Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Construction a Novel Predictive Model for Papillary Thyroid Carcinoma Based on TyG Index

Qianqian Li
First Affiliated Hospital of Zhengzhou University
Hao Wang
First Affiliated Hospital of Zhengzhou University
Jie Wang
First Affiliated Hospital of Zhengzhou University
Shuo Wang
First Affiliated Hospital of Zhengzhou University
Zhen Su
First Affiliated Hospital of Zhengzhou University
Wenjing Yao
First Affiliated Hospital of Zhengzhou University
Jianfeng Liu
First Affiliated Hospital of Zhengzhou University
Yanshu Li
First Affiliated Hospital of Zhengzhou University
Jing Wu
First Affiliated Hospital of Zhengzhou University
Meng Jia
fccjiam@zzu.edu.cn

First Affiliated Hospital of Zhengzhou University

Research Article

Keywords: Papillary thyroid carcinoma, Machine learning model, Triglyceride-glucose index, TyG-Th, Area under the curve

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11194231/v1

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

Additional Declarations: No competing interests reported.

Construction a Novel Predictive Model for Papillary Thyroid Carcinoma Based on TyG Index

Qianqian Li[1#], Hao Wang[1#], Jie Wang[1], Shuo Wang[1], Zhen Su [1], Wenjing Yao[1], Jianfeng Liu[1], Yanshu Li[1], Jing Wu[2*], Meng Jia[1*]

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

[2] Department of Pediatric, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China

[#]Co-author.

[*]Corresponding author.

Meng Jia: fccjiam@zzu.edu.cn, Jing Wu: fccwj@zzu.edu.cn

Abstract

Background: The incidence of papillary thyroid carcinoma (PTC) is continuously increasing, and conventional examinations have obvious limitations in early PTC screening. Although the triglyceride-glucose (TyG) index is associated with multiple malignancies, its predictive value for PTC risk is still undefined.This study aimed to identify independent PTC risk factors, construct effective machine learning predictive models, and develop a novel composite indicator to provide evidence for early screening and risk stratification of PTC.

Methods: A total of 869 patients (447 PTC and 422 benign thyroid nodule cases) from 2022 to 2025 were included for retrospective analysis. Logistic regression analyses were used to screen PTC risk factors. Nine machine learning models were constructed and evaluated by AUC and sensitivity, and SHAP analysis was performed for model interpretation. A novel composite indicator TyG-Th was established and validated.

Results: Age and thyroglobulin (Tg) were protective factors for PTC, while TSH, BMI, and TyG index were independent risk factors. The ANN model exhibited optimal predictive performance with Tg as the core feature. This study innovatively constructed a novel composite indicator, TyG-Th, which achieved a standalone AUC of 0.754 and outperformed conventional single indicators. The combination of TyG-Th with core variables further increased the AUC to 0.804. Stratified analyses validated its stable predictive ability across age and gender subgroups, and higher TyG-Th was correlated with PTC clinical characteristics, demonstrating its promising predictive value.

Conclusions: TyG-Th is a reliable PTC biomarker across all ages and genders. Age and Tg were protective factors against PTC, while TSH, BMI and TyG index were independent risk factors. The models performed well, with Tg as the core ANN feature.

Keywords: Papillary thyroid carcinoma; Machine learning model; Triglyceride-glucose index; TyG-Th; Area under the curve

Introduction

Thyroid carcinoma (TC) is one of the most common malignancies of the endocrine system. Its incidence has risen continuously worldwide in recent years, rendering it one of the solid tumors with the fastest growing incidence rate[1]. PTC accounts for approximately 80% of all TC cases and represents the predominant pathological type[2]. At present, the preoperative diagnosis of PTC mainly relies on high-resolution ultrasonography and fine-needle aspiration (FNA). Ultrasonography plays a vital role

in screening and risk stratification of thyroid nodules; nevertheless, its interpretation is highly dependent on the experience of sonographers[3]. As the diagnosis and treatment of PTC are evolving towards precision and individualization, the limitations of the current evaluation system have become increasingly prominent. Accordingly, it is necessary to explore predictive markers with improved sensitivity and specificity to facilitate the early diagnosis, treatment and even prevention of PTC.

Accumulating evidence indicates that dyslipidaemia is not only closely linked to cardiovascular diseases[4, 5], but also potentially participates in the initiation, progression and metastasis of various malignancies[5-7]. In terms of PTC, a recent study by Lakshya et al.[8] verified an association between statin use and reduced thyroid carcinoma risk. The triglyceride-glucose (TyG) index is a convenient metabolic parameter calculated from fasting triglycerides and fasting blood glucose, which has attracted extensive attention in research on metabolism-related disorders and cancers in recent years[9, 10]. In recent years, research concerning the TyG index in thyroid disorders has gained increasing interest[11-13]. Xu et al.[14] reported that a higher TyG level constitutes a risk factor for thyroid nodules, and this association is more pronounced in specific populations. Several studies[15, 16] have identified a correlation between the TyG index and PTC occurrence. Furthermore, studies investigating the capacity of the TyG index to predict disease severity in patients with PTC remain scarce.

In postoperative management of PTC, TSH suppression therapy is one of the essential therapeutic strategies[17]. A study conducted by Zafon et al.[18] demonstrated that elevated TSH levels may indicate a higher probability of malignant thyroid nodules,

while a cohort study by Rinaldi et al.[19] revealed that low TSH levels are associated with thyroid carcinogenesis. Although several studies have explored the relationship between the TyG index and PTC occurrence[20], there remains a research gap regarding the construction of clinical predictive models combining the TyG index and thyroid function indicators[21, 22].

Accordingly, this study aimed to screen independent risk factors for PTC, construct and optimize machine learning prediction models, establish a novel combined diagnostic indicator TyG-Th, and validate its clinical value in differentiating PTC from benign thyroid nodules. This study intends to provide an efficient and novel biochemical basis for individualized and precise clinical diagnosis of PTC.

Methods

Study Population

This retrospective study enrolled patients with thyroid nodules who underwent preoperative pathological puncture or surgical treatment at the Department of Thyroid Surgery, the First Affiliated Hospital of Zhengzhou University, from January 2022 to October 2025. Patients with any of the following conditions were excluded: (1) a history of previous thyroid-related treatment; (2) comorbid hematological diseases, acute or chronic infectious diseases, autoimmune diseases (excluding thyroid diseases), or other malignant tumors; (3) preoperative interventions including radiotherapy, chemotherapy, and hormone therapy that could interfere with thyroid function and metabolic indicators; (4) missing key data such as medical records, laboratory test

results, or pathological reports that failed to meet the requirements for statistical analysis. After strict screening according to the above criteria, a total of 869 eligible patients were finally included in the study, consisting of 447 patients in the PTC group and 422 patients in the benign thyroid nodule group (Figure 1).

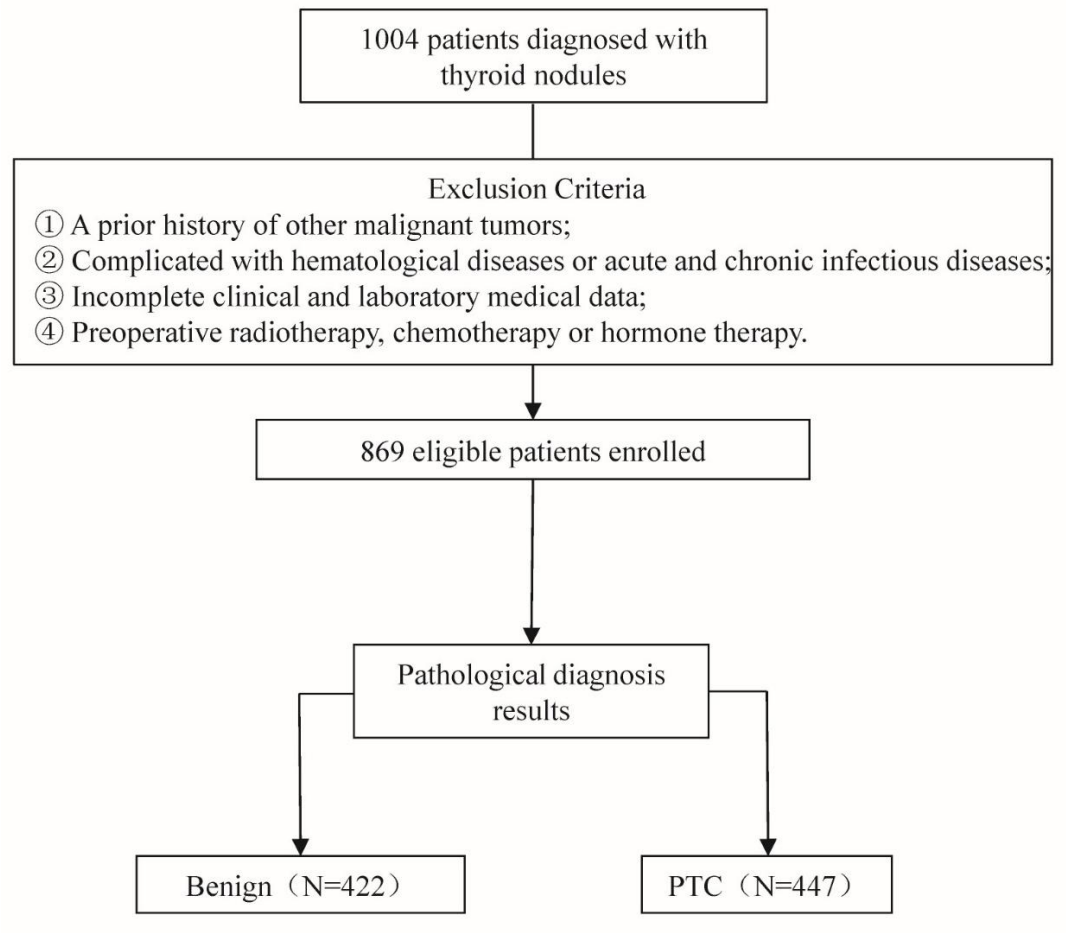


Figure1 Participant selection flow diagram.

All participants were grouped based on postoperative pathological findings with unified diagnostic criteria: (1) patients with negative fine-needle aspiration cytology and no BRAF V600E mutation on genetic testing were assigned to the benign nodule group; (2) patients with postoperative pathological results confirming benign lesions such as nodular goiter without tumor cells were included in the benign nodule group;

(3) patients with a definitive pathological diagnosis of papillary thyroid carcinoma after surgery were enrolled in the PTC group. All pathological results were independently reviewed by no less than two senior pathologists to ensure grouping accuracy.

The study protocol was approved by the Ethics Committee of the First Affiliated Hospital of Zhengzhou University (protocol code 2026-KY-1550). The requirement for informed consent was waived due to the retrospective nature of the study. All raw data were anonymized and strictly confidential, and the entire research procedure complied with the ethical principles of the Declaration of Helsinki. All clinical data of the enrolled patients were extracted from the electronic medical record system of the First Affiliated Hospital of Zhengzhou University. Baseline demographic characteristics and preoperative laboratory indicators were strictly collected to ensure the authenticity and integrity of the research data. The collected variables included demographic characteristics (gender, age), tumor-related features (maximum tumor diameter, tumor stage, lymph node metastasis status), preoperative serum thyroid function indicators (free triiodothyronine [FT3], free thyroxine [FT4], thyroid-stimulating hormone [TSH], thyroid peroxidase antibody [TPOAb], thyroglobulin [Tg], anti-thyroglobulin antibody [TgAb], thyroid-stimulating hormone receptor antibody [TRAb]), and preoperative metabolic-related laboratory indicators (fasting plasma glucose [FPG], triglyceride [TG]). The triglyceride-glucose (TyG) index was adopted as one of the core predictive indicators in this study, which was calculated using the following formula: TyG index = ln [TG (mg/dL) × FPG (mg/dL)] / 2[23].

Model Construction and Validation

Based on the variables with significant intergroup differences, multivariate logistic regression analysis was performed to screen independent predictive factors for PTC. A visualized nomogram prediction model was constructed and optimized using R software and corresponding packages. Internal model validation was conducted via 1000-bootstrap resampling calibration curves and the Hosmer-Lemeshow test to evaluate the predictive accuracy and goodness-of-fit of the model.

Furthermore, multiple machine learning models, including logistic regression, random forest, support vector machine, k-nearest neighbor, and artificial neural network, were established based on the core predictive factors. The original dataset was randomly divided into a training set and a validation set at a stratified ratio of 8:2. Five-fold cross-validation was applied to optimize model parameters and prevent overfitting. Multiple evaluation metrics, including the receiver operating characteristic (ROC) curve, area under the curve (AUC), F1 score, sensitivity, specificity, and accuracy, were used to comprehensively compare the predictive efficacy of single indicators and select the optimal model.

Indicator Construction and Validation

To further improve the differential diagnostic efficacy for PTC, a novel composite indicator was developed by integrating the screened core serological markers. The calculation formula of the composite indicator was defined, and participants were stratified into different groups using the quartile method. Multivariate logistic regression was used to calculate the odds ratio and 95% confidence interval of PTC risk across subgroups. Linear trend tests were performed to explore the dose-response

relationship between the novel composite indicator and PTC risk, so as to systematically verify its superior predictive value compared with traditional single indicators.

Statistical Analysis

All data processing and statistical analyses were performed using R software (version 4.3.2). The normality of continuous variables was assessed using the Shapiro-Wilk test, and homogeneity of variance was verified using the Levene test. Continuous variables conforming to normal distribution and homogeneous variance were presented as mean ± standard deviation (x±s) and compared using the independent t-test. Variables with non-normal distribution were expressed as median (interquartile range) and analyzed using nonparametric rank-sum tests.

Categorical variables were described as composition ratio or percentage (%). The chi-square test was used for intergroup comparison of categorical variables. Specifically, the corrected chi-square test was adopted when the theoretical frequency was less than 5 but the total sample size was no less than 40, and Fisher's exact test was applied when the theoretical frequency was less than 1 or the total sample size was less than 40. Variables with statistically significant intergroup differences were included in the multivariate logistic regression model to screen independent risk factors for PTC. The results were presented as odds ratio (OR) with 95% confidence interval (CI). A two-tailed P value < 0.05 was considered statistically significant throughout the study.

Results

Baseline Characteristics

A total of 869 patients were enrolled in the study cohort, including 163 males (18.8%) and 706 females (81.2%). According to pathological diagnosis, 447 patients were assigned to the PTC group and 422 to the benign nodule group. Detailed baseline clinical characteristics of the two groups are presented in Table 1.

Table1. Baseline characteristics of enrolled patients.

Variable	PTC (n=447)	Benign (n=422)	χ^2/Z/t	*P*
Gender			0.571	0.450
Man(163)	79(17.7%)	84(19.9%)		
Woman(706)	368(82.3%)	338(80.1%)		
Age	42.009±11.307	45.944±14.892	-4.680	<0.001
High	163.922±8.087	163.094±6.953	-1.621	0.105
Weight	68.013±13.875	63.174±11.367	-5.171	<0.001
BMI	25.229±4.277	23.701±3.654	-5.270	<0.001
FT3	4.766±0.661	5.824±22.312	-1.493	0.135
FT4	11.271±2.051	14.112±57.931	-0.034	0.973
TSH	2.746±2.569	2.353±1.995	-3.015	0.003
TgAb	123.258±419.784	57.653±287.017	-9.699	<0.001
Tg	40.536±106.470	647.163±2192.080	-14.293	<0.001
TRAb	0.984±0.657	1.031±2.535	-9.506	<0.001
TPOAb	41.568±83.576	28.545±82.050	2.317	0.021
TG	1.657±2.249	1.237±0.831	-4.492	<0.001
FPG	4.770±1.234	4.608±1.116	-2.856	0.004
TyG	1.105±0.687	0.876±0.609	-4.737	<0.001

In terms of gender distribution, the PTC group consisted of 79 males (17.7%) and 368 females (82.3%), while the benign nodule group included 84 males (19.9%) and

338 females (80.1%). Statistical analysis showed no significant intergroup difference in gender composition ($\chi^2 = 0.571$, $P = 0.45$). The mean age was significantly lower in the PTC group (42.009 ± 11.307 years) than in the benign nodule group (45.944 ± 14.892 years, $P < 0.001$).

Comparative analysis of clinical and laboratory parameters revealed significant differences between the two groups in body weight (t = −5.171, $P < 0.001$), BMI (t = −5.270, $P < 0.001$), TSH (t = −3.015, $P = 0.003$), TgAb (Z = −9.699, $P < 0.001$), Tg (Z = −14.293, $P < 0.001$), TRAb (Z = −9.506, $P < 0.001$), TPOAb (t = 2.317, $P = 0.021$), TG (t = −4.492, $P < 0.001$), FPG (t = −2.856, $P = 0.004$), and TyG index (Z = −4.737, $P < 0.001$). In contrast, no significant differences were observed in height (t = −1.621, $P = 0.105$), FT3 (t = −1.493, $P = 0.135$), and FT4 (t = −0.034, $P = 0.973$).

PTC Risk Factors Screening and Nomogram Construction

Taking pathological results of thyroid nodules as the dependent variable, 11 indicators with significant intergroup differences in univariate analysis were included as candidate independent variables. To eliminate multicollinearity, variables with computational correlations were excluded, including body weight (used for BMI calculation), TG, and FPG (used for TyG calculation). Gender was further incorporated to control confounding effects. Finally, nine variables were included in the multivariate regression model: age, gender, BMI, TSH, TgAb, Tg, TRAb, TPOAb, and TyG index.

Multivariate logistic regression analysis demonstrated that TyG index (OR = 1.787, 95%CI: 1.349–2.393, $P < 0.001$), age (OR = 0.958, 95%CI: 0.945–0.971, $P < 0.001$), Tg (OR = 0.994, 95%CI: 0.992–0.995, $P < 0.001$), TSH (OR = 1.082, 95%CI: 1.004–

1.169, P = 0.042), and BMI (OR = 1.092, 95%CI: 1.044–1.144, P < 0.001) were independently associated with PTC risk. VIF tests confirmed no multicollinearity among all included variables (all VIF values below the critical threshold, Table 2).

Table 2 Multivariate logistic regression results.

Variable	*P*	OR（95%CI）	VIF	Multicollinearity assessment
Gender	0.454	1.18(0.764-1.824)	-	-
Age	<0.001	0.958(0.945-0.971)	1.072	No multicollinearity
BMI	<0.001	1.092(1.044-1.144)	1.179	No multicollinearity
TSH	0.042	1.082(1.004-1.169)	1.042	No multicollinearity
TgAb	0.114	1.001(1.000-1.001)	1.158	No multicollinearity
Tg	<0.001	0.994(0.992-0.995)	1.004	No multicollinearity
TRAb	0.224	0.945(0.834-1.026)	1.099	No multicollinearity
TPOAb	0.443	1.001(0.999-1.003)	1.17	No multicollinearity
TyG	<0.001	1.787(1.349-2.393)	1.233	No multicollinearity

A nomogram prediction model was established based on the five independent risk factors (Tg was standardized before modeling) with pathological diagnosis (PTC vs. benign nodules) as the dependent variable. Bootstrap resampling (1000 iterations) and calibration curves were applied for internal validation, and the AUC value was used to evaluate discriminatory ability. The initial nomogram showed good discriminatory efficacy with an AUC of 0.799 (95%CI: 0.770–0.828) but exhibited significant calibration deviation (Hosmer–Lemeshow P = 0.001) with a calibration slope of 5.799, indicating substantial underestimation of PTC risk.

After logistic regression-based calibration correction, the revised nomogram (Figure 2A) achieved favorable calibration (Hosmer–Lemeshow P = 0.087) with a reduced

calibration slope of 5.011 (Figure 2C), which effectively improved the risk underestimation. The corrected model maintained an AUC of 0.799 (95%CI: 0.770–0.828). Detailed validation results are shown in Table S1[see Additional file 1].

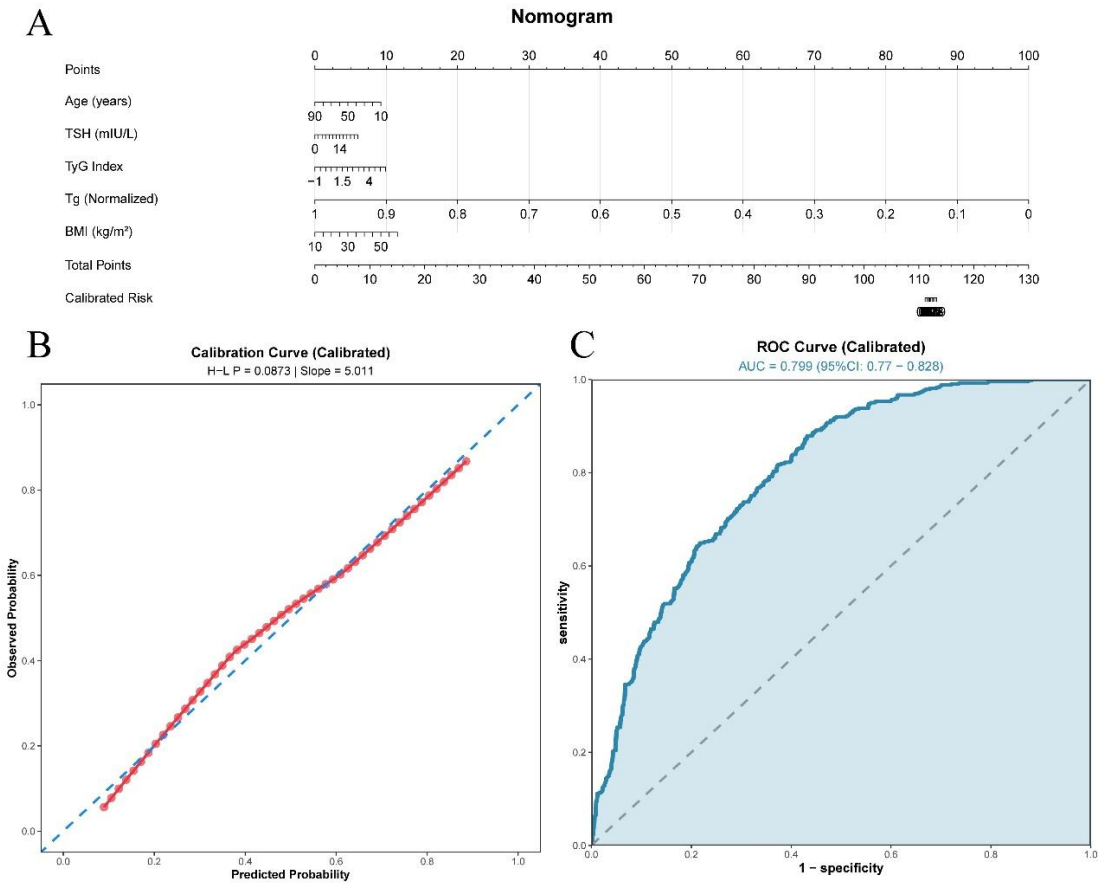


Figure2 Calibrated nomogram demonstrated good discrimination and calibration for predicting PTC. (A) Calibrated nomogram (B) ROC curve for the calibrated nomogram (C) Calibration curve for the calibrated nomogram

Machine Learning PTC Prediction Models

Nine machine learning models, including RF, SVM, LR, KNN, PLS, ANN, NB, LDA, and LASSO, were constructed using five core indicators (age, TSH, Tg, BMI, and TyG) to predict PTC. The ROC curves of training and validation cohorts for all models are

displayed in Figures 3A and 3B. The ANN model yielded the optimal comprehensive performance.

As summarized in Table S2[see Additional file 1], the ANN model achieved a validation cohort AUC of 0.791, with an F1-score of 0.713, accuracy of 0.717, specificity of 0.726, and sensitivity of 0.708. In the training cohort, the ANN model obtained an AUC of 0.818, a sensitivity of 0.767, and a specificity of 0.679. Although the NB model showed the highest training AUC (0.849), its specificity was only 0.552. Both PLS and LDA exhibited poor performance with validation AUC values below 0.700. Collectively, the ANN model based on the five indicators demonstrated high accuracy and stability for PTC prediction.

Confusion matrices were used to evaluate the stability of the ANN model (Figure 3D), where blue represents correct predictions and red represents incorrect predictions. The overall accuracy was 74.6% in the training set and 69.4% in the validation set, indicating stable generalization without significant performance decline.

SHapley Additive exPlanations (SHAP) analysis was performed to interpret the ANN model. The feature importance ranking indicated that the predictive contribution from highest to lowest was Tg (mean |SHAP| = 0.149), age (mean |SHAP| = 0.116), BMI (mean |SHAP| = 0.054), TyG index (mean |SHAP| = 0.049), and TSH (mean |SHAP| = 0.019). The SHAP summary plot revealed that elevated TyG and BMI corresponded to positive SHAP values and increased PTC probability. Age exhibited a bidirectional nonlinear effect, while TSH showed the weakest independent contribution. Tg served as the core linear protective factor, whereas TyG acted as a weak nonlinear risk factor.

The SHAP waterfall plot visualized individual prediction processes: the baseline predictive value was 0.494, and the final predictive value was 0.428. TyG exerted the dominant negative contribution (SHAP = −0.155), followed by negative effects of BMI and TSH, while age and Tg provided positive contributions (Figure 3C).

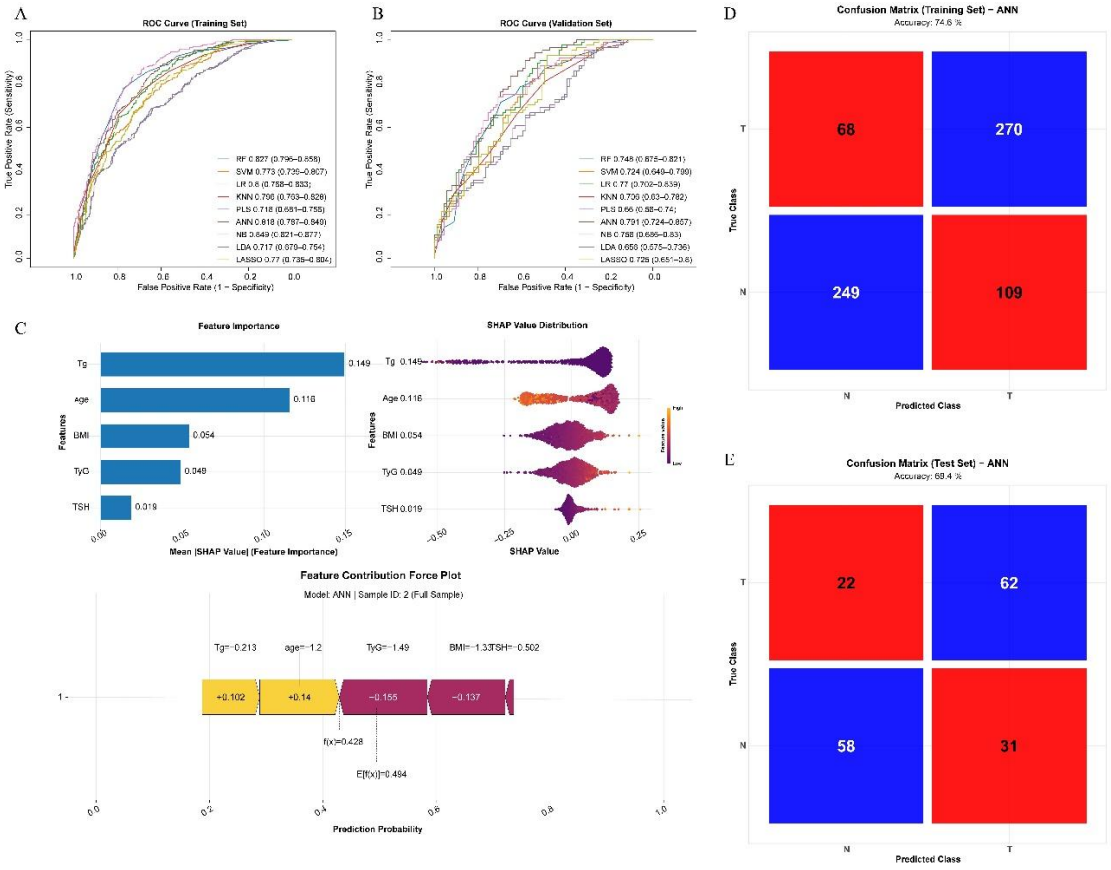


Figure3 Machine learning PTC prediction models. (A) ROC curves of nine machine learning models in the training set and(B)test set (C) SHAP interpretation analysis of the ANN model (D) Confusion matrix of the ANN model

Construction and Predictive Performance of TyG-Th

Quartile stratification analyses were performed for TSH, TyG, and Tg to explore the dose–response relationships between indicator levels and PTC risk (Figures 4A–4C). Taking the lowest quartile (Q1) as the reference group, higher quartiles of TSH and TyG

were significantly associated with increased PTC prevalence (all $P < 0.01$), indicating positive dose–response trends. In contrast, higher Tg levels were correlated with lower PTC prevalence ($P < 0.001$), demonstrating a protective effect.

Since TyG and TSH were identified as independent risk factors (positive association), whereas Tg was a protective factor (negative association), we hypothesized that combining these biomarkers by multiplying positively associated variables and dividing by negatively associated variables may better reflect the overall metabolic-thyroid imbalance. We innovatively constructed a novel composite index, TyG-Th, calculated as TyG-Th = TyG × TSH / Tg. Quartile analysis of TyG-Th (Figure 4D) showed that patients in the highest quartile (Q4) had a significantly higher PTC proportion than those in Q1 ($P < 0.001$). Trend analysis confirmed a significant incremental dose–response relationship between elevated TyG-Th levels and increased PTC risk (P trend < 0.01).

To further validate the predictive value of TyG-Th, nine machine learning models were established using single variables including age, TSH, BMI, TyG, Tg, and TyG-Th. The ROC curves of all single-variable models are shown in Figure S1[see Additional file 1]. TyG-Th yielded the highest AUC across all models(Table 3). Based on F1-score optimization, the SVM model using TyG-Th alone achieved the best predictive performance (AUC = 0.754), with a sensitivity of 0.690, specificity of 0.742, accuracy of 0.717, and F1-score of 0.703. The LDA model based on Tg showed the highest sensitivity (0.881) but low specificity (0.472), indicating a high false-positive

rate. All other single-variable models exhibited limited predictive efficacy with AUC values below 0.65.

Table 3 Performance of optimal predictive models.

Variable	Model	Sensitivity	Specificity	Accuracy	F1	AUC
Tg	LDA	0.881	0.472	0.671	0.722	0.747
age	KNN	0.690	0.494	0.590	0.620	0.601
TSH	NN	0.762	0.303	0.526	0.610	0.559
TyG	NN	0.571	0.562	0.566	0.561	0.608
BMI	Adaboost	0.452	0.573	0.514	0.475	0.518
TyG-Th	SVM	0.690	0.742	0.717	0.703	0.754

To further enhance predictive accuracy, TyG-Th was integrated into the original five-indicator combined model. The optimized ANN model with TyG-Th exhibited superior predictive performance (Figures 4E, 4F), with a validation AUC of 0.804, which was 0.013 higher than that of the original five-indicator model (AUC = 0.791).

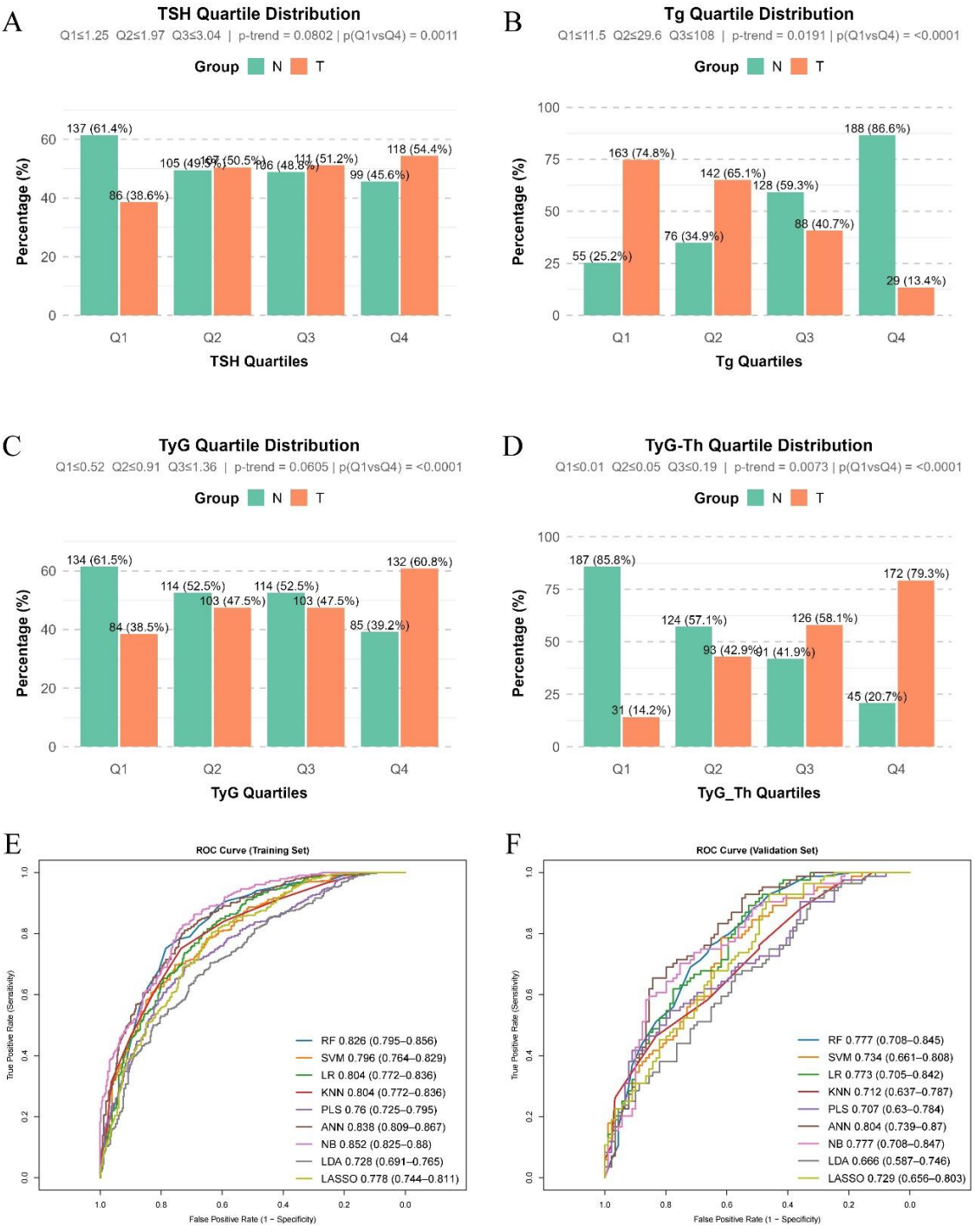


Figure4 Construction and predictive performance of TyG-Th. Quartile analysis of each indicator (A) TSH indicator; (B) TyG indicator; (C) Tg indicator; (D) TyG-Th indicator; ROC Curves for the TyG-Th machine learning model: training set (E) and validation set (F).

Stratified Validation of TyG-Th for PTC

Stratified analyses by age (<45 years vs. ≥45 years) and gender (female vs. male) were performed to verify the stability of the TyG-Th predictive effect (Figure 5). In all subgroups, elevated TyG-Th was significantly and positively correlated with PTC risk (all $P < 0.001$).

In age-stratified subgroups, the Q4 HR was 22.000 (95%CI: 11.160–43.390) for patients aged <45 years and 57.270 (95%CI: 19.260–170.270) for patients aged ≥45 years. In gender-stratified subgroups, the Q4 HR was 25.340 (95%CI: 14.380–44.640) in females and 17.720 (95%CI: 5.690–55.200) in males. Interaction analysis showed no significant interactions between TyG-Th and age or gender (both $P > 0.05$), indicating that the predictive value of TyG-Th was independent of age and gender. Although numerical differences existed across subgroups, the positive association between TyG-Th elevation and increased PTC risk remained consistent without obvious heterogeneity, confirming the favorable stability and generalizability of TyG-Th.

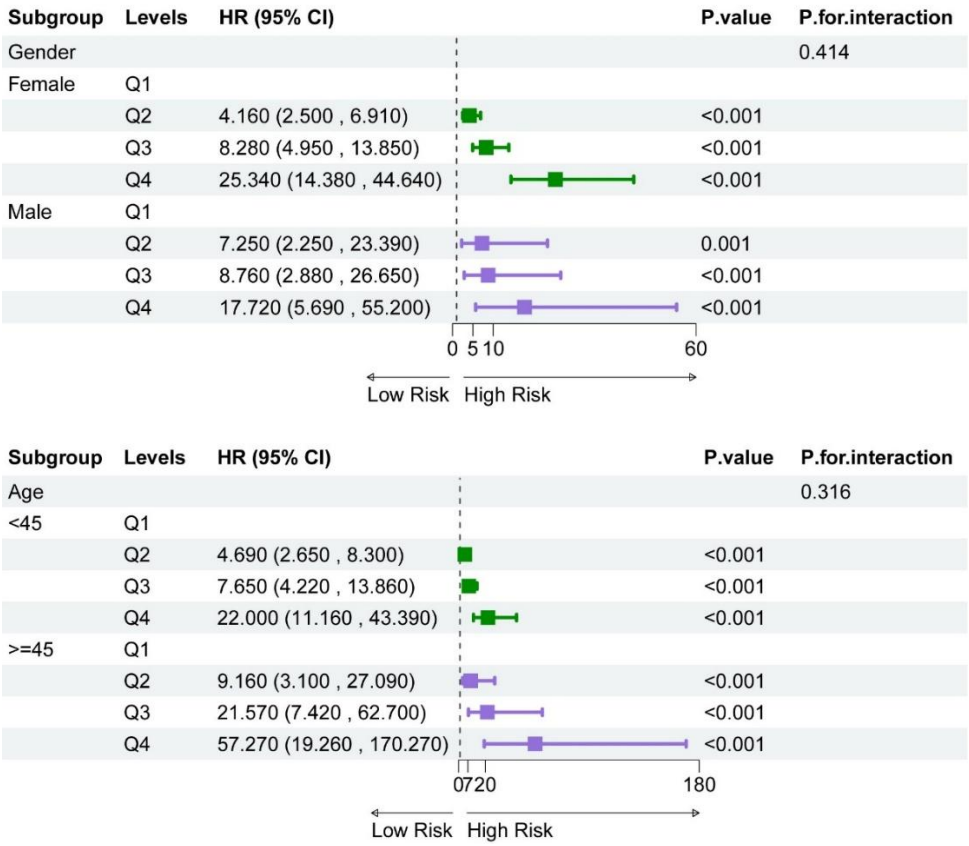


Figure5. Age- and gender-stratified analysis of TyG-Th and PTC risk.

To analyze the correlation between TyG-Th and PTC, a total of 422 PTC patients were stratified into four quartile groups according to TyG-Th levels: Q1 (<0.01, n = 37), Q2 (0.01–0.05, n = 85), Q3 (0.05–0.19, n = 126), and Q4 (>0.19, n = 174). Baseline characteristics across quartiles are summarized in Table S3[see Additional file 1]. With the increase in TyG-Th levels, age, body weight, and BMI increased significantly (P < 0.001, P = 0.012, P = 0.037), while gender and height showed no significant differences. In terms of thyroid function and autoimmune indicators, higher TyG-Th was associated with decreased FT4 and Tg, as well as increased TSH, TPOAb, and TgAb (all P < 0.05). No significant differences were observed in FT3 and TRAb across groups. For glycolipid metabolic indicators, TG and TyG increased significantly with elevated TyG-Th (P < 0.05), whereas FPG remained stable (P = 0.086). In terms of tumor

characteristics, higher TyG-Th levels were associated with significantly smaller maximum tumor diameter and a higher proportion of T1 stage tumors (both $P < 0.001$). All patients were M0 stage, and no significant differences were found in N stage or overall tumor stage across quartiles ($P > 0.05$).

Discussion

We systematically explored the associations of demographic characteristics, thyroid function parameters, and metabolic indicators with PTC, screened independent risk factors, constructed predictive models, and innovatively established a novel composite index (TyG-Th) to provide new evidence for early PTC risk assessment. Of the 869 enrolled patients, 81.2% were female, and no significant difference in gender composition was observed between the PTC group and the benign nodule group, which is consistent with the female predominance of thyroid diseases reported in previous studies[24]. Patients in the PTC group had a significantly younger mean age than those in the benign nodule group. This finding is consistent with multiple domestic and international studies[25].

As core obesity-related indicators, body weight and BMI were significantly higher in the PTC group, suggesting that obesity may promote abnormal thyroid cell proliferation and malignant transformation via insulin resistance, inflammatory activation, and sex hormone disorders[26]. In terms of thyroid function and autoimmune indicators, higher TSH levels and lower Tg levels were detected in the PTC group, which is in line with previous findings[27]. TSH was identified as an

independent risk factor for PTC. Elevated TSH may upregulate vascular endothelial growth factor (VEGF) expression and promote tumor angiogenesis[28], reminding clinicians to be alert to PTC in nodule patients with increased TSH levels. BMI also served as an independent risk factor for PTC, indicating a higher PTC risk in obese patients with thyroid nodules. In contrast, Tg was an independent protective factor for PTC. Closely related to thyroid structural integrity and functional status, Tg synthesis and secretion are impaired when normal follicular structures are destroyed by PTC tumors[27]. Reduced Tg levels therefore suggest a higher possibility of PTC.

The TyG index was a powerful independent risk factor for PTC. Calculated based on triglyceride and fasting glucose levels, the TyG index reliably reflects insulin resistance[15]. Insulin resistance promotes tumor cell proliferation and survival via insulin receptor substrate (IRS)-mediated signaling pathways[29] and exacerbates oxidative stress, leading to reactive oxygen species (ROS) accumulation, DNA damage, and gene mutation. These mechanisms further validate the close link between metabolic disorders and PTC[30].

Based on five independent risk factors (age, TSH, Tg, BMI, and TyG index), a visual PTC risk nomogram was constructed and internally validated using Bootstrap resampling. Nomograms convert complex multivariate regression models into intuitive graphical tools, enabling rapid individualized PTC risk evaluation in clinical practice[31]. Nine machine learning models were established based on the five screened risk factors. According to training and validation cohort evaluations, the artificial neural network[32] (ANN) model yielded the optimal comprehensive performance. Individual

SHAP waterfall analysis revealed that TyG exerted the predominant negative predictive contribution, followed by negative effects of BMI and TSH, whereas age and Tg provided positive contributions. These findings provide novel insights into the pathogenesis of PTC. Notably, PTC development is inherently accompanied by systemic lipid metabolic disorders[30], and thyroid function indicators can partially reflect the overall metabolic status of the human body[33]. The intrinsic correlation between lipid metabolism and thyroid function in metabolic regulation provides a solid theoretical basis for their combined application in PTC risk assessment.

Given the differential associations of TSH, TyG, and Tg with PTC risk, we innovatively constructed a composite index, TyG-Th (TyG-Th = TyG × TSH / Tg). Quartile stratification analysis revealed a significant upward trend in PTC prevalence with increasing TyG-Th levels, confirming a robust dose–response relationship. Elevated TSH and TyG levels promote PTC tumorigenesis by accelerating thyroid cell proliferation and exacerbating metabolic disorders, while decreased Tg levels reflect impaired thyroid structural integrity and function, indicating potential malignant lesions[27]. Age- and gender-stratified analyses showed that elevated TyG-Th was significantly positively correlated with PTC risk in all subgroups. No significant interaction was observed between TyG-Th and age or gender, indicating that the predictive value of TyG-Th was stable and consistent across different age and gender populations with good generalizability. Baseline characteristic analysis across TyG-Th quartiles demonstrated that increased TyG-Th levels were accompanied by significant increases in age, body weight, and BMI. In terms of tumor features, patients in the

higher TyG-Th quartiles (Q3, Q4) presented significantly smaller maximum tumor diameters and earlier T stage (higher proportion of T1 tumors). These results suggest that elevated TyG-Th is correlated with smaller tumor size and less invasive early-stage PTC, indicating its potential value for early PTC identification.

Although the optimized ANN model incorporating TyG-Th exhibited superior predictive performance, with a validation AUC of 0.804 which was only 0.013 higher than that of the original five-indicator model (AUC = 0.791), the stable predictive performance of TyG-Th for PTC across subgroups stratified by gender and age, as well as the simplicity of TyG-Th calculation, endows it with potential value for clinical application.

The present study has several limitations. This single-center retrospective study lacked multi-center external validation, which may reduce the generalizability of the results.

Conclusion

TyG-Th is the novel composite index positively associated with PTC risk. It outperforms conventional indicators and remains stable across age and gender subgroups, supporting early PTC risk assessment. This study identified age and Tg as independent protective factors, while TSH, BMI, and TyG index served as independent risk factors for PTC. The ANN model achieved good predictive performance, with Tg as its core feature. These findings may provide references for early-stage screening, risk stratification and individualized clinical management of PTC.

List of abbreviations

AdaBoost Adaptive Boosting
ANN Artificial Neural Network
ATC Anaplastic Thyroid Carcinoma
AUC Area Under the Curve
BMI Body Mass Index
CI Confidence Interval
C-index Consistency Index
FNA Fine Needle Aspiration
FPG Fasting Plasma Glucose
FT3 Free Triiodothyronine
FT4 Free Thyroxine
FTC Follicular Thyroid Carcinoma
H-L Test Hosmer-Lemeshow Test
HOMA-IR Homeostasis Model Assessment of Insulin Resistance
HT Hashimoto Thyroiditis
IR Insulin Resistance
KNN K-Nearest Neighbors
LASSO Least Absolute Shrinkage and Selection Operator
LDA Linear Discriminant Analysis
LR Logistic Regression
MS Metabolic Syndrome
MTC Medullary Thyroid Carcinoma
MTD Maximum Tumor Diameter
NB Naive Bayes
OR Odds Ratio
PLS Partial Least Squares
PTC Papillary Thyroid Carcinoma
RF Random Forest
ROC Receiver Operating Characteristic
SHAP SHapley Additive exPlanations
SVM Support Vector Machine
TC Thyroid Carcinoma
Tg Thyroglobulin

TG Triglyceride
TgAb Thyroglobulin Antibody
TPOAb Thyroid Peroxidase Antibody
TRAb Thyrotropin Receptor Antibody
TSH Thyroid-Stimulating Hormone
TyG Triglyceride-Glucose Index
TyG-Th Triglyceride-Glucose-Thyroid Hormone Index
XGBoost eXtreme Gradient Boosting

Declarations

Ethics approval and consent to participate

The studies involving humans were approved by the First Affiliated Hospital of Zhengzhou University (protocol code 2026-KY-1550). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin due to the retrospective nature of the study.

Clinical trial registration

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Fundings

This work was supported by the Key Scientific Research Project Plan of Institutions of Higher Education in Henan Province [No. 26A320056]; Zhengzhou Municipal

Project for Basic Research and Applied Basic Research [No. 2024ZZJCYJ021]; Key Scientific and Technological Tackling Program of Henan Province from the Department of Science and Technology of Henan Province [No. 262102311070]; Leading talents Plan in Henan Province [No. ZYYCYU202012116]. Natural Science Foundation of Henan Province [No.252300423216].

Author contributions

QQL: Writing – original draft, Writing – review & editing. HW: Writing – original draft, Writing – review & editing. JW: Writing review & editing, Methodology. SW: Methodology, Writing – review & editing. ZS: Writing – review & editing, Software. WJY: Writing – review & editing. YSL: Writing – review & editing. JW: Writing review & editing, Writing – original draft. MJ: Writing review & editing, Writing – original draft.

Acknowledgments

The authors thank the staff of the Department of Thyroid Surgery of the First Affiliated Hospital of Zhengzhou University for their valuable support and dedicated work throughout this study.

Availability of data and materials

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher,

the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Zheng RS, Chen R, Han BF, Wang SM, Li L, Sun KX, et al. [Cancer incidence and mortality in china, 2022]. Zhonghua Zhong Liu Za Zhi. 2024 2024 Mar 23;46(3):221-31
2. Chen DW, Lang BHH, McLeod DSA, Newbold K, Haymart MR. Thyroid cancer. Lancet. 2023 2023 May 6;401(10387):1531-44
3. Wang C, Li Y, Zhang M, Yu Z, Wu Z, Li C, et al. Analysis of differential diagnosis of benign and malignant partially cystic thyroid nodules based on ultrasound characterization with a TIRADS grade-4a or higher nodules. Front Endocrinol (Lausanne). 2022 2022;13:861070
4. Li J, Zhao S, Zhao D, Lu G, Peng D, Liu J, et al. 2023 chinese guideline for lipid management. Front Pharmacol. 2023 2023;14:1190934
5. Xiong X, Cheng F, Qi W, Cheng J, Jiang W, Wang Y. Association between the TyG/HDL-c ratio and non-alcoholic fatty liver disease: a single-centre retrospective study. BMC Gastroenterol. 2026 2026 Jul 15
6. Yao X, Tian Z. Dyslipidemia and colorectal cancer risk: a meta-analysis of prospective studies. Cancer Causes Control. 2015 2015 Feb;26(2):257-68
7. Shao Y, Li W, Wen CP, Yang M, Ning X, Hu K, et al. Dyslipidemia progression and increased lung cancer risk: a prospective cohort study. Eur J Epidemiol. 2024 2024 Dec;39(12):1363-71
8. Trivedi LU, Femnou Mbuntum L, Halm EA, Mansi I. Is statin use associated with risk of thyroid diseases? Results of a retrospective cohort study. Ann Pharmacother. 2021 2021 Sep;55(9):1110-19
9. Mo Y, Chen S, Chen Y, Zhang C, Deng S, Mao Z, et al. Prognostic value of the TyG index and CHG index in gastric cancer. J Evid Based Med. 2026 2026 Jul 1:e70142
10. Li Y, Guo L, Zhu X, Lin C, Liu X, Huang Z, et al. TyG index is associated with colorectal adenomatous polyps and genetically linked to colon polyps: a cross-sectional and mendelian randomization study. Front Oncol. 2026 2026;16:1756399
11. Frandes M, Gherbon A, Tudor A, Albai O, Vlad MM, Muntean C. TyG-centered endocrine-metabolic architecture in patients with thyroid dysfunction: a systems phenotyping study. Metabolites. 2026 2026 Jul 12;16(7)
12. Li C, Dionigi G, Guan H, Sun H, Zhang J. The triglyceride-glucose body mass index paradox: dual metabolic effects on tumor aggressiveness and recurrence risk in 11 317 papillary thyroid carcinoma patients. Int J Surg. 2026 2026 Feb 1;112(2):4356-65
13. Li G, Ren Q, Zhou J, Zhou H, Wu Y, Chen L, et al. Triglyceride-glucose-body mass index is linked to papillary thyroid carcinoma odds in china: a retrospective study. Front Endocrinol (Lausanne). 2026 2026;17:1766186
14. Xu H, Si C. Association between triglyceride-glucose index and thyroid nodule: a cross-sectional study in western liaoning province, china. Sci Rep. 2025 2025 Jul 2;15(1):23100
15. Alkurt EG, Sahin F, Tutan B, Canal K, Turhan VB. The relationship between papillary thyroid cancer and triglyceride/glucose index, which is an indicator of insulin resistance. Eur Rev Med

Pharmacol Sci. 2022 2022 Sep;26(17):6114-20
16. Song C, Ping M, Lin L, Meng X, Lan Y, Tong H. Association between triglyceride-glucose index and papillary thyroid carcinoma among chinese adults with thyroid nodules. Front Endocrinol (Lausanne). 2025 2025;16:1616350
17. Gao L, Bao Y, Li T, Liu G, Zhang F, Chen Y, et al. Detection of thyroglobulin, TSH receptor and sodium-iodine symporter mRNA for identification of lymph node metastases in patients with papillary thyroid carcer. BMC Cancer. 2026 2026 Jul 9
18. Zafon C, Obiols G, Baena JA, Castellvi J, Dalama B, Mesa J. Preoperative thyrotropin serum concentrations gradually increase from benign thyroid nodules to papillary thyroid microcarcinomas then to papillary thyroid cancers of larger size. J Thyroid Res. 2012 2012;2012:530721
19. Rinaldi S, Plummer M, Biessy C, Tsilidis KK, Ostergaard 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 2014 Jun;106(6):dju097
20. Song C, Ping M, Lin L, Meng X, Lan Y, Tong H. Association between triglyceride-glucose index and papillary thyroid carcinoma among chinese adults with thyroid nodules. Front Endocrinol (Lausanne). 2025 2025;16:1616350
21. Dogan UK, Aslan S. Triglyceride-glucose index as a marker of metabolic and inflammatory risk across different thyroid function states. Sci Rep. 2026 2026 May 28;16(1)
22. Su Y, Chen L, Huang J, Lyu H, Feng W, Zhan J, et al. FT3 combined with the TyG index for risk stratification of mild cognitive impairment in euthyroid patients with type 2 diabetes. Front Endocrinol (Lausanne). 2026 2026;17:1802217
23. Simental-Mendia LE, Rodriguez-Moran M, Guerrero-Romero F. The product of fasting glucose and triglycerides as surrogate for identifying insulin resistance in apparently healthy subjects. Metab Syndr Relat Disord. 2008 2008 Dec;6(4):299-304
24. Pizzato M, Li M, Vignat J, Laversanne M, Singh D, La Vecchia C, et al. The epidemiological landscape of thyroid cancer worldwide: GLOBOCAN estimates for incidence and mortality rates in 2020. Lancet Diabetes Endocrinol. 2022 2022 Apr;10(4):264-72
25. Chen DW, Carr G, Worden FP, Veenstra CM, Haymart MR, Banerjee M. Survival prognostication in patients with differentiated thyroid cancer and distant metastases: a SEER population-based study. Thyroid. 2024 2024 Jul;34(7):837-45
26. Li X, Zhang X, Sun L, Yang L, Li Q, Wang Z, et al. Associations between metabolic obesity phenotypes and pathological characteristics of papillary thyroid carcinoma. Endocr Pract. 2024 2024 Jul;30(7):624-30
27. Li S, Ren C, Gong Y, Ye F, Tang Y, Xu J, et al. The role of thyroglobulin in preoperative and postoperative evaluation of patients with differentiated thyroid cancer. Front Endocrinol (Lausanne). 2022 2022;13:872527
28. Li J, Shi M, Sun X, Zhang B, Cao Y, Su J, et al. Correlation between ultrasound characteristics of thyroid microcarcinoma and levels of thyroid-stimulating hormone and vascular endothelial growth factor. Am J Transl Res. 2025 2025;17(3):1938-49
29. Xu X, Nie J, Lu L, Du C, Meng F, Song D. YAP-TEAD up-regulates IRS2 expression to induce and deteriorate oesophageal cancer. J Cell Mol Med. 2021 2021 Mar;25(5):2584-95
30. Lukasiewicz M, Zwara A, Kowalski J, Mika A, Hellmann A. The role of lipid metabolism disorders in the development of thyroid cancer. Int J Mol Sci. 2024 2024 Jun 28;25(13)

31. Chang Q, Zhang J, Wang Y, Li H, Du X, Zuo D, et al. Nomogram model based on preoperative serum thyroglobulin and clinical characteristics of papillary thyroid carcinoma to predict cervical lymph node metastasis. Front Endocrinol (Lausanne). 2022 2022;13:937049
32. Helin R, Indahl U, Tomic O, Liland KH. Non-linear shrinking of linear model errors. Anal Chim Acta. 2023 2023 Jun 1;1258:341147
33. Wang X, Wu Z, Liu Y, Wu C, Jiang J, Hashimoto K, et al. The role of thyroid-stimulating hormone in regulating lipid metabolism: implications for body-brain communication. Neurobiol Dis. 2024 2024 Oct 15;201:106658

Supplementary Files

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

- Additionalfile1.docx