scientific reports

Check for updates

OPEN

Integrated intraoperative predictive model for malignancy risk assessment of thyroid nodules with atypia of undetermined significance cytology

Cheng Li[1]✉, Yong Luo[1], Yan Jiang[2] & Qi Li[1]

Management of thyroid nodules with atypia of undetermined significance/follicular lesion of undetermined significance (AUS/FLUS) cytology is challenging because of uncertain malignancy risk. Intraoperative frozen section pathology provides real-time diagnosis for AUS/FLUS nodules undergoing surgery, but its accuracy is limited. This study aimed to develop an integrated predictive model combining clinical, ultrasound and IOFS features to improve intraoperative malignancy risk assessment. A retrospective cohort study was conducted on patients with AUS/FLUS cytology and negative *BRAF*^{V600E} mutation who underwent thyroid surgery. The cohort was randomly divided into training and validation sets. Clinical, ultrasound, and pathological features were extracted for analysis. Three models were developed: an IOFS model with IOFS results as sole predictor, a clinical model integrating clinical and ultrasound features, and an integrated model combining all features. Model performance was evaluated using comprehensive metrics in both sets. The superior model was visualized as a nomogram. Among 531 included patients, the integrated model demonstrated superior diagnostic ability, predictive performance, calibration, and clinical utility compared to other models. It exhibited AUC values of 0.92 in the training set and 0.95 in the validation set. The nomogram provides a practical tool for estimating malignancy probability intraoperatively. This study developed an innovative integrated predictive model for intraoperative malignancy risk assessment of AUS/FLUS nodules. By combining clinical, ultrasound, and IOFS features, the model enhances IOFS diagnostic sensitivity, providing a reliable decision-support tool for optimizing surgical strategies.

Keywords AUS/FLUS, Intraoperative, Thyroid Cancer, Malignancy risk assessment, Nomogram predictive model

Thyroid cancer represents one of the most rapidly proliferating malignancies globally, with an estimated 586,000 new cases reported worldwide in 2020, constituting 3% of all new cancer diagnoses[1,2]. Precise and timely diagnosis is paramount for the effective treatment and management of thyroid cancer. Fine-needle aspiration (FNA) biopsy, interpreted using the Bethesda System for Reporting Thyroid Cytopathology (TBSRTC)[3], is widely acknowledged as the gold standard for diagnosing ultrasound-detected suspicious thyroid nodules. This method has demonstrated remarkable diagnostic accuracy, with reported sensitivities ranging from 0.72 to 0.93 and specificities from 0.96 to 0.99[4–6]. However, notable challenges emerge in the management of patients with FNA results indicating "atypia of undetermined significance/follicular lesion of undetermined significance" (AUS/FLUS) (TBSRTC category III). The reported malignancy risk for AUS/FLUS nodules spans from 22.6% to 37.8%[7–9], introducing substantial uncertainty in clinical decision-making.

In cases where AUS/FLUS nodules require surgical intervention, such as those exhibiting recurrent AUS/FLUS results, rapid growth, compressive symptoms, or cosmetic concerns, intraoperative frozen section pathology (IOFS) is typically performed to guide surgeons in determining the optimal extent of surgery. Although IOFS provides immediate diagnostic information, enabling surgeons to make immediate decisions regarding the necessity for total thyroidectomy, lobectomy, or central lymph node dissection, the diagnostic accuracy of IOFS

[1]Department of Thyroid Surgery, Ningbo Medical Center Lihuili Hospital, Ningbo, China. [2]Department of Ultrasound, Ningbo Medical Center Lihuili Hospital, Ningbo, China. ✉email: dr.cheng_li@outlook.com

for AUS/FLUS nodules remains limited. Previous studies have reported a wide range of sensitivity values for IOFS in diagnosing malignancy in AUS/FLUS nodules, with figures varying from 0.22 to 0.54[10–12]. This low sensitivity may result in the misdiagnosis of malignant cases during surgery, which could lead to suboptimal surgical intervention, an increased risk of reoperation, and elevated healthcare costs. It is therefore imperative to enhance the sensitivity of IOFS in diagnosing malignancy in AUS/FLUS nodules in order to optimize surgical decision-making and improve patient outcomes.

It was postulated that an integrated predictive model, which would combine clinical features, imaging findings, and IOFS results, could significantly enhance the sensitivity of IOFS alone in diagnosing malignancy in AUS/FLUS nodules intraoperatively. This hypothesis is based on previous research that has identified correlations between specific clinical features (such as nodule size[13] and patient age[14,15]) and ultrasound characteristics (including echogenicity, margins, and calcifications[13–18]) with the malignancy risk of AUS/FLUS nodules. While previous studies have attempted to develop predictive models that incorporate clinical and ultrasound features[14,18], a comprehensive model that integrates all three essential diagnostic components-clinical features, ultrasound features, and IOFS results-remains conspicuously absent from the literature. This represents a significant gap in clinical decision support tools, particularly because these diagnostic modalities offer complementary perspectives that, when synthesized, could yield significantly more accurate risk assessments than current single-modality or dual-modality approaches. Our study addresses this crucial gap by developing the first integrated predictive model that systematically combines these three key diagnostic elements for real-time intraoperative decision-making in AUS/FLUS nodules. This novel approach strategically leverages the complementary strengths of each modality while mitigating their individual limitations, thereby potentially offering a more robust and reliable clinical assessment tool.

Accordingly, the objective of this study was to develop an integrated intraoperative predictive model specifically for thyroid nodules with FNA-indicated AUS/FLUS cytology, where diagnostic uncertainty has the greatest impact on surgical decision-making. This innovative model incorporates clinical features, ultrasound imaging features, and IOFS results, and is designed to provide surgeons with a more reliable decision-support tool, optimize surgical strategies, and ultimately improve both immediate surgical management and long-term patient prognosis in cases of AUS/FLUS thyroid nodules.

Methods

This retrospective cohort study was conducted with approval from the Ethics Committee of Ningbo Medical Center Lihuili Hospital. All methods were carried out in accordance with relevant guidelines and regulations. The requirement for informed consent was waived by the Ethics Committee of Ningbo Medical Center Lihuili Hospital due to the retrospective nature of the study, the minimal risk to participants, the fact that all data were collected from routine clinical care, and that all patient data were analyzed in a de-identified manner with no individual patient data presented. We rigorously adhered to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement for model development, validation, and reporting[19]. The study cohort comprised patients who underwent thyroid surgery at Ningbo Medical Center Lihuili Hospital, a comprehensive medical, teaching, and research institution in Ningbo, China, which performs approximately 2,000 thyroid surgeries annually.

The diagnosis and treatment protocols for thyroid nodules at our institution adhere the guidelines of the American Thyroid Association (ATA)[20]. Ultrasound reports adhere to the standardized Thyroid Imaging Reporting and Data System (TI-RADS)[21]. FNA is recommended for nodules classified as TI-RADS category 4A or higher and exceeding 5 mm in size. In cases where patients have been diagnosed with AUS/FLUS via FNA and have a negative $BRAF^{V600E}$ mutation status, surgical intervention is advised if they meet specific criteria, including persistent AUS/FLUS classification following repeated FNA, rapid nodule growth, compressive symptoms, cosmetic concerns, or other clinical indications necessitating surgical intervention. In these cases, IOFS analysis is conducted as a routine procedure to inform surgical decision-making.

A comprehensive physical examination, high-resolution ultrasound imaging, and thyroid function tests is performed as part of the preoperative evaluation, which also includes assessments of thyroglobulin antibody (TgAb) and thyroid peroxidase antibody (TPOAb). Ultrasound examinations are conducted by experienced institutional sonographers, with images stored in DICOM format in the hospital's Picture Archiving and Communication Systems (PACS). All pathological diagnostic services, including FNA cytology, molecular biology, IOFS, and postoperative paraffin section (PS), are provided by the Ningbo Clinical Pathology Diagnostic Center. Cytological results are classified in accordance with the latest Bethesda System for Reporting Thyroid Cytopathology (TBSRTC)[3], while pathological assessment adheres to the World Health Organizatio (WHO) classification criteria[22].

Study cohort establishment

We conducted a comprehensive retrospective analysis of all patients who underwent thyroid surgery at our institution between January 1, 2019, and December 31, 2023. The selection of study population was guided by comprehensive clinical considerations, including surgical indications, molecular profiles, imaging characteristics, and pathological features. Specifically, we focused on patients with AUS/FLUS cytology who required surgical intervention due to clinical indications such as persistent diagnostic uncertainty, compressive symptoms, or nodule progression. Among these patients, we specifically selected $BRAF^{V600E}$ negative cases where diagnostic uncertainty has the greatest clinical impact, as $BRAF^{V600E}$ positive nodules demonstrate exceptionally high positive predictive value for malignancy[23]. While other molecular markers exist, their limited independent predictive value[24,25] and non-routine implementation in our institutional practice led to their exclusion from our analysis.

Patients were included in the study cohort if they met all the following inclusion criteria and did not meet any exclusion criteria. The inclusion criteria were: 1) preoperative FNA result of AUS/FLUS (TBSRTC category III); 2) surgery performed due to repeated AUS/FLUS results, rapid growth, compressive symptoms, cosmetic concerns, or other clinical indications necessitating surgical intervention; 3) negative $BRAF^{V600E}$ mutation status; 4) IOFS performed intraoperatively. The exclusion criteria were: 1) incomplete clinical data; 2) missing ultrasound imaging data for the AUS/FLUS nodule; 3) inconclusive pathological results by PS; 4) the presence of other nodules in the ipsilateral or contralateral thyroid classified as TI-RADS 4A or higher, FNA result of Bethesda III or higher, positive $BRAF^{V600E}$ mutation, or IOFS confirming indeterminate or malignant status; 5) preoperative confirmation of central (level VI) or lateral neck (levels II-V) lymph node metastasis; 6) preoperative confirmation of distant thyroid cancer metastasis; 7) history of ipsilateral or contralateral thyroid surgery (whether benign or malignant).

Data collection and processing

Following an extensive literature review, we identified key clinical and ultrasound features that may serve as predictors of malignancy risk in AUS/FLUS nodules. These included age[14,15,26,27], nodule size[13,15], gender[13], presence of Hashimoto's thyroiditis[27], and ultrasound features such as echogenicity[13,16], composition[13,18], shape[13,16,27], margin[13,14,16–18,27], and calcifications[13,14,16–18].

The clinical, ultrasound, and pathological features were meticulously extracted from the hospital information system. The ultrasound features reported in the or original reports were independently reviewed and evaluated by a specialist in thyroid ultrasound with 15 years of experience, who was unaware of the pathology results, using the original images from the PACS system. Any discrepancies between the original reports and the re-evaluation were resolved through consensus with a senior sonographer. The presence of Hashimoto's thyroiditis was determined based on the presence of positive TgAb or TPOAb results. The quantification of these tests was conducted using the Siemens ADVIA Centaur XP Immunoassay System via an immunochemiluminescent assay. Patients with TgAb levels exceeding 4.5 IU/mL or TPOAb levels above 60 U/mL were classified as having Hashimoto's thyroiditis. The size of the nodule was determined by the maximum diameter measured by ultrasound and categorized as $\leq$ 1 cm (T1a), >1 cm and $\leq$2 cm (T1b), and >2 cm (>T2). IOFS results were classified as indeterminate in cases where the diagnosis was described as "suspicious but inconclusive," "deferred diagnosis," "further evaluation required," or "definitive resection recommended." PS diagnosis adhered to the WHO classification of thyroid tumors, 5th edition. Tumors of uncertain malignant potential were classified as benign. To minimize the potential for data entry errors, all data were independently entered by two investigators into a Microsoft Access database and automatically cross-checked for discrepancies.

Model development

The study cohort was randomly divided into a training (70%) and a validation (30%) sets using a 7:3 ratio. The training set was employed for the development of the model, while the validation set was utilized for the evaluation of the model's performance. We initially conducted univariate and multivariate analyses to assess the ability of IOFS to independently predict malignancy risk. Subsequently, the Least Absolute Shrinkage and Selection Operator (LASSO) regression model[28] was employed to identify the most significant predictors for the clinical (incorporating clinical and ultrasound features) and integrated models (combining clinical, ultrasound, and IOFS features). LASSO utilizes the penalty parameter λ to shrink the coefficients of less important predictors to zero, thereby identifying the most relevant predictors while mitigating the risk of overfitting, particularly in scenarios with numerous predictors and limited samples[29]. The optimal value of λ was identified as the point of minimal binomial deviance using 10-fold cross-validation, as this value optimally balances model complexity and goodness of fit. The variables with non-zero coefficients at the optimal λ were selected for inclusion in the clinical and integrated models, respectively, and subsequently used to construct logistic regression models with PS results as the dependent variable. The IOFS model was constructed as a logistic regression model, with IOFS results serving as the sole predictor variable.

The optimal threshold probability for each model was determined by identifying the point on the Receiver Operating Characteristic (ROC) curve that was closest to the top-left corner in the training set, maximizing Youden's J statistic and balancing the trade-off between sensitivity and specificity. We calculated The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1 score were calculated for each model in both the training and validation sets.

Model evaluation and comparison

We employed a comprehensive set of metrics to evaluate model performance, including goodness of fit, predictive ability, diagnostic accuracy, calibration, clinical utility, and predictive improvement.

The Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and McFadden's pseudo-R^2 were used to assess the goodness of fit of the models in the training set. A lower AIC and BIC value, coupled with a higher McFadden's pseudo-R^2, indicate superior model fit. The Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) were calculated for both the training and validation sets to evaluate predictive ability. A lower MSE and RMSE value in both sets suggest enhanced predictive ability and generalizability.

To assess diagnostic accuracy, both training and validation sets were evaluated using ROC curves and the area under the curve (AUC). Higher AUC values indicate superior discrimination between benign and malignant cases. Calibration curves were employed to assess the calibration of the models in both sets. The ideal calibration is represented by a 45-degree diagonal line, indicating that the model predictions perfectly align with the observed results. A decision curve analysis (DCA) was conducted to assess the clinical utility of the models in both sets. DCA assesses the net benefit of utilizing the model to inform clinical decision-making in

comparison to the alternative approaches of treating all patients or none. A model that demonstrates a higher net benefit across a range of threshold probabilities is considered to have superior clinical utility.

Lastly, the AUC difference (ΔAUC), continuous Net Reclassification Improvement (NRI), and Integrated Discrimination Improvement (IDI) were calculated to assess the predictive improvement between the models. Significantly higher ($p < 0.05$) ΔAUC, NRI, and IDI values indicate that the model provides enhanced discrimination, reclassification, and prediction probability, respectively, compared to other models.

Statistical analysis

All statistical analyses were performed using R software (version 4.3.1)[30]. We utilized several R packages for specific analyses: `gtsummary`[31] for univariate and multivariate analyses, `networkD3`[32] for Sankey diagram generation, `glmnet`[33] for LASSO regression, `rms`[34] for logistic regression and nomogram construction, `Metrics` for RMSE and MSE calculations[35], `pROC`[36] for ROC curve analysis, `probably` for calibration curves[37], `PredictABEL` for NRI and IDI calculations[38], and `rmda` for DCA[39].

Continuous variables were presented as mean ± standard deviation (SD), while categorical variables were presented as frequencies and percentages. The Student's *t*-test was employed to compare continuous variables, while the χ^2 test or Fisher's exact test was used for categorical variables. A two-tailed p value < 0.05 was considered statistically significant.

Results

Study cohort and baseline characteristics

A total of 531 patients met the inclusion criteria and were included in the study cohort. The inclusion and exclusion flowchart is shown in Figure 1. The mean age of the cohort was 48.2 ± 10.9 years, with 427 (80.4%) female participants. The concordance between IOFS and PS results is statistically presented in Table 1 and visualized in Figure 2.

The IOFS diagnosis identified 292 patients (55.0%) as malignant, 129 (24.3%) as indeterminate, and 110 (20.7%) as benign. PS confirmed 397 patients (74.8%) as malignant and 134 (25.2%) as benign. Consequently, IOFS demonstrated a sensitivity of 0.73 (292/397) and a PPV of 0.99 (292/295) for malignant case diagnosis, while exhibiting a specificity of 0.60 (81/134) and an NPV of 0.74 (81/110) for benign cases. Among the indeterminate cases diagnosed by IOFS, 61.2% (79/129) were subsequently confirmed as malignant and 38.8% (50/129) as benign by PS.

The cohort was randomly divided into a training set ($n = 372$) and a validation set ($n = 159$). No significant differences in baseline characteristics were observed between the two sets, as detailed in Table 2.

Model development

The univariate analysis revealed that age, nodule size, composition, shape, margin, and IOFS results were significantly associated with PS results (all $p < 0.05$). The results of the multivariate analysis demonstrated that IOFS results and nodule shape were independently associated with PS scores (both $p < 0.05$), as shown in Table 3. These analyses confirmed that IOFS results are the most significant predictor of malignancy risk in AUS/FLUS nodules.

The cross-validation of the LASSO regression models for the clinical and integrated models showed that the optimal λ values were 0.026 and 0.014, respectively, at which the binomial deviance reached its minimum, as shown in Figure 3A and Figure 3B. At these optimal λ values, the LASSO regression models selected six non-zero coefficients (age, nodule size, composition, shape, margin, and echogenicity) for the clinical model and seven non-zero coefficients (age, composition, shape, echogenicity, macro-calcification, peripheral calcification, and IOFS results) for the integrated model, respectively, as shown in Figure 3C and Figure 3D. The logistic regression models of the clinical, IOFS, and integrated models were constructed as shown in Table 4.

Model evaluation and comparison

The goodness of fit and predictive ability metrics of the clinical, IOFS, and integrated models in the training and validation sets are shown in Table 5. The integrated model exhibited the highest McFadden's pseudo-R^2 (0.48) and relatively lower AIC (241.03) and BIC (288.06) in the training set, indicating a comparable goodness of fit to that of the other models. Additionally, the integrated model showed the lowest RMSE (0.31 in the training set and 0.29 in the validation set) and MSE (0.19 in the training set and 0.20 in the validation set), indicating the best predictive ability.

The ROC curves for the clinical, IOFS, and integrated models in the training and validation sets are shown in Figure 4. The AUC values of the clinical, IOFS, and integrated models in the training set were 0.71, 0.89, and 0.92, respectively, and in the validation set were 0.77, 0.92, and 0.95, respectively. The integrated model exhibited the highest AUC values in both the training and validation sets, indicating the best diagnostic ability.

The optimal threshold probabilities for the clinical, IOFS, and integrated models were 0.77, 0.76, and 0.65, respectively. The diagnostic performance metrics at these specified thresholds are presented in Table 6. The integrated model demonstrated superior performance across several key metrics, achieving the highest sensitivity, NPV, and F1 score in both sets. Although it exhibited slightly diminished specificity and PPV in the training set compared to the IOFS model, its performance in the validation set was comparable for these metrics.

The calibration curves are shown in Figure 5. The integrated model exhibited satisfactory calibration in both datasets, particularly for low-risk patients. However, it demonstrated less confidence in predicting high-risk patients, particularly in the validation set.

The DCA curves are presented in Figure 6. All models showed higher net benefit compared to treat-all or treat-none strategies, with the integrated model consistently demonstrating the highest net benefit across the range of threshold probabilities in both sets.

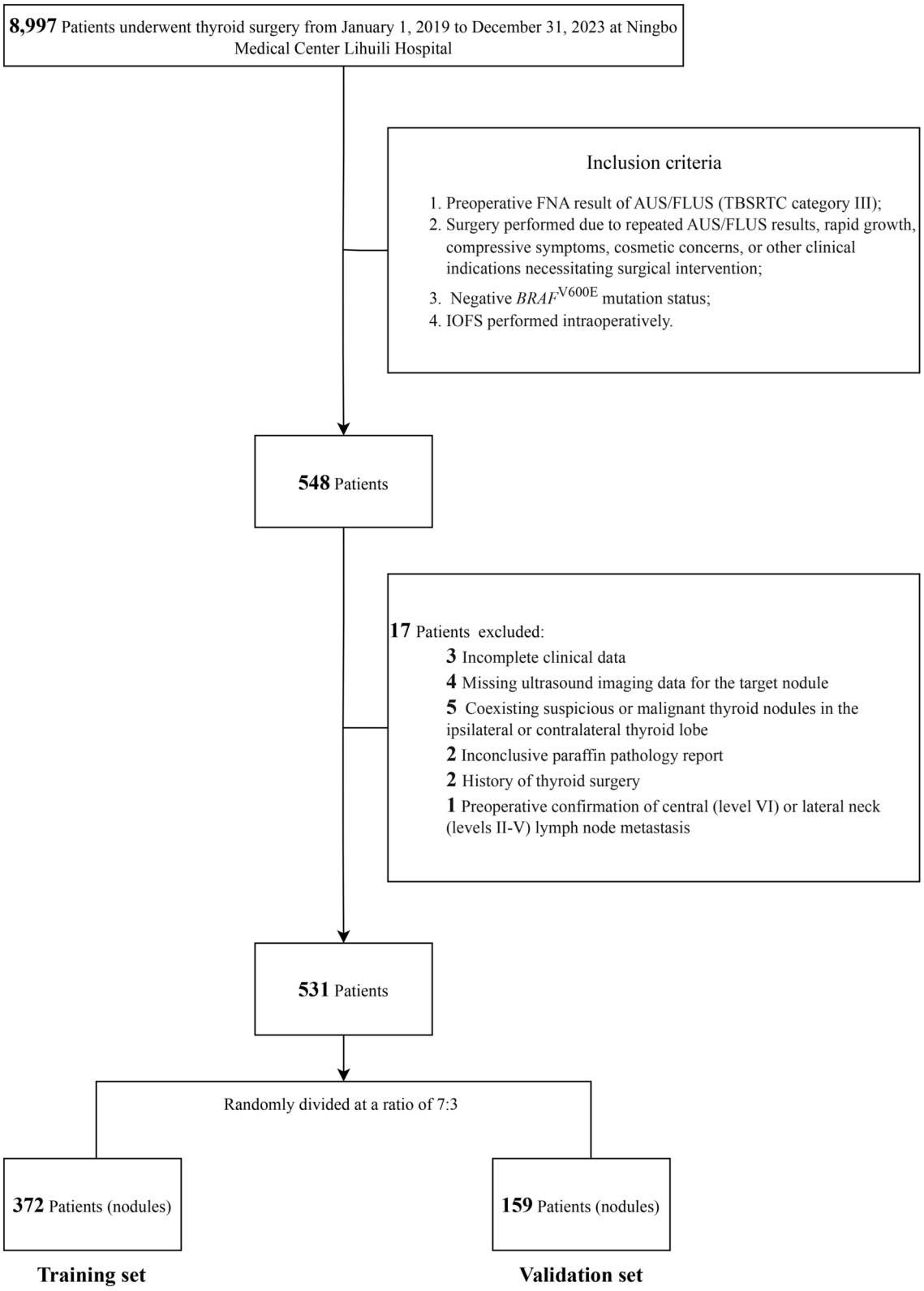

Fig. 1. Patient Selection Flowchart.

A detailed comparison of the predictive improvement between the models can be found in Table 7. The integrated model showed statistically significant improvements in ΔAUC, NRI, and IDI compared to both the clinical and IOFS models in both sets, indicating enhanced discrimination, reclassification, and predictive accuracy.

PS report	Cases	Percentage(%)
IOFS report: benign	**110**	
PS report: benign		
Follicular nodular disease	66	60.00
Hashimoto's thyroiditis[a]	13	11.82
Thyroid tumors of uncertain malignant potential	2	1.82
PS report: malignant		
Papillary thyroid carcinoma	27	24.55
Follicular thyroid carcinoma	1	0.91
Invasive encapsulated follicular variant papillary thyroid carcinoma	1	0.91
IOFS report: indeterminate[b]	**129**	
PS report: benign		
Follicular nodular disease	35	27.13
Hashimoto's thyroiditis[a]	7	5.43
Thyroid tumors of uncertain malignant potential	7	5.43
Hyalinizing trabecular tumor	1	0.78
PS report: malignant		
Papillary thyroid carcinoma	72	55.81
Invasive encapsulated follicular variant papillary thyroid carcinoma	4	3.10
Medullary thyroid carcinoma	2	1.55
Follicular thyroid carcinoma	1	0.78
IOFS report: malignancy	**292**	
PS report: benign		
Follicular nodular disease	3	1.03
PS report: malignant		
Papillary thyroid carcinoma	287	98.29
Follicular thyroid carcinoma	1	0.34
Medullary thyroid carcinoma	1	0.34

Table 1. Concordance Between IOFS Results and PS Results for AUS/FLUS Thyroid Nodules Percentage calculated based on the total number of cases in each PS report category. Some percentages may not add up to 100% due to rounding. [a]Hashimoto's thyroiditis was defined as a positive thyroglobulin antibody(>4.5IU/mL) or a positive thyroid peroxidase antibody(>60 U/mL) result via immunochemiluminescent assay. [b]In cases IOFS reported as "suspicious but inconclusive," "deferred diagnosis," "further evaluation required," or "definitive resection recommended."

Model visualization

The nomogram of the integrated model is presented in Figure 7. This visual tool facilitates intraoperative estimation of the probability of malignancy for AUS/FLUS nodules based on clinical, ultrasound, and IOFS features. The threshold probability of 0.65 is equivalent to a total score of 110 on the nomogram. Patients with a total point score exceeding 110 are deemed to have a high probability of malignancy and may necessitate more extensive surgical intervention.

Discussion

This study presents an innovative integrated predictive model for intraoperative malignancy risk assessment of thyroid nodules with AUS/FLUS cytology and negative $BRAF^{V600E}$ mutation status. While Bethesda categories III-V are all considered indeterminate cytology, we specifically focused on category III due to several compelling reasons. First, these categories exhibit markedly different risk profiles for malignancy (13-30% for category III, 23-34% for category IV, and 67-83% for category V[3]), which necessitates differentiated management approaches. Second, category IV nodules, particularly follicular neoplasms, present unique diagnostic challenges during intraoperative frozen section analysis, which our institution has previously addressed through a dedicated predictive model[40]. Third, while categories IV and V typically proceed directly to diagnostic surgical excision due to their higher malignancy risk, category III nodules require a more selective approach including repeat FNA or molecular testing, making the decision-making process particularly challenging.

By integrating clinical features, ultrasound imaging characteristics, and IOFS results, our model significantly enhances the sensitivity of IOFS in diagnosing malignancy intraoperatively for these challenging category III nodules. Our findings demonstrate that the integrated model exhibits superior diagnostic ability, predictive performance, calibration, and clinical utility compared to both the clinical and IOFS models when considered independently. This improved diagnostic accuracy has direct implications for surgical decision-making: nodules determined to be benign may be appropriately treated with lobectomy alone, while confirmation of malignancy may indicate the need for central lymph node dissection in selected cases[41]. Additionally, surgical planning

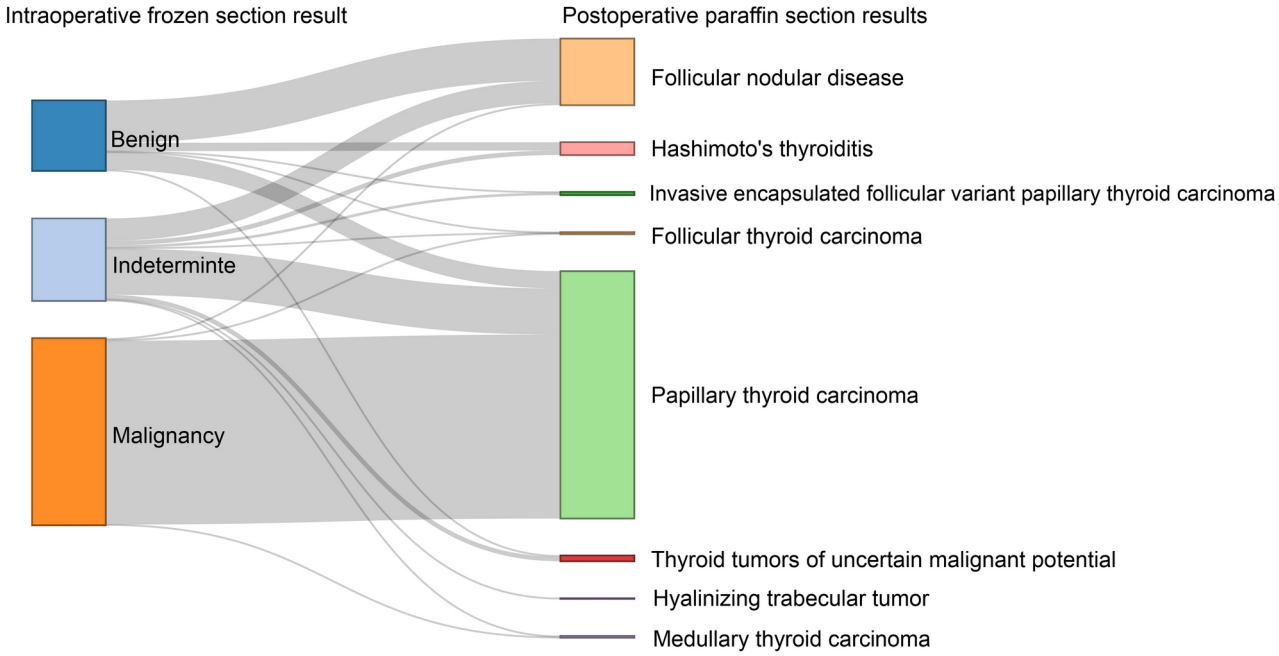

Fig. 2. Sankey Diagram of Intraoperative Frozen Section Results and Postoperative Paraffin Section Results in AUS/FLUS Thyroid Nodules. This Sankey diagram visualizes the diagnostic concordance between intraoperative frozen section results and postoperative paraffin section results in AUS/FLUS thyroid nodules. The width of the flow lines represents the proportion of nodules in each category. The detailed numbers of nodules in each category are provided in Table 1. The "Indeterminate" category includes cases where intraoperative frozen section results were reported as "suspicious but inconclusive," "deferred diagnosis," "further evaluation required," or "definitive resection recommended".

may be further modified based on factors such as multifocality or tumor size exceeding 4 cm, which might necessitate total thyroidectomy[20]. To the best of our knowledge, this represents the first study to develop such a comprehensive integrated model for intraoperative risk assessment of category III (AUS/FLUS) nodules.

Our analysis revealed that for AUS/FLUS nodules, IOFS diagnosed 55.0% patients (292/531) as malignant, with a sensitivity of 0.74 (292/397). Although this sensitivity exceeds that reported in previous studies[10–12], there is still considerable scope for improvement in order to minimize the number of missed diagnoses of malignant cases during surgery. This highlights the need for a more comprehensive approach to intraoperative decision-making in cases of AUS/FLUS. Our findings corroborate those of previous research, which identified clinical features such as age and nodule size, as well as ultrasound characteristics including composition, shape, and margin, as significant predictors of malignancy risk in AUS/FLUS nodules[13–18]. However, few studies have constructed predictive models integrating these features. Choi et al. developed a model based on clinical and ultrasound features, achieving an AUC of 0.83[18], while Öcal et al. reported a similar model with an AUC of 0.78[14]. These findings are consistent with the AUC of our clinical model, which was 0.71 in the training set and 0.77 in the validation set.

Our study builds upon previous models by incorporating IOFS results, thereby markedly enhancing diagnostic accuracy in comparison to both the clinical model and IOFS alone. The integrated model demonstrated superior AUC values (0.92 in the training set and 0.95 in the validation set) and exhibited enhanced calibration and clinical utility. This improvement in diagnostic accuracy may facilitate more appropriate surgical decision-making, which could potentially reduce both under- and over-treatment of thyroid nodules.

The nomogram generated from our integrated model offers a practical tool for clinicians to estimate the likelihood of malignancy intraoperatively. This visual aid may assist in making real-time decisions regarding the extent of thyroid surgery, which could potentially reduce the necessity for total thyroidectomy and improve patient outcomes.

It should be acknowledged that, while our study presents a novel and comprehensive approach, it is not without limitations. Firstly, as a single-center, retrospective study, there is a possibility of selection bias, which may affect the generalizability of the findings. To validate the generalizability of our findings, further multi-center prospective studies are required. Secondly, although the sample size was substantial, it was relatively limited. The integration of larger datasets could potentially enhance the model's performance and reliability. Thirdly, the model has not been externally validated, which is essential for confirming its reliability and generalizability across diverse patient populations and clinical settings. Fourthly, both the evaluation of ultrasound features and frozen section interpretation were necessarily subjective, which could have introduced observer bias. However, efforts were made to minimize these potential sources of bias through centralized pathology services, standardized WHO diagnostic criteria for frozen sections, and blinded scoring with senior review for ultrasound

Characteristic	Overall N = 531	Validation set N = 159	Training set N = 372	*p*-value
IOFS result, No. (%)				0.10
Benign	110 (20.7%)	33 (20.8%)	77 (20.7%)	
Indeterminate[a]	129 (24.3%)	48 (30.2%)	81 (21.8%)	
Malignancy	292 (55.0%)	78 (49.1%)	214 (57.5%)	
PS report, No. (%)				0.68
Benign	134 (25.2%)	42 (26.4%)	92 (24.7%)	
Malignant	397 (74.8%)	117 (73.6%)	280 (75.3%)	
Age, mean(SD), years	48.2 ± 10.9	48.6 ± 11.8	48.0 ± 10.6	0.54
Sex, No. (%)				0.79
Female	427 (80.4%)	129 (81.1%)	298 (80.1%)	
Male	104 (19.6%)	30 (18.9%)	74 (19.9%)	
Hashimoto's thyroiditis[b], No. (%)				0.57
Absent	375 (70.6%)	115 (72.3%)	260 (69.9%)	
Present	156 (29.4%)	44 (27.7%)	112 (30.1%)	
Size, No. (%)				0.73
<1cm	412 (77.6%)	120 (75.5%)	292 (78.5%)	
1-2cm	96 (18.1%)	31 (19.5%)	65 (17.5%)	
>2cm	23 (4.33%)	8 (5.03%)	15 (4.03%)	
Composition, No. (%)				0.51
Mixed cystic and solid	25 (4.71%)	6 (3.77%)	19 (5.11%)	
Solid or almost completely solid	506 (95.3%)	153 (96.2%)	353 (94.9%)	
Echogenicity, No. (%)				0.87
Hypoechoic	520 (97.9%)	157 (98.7%)	363 (97.6%)	
Isoechoic	7 (1.32%)	1 (0.63%)	6 (1.61%)	
Very hypoechoic	4 (0.75%)	1 (0.63%)	3 (0.81%)	
Shape, No. (%)				0.84
Taller-than-wide	237 (44.6%)	72 (45.3%)	165 (44.4%)	
Wider-than-tall	294 (55.4%)	87 (54.7%)	207 (55.6%)	
Margin, No. (%)				0.44
Ill-defined	132 (24.9%)	35 (22.0%)	97 (26.1%)	
Lobulated or irregular	256 (48.2%)	76 (47.8%)	180 (48.4%)	
Smooth	143 (26.9%)	48 (30.2%)	95 (25.5%)	
Large comet-tail artifacts, No. (%)				0.78
Absent	516 (97.2%)	154 (96.9%)	362 (97.3%)	
Present	15 (2.82%)	5 (3.14%)	10 (2.69%)	
Macrocalcifications				0.70
Absent	488 (91.9%)	145 (91.2%)	343 (92.2%)	
Present	43 (8.10%)	14 (8.81%)	29 (7.80%)	
Peripheral calcifications, No. (%)				0.25
Absent	516 (97.2%)	157 (98.7%)	359 (96.5%)	
Present	15 (2.82%)	2 (1.26%)	13 (3.49%)	
Punctate echogenic foci, No. (%)				0.67
Absent	328 (61.8%)	96 (60.4%)	232 (62.4%)	
Present	203 (38.2%)	63 (39.6%)	140 (37.6%)	

Table 2. Baseline Characteristics and Comparison Between Training and Validation Sets for AUS/FLUS Thyroid Nodules Some percentages may not add up to 100% due to rounding. [a] In cases IOFS reported as "suspicious but inconclusive," "deferred diagnosis," "further evaluation required," or "definitive resection recommended." [b] Hashimoto's thyroiditis was defined as a positive thyroglobulin antibody(>4.5 IU/mL) or a positive thyroid peroxidase antibody(>60 U/mL) result via immunochemiluminescent assay

features. Future studies may benefit from the implementation of standardized, quantitative assessment methods, multi-observer validation approaches, or the incorporation of artificial intelligence-based image analysis.

Furthermore, while our model enhances intraoperative malignancy detection, a significant limitation is that it does not directly guide surgical extent decisions. The current model focuses primarily on the binary classification of malignancy status, which, while valuable, represents only one factor in the complex decision-making process regarding surgical approach. Future research should aim to develop more sophisticated predictive tools that

Characteristic	Summary			Univariable analysis			Multivariable analysis		
	Overall N = 531	Benign N = 134	Malignant N = 397	OR	95% CI	*p*-value	OR	95% CI	*p*-value
Age, mean(SD), years	48.20 ± 10.92	51.13 ± 10.85	47.21 ± 10.78	0.97	0.95, 0.98	<**0.001*****	0.99	0.96, 1.02	0.42
Sex, No. (%)									
Female	427 (80.41%)	106 (79.10%)	321 (80.86%)	1.00	-				
Male	104 (19.59%)	28 (20.90%)	76 (19.14%)	0.90	0.56, 1.47	0.66			
Hashimoto's thyroiditis, No. (%)									
Absent[a]	375 (70.62%)	97 (72.39%)	278 (70.03%)	1.00	-				
Present	156 (29.38%)	37 (27.61%)	119 (29.97%)	1.12	0.73, 1.75	0.60			
Size, No. (%)									
<1cm	412 (77.59%)	86 (64.18%)	326 (82.12%)	1.00	-		1.00	-	
1-2cm	96 (18.08%)	34 (25.37%)	62 (15.62%)	0.48	0.30, 0.78	**0.003****	1.32	0.65, 2.72	0.44
>2cm	23 (4.33%)	14 (10.45%)	9 (2.27%)	0.17	0.07, 0.40	<**0.001*****	0.82	0.23, 2.74	0.75
Composition, No. (%)									
Mixed cystic and solid	25 (4.71%)	13 (9.70%)	12 (3.02%)	1.00	-		1.00	-	
Solid or almost completely solid	506 (95.29%)	121 (90.30%)	385 (96.98%)	3.45	1.52, 7.86	**0.003****	1.22	0.33, 4.94	0.77
Echogenicity, No. (%)									
Hypoechoic	520 (97.93%)	128 (95.52%)	392 (98.74%)	1.00	-				
Isoechoic	7 (1.32%)	3 (2.24%)	4 (1.01%)	0.44	0.09, 2.23	0.28			
Very hypoechoic	4 (0.75%)	3 (2.24%)	1 (0.25%)	0.11	0.01, 0.86	0.056			
Shape, No. (%)									
Taller than wide	237 (44.63%)	33 (24.63%)	204 (51.39%)	1.00	-		1.00	-	
Wider than tall	294 (55.37%)	101 (75.37%)	193 (48.61%)	0.31	0.20, 0.48	<**0.001*****	0.25	0.13, 0.47	<**0.001*****
Margin, No. (%)									
Ill-defined	132 (24.86%)	26 (19.40%)	106 (26.70%)	1.00	-		1.00	-	
Lobulated or irregular	256 (48.21%)	48 (35.82%)	208 (52.39%)	1.06	0.62, 1.80	0.82	1.34	0.64, 2.81	0.44
Smooth	143 (26.93%)	60 (44.78%)	83 (20.91%)	0.34	0.19, 0.58	<**0.001*****	1.01	0.46, 2.27	0.97
Large comet-tail artifacts, No. (%)									
Absent	516 (97.18%)	129 (96.27%)	387 (97.48%)	1.00	-				
Present	15 (2.82%)	5 (3.73%)	10 (2.52%)	0.67	0.23, 2.17	0.47			
Macrocalcifications, No. (%)									
Absent	488 (91.90%)	118 (88.06%)	370 (93.20%)	1.00	-				
Present	43 (8.10%)	16 (11.94%)	27 (6.80%)	0.54	0.28, 1.05	0.063			
Peripheral calcifications, No. (%)									
Absent	516 (97.18%)	129 (96.27%)	387 (97.48%)	1.00	-				
Present	15 (2.82%)	5 (3.73%)	10 (2.52%)	0.67	0.23, 2.17	0.47			
Punctate echogenic foci, No. (%)									
Absent	328 (61.77%)	84 (62.69%)	244 (61.46%)	1.00	-				
Present	203 (38.23%)	50 (37.31%)	153 (38.54%)	1.05	0.71, 1.58	0.80			
IOFS result, No. (%)									
Benign	110 (20.72%)	81 (60.45%)	29 (7.31%)	1.00	-		1.00	-	
Indeterminate[b]	129 (24.29%)	50 (37.31%)	79 (19.90%)	4.41	2.56, 7.76	<**0.001*****	3.88	2.16, 7.13	<**0.001*****
Malignancy	292 (54.99%)	3 (2.24%)	289 (72.80%)	269	93.2, 1146	<**0.001*****	265	887.9, 1157	<**0.001*****

Table 3. Univariate and Multivariate Logistic Regression Analysis for Predicting Malignancy in AUS/FLUS Thyroid Nodules in the Entire Cohort Multivariable analysis only included variables with $p < 0.05$ in the univariate analysis. The reference category for each categorical variable is indicated by "1.00" in the OR column and "-" in the 95% CI column. Some percentages may not add up to 100% due to rounding. [a] Hashimoto's thyroiditis was defined as a positive thyroglobulin antibody (>4.5 IU/mL) or a positive thyroid peroxidase antibody (>60 U/mL) result via immunochemiluminescent assay. [b] In cases IOFS reported as "suspicious but inconclusive," "deferred diagnosis," "further evaluation required," or "definitive resection recommended." ***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$

can directly inform surgical extent decisions by incorporating additional critical parameters that influence the choice between conservative and extensive surgery, including but not limited to tumor size, extrathyroidal extension, lymph node status, patient age, and comorbidities. These advanced models could potentially provide more specific recommendations regarding the appropriateness of procedures such as total thyroidectomy versus lobectomy, or the necessity of central compartment lymph node dissection.

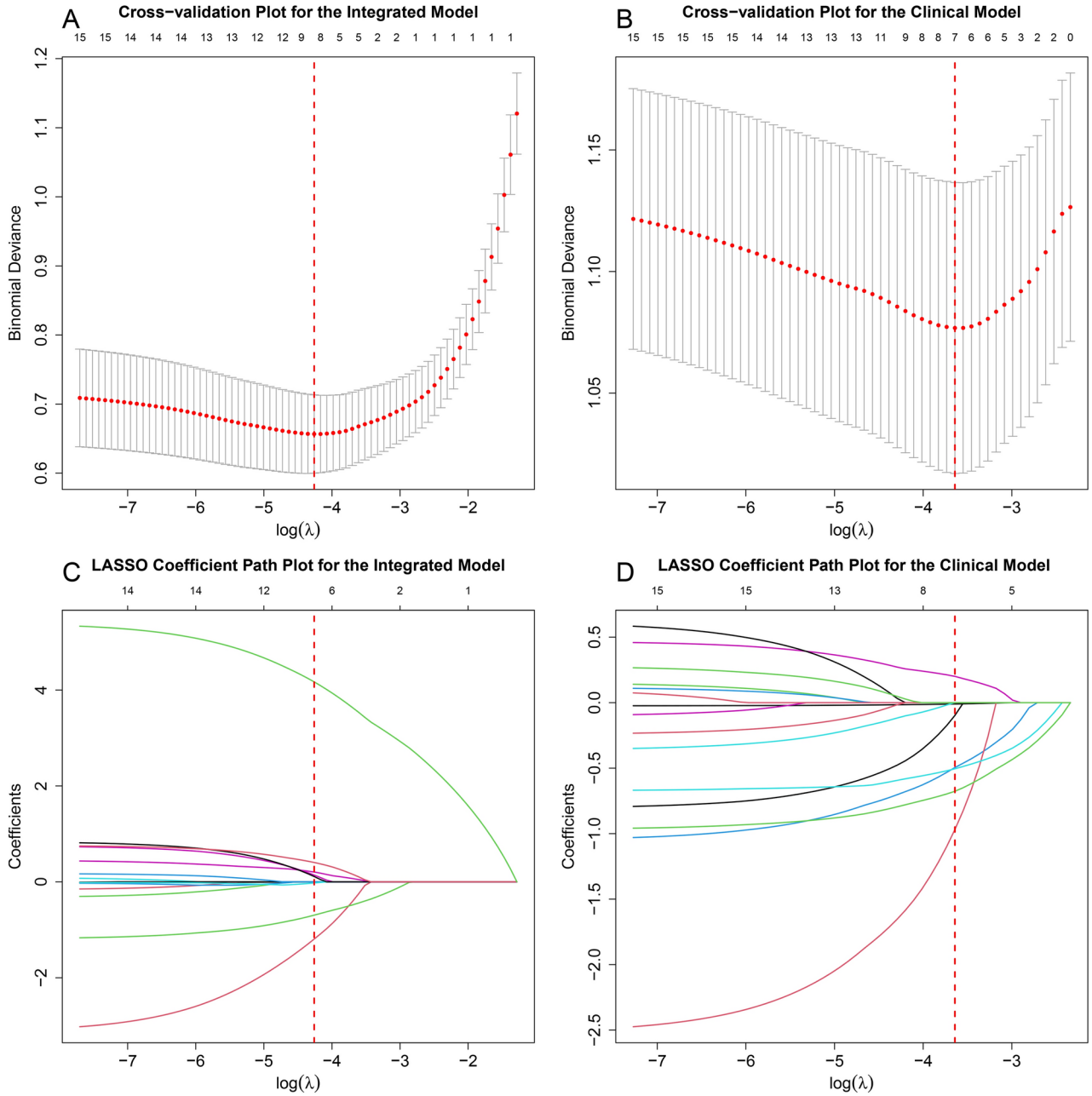

Fig. 3. LASSO Regression Analysis for the Selection of Variables for Models in the Prediction of Malignancy in AUS/FLUS Thyroid Nodules. (**A**) and (**B**) depict the cross-validation process of the LASSO regression for the integrated and clinical models, respectively. These plots show the change in binomial deviance (a measure of model fit) across different λ values. The optimal λ value, determined at the minimum binomial deviance, is marked by the vertical red dashed line in each plot. (**C**) and (**D**) illustrate the LASSO coefficient path plots for the integrated and clinical models, respectively. These plots show how the coefficients of various predictors change as the penalty parameter λ increases (moving from right to left on the x-axis). The numbers at the top indicate the count of non-zero coefficients at each λ value.

Additionally, while our deliberate focus on $BRAF^{V600E}$ negative nodules allowed us to address the most challenging diagnostic scenarios, this approach necessarily excludes a subset of thyroid nodules from our analysis. It should be noted that not all patients underwent preoperative molecular testing, which may introduce a potential selection bias. As our understanding of molecular markers in thyroid cancer continues to evolve, future research should not only examine the model's predictive capacity in nodules tested for $BRAF^{V600E}$ but also explore the potential value of incorporating additional molecular markers that may influence malignancy risk. Such investigations could potentially expand the model's applicability and further enhance its diagnostic utility across a broader spectrum of thyroid nodules.

Variables	Integrated model			Clinical model			IOFS model		
	Coefficients(B)	SE	95% CI	Coefficients(B)	SE	95% CI	Coefficients(B)	SE	95% CI
Age (years)	-0.01	0.016	-0.04, 0.02	-0.02	0.012	-0.05, 0.00			
Composition									
Mixed cystic and solid	-	-	-	-	-	-			
Solid or almost completely solid	0.49	0.762	-0.94, 2.1	0.34	0.546	-0.75, 1.4			
Echogenicity									
Hypoechoic	-	-	-	-	-	-			
Isoechoic	-0.08	1.19	-2.9, 2.1	-0.89	0.874	-2.7, 0.89			
Very hypoechoic	-3.2	2.54	-7.8, 0.64	-2.5	1.29	-5.7, -0.03			
Shape									
Taller-than-wide	-	-	-	-	-	-			
Wider-than-tall	-1.1	0.379	-1.9, -0.42	-0.93	0.302	-1.5, -0.35			
Margin									
Ill-defined	-	-	-	-	-	-			
Lobulated or irregular	0.19	0.433	-0.66, 1.0	0.11	0.330	-0.55, 0.75			
Smooth	-0.04	0.464	-0.95, 0.88	-0.62	0.355	-1.3, 0.07			
Macrocalcifications									
Absent	-	-	-						
Present	0.78	0.572	-0.34, 1.9						
Peripheral calcifications									
Absent	-	-	-						
Present	0.84	0.703	-0.55, 2.3						
Size									
<1cm				-	-	-			
1-2cm				-0.25	0.332	-0.90, 0.41			
>2cm				-0.95	0.599	-2.2, 0.21			
IOFS result									
Benign	-	-	-				-	-	-
Indeterminate[a]	0.74	0.362	0.03, 1.5				0.86	0.330	0.22, 1.5
Malignancy	5.3	0.774	4.1, 7.2				5.4	0.751	4.1, 7.2

Table 4. Coefficients of Variables in The Models for Predicting Malignancy in AUS/FLUS Thyroid Nodules The integrated model combined features from clinical, ultrasound, and IOFS results. The clinical model combined features from clinical and ultrasound results. The IOFS model included only IOFS results. Coefficients (B) represent the change in log odds of malignancy for a one-unit increase in the predictor variable. "–" indicates the reference category for each categorical variable. Empty cells indicate that the variable was not included in the model. [a] In cases IOFS reported as "suspicious but inconclusive," "deferred diagnosis," "further evaluation required," or "definitive resection recommended."

It is also important to recognize that our model, which relies in part on IOFS results, is specifically designed for intraoperative decision-making in cases where surgery has already been determined to be the appropriate intervention. In the context of evolving clinical practice guidelines that increasingly adopt de-escalation strategies for the management of differentiated thyroid cancer, including active surveillance protocols, our model cannot assist in the critical upstream decision between observation and surgical intervention. This reflects an important limitation of our work, particularly as the field moves toward more conservative management approaches. Future research should focus on developing predictive tools that can help guide initial management decisions, including patient selection for active surveillance programs.

In conclusion, this study developed a novel integrated predictive model for intraoperative malignancy risk assessment of thyroid nodules with AUS/FLUS cytology and negative *BRAF*^{V600E} mutation status. The model demonstrated superior diagnostic capability, predictive performance, calibration, and clinical utility compared to conventional approaches. It provides a more reliable decision support tool for surgeons to optimize surgical strategies and potentially improve patient outcomes in AUS/FLUS thyroid nodule management. Future research should focus on external validation of this model and exploration of its impact on clinical decision-making and long-term patient outcomes.

Models	R^2	AIC	BIC	RMSE	MAE
Training set					
Integrated model	0.48	241.03	288.06	0.31	0.19
Clinical model	0.10	393.41	432.60	0.41	0.33
IOFS model	0.44	237.72	249.48	0.32	0.21
Validation set					
Integrated model	NA	NA	NA	0.29	0.20
Clinical model	NA	NA	NA	0.40	0.33
IOFS model	NA	NA	NA	0.33	0.23

Table 5. Model Fit and Predictive Performance Metrics in The Models for Predicting Malignancy in AUS/FLUS Thyroid Nodules R^2 is a goodness-of-fit measure for logistic regression models; higher values indicate better fit. AIC and BIC are model selection tools that balance fit and complexity; lower values are preferred. R^2, AIC, and BIC are not applicable for the validation set. RMSE and MAE measure prediction errors; lower values indicate better predictive accuracy. The integrated model combined features from clinical, ultrasound, and IOFS results. The clinical model combined features from clinical and ultrasound results. The IOFS model included only IOFS results.

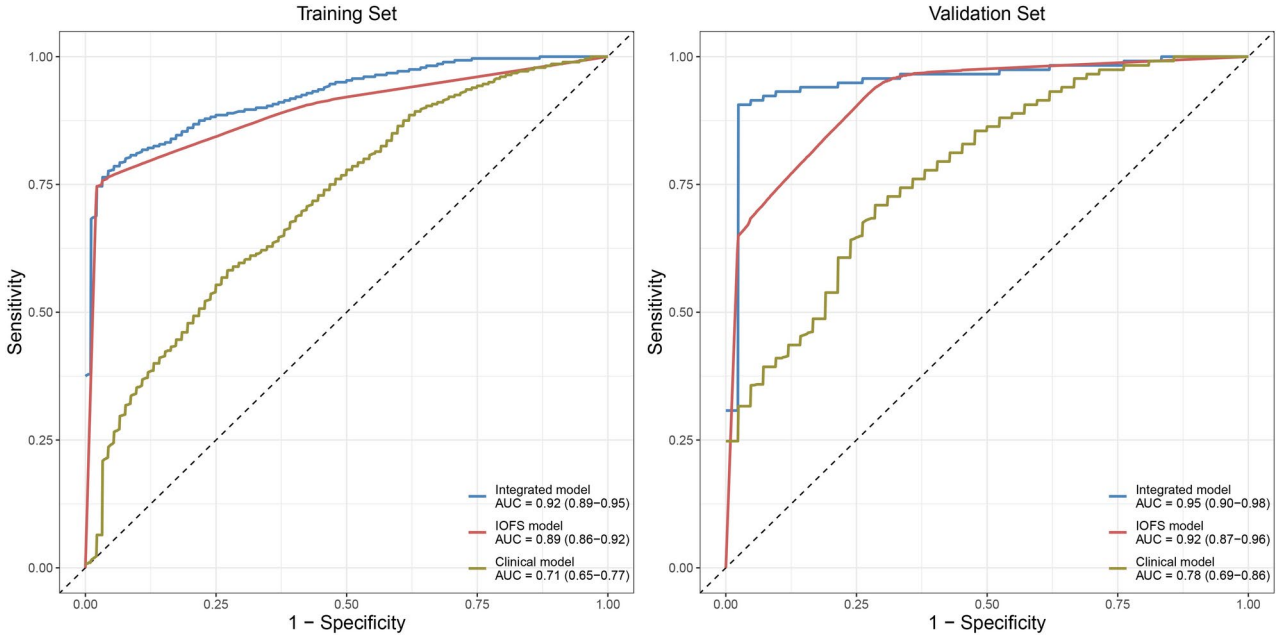

Fig. 4. ROC Curves of the Models for Predicting Malignancy in AUS/FLUS Thyroid Nodules in the Training and Validation Sets. The integrated model combined features from clinical, ultrasound, and IOFS results. The clinical model combined features from clinical and ultrasound results. The IOFS model included only IOFS results. The AUC is provided with 95% confidence intervals in parentheses.

Models	Sensitivity	Specificity	PPV	NPV	F1 Score
Training Set					
Integrated model	0.81	0.90	0.96	0.62	0.88
Clinical model	0.60	0.72	0.87	0.37	0.71
IOFS model	0.76	0.98	0.99	0.57	0.86
Validation Set					
Integrated model	0.90	0.98	0.99	0.77	0.94
Clinical model	0.63	0.76	0.88	0.43	0.74
IOFS model	0.66	0.98	0.99	0.51	0.79

Table 6. Diagnostic Performance Metrics of Models for Predicting Malignancy in AUS/FLUS Thyroid Nodules The integrated model combined features from clinical, ultrasound, and IOFS results. The clinical model combined features from clinical and ultrasound results. The IOFS model included only IOFS results. The thresholds for the integrated, clinical, and IOFS models are 0.65, 0.77, and 0.76, respectively, determined by the point on the Receiver Operating Characteristic curve closest to the top-left corner in the training set for each model

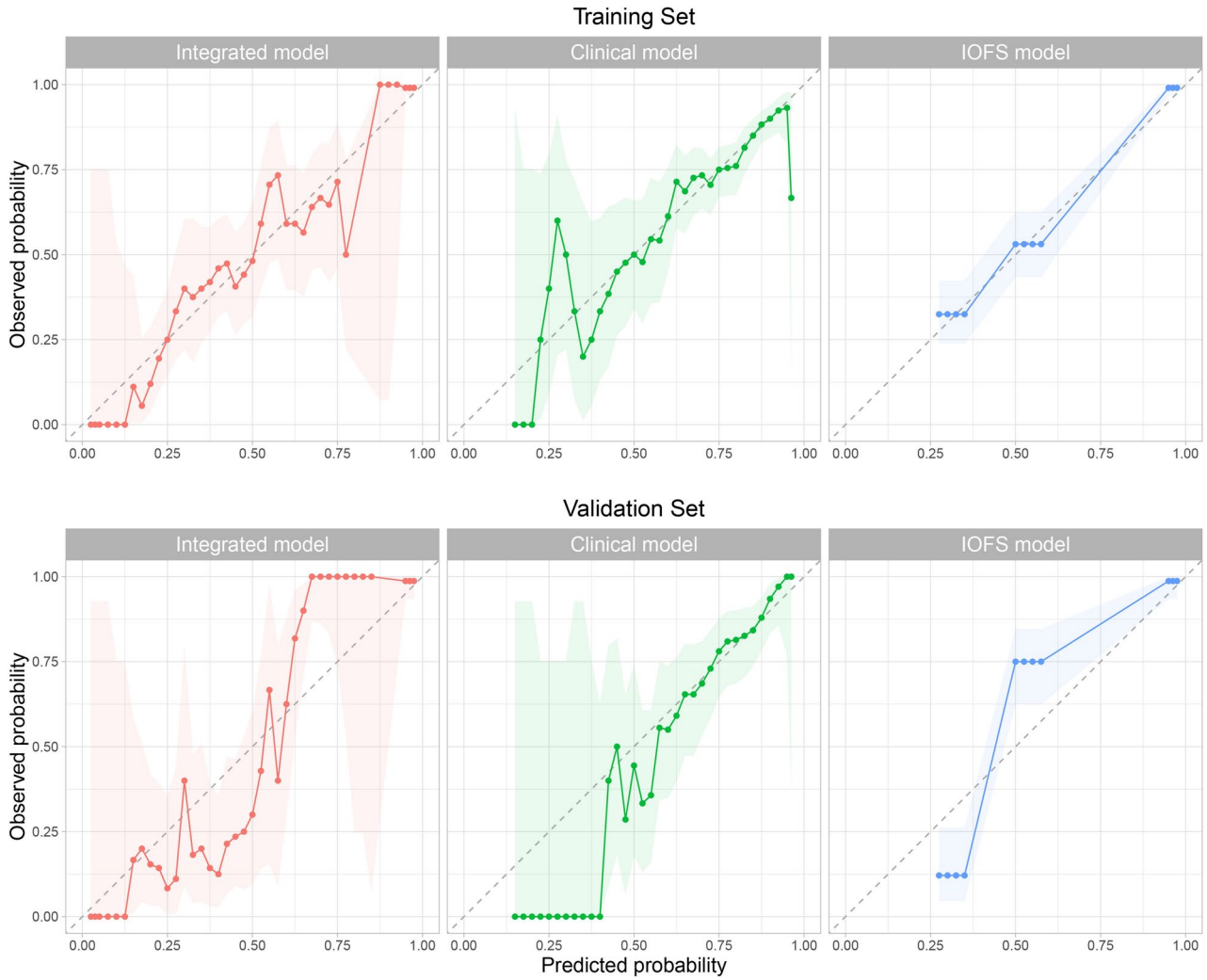

Fig. 5. Calibration Curves of the Three Models for Predicting Malignancy in AUS/FLUS Thyroid Nodules in the Training and Validation Sets. The integrated model combined features from clinical, ultrasound, and IOFS results. The clinical model combined features from clinical and ultrasound results. The IOFS model included only IOFS results. The diagonal dashed line represents a perfect calibration curve, where the predicted probabilities are identical to the observed probabilities. The shaded areas represent the 95% confidence intervals of the calibration curves.

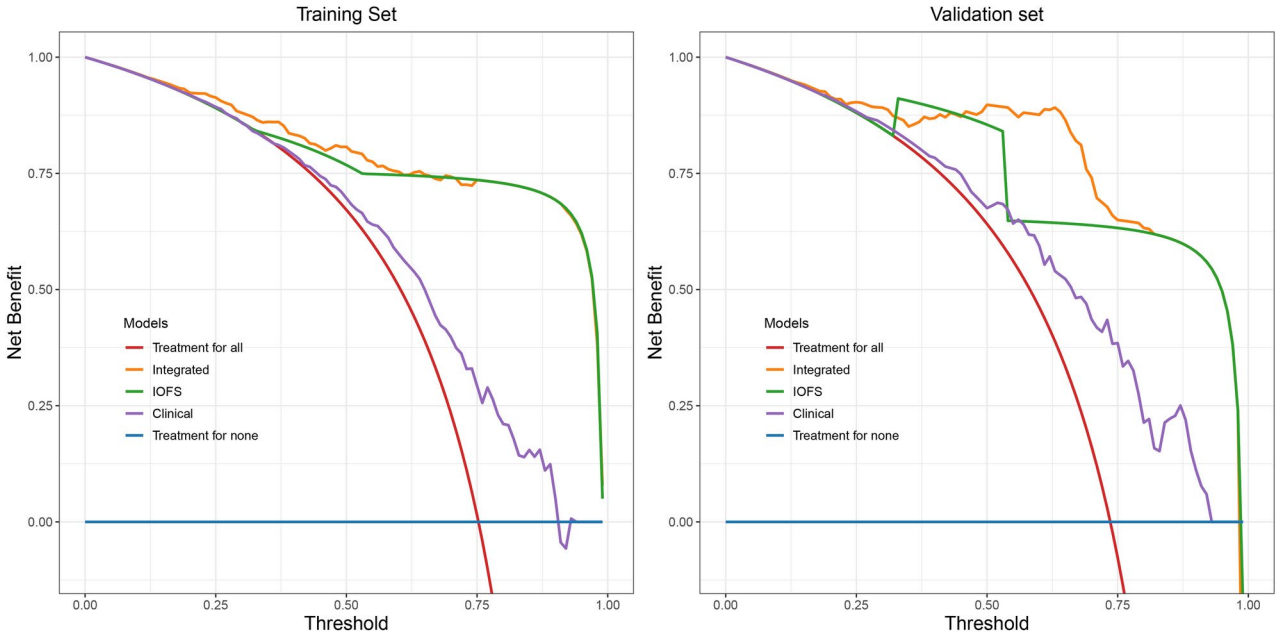

Fig. 6. Decision Curve Analysis of the Models for Predicting Malignancy in AUS/FLUS Thyroid Nodules. Decision Curve Analysis (DCA) illustrates the net benefit across different threshold probabilities. The net benefit is calculated by subtracting the proportion of false positives, weighted by the odds at each threshold probability, from the proportion of true positives. This weighting reflects the trade-off between the benefit of correctly identifying true positives and the harm of incorrectly classifying false positives. The "treat for all" curve assumes that all patients are treated as malignant, while the "treat for none" curve assumes that no patients are treated as malignant. These two curves represent the extreme scenarios of overtreatment and undertreatment, respectively. The integrated model combined features from clinical, ultrasound, and IOFS results. The clinical model combined features from clinical and ultrasound results. The IOFS model included IOFS results only.

	Δ AUC			NRI			IDI		
Models	**Mean**	**95%CI**	***p*-value**	**Mean**	**95%CI**	***p*-value**	**Mean**	**95%CI**	***p*-value**
Training set									
Integrated model vs clinical model	0.21	0.16, 0.26	<0.001	1.44	1.31, 1.57	<0.001	0.36	0.32, 0.40	<0.001
Integrated model vs IOFS model	0.03	0.01, 0.05	0.006	0.43	0.20, 0.65	0.0001	0.05	0.02, 0.07	0.002
Validation set									
Integrated model vs clinical model	0.17	0.09, 0.26	<0.001	1.19	0.96, 1.42	<0.001	0.35	0.29, 0.42	<0.001
Integrated model vs IOFS model	0.03	0.001, 0.06	0.042	0.40	0.06, 0.74	0.02	0.08	0.03, 0.12	<0.001

Table 7. Predictive Improvement Between Models for Predicting Malignancy in AUS/FLUS Thyroid Nodules The integrated model combined features from clinical, ultrasound, and IOFS results. The clinical model combined features from clinical and ultrasound results. The IOFS model included only IOFS results. Δ AUC represents the difference in AUC between the two models, which measures the improvement in predictive accuracy. The *p* value for Δ AUC is calculated by the DeLong test. The NRI quantifies the improvement in reclassification between models, using a continuous scale in this study. The IDI measures the improvement in discrimination between models

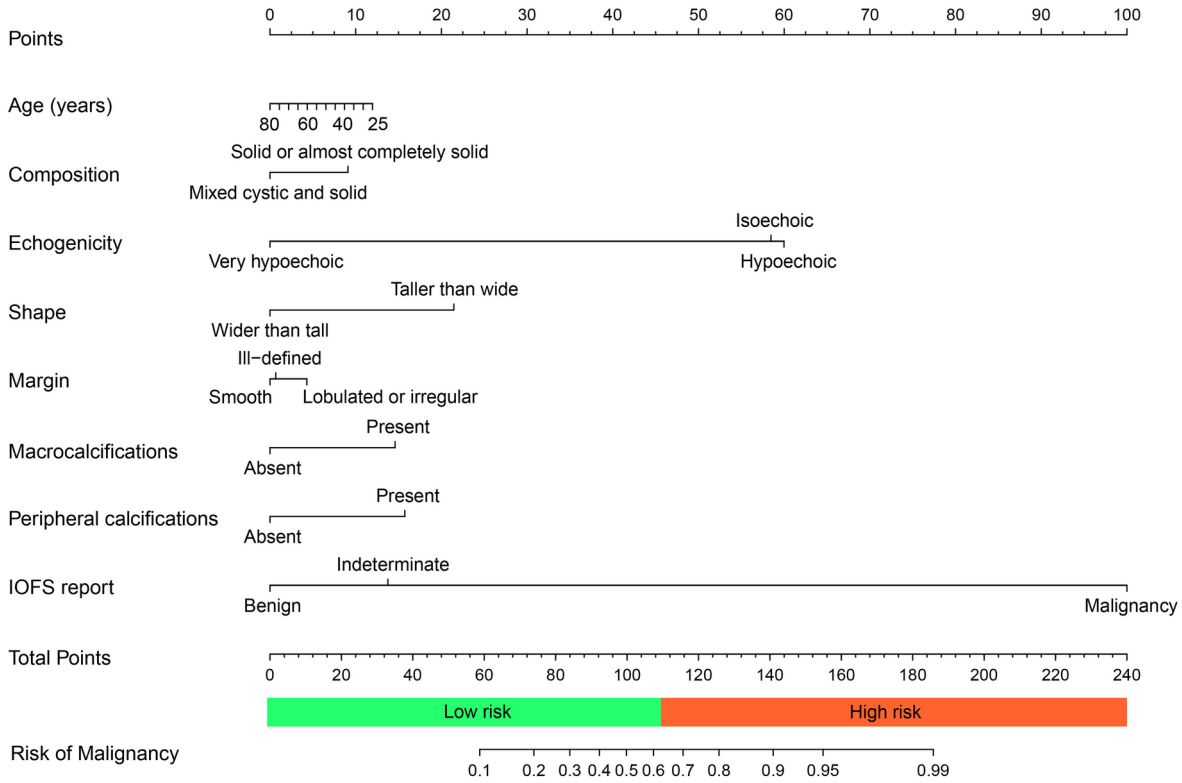

Fig. 7. Nomogram of the integrated Model for Predicting Malignancy Risk in AUS/FLUS Thyroid Nodules. This nomogram of the integrated model presents a visual tool for predicting the risk of malignancy in thyroid nodules classified as AUS/FLUS. It incorporates clinical, ultrasound, and IOFS features to estimate the probability of malignancy. To use the nomogram, locate the patient's value on each variable axis and draw a line straight up to the "points" axis to determine the score for each variable. Sum the scores for all variables and locate the sum on the "total points" axis. Draw a line straight down to the "risk of malignancy" axis to determine the patient's probability of malignancy. Nodules with a total point score of 110 or higher, corresponding to a risk of malignancy of 0.65 or higher, are considered having a high risk of malignancy.

Data availability

The datasets generated and/or analysed during the current study are not publicly available due to restrictions related to patient privacy and confidentiality as per our institutional ethics review board requirements and the sensitive nature of the clinical data involved. However, de-identified data are available from the corresponding author on reasonable request.

Received: 26 October 2024; Accepted: 26 December 2024
Published online: 13 January 2025

References

1. Siegel, R. L., Miller, K. D., Wagle, N. S. & Jemal, A. Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians* **73**, 17–48. https://doi.org/10.3322/caac.21763 (2023).
2. Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians* **71**, 209–249. https://doi.org/10.3322/caac.21660 (2021).
3. Ali, S. Z. et al. The 2023 Bethesda System for Reporting Thyroid Cytopathology. *Thyroid: Official Journal of the American Thyroid Association* **33**, 1039–1044. https://doi.org/10.1089/thy.2023.0141 (2023).
4. Lan, L. et al. Comparison of Diagnostic Accuracy of Thyroid Cancer With Ultrasound-Guided Fine-Needle Aspiration and Core-Needle Biopsy: A Systematic Review and Meta-Analysis. *Frontiers in Endocrinology* **11**, https://doi.org/10.3389/fendo.2020.00044 (2020).
5. Singh Ospina, N. et al. Diagnostic accuracy of ultrasound-guided fine needle aspiration biopsy for thyroid malignancy: Systematic review and meta-analysis. *Endocrine.* **53**, 651–661. https://doi.org/10.1007/s12020-016-0921-x (2016).
6. Rotondi, M. et al. The diagnostic accuracy of fine-needle aspiration cytology for thyroid nodules is not affected by coexistent chronic autoimmune thyroiditis: Results from a cyto-histological series of patients with indeterminate cytology. *European Journal of Endocrinology*. **185**, 201–208. https://doi.org/10.1530/EJE-21-0094 (2021).
7. Park, V. Y., Kim, E.-K., Kwak, J. Y., Yoon, J. H. & Moon, H. J. Malignancy risk and characteristics of thyroid nodules with two consecutive results of atypia of undetermined significance or follicular lesion of undetermined significance on cytology. *European Radiology*. **25**, 2601–2607. https://doi.org/10.1007/s00330-015-3668-5 (2015).
8. Rosario, P. W. Thyroid Nodules with Atypia or Follicular Lesions of Undetermined Significance (Bethesda Category III): Importance of Ultrasonography and Cytological Subcategory. *Thyroid*. **24**, 1115–1120. https://doi.org/10.1089/thy.2013.0650 (2014).
9. Ho, A. S. et al. Malignancy Rate in Thyroid Nodules Classified as Bethesda Category III (AUS/FLUS). *Thyroid*. **24**, 832–839. https://doi.org/10.1089/thy.2013.0317 (2014).

10. Roychoudhury, S. et al. Utility of intraoperative frozen sections for thyroid nodules with prior fine needle aspiration cytology diagnosis. *Diagnostic Cytopathology*. **45**, 789–794. https://doi.org/10.1002/dc.23765 (2017).
11. Trosman, S. J., Bhargavan, R., Prendes, B. L., Burkey, B. B. & Scharpf, J. The contemporary utility of intraoperative frozen sections in thyroid surgery. *American Journal of Otolaryngology*. **38**, 614–617. https://doi.org/10.1016/j.amjoto.2017.07.003 (2017).
12. Huang, J. et al. Intraoperative frozen section can be reduced in thyroid nodules classified as Bethesda categories V and VI. *Scientific Reports*. **7**, 5244. https://doi.org/10.1038/s41598-017-05459-x (2017).
13. Valderrabano, P. & Mciver, B. Evaluation and management of indeterminate thyroid nodules. *Cancer Control : Journal of the Moffitt Cancer Center*. **24**, https://doi.org/10.1177/1073274817729231(2017).
14. Öcal, B. et al. The Malignancy Risk Assessment of Cytologically Indeterminate Thyroid Nodules Improves Markedly by Using a Predictive Model. *European Thyroid Journal*. **8**, 83–89. https://doi.org/10.1159/000494720 (2019).
15. Scerrino, G. et al. Improving diagnostic performance for thyroid nodules classified as bethesda category III or IV: How and by whom ultrasonography should be performed. *The Journal of surgical research*. **262**, 203–211. https://doi.org/10.1016/j.jss.2020.12.009 (2021).
16. Yoon, J. H., Kwon, H., Kim, E.-K., Moon, H. & Kwak, J. Subcategorization of atypia of undetermined significance/follicular lesion of undetermined significance (AUS/FLUS): A study applying Thyroid Imaging Reporting and Data System (TIRADS). *Clinical Endocrinology*. **85**, https://doi.org/10.1111/cen.12987(2016).
17. Yousefi, E., Sura, G. H. & Somma, J. The gray zone of thyroid nodules: Using a nomogram to provide malignancy risk assessment and guide patient management. *Cancer Medicine*. **10**, 2723–2731. https://doi.org/10.1002/cam4.3866 (2021).
18. Choi, Y. et al. Web-based thyroid imaging reporting and data system: Malignancy risk of atypia of undetermined significance or follicular lesion of undetermined significance thyroid nodules calculated by a combination of ultrasonography features and biopsy results. *Head & Neck*. **40**, 1917–1925. https://doi.org/10.1002/hed.25173 (2018).
19. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. M. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD). *Circulation*. **131**, https://doi.org/10.1161/circulationaha.114.014508(2015).
20. Haugen, B. R. et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid: Official Journal of the American Thyroid Association* **26**, 1–133, https://doi.org/10.1089/thy.2015.0020(2016).
21. Tessler, F. N. et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. *Journal of the American College of Radiology*. **14**, 587–595. https://doi.org/10.1016/j.jacr.2017.01.046 (2017).
22. WHO Classification of Tumours Editorial Board. *WHO Classification of Tumours: Endocrine and Neuroendocrine Tumours*, vol. 8 (International Agency for Research on Cancer, Lyon, France, 2022), 5 edn.
23. Zatelli, M. C. et al. Braf v600e mutation analysis increases diagnostic accuracy for papillary thyroid carcinoma in fine-needle aspiration biopsies. *European journal of endocrinology*. **161**, 467–473 (2009).
24. Clinkscales, W. B., Ong, A. A., Nguyen, S., Harruff, E. & Gillespie, M. B. Diagnostic value of ras mutations in indeterminate thyroid nodules: Systematic review and meta-analysis. *Otolaryngology-Head and Neck Surgery*. **156**, 472–479. https://doi.org/10.1177/0194599816685697 (2017).
25. Liu, R. & Xing, M. Diagnostic and prognostic tert promoter mutations in thyroid fine-needle aspiration biopsy. *Endocrine-related Cancer*. **21**, 825–830. https://doi.org/10.1530/ERC-14-0359 (2014).
26. Turkyilmaz, S. et al. Thyroid nodules classified as atypia or follicular lesions of undetermined significance deserve further research: Analysis of 305 surgically confirmed nodules. *Cytopathology: official journal of the British Society for Clinical Cytology*. **28**, 391–399. https://doi.org/10.1111/cyt.12438 (2017).
27. Pikis, G. et al. Prevalence of thyroid carcinoma in nodules with thy 3 cytology: The role of preoperative ultrasonography and strain elastography. *Thyroid Research*. **14**, https://doi.org/10.1186/s13044-021-00098-x(2021).
28. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*. **58**, https://doi.org/10.1111/j.2517-6161.1996.tb02080.x(1996).
29. Scherr, S. & Zhou, J. Automatically Identifying Relevant Variables for Linear Regression with the Lasso Method: A Methodological Primer for its Application with R and a Performance Contrast Simulation with Alternative Selection Strategies. *Communication Methods and Measures*. **14**, 204–211. https://doi.org/10.1080/19312458.2019.1677882 (2020).
30. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria (2024).
31. Sjoberg, D. D., Whiting, K., Curry, M., Lavery, J. A. & Larmarange, J. Reproducible summary tables with the gtsummary package. *The R Journal*. **13**, 570–580. https://doi.org/10.32614/RJ-2021-053 (2021).
32. Allaire, J., Gandrud, C., Russell, K. & Yetman, CJ. *networkD3: D3 JavaScript Network Graphs from R* (2017).
33. Tay, J. K., Narasimhan, B. & Hastie, T. Elastic net regularization paths for all generalized linear models. *Journal of Statistical Software*. **106**, 1–31. https://doi.org/10.18637/jss.v106.i01 (2023).
34. Harrell Jr, F. E. *Rms: Regression Modeling Strategies* (2023).
35. Hamner, B. & Frasco, M. Metrics: Evaluation Metrics for Machine Learning, (2018).
36. Robin, X. et al. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. **12**, 77 (2011).
37. Kuhn, M. & Vaughan, D. & Ruiz, E. Tools for Post-Processing Predicted Values, Probably, (2024).
38. Kundu, S., Aulchenko, Y. S. & Janssens, A. C. J. *PredictABEL: Assessment of Risk Prediction Models* (2020).
39. Brown, M. *Rmda: Risk Model Decision Analysis* (2018).
40. Li, C. et al. Development and validation of a predictive model for assessing the risk of follicular carcinoma in thyroid nodules identified as suspicious by intraoperative frozen section. *Frontiers in Endocrinology*. **15**, https://doi.org/10.3389/fendo.2024.1431247(2024).
41. Medas, F. et al. Prophylactic central lymph node dissection improves disease-free survival in patients with intermediate and high risk differentiated thyroid carcinoma: A retrospective analysis on 399 patients. *Cancers*. **12**, https://doi.org/10.3390/cancers12061658(2020).

Author contributions

C.Li performed data collection, statistical analysis, and drafting of the manuscript; Y.Luo and Q.Li performed the study design; Y.Jiang reviewed and classified the ultrasound features of the masses; all authors read and approved the final manuscript.

Funding

This research is funded by the Medical and Health Research Project of Zhejiang Province (2024KY1484).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.