Peer Review File

Article information: https://dx.doi.org/10.21037/gs-2024-508

Reviewer A

Good paper.

Reply: Thank you for your comment, we will continue to maintain our efforts and strive for improvement in our future research. Thank you.

Reviewer B

Comment 1: In the title, I suggest the authors to indicate the development and validation of nomogram.

Reply 1: Thank you for your suggestion. We have revised the article title to : **Development and Validation of a Multidimensional Machine Learning-Based Nomogram for Predicting Central Lymph Node Metastasis in Papillary Thyroid Microcarcinoma (see Page 1, line 1-4)**

Comment 2: In the abstract, the authors did not describe the current knowledge gap on the predictive model for CLNM in the background, did not describe the inclusion of subjects, assessment of multidimentional predictors, and gold diagnosis of CLNM in the methods, did not describe the patient sample characteristics and the specificity and sensitivity of the nomogram in both the training and validation samples in the results, and did not have more detailed comments for the clinical implications of the findings.

Reply 2: Thank you for the suggestion. We've carefully revised each of the recommendations in the Background, Methods, Results section. We mark the change with red.**(see Page 2)**

Background: Papillary thyroid microcarcinoma (PTMC), a subset of papillary thyroid carcinoma, is characterized by tumors ≤10 mm in size. While generally indolent, central lymph node metastasis (CLNM) is associated with higher risks of recurrence and distant metastasis. Existing prediction models for CLNM predominantly depend on isolated clinical or imaging parameters, failing to integrate multidimensional predictors such as clinicopathological, ultrasonographic, and serological features. This limitation significantly undermines their clinical applicability. Therefore, we developed a machine learning-based nomogram that integrates comprehensive predictors to enhance preoperative risk stratification and facilitate personalized surgical decision-

making.

Methods:A retrospective study was conducted on 503 PTMC patients who underwent thyroidectomy in our hospital between 2020 and 2023. Patients were randomly divided into training (n=352) and validation (n=151) cohorts. Inclusion criteria required preoperative imaging to confirm no cervical lymph node metastasis, complete clinicopathologic data, and initial surgery with central lymph node dissection, as well as postoperative pathology confirming papillary thyroid carcinoma. Multidimensional predictors (clinical demographics, ultrasonographic features, serological markers, and histopathological characteristics) were analyzed. Central lymph node metastasis (CLNM) was definitively diagnosed via postoperative histopathology. Least absolute shrinkage and selection operator (LASSO) regression was used to identify key predictors, which were incorporated into a logistic regression model. The model's performance was evaluated using ROC curves, calibration plots, and decision curve analysis (DCA).
Results: Among 503 enrolled patients (mean age: 48.5 years; male: 24%, female: 76%), central lymph node metastasis (CLNM) was pathology confirmed in 28.8% (145/503). Ultimately, age, gender, tumor size, tumor location, and extrathyroidal invasion (ETE) were identified as independent predictors of CLNM. The nomogram achieved an AUC of 0.88 (sensitivity 0.84, specificity 0.76)in the training cohort and 0.78(sensitivity 0.80, specificity 0.70) in the validation cohort. Calibration plots indicated excellent agreement between predicted and observed probabilities, with mean absolute errors below 0.05. DCA demonstrated clinical utility for threshold probabilities ranging from 15% to 88%.These results suggest that the nomogram has good predictive performance and clinical applicability in assessing the risk of CLNM in PTMC patients.

Comment 3: In the introduction, it is necessary to review the factors and biomarkers associated with CLNM and the available predictive models of CLNM based on them, as well as have comments on the limitations of prior studies. The authors emphasized "multidimentional" in the title, so they need to explain why multidimentional predictors are important and helpful for the prediction. In general, it is difficult to collect data on multidimentional predictors. BTW, to achieve the goal of "help patients and doctors make educated surgical decisions", to what extent the accuracy of the predictive model would have. I do not think a 0.88 AUC is acceptable.

Reply 3: Thank you for your valuable comments, I have added factors and biomarkers associated with CLNM in the introduction, as well as predictive models from previous studies, explaining the multidimensionality. The revised text

has been marked in red font in the original manuscript.(see Page 3-4, line 86-107)

Regarding your query about the AUC value of the prediction model (0.88 for the training set), our response is as follows:
Although we recognize that no model can achieve an absolutely perfect level of prediction, the training set AUC of 0.88 and the validation set AUC of 0.78 both indicate that the model has a strong discriminative ability. This level of performance is among the more desirable in similar studies to date. HIn addition to the AUC metrics, we also used decision curve analysis (DCA), which showed that the model can provide net benefits for clinical decision making across the threshold range of 15% to 88%, which further validates the potential value of the model in practical applications. We acknowledge that there is still room for improvement of the model and plan to conduct more external validation and model optimization in future studies with a view to further improving the predictive accuracy and clinical applicability.
We hope the above explanations have answered your questions and thank you again for your careful review of our work and valuable suggestions.

Comment 4: In the methodology, the authors need to report the clinical research design and sample size estimation procedures of this study. Please describe the data collection of clinical variables and how CLNM was diagnosed. In statistics, please report the P value for statistical significance and provide the threshold AUC values, as well as sensitivity and specificity, for a nomogram, that can be used for clinical decision.

Reply 4: Thank you for your valuable comments, we have added to the methodology and results that describe the collectiHon of variables as well as diagnostic criteria for CLNM, P-values, sensitivity and specificity of nomograms. Study design and AUC values have been mentioned in the original text. The revised text has been marked in red font in the original manuscript.(see Page 5, line 134-135; Page 6, line 173; Page 8, line 226)

Study Design and Sample Size Estimation
This study was designed as a retrospective observational cohort study conducted at Liaoyang Central Hospital from January 2020 to December 2023. The study aimed to develop and validate a machine learning-based nomogram for predicting central lymph node metastasis (CLNM) in papillary thyroid microcarcinoma (PTMC) patients. To determine the appropriate sample size, we conducted a priori power analysis based on previous studies assessing risk factors for CLNM in PTMC. Using the formula for logistic regression sample size estimation proposed

by Peduzzi et al. (1), the minimum required sample size was calculated as follows: N=10×k/P
where k represents the number of predictor variables, and P is the expected event rate (i.e., the proportion of patients with CLNM). Based on previous literature and preliminary data, we anticipated an event rate of 28%, and five predictor variables were considered in the final model. Thus, the minimum required sample size was: N=10×5/0.28 ≈ 179.
To ensure sufficient statistical power and improve model generalizability, we included a total of 503 patients, significantly exceeding the minimum requirement. Patients were randomly divided into a training set (n = 352, 70%) and a validation set (n = 151, 30%) to assess model performance. The adequacy of the sample size was further validated by checking the stability of the model performance metrics, including the area under the receiver operating characteristic (ROC) curve (AUC) and calibration plots.
Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996 Dec;49(12):1373-9.

Comment 5: Please consider to cite several related papers: 1. Chen Z, Wang JJ, Du JB, Li JF, Zheng RT, Yuan SM, Wu T, Guo DM, Zhai YX. Development and validation of a dynamic nomogram for predicting central lymph node metastasis in papillary thyroid carcinoma patients based on clinical and ultrasound features. Quant Imaging Med Surg 2025;15(2):1555-1570. doi: 10.21037/qims-24-618. 2. Gao Y, Tian M, Hou X, Hao W, Zhang Y, Hu L, Kim JM, Gao M, Li D. Multifocality increases the risk of central compartment lymph node metastasis but is not related to the risk of recurrence and death in papillary thyroid carcinoma. Gland Surg 2024;13(12):2383-2394. doi: 10.21037/gs-2024-505. 3. Zhan H, Hong Y, Zhang L, Huang K, Zheng M, Zhang F. Impact of location and size of minimal extrathyroidal extension on lymph node metastasis in papillary thyroid cancer: a retrospective analysis. Gland Surg 2024;13(9):1619-1627. doi: 10.21037/gs-24-273

Reply 5: Thank you for the suggested references, all the references you provide have been cited in my manuscript. In the manuscript the sequence of references is 13.18.22 and is marked in red. (see line 370)

Comment 6: There are some minor language problems. For example, in the sentence "The increase in public health awareness has led to a rise in the practice of routine check - ups, facilitating the prompt detection of PTMC in its incipient stages, thereby preventing cases where diagnosis occurs only after the onset of symptoms.", it could be

rephrased more concisely for better readability. Another issue is that some abbreviations like “CNLM” are used before they are fully introduced in the text, which might cause confusion for readers at first. Additionally, in the Results section, when presenting data in tables, some column headers could be more clearly labeled to improve data interpretation.

Reply 6: (1)Thank you for your suggestion, in order to improve readability, I have simplified and rewritten the sentence you mentioned in the introduction.(see Page 3, line 75-76)

(2) Regarding the abbreviation 'CNLM', the issue was due to my misspelling, which has been corrected to 'CLNM' in Table 2. Additionally, I carefully reviewed the abbreviations throughout the text and found that 'A/T' was not expressed in detail, and it has been changed to anteroposterior and transverse diameter ratio (A/T) in the original text. (see Page 5, line 147)

(3)Shorthand for nouns in the table I've added notes below the table. In addition, we have added more details about the captions and acronyms of the Figure

Thank you again for your careful revisions of my article. (see line 463)

Reviewer C

Comment 1: Tables

a) Please provide the unit of age in tables.

b) Please add the description to the table footnote that how the data are presented in table 1.

CLNM				χ^2=0.01	0.91
No	358 (71.17)	250 (71.02)	108 (71.52)		
Yes	145 (28.83)	102 (28.98)	43 (28.48)		
Gender				χ^2=0.20	0.652
Male	120 (23.86)	82 (23.30)	38 (25.17)		
Female	383 (76.14)	270 (76.70)	113 (74.83)		
Age	48.50 ± 10.07	48.39 ± 9.82	48.78 ± 10.66	t=0.40	0.687
Tumor diameter	0.60 (0.50, 0.80)	0.60 (0.50, 0.80)	0.70(0.40,0.80)	Z=-0.23	0.818

Reply 5 :After carefully reading your journal's requirements for the graphs and tables section of the Author Submission Guidelines and listening to your comments, I have made the following changes to the table:

(a) In Tables 1, 2, and 3 we added the unit of age: years and added the unit of tumor diameter: cm

(b)In the footnote section, it is explained how the variables are presented:

t: t-test, Z: Mann-Whitney test, χ^2: Chi-square test, -: Fisher exact;

SD: standard deviation, M: Median, Q_1: 1st Quartile, Q_3: 3st Quartile;

Continuous variables that follow a normal distribution are typically presented as mean ± standard deviation (SD). For continuous variables that do not follow a normal distribution, the median and interquartile range (IQR) are used. Categorical variables are expressed as n (%);

To be consistent with Table 2, add: n(%), M (Q_1, Q_3), Mean ± SD after the name of each variable in Table 1.

e.g. CLNM, n(%),

Gender, n(%)

Age(years), Mean ± SD

Tumor diameter(cm), M (Q_1, Q_3).

(c)There were also a few minor errors that I found myself and changed:

1. Table 2:no "n (%) "after the variable "HT" 'HT was changed to ---- HT, n (%)'

2. Table 3 was further optimized

The title of Table 3: 'Multivariate Logistic Regression' was changed to "Multivariate Logistic Regression Analysis in the Training Group to be more easily understood".

Removing the original statistic 'Z' column from Table 3, as he was essentially useless would not have had any effect on the results.

Added footnotes: β: coefficient of regression; SE: standard error; OR: odds ratio; CI: confidence interval.

3.We modified Figure 3

changing Age to Age(years) diameter(mmm) to Tumor diameter(cm).