Name of Journal: World Journal of Gastrointestinal Oncology
Manuscript NO: 94901
Manuscript Type: ORIGINAL ARTICLE

Retrospective Study
Combining lymph node ratio to develop prognostic models for postoperative gastric neuroendocrine neoplasm patients
Abstract

BACKGROUND

Lymph node ratio (LNR) was demonstrated to play a crucial role in the prognosis of many tumors. However, research concerning the prognostic value of LNR in postoperative gastric neuroendocrine neoplasm (NEN) patients was limited.

AIM

To explore the prognostic value of LNR in postoperative gastric NEN patients and to combine LNR to develop prognostic models.

METHODS

A total of 286 patients from the Surveillance, Epidemiology, and End Results database were divided into the training set and validation set at a ratio of 8:2. 92 patients from the First Affiliated Hospital of Soochow University in China were designated as a test set. Cox regression analysis was used to explore the relationship between LNR and disease-specific survival (DSS) of gastric NEN patients. Random survival forest (RSF) algorithm and Cox proportional hazards (CoxPH) analysis were applied to develop models to predict DSS respectively, and compared with the 8th edition American Joint Committee on Cancer (AJCC) tumor-node-metastasis (TNM) staging.

RESULTS

Multivariate analyses indicated that LNR was an independent prognostic factor for postoperative gastric NEN patients and a higher LNR was accompanied by a higher risk of death. The RSF model exhibited the best performance in predicting DSS, with the C-index in the test set being 0.769 [95% confidence interval (CI): 0.691-0.846] outperforming the CoxPH model (0.744, 95%CI: 0.665-0.822) and the 8th edition AJCC TNM staging (0.723, 95%CI: 0.613-0.833). The calibration curves and decision curve analysis (DCA) demonstrated the RSF model had good calibration and clinical benefits.
Furthermore, the RSF model could perform risk stratification and individual prognosis prediction effectively.

CONCLUSION
A higher LNR indicated a lower DSS in postoperative gastric NEN patients. The RSF model outperformed the CoxPH model and the 8th edition AJCC TNM staging in the test set, showing potential in clinical practice.

Key Words: Gastric neuroendocrine neoplasm; Lymph node ratio; Disease-specific survival; Random survival forest; Predictive model

Liu W, Wu HY, Lin JX, Qu ST, Gu YJ, Zhu JZ, Xu CF. Combining lymph node ratio to develop prognostic models for postoperative gastric neuroendocrine neoplasm patients. World J Gastrointest Oncol 2024; In press

Core Tip: The prognostic value of lymph node ratio (LNR) in postoperative gastric neuroendocrine neoplasm (NEN) patients was explored in this research. A higher LNR indicated a lower disease-specific survival (DSS) in postoperative gastric NEN patients. In addition, we combined LNR to develop prognostic models to predict DSS for postoperative gastric NEN patients, using the random survival forest (RSF) algorithm and Cox proportional hazards (CoxPH) analysis. The RSF model outperformed the CoxPH model and the 8th edition American Joint Committee on Cancer tumor-node-metastasis staging in the test set. Also, the RSF model demonstrated value in risk stratification and individual prognosis prediction.

INTRODUCTION
Neuroendocrine neoplasms (NENs) are rare tumors originating from secretory cells of the diffuse endocrine system, which typically produce bioactive amines or peptide hormones\(^1\). A majority of NENs occur in the gastroenteropancreatic system, and
gastric NENs account for 6.9% to 23% of all gastroenteropancreatic NENs\textsuperscript{[2-5]}. In recent years, with the popularization of physical examination and the improvement of detection methods (computed tomography and endoscopy), the detection rate of gastric NENs has increased 7-fold to 10-fold\textsuperscript{[6]}. Based on the 2019 revision of the World Health Organization (WHO) classification criteria for digestive system tumors, gastric NENs were divided into three types: Well-differentiated neuroendocrine tumor (NET), poorly-differentiated neuroendocrine carcinoma (NEC), and mixed neuroendocrine-non-NEN. The outcomes of NENs were significantly different between gastric NET and gastric NEC. The latter presented with more aggressive behavior, a shorter survival rate, and was prone to relapse after surgery\textsuperscript{[7-9]}. Based on the 5\textsuperscript{th} edition Japanese gastric cancer treatment guidelines\textsuperscript{[10]}, standard surgical procedure combined with adjuvant chemotherapy was the common treatment strategy for gastric NEC. With the increasing detection rate of gastric NENs, more and more researchers have begun to explore the risk factors that affect the prognosis of gastric NENs, including tumor classification, tumor-node-metastasis (TNM) stage, tumor size, and so on.

Lymph node ratio (LNR) has been proven to be an important prognostic factor in most tumors\textsuperscript{[11-13]}. LNR is defined as the ratio of the number of positive lymph nodes to the number of examined lymph nodes. Due to combining the number of lymph nodes examined during the operation, LNR was more beneficial to be applied in the prognostic analysis of patients with the same number of lymph nodes involved. Many studies have demonstrated that the prognostic value of LNR was superior to the absolute number of involved lymph nodes\textsuperscript{[14,15]}. As extremely rare types of gastric neoplasm, research concerning the prognostic value of LNR in patients with gastric NENs was limited.

Artificial intelligence (AI) has undergone rapid development over the past decade and its application in medicine is currently a hot topic. Machine learning (ML), as a subgroup of AI, presented strengths in tackling larger sample sizes and high-dimensional data compared with the traditional statistical methods\textsuperscript{[16]}. Random forest (RF), as one of the ML algorithms, showed good performance in the application of
oncology because RF was suitable for the common medium-sized datasets in clinical\textsuperscript{17}. Random survival forest (RSF), as a combination of RF and traditional survival analysis, was first proposed in 2008\textsuperscript{18}. RSF can be applied to analyze right-censored survival data. Numerous researches have demonstrated that the RSF model could predict cancer prognosis effectively and outperformed the traditional Cox proportional hazards (CoxPH) model, such as in pancreatic cancer, breast cancer, cervical adenocarcinoma, and so on\textsuperscript{19-21}. Our research aims to explore the relationship between LNR and disease-specific survival (DSS) in postoperative gastric NEN patients and to establish and validate prognostic models using the RSF algorithm and CoxPH analysis.

**MATERIALS AND METHODS**

**Study population**

We obtained patients’ data from the Surveillance, Epidemiology, and End Results (SEER) database between 2000 and 2019. Cases were selected based on the primary site code (C16.0-C16.9, stomach) and third edition (ICD-O-3) histology codes (8013, large cell NEC; 8153, gastrinoma; 8240, carcinoid tumor; 8241, enterochromaffin cell carcinoid; 8242, enterochromaffin-like cell tumor; 8244, mixed adenoneuroendocrine carcinoma (MANEC); 8245, adenocarcinoid tumor; 8246, NEC; 8249, atypical carcinoid tumor; 8156, somatostatinoma). Exclusion criteria included: (1) Cases without histopathological evidence; (2) Cases with a history of other malignancies; (3) Cases without detailed clinical data information: Lack of differentiation grade, size, or TNM stage; (4) Cases without information on survival months or death within one month; (5) Cases without surgery performed or performing local surgery; and (6) Cases without information on the number of examined lymph nodes and positive lymph nodes. Finally, 286 cases with gastric NENs were included. Moreover, 92 gastric NENs patients from the First Affiliated Hospital of Soochow University were enrolled in the study as the test set. These patients were diagnosed by histopathology from 2011 to 2020 and experienced the screening process as above. We excluded patients who died within one month because these patients died due to postoperative complications rather than
reccurent diseases[22]. The concrete screening process is shown in Figure 1. The ethics committee of the First Affiliated Hospital of Soochow University approved this retrospective study (Number: 2024-145). We conducted postoperative follow-ups for our hospital patients, including outpatient reviews, inpatient medical reviews, and telephone interviews. The last follow-up time was December 2022. DSS was calculated from the date of surgery to the date of last contact or death caused by the primary tumor.

**Variables definition and selection**

The following information were identified from the SEER dataset: Sex, age at diagnosis, race, marital status, primary site, differentiation grade, tumor size, American Joint Committee on Cancer (AJCC) T stage, AJCC N stage, AJCC M stage, the number of regional nodes positive, the number of regional nodes examined, surgery for primary site, radiotherapy recode, chemotherapy recode, cause-specific death classification and survival months. LNR was defined as the ratio of positive lymph nodes to the total number of examined lymph nodes. We turned the LNR from a continuous variable into a categorical variable with the X-tile software. We classify the separated, divorced, unmarried, domestic partner, single (never married), and widowed into unmarried groups[23]. Based on the latest 2019 5th WHO classification for tumors of the digestive system, the gastroenteropancreatic NENs were divided into three types: NET, NEC, and MANEC. NET is defined as well differentiated tumors and NEC is defined as poorly differentiated tumors. In this research, we define well and moderately differentiated gastric NENs as ‘NET’, defining poorly differentiated and undifferentiated gastric NENs as ‘NEC’[24]. Considering the partial overlapping relationship between the N stage and LNR, we did not admit the N stage into the univariate analysis and the construction of the RSF model. Moreover, we restage the TNM stage in the SEER database based on the latest 8th edition of the AJCC stage. Based on the SEER Program Coding and Staging Manual 2021, the 20 to 27code in the column of ‘surgery primary site’ was defined as local tumor excision, and the 30 to 80 code was defined as open
surgery[25]. We only admitted patients undergoing open surgery, ensuring the accuracy of the results of the regional nodes examined.

**The construction of prognostic models**

We randomly split patients from the SEER database into the training set and validation set at a ratio of 8:2. Patients from our hospital were designated as a test set. The training set was applied to develop models. The validation set was used for tuning hyperparameters of the RSF model. The test set was applied to confirm the optimal predictive model.

For the CoxPH model, we first conducted univariable and multivariable regression analyses to select independent risk factors. The selected independent risk factors were then applied to develop the CoxPH model and presented in the format of a nomogram. The criteria for variable inclusion in univariable and multivariate analysis were both $P$ values less than 0.05. The scaled Schoenfeld residual test was used to check the PH assumption.

For the RSF model, we directly admitted all variables. Optuna was applied to determine the hyperparameters of the RSF model. Optuna presented advantages in adopting an effective sampling and pruning strategy to construct hyperparameter search space and could achieve good performance under limited resources[26]. For the RSF model, we explored the hyperparameters’ combination of the number of estimators (from 100 to 1000, range 10), the minimum of samples split (from 1 to 29, range 2), and the minimum of samples leaf (from 1 to 29, range 2).

**The assessment and explanation of the developed models**

We evaluated the performances of the CoxPH model and the RSF model in the test set. The C-index and areas under the receiver operating characteristic curve (AUCs) of 1-, 3-, and 5-year were used to evaluate the discrimination ability of models. The calibration curves were used to evaluate the model’s calibration. The DCA was adopted to calculate the clinical net benefit of the models.
The SHapley Additive exPlanations (SHAP) plot, which adopts a game theoretic approach, was applied to explain the RSF model. SHAP can either explain the impact of each variable on the model output as a whole or locally analyze how each feature affects the outcome of a single patient.

**The risk stratification of patients**

We calculated each patient's risk score with the optimal model. Then, we stratify patients into the low-risk group, the medium-risk group, and the high-risk group with the tool of X-tile. Kaplan-Meier survival analysis was applied to calculate survival rates between different risk groups and compared with the method of log-rank test.

**The individual prediction**

The individual prediction was composed of the survival probability curve and local SHAP plot. The survival probability curve could exhibit the survival probability at each time point. The local SHAP plot could present the contribution of each admitted variable to the outcome of the individual.

**Statistical analysis**

R software (version 4.1.0) and Python (Version 3.8.8) were applied to perform statistical analysis. We expressed the continuous variables in the format of mean standard deviation. Categorical variables were reported as numbers and percentages. Kaplan-Meier survival analysis and log-rank test were used to explore the relationship between LNR and DSS of gastric NENs. Two-tailed $P$ values less than 0.05 were considered statistically significant. The “survival” package of R software (Version 4.1.0) was used for Kaplan-Meier survival analysis. The “survminer” package of R software was used for plotting Kaplan-Meier survival curves and performing the log-rank test. The “rms” package of R software was used for CoxPH regression and calibration curves. The “dcurves” package of R software was used for plotting DCA curves. The “scikit-
survival” module in Python (Version 3.8.8) was applied to establish the RSF model. The ‘optuna’ module in Python was used to tune hyperparameters.

RESULTS

Baseline characteristics of patients

According to the inclusion and exclusion criteria, 286 patients from the SEER database and 92 patients from the First Affiliated Hospital of Soochow University were enrolled into the research. We randomly split patients from the SEER database into the training set (n = 233) and validation set (n = 53) at a ratio of 8:2. Our hospital were assigned as the test set. The detailed demographic and clinical information of these patients were summarized in Table 1. The median follow-up period in training, validation and test set were 51, 40, and 23 months respectively.

The univariable and multivariate regression analysis

We confirmed optimal cutoff value for LNR was 0.20 by X-tile tool and patients was stratified into three groups (LNR0, LNR0.2, LNR0.2). Then, we conduct univariable and multivariate regression analysis. The univariable regression analysis showed that age, primary site, histology type, size, M stage, LNR, radiation, and chemotherapy were risk factors. The multivariate regression analysis showed that primary site, histology type, size, M stage, and LNR were independent risk factors for gastric NENs. Figure 2 was the forest plot of multivariate cox regression analysis.

We found patients with a higher LNR were accompanied by a higher risk of death [LNR = 0 vs LNR0.2: odds ratio (OR) = 2.214, 95% confidence interval (CI): 1.005-4.880, P = 0.0485; LNR = 0 vs LNR: OR = 4.774, 95%CI: 2.369-9.620, P = 0.001]. Kaplan-Meier survival curves presented that DSS of different LNR groups were significant different both in the training set and test set (Figure 3A and B). The log-rank test revealed P value less than 0.05 both. The 1-, 3- and 5- years DSS of different LNR groups in the training set were observed to be 95% vs 87% vs 84%, 81% vs 57% vs 51%, 55% vs 34% vs 29%
respectively. The 1-, 3- and 5-years DSS of different LNR groups in the test set were observed to be 92% vs 77% vs 52%, 67% vs 29% vs 12%, 46% vs 17% vs 6% respectively.

The construction of prognostic models

CoxPH model: We used the above independent risk factors to develop the CoxPH model and visualized in the format of a nomogram (Supplementary Figure 1). The $\chi^2$ test of the Schoenfeld residuals demonstrated the Cox model satisfied the PH assumption, with $P$ values of all variables above 0.05 (Supplementary Figure 2).

RSF model: We treated age as a continuous variable, gender, marital status, race, primary site, histology type, size, T stage, M stage, LNR, radiotherapy, chemotherapy as categorical variables, and directly admitted them in the establishment of the RSF model. Then, we took advantage of the Optuna algorithm to confirm the optimal hyperparameter combination of RSF model: 330 estimators, 5 minimum of samples split, and 1 minimum of samples leaf. Supplementary Figure 3A depicted the optimization history, Supplementary Figure 3B depicted the hyperparameters importance, and Supplementary Figure 3C depicted the high-dimensional parameter relationships.

The performance of the developed models

CoxPH model: The C-index of the CoxPH model in the training set, validation set, and the test set were 0.834 (95% CI: 0.789-0.879), 0.871 (95% CI: 0.802-0.940), and 0.744 (95% CI: 0.665-0.822) respectively. AUCs for 1-, 3-, and 5-year DSS in the training set were 0.848 (95% CI: 0.763-0.930), 0.881 (95% CI: 0.831-0.932), and 0.875 (95% CI: 0.822-0.927). AUCs for 1-, 3-, and 5-year DSS in the validation set were 0.843 (95% CI: 0.717-0.969), 0.948 (95% CI: 0.892-1.000), and 0.990 (95% CI: 0.969-1.000). AUCs for 1-, 3-, and 5-year DSS in the test set were 0.786 (95% CI: 0.622-0.889), 0.834 (95% CI: 0.735-0.934), and 0.810 (95% CI: 0.688-0.931). The 1-, 3-, and 5-year calibration curves of the CoxPH model in the test set (Supplementary Figure 4A-C) indicated a good calibration.
**RSF model:** The C-index of the RSF model in the training set, validation set, and the test set were 0.940 (95%CI: 0.924-0.956), 0.870 (95%CI: 0.818-0.921), and 0.769 (95%CI: 0.691-0.846) respectively. AUCs for 1-, 3-, and 5-year DSS in the training set were 0.962 (95%CI: 0.938-0.989), 0.979 (95%CI: 0.963-0.995), and 0.971 (95%CI: 0.951-0.992). AUCs for 1-, 3-, and 5-year DSS in the validation set were 0.867 (95%CI: 0.761-0.973), 0.955 (95%CI: 0.899-1.000), and 0.986 (95%CI: 0.960-1.000). AUCs for 1-, 3-, and 5-year DSS in the test set were 0.803 (95%CI: 0.608-0.891), 0.895 (95%CI: 0.814-0.976), and 0.869 (95%CI: 0.769-0.970). The 1-, 3-, and 5-year calibration curves of the RSF model (Figure 4A-C) in the test set also indicated a good calibration. We summarized the performance of the CoxPH model and the RSF model in Table 2.

**Comparison between the AJCC TNM staging, the CoxPH model, and the RSF model**

We compared the performance of the 8th AJCC TNM staging, the CoxPH model, and the RSF model in the test set. The C-index of the 8th AJCC TNM staging, the CoxPH model, and the RSF model in the test set were 0.723 (95%CI: 0.613-0.833), 0.744 (95%CI: 0.665-0.822), and 0.769 (95%CI: 0.691-0.846) respectively. The RSF model owned the best C-index. Figure 5A-C depicted the receiver operating characteristic curves of 1-, 3-, and 5-year DSS in the test set. The 1-, 3-, and 5-year AUCs of the AJCC TNM staging were 0.690, 0.769, and 0.77. The 1-, 3-, and 5-year AUCs of the CoxPH model were 0.786, 0.834, and 0.810. The 1-, 3-, and 5-year AUCs of the RSF model were 0.803, 0.895, and 0.869. We could find the RSF model still owned the higher 1-, 3-, and 5-year AUCs than the other two. Furthermore, we depicted the 1-, 3-, and 5-year DCA of the above three models in the test set (Figure 6A-C). We found the RSF owned higher clinical net benefits in 3-year and 5-year.

**The interpretation for the RSF model**

We interpreted the RSF model with SHAP plot (Figure 7). A dot on the SHAP plot represents a sample. Y-axis variables are sorted according to the variables' importance.
We found histology type was thought the most valuable variable, followed by LNR, T stage, M stage and so on. The X-axis represents the influence of variables on the model output outcome. The left side of \( X = 0.0 \) represents the negative effect of variables on the outcome, and the right side of \( X = 0.0 \) represents the positive effect of variables on the outcome. Taking LNR as an example, the higher the LNR level, the worse the prognosis of patients.

**Patients' risk stratification**

We computed patients' risk scores based on the RSF model and divided them into low-risk group (risk score: 9.61), medium-risk group (9.61 risk score 26.10), and high-risk group (risk score: 26.10) with X-tile tool (Supplementary Figure 5). The Kaplan-Meier survival analysis in training set, internal validation set and test set demonstrated that the DSS between the high-risk group, medium-risk group, and low-risk group were significant different (Figure 8A-C). The log-rank test revealed \( P \) values were all less than 0.05. The 1-, 3- and 5- years DSS of different risk groups in the training set were observed to be 99% vs 87% vs 45%, 84% vs 48% vs 3%, 57% vs 20% vs 0% respectively. The 1-, 3- and 5- years DSS of different risk groups in the validation set were observed to be 96% vs 68% vs 50%, 79% vs 32% vs 0%, 64% vs 5% vs 0% respectively. The 1-, 3- and 5- years DSS of different risk groups in the test set were observed to be 94% vs 80% vs 33%, 81% vs 31% vs 0%, 62% vs 16% vs 0%, respectively.

**Individual prediction**

We randomly select three patients from the test set, and then draw the survival probability curves and local SHAP plots of the three patients respectively. Figure 9A exhibited three patients' survival probability in each time point. Figure 9B-D explained the prognosis of the three patients from each variable's contribution to the outcome in detail.
**Patient #1:** Age: 63 years, sex: Male, married: Yes, race: Asian, primary site: Middle 1/3, histology: NEC, tumor size: 4 cm, T stage: T4, M stage: M1, LNR: 0.667, radiation: No, chemotherapy: Yes. The follow-up time was 19 months. The status was dead at the deadline of follow-up.

**Patient #2:** Age: 61 years, sex: Male, married: Yes, race: Asian, primary site: Upper 1/3, histology: MANEC, tumor size: 6 cm, T stage: T3, M stage: M0, LNR: 0.167, radiation: Yes, chemotherapy: No. The follow-up time was 42 months. The status was dead at the deadline of follow-up.

**Patient #3:** Age: 76 years, sex: Male, married: Yes, race: Asian, primary site: Upper 1/3, histology: NEC, tumor size: 2.8 cm, T stage: T2, M stage: M0, LNR: 0, radiation: No, chemotherapy: No. The follow-up time was 127 months. The status was alive at the deadline of follow-up.

**DISCUSSION**

With the increasing incidence of gastric NENs, more attention was paid to the prognosis of the tumor. Multiple researches had confirmed that lymph node metastasis plays a crucial role in the prognosis of gastric NENs\(^\text{[27-29]}\). Recently, the relationship between metastatic lymph node number and LNR has been analyzed in gastric cancer\(^\text{[30]}\). Compared with the factor of lymph node metastasis, the factor of LNR is less influenced by the number of lymph nodes resected and may be more suitable for inclusion in prognosis evaluation\(^\text{[31]}\). However, researches about the influence of LNR on the prognosis of gastric NENs are still limited due to the rarity of the disease. In this research, we collected patients from the SEER dataset and the First Affiliated Hospital of Soochow University to explore the relationship between LNR and the prognosis of postoperative gastric NENs patients. Moreover, we respectively developed the CoxPH model and the RSF model to predict DSS of postoperative gastric NEN patients.
LNR has been proved to be crucial in prognosis evaluation in many cancers, such as gastric, colon, rectal, and breast cancer\textsuperscript{[32-34]}. A large retrospective single-institutional studies divide gastric cancer patients into two parts based on the 0.25 cut-off value of LNR and demonstrated that LNR was negatively associated with overall survival\textsuperscript{[32]}. The prognostic value of LNR in other sites of NENs has also been researched and demonstrated, such as lung, small intestinal, pancreas, and colon\textsuperscript{[35-38]}. We find out two articles concerning the effect of LNR on gastric NENs. However, due to the small simple sizes or lack of the external validation set, researches have no enough persuasion\textsuperscript{[31,39]}. In our study, we select the 286 patients from the SEER database as a training set and validation set, 92 patients from our hospital as an external test set. The optimal cut-off value for LNR was defined as 0.20 by using X-tile software, and patients were divided into three groups (LNR0, LNR0.2, LNR0.2). Our research shows that the DSS have significant differences in the three LNR groups, no matter in the training set and the test set. The higher the LNR, the lower the DSS rates.

Researchers had previously built various prognostic models for gastric NENs, using traditional Cox regression analysis\textsuperscript{[31,40-42]}. Cao et al\textsuperscript{[41]} develop a nomogram to predict DSS for gastric NENs, with a higher C-index compared with the traditional AJCC TNM staging (C-index: 0.899 vs 0.864). Song et al\textsuperscript{[42]} construct a nomogram to predict DSS for gastric NEC, also outperforming the traditional AJCC TNM staging. However, the significant drawback is that the Cox regression analysis is limited by the PH assumption and can only explore linear relationships between variables, even though sometimes the relationship between variables and survival outcomes is complex and non-linear\textsuperscript{[43]}.

The repaid development in AI facilitated the establishment of the required nonlinear prognostic models. The RSF model, as a novel nonlinear ML model for survival analysis, adopted an ensemble tree method to analyze censored survival data\textsuperscript{[18]}. The RSF model was no need assuming the influence of all variables on the risk function is linear and applied the internal cross-validation to avoid overfitting, ensuring the high predictive accuracy\textsuperscript{[44]}. Lin et al\textsuperscript{[39]} developed a RSF model to predict the DSS in
pancreatic cancer, outperforming the Cox regression model (C-index: 0.723 vs 0.670). Kar et al[45] also found the RSF model had better performance in predicting relapse in stage I non-small cell lung cancer patients, compared with the CoxPH model. In our research, we concluded the RSF model had favorable performance in predicting DSS of postoperative gastric NEN patients, with C-index in the external test set was 0.769 (95%CI: 0.691-0.846), superior to the CoxPH model which was 0.744 (95%CI: 0.665-0.822).

In our research, we made use of the SHAP plot to rank the importance of variables in the RSF model. We found histology, LNR and T stage ranked the top three. Compared with the well-differentiated NET, the poorly-differentiated NEC and MANEC tend to have more aggressive behavior and a worse prognosis[46]. Gastric MANEC was defined as containing gastric adenocarcinoma and NEC cells, with each proportion accounting for at least 30%[24]. Several researches had evaluated the prognosis of gastric MANEC. Fernandes et al[47] concluded that gastric MANEC prognosis is decided by the more aggressive component. Chen et al[48] concluded that a higher rate of NEC component indicating a poorer prognosis. We found LNR ranked second. This also demonstrated LNR plays a crucial role in the construction of the RSF model and evaluating patients’ prognosis. The higher the LNR, the lower the DSS of gastric NENs patients. T stage represented the infiltration depth of tumor. T stage as was an independent risk factor for gastric NENs has been demonstrated in many researches[42,49,50].

In addition, we made use of the RSF model for risk stratification and individualized survival prediction. When risk stratification is performed, it is easier for physicians to identify high-risk groups (risk score: 26.10) and give patients clinical intervention in time. For individualized survival prediction, we combined survival probability curves and local SHAP plots. The survival curves can dynamically reflect the change of patients’ survival probability over time, which are more flexible and intuitive. The local SHAP clearly shows the impact of all admitted variables on each patient’s outcome, which is easy to understand.
We are also aware of the potential limitations of the research. Firstly, our research was retrospective and may exist selection bias\textsuperscript{51}. Secondly, some important histopathological information cannot be found in SEER dataset, such as Ki-67 index, vascular invasion, neural invasion which are important for evaluating tumor prognosis\textsuperscript{52,53}. Thirdly, the training set and the internal validation set are all derived from the SEER database and lack multicenter external validation sets.

CONCLUSION

In conclusion, our research demonstrated that the LNR was negatively correlated with the prognosis of postoperative gastric NEN patients. The novel constructed RSF model presented the best performance in predicting DSS of postoperative gastric NEN patients, compared with the CoxPH model and the 8\textsuperscript{th} AJCC TNM staging. Furthermore, we performed risk stratification and individual prognosis prediction with the RSF model. The RSF algorithm showed potential in clinical practice.
<table>
<thead>
<tr>
<th>PRIMARY SOURCES</th>
<th>SIMILARITY INDEX</th>
<th>PRIMARY SOURCES DETAILS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7%</td>
<td>Lu Liu, Wen Liu, Zhenyu Jia, Yao Li, Hongyu Wu, Shuting Qu, Jinzhou Zhu, Xiaolin Liu, Chunfang Xu. &quot;Application of machine learning algorithms to predict lymph node metastasis in gastric neuroendocrine neoplasms&quot;, Heliyon, 2023. Crossref</td>
</tr>
<tr>
<td>2</td>
<td>7%</td>
<td><a href="http://www.nature.com">www.nature.com</a> Internet</td>
</tr>
<tr>
<td>3</td>
<td>7%</td>
<td>discovery.researcher.life Internet</td>
</tr>
<tr>
<td>5</td>
<td>7%</td>
<td><a href="http://www.china-oncology.com">www.china-oncology.com</a> Internet</td>
</tr>
<tr>
<td>6</td>
<td>7%</td>
<td>bmccancer.biomedcentral.com Internet</td>
</tr>
<tr>
<td>7</td>
<td>7%</td>
<td><a href="http://www.frontiersin.org">www.frontiersin.org</a> Internet</td>
</tr>
<tr>
<td>Feature</td>
<td>Setting</td>
<td>Description</td>
</tr>
<tr>
<td>------------------------</td>
<td>---------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Exclude Quotes</td>
<td>ON</td>
<td></td>
</tr>
<tr>
<td>Exclude Bibliography</td>
<td>ON</td>
<td></td>
</tr>
<tr>
<td>Exclude Sources</td>
<td>&lt; 15 Words</td>
<td></td>
</tr>
<tr>
<td>Exclude Matches</td>
<td>&lt; 15 Words</td>
<td></td>
</tr>
</tbody>
</table>