Size-specific predictors for malignancy risk in follicular thyroid neoplasms: A machine-learning analysis and literature review

Wenyu Yang

Submitted to: JMIR Cancer
on: February 25, 2025

Table of Contents

Size-specific predictors for malignancy risk in follicular thyroid neoplasms: A machine-learning analysis and literature review

Wenyu Yang[1]

[1] Department of Maternal and Child Health, School of Public Health, Peking University Beijing CN

Corresponding Author:
Wenyu Yang

Department of Maternal and Child Health, School of Public Health, Peking University
Haidian No 38, Xueyuan Road, Haidian District Beijing
Beijing
CN

Abstract

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

Objective: This study aimed to identify the size-specific predictors for 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 carcinoma (FTC) were included. FTNs were classified into small- and large-sized categories based on the cutoff value of 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 odds ratios (OR) with 95% confidence intervals (CI) for these predictors in both small- and large-sized FTNs. To synthesize existing evidence on this topic, a literature review of clinical guidelines and research papers was conducted.

Results: Altogether, we included 1494 FTNs, comprising 1266 FTAs and 228 FTCs. FTNs with a maximum diameter smaller than 3.0 cm were grouped as small-sized (n = 715), while those with larger diameters were categorized as large-sized (n = 779). In the small-sized group, tumors appearing macrocalcification [OR (95%CI): 2.90 (1.50, 5.60)], peripheral calcification [4.50 (1.50, 13.00)], or in younger patients [1.33 (1.05, 1.69)] showed a higher malignancy risk. In the large-sized group, tumors presenting a nodule-in-nodule appearance [3.30 (1.30,7.90)] exhibited a higher malignancy risk. In both groups, lower TSH levels [small-sized FTNs: 1.49 (1,20, 1,85); large-sized FTNs: 1.61 (1.37, 1.96)] and larger mean diameter [small-sized FTNs: 1.40 (1.10, 1.70); large-sized FTNs: 1.50 (1.20, 1.70)] were associated with the malignancy risk of FTNs. Our review identified a research gap in using tumor size thresholds as a stratification factor for assessing malignancy risk in FTNs before diagnostic surgery, which is not addressed by current clinical guidelines or existing literature.

Conclusions: This study identified size-specific predictors for malignancy risk in FTNs, highlighting the importance of stratified prediction based on tumor size. Clinical Trial: This study has been registered publicly.

(JMIR Preprints 25/02/2025:73069)
DOI: https://doi.org/10.2196/preprints.73069

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
No. Please do not make my accepted manuscript PDF available to anyone.

Original Manuscript

Size-specific predictors for malignancy risk in follicular thyroid neoplasms: A machine-learning analysis and literature review

Author list:

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[#]

[*]These authors contributed equally as co-first authors.

[#]These authors contributed equally as corresponding authors.

Corresponding to:

Prof. Zheng Liu, Department of Maternal and Child Health, School of Public Health, Peking University; E-mail: liuzheng@bjmu.edu.cn; ORCID: https://orcid.org/0000-0002-0405-2348; Phone: 010-82801222.

Author affiliations:

Department of General Surgery, Peking University Third Hospital (Xin Li, Shi-Bing Song, Chun-Hui Yuan); China Center for Health Development Studies, Peking University (Wen-Yu Yang); Department of Maternal and Child Health, School of Public Health, Peking University (Rui Shan, Jing Chen, Run-Ze Hu, Yang Yang, Yi-Hang Yang, Jing-Yao Liu, Zheng Liu); Department of Ultrasound, Peking University Third Hospital (Fan Zhang); Department of Pathology, Peking University Third Hospital (Fang Mei); Information Management and Big Data Center, Peking University Third Hospital (Bang-Kai Sun)

Abstract

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

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

Materials and 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 carcinoma (FTC) were included. FTNs were classified into small- and large-sized categories based on the cutoff value of 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 odds ratios (OR) with 95% confidence intervals (CI) for these predictors in both small- and large-sized FTNs. To synthesize existing evidence on this topic, a literature review of clinical guidelines and research papers was conducted.

Results: Altogether, we included 1494 FTNs, comprising 1266 FTAs and 228 FTCs. FTNs with a maximum diameter smaller than 3.0 cm were grouped as small-sized (n = 715), while those with larger diameters were categorized as large-sized (n = 779). In the small-sized group, tumors appearing macrocalcification [OR (95%CI): 2.90 (1.50, 5.60)], peripheral calcification [4.50 (1.50, 13.00)], or in younger patients [1.33 (1.05, 1.69)] showed a higher malignancy risk. In the large-sized group, tumors presenting a nodule-in-nodule appearance [3.30 (1.30,7.90)] exhibited a higher malignancy risk. In both groups, lower TSH levels [small-sized FTNs: 1.49 (1,20, 1,85); large-sized FTNs: 1.61 (1.37, 1.96)] and larger mean diameter [small-sized FTNs: 1.40 (1.10, 1.70); large-sized FTNs: 1.50 (1.20, 1.70)] were associated with the malignancy risk of FTNs. Our review identified a research gap in using tumor size thresholds as a stratification factor for assessing malignancy risk in FTNs before diagnostic surgery, which is not addressed by current clinical guidelines or existing literature.

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

Keywords: Follicular thyroid neoplasm; Tumor size; Machine learning; Malignancy; Literature review.

Introduction

In recent years, the incidence of thyroid cancer has been fast growing[1], and expected to continue to increase in a pronounced manner[2]. Follicular thyroid neoplasms (FTNs) are one important type of thyroid tumors in addition to papillary thyroid tumors. Notably, FTNs are much more challenging for clinical management compared with papillary thyroid tumors[3, 4]. This situation results from the fact that over 95% of FTNs cannot be accurately diagnosed as benign or malignancy type, regardless of the approaches of ultrasound, cytology, or biomarkers. Currently, clinicians often use diagnostic surgery to distinguish between benign and malignancy type of FTNs. Still, this might lead to unnecessary diagnostic surgery for those finally diagnosed as the benign type of FTN [i.e., follicular thyroid adenoma (FTA)] or a second surgery after initial diagnostic surgery for those finally diagnosed as the malignant type of FTN [i.e., follicular thyroid carcinoma (FTC)][5]. It is thus crucial to improve the accuracy for the prediction of 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 decision for patients with FTC.

In the clinical context, three approaches are typically employed to predict the malignancy risk of FTNs prior to the diagnostic surgery: cytology, biomarkers, and ultrasound. Concerning cytology, neither the fine needle aspiration cytopathology nor the core needle histopathology can reliably differentiate FTA from FTC. The reason is that the pathological diagnosis of FTN requires a comprehensive sampling of the entire tumor following surgical dissection to determine the presence of capsular or vascular invasion across all tumor margins. Regarding the biomarker, its 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, ultrasound examination of FTNs is of great significance in facilitating the evaluation of the malignant risk of FTNs and the necessity for further diagnostic surgery.

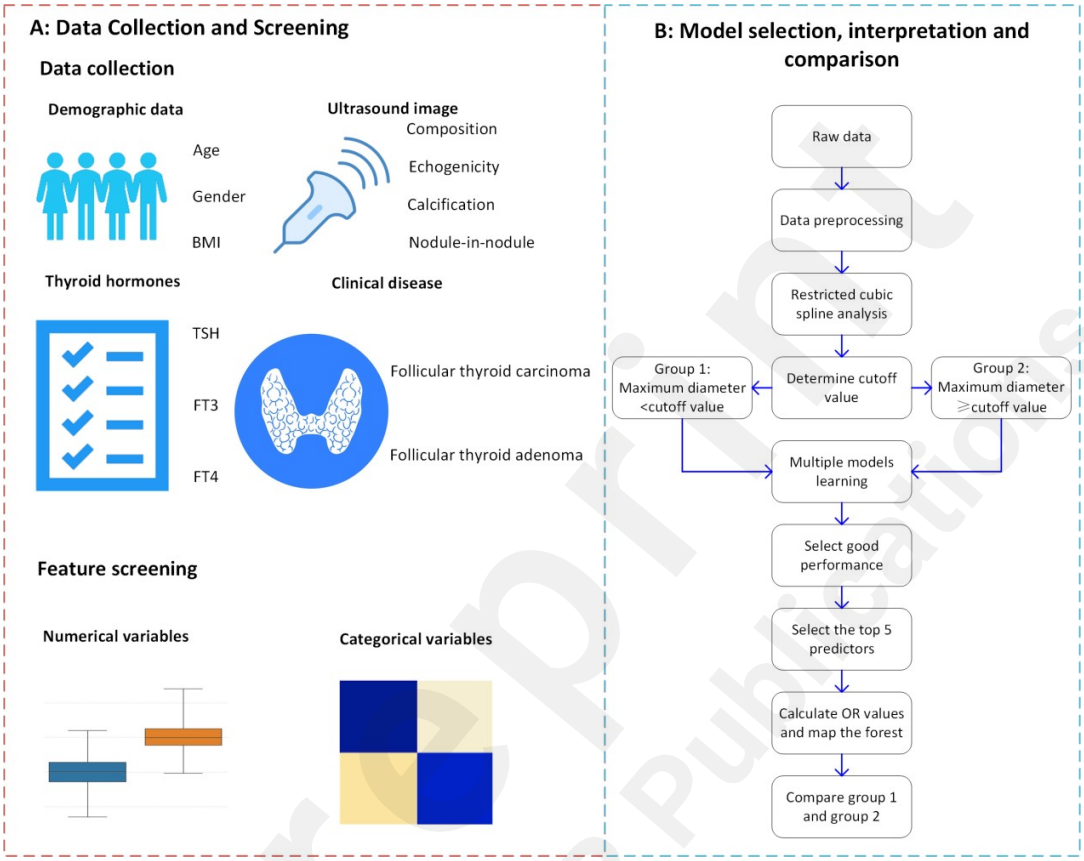
Despite considerable efforts to preoperatively distinguish between FTA and FTC, research gaps remained, particularly concerning small-sized FTNs[6-9]. This is attributable to the fact that, on average, FTC exhibits a larger tumor diameter compared to FTA[9, 10]; nevertheless, clinicians have reported the existence of small-sized FTC in routine clinical care[11]. If the suboptimal performance of existing prediction models for small-sized FTNs remains inadequately elucidated[12], there would continue to be a significant risk of misdiagnosis and undertreatment of small-sized FTN. 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) utilize 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

We reported this retrospective cohort study following the suggestion of the TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis Statement[13]. Framework of the study is shown in Figure 1.

Figure 1 Framework of the study



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. Patients were included if they were pathologically diagnosed as 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[14]. This study was approved by the Medical Research Ethics Committee of Peking University Third Hospital (No. IRB00006761-M2023168).

Classification of FTNs into small- and large-sized categories

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[15]), tRMSSD of TSH (the time-adjusted root mean square of successive differences of TSH[16]), mean TSH (mean value of preoperative TSH), and coefficient of variation of TSH (the ratio of standard deviation of preoperative TSH to mean value of preoperative TSH). To determine the optimal number of nodes for the restricted cubic splines, we employed the Akaike information criterion (AIC) to strike the balance between model goodness of fit and complexity that most effectively aligns with the data[17]. Finally, we divided FTNs into two groups: small- (maximum diameter < cutoff value) and large-sized (maximum diameter ≥ cutoff value) FTNs.

Predictors for the malignancy risk of FTN

We selected the predictors for the malignancy risk of FTN based on our domain knowledge [18, 19] and data available. The predictors mainly included patients' gender, age, body mass index (BMI), ultrasound features, thyroid hormones, and Hashimoto's 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 anechoric), 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 Thyroid-stimulating hormone (TSH), FT3, FT4, and TSH related features. As listed in Supplemental Table 1, TSH related features included standardized interval-adjusted detailed thyroid-stimulating hormone score (mean TSH score), the time-adjusted root mean square of successive differences of thyroid-stimulating hormone(tRMSSD of TSH), mean value of preoperative thyroid-stimulating hormone (mean TSH), and coefficient of variation of thyroid-stimulating hormone (coefficient of variation of TSH). The diagnostic criteria for Hashimoto's thyroiditis referred to the ultrasonography describing the thyroid tissue as substantial diffuse lesion and the pathology report describing the thyroid tissue as Hashimoto's thyroiditis.

Development and validation of machine-learning-based models

We established the machine-learning-based models as shown in Figure 1. We trained 8 classification models including logistic regression, lasso regression, weighted k-nearest neighbor, decision tree, random forest, naive bayes, XGBoost, and SVM. We stepwise developed and validated the models 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 FTN). 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[20]. 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 mlr3[21] 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[22]. This approach helps to determine which features most effectively enhance classification accuracy by calculating the difference in information entropy before and after feature classification[22]. After selection of important predictors, we calculated the odds ratio (OR) with 95% confidence intervals (CI) and drew the forest maps using multivariate logistic regression models.

Review for clinical guidelines and research papers

We conducted a systematic review of both clinical guidelines and research papers of the present topic. According to the PRISMA guidelines[23], we searched PubMed, Web of Science, and Embase using the terms “follicular thyroid cancer”, “follicular adenocarcinoma”, “tumor size” and “predict” for articles published up to December 25, 2024. Eligible studies included those that tumor size was among the preoperative predictors for malignancy risk of FTNs. Studies were excluded if predictors of FTNs did not incorporate tumor size or if the articles were not written in English.

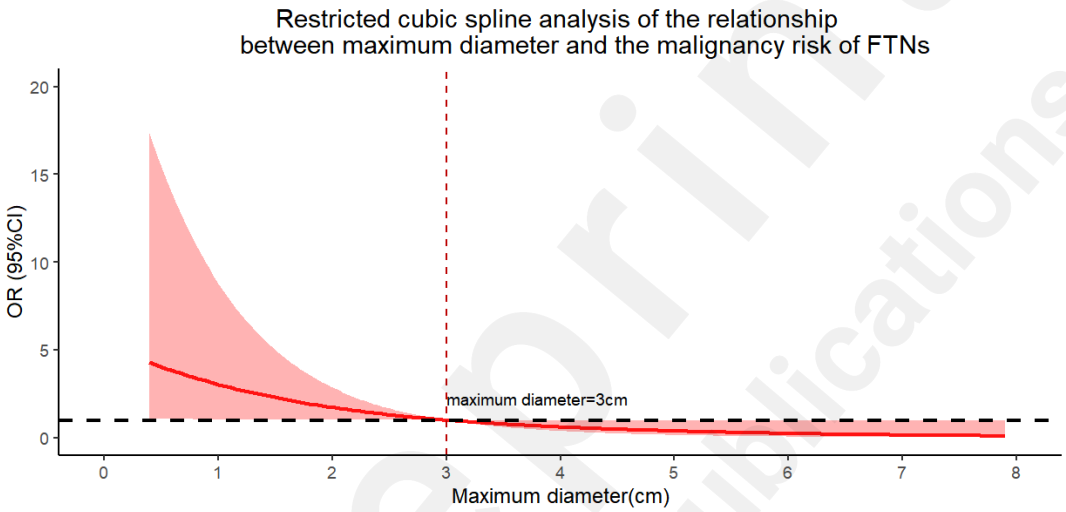
Results

Selection of cut-off value for the size of FTNs

Based on the AIC criteria, we employed a restricted cubic spline model with three nodes that

optimally balanced fitting the nonlinear relationship within the data while minimizing the overfitting risk. Analysis of the restricted cubic splines revealed a key turning point: the slope of the curve changes distinctly at a maximum diameter of 3.0 cm for FTNs (Figure 2), indicating that the influence of maximum diameter on the risk of FTN malignancy shifts at this threshold.

Figure 2 Restricted cubic spline regression analysis of the association between maximum diameter and the malignancy risk of FTNs



Characteristics of the study population

Among the included 1494 patients, 1266 (84.7 %) of them were diagnosed as FTA and the remaining was diagnosed as FTC. Of them, 715 (47.9 %) were classified as small-sized FTNs (maximum diameter < 3.0cm) while 779 (52.1%) were large-sized FTNs (maximum diameter ≥ 3.0cm). The characteristics of the study population and FTNs are listed in Table 1 and Table 2, respectively.

Table 1 Characteristics of the study population

Characteristics	Participants having FTNs[a] with maximum diameter < 3.0 cm			Participants having FTNs[a] with maximum diameter ≥ 3.0cm		
	FTA[b] (630 cases)	FTC[c] (85 cases)	*P* value	FTA[b] (636 cases)	FTC[c] (143 cases)	*P* value
Female, n (%)	482 (77.2)	72 (84.7)	0.155[d]	478 (75.4)	95 (69.9)	0.217[d]
Age (year)	49.20±13.32	45.40±14.75	0.318[e]	47.32±14.81	49.85±15.01	0.422[e]
Hashimoto's Thyroiditis, n (%)	176 (29.5)	32 (43.8)	**0.018**[d]	168 (28.0)	43 (35.8)	0.107[d]
BMI[f] (kg/m^2)	24.16±3.57	24.18±3.52	0.445[e]	24.14±3.78	24.85±3.90	0.708[d]
TSH[g] (μIU/mL)	1.82±1.14	2.31±1.76	0.427[e]	1.70±2.30	1.86±1.42	0.185[e]
Mean TSH score[h]	0.34±1.19	-0.26±1.17	0.511[e]	0.51±1.06	-0.27±1.27	0.461[e]
TRMSSD of TSH[i]	0.16±0.65	0.12±0.23	0.534[e]	0.14±0.41	0.12±0.26	0.454[e]
Mean TSH[j]	3.48±4.94	2.72±4.58	0.479[e]	3.52±4.47	2.62±4.78	0.480[e]
Coefficient of variation of TSH[k]	0.78±0.52	0.90±0.56	0.512[e]	0.69±0.46	0.94±0.55	0.484[e]
FT3[l] (pg/mL)	3.23±0.64	3.29±0.52	0.125[e]	3.31±0.67	3.33±0.76	0.532[e]
FT4[m] (ng/dL)	1.28±0.19	1.25±0.22	0.273[e]	1.26±0.22	1.24±0.30	0.474[e]

[a]FTNs: Follicular thyroid neoplasms

[b]FTA: Follicular thyroid adenoma

[c]FTC: Follicular thyroid carcinoma

[d]: Used Kruskal-Wallis test

[e]: Used Pearson Chi-square test

[f]BMI: Body mass index

[g]TSH: Thyroid-stimulating hormone

[h]Mean TSH score: Standardized interval-adjusted detailed thyroid-stimulating hormone score

[i]TRMSSD of TSH: The time-adjusted root mean square of successive differences of thyroid-stimulating hormone

[j]Mean TSH: Mean value of preoperative TSH

[k]Coefficient of variation of TSH: Coefficient of variation of thyroid-stimulating hormone

[l]FT3: Free triiodothyronine

[m]FT4: Free thyroxine

Table 2 Characteristics of the small- and large-sized follicular thyroid neoplasms

Characteristics		FTNs[a] with maximum diameter < 3.0cm, n (%)			FTNs[a] with maximum diameter ≥ 3.0cm, n (%)		
		FTA[b] (630 cases)	FTC[c] (85 cases)	*P* value	FTA[b] (636 cases)	FTC[c] (143 cases)	*P* value
Composition	Solid	365 (63.5)	50 (63.2)	0.193[d]	220 (38.8)	73 (57.5)	**<0.001**[e]
	Predominantly solid	159 (27.7)	27 (34.2)		239 (42.2)	43 (33.9)	
	Predominantly cystic	39 (6.8)	1 (1.3)		105 (18.5)	11 (8.6)	
	Cystic	12 (2.0)	1 (1.3)		3 (0.5)	0 (0.0)	
Echogenicity	Anechoic	5 (0.9)	0 (0.0)	**0.006**[e]	2 (0.4)	0 (0.0)	**0.048**[e]
	Hyperechoic	8 (1.4)	4 (4.8)		19 (3.6)	2 (1.5)	
	Isoechoic	292 (50.1)	28 (33.3)		311 (58.1)	62 (47.7)	
	Hypoechoic	278 (47.6)	52 (61.9)		203 (37.9)	66 (50.8)	
Margin	Circumscribed	463 (78.7)	46 (56.1)	**<0.001**[d]	508 (87.6)	102 (75.0)	**<0.001**[d]
	Ill-defined	22 (3.7)	4 (4.9)		14 (2.4)	1 (0.7)	
	Irregular	72 (12.3)	18 (22.0)		28 (4.8)	17 (12.5)	
	Lobulated	31 (5.3)	14 (17.0)		30 (5.2)	16 (11.8)	
Halo	Absence of halo	285 (49.2)	36 (46.8)	**<0.001**[e]	244 (42.9)	55 (42.0)	**0.008**[f]
	Presence of halo, even thickness	194 (33.5)	20 (26.0)		214 (37.6)	37 (28.2)	
	Presence of halo, evenness of thickness unknown	33 (5.7)	0 (0.0)		43 (7.5)	9 (6.9)	

	Presence of halo, uneven thickness	67 (11.6)	21 (27.2)		68 (12.0)	30 (22.9)	

Characteristics		FTNs[a] with maximum diameter < 3.0cm, n (%)			FTNs[a] with maximum diameter ≥ 3.0cm, n (%)		
		FTA[b] (630 cases)	FTC[c] (85 cases)	*P* value	FTA[b] (636 cases)	FTC[c] (143 cases)	*P* value
Taller-than-Wide	Absence	514 (90.2)	67 (84.8)	0.207[f]	541 (85.1)	122 (85.3)	0.101[f]
	Presence	56 (9.8)	12 (15.2)		9 (1.4)	5 (3.5)	
Calcification	No echogenic foci	488 (77.5)	52 (61.2)	**0.004**[e]	528 (83.0)	103 (72.0)	**0.008**[e]
	Microcalcification	59 (9.4)	11 (12.9)		31 (4.9)	13 (9.1)	
	Macrocalcification	56 (8.9)	15 (17.6)		54 (8.5)	22 (15.4)	
	Peripheral calcification	12 (1.9)	6 (7.1)		2 (0.3)	2 (1.4)	
	Microcalcification with comet-tail artifacts	8 (1.2)	1 (1.2)		13 (2.0)	3 (2.1)	
	Punctate echogenic foci of undetermined significance	7 (1.1)	0 (0.0)		8 (1.3)	0 (0.0)	
Internal blood flow	Absence	97 (16.1)	10 (12.2)	0.362[f]	51 (8.4)	9 (6.6)	0.480[f]
	Presence	506 (83.9)	72 (87.8)		557 (91.6)	128 (93.4)	
Vascularity	Mainly central vascularity	29 (7.2)	5 (7.9)	0.079[e]	34 (5.9)	9 (6.6)	0.064[d]
	Mainly peripheral vascularity	2 (0.5)	0 (0.0)		2 (0.3)	0 (0.0)	
	Mixed vascularity	219 (54.3)	44 (69.8)		308 (53.6)	89 (65.0)	

	Avascularity	153 (38.0)	14 (22.3)		231 (40.2)	39 (28.4)	
Trabecular formation	Absence	520 (98.3)	72 (96.0)	0.372[d]	587 (96.7)	129 (92.1)	**0.015**[f]
	Presence	9 (1.7)	3 (4.0)		20 (3.3)	11 (7.9)	
Characteristics		FTNs[a] with maximum diameter < 3.0cm, n (%)			FTNs[a] with maximum diameter ≥ 3.0cm, n (%)		
		FTA[b] (630 cases)	FTC[c] (85 cases)	*P* value	FTA[b] (636 cases)	FTC[c] (143 cases)	*P* value
Nodule-in-nodule appearance	Absence	525 (99.2)	71 (94.7)	**0.007**[d]	593 (97.7)	131 (93.6)	**0.011**[f]
	Presence	4 (0.8)	4 (5.3)		14 (2.3)	9 (6.4)	
Mean diameter (cm)	Mean ± SD[h]	1.43±0.54	1.56±0.60	0.086[g]	3.23±0.85	3.68±1.09	0.258[g]

[a]FTNs: Follicular thyroid neoplasms

[b]FTA: Follicular thyroid adenoma

[c]FTC: Follicular thyroid carcinoma

[d]: Used Pearson Chi-square test using Yates continuity correction formula

[e]: Used Fisher's precision probability test

[f]: Used Kruskal-Wallis test

[g]: Used Pearson Chi-square test

[h]SD: Standard deviation

Distinct predictors for malignancy risk in small- and large-sized FTNs

We compared performance in discrimination among 8 models (logistic regression, weighted k-nearest neighbor, lasso regression, decision tree, random forest, naive bayes, XGBoost, SVM). The XGBoost and random forest model performed broadly better in the small-, and large-sized FTN groups, respectively (Table 3, Supplemental Table 2).

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

Models	Small-sized FTNs[a] (maximum diameter < 3.0cm)			Large-sized FTNs[a] (maximum diameter≥ 3.0cm)		
	Accuracy	F1 score	AUPRC[b]	Accuracy	F1 score	AUPRC[b]
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	0.643	0.226	0.151	0.691	0.312	0.199

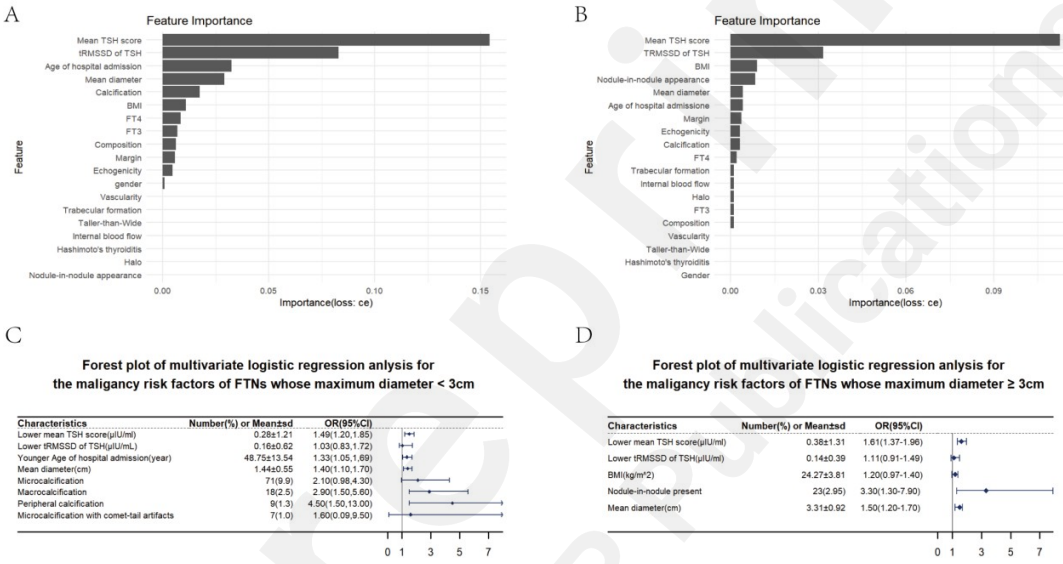
[a]FTNs: Follicular thyroid neoplasms

[b]AUPRC: Area under precision-recall curve

In the small-sized FTN group, the top five predictors were mean TSH score, tRMSSD of TSH, age of hospital admission, mean diameter, and calcification in the XGBoost model (Figure 3). Comparing with no echogenic foci in the ultrasonic image, the small-sized FTNs expressing microcalcification [OR (95%CI): 2.10 (0.98, 4.30)], macrocalcification [2.90 (1.50, 5.60)], peripheral calcification [4.50 (1.50, 13.00)], and microcalcification with comet-tail artifacts [1.60 (0.09, 9.50)] had a higher risk of malignancy. Additionally, the risk of malignancy was higher in the small-sized FTNs with lower mean TSH score [1.49 (1.20, 1.85)], lower tRMSSD of TSH [1.03 (0.83, 1.72)], younger patients [1.33 (1.05, 1.69)], and larger mean diameter [1.40 (1.10, 1.70)].

By contrast, in the large-sized FTN group, the top 5 predictors were 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 performed lower mean TSH score [OR (95%CI): 1.61 (1.37, 1.96)], lower tRMSSD of TSH [1.11 (0.91, 1.49)], higher BMI [1.20 (0.97, 1.40)], the presence of nodule-in-nodule [3.30 (1.30, 7.90)], and larger mean diameter [1.50 (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 FTNs



Notes: **A**: the feature importance using XGBoost model in small follicular tumors, and loss: ce is the cross-entropy loss. **B:** the feature importance using random forest model in large follicular tumors, and loss: ce is the cross-entropy loss. C, the forest map of top 5 predictors using multivariate logistic regression in small follicular tumors, and the OR and 95%CI of Microcalcification, Macrocalcification, Peripheral calcification, and Microcalcification with comet-tail artifacts Is based on No echogenic foci compute. D, the forest map of top 5 predictors using multivariate logistic regression in large follicular tumors, and the OR and 95%CI of nodule-in-nodule presence is based on nodule-in-nodule absence compute.

Review results for clinical guidelines and research papers

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[24-26]. 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 contradictory. Four studies[27-30] employed 4 cm as the discriminatory cutoff, while one study[31] used 2 cm. Four of the five included studies reported the associations between tumor size and malignancy risk of FTNs, but all of them lacked a solid basis for their tumor - size classification criteria prior to surgical treatment.

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 FTN between the small- and large-sized ones. Specifically, tumor's calcification appearance, mean diameter, and patients' age were more important in predicting the malignancy risk of small-sized FTNs, whereas tumor's nodule-in-nodule appearance and patients' BMI were more important in that of large-sized FTNs.

Comparison to Prior Work

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 tumor size. 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 performance in differentiating between FTA and FTC (AUROC: 0.81)[32]. 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[33]. Furthermore, we evaluated the follicular TI-RADS scoring criteria developed by Li at al.[32] 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 valuating 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 malignancy risk.

Our finding suggests 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[34]. 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 FTNs, complicating the identification of small FTCs. Calcification, however, demonstrate 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 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[35]. Similarly, Gudmundsson et al. proposed that low TSH levels may reduce the differentiation of thyroid epithelium, potentially increasing the predisposition to malignant cell transformation[36].

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.[32] 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 [33] (the ratio of FTA to FTC: 4.00). However, neither study

incorporated 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[37]. 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[38].

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[39]. Importantly, the distribution of FTA and FTC in our study population was fully consistent with that of FTN patients 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 FTNs based on a threshold of 3.0 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 [40]. 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[41].

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 observation period (i.e., 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 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 FTNs categories, suggesting that further stratification within FTNs based on tumor size can facilitate more precise diagnostic and treatment decisions.

Conclusion

In our study, we identified differences in predictors among follicular tumors of varying sizes. Clinically, these findings emphasize 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 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 (82373694), Young Elite Scientists Sponsorship Program by CAST (2023QNRC001), Beijing Education Sciences Planning Program during the 14th Five-Year Plan (BECA23111), and Proof of Concept Program of Zhongguancun Science City and Peking University Third Hospital (HDCXZHKC2022210).

Data Availability

Data used in this study is available upon reasonable request.

Authors' Contributions

Study concept and design: Zheng Liu, Fan Zhang, Xin Li

Data collection: Xin Li, Wen-Yu Yang, 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, Fan Zhang

Data analysis and interpretation: Xin Li, Wen-Yu Yang, Fan Zhang, Zheng Liu

Writing the paper: Wen-Yu Yang, Zheng Liu, Xin Li, Fan Zhang

Critical revision of the paper: Zheng Liu

Conflicts of Interest

None declared.

Abbreviations:

AIC: Akaike information criterion
AUPRC: Area under precision-recall curve
BMI: Body mass index
CI: confidence intervals
Coefficient of variation of TSH: Coefficient of variation of thyroid-stimulating hormone
EPIC: European Prospective Investigation into Cancer and Nutrition
FTA: follicular thyroid adenoma
FTC: follicular thyroid carcinoma
FTNs: follicular thyroid neoplasms
Mean TSH: Mean value of preoperative TSH
Mean TSH score: Standardized interval-adjusted detailed thyroid-stimulating hormone score
OR: odds ratios
SD: Standard deviation
SMOTE: Synthetic Minority Over-sampling Technique
TI-RADS: Thyroid Imaging Reporting and Data System
TRMSSD of TSH: The time-adjusted root mean square of successive differences of thyroid-stimulating hormone
TSH: Thyroid-stimulating hormone

References

1. Issa PP, Munshi R, Albuck AL, Omar M, Abu Alhuda RF, Metz T, et al. Recommend with caution: A meta-analysis investigating papillary thyroid carcinoma tumor progression under active surveillance. Am J Otolaryngol. 2023 Nov-Dec;44(6):103994. PMID: 37607459. doi: 10.1016/j.amjoto.2023.103994.
2. Li M, Hu M, Jiang L, Pei J, Zhu C. Trends in Cancer Incidence and Potential Associated Factors in China. JAMA Netw Open. 2024 Oct 1;7(10):e2440381. PMID: 39432306. doi: 10.1001/jamanetworkopen.2024.40381.
3. 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. 2015 Jun 4;13:193. PMID: 26041024. doi: 10.1186/s12957-015-0612-8.
4. 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. 1991 Jan;66(1):11-22. PMID: 1988751. doi: 10.1016/s0025-6196(12)61170-7.
5. 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. 2018 Jun;6(6):500-14. PMID: 29102432. doi: 10.1016/s2213-8587(17)30325-x.
6. 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. PMID: 29527533. doi: 10.1155/2017/3098293.
7. Shin I, Kim YJ, Han K, Lee E, Kim HJ, Shin JH, et al. Application of machine learning to ultrasound images to differentiate follicular neoplasms of the thyroid gland. Ultrasonography. 2020 Jul;39(3):257-65. PMID: 32299197. doi: 10.14366/usg.19069.
8. 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. 2020 Dec 22;65(24):245040. PMID: 33137800. doi: 10.1088/1361-6560/abc6f2.
9. 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. 2018 Mar;46(3):244-9. PMID: 29266871. doi: 10.1002/dc.23880.
10. Lee SH, Baek JS, Lee JY, Lim JA, Cho SY, Lee TH, et al. Predictive factors of malignancy in thyroid nodules with a cytological diagnosis of follicular neoplasm. Endocr Pathol. 2013 Dec;24(4):177-83. PMID: 24014040. doi: 10.1007/s12022-013-9263-x.
11. Huang CC, Hsueh C, Liu FH, Chao TC, Lin JD. Diagnostic and therapeutic strategies for minimally and widely invasive follicular thyroid carcinomas. Surg Oncol. 2011 Mar;20(1):1-6. PMID: 19596568. doi: 10.1016/j.suronc.2009.06.006.
12. Shan R LX, Zhang F, Liu Z, er al. Interpretable machine learning to predict the malignancy risk of follicular thyroid neoplasms in the extremely unbalanced data:

Experiences from a real-world study and literature review. JMIR Preprints. 2024 09/09:66269. doi: doi: 10.2196/66269.
13. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Bmj. 2015 Jan 7;350:g7594. PMID: 25569120. doi: 10.1136/bmj.g7594.
14. 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 Mar;33(1):27-63. PMID: 35288841. doi: 10.1007/s12022-022-09707-3.
15. Ito Y, Miyauchi A, Fujishima M, Noda T, Sano T, Sasaki T, 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. 2023 Feb;47(2):392-401. PMID: 36182976. doi: 10.1007/s00268-022-06770-z.
16. 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. 1996 Mar;17(3):354-81. PMID: 8737210.
17. FE H. Regression modeling strategies: with applications to linear models,logistic regression, and survival analysis. Switzerland: Springer International Publishing; 2015. ISBN: 978-3-319-19424-0.
18. Li X, Xiao WC, Mei F, Shan R, Song SB, Sun BK, 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). 2023 Nov;32(11):1174-81. PMID: 37638829. doi: 10.1089/jwh.2023.0172.
19. Xiao WC, Li X, Shan R, Mei F, Song SB, Chen J, et al. Pregnancy and Progression of Differentiated Thyroid Cancer: A Propensity Score-Matched Retrospective Cohort Study. J Clin Endocrinol Metab. 2024 Feb 20;109(3):837-43. PMID: 37738427. doi: 10.1210/clinem/dgad557.
20. Dablain D, Krawczyk B, Chawla NV. DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data. IEEE Trans Neural Netw Learn Syst. 2023 Sep;34(9):6390-404. PMID: 35085094. doi: 10.1109/tnnls.2021.3136503.
21. Bischl LKaRSaNFaB. Introduction and Overview. In: Lang BBaRSaLKaM, editor. Applied Machine Learning Using mlr3 in R: CRC Press; 2024.
22. Mukhopadhyay S. InfoGram and admissible machine learning. Machine Learning. 2022 2022/01/01;111(1):205-42. doi: 10.1007/s10994-021-06121-4.
23. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. Bmj. 2009 Jul 21;339:b2700. PMID: 19622552. doi: 10.1136/bmj.b2700.
24. Oncology CA-CACoT. China Anti-Cancer Association guidelines for holistic integrative management of thyroid cancer (2022 abridged version). CHINESE JOURNAL OF CLINICAL ONCOLOGY. 2023:325-30. doi: 10.12354/j.issn.1000-8179.2023.20221410.

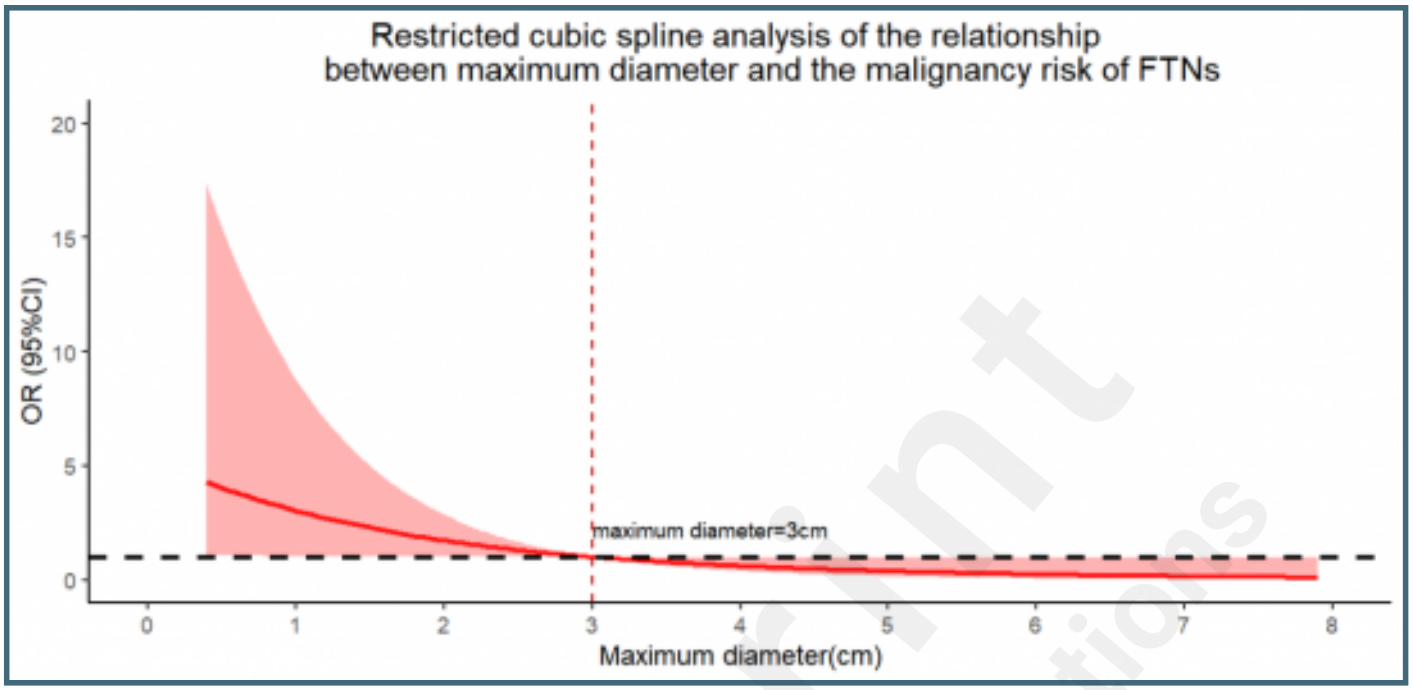
25. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid. 2016 Jan;26(1):1-133. PMID: 26462967. doi: 10.1089/thy.2015.0020.
26. Haddad RI, Bischoff L, Ball D, Bernet V, Blomain E, Busaidy NL, et al. Thyroid Carcinoma, Version 2.2022, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw. 2022 Aug;20(8):925-51. PMID: 35948029. doi: 10.6004/jnccn.2022.0040.
27. Boonrod A, Akkus Z, Castro MR, Zeinodini A, Philbrick K, Stan M, et al. Thyroid Nodule Size as a Predictor of Malignancy in Follicular and Hurthle Neoplasms. Asian Pac J Cancer Prev. 2021 Aug 1;22(8):2597-602. PMID: 34452575. doi: 10.31557/apjcp.2021.22.8.2597.
28. 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. 2018 Jun;46(6):489-94. PMID: 29524316. doi: 10.1002/dc.23923.
29. 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 Sep;22(9):3007-13. PMID: 25564170. doi: 10.1245/s10434-014-4324-z.
30. Lubitz CC, Faquin WC, Yang J, Mekel M, Gaz RD, Parangi S, et al. Clinical and cytological features predictive of malignancy in thyroid follicular neoplasms. Thyroid. 2010 Jan;20(1):25-31. PMID: 20025540. doi: 10.1089/thy.2009.0208.
31. Kamran SC, Marqusee E, Kim MI, Frates MC, Ritner J, Peters H, et al. Thyroid nodule size and prediction of cancer. J Clin Endocrinol Metab. 2013 Feb;98(2):564-70. PMID: 23275525. doi: 10.1210/jc.2012-2968.
32. Li J, Li C, Zhou X, Huang J, Yang P, Cang Y, et al. US Risk Stratification System for Follicular Thyroid Neoplasms. Radiology. 2023 Nov;309(2):e230949. PMID: 37987664. doi: 10.1148/radiol.230949.
33. 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 Jan-Feb 01;45(1):128-34. PMID: 33475318. doi: 10.1097/rct.0000000000001078.
34. 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. 2020 Jan;43(1):339-46. PMID: 31182260. doi: 10.1016/j.asjsur.2019.04.016.
35. Rinaldi S, Plummer M, Biessy C, Tsilidis KK, Østergaard JN, Overvad K, et al. Thyroid-stimulating hormone, thyroglobulin, and thyroid hormones and risk of differentiated thyroid carcinoma: the EPIC study. J Natl Cancer Inst. 2014 Jun;106(6):dju097. PMID: 24824312. doi: 10.1093/jnci/dju097.
36. Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Masson G, He H, et al. Discovery of common variants associated with low TSH levels and thyroid cancer risk. Nat Genet. 2012 Jan 22;44(3):319-22. PMID: 22267200. doi: 10.1038/ng.1046.

37. 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. PMID: 25738806. doi: 10.1371/journal.pone.0118432.
38. Fan Z, Zhang X, Ruan S, Thorstad W, Gay H, Song P, et al. A medical image classification method based on self-regularized adversarial learning. Med Phys. 2024 Nov;51(11):8232-46. PMID: 39078069. doi: 10.1002/mp.17320.
39. Hashimoto DA, Varas J, Schwartz TA. Practical Guide to Machine Learning and Artificial Intelligence in Surgical Education Research. JAMA Surg. 2024 Apr 1;159(4):455-6. PMID: 38170510. doi: 10.1001/jamasurg.2023.6687.
40. Lee YW, Choi JW, Shin EH. Machine learning model for predicting malaria using clinical information. Comput Biol Med. 2021 Feb;129:104151. PMID: 33290932. doi: 10.1016/j.compbiomed.2020.104151.
41. Ong Ly C, Unnikrishnan B, Tadic T, Patel T, Duhamel J, Kandel S, et al. Shortcut learning in medical AI hinders generalization: method for estimating AI model generalization without external data. NPJ Digit Med. 2024 May 14;7(1):124. PMID: 38744921. doi: 10.1038/s41746-024-01118-4.

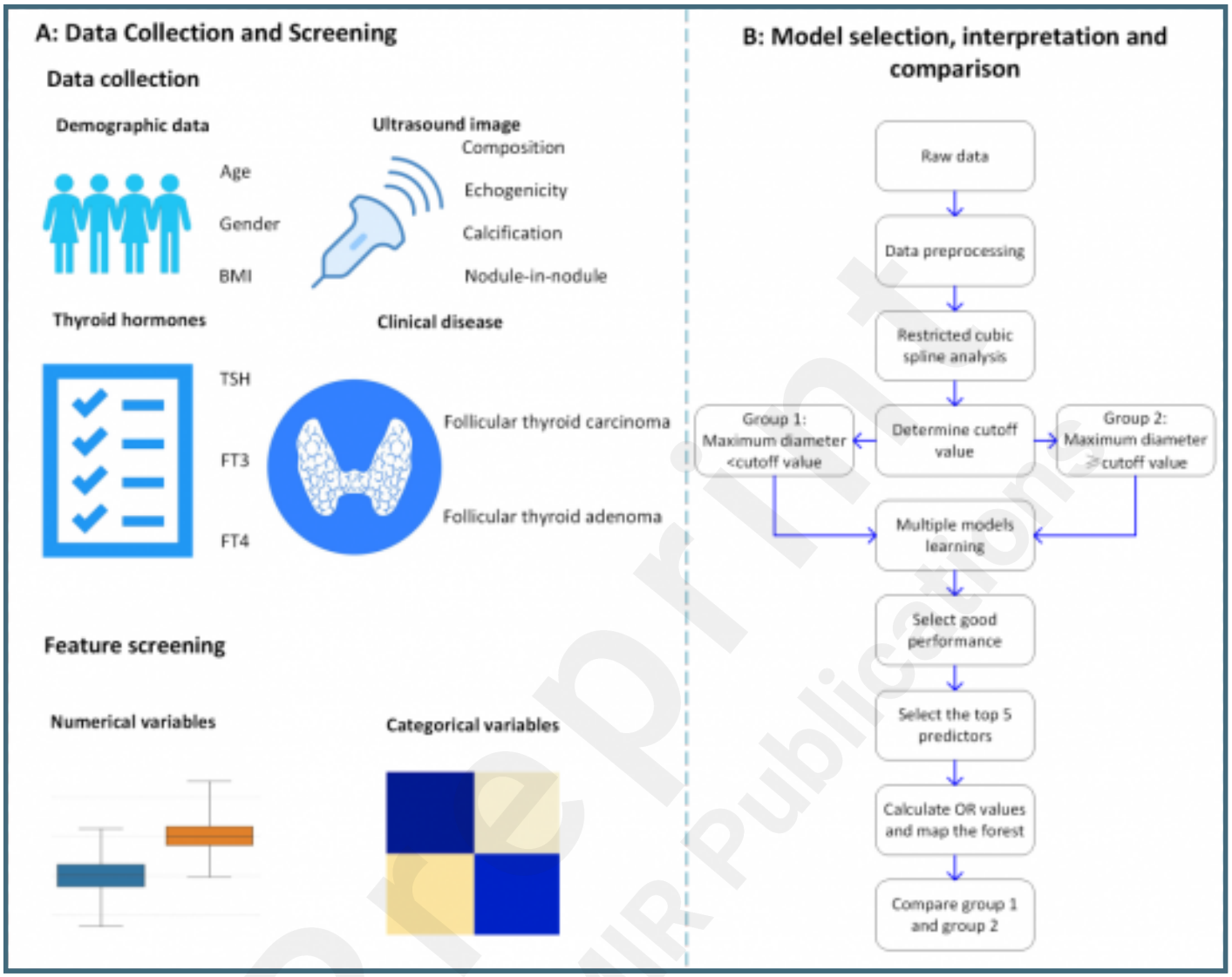
Supplementary Files

Figures

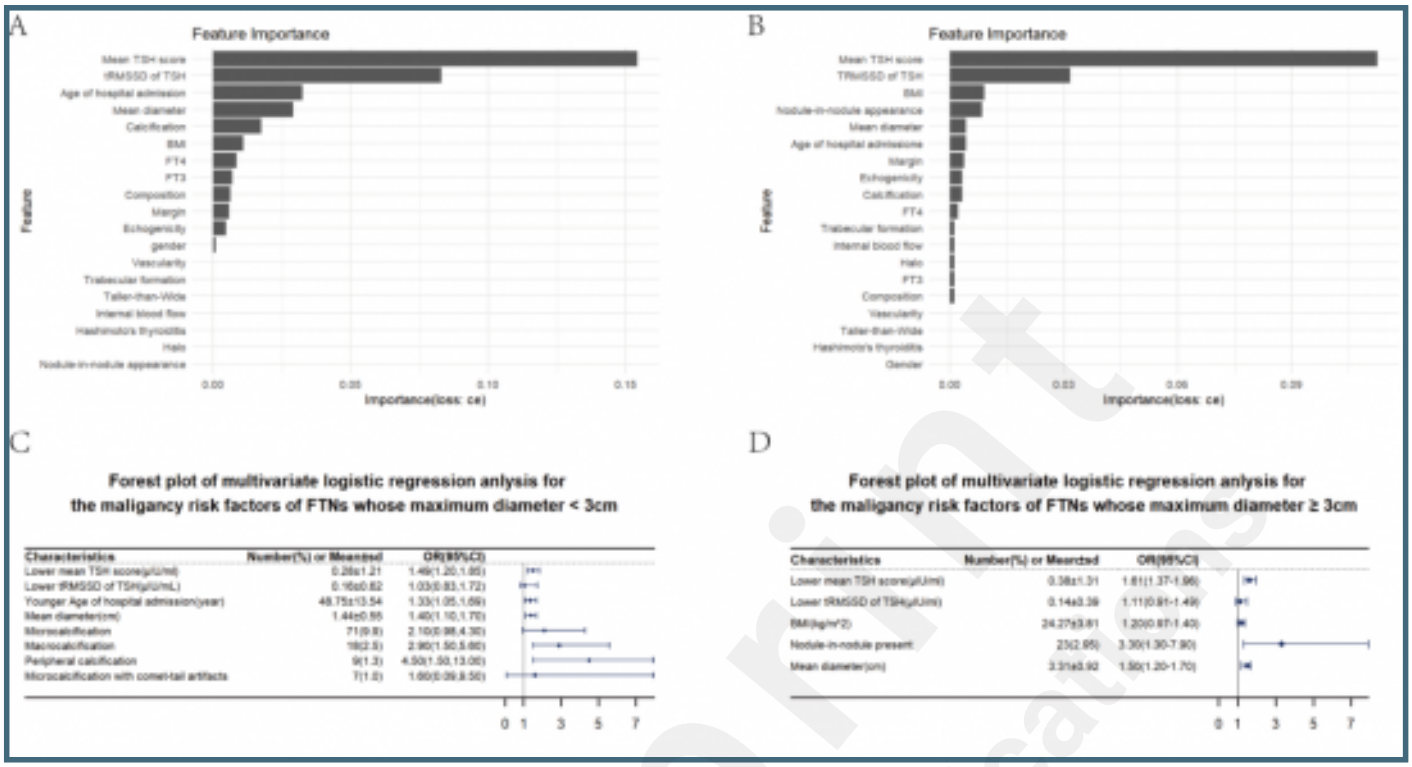
Untitled.



Untitled.



Untitled.



Multimedia Appendixes

Untitled.
URL: http://asset.jmir.pub/assets/026fdf5949145c9c1ff9a6b8c675b55f.docx