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 Diagnostic and Prognostic Prediction Model for Follicular Thyroid Carcinoma Based on Multicenter Data

Ziyi Cao
The First Affiliated Hospital of Zhengzhou University
Tianlong Gao
The First Affiliated Hospital of Zhengzhou University
Pu Xi
Henan Cancer Hospital
Yihao Liu
The First Affiliated Hospital of Zhengzhou University
Yu Chen
The First Affiliated Hospital of Zhengzhou University
Qungang Chang
The First Affiliated Hospital of Zhengzhou University
Qian Zhao
The First Affiliated Hospital of Zhengzhou University
Detao Yin
detaoyin@zzu.edu.cn
The First Affiliated Hospital of Zhengzhou University

Research Article

Keywords: Follicular thyroid carcinoma, Follicular thyroid adenoma, multicenter study, Diagnostic predictive model, Prognostic model, Progression-free survival

Posted Date: September 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10344321/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 Diagnostic and Prognostic Prediction Model for Follicular Thyroid Carcinoma Based on Multicenter Data

Ziyi Cao[1], Tianlong Gao[1], Pu Xi[2], Yihao Liu[1], Yu Chen[1], Qungang Chang[1], Qian Zhao[1], Detao Yin[1]*

[1] Department of Thyroid Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, 450052 China

[2] Henan Cancer Hospital, Zhengzhou, 450008 China

*Corresponding author: Detao Yin

Email: detaoyin@zzu.edu.cn

Postal address: The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China

Telephone: 13693717963

Abstract

Background

Follicular thyroid carcinoma (FTC) is often indistinguishable from benign follicular thyroid adenoma (FTA) on preoperative fine-needle aspiration, frequently leading to unnecessary diagnostic surgery. Moreover, reliable tools for postoperative risk stratification in FTC remain limited. We therefore developed and validated a preoperative diagnostic nomogram and a postoperative progression-free survival (PFS) prognostic model for FTC using multicenter data.

Methods

This retrospective study included cases of thyroid follicular tumors from the First Affiliated Hospital of Zhengzhou University (2012-2024) and the Henan Cancer Hospital (2018-2025), which served as the training and validation centers, respectively. Using postoperative pathology as the gold standard, multivariate logistic regression was employed to identify independent diagnostic predictors and construct a diagnostic prediction model; for FTC patients, multivariate Cox regression was used following follow-up to identify independent risk factors for PFS and construct a prognostic model. The diagnostic cohort at the training center comprised 299 cases (197 FTC, 102 FTA), while the diagnostic cohort at the validation center comprised 273 cases (120 FTC, 153 FTA); the prognostic development cohort comprised 161 FTC patients, and the prognostic validation cohort comprised 87 FTC patients.

Results

Multivariate logistic regression analysis revealed that margin (OR=12.821, 95% CI: 5.009-32.815, P<0.001), calcification (OR=4.220, 95% CI: 2.061-8.639, P<0.001), coexisting

Hashimoto's thyroiditis (OR=7.786, 95% CI: 2.921-20.751, P < 0.001), echogenicity (OR = 4.244, 95% CI: 1.740-10.349, P = 0.001), echotexture (OR = 3.086, 95% CI: 1.465-6.500, P=0.003), sex (OR=2.441, 95% CI: 1.176-5.069, P = 0.017) and shape (OR = 2.491, 95% CI: 1.150-5.400, P = 0.021) were identified as independent predictors of FTC, and a preoperative diagnostic prediction model was constructed based on these findings. Multivariate Cox regression analysis revealed that pathological subtype (HR = 11.704, 95% CI: 4.851-28.236, P < 0.001) and M stage (HR = 2.879, 95% CI: 1.143-7.239, P = 0.025) were identified as independent adverse prognostic factors for PFS in FTC patients, and a postoperative prognostic prediction model was constructed accordingly.

Conclusions

FTC prediction and prognostic models developed using multicenter data can provide guidance for preoperative risk assessment, postoperative risk stratification, and personalized management.

Keywords: Follicular thyroid carcinoma; Follicular thyroid adenoma; multicenter study; Diagnostic predictive model; Prognostic model; Progression-free survival

1 Introduction

Follicular thyroid carcinoma (FTC) is the second most common thyroid malignancy and accounts for 10-15% of all malignant thyroid tumors(1). Unlike papillary thyroid carcinoma, FTC is more likely to metastasize via the bloodstream; some patients already present with bone or lung metastases at the time of diagnosis, resulting in greater invasiveness and a relatively poorer prognosis(2). Follicular thyroid adenoma (FTA), on the other hand, is a common benign follicular neoplasm. However, as FTC and FTA are highly similar in their

cytological presentation, it is often difficult to accurately distinguish between benign and malignant tumors based solely on preoperative fine-needle aspiration biopsy(3).

Currently, preoperative assessment of thyroid nodules relies primarily on ultrasound, fine-needle aspiration cytology and necessary ancillary investigations. Ultrasound provides information on nodule size, shape, margin, echogenicity, calcification and blood flow, and is used in conjunction with TI-RADS for risk stratification. However, the diagnosis of FTC still relies primarily on postoperative pathological evidence of capsular and/or vascular invasion, and these features cannot typically be directly identified preoperatively through routine imaging or cytology(4). Consequently, many follicular nodules still require diagnostic surgery to confirm the pathology, even after preoperative assessment, with a significant proportion ultimately proving to be benign(5). This not only increases the burden of unnecessary surgery but may also affect initial surgical decisions and subsequent management due to inadequate preoperative risk assessment. Although FTC lacks specific preoperative features, existing studies suggest that it may differ from FTA in terms of margin, shape, echogenicity, calcification and internal structure(6). However, the diagnostic value of a single finding is limited; it is therefore necessary to integrate multiple ultrasound and clinical findings to establish a reliable preoperative predictive model, thereby improving the diagnostic accuracy for FTC.

In addition to preoperative diagnosis, prognostic assessment of FTC is equally important. Previous studies have shown that factors such as age, tumor size, capsular invasion, vascular invasion and distant metastasis may be associated with poor outcomes in FTC(7). However, such information is often only confirmed through postoperative pathology or

long-term follow-up, which limits early risk stratification and personalized management. Therefore, the development of reliable prognostic prediction models is of significant clinical importance for identifying high-risk patients and optimizing follow-up strategies.

Accordingly, this study used postoperative pathology as the gold standard to integrate preoperative ultrasound findings and multidimensional preoperative examination data in order to construct and validate a preoperative diagnostic prediction model for FTC; simultaneously, it analyses factors associated with the prognosis of FTC and establishes a prognostic prediction model, with the aim of providing a quantitative basis for the precise diagnosis and treatment of FTC and for risk stratification management.

2 Materials and Methods

2.1 Study Design and Data Sources

This study is a multicenter retrospective cohort study. Case data were collected on patients with follicular thyroid neoplasms(FTNs) who underwent surgical treatment and received a definitive pathological diagnosis at the First Affiliated Hospital of Zhengzhou University between January 2012 and October 2024, and at the Henan Cancer Hospital between January 2018 and October 2025. The study employed a center-based development-external validation design: The First Affiliated Hospital of Zhengzhou University served as the training center for model development, while the Henan Cancer Hospital acted as the validation center for external validation. The diagnostic model included cases from both centers that were confirmed as FTC or FTA by postoperative pathology; the prognostic model was subsequently established within the FTC patient cohort. This study was approved by the Ethics Committee of the First Affiliated Hospital of Zhengzhou University and Henan Cancer Hospital (No.

2020-KY-0075-002).

Inclusion criteria: (1) Undergoing thyroid surgery for the first time and with a complete surgical record; (2) Final postoperative pathological diagnosis of FTC or FTA; (3) In cases of multiple lesions, only the primary lesion with the most pronounced malignant features will be included; if malignant features are similar, the largest lesion will be selected; (4) A routine ultrasound examination must have been performed prior to surgery, and the report must contain the key ultrasound features required for the study; (5) For cases used in prognostic analysis, data must be available to determine the outcome status and duration of follow-up.

Exclusion criteria: (1) Patients who have not undergone surgery or for whom this is not the first surgery; (2) Patients with a history of radiotherapy to the neck; (3) Patients with concurrent thyroid cancer of another type or other malignant tumors; (4) Poor-quality imaging data where key ultrasound findings cannot be interpreted or where there are significant inconsistencies in the records; (5) Patients with missing key variables that cannot be used for modeling.

Clinical data were primarily obtained from electronic medical record systems, preoperative examination reports, surgical and pathology records, and follow-up records. The data collected included demographic information: age, sex, smoking history, etc; laboratory parameters: free triiodothyronine (FT3), free thyroxine (FT4), thyroid-stimulating hormone (TSH), thyroglobulin (Tg), and thyroglobulin antibodies (TgAb), etc; clinical comorbidities: Hashimoto's thyroiditis and hyperthyroidism; preoperative ultrasound findings; surgery-related information; and postoperative pathological information, etc.

2.2 Diagnostic Prediction Model

2.2.1 Research Subjects and Variable Definitions

In the FTC diagnostic prediction model, cases from the First Affiliated Hospital of Zhengzhou University were used as the development cohort, with postoperative pathology serving as the diagnostic gold standard; a total of 299 cases were included, comprising 197 cases of FTC and 102 cases of FTA. Cases from the Henan Cancer Hospital were used as the external validation cohort, comprising a total of 273 cases, including 120 cases of FTC and 153 cases of FTA.

The candidate predictors are primarily derived from information available prior to surgery, including demographic data, laboratory parameters, and ultrasound imaging characteristics. To enhance the model's robustness and clinical interpretability, some of the original multi-class variables were recoded into binary categories based on their clinical significance. Specifically, these included: margin (poorly defined/ill defined vs fairly defined/clearly defined), echogenicity (very hypoechoic/hypoechoic vs isoechoic/hyperechoic), shape (abnormal/irregular/slightly irregular vs fairly regular/regular/oval), echotexture (heterogeneous/poorly homogeneous vs fairly homogeneous/homogeneous), blood flow signal grade (Grade 2-3 vs Grade 0-1), composition (solid vs cystic/cystic-solid/mixed), TI-RADS classification (Grade 4-5 vs Grade 2-3) and multifocality ($\geqslant$2 vs 1), among others.

2.2.2 Model Construction

At the training center, univariate logistic regression analyses were first performed on each candidate variable to identify factors associated with the occurrence of FTC. Subsequently, variables found to be statistically significant in the univariate analyses were incorporated into

a multivariate logistic regression model, taking clinical plausibility into account, and variable selection was carried out using the backward stepwise method, ultimately resulting in the establishment of a preoperative diagnostic prediction model for FTC.

Given that the cases in the training center span a considerable period of time, this study further conducted tests for temporal covariate analysis to assess whether the model was influenced by changes in the enrolment period. Specifically, the years of surgery at the training center were stratified into time periods (2012-2015, 2016-2019 and 2020-2024), and the time stratification variable was forced into the multivariate analysis to test for interactions between time stratification and key predictors, thereby assessing whether there was any significant temporal drift in the model coefficients.

2.2.3 Model Validation and Evaluation Metrics

Model performance was evaluated across three dimensions: discriminatory power, calibration, and clinical utility. Discriminatory power was assessed using the receiver operating characteristic curve (ROC) and the area under the curve (AUC); calibration was evaluated using calibration curves to assess the consistency between the model's predicted probabilities and actual outcomes; and clinical utility was assessed using decision curve analysis (DCA) to evaluate the net benefit at different threshold probabilities. To enhance the clinical interpretability of the results, a fixed threshold was further determined in the development cohort, and confusion matrices were constructed in the development and external validation cohorts to calculate metrics such as sensitivity(Se), specificity(Sp), positive predictive value(PPV), negative predictive value(NPV), positive likelihood ratio(LR+), negative likelihood ratio(LR-) and accuracy(ACC).

2.3 Prognostic Model

2.3.1 Cohort Definition and Follow-up

The study subjects of the prognostic model were FTC patients. All FTC cases in the training center were screened according to inclusion and exclusion criteria, with progression-free survival (PFS) as the primary outcome. After excluding cases with incomplete follow-up data, unclear PFS status, or unavailable follow-up duration, a final cohort of 161 FTC patients was established for the development of a prognostic model, including 21 cases with PFS events. FTC cases from the validation center were screened for follow-up using the same criteria, with further exclusion of cases after October 2024 to ensure all patients had at least 1 year of follow-up. This ultimately formed the external validation cohort for prognosis, comprising 87 FTC patients, including 11 PFS events.

Follow-up commenced on the date of surgery, with data collected through outpatient reviews and telephone follow-ups, and other methods, focusing on the recording of local recurrence, lymph node recurrence, distant metastasis and mortality. The primary endpoint was PFS, defined as the time from surgery to the first occurrence of death, local recurrence, lymph node recurrence, or distant metastatic progression/new metastasis; if a patient experiences multiple events simultaneously, the earliest occurrence is taken as the time of the outcome event. The follow-up cut-off date is set as 1 October 2025.

2.3.2 Prediction Time Point Selection

Given the clinical management needs regarding the risk of recurrence following FTC surgery and the available follow-up duration in the external validation cohort of this study, 3-year and 5-year PFS were predefined as the primary prediction time points for model development and

external validation. Given the long time span of cases at the training centers and the availability of extended follow-up data, 10-year PFS was included as an exploratory long-term prediction endpoint. It was analyzed only in the development cohort and was not designated as a primary external validation endpoint for this study.

2.3.3 Variable Screening and Model Building

Candidate variables for the prognostic model include patient demographics, tumor pathological characteristics, staging information, treatment modalities, and postoperative follow-up data. The original multiclass variables were recoded into binary categories based on their clinical significance, specifically: pathological subtype (extensive infiltration vs. encapsulated vascular infiltration/microscopic infiltration); and T-stage (1-2 vs. 3-4).

In the development cohort, univariate Cox proportional hazards regression was first used to identify candidate variables associated with PFS. These variables were then incorporated into a multivariate Cox regression model using a backward stepwise approach, taking into account clinical significance and the limited number of events, to construct a prognostic model for PFS following FTC surgery. To test the proportional hazards assumption and assess potential time-dependence, this study conducted temporal covariate analysis on the candidate variables to evaluate whether their effects changed over the course of follow-up. Due to the limited number of events in the development cohort, the number of variables in the multivariate model was kept to a minimum to reduce the risk of overfitting. After model development, individual linear predictors were calculated based on the regression coefficients, and these were used to generate risk stratification indicators or visual prediction tools.

2.3.4 Temporal Sensitivity Analysis

To assess the robustness of the prognostic model across different clinical periods, this study conducted a temporal sensitivity analysis. Specifically, cases from the training center were divided into Training Subset 1 (2012-2017) and Training Subset 2 (2018-2024) based on the year of surgery. The modeling process was repeated for each subset, and the performance of the models from each period was compared using the same external validation cohort.

2.3.5 Model Validation and Evaluation Metrics

The validation of the prognostic model was conducted in a cohort from Henan Cancer Hospital. Model discriminative power was assessed using the C-index and the time-dependent area under the ROC curve (AUC); model calibration was evaluated using calibration curves at different time points to assess the consistency between predicted and observed probabilities; and clinical utility was assessed using DCA to evaluate the net benefit of the model at different threshold probabilities. Patients were classified into high- and low-risk groups based on linear predicted values or preset cutoffs. Survival curves were plotted using the Kaplan-Meier method, and differences between groups were compared using the log-rank test.

2.4 Statistical Methods

All statistical analyses were performed using R software(4.5.2). Categorical variables are presented as frequencies (percentages), while continuous variables are presented as mean ± standard deviation or median [interquartile range], depending on their distribution. For comparisons between groups, categorical variables were analyzed using the chi-square test or Fisher's exact test; continuous variables were analyzed using the independent samples t-test or the Mann-Whitney U test. All statistical tests were two-sided, and a P-value of <0.05 was considered statistically significant.

3. Results

3.1 Baseline Characteristics

This study included a total of 572 patients, with the development cohort consisting of 299 cases from the First Affiliated Hospital of Zhengzhou University (197 FTC cases and 102 FTA cases), and the external validation cohort comprising 273 cases from Henan Cancer Hospital (120 FTC cases and 153 FTA cases). The screening process of study subjects and the cohort allocation for diagnostic prediction models and prognostic models are shown in Figure 1.

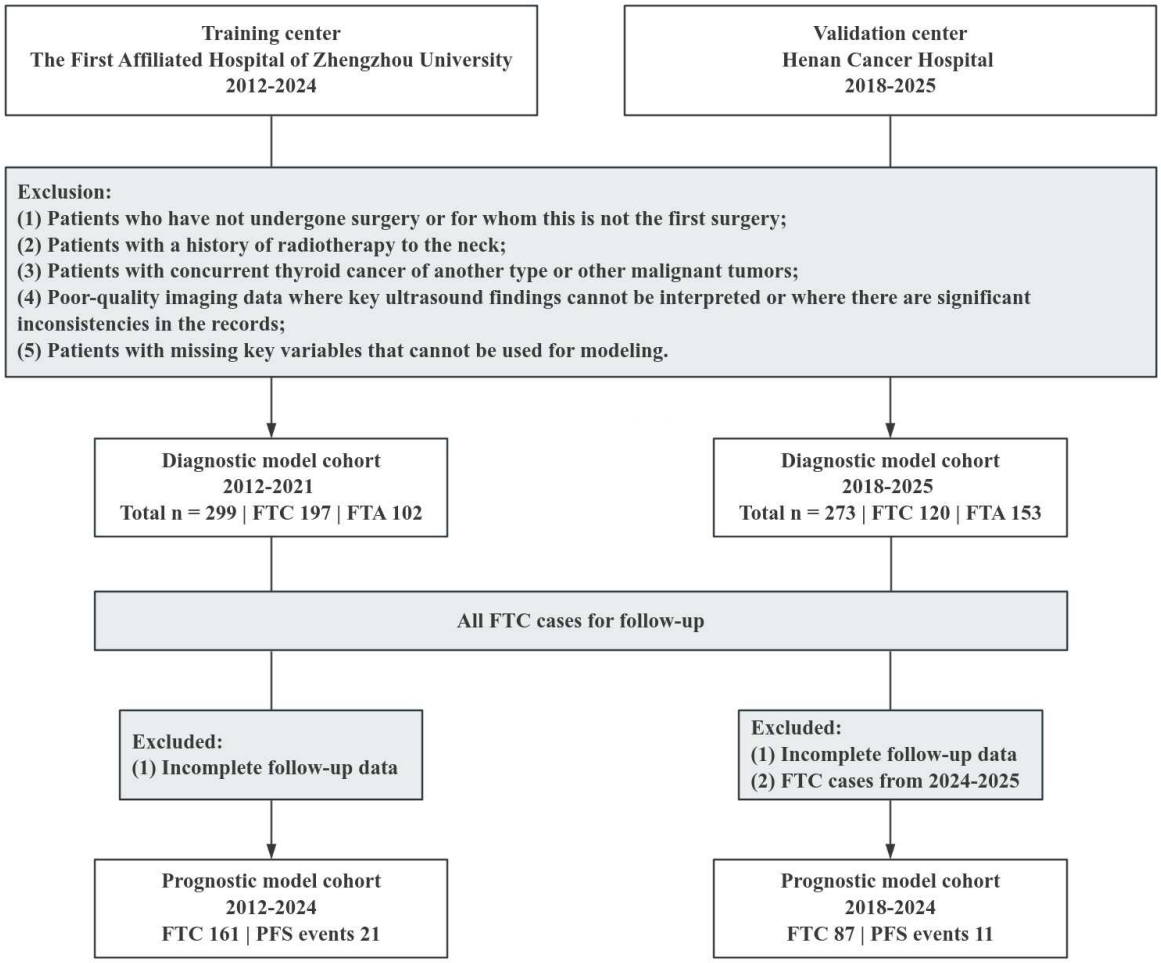


Figure 1 Flowchart

In the development cohort of the diagnostic prediction model, the FTC group had

significantly higher smoking rates, Tg, and TgAb levels than the FTA group, and also exhibited a greater tendency toward adverse features in terms of margin, shape, echogenicity, echotexture, calcification, the convex capsule sign, and coexisting Hashimoto's thyroiditis (all $P < 0.05$); however, there were no statistically significant differences between the two groups in terms of age, FT3, FT4, TSH, nodule size, blood flow grade, composition, TI-RADS classification, multifocality, and combined hyperthyroidism (all $P > 0.05$). Additionally, the proportion of females was lower in the FTC group than in the FTA group ($P = 0.021$).

In the external validation cohort, Tg levels in the FTC group were significantly higher than those in the FTA group; additionally, the FTC group exhibited a higher prevalence of adverse features, including margin, shape, echogenicity, calcification, convex capsule sign, composition, and coexisting Hashimoto's thyroiditis (all $P < 0.05$); in contrast, there were no statistically significant differences between the two groups in terms of age, sex, smoking history, FT3, FT4, TSH, TgAb, nodule size, echotexture, blood flow grade, TI-RADS classification, multifocality, or the presence of combined hyperthyroidism (all $P > 0.05$).

Overall, the baseline characteristics of the cases from the two centers show good consistency and are generally comparable, though some heterogeneity is present. This diversity enhances the external validity of the study's findings and supports the robustness of subsequent analyses (Table 1).

A total of 21 PFS events (13.0%) occurred in the development cohort of prognostic model, comprising 11 deaths, 5 local recurrences, 4 lymph node recurrences, and 7 cases of distant metastatic progression/new metastasis; in the external validation cohort, a total of 11 PFS events (12.6%) occurred, comprising 3 deaths, 1 local recurrence, 7 lymph node

recurrences, and 4 cases of distant metastatic progression/new metastasis. As some patients may experience overlapping events, the sum of the numbers of events across each category exceeds the total number of PFS events.

Comparison between the two cohorts revealed no statistically significant differences in variables such as age, sex, symptom, smoking history, maximum tumor diameter, multifocality, lymph node metastasis, capsular invasion, vascular invasion, T/N/M stage, combined hyperthyroidism, extraglandular invasion, and Hürthle cell subtype (all $P > 0.05$). However, there were significant differences between the two cohorts regarding the receipt of I-131 therapy, the prevalence of coexisting Hashimoto's thyroiditis, and the distribution of pathological subtype (all $P < 0.001$). Notably, the external validation cohort had a lower proportion of patients receiving I-131 therapy, a higher prevalence of coexisting Hashimoto's thyroiditis, and a significantly higher proportion of extensively infiltrating tumors compared with the development cohort. This suggests that, while the two cohorts are generally comparable, clinical and pathological heterogeneity persists, indicating good value for independent external validation (Table 2).

3.2 Diagnostic Prediction Model

3.2.1 Predictor screening

In the development cohort, univariate logistic regression analysis revealed that margin, calcification, coexisting Hashimoto's thyroiditis, echogenicity, echotexture, sex, shape, smoking, Tg levels, and the convex capsule sign were significantly associated with the occurrence of FTC (all $P < 0.05$); the remaining variables were not statistically significant. Variables found to be statistically significant in the univariate analysis were further included

in a multivariate logistic regression model, and variables were selected using a backward stepwise method. The results showed that margin (OR = 12.821, 95% CI: 5.009-32.815, P < 0.001), calcification (OR = 4.220, 95% CI: 2.061-8.639, P < 0.001), coexisting Hashimoto's thyroiditis (OR = 7.786, 95% CI: 2.921-20.751, P < 0.001), and echogenicity (OR = 4.244, 95% CI: 1.740-10.349, P=0.001), echotexture (OR=3.086, 95% CI: 1.465-6.500, P=0.003), sex (OR=2.441, 95% CI: 1.176-5.069, P = 0.017), and shape (OR = 2.491, 95% CI: 1.150-5.400, P = 0.021) were identified as independent predictors of FTC and were included in the subsequent diagnostic prediction model. Results of the time covariate test showed that the time period variable was not statistically significant in univariate analysis (P = 0.329) and was not retained after forced inclusion in the multivariate model, suggesting that differences in the time range of case collection did not significantly affect the main effects of the model (Table 3).

3.2.2 Model Development and External Validation

A preoperative diagnostic prediction model for FTC was developed based on the above seven independent predictors, and a nomogram was constructed (Figure 2).

Analysis of the model's performance showed that it demonstrated good discriminatory ability in the development cohort, with an AUC of 0.889 (95% CI: 0.852-0.927); after applying the optimal cut-off value of 0.362 determined from the development cohort, the AUC in the external validation cohort was 0.808 (95% CI: 0.757-0.860), indicating that the model retains good discriminatory performance in an external independent population (Figure 3). Calibration analysis showed Brier scores of 0.128 and 0.190 for the development and external validation cohorts, respectively. The calibration curves for both cohorts demonstrated

overall consistency between predicted and observed probabilities; while the external validation cohort exhibited some deviation in the medium-to-high predicted probability region, the overall calibration performance remained acceptable (Figure 4). DCA further indicated that the model yielded a net benefit superior to the “treat all” or “treat none” strategies across a wide range of threshold probabilities in both the development and external validation cohorts, suggesting that it has certain clinical utility (Figure 5).

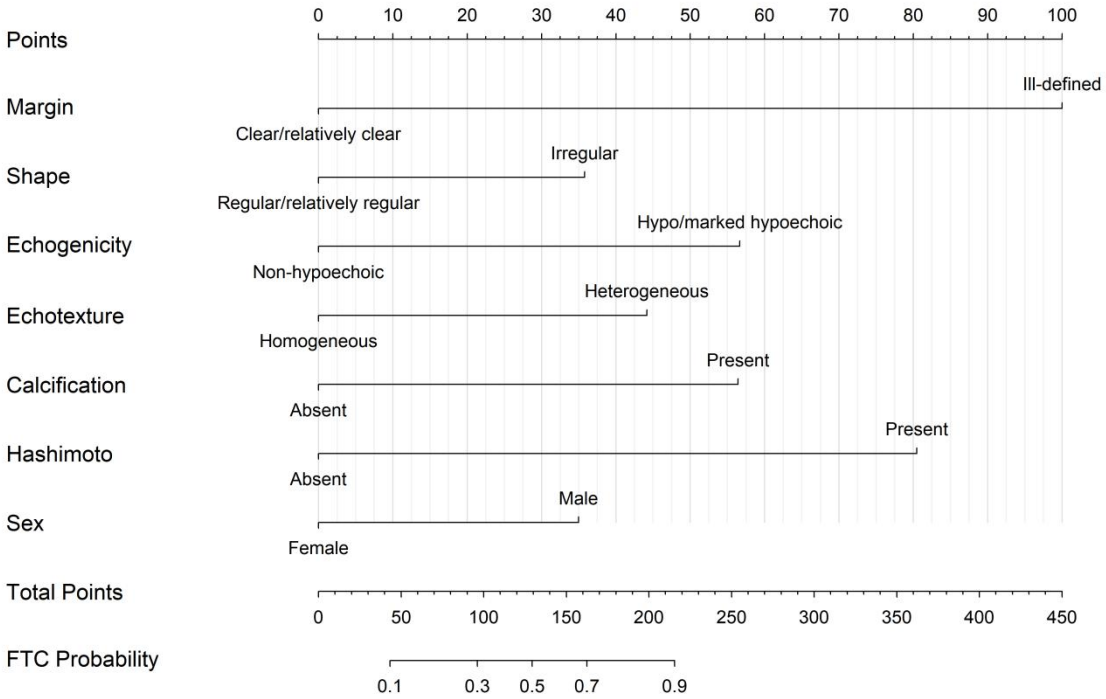


Figure 2 Diagnostic Prediction Model Nomogram

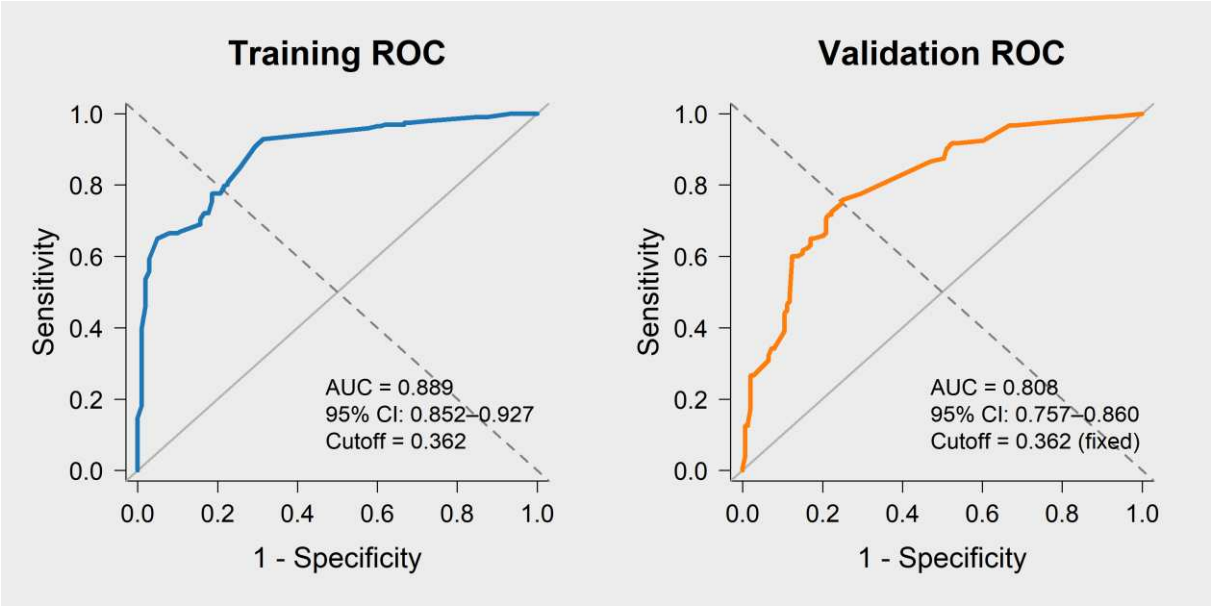


Figure 3 FTC Diagnostic Prediction Model ROC Curve

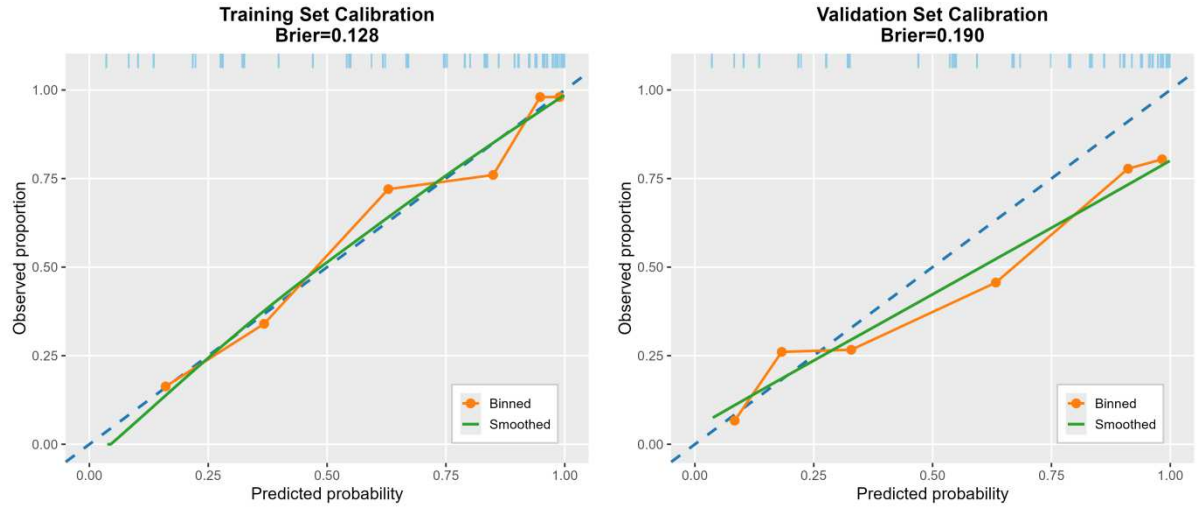


Figure 4 FTC Diagnostic Prediction Model Calibration Curve

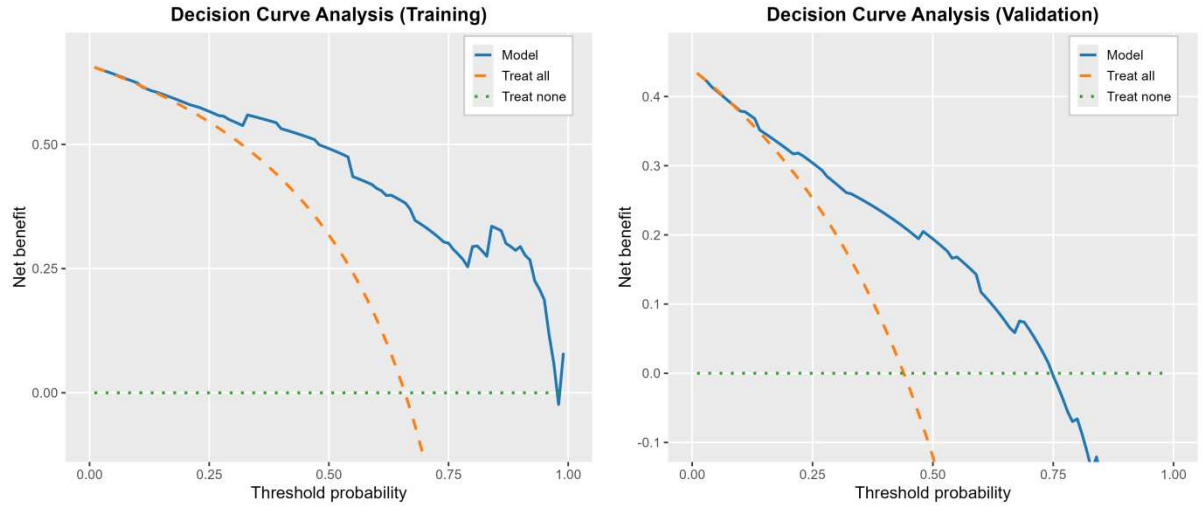


Figure 5 FTC Diagnostic Prediction Model DCA Curve

3.2.3 Classification Performance and External Validation of Diagnostic Models with Fixed Cut-off Values

Using postoperative pathology as the gold standard, the optimal cutoff value was determined to be 0.362 in the development cohort based on the Youden index, and this threshold was applied to the external validation cohort for classification performance evaluation. In the development cohort, the confusion matrix of this model was: TP=183, FN=14, FP=32, TN=70; with corresponding sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, negative likelihood ratio, and accuracy of 0.929, 0.686, 0.851, 0.833, 2.961, 0.104, and 0.846 respectively. In the external validation cohort, the confusion matrix was: TP=93, FN=27, FP=45, TN=108; the corresponding metrics were 0.775, 0.706, 0.674, 0.800, 2.635, 0.319, and 0.736 respectively(Figure 6, Table 4). The results indicate that when the optimal cutoff value determined in the development cohort was applied to the external validation cohort, the model still maintained good classification performance. Although the sensitivity, positive predictive value, and overall accuracy in the external validation cohort were lower than those in the development cohort, the specificity remained relatively stable, suggesting that the model still possesses good generalization ability and practical application value in an independent external population.

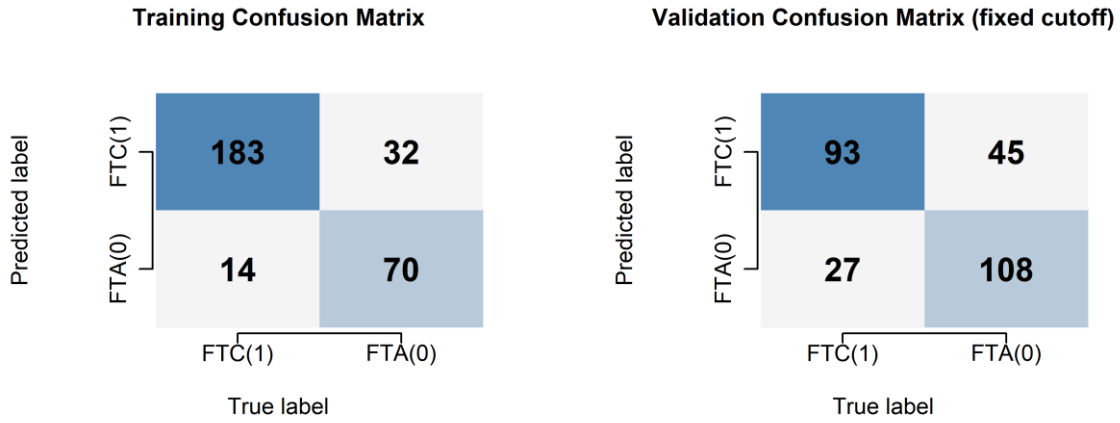


Figure 6 Confusion Matrix

3.3 Prognostic Model

3.3.1 Independent Risk Factors

Univariate Cox regression analysis was first performed in the development cohort. The results showed that pathological subtype, M stage, maximum tumor diameter, age and vascular invasion were significantly associated with PFS (all $P < 0.05$). Based on the univariate results and clinical significance, Kaplan-Meier survival curves were further plotted for groups stratified by age, maximum tumor diameter, pathological subtype and M stage. The results showed that differences in PFS between groups were statistically significant, with Log-rank test p-values of 0.011, 0.002, <0.001 and 0.003, respectively (Figure 7).

Subsequently, variables with $P < 0.05$ in the univariate analysis were included in a multivariate Cox regression model. Given the limited number of outcome events (n = 21), variables exhibiting multicollinearity were first excluded to avoid overfitting. As pathological subtype and vascular invasion were collinear, only the former was retained for the multivariate analysis. Concurrently, the surgical period (2012-2017 vs 2018-2024) was forced into the model as a temporal covariate analysis to assess the potential impact of period

differences on prognostic outcomes. Ultimately, five variables—age, maximum tumor diameter, pathological subtype, M stage and surgical period—were included, and a final model was constructed using backward stepwise regression. The results showed that pathological subtype (HR = 11.704, 95% CI: 4.851-28.236, P < 0.001) and M stage (HR = 2.879, 95% CI: 1.143-7.239, P = 0.025) were independent adverse prognostic factors for PFS (Table 5). Age, maximum tumor diameter and time period variables were not retained in the final model. The independently identified prognostic variables were subsequently used for the construction of survival curves and validation of model performance.

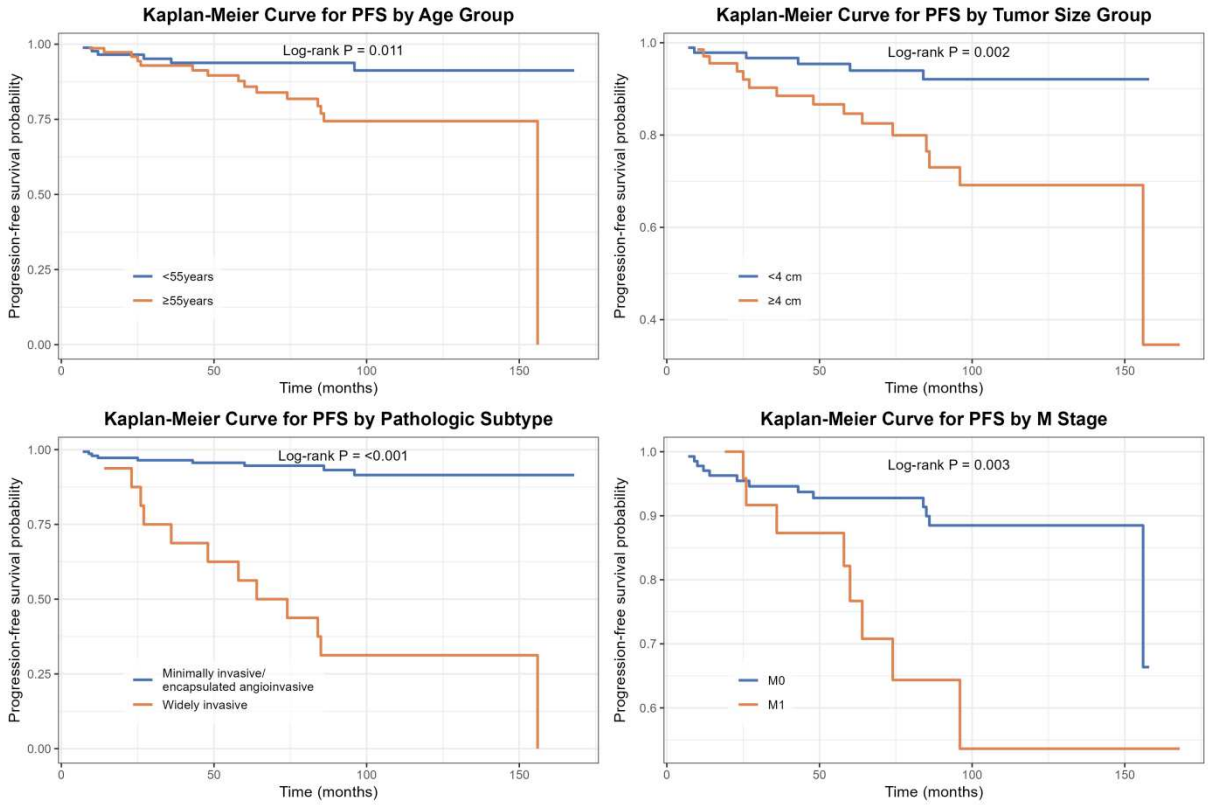


Figure 7 KM Survival Curve

3.3.2 Model Development and External Validation

Based on the two independent adverse prognostic factors identified through multivariate Cox regression analysis—namely, pathological subtype and M stage—a PFS prediction formula

was established using their regression coefficients, and a nomogram was further constructed to enable individualized prediction of the 3-year and 5-year PFS probabilities for FTC patients (Figure 8). The nomogram shows that different predictors are assigned corresponding scores based on their effect size. Clinically, the total score can be obtained by summing the scores of each variable, which is then converted into the patient's PFS probability at the corresponding time point.

To evaluate the model's predictive performance and stability, an assessment was conducted across three dimensions: discriminatory power, calibration, and clinical utility. Regarding discriminatory power, the model's C-index was 0.750 in the development cohort and 0.777 in the validation cohort, indicating that the model possesses good discriminatory ability and a low risk of overfitting. Time-dependent ROC curves were further plotted for 3-year and 5-year PFS. The results showed that the AUCs for 3-year and 5-year PFS in the development cohort were 0.707 (95% CI: 0.528-0.885) and 0.754 (95% CI: 0.607-0.902), respectively; in the external validation cohort, the corresponding AUCs were 0.715 (95% CI: 0.528-0.901) and 0.813 (95% CI: 0.656-0.970), suggesting that the model possesses stable discriminatory power in an independent external cohort (Figure 9). An exploratory ROC analysis for 10-year PFS was additionally conducted in the development cohort, yielding an AUC of 0.822 (95% CI: 0.713-0.932); due to the lack of sufficient long-term follow-up data in the external validation cohort, no external validation for 10-year PFS was performed, and the relevant results are presented as supplementary material (Supplementary Figure S1). Regarding calibration, the 3-year and 5-year calibration curves show that the model's predicted probabilities generally align with the observed probabilities in the development

cohort, with the 5-year calibration performing relatively closer to the ideal reference line. In the external validation cohort, the 3-year and 5-year calibrated curves also demonstrated overall consistency; although some deviation was observed in certain probability intervals, the overall calibration performance was acceptable (Figure 10). Regarding clinical utility, DCA indicated that in the development cohort, the model yielded a net benefit superior to the “treat all” and “treat none” strategies across several threshold probability ranges at both the 3-year and 5-year time points; in the external validation cohort, the model still demonstrated a relatively clear net benefit in 5-year PFS prediction, while the net benefit for 3-year PFS prediction was primarily concentrated within lower threshold probability ranges, suggesting that the model possesses certain clinical utility in the external validation population (Figure 11).

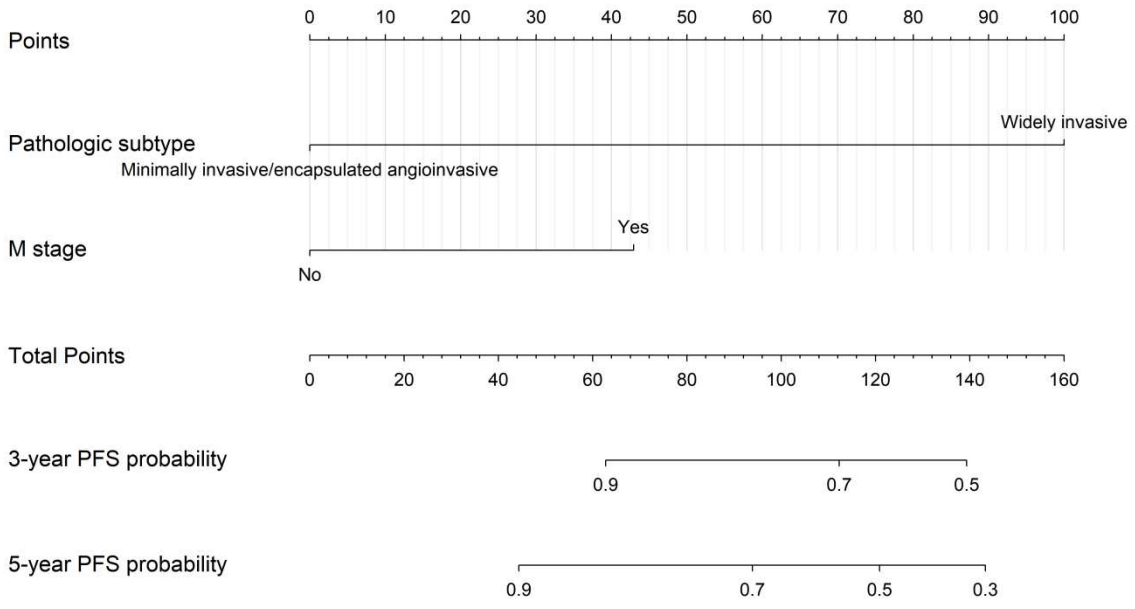


Figure 8 Prognostic Model Nomogram

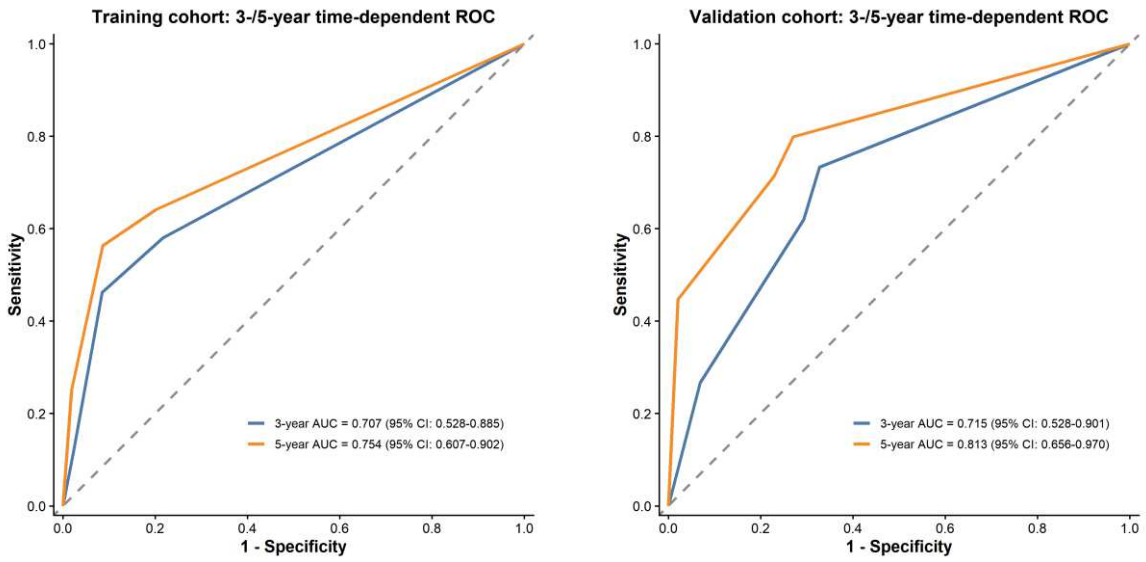


Figure 9 Prognostic Model ROC Curve

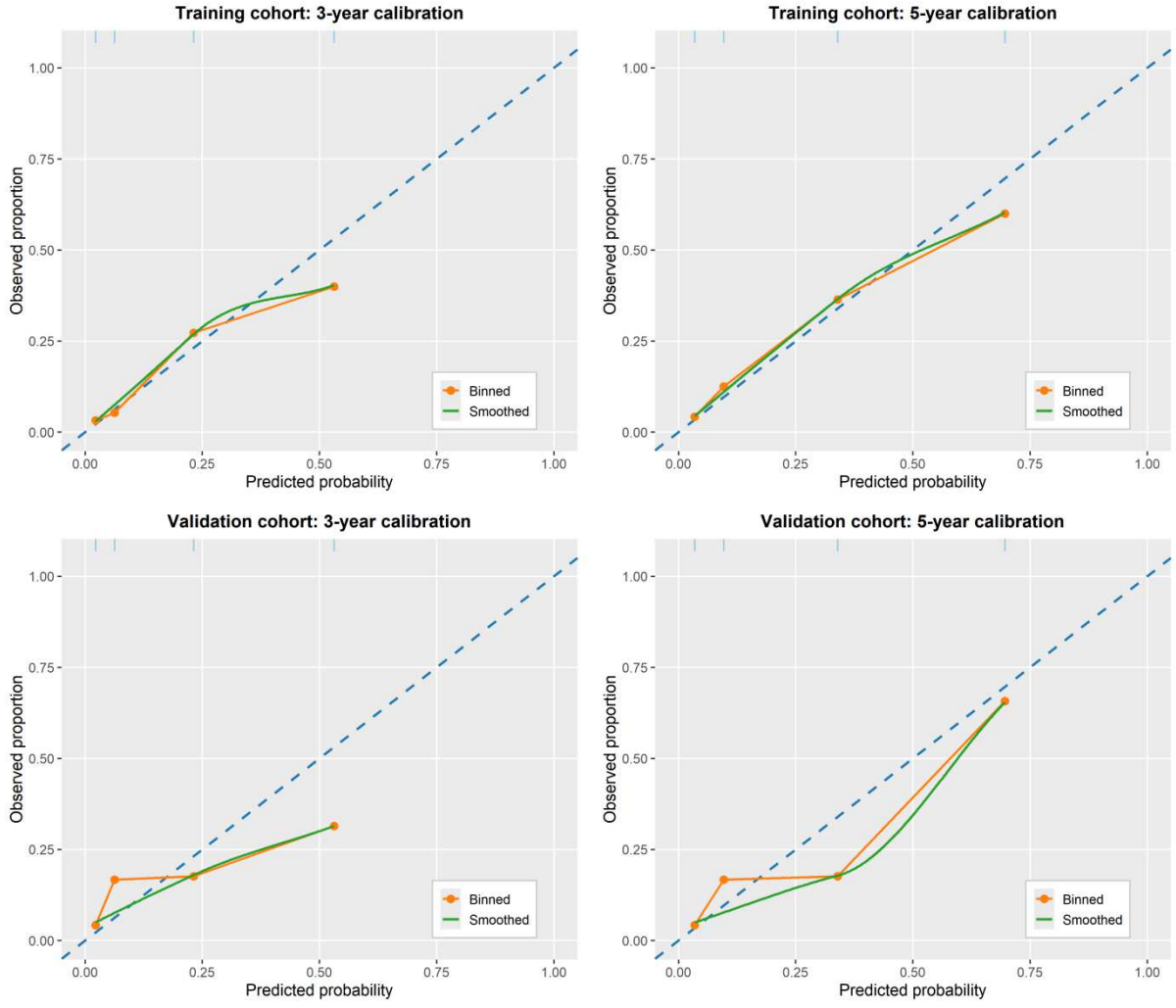


Figure 10 Prognostic Model Calibration Curve

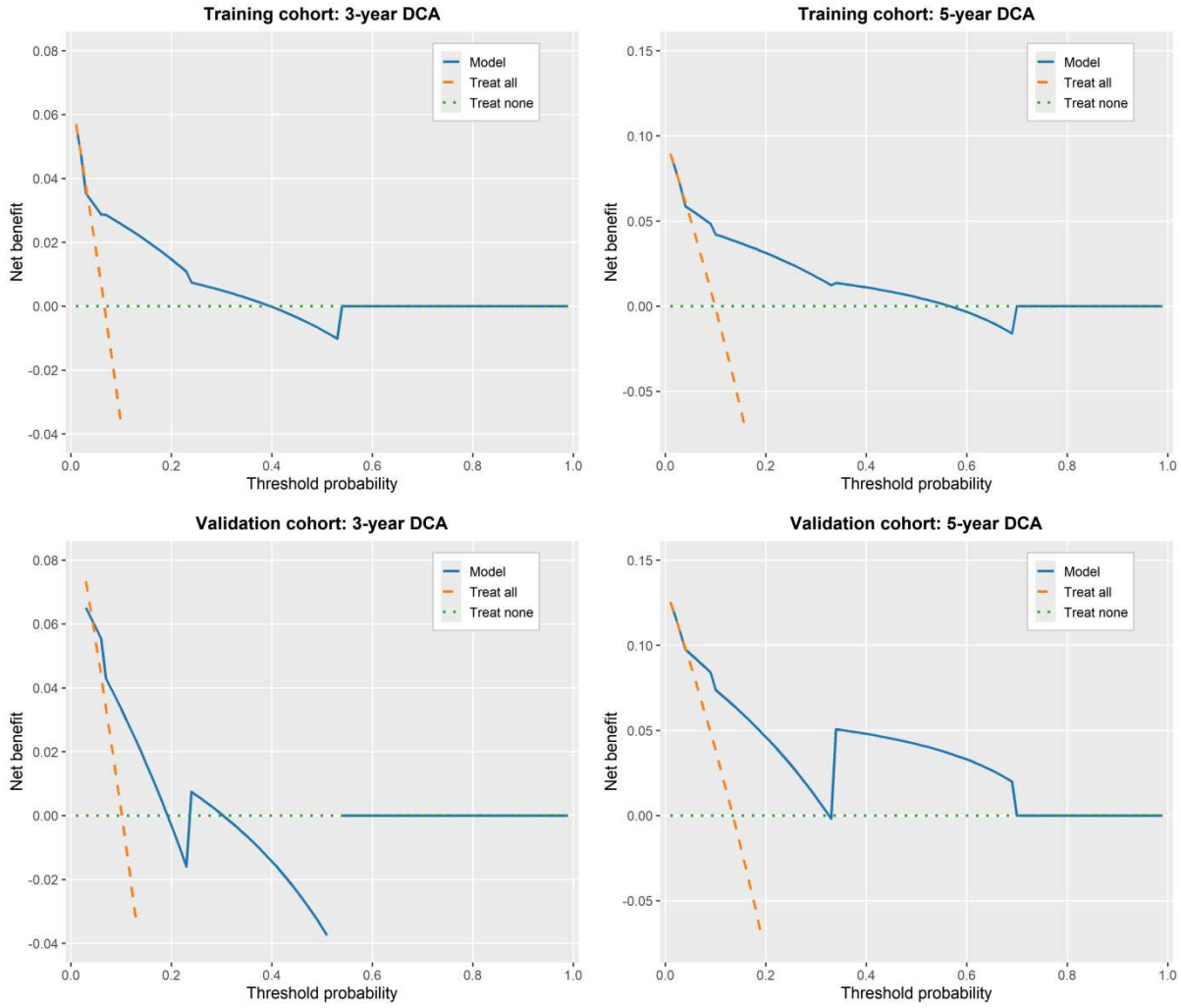


Figure 11 Prognostic Model DCA Curve

3.3.3 Temporal Sensitivity Analysis

Given that the development cohort spans the period 2012-2024, while the external validation cohort comprises cases from 2018-2024, a temporal sensitivity analysis was conducted in this study to further assess the model's stability when trained on data from different time periods. The development cohort was divided into two training subsets based on the date of surgery: Training Subset 1 (2012-2017) and Training Subset 2 (2018-2024). Prognostic models were constructed for each subset using the same modeling strategy and validated against the same external validation cohort.

The results showed that the discriminatory performance of the models built from the

two training subsets was largely consistent in the external validation cohort. The AUCs for 3-year and 5-year PFS predicted by the model built from Training Subset 1 in the external validation cohort were 0.715 (95% CI: 0.528-0.901) and 0.813 (95% CI: 0.656-0.970), respectively; the model built using training subset 2 also yielded AUCs of 0.715 (95% CI: 0.528-0.901) and 0.813 (95% CI: 0.656-0.970) for 3-year and 5-year PFS, respectively, in the external validation cohort (Figure 12). Furthermore, the C-index for both models in the external validation cohort was 0.777. This suggests that the prognostic model demonstrates stable external validation performance across different periods of training data, indicating good temporal robustness. These results are consistent with the findings of the temporal covariate analysis, further supporting the conclusion that differences in the period of case collection did not have a significant impact on the model's primary predictive performance.

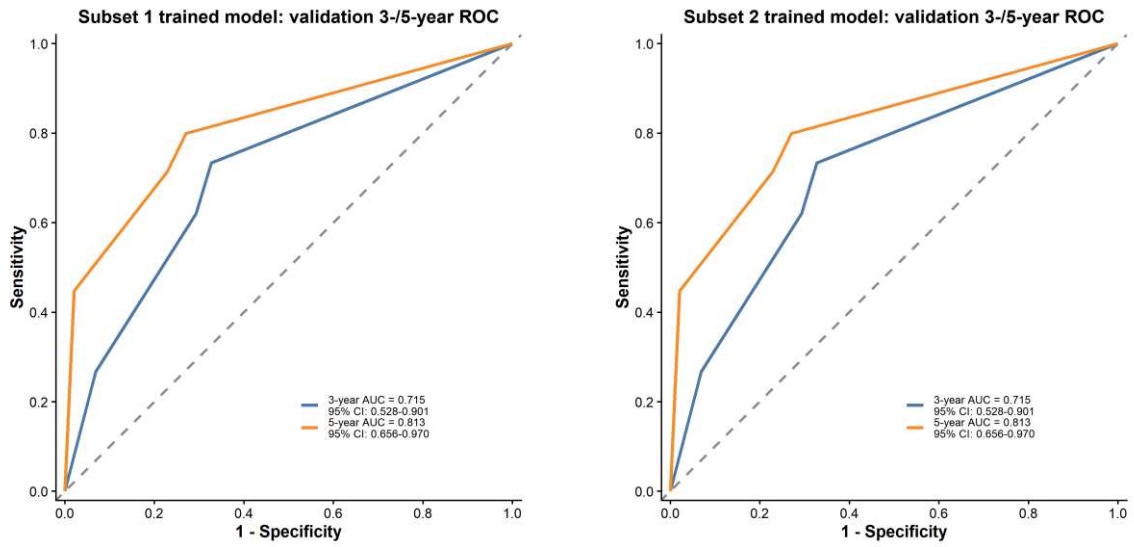


Figure 12 Temporal Sensitivity Analysis

4. Discussion

Based on multicenter clinical data, this study addresses two core clinical issues related to FTC: first, how to more accurately distinguish between FTC and FTA preoperatively; and second,

how to more effectively stratify PFS risk in FTC patients postoperatively. Unlike previous studies that focused solely on diagnosis or prognosis, this study simultaneously established and completed external validation of both diagnostic and prognostic models within the same research framework. More importantly, the significance of this study lies not only in the predictive performance of the models themselves, but also in the fact that the variables ultimately retained reveal several key characteristics of FTC in terms of its radiological presentation and biological behavior.

In the diagnostic prediction model, margin had the strongest effect, suggesting that it remains one of the most valuable clues for preoperative differentiation of follicular tumors. Radiologically, indistinct margin typically reflects disruption of the tumor-capsule-gland interface, which may be associated with capsular invasion, infiltrative growth, or localized microinvasion(8). Several recent studies also support the value of margin in the differential diagnosis of FTC(9-11). Calcification, echogenicity, echotexture, and shape were also included in the final model; these features reflect the complexity of the tumor microenvironment and potential invasiveness, indicating that preoperative identification of FTC does not rely on a single, unique characteristic, but rather manifests as a combination of multiple suspicious yet nonspecific features(12). Recent ultrasound predictive studies on FTC have also shown that features such as margin, calcification, hypoechogenicity, and internal structural abnormalities have greater diagnostic value when modeled in combination than when considered as single indicators, which is consistent with the findings of this study(13).

The independent significance of coexisting Hashimoto’s thyroiditis warrants attention. It may not merely serve as a background condition but, by altering the glandular parenchyma,

influencing nodule imaging, and affecting the local microenvironment, provides additional information for the preoperative identification of FTC(14). Studies have found that chronic lymphocytic thyroiditis itself is an independent risk factor for FTC (OR = 9.195, P = 0.006), and when combined with other ultrasound features, it yields high diagnostic accuracy (AUC $\approx$ 0.914)(15), suggesting that an autoimmune background may, to some extent, contribute to or indicate tumor-related risk(16). In contrast, sex is more likely to reflect differences in the epidemiological distribution between FTC and FTA rather than a direct pathological mechanism(17, 18). On the other hand, although smoking, Tg, and the convex capsule sign were significant in univariate analysis, they were not retained in the final model, suggesting that these variables may overlap with stronger imaging predictors rather than independently providing diagnostic information(19, 20).

In the prognostic model, pathological subtype and M stage were the two independent adverse prognostic factors ultimately retained; this result is highly clinically plausible. Pathological subtype reflects the tumor's inherent aggressiveness. The minimally invasive and encapsulating vascular invasion subtypes generally belong to the encapsulating/focal invasion spectrum. The extensively invasive subtype is more likely to be accompanied by significant vascular invasion and a tendency for hematogenous dissemination, leading to a higher risk of distant metastasis and disease recurrence(21). Recent evidence is consistent with the conclusions of this study(21, 22). M stage represents the burden of distant dissemination, indicating that the tumor has crossed the barrier of the primary tissue and exhibits stronger systemic invasiveness; therefore, it is closely associated with shorter progression-free survival(2). Recent studies have also pointed out that distant metastasis is not only an

independent poor prognostic factor for FTC patients but also significantly increases the risk of cancer-specific mortality(23, 24). Compared to FTC without distant metastasis, the presence of distant metastasis often implies a more complex treatment pathway—such as reduced uptake of radioactive iodine and the need for multimodal therapy—which consequently affects long-term PFS and overall outcomes(25).

It is worth noting that although age, maximum tumor diameter, and vascular invasion were significantly associated with PFS in univariate analysis, they were not ultimately retained in the multivariate model. This does not imply that these factors are unimportant; rather, it likely indicates that their independent contribution is diminished once pathological subtype and M stage are included in the model(26-29). This is particularly true for vascular invasion, recent studies suggest that its extent is closely associated with the prognosis of FTC; however, when it is highly correlated with histological aggressiveness, continuing to include it in the model may increase instability without necessarily improving predictive power(30). Given the limited number of PFS events in the cohort developed for this study, adopting a more parsimonious model is both reasonable and necessary.

From the perspective of clinical pathways, the two models in this study correspond to the two key phases of FTC management—preoperative and postoperative, respectively. Compared with previous studies that often focused on a single endpoint, this study integrates preoperative differentiation and postoperative stratification into a single multicenter framework, thereby facilitating a more continuous approach to individualized management(10, 23, 31).

This study still has several limitations. First, as this is a retrospective study, selection

bias and information bias are difficult to avoid. Second, the limited sample size and number of PFS events in the prognostic model restrict the model's complexity and parameter stability. Third, there remains some clinical and pathological heterogeneity between the external validation cohort and the development cohort; while this enhances the value of external validation, it may also affect certain performance metrics. Finally, the diagnostic model has not yet incorporated molecular biomarkers, cytological subtypes, or higher-level radiomic information. Future studies should further validate the model in larger, prospective, multicenter cohorts and attempt to integrate molecular testing, cytological stratification, and radiomic information to enhance the model's generalizability and long-term clinical utility.

5. Conclusion

The predictive and prognostic models developed in this study may serve as useful tools for the personalized management of FTC, aiding in the accurate preoperative assessment of malignant nodules, effectively predicting the risk of disease progression postoperatively, and supporting follow-up and treatment decisions.

Abbreviations

AUC	Area Under the Curve
CI	Confidence Interval
DCA	Decision Curve Analysis
FTA	Follicular Thyroid Adenoma
FTC	Follicular Thyroid Carcinoma
FTN	Follicular Thyroid Neoplasm

HR	Hazard Ratio
OR	Odds Ratio
PFS	Progression-Free Survival
ROC	Receiver Operating Characteristic

Declarations

Ethics approval and consent to participate

This study was approved by The First Affiliated Hospital of Zhengzhou University and Henan Cancer Hospital Ethics Review Committee (Ethics No.2020-KY-0075-002). This study was performed in accordance with the Declaration of Helsinki. Written informed consent was obtained from all patients enrolled in this study.

Consent for publication

Not applicable.

Availability of data and materials

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

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the Henan Provincial Key Research and Development Program (No.25111312400).

Authors' contributions

Z.C. participated in the design of the study, data collection, statistical analysis and paper writing. T.G. participated in data collection and statistical analysis. P.X. participated in external validation. Y.L. and Y.C. participated in revising the manuscript. Q.C. participated in data sorting. Q.Z. participated in data collection. D.Y. participated in the design of the study and helped to revise the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We extend our sincere gratitude to all the participants involved in this study.

References

1. Su DH, Chang TC, Chang SH. Prognostic factors on outcomes of follicular thyroid cancer. J Formos Med Assoc. 2019;118(7):1144-53.
2. Ito Y, Hirokawa M, Masuoka H, Higashiyama T, Kihara M, Onoda N, et al. Prognostic factors for follicular thyroid carcinoma: the importance of vascular invasion. Endocr J. 2022;69(9):1149-56.
3. Deveci MS, Deveci G, LiVolsi VA, Baloch ZW. Fine-needle aspiration of follicular lesions of the thyroid. Diagnosis and follow-Up. Cytojournal. 2006;3:9.
4. Baloch ZW, Asa SL, Barletta JA, Ghossein RA, Juhlin CC, Jung CK, et al. Overview of the 2022 WHO Classification of Thyroid Neoplasms. Endocr Pathol. 2022;33(1):27-63.

5. Zahid A, Shafiq W, Nasir KS, Loya A, Abbas Raza S, Sohail S, et al. Malignancy rates in thyroid nodules classified as Bethesda categories III and IV; a subcontinent perspective. J Clin Transl Endocrinol. 2021;23:100250.
6. Zhang Y, Ding W, Han M, Wu X, Agyekum EA, Xu S, et al. A practical ultrasound-based risk stratification system for differentiating follicular thyroid carcinoma and adenoma. Gland Surg. 2025;14(10):1873-85.
7. Shen J, Yan M, Chen L, Ou D, Yao J, Feng N, et al. Prognosis and influencing factors of follicular thyroid cancer. Cancer Med. 2024;13(1):e6727.
8. Li W, Song Q, Lan Y, Li J, Zhang Y, Yan L, et al. The Value of Sonography in Distinguishing Follicular Thyroid Carcinoma from Adenoma. Cancer Manag Res. 2021;13:3991-4002.
9. Ringel MD, Sosa JA, Baloch Z, Bischoff L, Bloom G, Brent GA, et al. 2025 American Thyroid Association Management Guidelines for Adult Patients with Differentiated Thyroid Cancer. Thyroid. 2025;35(8):841-985.
10. Zhang F, Mei F, Chen W, Zhang Y. Role of Ultrasound and Ultrasound-Based Prediction Model in Differentiating Follicular Thyroid Carcinoma From Follicular Thyroid Adenoma. J Ultrasound Med. 2024;43(8):1389-99.
11. Ma C, Zhong M, Chen D, Ye J, Cheng S, Fu J, et al. Independent predictors and ultrasound-based risk stratification for follicular thyroid tumors: a multicenter retrospective study. World J Surg Oncol. 2025;24(1):32.
12. Borowczyk M, Wolinski K, Wieckowska B, Jodlowska-Siewert E, Szczepanek-Parulska E, Verburg FA, et al. Sonographic Features Differentiating Follicular Thyroid Cancer from Follicular Adenoma-A Meta-Analysis. Cancers (Basel). 2021;13(5).
13. Li C, Luo Y, Gan Y, Jiang Y, Li Q, Huang J. Development and validation of a predictive model for assessing the risk of follicular carcinoma in thyroid nodules identified as suspicious by intraoperative frozen section. Front Endocrinol (Lausanne). 2024;15:1431247.
14. Silva de Morais N, Stuart J, Guan H, Wang Z, Cibas ES, Frates MC, et al. The Impact of Hashimoto Thyroiditis on Thyroid Nodule Cytology and Risk of Thyroid Cancer. J Endocr Soc. 2019;3(4):791-800.
15. Alyusuf EY, Alhmayin L, Albasri E, Enani J, Altuwaijri H, Alsomali N, et al. Ultrasonographic predictors of thyroid cancer in Bethesda III and IV thyroid nodules. Front Endocrinol (Lausanne). 2024;15:1326134.
16. Yao XY, Li X, Mei F, Yu B, Song SB, Cui LG, et al. Prospective Validation of an Ultrasound Capsule-Based Model for Predicting Follicular Thyroid Carcinoma Without High-Risk Features. Cancer Med. 2025;14(17):e71190.

17. Najafian A, Olson MT, Schneider EB, Zeiger MA. Clinical presentation of patients with a thyroid follicular neoplasm: are there preoperative predictors of malignancy? Ann Surg Oncol. 2015;22(9):3007-13.
18. Wu Q, Qu Y, Li Y, Liu Y, Shen J, Wang Y. Logistic regression analysis of contrast-enhanced ultrasound and conventional ultrasound of follicular thyroid carcinoma and follicular adenoma. Gland Surg. 2021;10(10):2890-900.
19. Jiang H, Li Y, Shen J, Lin H, Fan S, Qiu R, et al. Cigarette smoking and thyroid cancer risk: A Mendelian randomization study. Cancer Med. 2023;12(19):19866-73.
20. Zhu X, Liu F, Liu J, Li Z, Ma Y. Thyroglobulin-to-tumor volume ratio combined with ultrasound features for diagnosing thyroid follicular neoplasms. Front Endocrinol (Lausanne). 2025;16:1626766.
21. Cheng J, Han B, Chen Y, Li Q, Xia W, Wang N, et al. Clinical risk factors and cancer risk of thyroid imaging reporting and data system category 4 A thyroid nodules. J Cancer Res Clin Oncol. 2024;150(6):327.
22. Liu X, Chen S, Wu C. Nomogram for predicting overall and cancer-specific survival in patients with postoperative follicular thyroid cancer. Braz J Otorhinolaryngol. 2026;92(1):101719.
23. Yamazaki H, Sugino K, Katoh R, Matsuzu K, Kitagawa W, Nagahama M, et al. Management of follicular thyroid carcinoma. Eur Thyroid J. 2024;13(5).
24. Ito Y, Miyauchi A. Prognostic factors of papillary and follicular carcinomas based on pre-, intra-, and post-operative findings. Eur Thyroid J. 2024;13(5).
25. Schlumberger M, Leboulleux S. Treatment of distant metastases from follicular cell-derived thyroid cancer. F1000Prime Rep. 2015;7:22.
26. Quan ML, Pasieka JL, Rorstad O. Bone mineral density in well-differentiated thyroid cancer patients treated with suppressive thyroxine: a systematic overview of the literature. J Surg Oncol. 2002;79(1):62-9; discussion 9-70.
27. Schneider TC, de Wit D, Links TP, van Erp NP, van der Hoeven JJ, Gelderblom H, et al. Everolimus in Patients With Advanced Follicular-Derived Thyroid Cancer: Results of a Phase II Clinical Trial. J Clin Endocrinol Metab. 2017;102(2):698-707.
28. Grønlund MP, Jensen JS, Hahn CH, Grønhøj C, Buchwald CV. Risk Factors for Recurrence of Follicular Thyroid Cancer: A Systematic Review. Thyroid. 2021;31(10):1523-30.
29. Leem DE, Yoo JH, Kim BR, Kim JS, Kim TH, Kim SW, et al. Clinical Outcomes of Follicular Thyroid Carcinoma Did Not Significantly Differ according to Tumor Size in an Iodine-Excessive Country. Endocrinol Metab. 2025;40(5):772-80.

30. Jin S, Ye X, Ye T, Chen X, Ji J, Wang J, et al. Nomogram models for predicting outcomes in thyroid cancer patients with distant metastasis receiving (131)iodine therapy. Sci Rep. 2025;15(1):2486.
31. Buzejic M, Bukumiric Z, Rovcanin B, Jovanovic M, Stojanovic M, Zoric G, et al. Prognostic Factors for Cancer-Specific Survival and Disease-Free Interval in 130 Patients with Follicular Thyroid Carcinoma: Single Institution Experience. Diagnostics (Basel). 2024;14(24).

Figure legends

Figure 1 Flowchart

Figure 2 Diagnostic Prediction Model Nomogram

Figure 3 FTC Diagnostic Prediction Model ROC Curve

Figure 4 FTC Diagnostic Prediction Model Calibration Curve

Figure 5 FTC Diagnostic Prediction Model DCA Curve

Figure 6 Confusion Matrix

Figure 7 KM Survival Curve

Figure 8 Prognostic Model Nomogram

Figure 9 Prognostic Model ROC Curve

Figure 10 Prognostic Model Calibration Curve

Figure 11 Prognostic Model DCA Curve

Figure 12 Temporal Sensitivity Analysis

Figure S1 Exploratory 10-year ROC analysis of the prognostic model

Tables

Table 1 Baseline characteristics of FTC and FTA patients in the development and external validation cohorts

Variable	Development cohort			External validation cohort		
	FTA (n=102)	FTC (n=197)	*P* value	FTA (n=153)	FTC (n=120)	*P* value
General characteristics						
Age (years)	47.5 (36, 54.75)	51 (37, 61)	0.060	52 (37, 58)	51 (40.75, 61.5)	0.340
Sex						
Female	80 (78.4%)	129 (65.5%)	0.021	129 (84.3%)	91 (75.8%)	0.079
Male	22 (21.6%)	68 (34.5%)		24 (15.7%)	29 (24.1%)	
Smoking						
Yes	6 (5.9%)	42 (21.3%)	<0.001	5 (3.3%)	8 (6.7%)	0.191
No	96 (94.1%)	155 (78.6%)		148 (96.7%)	112 (93.3%)	
Laboratory parameters						
FT3	4.81 (4.38, 5.46)	4.85 (4.35, 5.51)	0.626	4.85 (4.41, 5.28)	4.95 (4.44, 5.37)	0.595
FT4	10.63 (9.18, 12.5)	10.53 (8.38, 12.53)	0.842	15 (13.24, 16.85)	15.43 (13.51, 17.23)	0.257
TSH	1.81 (1.06, 3.26)	1.88 (0.75, 3.86)	0.892	2.52 (1.57, 3.34)	2.23 (1.46, 3.19)	0.283
Tg	35 (9.22, 200.41)	220.26 (49.4, 692)	<0.001	89.67 (24.5, 314.43)	250.55 (53, 500)	0.002
TgAb	16 (10.28, 24.35)	20.9 (12.02, 55.92)	0.005	17.1 (13.14, 21.79)	18.2 (11.95, 38.55)	0.265
Ultrasound features						
Size (mm)	41 (28.12, 54)	37 (25, 50)	0.168	37 (22, 47)	36.5 (24.5, 52)	0.319
Margin						
Poorly defined/ill-defined margin	7 (6.9%)	101 (51.3%)	<0.001	20 (13.1%)	78 (65.0%)	<0.001
Fairly defined/clearly defined	95 (93.1%)	96 (48.7%)		133 (86.9%)	42 (35.0%)	
Shape						

Variable	Development cohort			External validation cohort		
	FTA (n=102)	FTC (n=197)	*P* value	FTA (n=153)	FTC (n=120)	*P* value
Abnormal/irregular/slightly irregular	14 (13.7%)	90 (45.7%)	<0.001	10 (6.5%)	35 (29.2%)	<0.001
Fairly regular/regular/oval	88 (86.3%)	107 (54.3%)		143 (93.5%)	85 (70.8%)	
Echogenicity						
Hypoechoic/very hypoechoic	72 (70.6%)	186 (94.4%)	<0.001	71 (46.4%)	95 (79.2%)	<0.001
Isoechoic/hyperechoic	30 (29.4%)	11 (5.5%)		82 (53.6%)	25 (20.8%)	
Echotexture						
Heterogeneous/poorly homogeneous	62 (60.8%)	145 (73.6%)	0.023	120 (78.4%)	97 (80.8%)	0.626
Fairly homogeneous/homogeneous	40 (39.2%)	52 (26.3%)		33 (21.6%)	23 (19.1%)	
Calcification						
Yes	22 (21.6%)	114 (57.9%)	<0.001	28 (18.3%)	36 (30.0%)	0.024
No	80 (78.4%)	83 (42.1%)		125 (81.7%)	84 (70.0%)	
Blood flow signal grade						
Grade 2-3	73 (71.6%)	127 (64.5%)	0.216	62 (40.5%)	47 (39.2%)	0.820
Grade 0-1	29 (28.4%)	70 (35.5%)		91 (59.5%)	73 (60.8%)	
Convex capsule sign						
Yes	5 (4.9%)	30 (15.2%)	0.008	2 (1.3%)	10 (8.3%)	0.005

Variable	Development cohort			External validation cohort		
	FTA (n=102)	FTC (n=197)	*P* value	FTA (n=153)	FTC (n=120)	*P* value
No	97 (95.1%)	167 (84.8%)		151 (98.7%)	110 (91.7%)	
Composition						
Solid	62 (60.8%)	115 (58.4%)	0.688	72 (47.1%)	74 (61.7%)	0.016
Cystic/cystic-solid/mixed	40 (39.2%)	82 (41.6%)		81 (52.9%)	46 (38.3%)	
TI-RADS classification						
Grade 4-5	41 (40.2%)	101 (51.3%)	0.069	61 (39.9%)	58 (48.3%)	0.162
Grade 2-3	61 (59.8%)	96 (48.7%)		92 (60.1%)	62 (51.6%)	
Multifocality						
Yes	13 (12.7%)	35 (17.8%)	0.262	17 (11.1%)	17 (14.2%)	0.448
No	89 (87.3%)	162 (82.2%)		136 (88.9%)	103 (85.8%)	
Hashimoto's thyroiditis						
Yes	8 (7.8%)	52 (26.4%)	<0.001	19 (12.4%)	35 (29.2%)	<0.001
No	94 (92.2%)	145 (73.6%)		134 (87.6%)	85 (70.8%)	
Hyperthyroidism						
Yes	6 (5.9%)	7 (3.6%)	0.378	2 (1.3%)	1 (0.8%)	1.000
No	96 (94.1%)	190 (96.4%)		151 (98.7%)	119 (99.2%)	

Note: Continuous variables are presented as median (interquartile range), and categorical variables as n (%). *P* values indicate comparisons between FTC and FTA within each center.

Table 2 Baseline characteristics of the development and external validation cohorts for the

1 prognostic model

Variable	Level/Statistic	Development cohort (n=161)	Validation cohort (n=87)	*P* value
General information				
Age (years)		52.00 (37.00, 61.00)	53.00 (41.00, 64.00)	0.499
Sex	Female	102 (63.4%)	61 (70.1%)	0.284
	Male	59 (36.6%)	26 (29.9%)	
Symptom	No	136 (84.5%)	72 (82.8%)	0.726
	Yes	25 (15.5%)	15 (17.2%)	
Smoking	No	139 (86.3%)	81 (93.1%)	0.108
	Yes	22 (13.7%)	6 (6.9%)	
Surgery and pathology				
Final extent of surgery	Lobectomy	47 (29.2%)	28 (32.2%)	0.625
	Total thyroidectomy	114 (70.8%)	59 (67.8%)	
Lymph node dissection	No	96 (59.6%)	56 (64.4%)	0.465
	Yes	65 (40.4%)	31 (35.6%)	
Surgical procedure	No	156 (96.9%)	87 (100.0%)	0.165
	Yes	5 (3.1%)	0 (0.0%)	
Maximum tumor diameter (cm)		3.00 (2.00, 4.50)	3.50 (2.50, 5.50)	0.082
Multifocality	No	131 (81.4%)	72 (82.8%)	0.786
	Yes	30 (18.6%)	15 (17.2%)	
Lymph node metastasis	No	44 (27.3%)	22 (25.3%)	0.728
	Yes	21 (13.0%)	9 (10.3%)	
	Not dissected	96 (59.6%)	56 (64.4%)	
Capsular invasion	No	25 (15.5%)	11 (12.6%)	0.538
	Yes	136 (84.5%)	76 (87.4%)	
Vascular invasion	No	82 (50.9%)	43 (49.4%)	0.821
	Yes	79 (49.1%)	44 (50.6%)	
T stage	T1-T2	89 (55.3%)	44 (50.6%)	0.478
	T3-T4	72 (44.7%)	43 (49.4%)	
N stage	N0	140 (87.0%)	78 (89.7%)	0.534
	N1	21 (13.0%)	9 (10.3%)	
M stage	M0	135 (83.9%)	74 (85.1%)	0.803

Variable	Level/Statistic	Development cohort (n=161)	Validation cohort (n=87)	*P* value
	M1	26 (16.1%)	13 (14.9%)	
I-131 therapy	No	69 (42.9%)	57 (65.5%)	<0.001
	Yes	92 (57.1%)	30 (34.5%)	
Hashimoto's thyroiditis	No	143 (88.8%)	61 (70.1%)	<0.001
	Yes	18 (11.2%)	26 (29.9%)	
Hyperthyroidism	No	156 (96.9%)	86 (98.9%)	0.668
	Yes	5 (3.1%)	1 (1.1%)	
Extraglandular invasion	No	154 (95.7%)	81 (93.1%)	0.388
	Yes	7 (4.3%)	6 (6.9%)	
Pathological subtype	Minimally invasive/encapsulated angioinvasive	145 (90.1%)	63 (72.4%)	<0.001
	Widely invasive	16 (9.9%)	24 (27.6%)	
Hürthle cell subtype	No	157 (97.5%)	87 (100.0%)	0.301
	Yes	4 (2.5%)	0 (0.0%)	
Radiotherapy/chemotherapy or seed implantation	No	159 (98.8%)	85 (97.7%)	0.614
	Yes	2 (1.2%)	2 (2.3%)	
Follow-up outcomes				
Death	No	150 (93.2%)	84 (96.6%)	0.39
	Yes	11 (6.8%)	3 (3.4%)	
Local recurrence	No	156 (96.9%)	86 (98.9%)	0.668
	Yes	5 (3.1%)	1 (1.1%)	
Lymph node recurrence	No	157 (97.5%)	80 (92.0%)	0.055
	Yes	4 (2.5%)	7 (8.0%)	
Distant metastatic progression/new metastasis	No	154 (95.7%)	83 (95.4%)	1
	Yes	7 (4.3%)	4 (4.6%)	
PFS event status	No	140 (87.0%)	76 (87.4%)	0.929
	Yes	21 (13.0%)	11 (12.6%)	
PFS time (months)		76.00 (47.00, 113.00)	64.00 (27.50, 85.00)	<0.001

Note: Continuous variables are presented as median (Q1, Q3) and were compared using the Mann-Whitney U test. Categorical variables are presented as n (%) and were compared using the chi-square test or Fisher exact test, as appropriate. Percentages were calculated using the total sample size of each hospital as the denominator.

Table 3 Univariate and Multivariate Logistic Regression Analysis

	Univariate analysis		Multivariate analysis	
Variables	OR（95%CI）	*P* value	OR（95%CI）	*P* value
Margin	14.278（6.308-32.317）	＜0.001***	12.821（5.009-32.815）	＜0.001***
Calcification	4.995（2.881-8.658）	＜0.001***	4.22（2.061-8.639）	＜0.001***
Hashimoto's thyroiditis	4.214（1.916-9.269）	＜0.001***	7.786（2.921-20.751）	＜0.001***
Echogenicity	7.045（3.353-14.804）	＜0.001***	4.244（1.74-10.349）	0.001***
Echotexture	1.799（1.082-2.991）	0.024*	3.086（1.465-6.5）	0.003**
Sex	1.917（1.100-3.342）	0.022*	2.441（1.176-5.069）	0.017*
Shape	5.287（2.816-9.925）	＜0.001***	2.491（1.15-5.4）	0.021*
Smoking	4.335（1.776-10.584）	0.001***		
Tg	1.000（1.000-1.001）	0.012*		
Convex capsule sign	3.485（1.309-9.279）	0.012*		
Time		0.329		
2012-2015(reference）	1			
2016-2019	0.666（0.368-1.204）	0.178		
2020-2024	0.684（0.377-1.240）	0.211		

Note: OR is presented as the odds ratio with its 95% CI. For categorical variables, the OR is displayed as the comparison of the highest category relative to the reference category (lowest

category/0); for continuous variables, the OR represents the comparison per 1-unit increase. Significance levels are denoted as: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Table 4 Diagnostic Performance of Models in the Development and Validation Cohorts

	Development	Validation
Cutoff	0.362	0.362
Se	0.929	0.775
Sp	0.686	0.706
PPV	0.851	0.674
NPV	0.833	0.8
LR+	2.961	2.635
LR-	0.104	0.319
Acc	0.846	0.736

Note: Se, sensitivity; Sp, specificity; PPV, positive predictive value; NPV, negative predictive value; LR+, positive likelihood ratio; LR-, negative likelihood ratio; Acc, accuracy. The validation cohort cutoff was determined based on the Youden index in the development cohort and applied consistently.

Table 5 Univariate and Multivariate Cox Regression Analysis

Variables	**Univariate analysis**		**Multivariate analysis**	
	HR(95%CI)	***P* value**	**HR(95%CI)**	***P* value**
Pathological subtype	12.939（5.412-30.935）	<0.001***	11.704(4.851-28.236)	<0.001***
M stage	3.597(1.482-8.728)	0.005**	2.879 (1.143-7.239)	0.025*
Age	1.054(1.018-1.092)	0.003**	1.012(0.979-1.047)	0.481
Maximum tumor diameter (cm)	1.288(1.056-1.570)	0.012*	1.201(0.936-1.539)	0.149
Vascular invasion	2.494(1.005-6.192)	0.049*		
Time	2.600(0.982-6.886)	0.055		
Symptom	2.358(0.914-6.084)	0.076		

Note: HR represents hazard ratio, and CI represents 95% confidence interval. Significance levels are denoted as: * $P<0.05$, ** $P<0.01$, *** $P<0.001$.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.pdf