligibility; two patients were excluded due to loss to follow-up, resulting in a final cohort of 278 patients.

The postoperative TSH target for each patient was determined according to the risk-adapted

management strategy documented in the medical record, taking into account the initial recurrence risk, extent of surgery, and potential adverse effects associated with TSH suppression. Generally, the target TSH level was < 0.1 mIU/L for patients with a high risk of recurrence, 0.1-0.5 mIU/L for those with an intermediate risk of recurrence, and 0.5-2.0 mIU/L for those with a low risk of recurrence or those managed after lobectomy, unless individualized targets were established due to clinically relevant contraindications to intensive TSH suppression.

TSH target attainment was defined as a serum TSH level within the prespecified, patient-specific target range at the designated postoperative follow-up assessment. Non-attainment was defined as a serum TSH level outside the corresponding target range. Based on these predefined criteria, 105 patients were classified into the non-attainment group and 173 into the attainment group. The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Fourth Affiliated Hospital of Nanjing Medical University. The requirement for written informed consent was waived because of the retrospective design and the use of de-identified clinical data. The detailed study flowchart in shown in **Figure 1**.

Inclusion and exclusion criteria

Inclusion criteria: (1) histopathologically confirmed DTC, including papillary thyroid carcinoma and follicular thyroid carcinoma; (2) patients who underwent thyroid surgery at the Fourth Affiliated Hospital of Nanjing Medical University during the study period; (3) receipt of postoperative levothyroxine therapy aimed at achieving individualized TSH targets; and (4) availability of complete baseline clinicopathological information and postoperative follow-up data for evaluation of TSH target achievement.

Exclusion criteria: (1) absence of definitive pathological verifications of DTC; (2) presence of other thyroid malignancies or non-differentiated histological subtypes, such as medullary thyroid carcinoma or anaplastic thyroid carcinoma; (3) history of previous thyroid surgery, radioactive iodine therapy, or other antitumor treatments that could substantially influence preoperative endocrine management; (4) severe dysfunction of major organs or other serious systemic diseases that might affect levothyroxine adherence or TSH assessment; (5) incomplete clinical, pathological, treatment-related, or follow-up data; or (6) loss to follow-up during the postoperative follow-up period.

Data collection

The extracted variables included demographic characteristics (age, sex, body mass index [BMI], smoking history, alcohol consumption, hypertension, and diabetes mellitus); pathological and tumor-related characteristics (pathological subtype, tumor diameter, number of lesions, extrathyroidal invasion, vascular invasion, lymph node metastasis, and BRAF V600E mutation status); treatment-related variables (surgical procedure, risk stratification for adverse effects associated with TSH suppression therapy, recurrence-risk stratification, and adherence to postoperative endocrine therapy); and thyroid function parameters measured before and after surgery, including free triiodothyronine (FT3), free thyroxine (FT4), thyroglobulin (Tg), and TSH.

The postoperative follow-up duration was calculated from the date of surgery to the designated biochemical assessment and was summarized descriptively. To minimize the influence of transient postoperative hormonal fluctuations and recent levothyroxine dose adjustments, the endpoint TSH value was defined as the measurement obtained after the patient had received an unchanged levothyroxine regimen for at least 6 weeks. When multiple eligible TSH measurements were available, the latest measurement within the study-defined follow-up period was selected, and the corresponding FT3, FT4, and Tg values measured at the same visit were recorded. This assessment was selected to characterize early postoperative biochemical control after adequate stabilization of levothyroxine dosage and to identify patients who might benefit from timely reinforcement of medication adherence, dose adjustment, or closer monitoring. Importantly, TSH target attainment at this designated assessment was not considered evidence of sustained long-term endocrine control, which would require serial TSH measurements over an extended follow-up period.

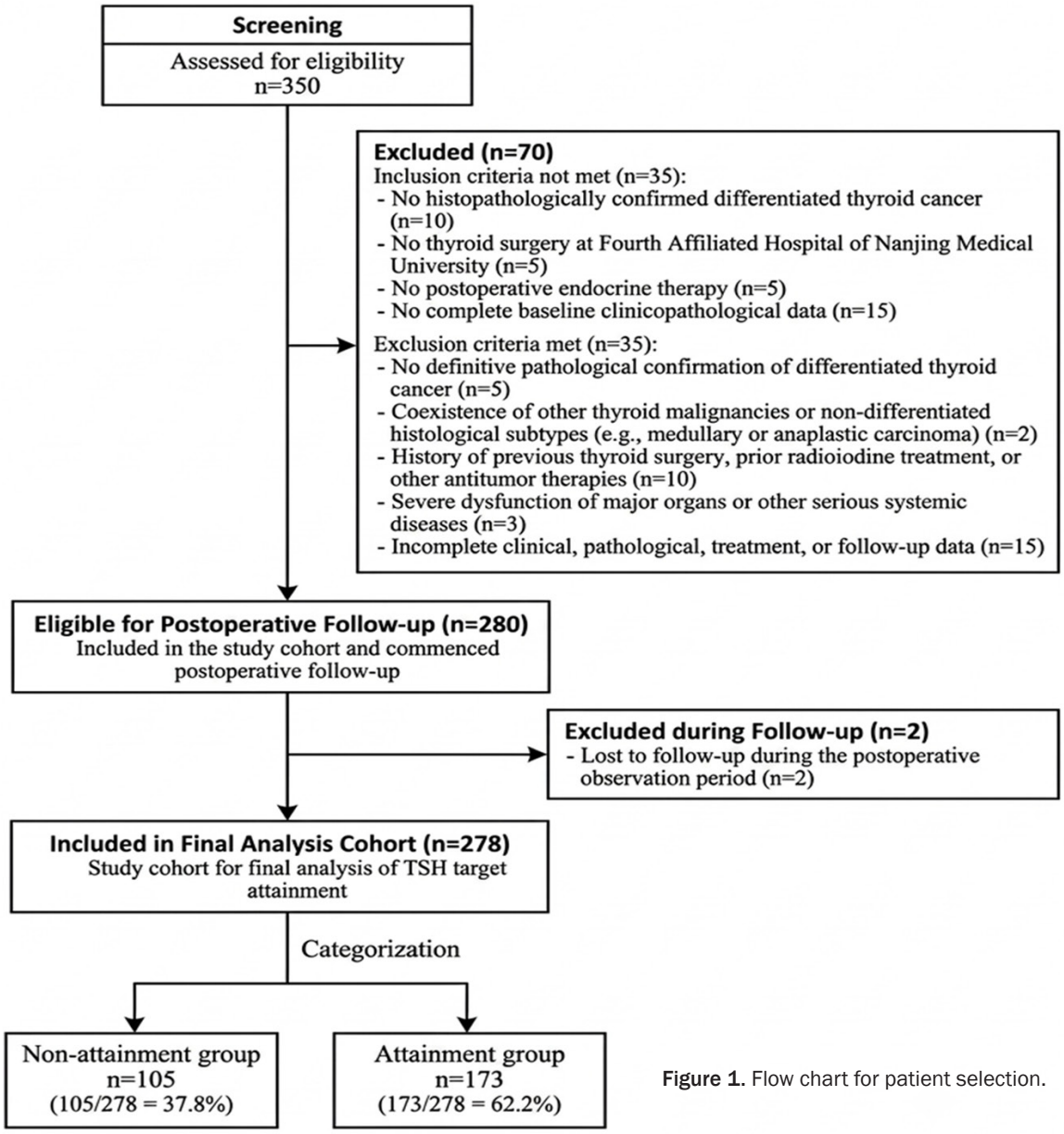


Figure 1. Flow chart for patient selection.

Adherence to postoperative endocrine therapy was assessed according to prespecified criteria documented in standardized postoperative follow-up records. The assessment incorporated the regularity of levothyroxine administration, documented missed or interrupted doses, compliance with the prescribed dose and recommended administration schedule, and attendance at scheduled medication-review visits. Based on the predefined grading criteria in the follow-up protocol, patients were categorized into three adherence groups: good, moderate, or poor adherence. Good adherence was defined as regular levothyroxine intake without documented treatment interruption or clinically relevant missed doses, appropriate compliance with the prescribed dose and administration schedule, and regular attendance at follow-up visits; Moderate adherence was defined as occasional missed doses, minor deviations from the recommended administration schedule, or intermittent noncompliance with follow-up visits without prolonged treatment interruption; Poor adherence was defined as frequent missed doses, documented interruption or discontinuation of levothyroxine therapy, persistent inappropriate medication administration, or repeated failure to attend

scheduled follow-up assessments. These criteria were applied consistently to all patients based on available follow-up documentation. Therefore, adherence classification was based on structured and reproducible clinical information rather than an unstructured subjective assessment by the treating physician.

For regression and machine-learning analyses, postoperative Tg was included as a continuous predictor. Categorical variables, including number of lesions, extrathyroidal invasion, vascular invasion, lymph node metastasis, BRAF V600E mutation status, risk stratification for adverse effects of TSH suppression therapy, recurrence-risk stratification, and adherence to endocrine therapy, were coded according to predefined binary or ordinal assignment rules. Continuous variables, including tumor diameter and postoperative Tg, were standardized using Z-scores. Eligible patients were classified into the TSH target-attainment or non-attainment group according to whether the selected endpoint TSH value fell within the prespecified patient-specific target range. Accordingly, the primary outcome reflected TSH target attainment at a predefined postoperative assessment rather than persistent long-term TSH control. The entire cohort was subsequently randomly divided into training and internal validation cohorts for the development and internal validation of the XGBoost model.

Outcome measures

The primary outcome of the study was postoperative TSH target attainment during follow-up in patients with DTC undergoing postoperative endocrine therapy. Follow-up TSH measurements and treatment records were used to categorize patients into the non-attainment and attainment groups for comparative analyses. Secondary outcomes included identifying independent predictors associated with failure to achieve postoperative TSH targets and developing a personalized prediction model for postoperative TSH target non-attainment. To identify independent risk factors, associations between clinicopathological characteristics, treatment-related variables, and TSH target attainment were assessed using regression analyses. For machine-learning analyses, model performance was evaluated in the training and validation cohorts with respect to discrimination, calibration, and clinical utility, using the area under the receiver operating characteristic curve (ROC-AUC), calibration plots, and decision curve analysis (DCA), respectively. Additional subgroup and sensitivity analyses were performed to evaluate the stability and robustness of model performance across clinically relevant patient subsets and under alternative analytical situations.

XGBoost model development and hyperparameter tuning

The XGBoost model was developed to predict postoperative TSH target non-attainment using predictors retained from the preceding feature-selection procedure. The entire cohort was randomly divided at a 2:1 ratio into a training cohort and a held-out internal validation cohort. All data preprocessing, feature selection, hyperparameter optimization, and model fitting procedures were performed exclusively within the training cohort, whereas the validation cohort remained inaccessible until final model evaluation. Hyperparameter tuning was conducted using a randomized search with stratified five-fold cross-validation. The mean cross-validated AUC was used as the primary criterion for model selection. A total of 100 randomly sampled parameter combinations were evaluated. The predefined search space included the following parameters: n_estimators (100-800), max_depth (2-6), learning_rate (0.01-0.20), min_child_weight (1-10), gamma (0-0.50), subsample (0.60-1.00), colsample_bytree (0.60-1.00), reg_alpha (0-1.00), and reg_lambda (0.50-10.00). These parameters were selected because max_depth, min_child_weight, and gamma, regulate tree complexity and reduce overfitting caused by splits based on limited observations; learning_rate and n_estimators jointly determine the rate and extent of boosting; subsample and colsample_bytree reduce variance by introducing row- and feature-level randomness; and reg_alpha and reg_lambda provide L1 and L2 regularization, respectively, to further limit overfitting.

Early stopping was applied when the cross-validated AUC failed to improve over 30 consecutive boosting rounds. The optimal hyperparameter configuration was n_estimators = 300, max_depth = 3, learning_rate = 0.05, min_child_weight = 3, gamma = 0.10, subsample = 0.80, colsample_bytree = 0.80, reg_alpha = 0.10, and reg_lambda = 1.00. This configura-

tion achieved the highest mean cross-validated AUC while maintaining relatively shallow trees, a conservative learning rate, stochastic subsampling, and explicit regularization. The final model was retrained using the complete training cohort and subsequently evaluated once in the held-out validation cohort. A fixed random seed of 2025 was applied throughout data partitioning, cross-validation, and model training to ensure reproducibility. Machine-learning analyses were performed using Python version 3.11.8, XGBoost version 2.0.3, and scikit-learn version 1.4.2. SHAP version 0.45.1 was used to calculate and visualize the direction and magnitude of individual predictor contributions.

Statistical analysis

All statistical analyses were performed using conventional statistical methods and machine-learning methods. Continuous variables were expressed as mean ± standard deviation (SD) and compared using independent-samples t-test. Categorical variables were summarized as frequencies and percentages and compared using the chi-square test or Fisher's exact test, as appropriate.

Candidate clinicopathological, treatment-related variables were initially screened using least absolute shrinkage and selection operator (LASSO) regression. Variables with non-zero coefficients were subsequently entered into multivariable logistic regression model to identify independent predictors. Results were presented as odds ratios and 95% confidence intervals (CIs). For machine-learning analyses, the entire cohort was randomly divided into training and validation cohorts at a 2:1 ratio. The training cohort was used for XGBoost model development, while the validation cohort was used for internal validation. Model interpretability was assessed using Shapley additive explanations (SHAP), including summary and waterfall plots.

Model performance, including discrimination, calibration, and clinical utility, was evaluated using AUC, calibration curves, and CAS, respectively. Sensitivity analyses and subgroup analyses were also conducted to assess model robustness and applicability across various clinical situations. All statistical tests were two-sided, and a *P* value of < 0.05 was considered statistically significant.

Results

Comparison of baseline clinical characteristics between the two groups

Baseline demographic characteristics were generally comparable between the non-attainment and attainment groups, with no significant differences observed in sex, age, BMI, smoking status, alcohol consumption, comorbid hypertension or diabetes, pathological subtype, or surgical procedure (all P > 0.05).

In contrast, patients in the non-attainment group exhibited significantly more adverse clinicopathological characteristics, including larger tumor size (P < 0.001), higher frequency of multifocal disease (P = 0.005), extrathyroidal invasion (P < 0.001), vascular invasion (P = 0.003), lymph node metastasis (P = 0.001), and BRAF V600E mutation (P = 0.008). In addition, patients in the non-attainment group were more likely to be classified as having intermediate-to-high risk for adverse effects associated with TSH suppression therapy (P = 0.004) and recurrence (P < 0.001). Endocrine therapy adherence also differed significantly between groups, with poorer adherence observed more frequently in the non-attainment group (P < 0.001) (**Table 1**).

Comparison of preoperative and postoperative thyroid function indicators between two groups

As shown in **Figure 2**, preoperative thyroid function parameters were comparable between the non-attainment and attainment groups, with no significant differences in FT3, FT4, Tg, or TSH (all P > 0.05). After surgery and postoperative endocrine therapy, FT3 and FT4 levels remained comparable between groups (both P > 0.05), whereas postoperative Tg and TSH differed significantly between groups (both P < 0.001). In addition, comparisons between preoperative and postoperative measurements demonstrated marked reductions in FT3, FT4, and Tg, accompanied by a significant increase in TSH levels after treatment (all P < 0.001).

LASSO-logistic regression analysis of TSH target attainment after postoperative endocrine therapy in patients with DTC

As illustrated in **Figure 3** and Supplementary Table 1, LASSO regression was applied to candidate clinicopathological and treatment-relat-

Table 1. Comparison of baseline characteristics between the two groups

Variables	Non-attainment group (n = 105)	Attainment group (n = 173)	t/χ^2	P
Sex			0.250	0.617
Male	21 (20.00)	39 (22.54)		
Female	84 (80.00)	134 (77.46)		
Age (years)	46.28 ± 8.13	44.79 ± 7.68	1.534	0.126
BMI (kg/m^2)	22.89 ± 1.42	23.03 ± 1.50	-0.770	0.442
Smoking history	36 (34.29)	55 (31.79)	0.185	0.667
Drinking history	32 (30.48)	50 (28.90)	0.078	0.780
Hypertension	18 (17.14)	26 (15.03)	0.219	0.640
Diabetes mellitus	15 (14.29)	21 (12.14)	0.267	0.605
Tumor diameter (cm)	2.48 ± 0.53	2.09 ± 0.47	6.389	< 0.001
Pathological type			0.258	0.611
Papillary carcinoma	103 (98.10)	171 (98.84)		
Follicular carcinoma	2 (1.90)	2 (1.16)		
Number of lesions			7.810	0.005
Solitary	58 (55.24)	124 (71.68)		
Multiple	47 (44.76)	49 (28.32)		
Extrathyroidal invasion			15.743	< 0.001
Yes	49 (46.67)	41 (23.70)		
No	56 (53.33)	132 (76.30)		
Vascular invasion			8.543	0.003
Yes	22 (20.95)	15 (8.67)		
No	83 (79.05)	158 (91.33)		
Lymph node metastasis			11.233	0.001
Yes	83 (79.05)	103 (59.54)		
No	22 (20.95)	70 (40.46)		
BRAF V600E mutation			6.997	0.008
Yes	71 (67.62)	89 (54.45)		
No	34 (32.38)	84 (48.55)		
Surgical procedure			0.615	0.893
Lobectomy	52 (49.52)	93 (53.76)		
Total thyroidectomy	48 (45.71)	72 (41.62)		
Lobectomy + partial lobectomy	4 (3.81)	7 (4.05)		
Risk stratification for adverse effects of TSH suppression therapy			8.469	0.004
Low risk	60 (57.14)	128 (73.99)		
Intermediate-high risk	45 (42.86)	45 (26.01)		
Recurrence risk stratification			26.725	< 0.001
Low risk	37 (35.24)	116 (67.05)		
Intermediate-high risk	68 (64.76)	57 (32.95)		
Adherence to endocrine therapy			5.908	< 0.001
Good	28 (26.67)	112 (64.74)		
Moderate	42 (40.00)	43 (24.86)		
Poor	35 (33.33)	18 (10.10)		

ed variables. The optimal penalty parameter identified a parsimonious subset of predictors, which were subsequently incorporated into a multivariable logistic regression model.

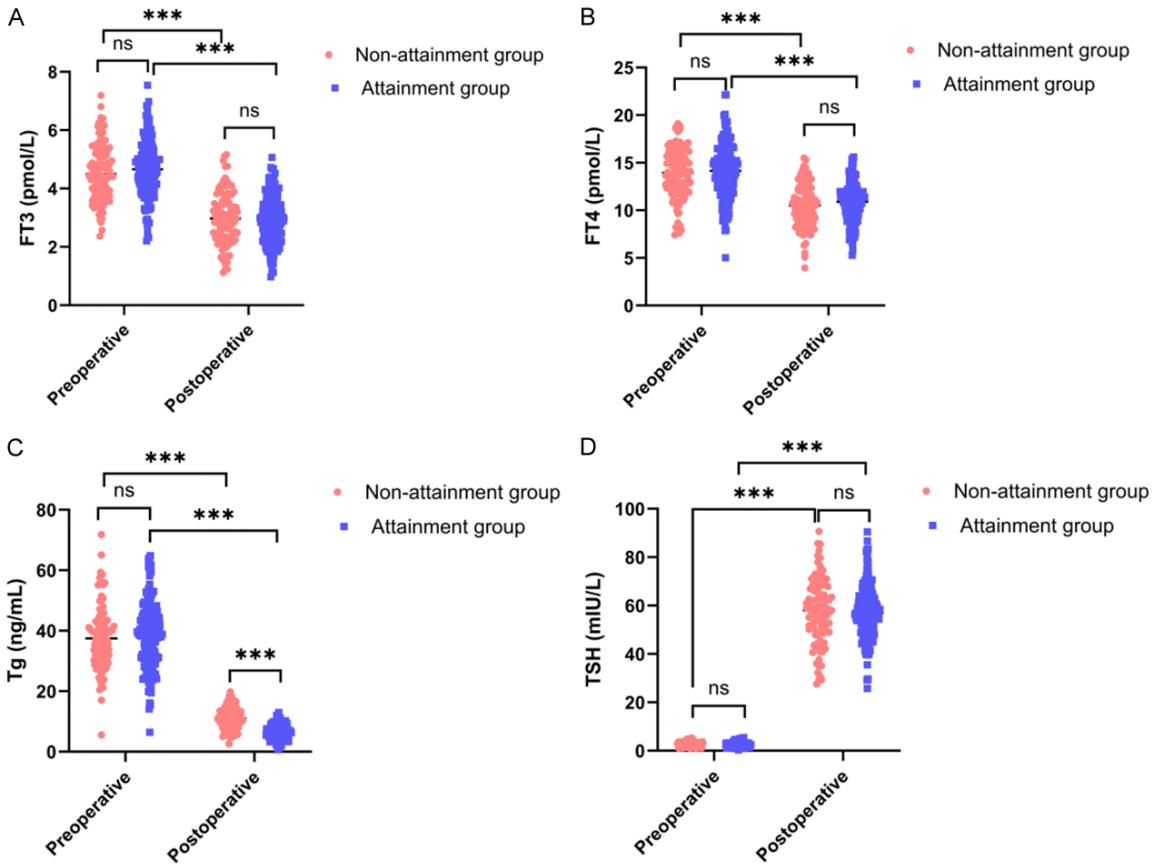


Figure 2. Comparison of preoperative and postoperative thyroid function parameters between two groups. A. FT3. B. FT4. C. Tg. D. TSH. Notes: FT3, free triiodothyronine; FT4, free thyroxine; Tg, thyroglobulin; TSH, thyroid-stimulating hormone. Compare with non-attainment group, ns, not significant ($P > 0.05$); ***$P < 0.001$.

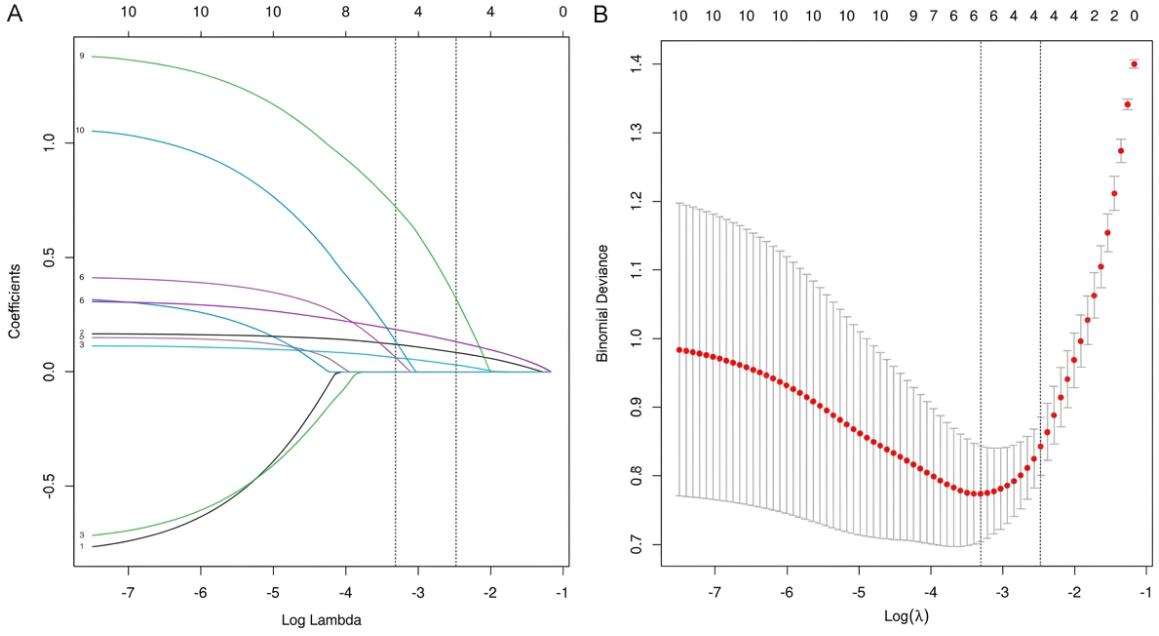


Figure 3. Selection of candidate predictors using the least absolute shrinkage and selection operator (LASSO) regression model. A. LASSO coefficient profiles of candidate variables plotted against log(λ). Each colored curve represents the trajectory of an individual regression coefficient as the penalty parameter increased. B. Ten-fold cross-validation for tuning parameter selection in the LASSO model. The red dots indicate the mean cross-validated binomial deviance at each value of log(λ), and the vertical bars represent standard errors. The two dotted vertical lines denote the optimal λ values according to the minimum criteria and the 1-standard-error criterion, respectively. Based on the optimal penalization parameter, variables with non-zero coefficients were retained for subsequent multivariable logistic regression analysis.

Table 2. LASSO-logistic regression analysis of factors associated with postoperative TSH target non-attainment in patients with DTC

Variable	β	S.E.	Wald χ^2	OR (95% CI)	*P* value
Extrathyroidal invasion	0.862	0.315	7.489	2.368 (1.277-4.391)	0.006
Lymph node metastasis	1.035	0.302	11.748	2.815 (1.557-5.089)	0.001
BRAF V600E mutation	0.658	0.281	5.485	1.931 (1.113-3.351)	0.019
Postoperative Tg	0.901	0.215	17.082	2.462 (1.606-3.775)	< 0.001
Recurrence risk stratification (intermediate/high vs. low)	1.168	0.309	14.287	3.215 (1.754-5.892)	< 0.001
Endocrine therapy adherence					
Moderate vs. Good	1.258	0.361	12.145	3.518 (1.734-7.138)	< 0.001
Poor vs. Good	2.128	0.418	25.921	8.399 (3.701-19.062)	< 0.001
Constant	-3.872	0.526	54.182	-	< 0.001

Table 3. Comparison of baseline characteristics between the training and validation cohorts

Variables	Training cohort (n = 194)	Validation cohort (n = 84)	t/χ^2	*P* value
Age, years	45.28 ± 7.89	45.86 ± 8.11	-0.540	0.590
Female sex, n (%)	153 (78.87)	65 (77.38)	0.077	0.781
Tumor diameter, cm	2.24 ± 0.51	2.27 ± 0.50	-0.448	0.655
Multiple lesions, n (%)	64 (32.99)	32 (38.10)	0.642	0.423
Extrathyroidal invasion, n (%)	64 (32.99)	26 (30.95)	0.112	0.738
Lymph node metastasis, n (%)	132 (68.04)	54 (64.29)	0.374	0.541
BRAF V600E mutation, n (%)	111 (57.22)	49 (58.33)	0.028	0.867
Postoperative Tg, ng/mL	9.12 ± 3.51	8.94 ± 3.48	0.390	0.697
Intermediate/high recurrence risk, n (%)	89 (45.88)	36 (42.86)	0.206	0.650
Good adherence, n (%)	98 (50.52)	42 (50.00)	0.006	0.938
Non-attainment events, n (%)	74 (38.14)	31 (36.90)	0.040	0.841

Multivariable analysis identified extrathyroidal invasion (P = 0.006), lymph node metastasis (P = 0.001), BRAF V600E mutation (P = 0.019), elevated postoperative Tg (P < 0.001), and intermediate/high recurrence risk stratification (P < 0.001) as independent risk factors for postoperative TSH target non-attainment. Furthermore, endocrine therapy adherence showed a strong graded association with TSH target attainment. Compared with good adherence, moderate adherence (P < 0.001) and poor adherence (P < 0.001) were both associated with reduced likelihood of achieving the target TSH range, with poor adherence exerting the strongest association (**Table 2**).

Comparison of baseline characteristics between the training and validation cohorts

The 278 patients included in this study were randomly divided into a training cohort (n = 194) and a validation cohort (n = 84) at a 2:1 ratio. As shown in **Table 3**, baseline clinicopathological characteristics were well balanced between the two cohorts, with no significant differences in age (P = 0.590), sex distribution (P = 0.781), tumor diameter (P = 0.655), number of lesions (P = 0.423), extrathyroidal invasion (P = 0.738), lymph node metastasis (P = 0.541), BRAF V600E mutation status (P = 0.867), postoperative Tg level (P = 0.697),

recurrence risk stratification (P = 0.650), or endocrine therapy adherence (P = 0.938). In addition, the proportion of TSH target non-attainment events was comparable between the training and validation cohorts (P = 0.841).

Development of the XGBoost prediction model for postoperative TSH target non-attainment in patients with DTC

Based on the variables retained from the preceding feature-selection procedure, an XGBoost prediction model was established to estimate the probability of postoperative TSH target non-attainment in DTC patients receiving postoperative endocrine therapy.

As illustrated in **Figure 4**, SHAP-based model interpretability analysis demonstrated that the final model effectively captured the relative importance and directional effects of individual predictors on model prediction. The SHAP summary plot showed clear heterogeneity in feature contributions among individual patients, indicating that the risk of TSH target non-attainment was jointly influenced by multiple clinicopathological and treatment-related factors rather than by any single variable alone.

In addition, representative SHAP waterfall plots illustrated how specific predictors contributed positively or negatively to individual predicted risks, highlighting the capacity of the XGBoost model to provide individualized risk stratification with improved interpretability.

Validation of the XGBoost prediction model for postoperative TSH target non-attainment in patients with DTC

The predictive performance of the XGBoost model was comprehensively evaluated in terms of discrimination, calibration, and clinical utility. As shown in **Figure 5**, the model demonstrated good discriminative performance in both the training and validation cohorts, with an AUC of 0.893 (95% CI: 0.854-0.932) in the training cohort and 0.870 (95% CI: 0.799-0.941) in the validation cohort, indicating satisfactory and stable predictive performance. Calibration plots (**Figure 6**) showed good agreement between predicted and observed probabilities in both cohorts, suggesting satisfactory calibration of the model. In addition, DCA (**Figure 7**) demonstrated that the XGBoost model achieved greater net clinical benefit than the treat-all or treat-none strategies across a broad range of threshold probabilities in both the training and validation cohorts, supporting its potential clinical utility for individualized postoperative endocrine management.

Subgroup and sensitivity analyses of the XGBoost model

As shown in **Figure 8**, the XGBoost model maintained stable discriminative performance across clinically relevant subgroups. Among female patients, the model achieved an AUC of 0.874 (95% CI, 0.819-0.929), while in male patients, the AUC was 0.848 (95% CI, 0.729-0.967), with corresponding accuracies of 0.821 and 0.800, respectively. Similarly, stable predictive ability was observed across age strata, with an AUC of 0.889 (95% CI, 0.824-0.954) in patients aged < 45 years and 0.867 (95% CI, 0.799-0.935) in those aged ≥ 45 years. In the pathology-restricted analysis including only patients with papillary thyroid carcinoma, the model maintained favorable classification performance, achieving an AUC of 0.882 (95% CI, 0.832-0.932), accuracy of 0.829, sensitivity of 0.798, and specificity of 0.848. Sensitivity analyses also demonstrated similar robustness after excluding patients with medication interruption (AUC, 0.871; 95% CI, 0.812-0.930) and those with poor adherence (AUC, 0.853; 95% CI, 0.790-0.916). Moreover, in the complete-case analysis of all 278 patients, the model maintained favorable overall performance, with an AUC of 0.881 (95% CI, 0.826-0.936), accuracy of 0.833, sensitivity of 0.806, and specificity of 0.849.

Comparative performance of the XGBoost and conventional logistic regression models in the validation cohort

In the validation cohort, the XGBoost model demonstrated consistently higher performance estimates than conventional logistic regression model across discrimination, classification, and calibration metrics. The XGBoost model achieved an AUC of 0.870 (95% CI, 0.799-0.941), compared with 0.823 (95% CI, 0.733-0.913) for logistic regression, although the difference was not statistically significant (P = 0.218). Similarly, XGBoost yielded higher accuracy (82.1% vs 77.4%; P = 0.449), sensitivity (77.4% vs 71.0%; P = 0.564), and specificity

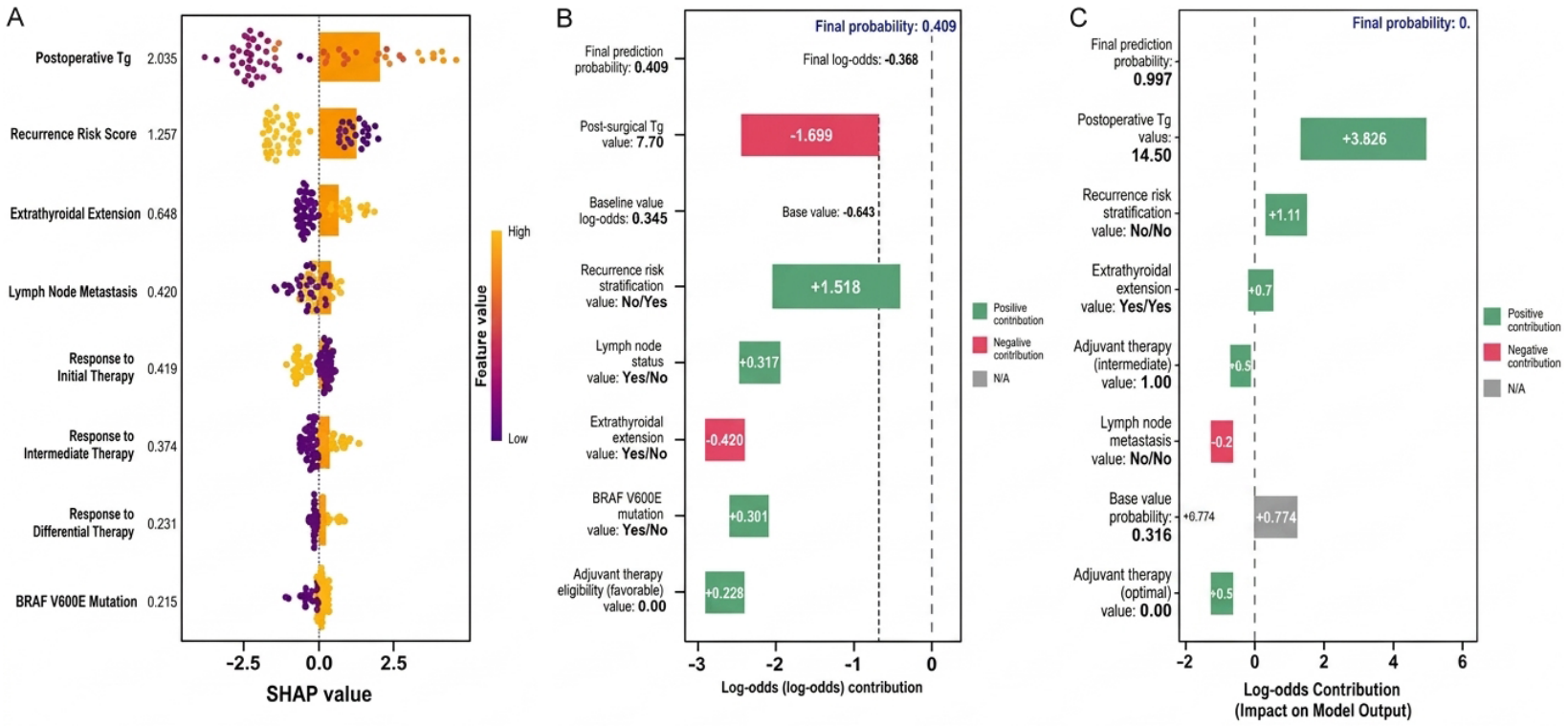


Figure 4. Development of the XGBoost prediction model for TSH target non-attainment after postoperative endocrine therapy in patients with DTC. A. SHAP summary plot. B. SHAP waterfall plot for a representative patient with low predicted probability of TSH target attainment. C. SHAP waterfall plot for a representative patient with high predicted probability of TSH target attainment.

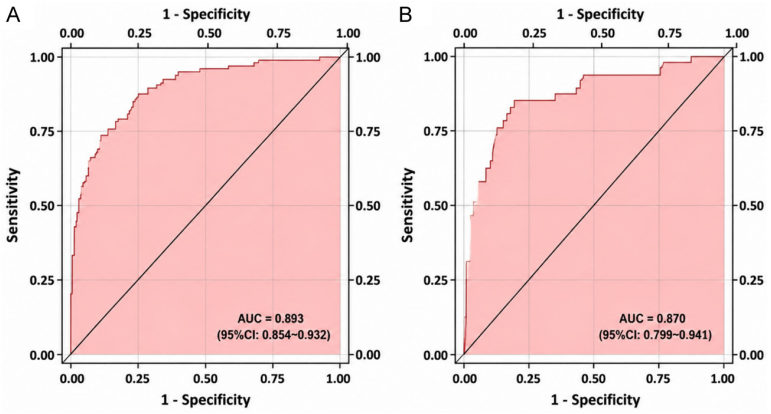


Figure 5. Receiver operating characteristic (ROC) curves of the XGBoost model in the training cohort (A) and validation cohort (B).

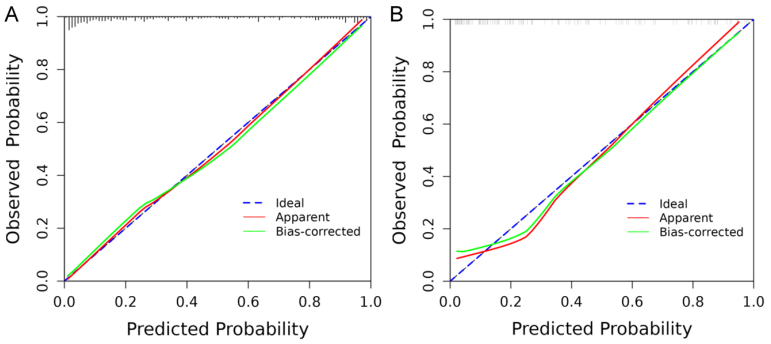


Figure 6. Calibration curves of the XGBoost model in the training cohort (A) and validation cohort (B).

(84.9% vs 81.1%; P = 0.599). Descriptively, XGBoost also achieved higher positive predictive value (75.0% vs 68.8%), negative predictive value (86.5% vs 82.7%), and F1 score (0.762 vs 0.698). Moreover, XGBoost model demonstrated a lower Brier score than logistic regression (0.149 vs 0.174), indicating improved probabilistic prediction performance, together with a calibration intercept closer to zero (-0.08 vs -0.14) and a calibration slope closer to 1 (0.93 vs 0.81) (**Table 4**). Collectively, these findings suggest that XGBoost provided modest incremental improvements over conventional logistic regression in the validation cohort, particularly in overall classification performance and calibration; however, the absence of statistically significant differences indicates that its added predictive value should be interpreted cautiously and requires confirmation in larger, independent cohorts.

Discussion

The current study identified a specific set of clinicopathological and treatment-related factors associated with postoperative TSH target non-attainment in patients with DTC and developed and internally validated an interpretable XGBoost model with high discrimination, calibration and clinical utility. Specifically, extrathyroidal invasion, lymph node metastasis, BRAF V600E mutation, elevated postoperative Tg levels, and intermediate/high recurrence risk stratification were independent risk factors for postoperative TSH target non-attainment, whereas suboptimal adherence to postoperative endocrine therapy represented a particularly strong adverse predictor.

Based on these findings, the XGBoost model demonstrated robust performance in both the training and validation cohorts and remained stable in subgroup and sensitivity analyses, highlighting its clinical utility in postoperative TSH management. Importantly, this study extends previous research by incorporating conventional clinicopathological risk factors with postoperative biochemical parameters, treatment-related behavioral factors, and machine-learning approaches.

Greater tumor burden and more aggressive biological characteristics were associated with an increased likelihood of post-operative TSH target non-attainment. This observation is broadly consistent with existing thyroid cancer risk stratification frameworks, in which extrathyroidal extension, lymph node involvement, and recurrence-risk stratification represent key determinants of postoperative management intensity and complexity of follow-up. Both the 2015 ATA and the updated 2025 ATA guidelines emphasize that the degree and duration of TSH

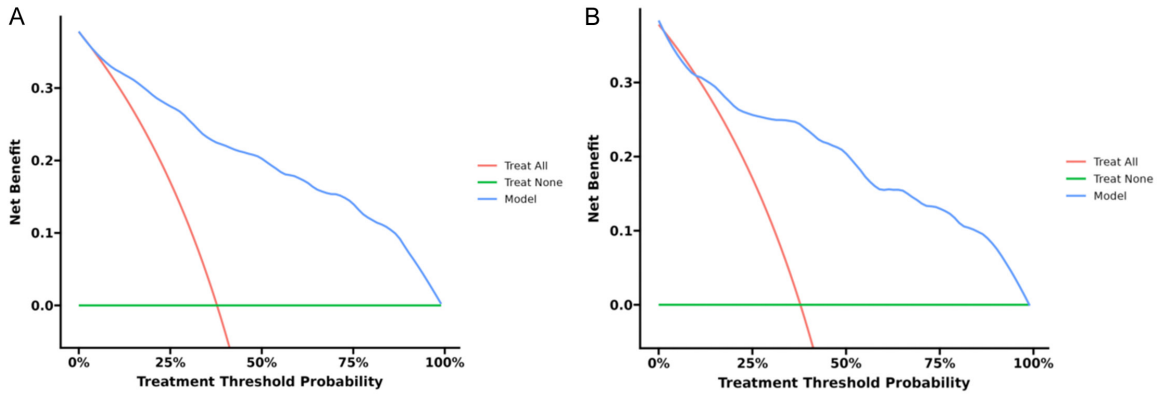


Figure 7. Decision curve analysis of the XGBoost model in the training cohort (A) and validation cohort (B).

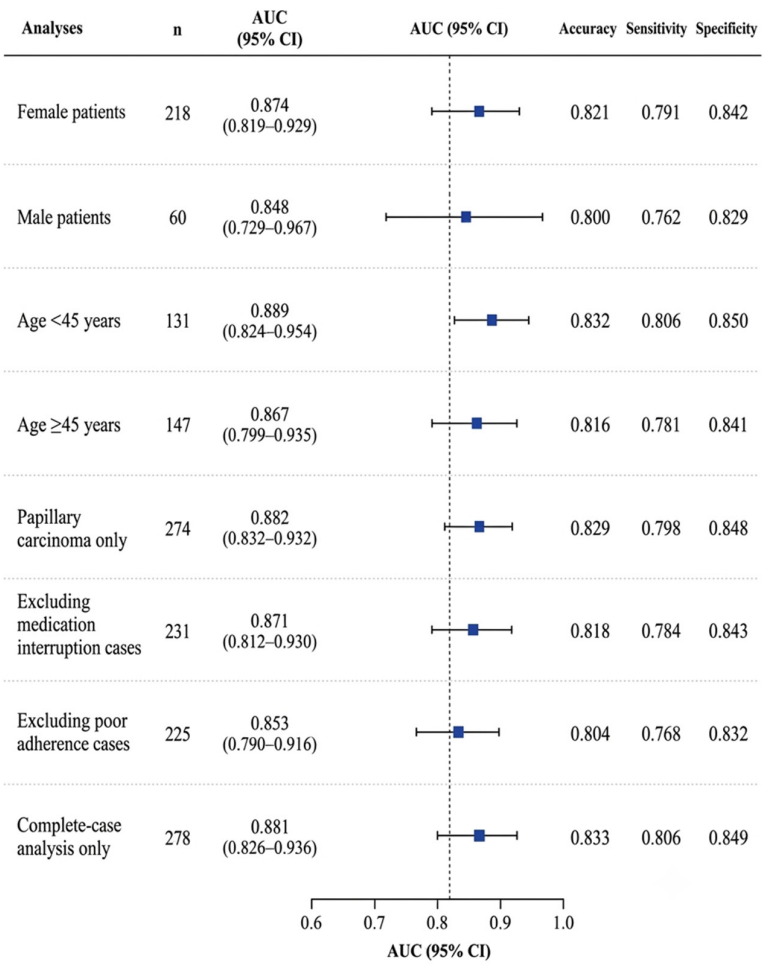


Figure 8. Subgroup and sensitivity analyses of the XGBoost model.

suppression should be tailored based on recurrence risk and response to therapy, rather than administered uniformly to all patients [16, 17]. One possible explanation is that patients with more progressive structural disease may require stricter TSH stages, more frequent dose adjustments, and closer biochemical surveillance, which inherently makes target achievement difficult to attain in practice. Clinically, failure to achieve TSH targets may not solely reflect inadequate treatment adherence but may also identify patients with a more complex disease-management profile who require closer monitoring and individualized reassessment of therapeutic strategies.

Although the prognostic role of BRAF V600E is context-dependent, substantial evidence has associated this mutation with more aggressive clinicopathological characteristics, increased recurrence risk, and altered tumor biology in papillary thyroid carcinoma [18-22]. The observed association in this study may therefore reflect the relationship between molecular aggressiveness and postoperative management complexity rather than a direct endocrine effect of the mutation itself. One possible explanation is that BRAF- positive tumors are more commonly accompanied by structural risk factors, persistent biochemical abnormalities, or more intensive clinical management strategies, all of

Table 4. Comparative performance of the XGBoost and conventional logistic regression models in the validation cohort

Performance metric	XGBoost model	Logistic regression model	*P* value
AUC (95% CI)	0.870 (0.799-0.941)	0.823 (0.733-0.913)	0.218
Accuracy, n/N (%)	69/84 (82.1)	65/84 (77.4)	0.449
Sensitivity, n/N (%)	24/31 (77.4)	22/31 (71.0)	0.564
Specificity, n/N (%)	45/53 (84.9)	43/53 (81.1)	0.599
Positive predictive value, %	75.0	68.8	-
Negative predictive value, %	86.5	82.7	-
F1 score	0.762	0.698	-
Brier score	0.149	0.174	-
Calibration intercept	-0.08	-0.14	-
Calibration slope	0.93	0.81	-

which can reduce the likelihood of achieving postoperative TSH targets. These findings suggest that molecular information may provide additional value for postoperative endocrine-management planning, particularly when it is interpreted in conjunction with conventional pathological risk factors.

Endocrine therapy adherence represents a particularly important predictor that should be given specific attention because it is a potentially modifiable determinant identified in the current study. Our results demonstrated a clear gradient association between adherence levels and postoperative TSH target attainment: both moderate and poor adherence were associated with a lower likelihood of postoperative TSH target attainment, the poor adherence showing the strongest association. These findings are consistent with emerging evidence suggesting that effective TSH suppression is highly dependent on patient knowledge and adherence to postoperative management recommendations [23, 24]. The underlying mechanism is clinically intuitive: irregular levothyroxine intake, inappropriate administration timing (such as taking medication close to meals or interacting medications), dose interruption, and insufficient follow-up compliance may compromise TSH control even with a correctly selected therapeutic strategy.

The machine-learning component of this study also provides important methodological implications. Compared with conventional regression model, XGBoost is capable of capturing nonlinear relationships and high-order interactions, which may be particularly relevant in postoperative endocrine management, where tumor burden, biochemical markers, recurrence risk, and treatment adherence may jointly influence the likelihood of achieving TSH targets. Previous studies in thyroid cancer have demonstrated that machine-learning methods, including XGBoost, may provide strong predictive performance for outcomes such as recurrence and may outperform conventional statistical methods when complex clinical relationships exist. In our study, SHAP summary and waterfall plots further enhanced model interpretability by illustrating both population-level feature importance and patient-level predictor contributions. The practical implication is that prediction tools for postoperative TSH target non-attainment may move beyond purely black-box performance measures toward transparent and personalized approaches for postoperative endocrine management.

The validation analyses further demonstrated that the proposed model showed not only statistical performance but also potential clinical relevance. The model achieved good discrimination in both training and validation cohorts, satisfactory calibration, and favorable net clinical benefit. In addition, its performance remained consistent across subgroup analyses stratified by sex, age, pathological characteristics, and adherence status. This consistency is promising because accurate calibration and decision-analytic performance are particularly important in endocrine follow-up. Both overestimation and underestimation of risk can cause inappropriate management strategies, including excessive TSH suppression, avoidable treatment-related adverse

effects, or inadequate surveillance intensity. These findings are in line with the evolving principles of ATA guidance, which increasingly emphasize risk-adapted and response-adapted management rather than uniform TSH suppression plans for all patients. In this context, an interpretable model capable of identifying patients at high risk of TSH target non-attainment may assist clinicians in optimizing follow-up intensity, implementing adherence-focused interventions, and adjusting levothyroxine management at an earlier stage after surgery.

Several limitations of this work should be acknowledged. First, the retrospective single-center design may have introduced selection bias and center-specific treatment patterns, which may limit the external generalizability of our findings. Second, although internal validation was performed, the lack of an independent external validation cohort prevents definitive assessment of model transportability across different institutions, patient populations, and follow-up settings. Therefore, prospective multicenter studies incorporating independent external validation cohorts are still required before broader clinical implementation of the model. Third, the sample size was moderate, and several potentially relevant determinants of TSH control, including detailed levothyroxine dosing histories, drug interactions, dietary interference, patient education, and variations in physician management strategies, were not incorporated into the model. Fourth, because this study focused on postoperative TSH target attainment rather than long-term oncological outcomes, the association between difficulty achieving endocrine control and subsequent recurrence or survival outcomes remains uncertain. Finally, although SHAP analysis improved model interpretability, the proposed model should be considered a decision-support tool rather than a replacement for guideline-based clinical judgment.

Conclusion

This study demonstrates that aggressive tumor characteristics and inadequate adherence to postoperative endocrine therapy are strongly associated with postoperative TSH target non-attainment in patients with DTC. The interpretable XGBoost model developed in this study, incorporating these variables, demonstrates favorable discrimination, calibration, and clinical utility, with stable performance across subgroup and sensitivity analyses.

Acknowledgements

This work was supported by 2023 Jiangsu Pharmaceutical Association Shiyao Hospital Pharmaceutical Research Fund (SY202304-1).

Disclosure of conflict of interest

None.

Address correspondence to: Li Shen, Department of Thyroid and Breast Surgery, The Fourth Affiliated Hospital of Nanjing Medical University, No. 298 Nanpu Road, Jiangbei New District, Nanjing 210031, Jiangsu, China. Tel: +86-025-56680746; E-mail: zhl49737542@163.com

References

[1] Moleti M, Aversa T, Crisafulli S, Trifirò G, Corica D, Pepe G, Cannavò L, Di Mauro M, Paola G, Fontana A, Calapai F, Cannavò S and Wasniewska M. Global incidence and prevalence of differentiated thyroid cancer in childhood: systematic review and meta-analysis. Front Endocrinol (Lausanne) 2023; 14: 1270518.

[2] Coca-Pelaz A, Rodrigo JP, Shah JP, Nixon IJ, Hartl DM, Robbins KT, Kowalski LP, Mäkitie AA, Hamoir M, López F, Saba NF, Nuyts S, Rinaldo A and Ferlito A. Recurrent differentiated thyroid cancer: the current treatment options. Cancers (Basel) 2023; 15: 2692.

[3] Silaghi H, Lozovanu V, Georgescu CE, Pop C, Nasui BA, Cătoi AF and Silaghi CA. State of the art in the current management and future directions of targeted therapy for differentiated thyroid cancer. Int J Mol Sci 2022; 23: 3470.

[4] Gimblet GR, Whitt J, Houson HA, Lin D, Guenter R, Rao TC, Wang D, Ness J, Gonzalez ML, Murphy MS, Gillis A, Chen H, Copland JA, Kenderian SS, Lloyd RV, Szkudlinski MW, Lapi SE and Jaskula-Sztul R. Thyroid-stimulating hormone receptor (TSHR) as a target for imaging differentiated thyroid cancer. Surgery 2024; 175: 199-206.

[5] Ryu HJ, Choi MS, Park H, Kim TH, Chung JH, Park SY and Kim SW. Adequate dose of levothyroxine for thyroid-stimulating hormone suppression after total thyroidectomy in patients with differentiated thyroid cancer. Endocrinol Metab (Seoul) 2024; 39: 615-621.

[6] Lee JC, Song BS, Kang YM, Kim YR, Kang YE, Lee JH, Shong M and Yi HS. Effect of thyroid-stimulating hormone suppression on muscle function after total thyroidectomy in patients

with thyroid cancer. Front Endocrinol (Lausanne) 2021; 12: 769074.
[7] Pitoia F, Bueno F, Urciuoli C, Abelleira E, Cross G and Tuttle RM. Outcomes of patients with differentiated thyroid cancer risk-stratified according to the American thyroid association and Latin American thyroid society risk of recurrence classification systems. Thyroid 2013; 23: 1401-7.
[8] Song X, Zhi X and Qian L. Tailoring TSH suppression in differentiated thyroid carcinoma: evidence, controversies, and future directions. Endocrine 2025; 89: 1-19.
[9] Lee YK, Lee H, Han S, Jung H, Shin DY, Nam KH, Chung WY and Lee EJ. Association between thyroid-stimulating hormone level after total thyroidectomy and hypercholesterolemia in female patients with differentiated thyroid cancer: a retrospective study. J Clin Med 2019; 8: 1106.
[10] Biondi B. TSH suppression in differentiated thyroid cancer patients. still more questions than answers after 30 years. Thyroid 2024; 34: 671-673.
[11] Gubbi S, Al-Jundi M, Foerster P, Cardenas S, Butera G, Auh S, Wright EC and Klubo-Gwiezdzinska J. The effect of thyrotropin suppression on survival outcomes in patients with differentiated thyroid cancer: a systematic review and meta-analysis. Thyroid 2024; 34: 674-686.
[12] Stewart LA and Kuo JH. Advancements in the treatment of differentiated thyroid cancer. Ther Adv Endocrinol Metab 2021; 12: 20420188211000251.
[13] Schindele A, Krebold A, Heiß U, Nimptsch K, Pfaehler E, Berr C, Bundschuh RA, Wendler T, Kertels O, Tran-Gia J, Pfob CH and Lapa C. Interpretable machine learning for thyroid cancer recurrence predicton: leveraging XGBoost and SHAP analysis. Eur J Radiol 2025; 186: 112049.
[14] Redlich A, Pfaehler E, Kunstreich M, Schmutz M, Lapa C and Kuhlen M. Machine learning prediction of recurrence in pediatric thyroid cancer: malignant endocrine tumors cohort analysis using XGBoost and SHAP. J Clin Endocrinol Metab 2026; 111: e844-e852.
[15] Su Y, Cai Y, Jin S, Ye X, Jeong J, Yuan Y and Yi H. Explainable multi-modal machine learning for predicting occult pulmonary metastases in differentiated thyroid cancer: a SHAP-based approach prior to radioactive iodine scans. Front Med Technol 2025; 7: 1685088.
[16] Uludag M, Caliskan O, Unlu MT, Cetinoglu I and Aygun N. What has changed in the 2025 American thyroid association management guidelines for adult patients with differentiated thyroid cancer? Part 3: long-term surveillance, advanced and novel treatments. Sisli Etfal Hastan Tip Bul 2025; 59: 284-297.
[17] Petranović Ovčariček P, Campennì A, D'Aurizio F, Imperiali M, Alibrandi A, Ruggeri RM, Bonanno L and Giovanella L. When guidelines meet reality: the combined impact of assay variability and prescribing differences on TSH management in thyroid cancer. Cancers (Basel) 2025; 17: 3912.
[18] Fernández Velasco P, Peciña Melgosa P, Torres Torres B, Torres Morientes LM, Fernández Rodríguez A, Alonso Mesonero M, de Uribe Viloria M, Álvarez Quiñones M, Santos Pérez J, de Luis Román D and Diaz-Soto G. Comparative evaluation of dynamic risk stratification according to ATA 2015 and ATA 2025 in low-risk differentiated thyroid cancer without radioiodine ablation. Endocrine 2026; 91: 77.
[19] Tabriz N, Grone J, Uslar V, Tannapfel A and Weyhe D. BRAF V600E mutation correlates with aggressive clinico-pathological features but does not influence tumor recurrence in papillary thyroid carcinoma-10-year single-center results. Gland Surg 2020; 9: 1902-1913.
[20] Abdulhaleem M, Bandargal S, Pusztaszeri MP, Rajab M, Greenspoon H, Krasner JR, Da Silva SD, Forest VI and Payne RJ. The impact of *BRAF V600E* mutation allele frequency on the histopathological characteristics of thyroid cancer. Cancers (Basel) 2023; 16: 113.
[21] Huang K, Gao N, Bian D, Zhai Q, Yang P and Zhang Y. Associations of BRAF V600E, clinical pathology and imaging factors with the recurrence rate of papillary thyroid microcarcinoma. Exp Ther Med 2020; 20: 243.
[22] Tallini G, de Biase D, Durante C, Acquaviva G, Bisceglia M, Bruno R, Bacchi Reggiani ML, Casadei GP, Costante G, Cremonini N, Lamartina L, Meringolo D, Nardi F, Pession A, Rhoden KJ, Ronga G, Torlontano M, Verrienti A, Visani M and Filetti S. BRAF V600E and risk stratification of thyroid microcarcinoma: a multicenter pathological and clinical study. Mod Pathol 2015; 28: 1343-59.
[23] Wu Y, Zhang Z, Qin W, Chen W and Xu T. Knowledge, attitudes, and practices regarding the postoperative management and TSH suppression therapy among patients with thyroid cancer. Front Oncol 2025; 15: 1441726.
[24] Praw SS and Brent GA. Approach to the patient with a suppressed TSH. J Clin Endocrinol Metab 2023; 108: 472-482.

Supplementary Table 1. Assignment method of independent variables

Variable	Coding Method
Tumor diameter	Continuous variable, standardized using Z-score
Number of lesions	Solitary = 0, Multiple = 1
Extrathyroidal extension	No = 0, Yes = 1
Vascular invasion	No = 0, Yes = 1
Lymph node metastasis	No = 0, Yes = 1
BRAF V600E mutation	No = 0, Yes = 1
Postoperative Tg	Continuous variable, standardized using Z-score
Risk stratification for adverse effects of TSH suppression therapy	Low risk = 0, Intermediate/high risk = 1
Recurrence risk stratification	Low risk = 0, Intermediate/high risk = 1
Adherence to endocrine therapy	Good = 1, Moderate = 2, Poor = 3