

Original Paper

Size-Specific Predictors for Malignancy Risk in Follicular Thyroid Neoplasms: Machine Learning Analysis

Xin Li^{1*}, MD; Wen-yu Yang^{2*}, BSc; Fan Zhang^{3*}, MD; Rui Shan⁴, MSc; Fang Mei⁵, PhD; Shi-Bing Song¹, MD; Bang-Kai Sun⁶, MD; Jing Chen⁴, BSc; Run-ze Hu⁴, BSc; Yang Yang⁴, BSc; Yi-hang Yang⁴, BSc; Jing-yao Liu⁴, BSc; Chun-Hui Yuan¹, MD; Zheng Liu⁴, PhD

¹Department of General Surgery, Peking University Third Hospital, Beijing, China

²China Center for Health Development Studies, Peking University, Beijing, China

³Department of Ultrasound, Peking University Third Hospital, Beijing, China

⁴Department of Maternal and Child Health, School of Public Health, Peking University, Beijing, China

⁵Department of Pathology, Peking University Third Hospital, School of Basic Medical Sciences, Peking University Health Science Center, Beijing, China

⁶Information Management and Big Data Center, Peking University Third Hospital, Beijing, China

*these authors contributed equally

Corresponding Author:

Zheng Liu, PhD

Department of Maternal and Child Health
School of Public Health, Peking University
38 Huayuan Road, Haidian District, Beijing
Beijing, 100191
China

Phone: 86 82801222

Email: liuzheng@bjmu.edu.cn

Abstract

Background: Surgeons often face challenges in distinguishing between benign and malignant follicular thyroid neoplasms (FTNs), particularly small tumors, until diagnostic surgery is performed.

Objective: This study aimed to identify the size-specific predictors for the malignancy risk of FTNs preoperatively.

Methods: A retrospective cohort study was conducted at Peking University Third Hospital in Beijing, China, from 2012 to 2023. Patients with a postoperative pathological diagnosis of follicular thyroid adenoma (FTA) or follicular thyroid carcinoma (FTC) were included. FTNs were classified into small- and large-sized categories based on the cutoff value of the tumor diameter derived from spline regression, which indicated the turning point of malignancy risk. We identified the 5 most important predictors from 22 variables including demography, sonography, and hormones, using machine learning methods. We also calculated the odds ratios (OR) with 95% CI for these predictors in both small- and large-sized FTNs.

Results: Altogether, we included 1494 FTNs, comprising 1266 FTAs and 228 FTCs. FTNs with a maximum diameter less than 3.0 cm were grouped as small-sized tumors (n=715), while those with larger diameters were categorized as large-sized tumors (n=779). In the small-sized group, tumors with macrocalcification (OR 2.90, 95% CI 1.50-5.60), those with peripheral calcification (OR 4.50, 95% CI 1.50-13.00), and those in younger patients (OR 1.33, 95% CI 1.05-1.69) showed a higher malignancy risk. In the large-sized group, tumors presenting with a nodule-in-nodule appearance (OR 3.30, 95% CI 1.30-7.90) exhibited a higher malignancy risk. In both groups, lower thyroid-stimulating hormone levels (OR 1.49, 95% CI 1.20-1.85 for small-sized FTNs; OR 1.61, 95% CI 1.37-1.96 for large-sized FTNs) and a larger mean diameter (OR 1.40, 95% CI 1.10-1.70 for small-sized FTNs; OR 1.50 95% CI 1.20-1.70 for large-sized FTNs) were associated with the malignancy risk of FTNs.

Conclusion: This study identified size-specific predictors for malignancy risk in FTNs, highlighting the importance of stratified prediction based on tumor size.

JMIR Cancer 2025;11:e73069; doi: [10.2196/73069](https://doi.org/10.2196/73069)

Keywords: follicular thyroid neoplasm; tumor size; machine learning; malignancy; follicular thyroid cancer; follicular thyroid adenoma; random forest; XGBoost

Introduction

In recent years, the incidence of thyroid cancer has been growing fast [1], and it is expected to continue to increase in a pronounced manner; the 5-year average annual percent change in the incidence of thyroid neoplasms from 1983 to 2017 in China was 7.82% in men and 8.59% in women [2]. Follicular thyroid neoplasms (FTNs) are one of the most important types of thyroid tumors in addition to papillary thyroid tumors; FTNs account for approximately 10-15% of all thyroid cancers [3]. Notably, FTNs are much more challenging for clinical management compared with papillary thyroid tumors [4,5]. This situation results from the fact that over 95% of FTNs cannot be accurately diagnosed as benign or malignant, regardless of using ultrasound, cytology, or biomarkers. Currently, clinicians often use diagnostic surgery to distinguish between benign and malignant FTNs. Still, this might lead to unnecessary diagnostic surgery for patients finally diagnosed with the benign type of FTN (ie, follicular thyroid adenoma (FTA)) or a second surgery after the initial diagnostic surgery for those finally diagnosed with the malignant type of FTN (ie, follicular thyroid carcinoma (FTC)) [6]. It is thus crucial to improve the accuracy for the prediction of the malignancy risk of FTNs prior to the diagnostic surgery, which could not only avoid the unnecessary diagnostic surgery for patients with FTA, but also provide timely clinical decisions for patients with FTC.

In the clinical context, three approaches are typically used to predict the malignancy risk of FTNs before the diagnostic surgery: cytology, biomarkers, and ultrasound. Concerning cytology, neither fine needle aspiration cytology nor core needle histopathology can reliably differentiate FTA from FTC. The reason is that the pathological diagnosis of FTN requires comprehensive sampling of the entire tumor following surgical dissection to determine the presence of capsular or vascular invasion across all tumor margins. Regarding the use of biomarkers, the routine clinical application, in most cases, is still hindered by the cost-ineffectiveness of testing, as well as suboptimal predictive performance in terms of sensitivity and specificity. Therefore,

the ultrasound examination of FTNs is of great significance in facilitating the evaluation of the malignancy risk of FTNs and the necessity for further diagnostic surgery.

Despite considerable efforts to preoperatively distinguish between FTA and FTC, research gaps remained, particularly concerning small-sized FTNs [7-10]. This is attributable to the fact that, on average, FTC exhibits a larger tumor diameter compared to FTA [10,11]; nevertheless, clinicians have reported the existence of small-sized FTC in routine clinical care [12]. If the suboptimal performance of existing prediction models for small-sized FTNs remains inadequately elucidated [13], there would continue to be a significant risk of misdiagnosis and undertreatment of small-sized FTNs. Consequently, it is imperative to identify important predictors (especially those from ultrasound examination) associated with the malignancy risk of both large- and small-sized FTNs.

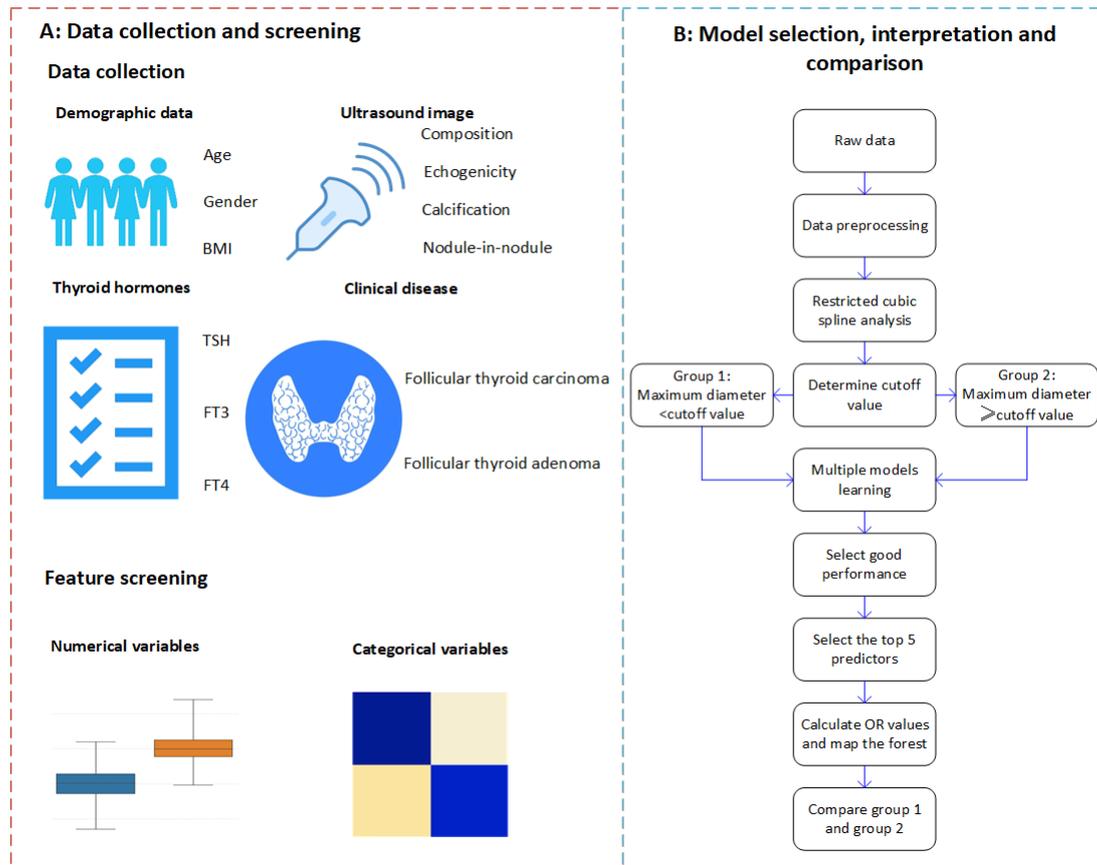
Our study aimed to (1) use machine learning to identify crucial predictors for the malignancy risk of both small- and large-sized FTNs, and (2) compare the differences in the direction and magnitude of predictors between the small- and large-sized FTNs. Findings from our study would facilitate the precision of differentiating between benign and malignant FTNs.

Methods

Overview

We reported this retrospective cohort study following the suggestion of the TRIPOD [14] (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis Statement), TRIPOD-AI [15] (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis-Artificial Intelligence Statement), and Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research [16] (Multimedia Appendix 1). The framework of the study is shown in Figure 1.

Figure 1. Framework of the study. FT3: free triiodothyronine; FT4: free thyroxine; TSH: thyroid-stimulating hormone.



Study Population

Our research team included multidisciplinary experts in the field of epidemiology, pathology, ultrasound, surgery, and endocrinology. We conducted a retrospective cohort study at Peking University Third Hospital in Beijing, China, from January 2012 to September 2023. The primary data source for this retrospective cohort study was the electronic medical record system of Peking University Third Hospital. Patients were included if they were pathologically diagnosed with FTC or FTA after surgical treatment; patients were excluded if they did not have ultrasound examinations before surgery. To ensure the accuracy of the pathological diagnosis of FTN, we invited the experienced pathologists to double-check all the pathological diagnoses of FTN based on the 2022, 5th edition WHO Classification of Thyroid Neoplasm [17].

Classification of FTNs Into Small- and Large-Sized Categories

We first scanned the entire dataset to identify and remove duplicate patient records by using a combination of unique identifiers such as patient ID, date of surgery, and specific pathological report numbers. For categorical variables with a relatively low proportion (5% or less) of missing values, we used imputation based on the majority class. For numerical variables such as patient age or tumor size, we used mean imputation. To classify the FTNs into small- and large-sized categories, we developed a restricted cubic spline model to identify the cutoff value. Specifically, we used the maximum tumor diameter as the continuous variable (predictor); the

malignancy risk of FTNs as the outcome; and the covariates included composition, echogenicity, margin, halo, taller-than-wide, calcification, internal blood flow, vascularity, trabecular formation, nodule-in-nodule appearance, mean diameter, thyroid-stimulating hormone (TSH), free triiodothyronine (FT3), free thyroxine (FT4), mean TSH score (interval-adjusted detailed TSH score [18]), tRMSSD of TSH (the time-adjusted root mean square of successive differences of TSH [19]), mean TSH (mean value of preoperative TSH), and coefficient of variation of TSH (the ratio of the SD of preoperative TSH to the mean value of preoperative TSH). To determine the optimal number of nodes for the restricted cubic splines, we employed the Akaike information criterion to strike the balance between the model goodness of fit and complexity that most effectively aligns with the data [20]. Finally, we divided FTNs into two groups: small- (maximum diameter less than the cutoff value) and large-sized (maximum diameter greater than the cutoff value) FTNs.

Predictors for the Malignancy Risk of FTNs

We selected the predictors for the malignancy risk of FTNs based on our domain knowledge [21,22] and data available. The predictors mainly included patients' gender, age, BMI, ultrasound features, thyroid hormones, and Hashimoto thyroiditis. To ensure the validity of measurements of predictors, both researchers and clinicians carefully checked the data source of the predictors.

Ultrasound features included composition (solid, predominantly solid, predominantly cystic, or cystic), echogenicity (hyperechoic, isoechoic, hypoechoic, or anechoic), margin (circumscribed, ill-defined, irregular, or lobulated), halo (absent halo, even thickness halo, present halo without evenness of thickness reported, or uneven thickness halo), taller-than-wide (absent or present), calcification (no echogenic foci, microcalcification, macrocalcification, peripheral calcification, microcalcification with comet-tail artifacts, or punctate echogenic foci of undetermined significance), internal blood flow (absent or present), vascularity (mainly central vascularity, mainly peripheral vascularity, mixed vascularity, or avascularity), trabecular formation (absent or present), nodule-in-nodule appearance (absent or present), and mean diameter.

The measurements of thyroid hormones included TSH, FT3, FT4, and TSH-related features. As listed in Table S1 in [Multimedia Appendix 2](#), TSH-related features included the mean TSH score, the tRMSSD of TSH, mean TSH, and coefficient of variation of TSH. The diagnostic criteria for Hashimoto thyroiditis referred to the ultrasonography describing the thyroid tissue as substantial diffuse lesions and the pathology report describing the thyroid tissue as Hashimoto thyroiditis.

Development and Validation of Machine Learning–Based Models

We established the machine learning–based models, as shown in [Figure 1](#). We trained eight classification models including logistic regression, lasso regression, weighted k-nearest neighbor, decision tree, random forest, naive bayes, XGBoost, and support vector machine (SVM). In total, 70% of the dataset was allocated for model training and selection, while the remaining 30% was reserved for internal validation. We developed and validated the models stepwise through predictor preprocessing, model training, hyperparameter tuning, and 5-fold cross-validation. It is important to note that our study population was from the real-world, naturally distributed population so that the outcome variable was slightly imbalanced (approximately 30% was FTC among all types of FTNs). We adopted the synthetic minority over-sampling technique (SMOTE), which increased the sample size of a few classes by creating new synthetic samples rather than simply copying existing ones [23]. We evaluated the model performance using accuracy, F1 score, the area under the receiver operating characteristic curve (AUROC), the area under the precision-recall curve (AUPRC), sensitivity, and specificity. We comprehensively considered the performance and interpretability of models and selected the most suitable model. We used the mlr3 [24] ecosystem in R 4.4.1 to conduct machine learning.

Comparison of Important Predictors for Malignancy Risk Between Small- and Large-Sized FTNs

We selected important predictors for small- and large-sized FTNs, respectively. We evaluated the importance

of predictors (feature importance) by computing the cross-entropy loss (loss: ce) of all features and visualized the importance of features. We identified the first five most important predictors that could both predict the outcome and not overlap with other predictors based on medical expertise and a novel information-gain approach. Based on the concept of entropy from information theory, the information gain approach is used to assess the extent to which features reduce uncertainty or increase the amount of information [25]. This approach helps to determine which features most effectively enhance classification accuracy by calculating the difference in information entropy before and after feature classification [25]. After the selection of important predictors, we calculated the odds ratio (OR) with 95% CI and drew the forest maps using multivariate logistic regression models.

Ethical Considerations

This study was approved by the Medical Research Ethics Committee of Peking University Third Hospital (No. IRB00006761- M2023168). As a retrospective analysis, the study was granted a waiver for additional informed consent. During the data extraction process, strict confidentiality measures were implemented to ensure patient privacy and data security. All extracted data were anonymized, with any information that could directly identify patients being removed.

Results

Characteristics of the Study Population and Selection of the Cut-Off Value for the Size of FTNs

Among the included 1494 patients, 1266 (84.7%) were diagnosed with FTA, and the remaining were diagnosed with FTC; the average (SD) age of the patients was 48.25 (0.75) years, and 1127/1494 patients (75.4%) were female. We used a restricted cubic spline model with three nodes that optimally balanced fitting the nonlinear relationship within the data while minimizing the overfitting risk. This analysis revealed a key turning point: the slope of the curve changed distinctly at a maximum diameter of 3 cm for FTNs ([Figure 2](#)), indicating that the influence of the maximum diameter on the risk of FTN malignancy shifted at this threshold. As such, 715/1494 tumors (47.9%) were classified as small-sized FTNs (maximum diameter <3 cm), while 779/1494 tumors (52.1%) were large-sized FTNs (maximum diameter ≥3 cm). The characteristics of the study population and FTNs are listed in [Tables 1](#) and [2](#), respectively.

Figure 2. Restricted cubic spline regression analysis of the association between the maximum diameter and the malignancy risk of follicular thyroid neoplasms (FTNs).

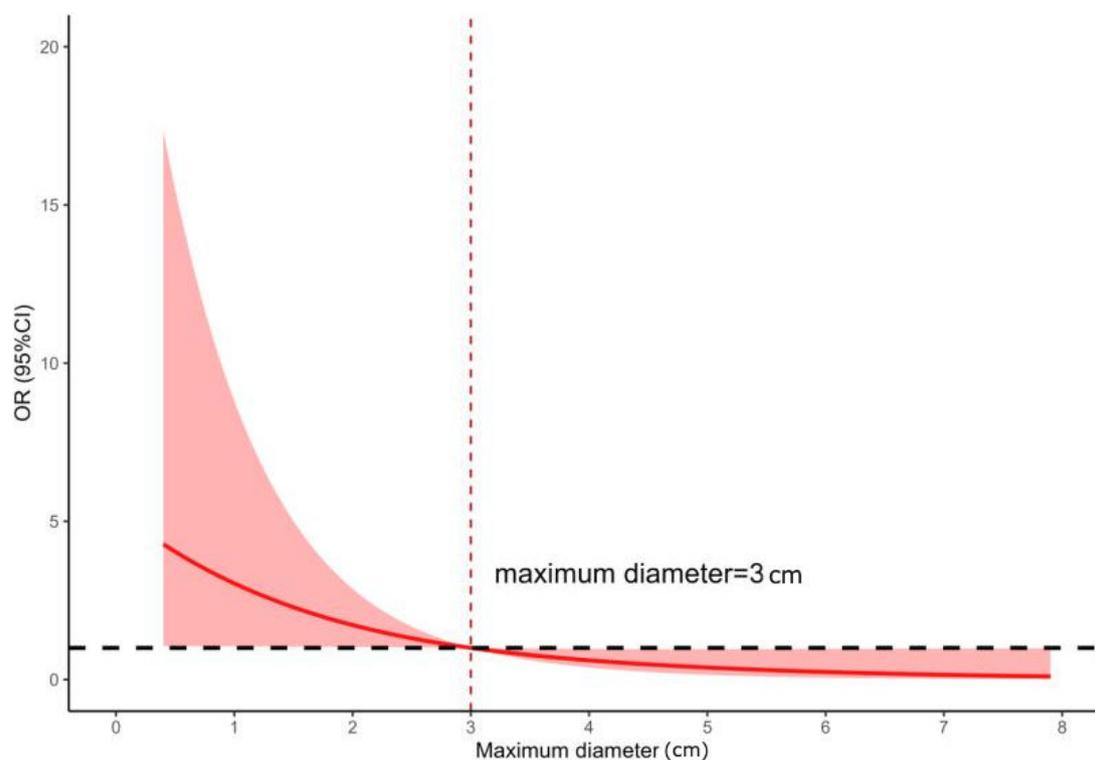


Table 1. Characteristics of the participants having FTNs^a with a maximum diameter of <3 cm.

Characteristics	FTA ^b (n=630)	FTC ^c (n=85)	Chi-square (df)	P value
Characteristics of the study population				
Female, n (%)	482 (76.5)	72 (84.7)	$\chi^2_{1}=2.44$.12 ^d
Age (years), mean (SD)	49.20 (13.32)	45.40 (14.75)	$\chi^2_{59}=63.61$.32 ^e
Hashimoto thyroiditis, n (%)	176 (27.9)	32 (37.6)	$\chi^2_{455}=458.55$.01 ^d
BMI (kg/m ²), mean (SD)	24.16 (3.57)	24.18 (3.52)	$\chi^2_{1}=2.44$.45 ^e
TSH ^f (μ IU/mL), mean (SD)	1.82 (1.14)	2.31 (1.76)	$\chi^2_{275}=278.65$.43 ^e
Mean TSH score ^g , mean (SD)	0.34 (1.19)	-0.26 (1.17)	$\chi^2_{563}=561.41$.51 ^e
tRMSSD of TSH ^h , mean (SD)	0.16 (0.65)	0.12 (0.23)	$\chi^2_{554}=550.51$.53 ^e
Mean TSH ⁱ , mean (SD)	3.48 (4.94)	2.72 (4.58)	$\chi^2_{533}=534.06$.48 ^e
Coefficient of variation of TSH ⁱ , mean (SD)	0.78 (0.52)	0.90 (0.56)	$\chi^2_{564}=562.36$.51 ^e
FT3 ^k (pg/mL), mean (SD)	3.23 (0.64)	3.29 (0.52)	$\chi^2_{152}=172.25$.13 ^e
FT4 ^l (ng/dL), mean (SD)	1.28 (0.19)	1.25 (0.22)	$\chi^2_{1}=2.44$.27 ^e
Characteristics of the FTNs, n (%)				
Composition				
Solid	365 (63.5)	50 (63.2)	$\chi^2_{3}=4.73$.19 ^m
Predominantly solid	159 (27.7)	27 (34.2)		
Predominantly cystic	39 (6.8)	1 (1.3)		
Cystic	12 (2.0)	1 (1.3)		
Echogenicity				
Anechoic	5 (0.9)	0 (0.0)	$\chi^2_{3}=12.71$.006 ⁿ
Hyperechoic	8 (1.4)	4 (4.8)		
Isoechoic	292 (50.1)	28 (33.3)		
Hypoechoic	278 (47.6)	52 (61.9)		
Margin				
Circumscribed	463 (78.7)	46 (56.1)	$\chi^2_{3}=25.07$	<.001 ^m

Characteristics	FTA ^b (n=630)	FTC ^c (n=85)	Chi-square (df)	P value
Ill-defined	22 (3.7)	4 (4.9)		
Irregular	72 (12.3)	18 (22.0)		
Lobulated	31 (5.3)	14 (17.0)		
Halo				
Absence of halo	285 (49.2)	36 (46.8)	$\chi^2_{3}=18.14$	<.001 ⁿ
Presence of halo, even thickness	194 (33.5)	20 (26.0)		
Presence of halo, evenness of thickness unknown	33 (5.7)	0 (0.0)		
Presence of halo, uneven thickness	67 (11.6)	21 (27.2)		
Taller-than-wide				
Absence	514 (90.2)	67 (84.8)	$\chi^2_{1}=2.13$.21 ^d
Presence	56 (9.8)	12 (15.2)		
Calcification				
No echogenic foci	488 (77.5)	52 (61.2)	$\chi^2_{5}=18.25$.004 ⁿ
Microcalcification	59 (9.4)	11 (12.9)		
Macrocalcification	56 (8.9)	15 (17.6)		
Peripheral calcification	12 (1.9)	6 (7.1)		
Microcalcification with comet-tail artifacts	8 (1.2)	1 (1.2)		
Punctate echogenic foci of undetermined significance	7 (1.1)	0 (0.0)		
Internal blood flow				
Absence	97 (16.1)	10 (12.2)	$\chi^2_{1}=0.83$.36 ^d
Presence	506 (83.9)	72 (87.8)		
Vascularity				
Mainly central vascularity	29 (7.2)	5 (7.9)	$\chi^2_{3}=6.44$.08 ⁿ
Mainly peripheral vascularity	2 (0.5)	0 (0.0)		
Mixed vascularity	219 (54.3)	44 (69.8)		
Avascularity	153 (38.0)	14 (22.3)		
Trabecular formation				
Absence	520 (98.3)	72 (96.0)	$\chi^2_{1}=0.80$.37 ^m
Presence	9 (1.7)	3 (4.0)		
Nodule-in-nodule appearance				
Absence	525 (99.2)	71 (94.7)	$\chi^2_{1}=7.32$.007 ^m
Presence	4 (0.8)	4 (5.3)		
Mean diameter (SD)	1.43 (0.54)	1.56 (0.60)		.09 ^e

^aFTNs: follicular thyroid neoplasms.

^bFTA: follicular thyroid adenoma.

^cFTC: follicular thyroid carcinoma.

^dUsed the Pearson χ^2 test.

^eUsed the Kruskal-Wallis test.

^fTSH: thyroid-stimulating hormone.

^gMean TSH score: standardized interval-adjusted detailed thyroid-stimulating hormone score.

^hTRMSSD of TSH: time-adjusted root mean square of successive differences of thyroid-stimulating hormone.

ⁱMean TSH: mean value of preoperative TSH.

^jCoefficient of variation of TSH: coefficient of variation of thyroid-stimulating hormone.

^kFT3: free triiodothyronine.

^lFT4: free thyroxine.

^mUsed the Pearson χ^2 test with the Yates continuity correction formula.

ⁿUsed the Fisher precision probability test.

Table 2. Characteristics of participants having FTNs^a with a maximum diameter of ≥ 3 cm.

Characteristics	FTA ^b (n=636)	FTC ^c (n=143)	Chi-square (<i>df</i>)	<i>P</i> value
Characteristics of the study population				
Female, n (%)	478 (75.2)	95 (66.4)	$\chi^2_{1}=1.81$.18 ^d
Age (years), mean (SD)	47.32 (14.81)	49.85 (15.01)	$\chi^2_{64}=65.58$.42 ^e
Hashimoto thyroiditis, n (%)	168 (26.4)	43 (30.1)	$\chi^2_{1}=2.96$.09 ^d
BMI (kg/m ²), mean (SD)	24.14 (3.78)	24.85 (3.90)	$\chi^2_{495}=477.31$.71 ^e
TSH ^f (μ IU/mL), mean (SD)	1.70 (2.30)	1.86 (1.42)	$\chi^2_{282}=303.13$.19 ^e
Mean TSH score ^g , mean (SD)	0.51 (1.06)	-0.27 (1.27)	$\chi^2_{625}=627.81$.46 ^e
tRMSSD of TSH ^h , mean (SD)	0.14 (0.41)	0.12 (0.26)	$\chi^2_{614}=617.36$.45 ^e
Mean TSH ⁱ , mean (SD)	3.52 (4.47)	2.62 (4.78)	$\chi^2_{575}=576.02$.48 ^e
Coefficient of variation of TSH ^j , mean (SD)	0.69 (0.46)	0.94 (0.55)	$\chi^2_{628}=628.80$.48 ^e
FT3 ^k (pg/mL), mean (SD)	3.31 (0.67)	3.33 (0.76)	$\chi^2_{170}=167.87$.53 ^e
FT4 ^l (ng/dL), mean (SD)	1.26 (0.22)	1.24 (0.30)	$\chi^2_{106}=106.28$.47 ^e
Characteristics of the FTNs, n (%)				
Composition				
Solid	220 (38.8)	73 (57.5)	$\chi^2_{3}=17.04$	<.001 ^m
Predominantly solid	239 (42.2)	43 (33.9)		
Predominantly cystic	105 (18.5)	11 (8.6)		
Cystic	3 (0.5)	0 (0.0)		
Echogenicity				
Anechoic	2 (0.4)	0 (0.0)	$\chi^2_{3}=8.11$.048 ^m
Hyperechoic	19 (3.6)	2 (1.5)		
Isoechoic	311 (58.1)	62 (47.7)		
Hypoechoic	203 (37.9)	66 (50.8)		
Margin				
Circumscribed	508 (87.6)	102 (75.0)	$\chi^2_{3}=21.10$	<.001 ⁿ
Ill-defined	14 (2.4)	1 (0.7)		
Irregular	28 (4.8)	17 (12.5)		
Lobulated	30 (5.2)	16 (11.8)		
Halo				
Absence of halo	244 (42.9)	55 (42.0)	$\chi^2_{3}=11.81$.008 ^d
Presence of halo, even thickness	214 (37.6)	37 (28.2)		
Presence of halo, evenness of thickness unknown	43 (7.5)	9 (6.9)		
Presence of halo, uneven thickness	68 (12.0)	30 (22.9)		
Taller-than-wide				
Absence	541 (85.1)	122 (85.3)	$\chi^2_{1}=2.70$.10 ^d
Presence	9 (1.4)	5 (3.5)		
Calcification				
No echogenic foci	528 (83.0)	103 (72.0)	$\chi^2_{5}=15.58$.008 ^m
Microcalcification	31 (4.9)	13 (9.1)		
Macrocalcification	54 (8.5)	22 (15.4)		
Peripheral calcification	2 (0.3)	2 (1.4)		
Microcalcification with comet-tail artifacts	13 (2.0)	3 (2.1)		
Punctate echogenic foci of undetermined significance	8 (1.3)	0 (0.0)		
Internal blood flow				
Absence	51 (8.4)	9 (6.6)	$\chi^2_{1}=0.50$.48 ^d
Presence	557 (91.6)	128 (93.4)		
Vascularity				

Characteristics	FTA ^b (n=636)	FTC ^c (n=143)	Chi-square (df)	P value
Mainly central vascularity	34 (5.9)	9 (6.6)	$\chi^2_{3}=7.13$.06 ⁿ
Mainly peripheral vascularity	2 (0.3)	0 (0.0)		
Mixed vascularity	308 (53.6)	89 (65.0)		
Avascularity	231 (40.2)	39 (28.4)		
Trabecular formation				
Absence	587 (96.7)	129 (92.1)	$\chi^2_{1}=5.95$.02 ^d
Presence	20 (3.3)	11 (7.9)		
Nodule-in-nodule appearance				
Absence	593 (97.7)	131 (93.6)	$\chi^2_{1}=6.48$.01 ^d
Presence	14 (2.3)	9 (6.4)		
Mean diameter (SD)	3.23 (0.85)	3.68 (1.09)	$\chi^2_{133}=143.19$.26 ^e

^aFTNs: follicular thyroid neoplasms.

^bFTA: follicular thyroid adenoma.

^cFTC: follicular thyroid carcinoma.

^dUsed the Pearson χ^2 test.

^eUsed the Kruskal-Wallis test.

^fTSH: thyroid-stimulating hormone.

^gMean TSH score: standardized interval-adjusted detailed thyroid-stimulating hormone score.

^htRMSSD of TSH: time-adjusted root mean square of successive differences of thyroid-stimulating hormone.

ⁱMean TSH: mean value of preoperative TSH.

^jCoefficient of variation of TSH: coefficient of variation of thyroid-stimulating hormone.

^kFT3: free triiodothyronine.

^lFT4: free thyroxine.

^mUsed the Fisher precision probability test.

ⁿUsed the Pearson χ^2 test with the Yates continuity correction formula.

Distinct Predictors for the Malignancy Risk in Small- and Large-Sized FTNs

We compared the performance in discrimination among the eight models (logistic regression, weighted k-nearest

neighbor, lasso regression, decision tree, random forest, naive bayes, XGBoost, and SVM). The XGBoost and random forest models performed broadly better in the small- and large-sized FTN groups, respectively (Table 3, Table S2 in Multimedia Appendix 2).

Table 3. Model performance in predicting the malignancy risk of small- and large-sized follicular thyroid neoplasms (FTNs).

Models	Small-sized FTNs (maximum diameter <3 cm)			Large-sized FTNs (maximum diameter ≥3 cm)		
	Accuracy	F1 score	AUPRC ^a	Accuracy	F1 score	AUPRC
Logistic regression	0.654	0.301	0.238	0.701	0.450	0.337
Weighted k-nearest neighbor	0.587	0.233	0.140	0.611	0.380	0.350
Lasso regression	0.752	0.299	0.227	0.669	0.411	0.368
Decision tree	0.614	0.205	0.134	0.646	0.381	0.274
Random forest	0.790	0.286	0.206	0.704	0.459	0.422
Naive bayes	0.577	0.220	0.163	0.623	0.297	0.251
XGBoost	0.811	0.330	0.248	0.695	0.403	0.380
SVM ^b	0.643	0.226	0.151	0.691	0.312	0.199

^aAUPRC: area under the precision-recall curve.

^bSVM: support vector machine.

In the small-sized FTN group, the top five predictors were the mean TSH score, tRMSSD of TSH, age at hospital admission, mean diameter, and calcification in the XGBoost model (Figure 3). Compared to tumors with no echogenic foci in the ultrasound image, the small-sized FTNs expressing microcalcification (OR 2.10, 95% CI 0.98-4.30), macrocalcification (OR 2.90, 95% CI 1.50-5.60), peripheral calcification (OR 4.50, 95% CI 1.50-13.00), and microcalcification with comet-tail artifacts (OR 1.60, 95% CI 0.09-9.50) had a higher risk of malignancy. Additionally, the risk of malignancy

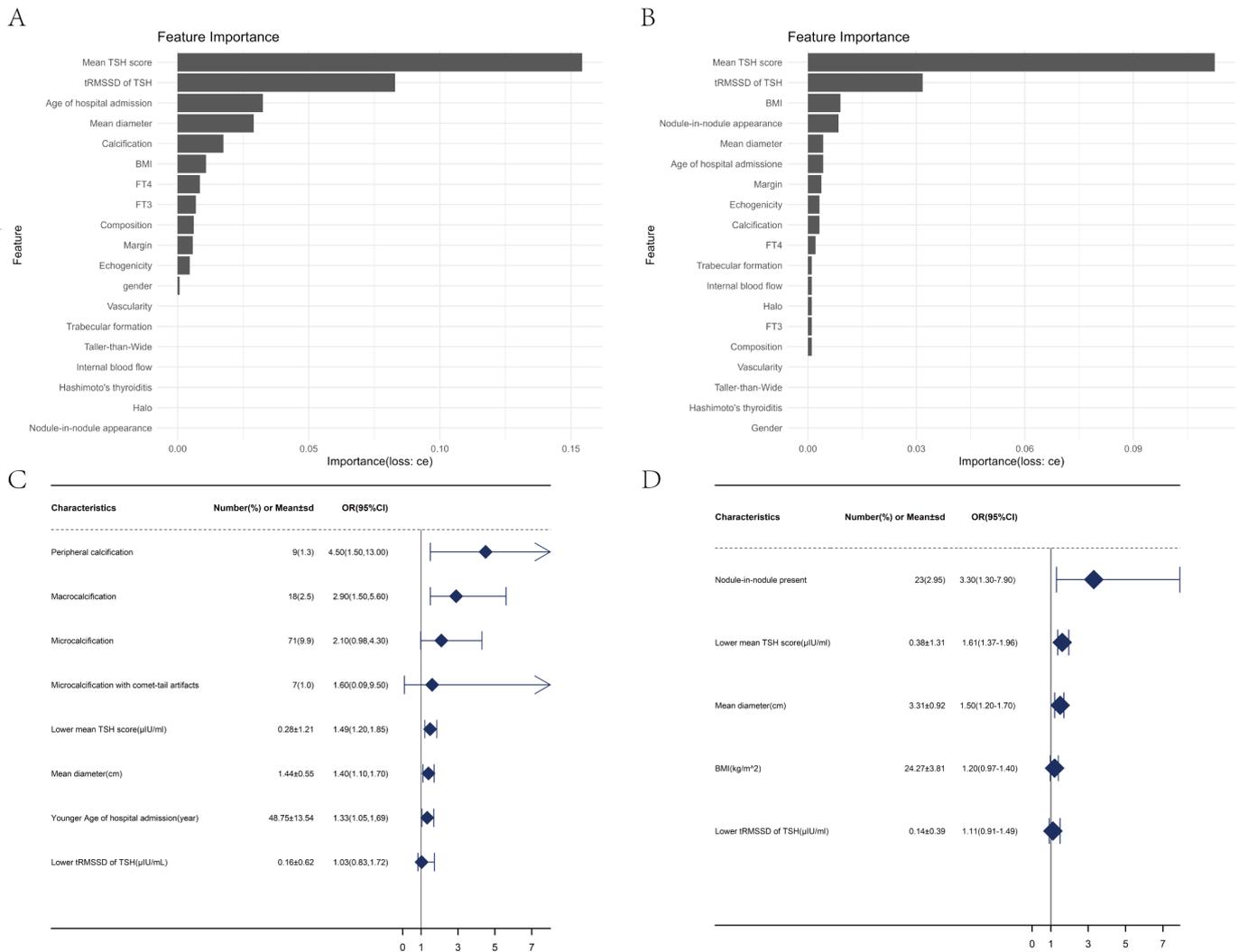
was higher in the patients with small-sized FTNs with a lower mean TSH score (OR 1.49, 95% CI 1.20-1.85), lower tRMSSD of TSH (OR 1.03, 95% CI 0.83-1.72), younger patients (OR 1.33, 95% CI 1.05-1.69), and larger mean diameter (OR 1.40, 95% CI 1.10-1.70).

By contrast, in the large-sized FTN group, the top 5 predictors were the mean TSH score, tRMSSD of TSH, BMI, nodule-in-nodule appearance, and mean diameter in the random forest model (Figure 3). The risk of malignancy was

higher in the large-sized FTN, which included the lower mean TSH score (OR 1.61, 95% CI 1.37-1.96), lower tRMSSD of TSH (OR 1.11, 95% CI 0.91-1.49), higher BMI (OR 1.20,

95% CI 0.97-1.40), the presence of nodule-in-nodule (OR 3.30, 95% CI 1.30-7.90), and larger mean diameter (OR 1.50, 95% CI 1.20-1.70).

Figure 3. The feature importance and forest map of the top 5 predictors for malignancy risk in small- and large-sized follicular thyroid neoplasms (FTNs). (A) The feature importance using the XGBoost model in small-sized follicular tumors, and loss: ce is the cross-entropy loss. (B) The feature importance using the random forest model in large-sized follicular tumors, and loss: ce is the cross-entropy loss. (C) The forest map of the top 5 predictors using multivariate logistic regression in small-sized follicular tumors, and the OR and 95% CI of microcalcification, macrocalcification, peripheral calcification, and microcalcification with comet-tail artifacts is based on no computed echogenic foci. (D) The forest map of the top 5 predictors using multivariate logistic regression in large-sized follicular tumors, and the OR and 95% CI of nodule-in-nodule presence is based on the computed nodule-in-nodule absence. FT3: free triiodothyronine; FT4: free thyroxine; tRMSSD: time-adjusted root mean square of successive differences; TSH: thyroid-stimulating hormone.



Discussion

Principal Findings

Taking advantage of a long-term cohort based on the real-world data, we found distinctive predictors for the malignancy risk of FTNs between the small- and large-sized tumors. Specifically, the tumor’s calcification appearance, mean diameter, and patients’ age were more important in predicting the malignancy risk of small-sized FTNs, whereas the tumor’s nodule-in-nodule appearance and patients’ BMI were more important in that of large-sized FTNs.

Comparison to Prior Work

It remained unclear regarding the cut-off value of tumor size that indicated the malignancy risk or surgical indications of FTNs. Concerning clinical guidelines, we observed that neither domestic nor international guidelines have recommended a cut-off value of tumor size for surgical treatment in FTNs [26-28]. Regarding existing research, among the 5 studies included in this review, the cut-off value of tumor size used to predict the malignancy risk of FTNs remained contradictor[29-33]. The included studies reported the associations between tumor size and the malignancy risk of FTNs, but all of them lacked a solid basis for their tumor-size classification criteria prior to surgical treatment.

Based on our literature review, previous studies on this topic often employed a one-size-fits-all model to predict the malignancy risk of FTN, without considering the tumor size [34,35]. For example, a retrospective multicenter study developed a prediction model for FTNs using the Thyroid Imaging Reporting and Data System (TI-RADS), and the validation dataset showed that the follicular TI-RADS model improved the performance in differentiating between FTA and FTC (AUROC 0.81) [34]. Additionally, a model combining the prior-based level set method with a deep convolutional neural network achieved an AUROC of 0.913 in distinguishing FTC from FTA using features extracted from ultrasound images [35]. Furthermore, we evaluated the follicular TI-RADS scoring criteria developed by Li et al [34] with our data and found a lower sensitivity (0.044 with a threshold for FTC risk set at >90%, and 0.269 with using a >50% FTC risk threshold) in evaluating the prediction model. These inconsistencies and the reduced sensitivity observed in external datasets may, at least in part, be attributed to the limitation of not considering tumor size when predicting the malignancy risk.

Our findings suggest that clinical differentiation between benign and malignant FTNs should be more precisely tailored according to the size of the FTNs. For example, the nodule-in-nodule appearance, a characteristic feature of FTC, was more frequently observed in larger FTNs [36]. This sign may reflect heterogeneous tumor cell proliferation, a phenomenon more commonly associated with larger tumors. Conversely, the incidence of this sign is relatively low in small-sized FTNs, complicating the identification of small-sized FTCs. Calcification, however, demonstrates a higher predictive value for malignancy in smaller FTNs. Our study focused on commonly used and easily obtainable predictors in clinical practice, including clinical features, TSH levels, and ultrasound characteristics. The limited number of predictors and the straightforward model design enhance its practicality for physicians. This research assists thyroid specialists in customizing predictor selection to assess the malignancy risk in FTNs of different sizes, ultimately improving the accuracy of FTC identification, patient outcomes, and quality of life while reducing postoperative complications.

Our finding also suggests that the mean TSH score ranked among the top 5 predictors for determining malignancy in tumors of both sizes. A cohort study from the EPIC (European Prospective Investigation into Cancer and Nutrition) cohort has revealed a negative association between elevated TSH levels and an increased risk of thyroid cancer [37]. Similarly, Gudmundsson et al [38] propose that low TSH levels may reduce the differentiation of thyroid epithelium, potentially increasing the predisposition to malignant cell transformation.

In evaluating the predictive performance of our machine learning models, we employed metrics distinct from those used in prior studies. For instance, Li et al [34] reported an AUROC of 0.76 for the LASSO regression model (the ratio of FTA to FTC: training set, 2.74; validation set, 3.70), and also in the LASSO regression model, the AUROC reached 0.913 for discriminating FTA from FTC [35] (the ratio of

FTA to FTC: 4.00). However, neither study incorporated the F1 score or AUPRC as evaluation metrics. Given the imbalanced nature of our data (FTA to FTC ratio: 7.00 for small follicular tumors and 5.00 for large follicular tumors), AUPRC provides a more informative and intuitive measure of model performance compared to AUROC [39]. Additionally, the F1 score offers a comprehensive evaluation by integrating precision and recall, ensuring a balanced assessment of model performance in the context of skewed category distributions [40].

Limitations and Strengths

Our study had several strengths. It was among the first to classify FTNs into two subgroups based on tumor diameter for machine learning analysis, revealing significant differences between these subgroups. Additionally, our models benefit from a large sample size, the use of clinically accessible and validated predictors, and a comprehensive evaluation using metrics appropriate for imbalanced data [41]. Importantly, the distribution of FTA and FTC in our study population was fully consistent with that of patients with FTNs in real-world settings, avoiding any exaggeration of sample sizes during model training.

However, our study had certain limitations. First, our classification of small- and large-sized FTNs based on a threshold of 3 cm was determined by our dataset rather than established guideline consensus, which may limit its generalizability to external datasets. We recommend that future clinical guidelines refine specific ultrasound risk indicators for FTNs based on the findings of this study and subsequent related research. Second, while we applied the SMOTE oversampling strategy to address imbalanced data, this approach may introduce bias when predicting new data [42]. Finally, the lack of external validation for our trained model limits our ability to assess its generalizability, potentially affecting its practical applicability and reliability in predicting new cases [43].

Future Directions

Our study had important clinical implications. In current clinical practice, it primarily relies on the ultrasonographic features of FTNs to assess the tumors' malignancy risk under the observation period (ie, before surgical treatment). It is thus crucial to emphasize the need for personalized predictive models in FTNs. Achieving this requires a more refined stratification of tumors based on the clinical characteristics of FTNs, which can enhance the customization of predictive models for tumors of individual cases. This tailored approach acknowledges the inherent variability within FTN categories, suggesting that further stratification within FTNs based on tumor size can facilitate more precise diagnostic and treatment decisions.

To advance this research, we intend to conduct a prospective study to validate the accuracy and reliability of the identified predictors in real-world clinical settings. The prospective evidence will enable updates and refinements to existing medical knowledge and practices for managing FTNs. For FTNs with tumor diameters <3 cm,

surgical decision-making should prioritize TSH levels, age at admission, tumor diameter, and calcification status. For FTNs ≥ 3 cm, TSH levels, BMI, nodule-in-nodule architecture, and tumor diameter should guide clinical recommendations. Furthermore, we advocate that professional medical societies and health care organizations collaboratively develop evidence-based practice guidelines integrating these predictors.

Conclusion

In our study, we identified differences in predictors among follicular tumors of varying sizes. Clinically, these findings

emphasize the importance of considering the size during the preoperative diagnosis of benign versus malignant FTNs. We found that both clinical guidelines and the existing research literature have not sufficiently addressed the optimal size of FTNs for surgical intervention or its correlation with malignancy risk. There is a significant research gap in precisely determining the preoperative size-based classification of FTNs. Thus, further investigations are imperative to address this knowledge deficit.

Acknowledgments

This work used the data of the Peking University Third Hospital. The authors sincerely thank the research team of the Thyroid Tumor Multi-Disciplinary Treatment. We appreciate the health professionals at Peking University Third Hospital for data collection and management. Funding support was received from the National Natural Science Foundation of China (grant no. 82373694), Young Elite Scientists Sponsorship Program by CAST (grant no. 2023QNRC001), Beijing Education Sciences Planning Program during the 14th Five-Year Plan (grant no. BECA23111), and Proof of Concept Program of Zhongguancun Science City and Peking University Third Hospital (grant no. HDCXZHKC2022210).

Data Availability

Data used in this study are available upon reasonable request.

Authors' Contributions

Study concept and design: ZL, FZ, XL

Data collection: XL, WYY, RS, FM, SBS, BKS, JC, RZH, YY, YHY, JYL, CHY, FZ

Data analysis and interpretation: XL, WYY, FZ, ZL

Writing the paper: WYY, ZL, XL, FZ

Critical revision of the paper: ZL

Conflicts of Interest

None declared.

Multimedia Appendix 1

Guidelines for developing and reporting machine learning predictive models in biomedical research.

[\[DOCX File \(Microsoft Word File\), 43 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Supplementary tables.

[\[DOCX File \(Microsoft Word File\), 18 KB-Multimedia Appendix 2\]](#)

Checklist 1

TRIPOD-AI Checklist.

[\[DOCX File \(Microsoft Word File\), 50 KB-Checklist 1\]](#)

References

1. Issa PP, Munshi R, Albuck AL, et al. Recommend with caution: a meta-analysis investigating papillary thyroid carcinoma tumor progression under active surveillance. *Am J Otolaryngol*. 2023;44(6):103994. [doi: [10.1016/j.amjoto.2023.103994](https://doi.org/10.1016/j.amjoto.2023.103994)] [Medline: [37607459](https://pubmed.ncbi.nlm.nih.gov/37607459/)]
2. Li M, Hu M, Jiang L, Pei J, Zhu C. Trends in cancer incidence and potential associated factors in China. *JAMA Netw Open*. Oct 1, 2024;7(10):e2440381. [doi: [10.1001/jamanetworkopen.2024.40381](https://doi.org/10.1001/jamanetworkopen.2024.40381)] [Medline: [39432306](https://pubmed.ncbi.nlm.nih.gov/39432306/)]
3. Jung CK, Bychkov A, Kakudo K. Update from the 2022 World Health Organization classification of thyroid tumors: a standardized diagnostic approach. *Endocrinol Metab*. Oct 2022;37(5):703-718. [doi: [10.3803/EnM.2022.1553](https://doi.org/10.3803/EnM.2022.1553)]
4. Podda M, Saba A, Porru F, Reccia I, Pisanu A. Follicular thyroid carcinoma: differences in clinical relevance between minimally invasive and widely invasive tumors. *World J Surg Oncol*. Jun 4, 2015;13(193):193. [doi: [10.1186/s12957-015-0612-8](https://doi.org/10.1186/s12957-015-0612-8)] [Medline: [26041024](https://pubmed.ncbi.nlm.nih.gov/26041024/)]

5. Brennan MD, Bergstralh EJ, van Heerden JA, McConahey WM. Follicular thyroid cancer treated at the Mayo Clinic, 1946 through 1970: initial manifestations, pathologic findings, therapy, and outcome. *Mayo Clin Proc.* Jan 1991;66(1):11-22. [doi: [10.1016/s0025-6196\(12\)61170-7](https://doi.org/10.1016/s0025-6196(12)61170-7)] [Medline: [1988751](https://pubmed.ncbi.nlm.nih.gov/1988751/)]
6. Grani G, Lamartina L, Durante C, Filetti S, Cooper DS. Follicular thyroid cancer and Hürthle cell carcinoma: challenges in diagnosis, treatment, and clinical management. *Lancet Diabetes Endocrinol.* Jun 2018;6(6):500-514. [doi: [10.1016/S2213-8587\(17\)30325-X](https://doi.org/10.1016/S2213-8587(17)30325-X)] [Medline: [29102432](https://pubmed.ncbi.nlm.nih.gov/29102432/)]
7. Seo JK, Kim YJ, Kim KG, Shin I, Shin JH, Kwak JY. Differentiation of the follicular neoplasm on the gray-scale US by image selection subsampling along with the marginal outline using convolutional neural network. *Biomed Res Int.* 2017;2017(3098293):3098293. [doi: [10.1155/2017/3098293](https://doi.org/10.1155/2017/3098293)] [Medline: [29527533](https://pubmed.ncbi.nlm.nih.gov/29527533/)]
8. Shin I, Kim YJ, Han K, et al. Application of machine learning to ultrasound images to differentiate follicular neoplasms of the thyroid gland. *Ultrasonography.* Jul 2020;39(3):257-265. [doi: [10.14366/usg.19069](https://doi.org/10.14366/usg.19069)] [Medline: [32299197](https://pubmed.ncbi.nlm.nih.gov/32299197/)]
9. Yang B, Yan M, Yan Z, Zhu C, Xu D, Dong F. Segmentation and classification of thyroid follicular neoplasm using cascaded convolutional neural network. *Phys Med Biol.* Dec 22, 2020;65(24):33137800. [doi: [10.1088/1361-6560/abc6f2](https://doi.org/10.1088/1361-6560/abc6f2)] [Medline: [33137800](https://pubmed.ncbi.nlm.nih.gov/33137800/)]
10. Savala R, Dey P, Gupta N. Artificial neural network model to distinguish follicular adenoma from follicular carcinoma on fine needle aspiration of thyroid. *Diagn Cytopathol.* Mar 2018;46(3):244-249. [doi: [10.1002/dc.23880](https://doi.org/10.1002/dc.23880)] [Medline: [29266871](https://pubmed.ncbi.nlm.nih.gov/29266871/)]
11. Lee SH, Baek JS, Lee JY, et al. Predictive factors of malignancy in thyroid nodules with a cytological diagnosis of follicular neoplasm. *Endocr Pathol.* Dec 2013;24(4):177-183. [doi: [10.1007/s12022-013-9263-x](https://doi.org/10.1007/s12022-013-9263-x)] [Medline: [24014040](https://pubmed.ncbi.nlm.nih.gov/24014040/)]
12. Huang CC, Hsueh C, Liu FH, Chao TC, Lin JD. Diagnostic and therapeutic strategies for minimally and widely invasive follicular thyroid carcinomas. *Surg Oncol.* Mar 2011;20(1):1-6. [doi: [10.1016/j.suronc.2009.06.006](https://doi.org/10.1016/j.suronc.2009.06.006)] [Medline: [19596568](https://pubmed.ncbi.nlm.nih.gov/19596568/)]
13. Shan R, Li X, Chen J, et al. Interpretable machine learning to predict the malignancy risk of follicular thyroid neoplasms in extremely unbalanced data: retrospective cohort study and literature review. *JMIR Cancer.* Feb 10, 2025;11:e66269. [doi: [10.2196/66269](https://doi.org/10.2196/66269)] [Medline: [39930991](https://pubmed.ncbi.nlm.nih.gov/39930991/)]
14. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ.* Jan 7, 2015;350:g7594. [doi: [10.1136/bmj.g7594](https://doi.org/10.1136/bmj.g7594)] [Medline: [25569120](https://pubmed.ncbi.nlm.nih.gov/25569120/)]
15. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* Apr 16, 2024;385:e078378. [doi: [10.1136/bmj-2023-078378](https://doi.org/10.1136/bmj-2023-078378)] [Medline: [38626948](https://pubmed.ncbi.nlm.nih.gov/38626948/)]
16. Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res.* Dec 16, 2016;18(12):e323. [doi: [10.2196/jmir.5870](https://doi.org/10.2196/jmir.5870)] [Medline: [27986644](https://pubmed.ncbi.nlm.nih.gov/27986644/)]
17. Baloch ZW, Asa SL, Barletta JA, et al. Overview of the 2022 WHO classification of thyroid neoplasms. *Endocr Pathol.* Mar 2022;33(1):27-63. [doi: [10.1007/s12022-022-09707-3](https://doi.org/10.1007/s12022-022-09707-3)] [Medline: [35288841](https://pubmed.ncbi.nlm.nih.gov/35288841/)]
18. Ito Y, Miyauchi A, Fujishima M, et al. Thyroid-stimulating hormone, age, and tumor size are risk factors for progression during active surveillance of low-risk papillary thyroid microcarcinoma in adults. *World J Surg.* Feb 2023;47(2):392-401. [doi: [10.1007/s00268-022-06770-z](https://doi.org/10.1007/s00268-022-06770-z)] [Medline: [36182976](https://pubmed.ncbi.nlm.nih.gov/36182976/)]
19. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Eur Heart J.* Mar 1996;17(3):354-381. [Medline: [8737210](https://pubmed.ncbi.nlm.nih.gov/8737210/)]
20. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.* Switzerland: Springer International Publishing; 2015. ISBN: 978-3-319-19424-0
21. Li X, Xiao WC, Mei F, et al. The association of pregnancy with disease progression in patients previously treated for differentiated thyroid cancer: a propensity score-matched retrospective cohort study. *J Womens Health (Larchmt).* Nov 2023;32(11):1174-1181. [doi: [10.1089/jwh.2023.0172](https://doi.org/10.1089/jwh.2023.0172)] [Medline: [37638829](https://pubmed.ncbi.nlm.nih.gov/37638829/)]
22. Xiao WC, Li X, Shan R, et al. Pregnancy and progression of differentiated thyroid cancer: a propensity score-matched retrospective cohort study. *J Clin Endocrinol Metab.* Feb 20, 2024;109(3):837-843. [doi: [10.1210/clinem/dgad557](https://doi.org/10.1210/clinem/dgad557)] [Medline: [37738427](https://pubmed.ncbi.nlm.nih.gov/37738427/)]
23. Dablain D, Krawczyk B, Chawla NV. DeepSMOTE: fusing deep learning and SMOTE for imbalanced data. *IEEE Trans Neural Netw Learn Syst.* Sep 2023;34(9):6390-6404. [doi: [10.1109/TNNLS.2021.3136503](https://doi.org/10.1109/TNNLS.2021.3136503)] [Medline: [35085094](https://pubmed.ncbi.nlm.nih.gov/35085094/)]
24. Bischl B, Sonabend R, Kotthoff L, Lang M, editors. Introduction and overview. In: *Applied Machine Learning Using Mlr3 in R.* CRC Press; 2024. [doi: [10.1201/9781003402848](https://doi.org/10.1201/9781003402848)]
25. Mukhopadhyay S. InfoGram and admissible machine learning. *Mach Learn.* Jan 2022;111(1):205-242. [doi: [10.1007/s10994-021-06121-4](https://doi.org/10.1007/s10994-021-06121-4)]

26. Chinese Anti-Cancer Association Thyroid Cancer Committee. China anti-cancer association guidelines for holistic integrative management of thyroid cancer (2022 abridged version). *Chinese Journal of Clinical Oncology*. 2023;325-330. [doi: [10.12354/j.issn.1000-8179.2023.20221410](https://doi.org/10.12354/j.issn.1000-8179.2023.20221410)]
27. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid*. Jan 2016;26(1):1-133. [doi: [10.1089/thy.2015.0020](https://doi.org/10.1089/thy.2015.0020)]
28. Haddad RI, Bischoff L, Ball D, et al. Thyroid carcinoma, version 2.2022, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. Aug 2022;20(8):925-951. [doi: [10.6004/jnccn.2022.0040](https://doi.org/10.6004/jnccn.2022.0040)] [Medline: [35948029](https://pubmed.ncbi.nlm.nih.gov/35948029/)]
29. Lubitz CC, Faquin WC, Yang J, et al. Clinical and Cytological Features Predictive of Malignancy in Thyroid Follicular Neoplasms. *Thyroid*. Jan 2010;20(1):25-31. [doi: [10.1089/thy.2009.0208](https://doi.org/10.1089/thy.2009.0208)]
30. 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*. Sep 2015;22(9):3007-3013. [doi: [10.1245/s10434-014-4324-z](https://doi.org/10.1245/s10434-014-4324-z)]
31. Kamran SC, Marqusee E, Kim MI, et al. Thyroid Nodule Size and Prediction of Cancer. *The Journal of Clinical Endocrinology & Metabolism*. Feb 2013;98(2):564-570. [doi: [10.1210/jc.2012-2968](https://doi.org/10.1210/jc.2012-2968)]
32. Boonrod A, Akkus Z, Castro MR, et al. Thyroid Nodule Size as a Predictor of Malignancy in Follicular and Hurthle Neoplasms. *Asian Pac J Cancer Prev*. Aug 1, 2021;22(8):2597-2602. [doi: [10.31557/APJCP.2021.22.8.2597](https://doi.org/10.31557/APJCP.2021.22.8.2597)]
33. Kuru B, Kefeli M. Risk factors associated with malignancy and with triage to surgery in thyroid nodules classified as Bethesda category IV (FN/SFN). *Diagn Cytopathol*. Jun 2018;46(6):489-494. URL: <https://onlinelibrary.wiley.com/doi/10.1097/0339/46/6> [doi: [10.1002/dc.23923](https://doi.org/10.1002/dc.23923)]
34. Li J, Li C, Zhou X, et al. US risk stratification system for follicular thyroid neoplasms. *Radiology*. Nov 2023;309(2):e230949. [doi: [10.1148/radiol.230949](https://doi.org/10.1148/radiol.230949)] [Medline: [37987664](https://pubmed.ncbi.nlm.nih.gov/37987664/)]
35. Tang P, Ren C, Shen L, Zhou Z. Development and validation of a diagnostic nomogram for the preoperative differentiation between follicular thyroid carcinoma and follicular thyroid adenomas. *J Comput Assist Tomogr*. 2021;45(1):128-134. [doi: [10.1097/RCT.0000000000001078](https://doi.org/10.1097/RCT.0000000000001078)] [Medline: [33475318](https://pubmed.ncbi.nlm.nih.gov/33475318/)]
36. Kuo TC, Wu MH, Chen KY, Hsieh MS, Chen A, Chen CN. Ultrasonographic features for differentiating follicular thyroid carcinoma and follicular adenoma. *Asian J Surg*. Jan 2020;43(1):339-346. [doi: [10.1016/j.asjsur.2019.04.016](https://doi.org/10.1016/j.asjsur.2019.04.016)] [Medline: [31182260](https://pubmed.ncbi.nlm.nih.gov/31182260/)]
37. Rinaldi S, Plummer M, Biessy C, et al. Thyroid-stimulating hormone, thyroglobulin, and thyroid hormones and risk of differentiated thyroid carcinoma: the EPIC study. *J Natl Cancer Inst*. Jun 2014;106(6):2482-2491. [doi: [10.1093/jnci/dju097](https://doi.org/10.1093/jnci/dju097)] [Medline: [24824312](https://pubmed.ncbi.nlm.nih.gov/24824312/)]
38. Gudmundsson J, Sulem P, Gudbjartsson DF, et al. Discovery of common variants associated with low TSH levels and thyroid cancer risk. *Nat Genet*. Jan 22, 2012;44(3):319-322. [doi: [10.1038/ng.1046](https://doi.org/10.1038/ng.1046)] [Medline: [22267200](https://pubmed.ncbi.nlm.nih.gov/22267200/)]
39. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*. 2015;10(3):e0118432. [doi: [10.1371/journal.pone.0118432](https://doi.org/10.1371/journal.pone.0118432)] [Medline: [25738806](https://pubmed.ncbi.nlm.nih.gov/25738806/)]
40. Fan Z, Zhang X, Ruan S, et al. A medical image classification method based on self-regularized adversarial learning. *Med Phys*. Nov 2024;51(11):8232-8246. [doi: [10.1002/mp.17320](https://doi.org/10.1002/mp.17320)] [Medline: [39078069](https://pubmed.ncbi.nlm.nih.gov/39078069/)]
41. Hashimoto DA, Varas J, Schwartz TA. Practical guide to machine learning and artificial intelligence in surgical education research. *JAMA Surg*. Apr 1, 2024;159(4):455-456. [doi: [10.1001/jamasurg.2023.6687](https://doi.org/10.1001/jamasurg.2023.6687)] [Medline: [38170510](https://pubmed.ncbi.nlm.nih.gov/38170510/)]
42. Lee YW, Choi JW, Shin EH. Machine learning model for predicting malaria using clinical information. *Comput Biol Med*. Feb 2021;129(104151):104151. [doi: [10.1016/j.compbiomed.2020.104151](https://doi.org/10.1016/j.compbiomed.2020.104151)] [Medline: [33290932](https://pubmed.ncbi.nlm.nih.gov/33290932/)]
43. Ong Ly C, Unnikrishnan B, Tadic T, et al. Shortcut learning in medical AI hinders generalization: method for estimating AI model generalization without external data. *NPJ Digit Med*. May 14, 2024;7(1):124. [doi: [10.1038/s41746-024-01118-4](https://doi.org/10.1038/s41746-024-01118-4)] [Medline: [38744921](https://pubmed.ncbi.nlm.nih.gov/38744921/)]

Abbreviations

- AUPRC:** area under precision-recall curve
- EPIC:** European Prospective Investigation into Cancer and Nutrition
- FTA:** follicular thyroid adenoma
- FTC:** follicular thyroid carcinoma
- FTNs:** follicular thyroid neoplasms
- OR:** odds ratio
- SMOTE:** synthetic minority over-sampling technique
- SVM:** support vector machine
- TI-RADS:** thyroid imaging reporting and data system

tRMSSD: time-adjusted root mean square of successive difference

TSH: thyroid-stimulating hormone

Edited by Naomi Cahill; peer-reviewed by Chandrashekar Br, Chinonso Gerald Udensi, Temitope Adebambo; submitted 25.02.2025; final revised version received 19.05.2025; accepted 19.05.2025; published 11.07.2025

Please cite as:

Li X, Yang WY, Zhang F, Shan R, Mei F, Song SB, Sun BK, Chen J, Hu RZ, Yang Y, Yang YH, Liu JY, Yuan CH, Liu Z
Size-Specific Predictors for Malignancy Risk in Follicular Thyroid Neoplasms: Machine Learning Analysis

JMIR Cancer 2025;11:e73069

URL: <https://cancer.jmir.org/2025/11/e73069>

doi: [10.2196/73069](https://doi.org/10.2196/73069)

© Xin Li, Wen-yu Yang, Fan Zhang, Rui Shan, Fang Mei, Shi-Bing Song, Bang-Kai Sun, Jing Chen, Run-ze Hu, Yang Yang, Yi-hang Yang, Jing-yao Liu, Chun-Hui Yuan, Zheng Liu. Originally published in JMIR Cancer (<https://cancer.jmir.org>), 11.07.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Cancer, is properly cited. The complete bibliographic information, a link to the original publication on <https://cancer.jmir.org/>, as well as this copyright and license information must be included.