Peer Review File

Article information: https://dx.doi.org/10.21037/gs-24-237

Response to Reviewer A

Comment 1: Overall, this article represents a valuable addition to the literature.
Reply 1: We greatly appreciate the time and effort you dedicated to reviewing our submission. Your support and encouragement inspire us to continue our research. Thank you once again for your kindness and consideration.

Response to Reviewer B

Comment 1: Since this is a retrospective study, the authors should explain how this model can be applied in common practice.
Reply 1: Thank you for raising this important point regarding the practical application of our model in clinical practice. We appreciate the opportunity to clarify how our retrospective study can inform and facilitate the use of the predictive model in common practice. Below, we provide a detailed explanation of how the model can be applied in common practice.
a. Preoperative Risk Prediction: The primary purpose of our model is to aid clinicians in preoperative risk prediction for patients with thyroid papillary carcinoma (PTC). By incorporating patient-specific preoperative ultrasound data and thyroperoxidase antibody (TPOAb) data into the model, clinicians can derive a personalized prediction of lymph node metastasis (LNM) risk. This information can inform decisions regarding the extent of lymph node dissection (e.g., whether to perform central neck lymph node dissection), the necessity for advanced imaging techniques (e.g., CT or MRI) to further evaluate suspicious lymph nodes, and the timing and intensity of follow-up for high-risk patients.
b. Integration into Clinical Workflow: To ensure seamless integration into clinical practice, the model can be developed into an online application to facilitate shared decision-making between clinicians and patients, providing recommendations based on the predicted risk of LNM.
c. Potential Impact on Patient Outcomes: Through accurate prediction of LNM risk, the model can contribute to: reducing unnecessary lymph node dissections in low-risk patients to minimize surgical complications such as hypoparathyroidism and recurrent laryngeal nerve injury; identifying high-risk patients who may benefit from more aggressive treatment or closer surveillance, thereby potentially enhancing long-term outcomes; and optimizing resource allocation by customizing diagnostic and therapeutic interventions according to individual patient requirements
Changes in the text:
We have added the following to the **Results**:
To delve deeper into the proportion of each risk factor for CLNM or HVCLNM, we visualize each factor as a line and create the corresponding nomogram. *Figure 3* showcases two new

nomograms, with each variable represented as a point ranging from 0 to 100 based on its regression coefficient for CLNM or HVCLNM. We draw a straight line for each variable according to its appropriate score, and the resulting scores are added up and placed on the total score axis to represent the variable value. The nomograms demonstrate that the maximal tumor size is the primary factor for both CLNM and HVCLNM. In clinical practice, the TPOAb value and ultrasound-related tumor characteristics are obtained preoperatively. Subsequently, the corresponding variable values are inputted into the nomogram model to derive the scores for each variable. The total score is calculated by summing these individual scores, which is then positioned on the total score axis to ascertain the probability of HVCLNM or CLNM occurrence. (see Page 8, line 193-198)

Comment 2: In the manuscript, the authors include training cohort and validation cohort. However, it is unclear how the validation cohort was analyzed. It is recommended to specify this point.

Reply 2: Thanks for the kind and insightful comments and suggestions. After constructing a nomogram using the training cohort in this study, to further evaluate the predictive performance of the nomogram, internal and external validations of the nomogram were carried out. In the internal validation, the bootstrap resampling method was applied. Specifically, N samples (where N is the size of the original dataset) were randomly drawn with replacement from the training dataset, and this process was repeated 1000 times to form the internal validation dataset for the internal validation of the model.

For the external validation cohort, the data collection period of the external validation cohort (from September 2022 to May 2023) was completely independent of that of the training cohort (from January 2020 to August 2022), which avoided the risk of model overfitting caused by time overlap. The inclusion and exclusion criteria were exactly the same as those of the training group (see Figure 1), and it was ensured that the external validation dataset contained all the predictive variables required by the model.

During the validation process, the data of each patient in the internal validation cohort and the external validation dataset were input into the nomogram model to calculate the predicted probability. Then, the corresponding AUC value was plotted to evaluate the discrimination ability of the model, a calibration curve was plotted to show the difference between the predicted values and the actual values of the model, and a DCA curve was plotted to quantify the clinical net benefit of the model at different threshold probabilities. This ultimately demonstrated the evaluation of the predictive effect of the model in practical applications. All processes were implemented using R version 4.3.1.

Changes in the text:

We have added the following to the **Statistical Analysis**:

The performance of the logistic regression model was internally validated using bootstrapping, with temporal validation (external validation) performed using data collected from patients between September 2022 and May 2023. To evaluate the nomogram's predictive ability, a ROC curve was plotted, and the area under the curve area under the curve (AUC)was calculated. A calibration plot was generated to show the difference between the nomogram's predicted outcomes and actual results, demonstrating the model's accuracy. Finally, decision curve analysis (DCA) was conducted to provide a more comprehensive

evaluation of the model. During the validation process, the data of patients in the internal validation cohort and the external validation dataset were input into the nomograms of the training cohort. Predicted probabilities were calculated, and corresponding ROC curves, calibration plots, and DCA curves were then generated to assess the predictive value of the model in practical applications. The model training and validation were both implemented using R version 4.3.1. (see Page 6, line 149-154)

Comment 3: Line 110 "At our institution, nearly all patients with TC routinely undergo CLND". The authors should better explain whether they always routinely perform a CND. Is this practice performed regardless of tumor size? This point should be clarified.
Reply 3: We appreciate the reviewer's insightful comments on the TC surgical approach and the question of whether CLND should be performed. We also recognize that current international guidelines, particularly the 2015 American Thyroid Association Guidelines and the 2024 American Thyroid Association Management Guidelines for Adult Patients with Differentiated Thyroid Cancer - Endorsement Response, recommend that prophylactic central compartment lymph node dissection should not be performed for most small, non - invasive, clinically node - negative (cT1 - T2, cN0) papillary thyroid carcinomas (PTC). However, the Chinese guidelines for the diagnosis and management of thyroid nodules and differentiated thyroid cancer (2nd edition) stipulate that during DTC surgery, ipsilateral central neck compartment lymph node dissection should be routinely performed when parathyroid gland preservation and recurrent laryngeal nerve protection are ensured (Strong Recommendation; High-Quality Evidence). As a result, prophylactic lymph node dissection is standard practice for almost all patients at our institution. Additionally, this study presents a comprehensive dataset on lymph node metastasis in PTC, serving as a key highlight.
Changes in the text:
We have added the following to the **Surgical Strategy**:
According to the guidelines for the diagnosis and management of thyroid nodules and differentiated thyroid cancer in China, nearly all patients with thyroid cancer at our institution routinely undergo CLND. (see Page 5, line 115-116)

Comment 4: Results, Line 147 "through postoperative pathological examination". What is meant by postoperative pathological examination? definitive histological examination?
Please specify.
Reply 4: We thank the reviewer for this detailed and constructive comment. We appreciate the reviewer's attention to detail. The term 'postoperative' refers to pathological examinations conducted after surgical resection, specifically definitive histological analysis of resected specimens. We have revised the text to clarify this terminology.
Changes in the text:
We have revised the text to clarify this terminology**:**
The training cohort of this study comprised 353 patients who were diagnosed with PTC through definitive histological examination. (see Page 7, line 159).