# Interpretable machine learning model based on the systemic inflammation response index and ultrasound features can predict central lymph node metastasis in cN0T1–T2 papillary thyroid carcinoma

**Jin Pang[1]^, Mohan Yang[2], Jun Li[1], Xiaoxiao Zhong[3], Xiangyu Shen[1], Ting Chen[1], Liyuan Qian[1]**

[1]Department of Breast and Thyroid Surgery, Third Xiangya Hospital, Central South University, Changsha, China; [2]Department of Urology, Third Xiangya Hospital, Central South University, Changsha, China; [3]Department of General Surgery, Third Xiangya Hospital, Central South University, Changsha, China

*Contributions:* (I) Conception and design: L Qian, J Pang; (II) Administrative support: L Qian; (III) Provision of study materials or patients: J Pang, M Yang; (IV) Collection and assembly of data: J Li, X Zhong, X Shen, T Chen; (V) Data analysis and interpretation: J Pang, M Yang, L Qian; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

*Correspondence to:* Liyuan Qian, MD. Department of Breast and Thyroid Surgery, Third Xiangya Hospital, Central South University, 138 Tongzipo Road, Changsha, China. Email: qianliyuan2014@126.com.

**Background:** It is arguable whether individuals with T1–T2 papillary thyroid cancer (PTC) who have a clinically negative (cN0) diagnosis should undergo prophylactic central lymph node dissection (pCLND) on a routine basis. Many inflammatory indices, including the neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), monocyte-to-lymphocyte ratio (MLR), and systemic immune-inflammatory index (SII), have been reported in PTC. However, the associations between the systemic inflammation response index (SIRI) and the risk of central lymph node metastasis (CLNM) remain unclear.

**Methods:** Retrospective research involving 1,394 individuals with cN0T1–T2 PTC was carried out, and the included patients were randomly allocated into training (70%) and testing (30%) subgroups. The preoperative inflammatory indices and ultrasound (US) features were used to train the models. To assess the forecasting factors as well as drawing nomograms, the least absolute shrinkage and selection operator (LASSO) and multivariate logistic regression were utilized. Then eight interpretable models based on machine learning (ML) algorithms were constructed, including decision tree (DT), K-nearest neighbor (KNN), support vector machine (SVM), artificial neural network (ANN), random forest (RF), extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), and categorical boosting (CatBoost). The performance of the models was evaluated by incorporating the area under the precision-recall curve (auPR) and the area under the receiver operating characteristic curve (auROC), as well as other conventional metrics. The interpretability of the optimum model was illustrated via the shapley additive explanations (SHAP) approach.

**Results:** Younger age, larger tumor size, capsular invasion, location (lower and isthmus), unclear margin, microcalcifications, color Doppler flow imaging (CDFI) blood flow, and higher SIRI (≥0.77) were independent positive predictors of CLNM, whereas female sex and Hashimoto thyroiditis were independent negative predictors, and nomograms were subsequently constructed. Taking into account both the auROC and auPR, the RF algorithm showed the best performance, and superiority to XGBoost, CatBoost and ANN. In addition, the role of key variables was visualized in the SHAP plot.

**Conclusions:** An interpretable ML model based on the SIRI and US features can be used to predict

^ ORCID: 0009-0003-9720-3647.

*Gland Surg* 2023;12(11):1485-1499 | https://dx.doi.org/10.21037/gs-23-349

CLNM in individuals with cN0T1–T2 PTC.

## Introduction

The primary histologic subtype of thyroid tumors is papillary thyroid cancer (PTC) (1,2). Individuals with lymph node metastasis (LNM) including central lymph node metastasis (CLNM) and lateral lymph node metastasis (LLNM) exhibit greater likelihood of disease persistence, recurrence, and re-operative surgery even though PTC commonly appears as an inert tumor (3-7). Meanwhile, an elevated incidence of postoperative complications is caused by prophylactic central lymph node dissection (pCLND) (8,9). pCLND should not be advised for individuals with the cN0T1–T2 subgroup following the principles of the 2015 American Thyroid Association (ATA) guidelines, whereas guidelines in the majority of East Asian nations prefer pCLND given the premise of ensuring sufficient

parathyroid gland and recurrent laryngeal nerve protection (10-12). Consequently, it is essential to establish a reliable preoperative forecasting algorithm model and to determine potential indicators of CLNM.

Studies have revealed that the inflammatory index, including the systemic immune-inflammatory index (SII), neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR) and monocyte-to-lymphocyte ratio (MLR), can serve as crucial indicators regarding the malignant biological behavior that occurs in several cancers (13-17). Recently, a new blood inflammatory index, the systemic inflammation response index (SIRI), has been utilized for forecasting the prognosis of breast cancer, nasopharyngeal carcinoma, cervical cancer, pancreatic cancer, and colorectal cancer (18-22). However, the relationship between preoperative SIRI levels in peripheral blood and the risk of CLNM in PTC remains unclear.

Machine learning (ML) is a robust collection of algorithms equipped to comprehend, modify, evaluate, and forecast records. It has been extensively utilized to investigate a wide range of illnesses (23). Common ML algorithms include the decision tree (DT), random forest (RF), artificial neural network (ANN), support vector machine (SVM), and K-nearest neighbors (KNN) algorithms (24). categorical boosting (CatBoost), light gradient boosting machine (LightGBM), and extreme gradient boosting (XGBoost) are the most recognized frameworks in the academic discipline of boosting. They stand out among the alternative boosting algorithms owing to the integration of weak classifiers that minimize the loss function. However, most ML algorithms trigger black box issues (25). To tackle the inexplicability dilemma, the shapley additive explanations (SHAP) approach is presented as an alternative solution (26).

Hence, by using the preoperative inflammatory index and ultrasound (US) features, we embarked on establishing and validating eight interpretable ML models to assess CLNM likelihood in individuals with cN0T1–T2 PTC.

---

**Highlight box**

**Key findings**
- The systemic inflammation response index (SIRI) has now been identified as a risk predictor for central lymph node metastasis (CLNM).
- We constructed and verified eight machine learning models based on SIRI and ultrasound features to evaluate CLNM risk in patients with cN0T1–T2 papillary thyroid cancer (PTC); the random forest model performed the best followed by extreme gradient boosting, categorical boosting, and artificial neural network.
- The interpretability of the models was illustrated via the SHapley Additive exPlanations approach.

**What is known and what is new?**
- Several researchers have discovered that monocyte-to-lymphocyte ratio, platelet-to-lymphocyte ratio, neutrophil-to-lymphocyte ratio, and systemic immune-inflammatory index are predictive factors for CLNM and lateral lymph node metastasis (LLNM).
- A pioneering exploration of the predictive value of SIRI in CLNM was carried out.

**What is the implication, and what should change now?**
- Preoperative prediction of CLNM may benefit patients with cN0 T1–T2 PTC.

---

We presented this article in accordance with the TRIPOD reporting checklist (available at https://gs.amegroups.com/article/view/10.21037/gs-23-349/rc).

## Methods

### Study population

A retrospective analysis was carried out on individuals with cN0T1–T2 PTC between January 2020 and December 2021. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of Third Xiangya Hospital, Central South University (No. quick23472), and individual consent for this retrospective analysis was waived.

### Data collection

The inclusion criteria were as follows: (I) T1–T2 (≤40 mm); (II) cN0 (preoperatively non-suspicious positive lymph node); (III) PTC (firstly diagnosed). The exclusion criteria were as follows: (I) cervical irradiation during childhood; (II) previously diagnosed with head and neck cancer or any other kind of tumors (n=36, including secondary metastases to the lungs, primary tumors in other sites, etc.); (III) individuals with preoperative infection or other inflammation (except Hashimoto thyroiditis) (n=16, including rheumatoid arthritis, Sjogren syndrome, etc.); (IV) individuals who suffered from multiple organ dysfunction such as heart failure, liver failure, or uremia (n=9); (V) primary or secondary illnesses causing abnormalities of the blood system (n=6, including aplastic anemia, etc.); (VI) incompleteness of medical records (n=58). Ultimately, a total of 125 individuals were excluded; thus, 1,394 individuals were included to establish and assess the model, and they were randomly allocated into training (70%) and testing (30%) subsets.

### Surgical strategy

Individuals diagnosed with PTC routinely undergo pCLND in our department. Thyroid lobectomy or total thyroidectomy corresponds to ipsilateral or bilateral pCLND, respectively.

### Inflammatory indices and US features

The 20 features were as follows: gender, age, tumor size, capsular invasion, laterality, multifocality, location, echogenicity, solid composition (almost 100% solid component), unclear margin, shape, aspect ratio, microcalcifications, color Doppler flow imaging (CDFI) blood flow, Hashimoto thyroiditis and inflammatory indices, including MLR, PLR, NLR, SII, and SIRI. The inflammatory index ranks were split into high and low subgroups, which were determined by the receiver operating characteristic (ROC) curve's optimal cutoff value for presence of lymph-node metastasis. By applying the aforementioned strategies, each of the optimum cut-off value was determined: NLR (low <1.83, high ≥1.83), PLR (low <146.58, high ≥146.58), MLR (low <0.27, high ≥0.27), SII (low <395.22, high ≥395.22), and SIRI (low <0.77, high ≥0.77).

### Construction of the nomogram

We ran 10-fold cross-validation codes to calculate the optimum punishment parameter and conduct dimensionality reduction procedures on the statistical framework. Then, we screened the least absolute shrinkage and selection operator (LASSO) regression to obtain nonzero coefficient features. Subsequently, a nomogram plot was drawn based on the results of the multivariate regression analysis. Currently recognized tools for nomogram evaluation include the ROC curve, calibration curve, and decision curve analysis (DCA), all of which were employed.

### Development, evaluation, and visual interpretation of ML models

There were eight supervised ML algorithms, including DT, KNN, SVM, ANN, RF, CatBoost, LightGBM, and XGBoost. Fivefold cross-validation was utilized to diminish overfitting, and then we performed repeated testing and tuning to obtain the optimal model parameters. The sensitivity, specificity, area under the ROC curve (auROC), accuracy, precision, recall, area under the precision-recall curve (auPR), F1-score, and Matthews correlation coefficient (MCC) of the ML algorithms were calculated. We used the confusion matrix as a visual illustration and employed DCA for assessing the clinical usefulness. Through game-theoretic tactics, SHAP presents a superior visual tool for evaluating the significance of the attributes.

### Statistical analysis

We ran the software R (version 4.3.0), Anaconda 3, and

1488

Pang et al. ML model predicts the CLNM in cN0T1–T2 PTC

Python (version 3.10.9) environments as statistical tools. The following packages were running in the generation of the code for algorithms: "pROC", "caret", "glmnet", "rms", "ggDCA", "ggplot2", "tidymodels", "fastshap", "bonsai", "treesnip", and "reticulate".

## Results

### Patient characteristics

*Table 1* displays the baseline traits of CLNM(+) and CLNM(–). There were no significant differences between the training and testing subsets (P>0.05) (*Table 2*).

### LASSO regression feature selection in the training set

LASSO regression yielded 15 nonzero coefficient features, including gender, age, tumor size, capsular invasion, laterality, multifocality, location, solid component, unclear margin, microcalcifications, CDFI blood flow, Hashimoto's thyroiditis, NLR, MLR, and SIRI (*Figure 1*).

### Construction and validation of the nomogram

Utilizing the 15 nonzero coefficient features, the multivariate analysis revealed that younger age, larger tumor size, capsular invasion, location (lower and isthmus), unclear margin, microcalcifications, CDFI blood flow, and higher SIRI (≥0.77) were independent positive predictors of CLNM, while female and Hashimoto thyroiditis were independent negative predictors (*Table 3*). Next, a nomogram plot was drawn on the basis of the training cohort's multivariate analysis (*Figure 2*). The ROC curve demonstrated a desirable discrimination capacity, with AUCs of 0.834 and 0.803 in the training and testing cohorts, respectively (*Figure 3A,3B*). The calibration curve exhibited notable consistency, regarding mean absolute errors in the two cohorts of 0.017 and 0.015, respectively (*Figure 3C,3D*). The DCA showed broad clinical utility when the threshold probability of an individual was between approximately 20% and 90% (*Figure 3E,3F*).

### Development and evaluation of ML models

Utilizing the 15 potential features selected by LASSO regression, 8 ML algorithm prediction models for CLNM were developed. The RF model performed optimally, with auROC values of 0.8177 and auPR values of 0.8029,

followed by XGBoost (0.8130 and 0.7987), CatBoost (0.8130 and 0.7985) and ANN (0.8105 and 0.7990). Detailed information concerning sensitivity, specificity, accuracy, precision, recall, f1 score and mcc is summarized in *Table 4* and Figure S1. The DCA plot proved that RF had better clinical suitability (Figure S1C).

### The RF model performance

*Figure 4A* shows the modeling process of the RF model, and choosing the appropriate parameters (mtry:2 and ntree:250) made the RF model perform best. Starting from the 100th DT, the error of the RF algorithm gradually flattened, indicating that the generalization ability of the RF algorithm gradually increased. We also used RF to explore the importance of variables. As illustrated in *Figure 4B*, the top 5 variables, CDFI blood flow, location, age, tumor size, and microcalcifications, are analogous when evaluated via two measures: a decrease in classification accuracy (mean lowered accuracy) and a decrease in node impurity (mean decreased Gini). The SIRI, which performed relatively better among the three inflammatory indices, ranked tenth and ninth in the mean decrease accuracy and Gini plot, respectively.

The confusion matrix of RF is displayed in Figure S2A,S2B. The respective auROCs were 85.48% and 81.77%, and DeLong's test between the training and testing cohorts revealed that there were no statistically significant differences (P>0.05) (Figure S2C). The learning curves indicate that the training and testing sets have a strong fitting ability and high stability (Figure S2D). In general, the RF model effectively prevents overfitting. Additionally, Figure S2E,S2F visually shows the predicted probability distribution of the RF model.

### Explanation of the ML model with the SHAP method

We performed interpretability manipulations using the SHAP tool in the RF and XGBoost models. Ranking of variable contributions was assessed by the mean absolute SHAP values (*Figure 5A,5B*). The top ten features in the RF model were age, CDFI blood flow, tumor size, location, microcalcifications, Hashimoto thyroiditis, unclear margin, SIRI, gender, and solid composition. In addition, we constructed scatter plots of SHAP summary plots, which visualized the relationship between eigenvalues and predicted probabilities by color (Figure S3A,S3B). The larger the absolute value on the x-axis, the more the

**Table 1** Baseline characteristics of the whole cohort grouped by lymph node status

| Variables | Total (n=1,394) | CLNM(−) (n=718) | CLNM(+) (n=676) | P | $\chi^2$ |
|---|---|---|---|---|---|
| Gender | | | | <0.001 | 37.345 |
| Male | 330 (23.7) | 121 (16.9) | 209 (30.9) | | |
| Female | | | | | |
| Age (years) | | | | <0.001 | 107.798 |
| >55 | 192 (13.8) | 139 (19.4) | 53 (7.8) | | |
| 40–55 | 596 (42.8) | 359 (50.0) | 237 (35.1) | | |
| <40 | 606 (43.5) | 220 (30.6) | 386 (57.1) | | |
| Tumor size (mm) | | | | <0.001 | 90.250 |
| <10 | 1,027 (73.7) | 606 (84.4) | 421 (62.3) | | |
| 10–20 | 299 (21.4) | 97 (13.5) | 202 (29.9) | | |
| 21–40 | 68 (4.9) | 15 (2.1) | 53 (7.8) | | |
| Capsular invasion | | | | <0.001 | 40.393 |
| No | 1,269 (91.0) | 688 (95.8) | 581 (85.9) | | |
| Yes | – | – | – | | |
| Laterality | | | | <0.001 | 18.010 |
| Unilateral | 1,095 (78.6) | 597 (83.1) | 498 (73.7) | | |
| Bilateral | – | – | – | | |
| Multifocality | | | | <0.001 | 15.323 |
| Solitary tumor | 894 (64.1) | 496 (69.1) | 398 (58.9) | | |
| Multifocal tumor | – | – | – | | |
| Location | | | | <0.001 | 71.795 |
| Upper | 257 (18.4) | 170 (23.7) | 87 (12.9) | | |
| Middle | 631 (45.3) | 360 (50.1) | 271 (40.1) | | |
| Lower | 373 (26.8) | 141 (19.6) | 232 (34.3) | | |
| Isthmus | 133 (9.5) | 47 (6.5) | 86 (12.7) | | |
| Echogenicity | | | | 0.006 | 12.584 |
| Hyper or isoechoic | 22 (1.6) | 11 (1.5) | 11 (1.6) | | |
| Mixed-echoic | 64 (4.6) | 28 (3.9) | 36 (5.3) | | |
| Hypo-echoic | 1,277 (91.6) | 672 (93.6) | 605 (89.5) | | |
| Very hypo-echoic | 31 (2.2) | 7 (1.0) | 24 (3.6) | | |
| Solid composition | | | | <0.001 | 20.749 |
| No | 939 (67.4) | 524 (73.0) | 415 (61.4) | | |
| Yes | – | – | – | | |
| Unclear margin | | | | <0.001 | 37.997 |
| No | 399 (28.6) | 258 (35.9) | 141 (20.9) | | |
| Yes | – | – | – | | |

**Table 1** (*continued*)

**Table 1** (*continued*)

| Variables | Total (n=1,394) | CLNM(−) (n=718) | CLNM(+) (n=676) | P | $\chi^2$ |
|---|---|---|---|---|---|
| Shape | | | | <0.001 | 19.172 |
|   Regular | 932 (66.9) | 519 (72.3) | 413 (61.1) | | |
| Irregular or lobulated | | | | | |
| Aspect ratio | | | | 0.003 | 8.692 |
|   A/T <1 | 757 (54.3) | 362 (50.4) | 395 (58.4) | | |
|   A/T ≥1 | – | – | – | | |
| Microcalcifications | | | | <0.001 | 95.395 |
|   No | 458 (32.9) | 322 (44.8) | 136 (20.1) | | |
|   Yes | – | – | – | | |
| CDFI blood flow | | | | <0.001 | 120.559 |
|   No | 1,072 (76.9) | 639 (89.0) | 433 (64.1) | | |
|   Yes | – | – | – | | |
| Hashimoto thyroiditis | | | | <0.001 | 28.898 |
|   No | 924 (66.3) | 428 (59.6) | 496 (73.4) | | |
|   Yes | – | – | – | | |
| NLR | | | | <0.001 | 14.022 |
|   Low (<1.83) | 617 (44.3) | 353 (49.2) | 264 (39.1) | | |
|   High (≥1.83) | – | – | – | | |
| PLR | | | | 0.116 | 2.472 |
|   Low (<146.58) | 875 (62.8) | 436 (60.7) | 439 (64.9) | | |
|   High (≥146.58) | – | – | – | | |
| MLR | | | | 0.007 | 7.396 |
|   Low (<0.27) | 1,123 (80.6) | 599 (83.4) | 524 (77.5) | | |
|   High (≥0.27) | – | – | – | | |
| SII | | | | <0.001 | 13.736 |
|   Low (<395.22) | 528 (37.9) | 306 (42.6) | 222 (32.8) | | |
|   High (≥395.22) | – | – | – | | |
| SIRI | | | | <0.001 | 17.874 |
|   Low (<0.77) | 791 (56.7) | 447 (62.3) | 344 (50.9) | | |
|   High (≥0.77) | – | – | – | | |

Data are presented as N (%). CLNM, central lymph node metastasis; A/T, aspect ratio (height divided by width on transverse views); CDFI, color Doppler flow imaging; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; MLR, monocyte-to-lymphocyte ratio; SII, systemic immune-inflammatory index; SIRI, systemic inflammation response index.

**Table 2** Baseline characteristics between the training and testing sets

| Variables | Total (n=1,394) | Training set (n=976) | Testing set (n=418) | P | $\chi^2$ |
|---|---|---|---|---|---|
| CLNM | | | | 0.834 | 0.044 |
| Negative group | 718 (51.5) | 505 (51.7) | 213 (51.0) | | |
| Positive group | – | – | – | | |
| Gender | | | | 0.147 | 2.104 |
| Male | 330 (23.7) | 220 (22.5) | 110 (26.3) | | |
| Female | – | – | – | | |
| Age (years) | | | | 0.408 | 1.794 |
| >55 | 192 (13.8) | 138 (14.1) | 54 (12.9) | | |
| 40–55 | 596 (42.8) | 425 (43.5) | 171 (40.9) | | |
| <40 | 606 (43.5) | 413 (42.3) | 193 (46.2) | | |
| Tumor size (mm) | | | | 0.963 | 0.075 |
| <10 | 1,027 (73.7) | 718 (73.6) | 309 (73.9) | | |
| 10–20 | 299 (21.4) | 211 (21.6) | 88 (21.1) | | |
| 21–40 | 68 (4.9) | 47 (4.8) | 21 (5.0) | | |
| Capsular invasion | | | | 0.542 | 0.372 |
| No | 1,269 (91.0) | 885 (90.7) | 384 (91.9) | | |
| Yes | – | – | – | | |
| Laterality | | | | 0.554 | 0.35 |
| Unilateral | 1,095 (78.6) | 762 (78.1) | 333 (79.7) | | |
| Bilateral | – | – | – | | |
| Multifocality | | | | 0.433 | 0.614 |
| Solitary tumor | 894 (64.1) | 619 (63.4) | 275 (65.8) | | |
| Multifocal tumor | – | – | – | | |
| Location | | | | 0.177 | 4.936 |
| Upper | 257 (18.4) | 183 (18.8) | 74 (17.7) | | |
| Middle | 631 (45.3) | 448 (45.9) | 183 (43.8) | | |
| Lower | 373 (26.8) | 263 (26.9) | 110 (26.3) | | |
| Isthmus | 133 (9.5) | 82 (8.4) | 51 (12.2) | | |
| Echogenicity | | | | 0.763 | 1.159 |
| Hyper or isoechoic | 22 (1.6) | 15 (1.5) | 7 (1.7) | | |
| Mixed-echoic | 64 (4.6) | 42 (4.3) | 22 (5.3) | | |
| Hypo-echoic | 1,277 (91.6) | 899 (92.1) | 378 (90.4) | | |
| Very hypo-echoic | 31 (2.2) | 20 (2.0) | 11 (2.6) | | |
| Solid composition | | | | >0.99 | 0 |
| No | 939 (67.4) | 657 (67.3) | 282 (67.5) | | |
| Yes | – | – | – | | |

**Table 2** (*continued*)

1492

Pang et al. ML model predicts the CLNM in cN0T1–T2 PTC

**Table 2** (*continued*)

| Variables | Total (n=1,394) | Training set (n=976) | Testing set (n=418) | P | $\chi^2$ |
|---|---|---|---|---|---|
| Unclear margin | | | | 0.592 | 0.287 |
| No | 399 (28.6) | 284 (29.1) | 115 (27.5) | | |
| Yes | – | – | – | | |
| Shape | | | | 0.135 | 2.233 |
| Regular | 932 (66.9) | 640 (65.6) | 292 (69.9) | | |
| Irregular or lobulated | – | – | – | | |
| Aspect ratio | | | | 0.77 | 0.085 |
| A/T <1 | 757 (54.3) | 533 (54.6) | 224 (53.6) | | |
| A/T ≥1 | – | – | – | | |
| Microcalcifications | | | | 0.52 | 0.413 |
| No | 458 (32.9) | 315 (32.3) | 143 (34.2) | | |
| Yes | – | – | – | | |
| CDFI blood flow | | | | 0.401 | 0.705 |
| No | 1,072 (76.9) | 744 (76.2) | 328 (78.5) | | |
| Yes | – | – | – | | |
| Hashimoto thyroiditis | | | | 0.846 | 0.038 |
| No | 924 (66.3) | 649 (66.5) | 275 (65.8) | | |
| Yes | – | – | – | | |
| NLR | | | | 0.518 | 0.417 |
| Low (<1.83) | 617 (44.3) | 426 (43.6) | 191 (45.7) | | |
| High (≥1.83) | – | – | – | | |
| PLR | | | | 0.459 | 0.549 |
| Low (<146.58) | 875 (62.8) | 606 (62.1) | 269 (64.4) | | |
| High (≥146.58) | – | – | – | | |
| MLR | | | | 0.112 | 2.527 |
| Low (<0.27) | 1,123 (80.6) | 775 (79.4) | 348 (83.3) | | |
| High (≥0.27) | – | – | – | | |
| SII | | | | 0.387 | 0.748 |
| Low (<395.22) | 528 (37.9) | 362 (37.1) | 166 (39.7) | | |
| High (≥395.22) | – | – | – | | |
| SIRI | | | | 0.182 | 1.782 |
| Low (<0.77) | 791 (56.7) | 542 (55.5) | 249 (59.6) | | |
| High (≥0.77) | – | – | – | | |

Data are presented as N (%). CLNM, central lymph node metastasis; A/T, aspect ratio (height divided by width on transverse views); CDFI, color Doppler flow imaging; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; MLR, monocyte-to-lymphocyte ratio; SII, systemic immune-inflammatory index; SIRI, systemic inflammation response index.
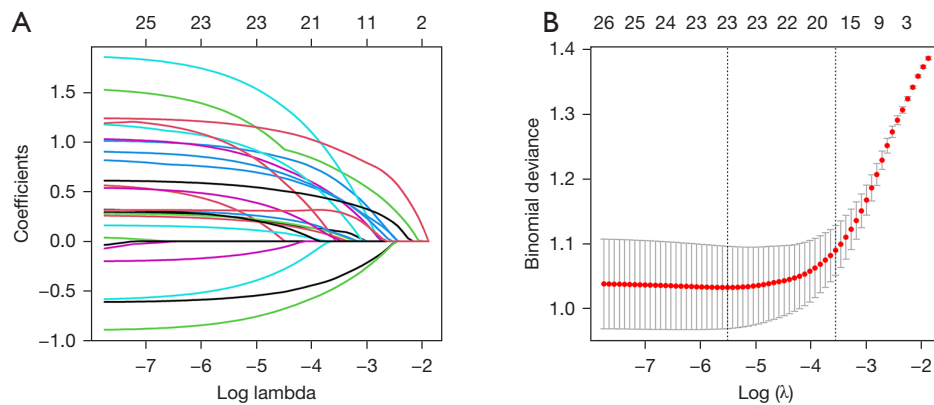
**Figure 1** Feature selection using LASSO regression. (A) LASSO coefficient profile plots of the 15 variables. (B) Selection of features using cross-validation; dotted vertical lines were drawn at the optimal values using the minimum criteria and the one-fold standard error of the minimum criteria. LASSO, least absolute shrinkage and selection operator.

**Table 3** Multivariate analysis

| Variables | OR | 95% CI | P |
|---|---|---|---|
| Gender | | | |
| Male | Reference | | |
| Female | 0.511 | 0.347–0.752 | 0.001 |
| Age (years) | | | |
| >55 | Reference | | |
| 40–55 | 1.772 | 1.073–2.926 | 0.025 |
| <40 | 4.794 | 2.877–7.988 | <0.001 |
| Tumor size (mm) | | | |
| <10 | Reference | | |
| 10–20 | 3.194 | 2.126–4.799 | <0.001 |
| 21–40 | 6.675 | 2.754–16.178 | <0.001 |
| Capsular invasion | | | |
| No | Reference | | |
| Yes | 2.862 | 1.598–5.126 | <0.001 |
| Laterality | | | |
| Unilateral | Reference | | |
| Bilateral | 1.276 | 0.748–2.177 | 0.371 |
| Multifocality | | | |
| Solitary tumor | Reference | | |
| Multifocal tumor | 1.296 | 0.825–2.037 | 0.261 |
| Location | | | |
| Upper | Reference | | |
| Middle | 1.077 | 0.7–1.658 | 0.735 |
| Lower | 2.288 | 1.429–3.663 | 0.001 |
| Isthmus | 3.373 | 1.738–6.545 | <0.001 |

**Table 3** (*continued*)

**Table 3** (*continued*)

| Variables | OR | 95% CI | P |
|---|---|---|---|
| Solid composition | | | |
| No | Reference | | |
| Yes | 1.369 | 0.974–1.923 | 0.07 |
| Unclear margin | | | |
| No | Reference | | |
| Yes | 2.43 | 1.697–3.479 | <0.001 |
| Microcalcifications | | | |
| No | Reference | | |
| Yes | 1.851 | 1.311–2.613 | <0.001 |
| CDFI blood flow | | | |
| No | Reference | | |
| Yes | 3.47 | 2.349–5.125 | <0.001 |
| Hashimoto thyroiditis | | | |
| No | Reference | | |
| Yes | 0.409 | 0.29–0.577 | <0.001 |
| NLR | | | |
| Low (<1.83) | Reference | | |
| High (≥1.83) | 1.337 | 0.917–1.95 | 0.131 |
| MLR | | | |
| Low (<0.27) | Reference | | |
| High (≥0.27) | 1.393 | 0.894–2.169 | 0.142 |
| SIRI | | | |
| Low (<0.77) | Reference | | |
| High (≥0.77) | 1.578 | 1.046–2.379 | 0.03 |

OR, odd ratio; CI, confidence interval; CDFI, color Doppler flow imaging; NLR, neutrophil-to-lymphocyte ratio; MLR, monocyte-to-lymphocyte ratio; SIRI, systemic inflammation response index.
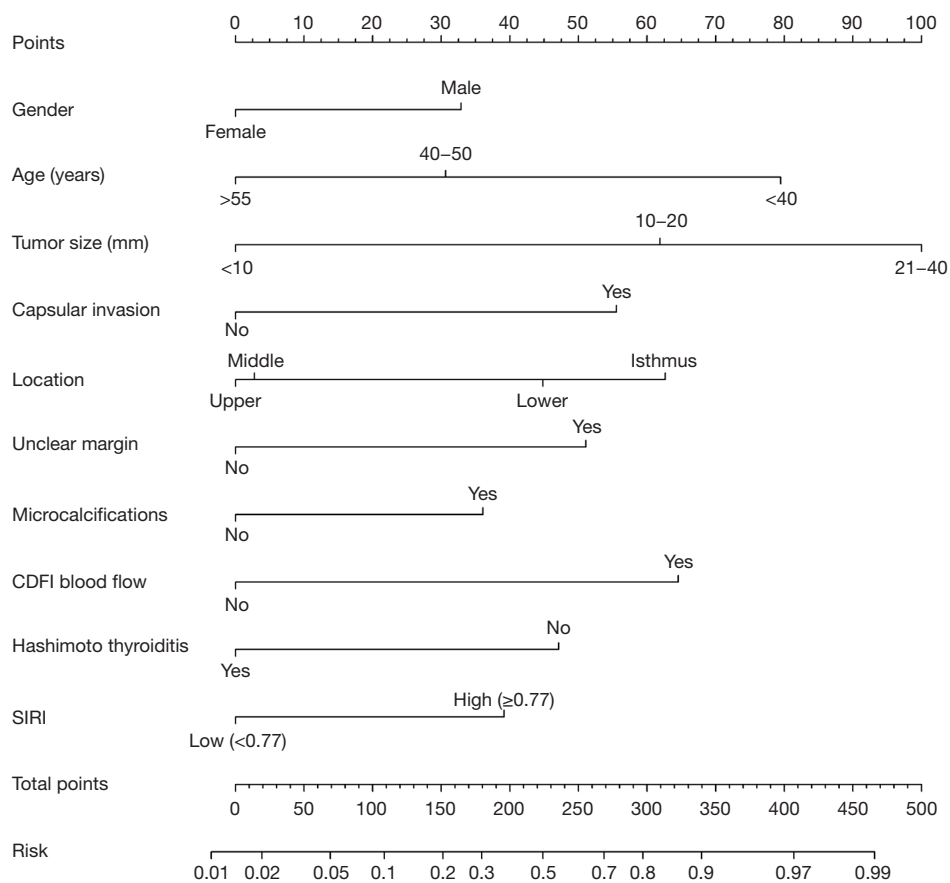
**Figure 2** Nomogram for predicting CLNM in individuals with cN0T1–T2 PTC. CDFI, color Doppler flow imaging; SIRI, systemic inflammation response index; CLNM, central lymph node metastasis; PTC, papillary thyroid cancer.

attribute affects the output, with colors representing high (red) and low (blue) raw eigenvalues. We can see that a higher SIRI has a positive impact, while Hashimoto thyroiditis has a negative impact. To visualize the contributions of individual variable levels, we implemented it with the help of the facet wrap method based on the SHAP value (Figure S4).

*Figure 6* presents a couple of classic scenarios that showcase the model's capacity for interpretation. The CLNM-absent individual obtained a poorer SHAP value (0.15) (*Figure 6A*), while CLNM-present individual obtained a stronger SHAP value (0.94) (*Figure 6B*).

## Discussion

### *Principal findings*

We had three major findings in this research. First, in addition to traditional US features, the inflammatory index,

especially the SIRI, was found to be a risk predictor for CLNM. Second, we established and verified eight ML models to assess CLNM likelihood in individuals with cN0T1–T2 PTC. The RF model performed the best (with maximum auROC and auPR), followed by XGBoost, CatBoost, and ANN. Third, the interpretability of the models was illustrated via the SHAP approach.

Consistent with numerous previous clinical studies, younger age, presence of CDFI blood flow, larger tumor, tumor located in the lower or isthmus, microcalcifications, absence of Hashimoto, unclear margin, male gender and capsular invasion were all found to be risk factors for CLNM (27-29). Several researchers have discovered that blood inflammatory indices such as MLR, PLR, NLR, and SII are predictive factors for CLNM and LLNM, and are even associated with poor prognosis and relapse (30-33). However, use of a new blood inflammatory index, SIRI, and research regarding its importance in PTC are still
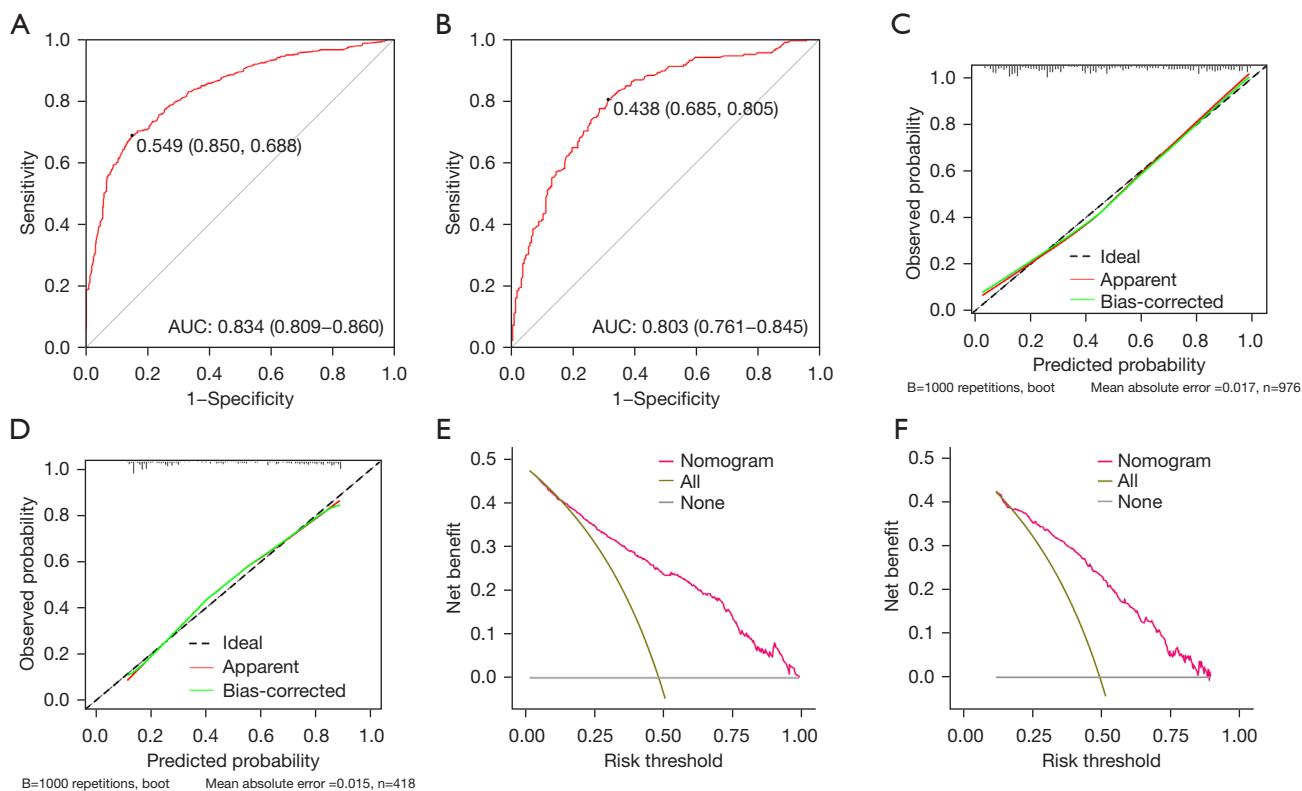
**Figure 3** The differential capability of the nomogram. (A) The ROC curve in the training set. (B) The ROC curve in the testing set. (C) The calibration curve in the training set. (D) The calibration curve in the testing set. (E) The DCA in the training set. (F) The DCA in the testing set. AUC, area under curve; ROC, receiver operating characteristic area; DCA, decision curve analysis.

**Table 4** Model capabilities of the eight ML algorithms in the testing set

| Model | ROC_AUC | PR_AUC | Sensitivity | Specificity | Accuracy | Precision | Recall | F1_score | MCC |
|---|---|---|---|---|---|---|---|---|---|
| DT | 0.7478 | 0.7168 | 0.6293 | 0.7465 | 0.6890 | 0.7049 | 0.6293 | 0.6649 | 0.3786 |
| CatBoost | 0.8130 | 0.7985 | 0.6537 | 0.8263 | 0.7416 | 0.7836 | 0.6537 | 0.7128 | 0.4880 |
| KNN | 0.7501 | 0.7311 | 0.7463* | 0.6150 | 0.6794 | 0.6511 | 0.7463* | 0.6955 | 0.3641 |
| LightGBM | 0.8057 | 0.7937 | 0.7415 | 0.7371 | 0.7392 | 0.7308 | 0.7415 | 0.7361 | 0.4785 |
| RF | 0.8177* | 0.8029* | 0.6732 | 0.7934 | 0.7344 | 0.7582 | 0.6732 | 0.7132 | 0.4705 |
| XGBoost | 0.8130 | 0.7987 | 0.6439 | 0.8310* | 0.7392 | 0.7857* | 0.6439 | 0.7078 | 0.4842 |
| SVM | 0.8088 | 0.7940 | 0.6927 | 0.7840 | 0.7392 | 0.7553 | 0.6927 | 0.7226 | 0.4791 |
| ANN | 0.8105 | 0.7990 | 0.7268 | 0.7887 | 0.7584* | 0.7680 | 0.7268 | 0.7469* | 0.5168* |

*, the maximum value of the column. ML, machine learning; ROC, receiver operating characteristic; AUC, area under curve; PR, precision-recall; MCC, Matthews correlation coefficient; DT, decision tree; CatBoost, categorical boosting; KNN, K-nearest neighbors; LightGBM, light gradient boosting machine; RF, random forest; XGBoost, extreme gradient boosting; SVM, support vector machine; ANN, artificial neural network.

lacking. This could be the first investigation regarding the relationship between SIRI and CLNM to the best of our knowledge.

In previous studies, scholars have used several ML algorithms to predict CLNM and LLNM (34,35). Different algorithms have their pros and cons (36). The RF model

1496

Pang et al. ML model predicts the CLNM in cN0T1–T2 PTC
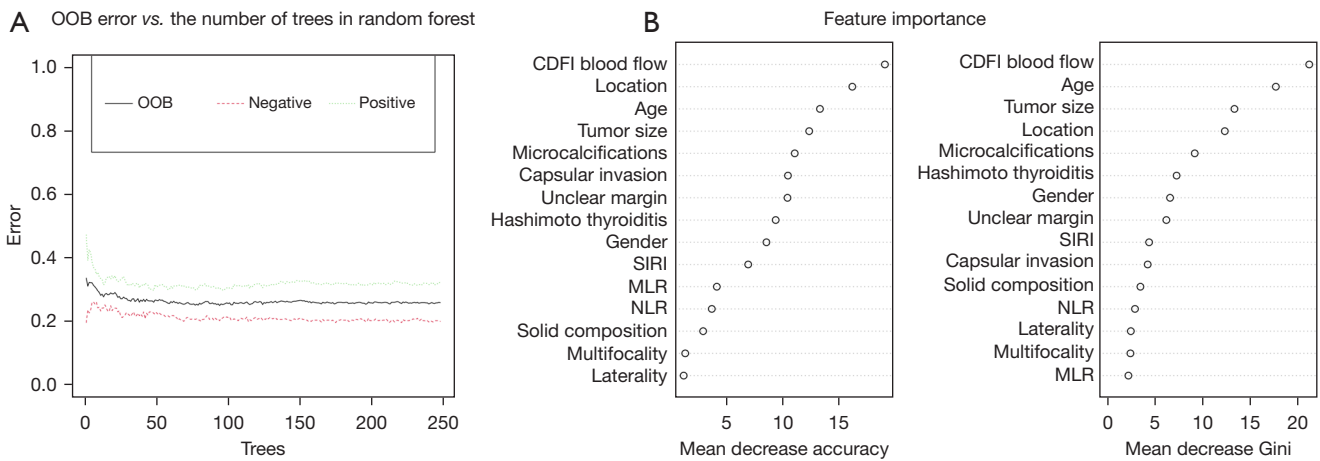
**Figure 4** The fitting tuning and variable importance of the RF model. (A) Diagram of model error rate and tree. (B) Variable importance is given by the mean decrease accuracy (left) and mean decrease Gini (right). OOB, out-of-bag; CDFI, color Doppler flow imaging; SIRI, systemic inflammation response index; MLR, monocyte-to-lymphocyte ratio; NLR, neutrophil-to-lymphocyte ratio; RF, random forest.
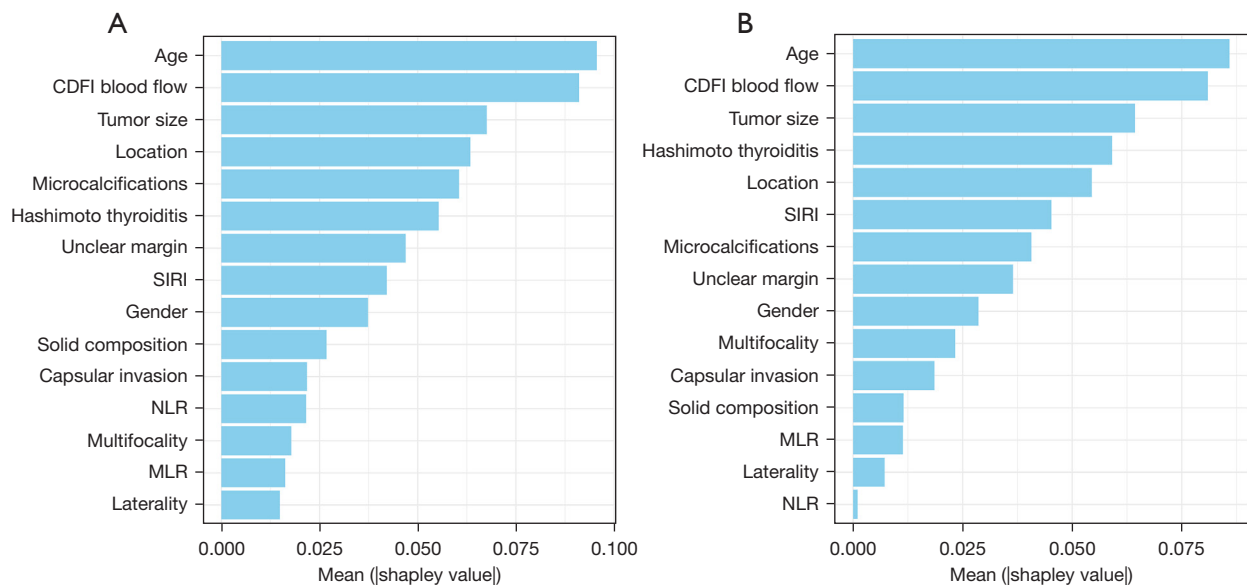


**Figure 5** Feature importance ranking as indicated by SHAP assessing eigenvalue contribution. (A) RF model; (B) XGBoost model. CDFI, color Doppler flow imaging; SIRI, systemic inflammation response index; NLR, neutrophil-to-lymphocyte ratio; MLR, monocyte-to-lymphocyte ratio; SHAP, shapley additive explanations; RF, random forest; XGBoost, extreme gradient boosting.

performed the best in this research. Several new ensemble learning algorithms have also shown good predictive performance, including XGBoost and CatBoost. Ensemble learning is mainly divided into bagging algorithms and boosting algorithms. Via bagging theory, the RF algorithm incorporates many DTs. The gradient boosting decision tree (GBDT) represents a broad family of algorithms

through the boosting theory of ensemble learning. LightGBM, XGBoost and CatBoost are the latest and most recognized algorithm members with enhanced capabilities in the GBDT theory family (37-39).

It is worth noting that a major flaw in most ML models is the black box problem. What we did to conquer this flaw was introduce the SHAP tool. In this study, we not
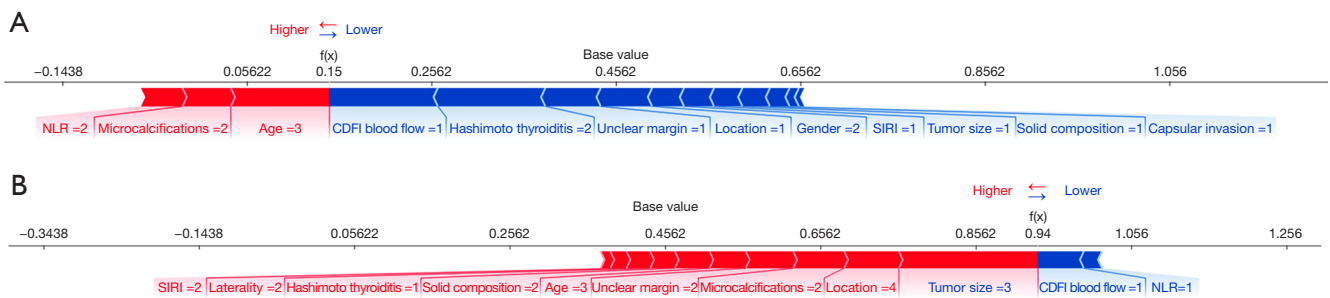
**Figure 6** SHAP individual force plot for the local interpretation in the RF model. The base value is the predicted value without providing input to the model, while f(x) is the probability forecast value of each observation; red indicates an increased risk of CLNM, and blue indicates a reduced risk of CLNM; the length of the arrows helps visualize the extent to which the prediction is affected; the longer the arrow, the greater the effect. NLR, neutrophil-to-lymphocyte ratio; CDFI, color Doppler flow imaging; SIRI, systemic inflammation response index; SHAP, shapley additive explanations; RF, random forest; CLNM, central lymph node metastasis.

only applied the global interpretations of SHAP to visually demonstrate the whole attributes but also delivered specific interpretations utilizing SHAP individual force plots encompassing both positive and negative effects.

### Limitations

First, there is a need for more prospective extensive investigations, considering that these findings relied on retrospective observations. Second, the models were randomly allocated into training and testing cohorts to diminish overfitting due to a lack of external validation. Next, we will build the database with the collaboration of multiple medical centers to further examine the model's capabilities.

### Conclusions

Interpretable ML models based on the SIRI and US features can be used to predict CLNM in individuals with cN0T1–T2 PTC.

### Acknowledgments

The authors would like to thank the numerous individuals who participated in this study.
*Funding:* None.

### Footnote

*Reporting Checklist:* The authors have completed the TRIPOD reporting checklist. Available at https://

gs.amegroups.com/article/view/10.21037/gs-23-349/rc

*Data Sharing Statement:* Available at https://gs.amegroups.com/article/view/10.21037/gs-23-349/dss

*Peer Review File:* Available at https://gs.amegroups.com/article/view/10.21037/gs-23-349/prf

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at https://gs.amegroups.com/article/view/10.21037/gs-23-349/coif). The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of Third Xiangya Hospital, Central South University (No. quick23472) and individual consent for this retrospective analysis was waived.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/

1498

Pang et al. ML model predicts the CLNM in cN0T1–T2 PTC

## References

1. Pizzato M, Li M, Vignat J, et al. The epidemiological landscape of thyroid cancer worldwide: GLOBOCAN estimates for incidence and mortality rates in 2020. Lancet Diabetes Endocrinol 2022;10:264-72.

2. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.

3. Pinheiro RA, Leite AK, Cavalheiro BG, et al. Incidental Node Metastasis as an Independent Factor of Worse Disease-Free Survival in Patients with Papillary Thyroid Carcinoma. Cancers (Basel) 2023;15:943.

4. Yu ST, Ge J, Wei Z, et al. The lymph node yield in the initial lateral neck dissection predicts recurrence in the lateral neck of papillary thyroid carcinoma: a revision surgery cohort study. Int J Surg 2023;109:1264-70.

5. Luo Z, Hei H, Qin J, et al. Lymph node ratio in lateral neck is an independent risk factor for recurrence-free survival in papillary thyroid cancer patients with positive lymph nodes. Endocrine 2022;78:484-90.

6. Hutchinson KA, Guerra A, Payne AE, et al. Risk Factors Associated With Reoperative Surgery for Thyroid Malignancies: A Retrospective Cohort Study. Otolaryngol Head Neck Surg 2023;168:392-7.

7. Wang W, Ding Y, Jiang W, et al. Can Cervical Lymph Node Metastasis Increase the Risk of Distant Metastasis in Papillary Thyroid Carcinoma? Front Endocrinol (Lausanne) 2022;13:917794.

8. Baud G, Jannin A, Marciniak C, et al. Impact of Lymph Node Dissection on Postoperative Complications of Total Thyroidectomy in Patients with Thyroid Carcinoma. Cancers (Basel) 2022;14:5462.

9. Privitera F, Centonze D, La Vignera S, et al. Risk Factors for Hypoparathyroidism after Thyroid Surgery: A Single-Center Study. J Clin Med 2023;12:1956.

10. Yang J, Han Y, Min Y, et al. Prophylactic central neck dissection for cN0 papillary thyroid carcinoma: is there any difference between western countries and China? A systematic review and meta-analysis. Front Endocrinol (Lausanne) 2023;14:1176512.

11. Wang Y, Xiao Y, Pan Y, et al. The effectiveness and safety of prophylactic central neck dissection in clinically node-negative papillary thyroid carcinoma patients: A meta-analysis. Front Endocrinol (Lausanne) 2022;13:1094012.

12. Feng JW, Ye J, Wu WX, et al. Management of cN0 papillary thyroid microcarcinoma patients according to risk-scoring model for central lymph node metastasis and predictors of recurrence. J Endocrinol Invest 2020;43:1807-17.

13. Duan X, Yang B, Zhao C, et al. Prognostic value of preoperative hematological markers in patients with glioblastoma multiforme and construction of random survival forest model. BMC Cancer 2023;23:432.

14. Huang C, Wang M, Chen L, et al. The pretherapeutic systemic inflammation score is a prognostic predictor for elderly patients with oesophageal cancer: a case control study. BMC Cancer 2023;23:505.

15. Huai Q, Luo C, Song P, et al. Peripheral blood inflammatory biomarkers dynamics reflect treatment response and predict prognosis in non-small cell lung cancer patients with neoadjuvant immunotherapy. Cancer Sci 2023. [Epub ahead of print]. doi: 10.1111/cas.15964.

16. Makino T, Izumi K, Iwamoto H, et al. Comparison of the Prognostic Value of Inflammatory and Nutritional Indices in Nonmetastatic Renal Cell Carcinoma. Biomedicines 2023;11:533.

17. Duque-Santana V, López-Campos F, Martin-Martin M, et al. Neutrophil-to-Lymphocyte Ratio and Platelet-to-Lymphocyte Ratio as Prognostic Factors in Locally Advanced Rectal Cancer. Oncology 2023;101:349-57.

18. Dong J, Sun Q, Pan Y, et al. Pretreatment systemic inflammation response index is predictive of pathological complete response in patients with breast cancer receiving neoadjuvant chemotherapy. BMC Cancer 2021;21:700.

19. Lu YF, Wu CY, Lo WC, et al. Postchemoradiotherapy systemic inflammation response index predicts treatment response and overall survival for patients with locally advanced nasopharyngeal cancer. J Formos Med Assoc 2023;122:1141-9.

20. Shan M, Deng Y, Zou W, et al. Salvage radiotherapy strategy and its prognostic significance for patients with locoregional recurrent cervical cancer after radical hysterectomy: a multicenter retrospective 10-year analysis. BMC Cancer 2023;23:905.

21. Pacheco-Barcia V, Mondéjar Solís R, France T, et al. A systemic inflammation response index (SIRI) correlates with survival and predicts oncological outcome for mFOLFIRINOX therapy in metastatic pancreatic cancer. Pancreatology 2020;20:254-64.

22. Cai H, Chen Y, Zhang Q, et al. High preoperative CEA and systemic inflammation response index (C-SIRI) predict unfavorable survival of resectable colorectal cancer.

World J Surg Oncol 2023;21:178.

23. Satapathy P, Pradhan KB, Rustagi S, et al. Application of machine learning in surgery research: current uses and future directions - editorial. Int J Surg 2023;109:1550-1.

24. Kufel J, Bargieł-Łączek K, Kocot S, et al. What Is Machine Learning, Artificial Neural Networks and Deep Learning?-Examples of Practical Applications in Medicine. Diagnostics (Basel) 2023;13:2582.

25. Ali S, Akhlaq F, Imran AS, et al. The enlightening role of explainable artificial intelligence in medical & healthcare domains: A systematic literature review. Comput Biol Med 2023;166:107555.

26. Nohara Y, Matsumoto K, Soejima H, et al. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. Comput Methods Programs Biomed 2022;214:106584.

27. Li J, Sun P, Huang T, et al. Preoperative prediction of central lymph node metastasis in cN0T1/T2 papillary thyroid carcinoma: A nomogram based on clinical and ultrasound characteristics. Eur J Surg Oncol 2022;48:1272-9.

28. Wang W, Ding Y, Meng C, et al. Patient's age with papillary thyroid cancer: Is it a key factor for cervical lymph node metastasis? Eur J Surg Oncol 2023;49:1147-53.

29. Meng C, Wang W, Zhang Y, et al. The influence of nodule size on the aggressiveness of thyroid carcinoma varies with patient's age. Gland Surg 2021;10:961-72.

30. Zhao L, Zhou T, Zhang W, et al. Blood immune indexes can predict lateral lymph node metastasis of thyroid papillary carcinoma. Front Endocrinol (Lausanne) 2022;13:995630.

31. Zhang Z, Xia F, Wang W, et al. The systemic immune-inflammation index-based model is an effective biomarker on predicting central lymph node metastasis in clinically

nodal-negative papillary thyroid carcinoma. Gland Surg 2021;10:1368-73.

32. Huang Y, Liu Y, Mo G, et al. Inflammation Markers Have Important Value in Predicting Relapse in Patients with papillary thyroid carcinoma: A Long-Term Follow-Up Retrospective Study. Cancer Control 2022;29:10732748221115236.

33. Song L, Zhu J, Li Z, et al. The prognostic value of the lymphocyte-to-monocyte ratio for high-risk papillary thyroid carcinoma. Cancer Manag Res 2019;11:8451-62.

34. Wu Y, Rao K, Liu J, et al. Machine Learning Algorithms for the Prediction of Central Lymph Node Metastasis in Patients With Papillary Thyroid Cancer. Front Endocrinol (Lausanne) 2020;11:577537.

35. Feng JW, Ye J, Qi GF, et al. LASSO-based machine learning models for the prediction of central lymph node metastasis in clinically negative patients with papillary thyroid carcinoma. Front Endocrinol (Lausanne) 2022;13:1030045.

36. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019;20:e262-73.

37. Ke G, Meng Q, Finley T, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: Guyon I, von Luxburg U, Bengio S, et al. editors. Advances in Neural Information Processing Systems. Curran Associates, Inc.; 2017.

38. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, California, USA: Association for Computing Machinery; 2016. doi: 10.1145/2939672.2939785.

39. Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. J Big Data 2020;7:94.
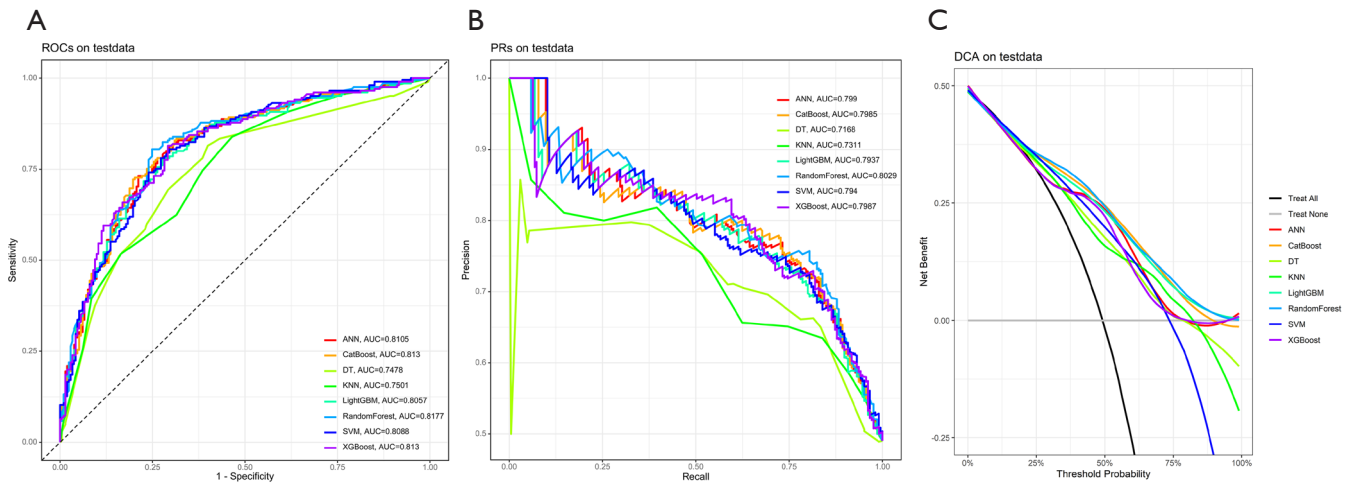
**Figure S1** Predictive performance of the eight ML algorithms in the testing set. (A) ROC curves display the predictive model discrimination based on auROC scores. (B) PR curves display the predictive model discrimination based on auPR scores. (C) Decision curve analysis revealed the net benefit of the models in terms of clinical utility. ML, machine learning; ROC, receiver operating characteristic; PR, precision-recall.
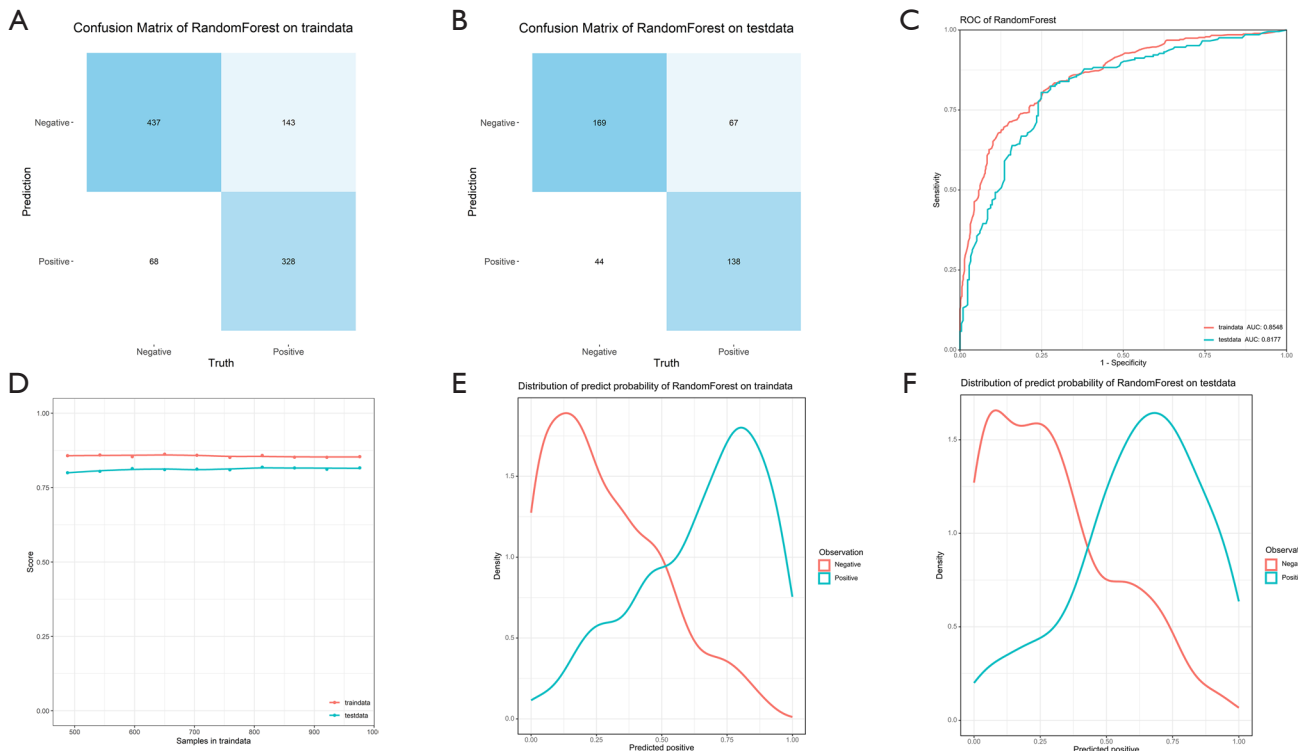


**Figure S2** RF model performance. (A,B) Confusion matrix of the RF model. (C) ROC curves of the RF model in the training and testing cohorts. (D) Learning curves of the RF model. (E,F) Predicted probability distribution curves of the RF model. RF, random forest; ROC, receiver operating characteristic.
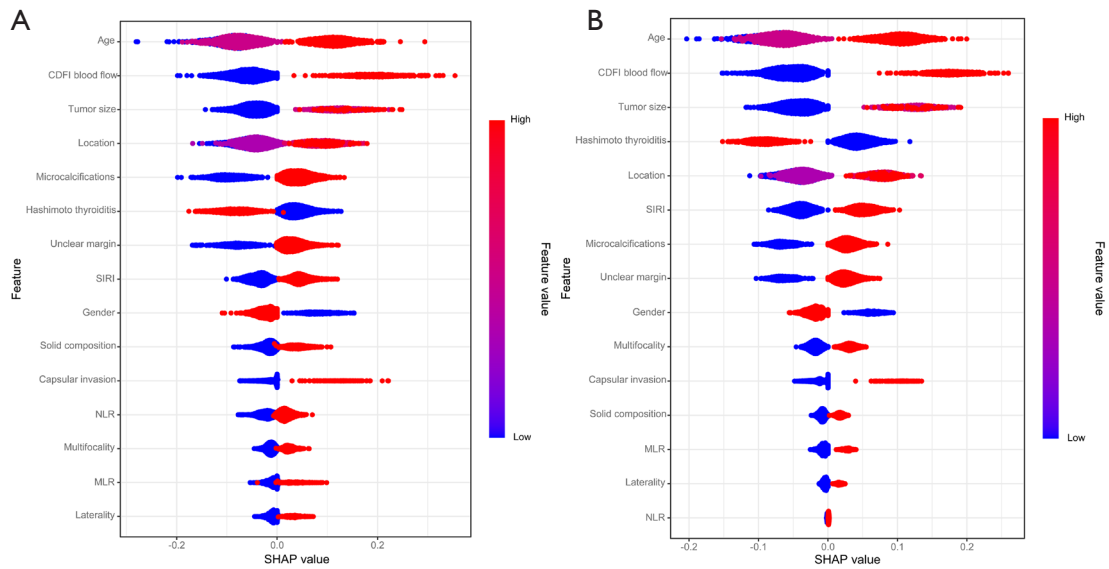
**Figure S3** SHAP summary plot's scatter plot. (A) RF model; (B) XGBoost model. RF, random forest.
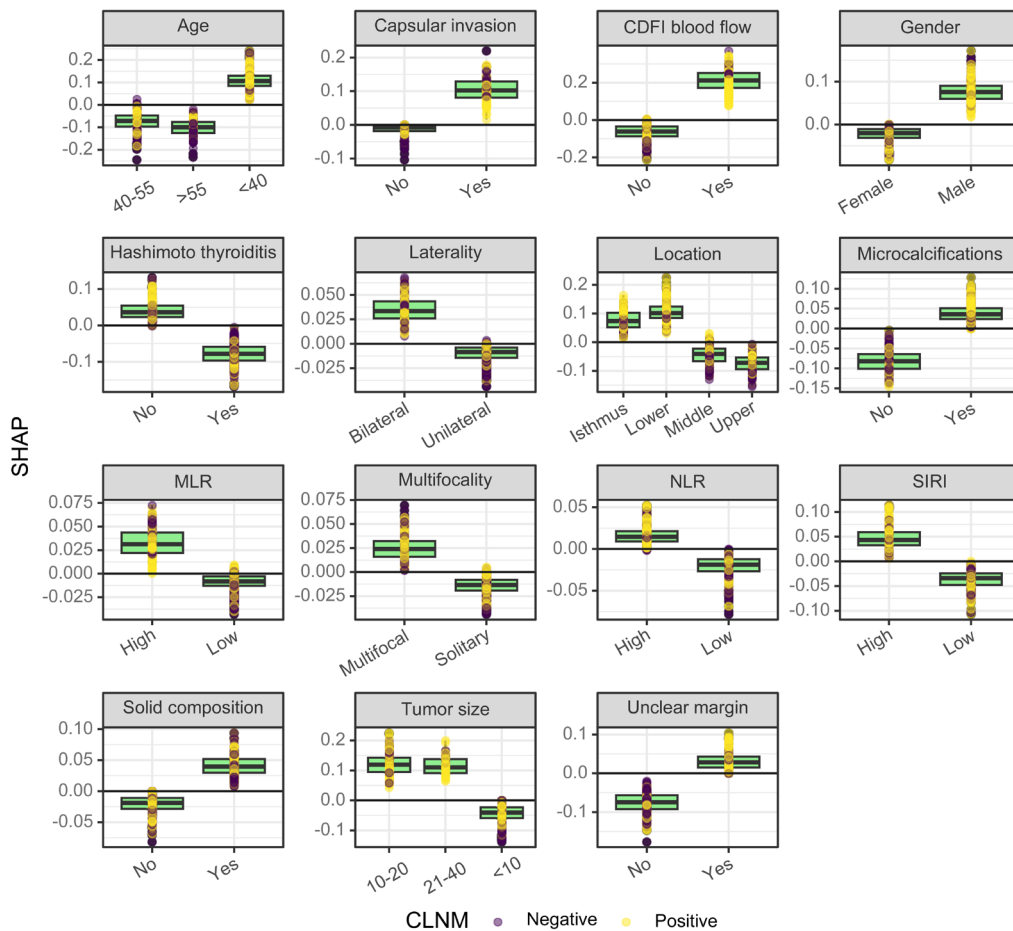


**Figure S4** The facet wrap plot based on the SHAP value in the RF model. RF, random forest; CLNM, central lymph node metastasis; CDFI, color Doppler flow imaging; MLR, monocyte-to-lymphocyte ratio; NLR, neutrophil-to-lymphocyte ratio; SIRI, systemic inflammation response index.