frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 06 November 2024
DOI 10.3389/fendo.2024.1366687

Check for updates

OPEN ACCESS

EDITED BY
Paula Soares,
University of Porto, Portugal

REVIEWED BY
Jincao Yao,
University of Chinese Academy of Sciences, China
Bilgin Kadri Aribas,
Bülent Ecevit University, Türkiye

*CORRESPONDENCE
Lu Zhang
zl11904@rjh.com.cn
Weiwei Zhan
shanghairuijin@126.com
Wei Zhou
zw11468@126.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 07 January 2024
ACCEPTED 21 October 2024
PUBLISHED 06 November 2024

CITATION
Cao Y, Yang Y, Chen Y, Luan M, Hu Y, Zhang L, Zhan W and Zhou W (2024) Optimizing thyroid AUS nodules malignancy prediction: a comprehensive study of logistic regression and machine learning models. *Front. Endocrinol.* 15:1366687. doi: 10.3389/fendo.2024.1366687

COPYRIGHT
© 2024 Cao, Yang, Chen, Luan, Hu, Zhang, Zhan and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Optimizing thyroid AUS nodules malignancy prediction: a comprehensive study of logistic regression and machine learning models

Yuan Cao[1†], Yixian Yang[1†], Yunchao Chen[2], Mengqi Luan[1], Yan Hu[1], Lu Zhang[1*], Weiwei Zhan[1*] and Wei Zhou[1*]

[1]Department of Ultrasound, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [2]Department of Ultrasound, Zhongshan Hospital (Xiamen Branch), Fudan University, Xiamen, Fujian, China

Background: The accurate diagnosis of thyroid nodules with indeterminate cytology, particularly in the atypia of undetermined significance (AUS) category, remains challenging. This study aims to predict the risk of malignancy in AUS nodules by comparing two machine learning (ML) and three conventional logistic regression (LR) models.

Methods: A retrospective study on 356 AUS nodules in 342 individuals from 6728 patients who underwent thyroid surgery in 2021. All the clinical, ultrasonographic, and molecular data were collected and randomly separated into training and validation cohorts at a ratio of 7: 3. ML (random forest and XGBoost) and LR (lasso regression, best subset selection, and backward stepwise regression) models were constructed and evaluated using area under the curve (AUC), calibration, and clinical utility metrics.

Results: Approximately 90% (321/356) of the AUS nodules were malignant, predominantly papillary thyroid carcinoma with 68.6% BRAF V600E mutations. The final LR prediction model based on backward stepwise regression exhibited superior discrimination with AUC values of 0.83 (95% CI: 0.73-0.92) and 0.80 (95% CI: 0.67-0.94) in training and validation, respectively. Well calibration, and clinical utility were also confirmed. The ML models showed moderate performance. A nomogram was developed on the final LR model.

Conclusions: The LR model developed using the backward stepwise regression, outperformed ML models in predicting malignancy in AUS thyroid nodules. The corresponding nomogram based on this model provides a valuable and practical tool for personalized risk assessment, potentially reducing unnecessary surgeries and enhancing clinical decision-making.

KEYWORDS

thyroid nodules, risk of malignancy, atypia of undetermined significance (AUS), machine learning (ML), logistic regression, prediction model, BRAF V600E

1 Introduction

The optimal management of thyroid nodules with indeterminate cytology, as classified under the 2023 Bethesda System for Reporting Thyroid Cytopathology (TBSRTC), poses a challenge in clinical practice (1). These indeterminate nodules encompass categories such as atypia of undetermined significance (AUS), follicular neoplasm (FN), and suspicious for malignancy (SUSP), each carrying a variable risk of malignancy from 6% to 94.2%, often influenced by institutional variations or calculation methods (2–6). Despite advancements in high resolution ultrasonography, molecular testing, and other modern diagnostic techniques, no consensus has been reached on the management of AUS nodules. Therefore, efficient risk assessment and intervention strategies are calling for.

Molecular testing has gained prominence in diagnosing indeterminate thyroid nodules and two molecular profiles have been identified: BRAF V600E-like and RAS-like (7, 8). Its applications extend from determining the presence of malignancy to offering prognostic insights and guiding therapeutic decisions (9–11). The BRAF V600E mutation, with its relatively high prevalence and specificity in papillary thyroid carcinoma (PTC), holds particular significance, especially in East Asian populations (12, 13). In China, it has been adopted as a routine element in pre-operative liquid-based biopsy following FNA in the majority of clinical centers. However, the diagnostic efficacy of BRAF V600E mutation alone in AUS nodules is not superior, possibly resulting from variant histopathology, particularly in nodules without the BRAF V600E mutation, where ruling out malignancy is challenging. The combination of demography, imaging, and molecular testing is likely to improve diagnostic accuracy. Therefore, adjunct preoperative assessments are still highly desirable. To the best of our knowledge, studies published to date specifically focusing on predicting the malignancy risk for category III nodules remain limited.

Recently, artificial intelligence (AI) has undergone rapid transformation in the medical field, with deep learning and machine learning emerging as the main algorithms. Numerous studies have been conducted to assess the nature of thyroid nodules and nodule classification using deep learning with ultrasonographic images as input (14, 15). Machine learning is another strategy for diagnosing or predicting thyroid nodules, given its distinctive discrimination efficiency (16). It harnesses data and algorithms to mimic the way that humans learn and make predictions or classifications (17, 18). Moreover, machine learning algorithms have emerged as an alternative to conventional logistic regression analysis for clinical risk prediction models. Despite the extensive exploration of AI in thyroid domain, there is scarce literature addressing the most challenging cohort, the AUS thyroid nodules (16). To date, no reported study has undertaken a comparison of the prediction performance between conventional logistic regression (LR) models and machine learning models (ML) in AUS thyroid nodules.

This study aims to investigate the potential of a multi-faceted approach, incorporating US findings, clinical data, biological features, and genetic information, to predict the risk of malignancy in AUS thyroid nodules. Through a comparative analysis of prediction performance between conventional logistic regression models and machine learning models, our goal is to develop a comprehensive prediction model. The final model will be presented in the form of a nomogram, providing clinicians with a practical tool for assessing AUS thyroid nodules following ultrasound-guided fine needle aspiration (US-FNA).

2 Methods

2.1 Study design and participants

This retrospective, single-center study involved consecutive patients who underwent thyroidectomy or thyroid lobectomy at Ruijin Hospital of Shanghai Jiao Tong University between January and December of 2021. The study protocol received approval from the Ethics Committee of Ruijin Hospital. All eligible patients were informed about the use of their data for study and had the option to decline to participate. Informed consent was obtained through a signed agreement before undergoing US-FNA and surgery. Inclusion criteria were: i) age> 18 years; ii) thyroid nodule classified as AUS following FNA cytology; iii) patients who underwent pre-operative US and molecular tests; iv) patients with documented surgery records and final pathology. Exclusion criteria comprised: i) refusal to participate; ii) patients who had undergone thyroid nodule ablation, radioactive iodine, or any other interventional procedure before surgery. The data meeting inclusion criteria were randomly separated into training and validation cohorts at a ratio of 7: 3. Both conventional LR and ML algorithms were applied to create the prediction models in the training dataset, followed by validation in the validation dataset. The final prediction model was chosen based on their performance, complexity, generalization capabilities, and practicality. Various metrics, including discrimination, calibration, and the clinical utility and value of the final prediction, were used for a comprehensive view. Additionally, a nomogram was developed for a graphical computation of the final prediction model. This study adhered to the Transparent Reporting of a multiple prediction model for Individual Prognosis or Diagnosis (TRIPOD) guidelines (19). The workflow of this study is depicted in Figure 1.

2.2 Ultrasonography

Preoperative neck US was performed in all patients using commercial diagnostic US equipment from various manufacturers (Mindray, Philips, GE Healthcare, Siemens, and HITACHI) with a high-frequency (5-14 MHz) linear array transducer. Settings such as depth, gain, and others were standardized as much as possible across all gray-scale and Color Doppler US examinations. Imaging of the neck, including transverse and longitudinal scanning of the thyroid glands, nodules, and bilateral lymph nodes, was all digitally stored in our hospital's picture archiving and communication system. Two radiologists reviewed and documented all sonographic characteristics of thyroid nodules using the terminology outlined in the I-TIRADS (20). Additionally, specific details such as size, echotexture, ultrasonography of thyroid

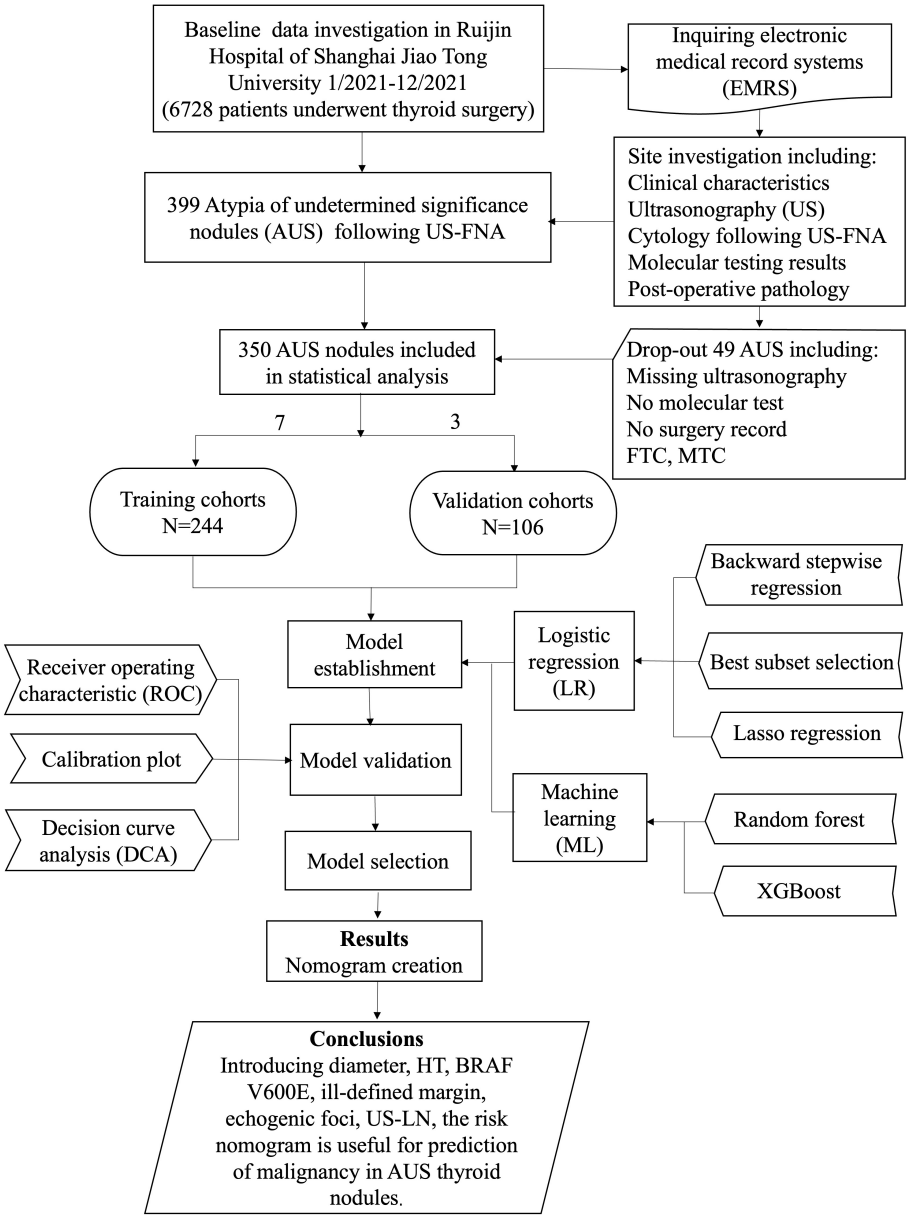

FIGURE 1
Flow diagram of study design. HT, Hashimoto's thyroiditis; US-LN, suspected cervical lymph nodes on the US; MTC, medullary thyroid carcinoma; FTC, follicular carcinoma.

parenchyma suggesting Hashimoto's thyroiditis (US-HT), and suspected cervical lymph nodes on US (US-LN) were also recorded. Any disagreements were resolved by discussion or by consultation with a third radiologist.

2.3 Cytological assessment and molecular testing following US-FNA

Nodules classified as C-TIRADS 4a or higher grades, or those exhibiting signs of suspected lymph nodes, were recommended for FNA (21). The targeted nodule underwent re-evaluation to determine its location, nearby vessels, and establish a proper needle path before aspiration. Following the guidelines (22), a 23 G needle was inserted through the skin into the nodule, guided by real-time US. Once the needle tip was positioned within the nodule, suction was applied to the needle, which rapidly moved back and forth within the nodule to obtain multiple samples of cells. The collected specimen was smeared and preserved in the liquid-based cytology preservative solution (Hangzhou HealthSky Biotechnology Co., Ltd).

Cytologists examined and analyzed the cells and cellular structures referring to the second edition of the Bethesda System for Reporting Thyroid Cytopathology (23). Molecular testing was done to provide more precise information about the nature of these thyroid nodules. Several molecular tests were available in our institution, including BRAF V600E, Ras, and TERT.

2.4 Prediction model establishment and evaluation

For the conventional LR model, several variable selection methods were applied to the training dataset for the best predictor features, including least absolute shrinkage and selection operator regression (LASSO), best subset selection, and backward stepwise regression, with all potential variables as input. Additionally, two popular machine learning algorithms, namely Random Forest (RF) and Extreme Gradient Boosting (XGBoost), were constructed to create a predictive model. To assess the discrimination capabilities of LR and ML models, the receiver operating characteristic area under the curve (ROC AUC) was used on both datasets. The forest plot was introduced to display the estimated coefficients (odds ratios) and their confidence intervals for various predictor variables in the LR model. The reliability and accuracy of the model were evaluated through a calibration curve, which compares the model's predicted outcomes or probabilities to the actual observed outcomes. This curve provides valuable insights into the model's calibration performance. The decision curve analysis (DCA) curve was utilized to evaluate the clinical utility and value of the model, offering a more comprehensive assessment of model performance than traditional metrics like sensitivity and specificity (24). DCA helped identify the most appropriate threshold probability for the model, considering user preferences for minimizing false positives or false negatives in the decision-making process and optimizing model application in practical scenarios. Finally, a nomogram was developed based on the final predictive model. This graphical tool simplifies the estimation of individual risk or probabilities based on the model's predictor variables (25).

2.5 Statistical analysis

All analyses were conducted in R, version 4.3.0 (R Project for Statistical Computing). The population characteristics were described using frequencies and proportions for categorical variables, and medians and interquartile ranges (IQR) for continuous variables. The chi-square test and Fisher's exact test were employed to compare categorical variables between the training and validation datasets. Sixteen features concerning patient demographics (age, gender), ultrasonography, and molecular testing (BRAF V600E) were recorded as potential variables to predict pathology outcomes. The "glmnet", "leaps", and "MASS" packages in R was used to select predictors by LASSO, best subset selection, and backward stepwise regression in the training dataset, respectively. Select the optimal model with the minimum criterion value (minimum AIC, BIC, or maximum adjusted R-squared). The "ROSE" package in R was applied to aid the task of binary classification in the presence of rare classes in ML establishment with over-sampling based on a bootstrap technique. The "caret" package in R was introduced to define optimal hyperparameters in the RF model using 5-fold cross-validation. The XGBoost model was fine-tuned by incorporating regularization parameters like gamma, lambda (L2 regularization term), and alpha (L1 regularization term) into a grid search to identify the optimal hyperparameters. For all tests, $P<0.05$ was considered statistically significant.

3 Results

3.1 Analysis of baseline information

From January to December 2021, a total of 6728 patients underwent thyroid surgery, and 399 thyroid nodules were diagnosed as AUS following US-FNA. Among them, 342 individuals (279 female and 77 male) with 356 AUS nodules and complete records were enrolled in this study. Approximately 90% (321/356) of the nodules were malignant, with the majority being PTC (315/321), along with a few cases of medullary thyroid carcinoma (5/321) and follicular carcinoma (1/321). Within the PTC cases, 68.6% (216/315) had BRAF V600E mutations. The remaining 10% (35/356) were benign lesions. The median age was 43.5 years (IQR 35.0-53.0), and the median nodule size was 5.1mm (IQR 3.98-7.93). Five medullary and one follicular carcinoma were excluded during predictors selection. As described in Table 1, these baseline features showed significant similarity in the training (n=244) and validation (n=106) cohorts ($p>0.05$).

3.2 Machine learning model

Figures 2A, B illustrate the relative importance of potential features in the RF and XGBoost models based on the training cohort. The performance of the random forest and XGBoost models were evaluated using AUC values on the validation set, resulting in values of 0.74 (95% CI: 0.62-0.87) for random forest (Figure 2C) and 0.74 (95% CI: 0.57-0.90) for XGBoost (Figure 2D). The RF model was figured with 100 trees (ntree = 150) and considered two random features at each split (mtry = 4). This configuration was fine-tuned based on testing and adjustments, ultimately resulting in the lowest estimation error rate of out-of-bag samples (OBB=9.02%). The XGBoost model was constructed with specific hyperparameters in terms of minimizing log loss.

3.3 Logistic regression model

In the training set, independent predictors of the LR prediction model were selected from 16 potential variables through lasso regression (Figures 3A, B), best subset selection (Figures 3C, D), and backward stepwise regression. The final predictable variables were diameter, Hashimoto's thyroiditis, BARF V600E, ill-defined margin, echogenic foci, and US-LN as described in the forest plot (Figure 3E). The AUC values of the model based on backward stepwise regression in the training cohort were 0.83 (95% CI: 0.73-0.92) (Figure 4A), and 0.80 (95% CI: 0.67-0.94) in the validation cohort (Figure 4B), respectively. As illustrated in Figures 4C, D, this model demonstrated superior calibration performance in both cohorts.

10.3389/fendo.2024.1366687

TABLE 1 Baseline characteristics of patients with AUS thyroid nodules in training and validation cohorts.

Variables	Total *N*=350	Training *N*=244 (70%)	Validation *N*=106 (n=30%)	P value
Age (median [IQR])	44.0 [35.0, 53.0]	44.0 [36.0, 54.0]	43.5 [34.25, 52.0]	0.683
Diameter (median [IQR])	5.1 [3.9, 7.6]	5.1 [4.1, 7.5]	5.3 [3.9, 8.47]	0.679
Gender (%)				
Female	277 (79.1)	191 (78.3)	86 (81.1)	0.645
Male	73 (20.9)	53 (21.7)	20 (18.9)	
Direction of growth (%)				
Wider-than-tall	94 (26.9)	68 (27.9)	26 (24.5)	0.605
Taller-than-wide	256 (73.1)	176 (72.1)	80 (75.5)	
Irregular margin (%)				
No	33 (9.4)	22 (9.0)	11 (10.4)	0.840
Yes	317 (90.6)	222 (91.0)	95 (89.6)	
Ill-defined margin (%)				
No	244 (69.7)	171 (70.1)	73 (68.9)	0.920
Yes	106 (30.3)	73 (29.9)	33 (31.1)	
Composition (%)				
Solid	329 (94.0)	229 (93.9)	100 (94.3)	1.00
Predominately solid	21 (6.0)	15 (6.1)	6 (5.7)	
Echogenicity (%)				
Markedly hypoechoic	34 (9.7)	25 (10.2)	9 (8.5)	0.799
Mildly hypoechoic	307 (87.7)	212 (86.9)	95 (89.6)	
Isoechoic	9 (2.6)	7 (2.9)	2 (1.9)	
Echotexture (%)				
Homogeneous	247 (70.6)	175 (71.7)	72 (67.9)	0.556
Heterogeneous	103 (29.4)	69 (28.3)	34 (32.1)	
Echogenic foci (%)				
No echogenic foci	198 (56.6)	142 (58.2)	56 (52.8)	0.793
Microcalcifications	104 (29.7)	70 (28.7)	34 (32.1)	
Macrocalcifications	22 (6.3)	15 (6.1)	7 (6.6)	
Micro-macro foci	26 (7.4)	17 (7.0)	9 (8.5)	
Color Doppler Flow Imaging (%)				
None	3 (0.9)	2 (0.8)	1 (0.9)	0.773
Slight	312 (89.1)	215 (88.1)	97 (91.5)	
Moderate	16 (4.6)	13 (5.3)	3 (2.8)	
Abundant	19 (5.4)	14 (5.7)	5 (4.7)	
Capsule contact (%)				
No	132 (37.7)	92 (37.7)	40 (37.7)	1.00
Yes	218 (62.3)	152 (62.3)	66 (62.3)	

(Continued)

10.3389/fendo.2024.1366687

TABLE 1 Continued

Variables	Total *N*=350	Training *N*=244 (70%)	Validation *N*=106 (n=30%)	P value
US-HT (%)				
No	255 (72.9)	182 (74.6)	73 (68.9)	0.329
Yes	95 (27.1)	62 (25.4)	33 (31.1)	
US-LN (%)				
No	278 (79.4)	195 (79.9)	83 (78.3)	0.842
Yes	72 (20.6)	49 (20.1)	23 (21.7)	
Pathology				
Benign	35 (10.0)	24 (9.8)	11 (10.4)	1.00
Malignant	315 (90)	220 (90.2)	95 (89.6)	
Hashimoto thyroiditis (%)				
No	287 (82.0)	202 (82.8)	85 (80.2)	0.667
Yes	63 (18.0)	42 (17.2)	21 (19.8)	
BRAF V600E mutation (%)				
No	125 (35.7)	85 (34.8)	40 (37.7)	0.690
Yes	225 (64.3)	159 (65.2)	66 (62.3)	

AUS, atypia of undetermined significance; US-HT, ultrasonographic of thyroid parenchyma suggesting Hashimoto's thyroiditis; US-LN, suspected cervical lymph nodes on US.

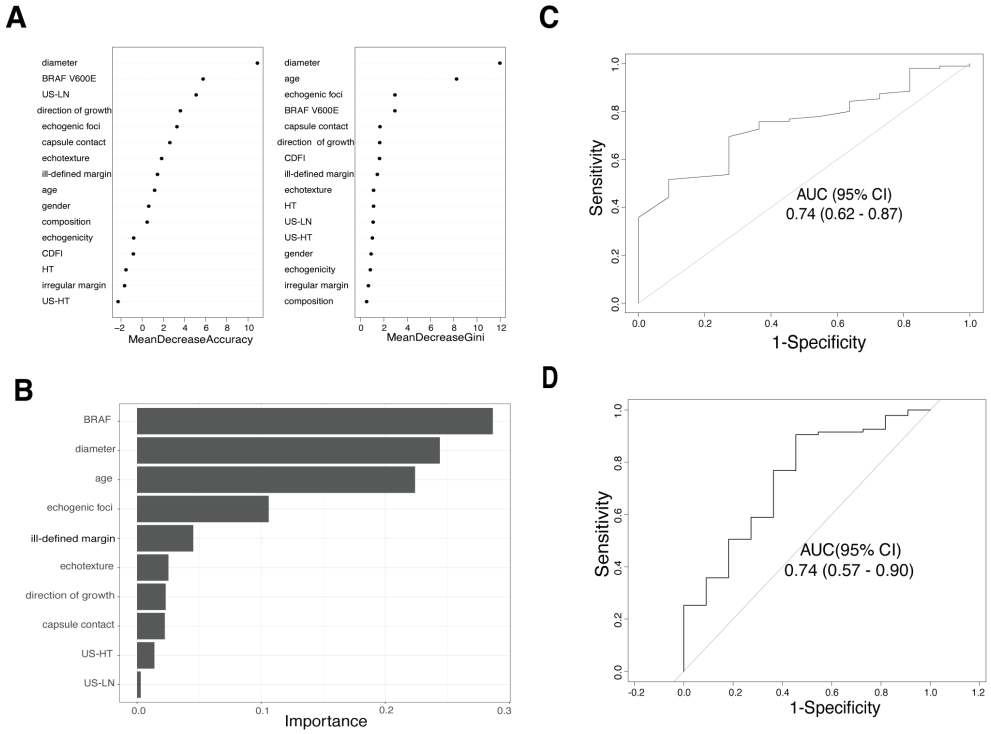

FIGURE 2
The relative contribution of predictor variables and the area under the receiver operating characteristic curve (AUC) in two Machine learning models. **(A)** Variable importance ranking plot for random forest model. **(B)** Out-of-bag variable importance ranking for the XGBoost model. **(C)** ROC curve for the random forest model in the validation cohort. **(D)** ROC curve for the XGBoost model in the validation cohort.

10.3389/fendo.2024.1366687

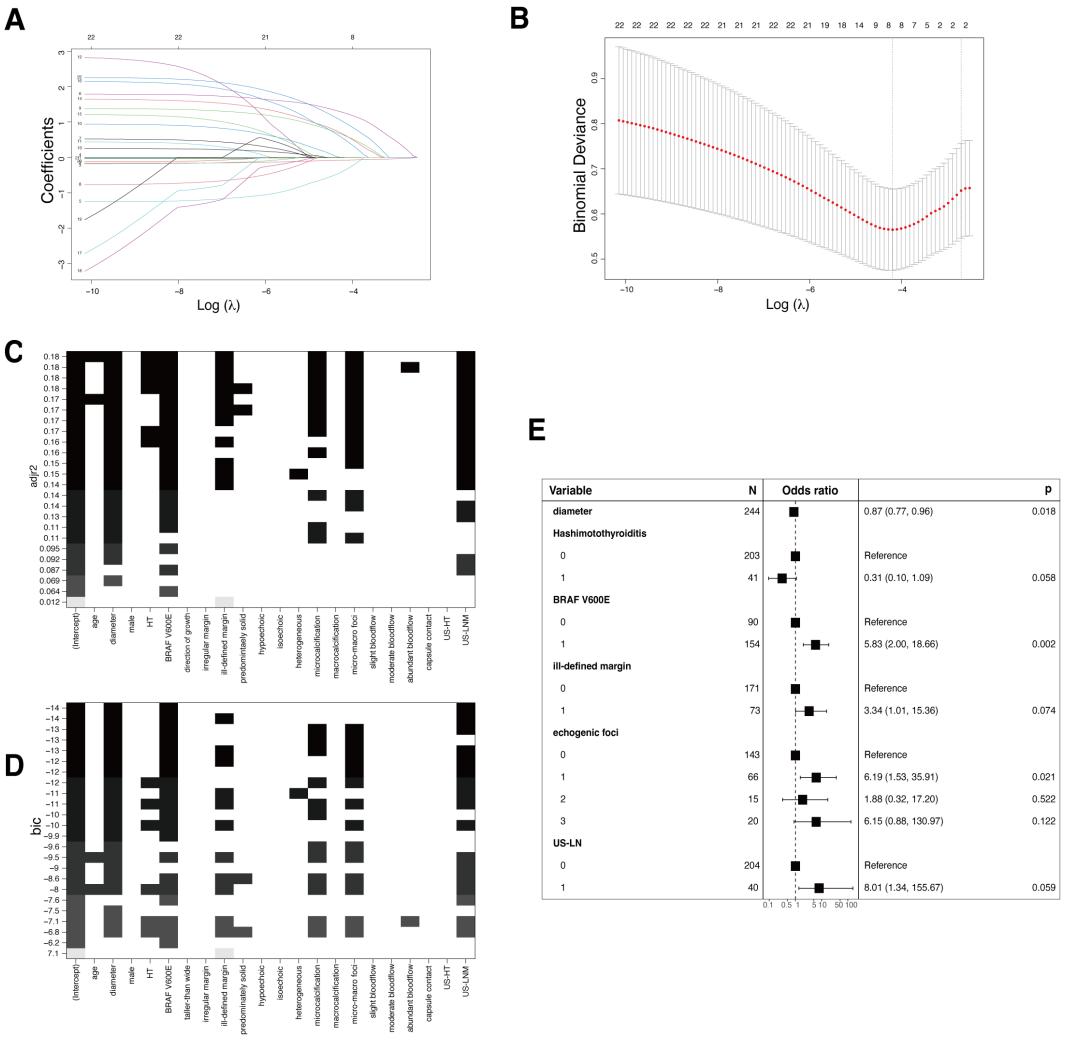

FIGURE 3
Conventional logistic regression (LR) models. **(A)** Variable screening via the LASSO binary logistic regression model is depicted, showing the coefficient profiles of 22 variables in the training cohort. **(B)** The 10-fold cross-validation process is visualized for the selection of the tuning parameter (λ) in the LASSO model. Vertical dotted lines denote the points of minimum mean square error (λ=0.015) and the standard error of the minimum distance (λ=0.067). **(C)** The process of variable selection with best subset selection regression on Adjusted R-Squared (adjr2). Seven variables were selected when the adjr2 was maximized, achieving the balance between explanatory power and model simplicity. **(D)** Best subset selection on the Bayesian Information Criteria (BIC). The same seven variables as selected based on adjr2 were chosen, as the BIC suggests a more efficient model in terms of explanatory power and complexity. **(E)** The forest plot shows the independent predictors in backward stepwise regression. Each square represents the point estimate of the effect size (e.g., odds ratio), the horizontal line indicates the 95% confidence interval, and the vertical dotted line represents the line of no effect, where the odds ratio equals 1.

3.4 Model selection

Both the RF and XGBoost models showed moderate AUC values on the validation set, indicating their ability to discriminate between malignancy and benignity, but without exceptionally high discrimination performance compared with LR models. Considering aspects such as model performance, complexity, generalization capabilities, and practicality, the LR based on backward stepwise regression prediction model was chosen finally. The forest plot depicted that certain variables with P-values exceeding 0.05, such as Hashimoto's thyroiditis, ill-defined margin, and US-LN, were retained in this logistic regression, as a result of their clinical significance in daily medical practice, and to allow for the interpretation of their sizes.

3.5 Comprehensive assessment of the final model

In both the training and internal validation set, the AUROC curve analysis yielded sensitivity values of 0.82 and 0.83, as well as specificity values of 0.75 and 0.64, respectively. The mean absolute error (MAE) for both data sets was 0.022 and 0.04 after performing bootstrapping 1000 times in the calibration plot, indicating that the model's predicted probabilities closely align with the actual outcomes across different samples (Figures 4C, D). The DCA curve was applied to both data sets to evaluate the clinical utility and value of the model. As Figure 4E revealed, the utilization of this prediction model within the threshold probability range of 20-95% in the training cohort demonstrated a higher net benefit compared

10.3389/fendo.2024.1366687

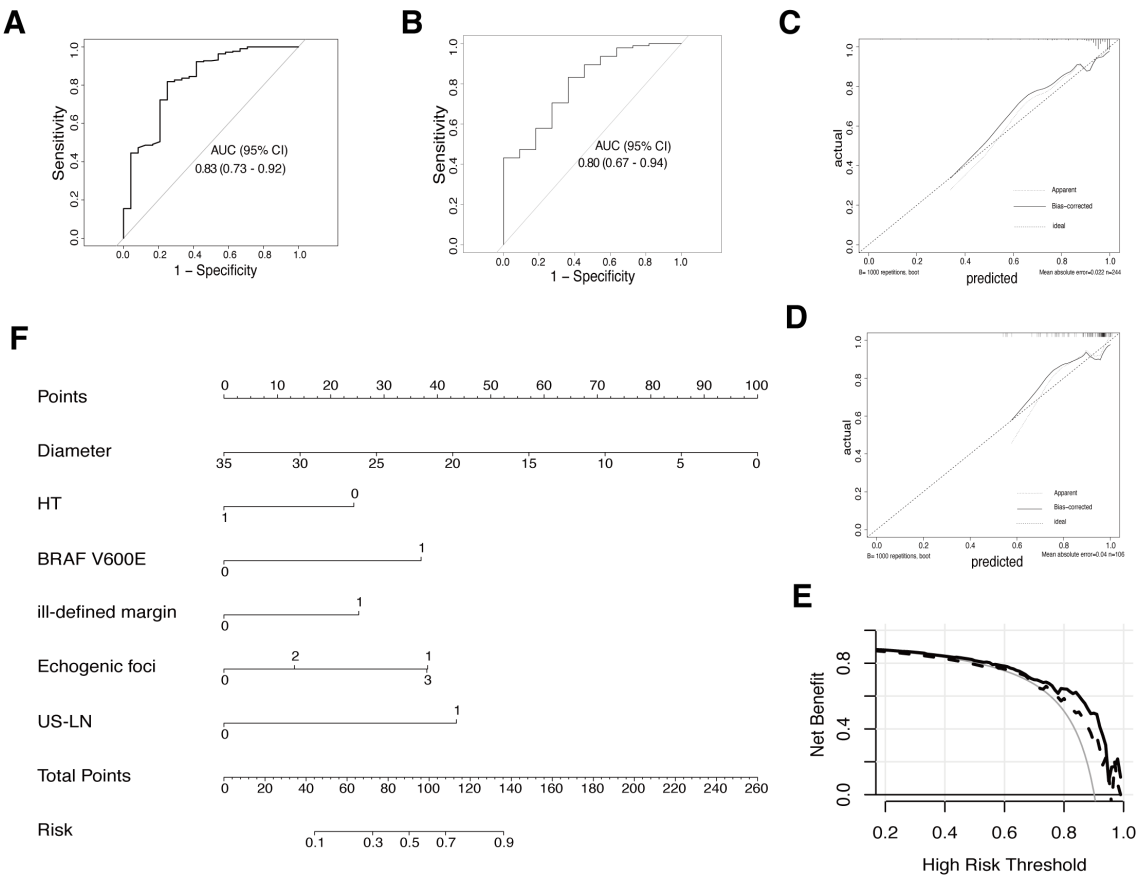

FIGURE 4
The final model (LR based on backward stepwise regression) evaluation and display. **(A)** ROC curve for final prediction model in the training cohort. **(B)** ROC curve in the validation cohort. **(C)** Calibration plot for this model in the training cohort. **(D)** The calibration plot in the validation cohort. **(E)** Decision curve analysis (DCA) for the model in both training (solid line) and validation cohorts (dashed line). **(F)** A nomogram was developed based on the final prediction model to estimate the risk of malignancy for AUS thyroid nodules. HT (0: without Hashimoto's thyroiditis, 1: the presence of Hashimoto's thyroiditis; BRAF V600E (0: wild type, 1: mutation); ill-defined margin (0: clear margin, 1: ill-defined margin); Echogenic foci (0: no echogenic foci; 1: micro-calcification; 2: macro-calcification; 3: both micro & macro calcification); US-LN (0: nonsuspicious lymph-node, 1: suspicious lymph node).

to both the screen-none and screen-all strategies. In the validation cohort, a threshold probability within 55-95% showed a favorable outcome.

3.6 Nomogram development

To depict this mathematical model vividly, a nomogram was created by assigning a weighted score to each of the predictors, as shown in Figure 4F. The total scores could be calculated as a summary of each predictor's scores, identified by drawing lines to the points axis. Refer to the risk line, the probability of malignant in AUS thyroid nodules was estimated. (For example, if a patient had Hashimoto's thyroiditis, a 20mm thyroid nodule with characteristics such as microcalcification and ill-defined margin, and the cervical lymph nodes were normal and AUS without BRAF V600E mutation after US-FNA, they would score 110 points, indicating a 72% estimated malignancy risk).

4 Discussion

Thyroid nodules, particularly those categorized as Bethesda III following US-FNA, remain a clinical challenge. Despite the TBSRTC 2023 reclassification of AUS/FLUS as AUS only, along with revised management recommendations, the accurate differentiation between benign and malignant nodules is still crucial for clinical decision-making (1). Therefore, we constructed a model to predict malignancy in AUS nodules, aiming to provide a reference for patients and physicians. Although several studies have been conducted to address this challenge, many have been limited by small sample sizes and a lack of comprehensive postoperative pathology results and other essential features (26–30). In our study, we take a more extensive approach by analyzing a large dataset of AUS nodules, incorporating comprehensive clinical, ultrasound, genetic, and pathological data. The nodules were randomly divided into training and validation cohorts to ensure the robustness of our analysis.

To develop more reliable predictive models for AUS nodules assessment, we considered several key variables selected through various methods, including LR with LASSO, backward stepwise regression, and best subset selection, as well as ML like RF and XGBoost, following a rigorous comparison. Despite the increasing popularity of ML-based models in predictive modeling, our study revealed that the LR-based nomogram exhibited superior performance in predicting malignancy of AUS thyroid nodules in terms of AUC, calibration, and clinical utility. This finding is in line with the claim that ML may not consistently outperform LR for prediction modeling (31, 32). We attempted to rectify the imbalanced training dataset in ML through oversampling and parameter adjustments, but it yielded few improvements. Furthermore, Silke Janitza and Roman Hornung suggest that OOB error may overestimate the true prediction error and raise uncertainty about its use for tuning random forest parameters, which could affect model performance (33). It remains difficult to predict whether LR would consistently outperform random forest on all future data, as we used OOB error for parameter selection.

Notably, in this study, we integrated a broad spectrum of features beyond typical ultrasound parameters outlined in I-TIRADS, recognizing the multifactorial nature of AUS nodules assessment, where clinical, imaging and genetic factors interplay intricately. The independent variables in our final predicting model encompassed diameter, Hashimoto's thyroiditis, BRAF V600E mutation status, ill-defined margin, echogenic foci, and the suspicion of cervical lymph node metastasis based on US.

With an odds ratio of 0.87 for the nodule diameter, larger AUS nodules following US-FNA seem to have a lower likelihood of malignancy compared to smaller ones. Yoon et al. introduced a nomogram incorporating clinical and ultrasound features to predict malignancy among AUS/FLUS nodules (26). Our study expands beyond typical ultrasound and clinical features, suggesting that the presence of Hashimoto's thyroiditis (odds ratio = 0.31) might contribute to the malignancy prediction model. In our predictive model, AUS nodules in the context of HT were assigned lower points compared to those without HT. Ultrasonography may be challenging for PTC in the context of HT due to its increased or decreased parenchyma echogenicity and coarsened echotexture with nodular margins. The association of HT and PTC has been a topic of controversy. On one hand, some studies have indicated that HT may elevate the risk of thyroid malignancy (34, 35). Several mechanisms have been proposed that TSH stimulation, proto-oncogenes such as BRAF mutations, and RET/PTC rearrangements may promote cancer development or growth. Chronic lymphocytic thyroiditis in HT may induce inflammation factors fueling cancer cell proliferation, while neoplastic cells could trigger a chronic inflammatory response as well (36). On the other hand, conflicting findings exist, with some reports confirming that the presence of Hashimoto's thyroiditis does not affect the risk of malignancy in thyroid nodules of category III (37). These differing findings reveal the complexity of thyroid nodule assessment coexisting with HT.

Simultaneously, the BRAF V600E mutation plays a vital role in assisting malignancy prediction. As the most extensively studied mutation in thyroid cancer, it has been reported to possess an approximate specificity of 100% in PTC (diagnostic role), leading to a significant reduction in unnecessary thyroid surgeries (7, 38). However, the isolated presence of BRAF V600E mutation is unlikely to discover the full picture of thyroid carcinogenesis due to its poor sensitivity. A meta-analysis revealed that its value as a single screening test alone in AUS is limited, with approximately 40% sensitivity in indeterminate nodules (39). Possible reasons for this limitation include the reported histopathology of carcinomas in AUS, which encompasses classic and tall cell variants of PTC, follicular thyroid carcinoma, Hurthle cell carcinoma, squamous cell, lymphoma, and others. Genetic mutations involved in these nodules also include BRAF K601E, RET-PTC, PAX8/PPARγ, RAS, etc. Incorporating a wider range of molecular markers could improve the predictive power of the models and provide deeper insights into the underlying biological mechanisms. Słowińska-Klencka et al. emphasized the importance of a holistic approach in managing category III nodules, highlighting the potential of combining miRNAs, BRAF V600E mutation, and EU-TIRADS (European Thyroid Imaging and Reporting Data System) to support clinical decision-making (40). Zhao et al. have also reported a 10.1% false-positive and a 7.1% false-negative rate of BRAF V600E mutations in thyroid FNA specimens, with a great improvement in diagnostic performance when combined with FNA cytology (41).

The nomogram, serving as a clear and intuitive visual tool, allows clinicians to easily understand and estimate probabilities without performing complex calculations. It was constructed based on the prediction model for assessing thyroid nodules. In cases where thyroid nodules are confirmed as indeterminate by US-FNA, this nomogram can be utilized as a reference for prediction.

Our models were developed with a wide range of data, including US findings, clinical data, biological features, and genetic information, and involved a thorough comparative analysis of prediction performance between conventional logistic regression models and machine learning models for a comprehensive prediction model. However, the interobserver variability and the evolving nature of medical guidelines remain inherent challenges, especially in ultrasonographic features collected. Nevertheless, many artificial intelligence (AI) technologies have emerged and are increasingly being applied to medical imaging (42, 43). Various deep learning-based AI models have demonstrated strong performance in feature extraction, serving as valuable aids for thyroid management (44). In our study, we specifically focused on patients with AUS nodules after FNA, who were initially assessed by ultrasonography and subsequently managed through surgery rather than active surveillance. All the ultrasonographic features were reported baesd on the established guidelines and clinical experience. Although LR based on backward stepwise regression was chosen finally, it is of great importance to continue research and refinement of the prediction models as new knowledge and guidelines emerge.

While this prediction model was established to reduce the need for unnecessary surgeries and alleviate patient anxiety associated with indeterminate nodules, several limitations should be considered. Firstly, although we analyzed a larger dataset of AUS nodules compared to many previous studies, a larger sample size could further enhance the generalizability of our findings. Secondly, due to a single-center, retrospective study, our results may be subject to selection bias. The availability of data, as well as

potential missing or incomplete information could have have influenced our results. Variations in clinical practices among institutions and regions may also play a role. The high malignancy rate in our cohort might have influenced the developments of the prediction model. Prospective multi-center investigations are needed to establish stronger causal relationships and external validity. Additionally, our focus on the BRAF V600E mutation, limited exploration of other prevalent molecular factors in thyroid cancer, and the observed discrepancies regarding HT and thyroid cancer risk highlight the need for more extensive molecular and mechanism research in this field. Future studies should aim to incorporate these additional mutations to improve the comprehensive and accuracy of the prediction models. Lastly, the potential for interobserver variability in ultrasound feature interpretation and the evolving nature of medical guidelines emphasize the need for ongoing research in this dynamic area.

5 Conclusion

In conclusion, our predictive model, incorporating US features, clinical data, and genetic information, offers a practical tool for assessing AUS thyroid nodules. Moreover, the model demonstrated promising performance and outperformed the machine learning algorithm. While this research represents a significant advancement in personalized medicine, reducing the ambiguity associated with indeterminate nodules and ultimately improving patient care, it highlights the need for larger-scale studies and further molecular investigations in this domain.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by The Ethics Committee of Shanghai Ruijin Hospital. The patients/participants provided their written informed consent to participate in this study. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

YCa: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. YY: Writing – review & editing, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation. YCh: Writing – review & editing, Visualization, Validation, Methodology, Investigation, Formal analysis. ML: Writing – review & editing, Visualization, Validation, Formal analysis. YH: Writing – review & editing, Validation, Methodology, Investigation, Data curation. LZ: Writing – review & editing, Visualization, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. WZ: Writing – review & editing, Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. WZ: Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by a grant from the National Natural Science Foundation of China (No. 82071923) and the Professional Department Construction Project of Huangpu District, Shanghai (No. 2023ZDZK02).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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. Ali SZ, Baloch ZW, Cochand-Priollet B, Schmitt FC, Vielh P, VanderLaan PA. The 2023 bethesda system for reporting thyroid cytopathology. *Thyroid*. (2023) 33:1039–44. doi: 10.1089/thy.2023.0141

2. Zhao H, Guo H, Zhao L, Cao J, Sun Y, Wang C, et al. (AUS/FLUS): A study of thyroid FNA cytology based on ThinPrep slides from the National Cancer Center in China. *Cancer Cytopathol.* (2021) 129:642–8. doi: 10.1002/cncy.v129.8

3. Iskandar ME, Bonomo G, Avadhani V, Persky M, Lucido D, Wang B, et al. Evidence for overestimation of the prevalence of Malignancy in indeterminate thyroid nodules classified as Bethesda category III. *Surgery*. (2015) 157:510–7. doi: 10.1016/j.surg.2014.10.004

4. Wu HH, Rose C, Elsheikh TM. The Bethesda system for reporting thyroid cytopathology: An experience of 1,382 cases in a community practice setting with the implication for risk of neoplasm and risk of Malignancy. *Diagn Cytopathol*. (2012) 40:399–403. doi: 10.1002/dc.21754

5. Huang J, Shi H, Song M, Liang J, Zhang Z, Chen X, et al. Surgical outcome and Malignant risk factors in patients with thyroid nodule classified as bethesda category III. *Front Endocrinol (Lausanne)*. (2021) 12:686849. doi: 10.3389/fendo.2021.686849

6. Kim J, Shin JH, Oh YL, Hahn SY, Park KW. Approach to Bethesda system category III thyroid nodules according to US-risk stratification. *Endocr J*. (2022) 69:67–74. doi: 10.1507/endocrj.EJ21-0300

7. Fumagalli C, Serio G. Molecular testing in indeterminate thyroid nodules: an additional tool for clinical decision-making. *Pathologica*. (2023) 115:205–16. doi: 10.32074/1591-951X-887

8. Vignali P, Macerola E, Poma AM, Sparavelli R, Basolo F. Indeterminate thyroid nodules: from cytology to molecular testing. *Diagnostics (Basel)*. (2023) 13:3008. doi: 10.3390/diagnostics13183008

9. Adeniran AJ, Hui P, Chhieng DC, Prasad ML, Schofield K, Theoharis C. BRAF mutation testing of thyroid fine-needle aspiration specimens enhances the predictability of Malignancy in thyroid follicular lesions of undetermined significance. *Acta Cytol*. (2011) 55:570–5. doi: 10.1159/000333274

10. Suh YJ, Choi YJ. Strategy to reduce unnecessary surgeries in thyroid nodules with cytology of Bethesda category III (AUS/FLUS): a retrospective analysis of 667 patients diagnosed by surgery. *Endocrine*. (2020) 69:578–86. doi: 10.1007/s12020-020-02300-w

11. Trimboli P, Scappaticcio L, Treglia G, Guidobaldi L, Bongiovanni M, Giovanella L. Testing for BRAF (V600E) mutation in thyroid nodules with fine-needle aspiration (FNA) read as suspicious for Malignancy (Bethesda V, thy4, TIR4): a systematic review and meta-analysis. *Endocr Pathol*. (2020) 31:57–66. doi: 10.1007/s12022-019-09596-z

12. Rashid FA, Munkhdelger J, Fukuoka J, Bychkov A. Prevalence of BRAF(V600E) mutation in Asian series of papillary thyroid carcinoma-a contemporary systematic review. *Gland Surg*. (2020) 9:1878–900. doi: 10.21037/gs-20-430

13. Chen H, Song A, Wang Y, He Y, Tong J, Di J, et al. BRAF(V600E) mutation test on fine-needle aspiration specimens of thyroid nodules: Clinical correlations for 4600 patients. *Cancer Med*. (2022) 11:40–9. doi: 10.1002/cam4.v11.1

14. Xu W, Jia X, Mei Z, Gu X, Lu Y, Fu CC, et al. Generalizability and diagnostic performance of AI models for thyroid US. *Radiology*. (2023) 307:e221157. doi: 10.1148/radiol.221157

15. Choi YJ, Baek JH, Park HS, Shim WH, Kim TY, Shong YK, et al. A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of thyroid nodules on ultrasound: initial clinical assessment. *Thyroid*. (2017) 27:546–52. doi: 10.1089/thy.2016.0372

16. Sorrenti S, Dolcetti V, Radzina M, Bellini MI, Frezza F, Munir K, et al. Artificial intelligence for thyroid nodule characterization: where are we standing? *Cancers (Basel)*. (2022) 14:3357. doi: 10.3390/cancers14143357

17. Yoon J, Lee E, Koo JS, Yoon JH, Nam KH, Lee J, et al. Artificial intelligence to predict the BRAFV600E mutation in patients with thyroid cancer. *PloS One*. (2020) 15: e0242806. doi: 10.1371/journal.pone.0242806

18. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. *Science*. (2015) 349:255–60. doi: 10.1126/science.aaa8415

19. 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) 350:g7594. doi: 10.1161/CIRCULATIONAHA.114.014508

20. Durante C, Hegedüs L, Na DG, Papini E, Sipos JA, Baek JH, et al. International expert consensus on US lexicon for thyroid nodules. *Radiology*. (2023) 309:e231481. doi: 10.1148/radiol.231481

21. Zhou J, Yin L, Wei X, Zhang S, Song Y, Luo B, et al. 2020 Chinese guidelines for ultrasound Malignancy risk stratification of thyroid nodules: the C-TIRADS. *Endocrine*. (2020) 70:256–79. doi: 10.1007/s12020-020-02441-y

22. Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedüs L, et al. American association of clinical endocrinologists, american college of endocrinology, and associazione medici endocrinologi medical guidelines for clinical practice for the diagnosis and management of thyroid nodules - 2016 update appendix. *Endocr Pract*. (2016) 22:622–39. doi: 10.4158/EP161208.GL

23. Cibas ES, Ali SZ. The 2017 bethesda system for reporting thyroid cytopathology. *Thyroid*. (2017) 27:1341–6. doi: 10.1089/thy.2017.0500

24. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and interpreting decision curve analysis: A guide for investigators. *Eur Urol*. (2018) 74:796–804. doi: 10.1016/j.eururo.2018.08.038

25. Kattan MW, Marasco J. What is a real nomogram? *Semin Oncol*. (2010) 37:23–6. doi: 10.1053/j.seminoncol.2009.12.003

26. Yoon JH, Lee HS, Kim EK, Moon HJ, Kwak JY. A nomogram for predicting Malignancy in thyroid nodules diagnosed as atypia of undetermined significance/follicular lesions of undetermined significance on fine needle aspiration. *Surgery*. (2014) 155:1006–13. doi: 10.1016/j.surg.2013.12.035

27. Guarnotta V, La Monica R, Ingrao VR, Di Stefano C, Salzillo R, Pizzolanti G, et al. Ultrasound parameters can accurately predict the risk of Malignancy in patients with "Indeterminate TIR3b" Cytology nodules: A prospective study. *Int J Mol Sci*. (2023) 24:8296. doi: 10.3390/ijms24098296

28. Batawil N, Alkordy T. Ultrasonographic features associated with Malignancy in cytologically indeterminate thyroid nodules. *Eur J Surg Oncol*. (2014) 40:182–6. doi: 10.1016/j.ejso.2013.11.015

29. Liu X, Wang J, Du W, Dai L, Fang Q. Predictors of Malignancy in thyroid nodules classified as bethesda category III. *Front Endocrinol (Lausanne)*. (2022) 13:806028. doi: 10.3389/fendo.2022.806028

30. Yousefi E, Sura GH, Somma J. The gray zone of thyroid nodules: Using a nomogram to provide Malignancy risk assessment and guide patient management. *Cancer Med*. (2021) 10:2723–31. doi: 10.1002/cam4.v10.8

31. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. (2019) 110:12–22. doi: 10.1016/j.jclinepi.2019.02.004

32. Pua YH, Kang H, Thumboo J, Clark RA, Chew ES, Poon CL, et al. Machine learning methods are comparable to logistic regression techniques in predicting severe walking limitation following total knee arthroplasty. *Knee Surg Sports Traumatol Arthrosc*. (2020) 28:3207–16. doi: 10.1007/s00167-019-05822-7

33. Janitza S, Hornung R. On the overestimation of random forest's out-of-bag error. *PloS One*. (2018) 13:e0201904. doi: 10.1371/journal.pone.0201904

34. Hu T, Li Z, Peng C, Huang L, Li H, Han X, et al. Nomogram to differentiate benign and Malignant thyroid nodules in the American College of Radiology Thyroid Imaging Reporting and Data System level 5. *Clin Endocrinol (Oxf)*. (2023) 98:249–58. doi: 10.1111/cen.14824

35. Silva de Morais N, Stuart J, Guan H, Wang Z, Cibas ES, Frates MC, et al. The impact of hashimoto thyroiditis on thyroid nodule cytology and risk of thyroid cancer. *J Endocr Soc*. (2019) 3:791–800. doi: 10.1210/js.2018-00427

36. Vita R, Ieni A, Tuccari G, Benvenga S. The increasing prevalence of chronic lymphocytic thyroiditis in papillary microcarcinoma. *Rev Endocr Metab Disord*. (2018) 19:301–9. doi: 10.1007/s11154-018-9474-z

37. Słowińska-Klencka D, Popowicz B, Klencki M. Hashimoto's thyroiditis does not influence the Malignancy risk in nodules of category III in the bethesda system. *Cancers (Basel)*. (2022) 14:1971. doi: 10.3390/cancers14081971

38. Ferraz C. Molecular testing for thyroid nodules: Where are we now? *Rev Endocr Metab Disord*. (2023) 25:149–59. doi: 10.1007/s11154-023-09842-0

39. Jinih M, Foley N, Osho O, Houlihan L, Toor AA, Khan JZ, et al. BRAF(V600E) mutation as a predictor of thyroid Malignancy in indeterminate nodules: A systematic review and meta-analysis. *Eur J Surg Oncol*. (2017) 43:1219–27. doi: 10.1016/j.ejso.2016.11.003

40. Słowińska-Klencka D, Popowicz B, Kulczycka-Wojdala D, Szymańska B, Duda-Szymańska J, Wojtaszek-Nowicka M, et al. Effective use of microRNA, BRAF and sonographic risk assessment in bethesda III thyroid nodules requires a different approach to nodules with features of nuclear atypia and other types of atypia. *Cancers (Basel)*. (2023) 15:4287. doi: 10.3390/cancers15174287

41. Zhao CK, Zheng JY, Sun LP, Xu RY, Wei Q, Xu HX. BRAF(V600E) mutation analysis in fine-needle aspiration cytology specimens for diagnosis of thyroid nodules: The influence of false-positive and false-negative results. *Cancer Med*. (2019) 8:5577–89. doi: 10.1002/cam4.v8.12

42. Yao J, Zhang Y, Shen J, Lei Z, Xiong J, Feng B, et al. AI diagnosis of Bethesda category IV thyroid nodules. *iScience*. (2023) 26:108114. doi: 10.1016/j.isci.2023.108114

43. Peng S, Liu Y, Lv W, Liu L, Zhou Q, Yang H, et al. Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study. *Lancet Digit Health*. (2021) 3:e250–e9. doi: 10.1016/S2589-7500(21)00041-8

44. Yao J, Lei Z, Yue W, Feng B, Li W, Ou D, et al. DeepThy-net: A multimodal deep learning method for predicting cervical lymph node metastasis in papillary thyroid cancer. *Advanced Intelligent Systems*. (2022) 4:2200100. doi: 10.1002/aisy.202200100