



OPEN ACCESS

EDITED BY

Dimitrios Schizas,
National and Kapodistrian University of
Athens, Greece

REVIEWED BY

Xue-Feng Leng,
University of Electronic Science and
Technology of China, China
Bin Zheng,
Fujian Medical University, China

*CORRESPONDENCE

Hai-ning Zhou
haining_zhou@zmu.edu.cn
Dong Tian
TianD_EATTS@nsmc.edu.cn

[†]These authors have contributed
equally to this work and share
first authorship

SPECIALTY SECTION

This article was submitted to
Surgical Oncology,
a section of the journal
Frontiers in Oncology

RECEIVED 05 July 2022

ACCEPTED 17 August 2022

PUBLISHED 08 September 2022

CITATION

Li D, Zhang L, Yan H, Zheng YB,
Guo X, Tang S, Hu H, Yan H, Qin C,
Zhang J, Guo H, Zhou H and Tian D
(2022) Machine learning models
predict lymph node metastasis in
patients with stage T1-T2 esophageal
squamous cell carcinoma.
Front. Oncol. 12:986358.
doi: 10.3389/fonc.2022.986358

COPYRIGHT

© 2022 Li, Zhang, Yan, Zheng, Guo,
Tang, Hu, Yan, Qin, Zhang, Guo, Zhou
and Tian. This is an open-access article
distributed under the terms of the
Creative Commons Attribution License
(CC BY). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Machine learning models predict lymph node metastasis in patients with stage T1-T2 esophageal squamous cell carcinoma

Dong-lin Li^{1†}, Lin Zhang^{2†}, Hao-ji Yan^{3,4†}, Yin-bin Zheng⁵
, Xiao-guang Guo⁶, Sheng-jie Tang¹, Hai-yang Hu¹,
Hang Yan¹, Chao Qin¹, Jun Zhang¹,
Hai-yang Guo¹, Hai-ning Zhou^{1*} and Dong Tian^{2,3,4*}

¹Department of Thoracic Surgery, Suining Central Hospital, Suining, China, ²Department of Thoracic Surgery, Affiliated Hospital of North Sichuan Medical College, Nanchong, China, ³Department of Thoracic Surgery, West China Hospital, Sichuan University, Chengdu, China, ⁴Academician (Expert) Workstation, Affiliated Hospital of North Sichuan Medical College, Nanchong, China, ⁵Department of Thoracic Surgery, Nanchong Central Hospital, Nanchong, China, ⁶Department of Pathology, Nanchong Central Hospital, Nanchong, China

Background: For patients with stage T1-T2 esophageal squamous cell carcinoma (ESCC), accurately predicting lymph node metastasis (LNM) remains challenging. We aimed to investigate the performance of machine learning (ML) models for predicting LNM in patients with stage T1-T2 ESCC.

Methods: Patients with T1-T2 ESCC at three centers between January 2014 and December 2019 were included in this retrospective study and divided into training and external test sets. All patients underwent esophagectomy and were pathologically examined to determine the LNM status. Thirty-six ML models were developed using six modeling algorithms and six feature selection techniques. The optimal model was determined by the bootstrap method. An external test set was used to further assess the model's generalizability and effectiveness. To evaluate prediction performance, the area under the receiver operating characteristic curve (AUC) was applied.

Results: Of the 1097 included patients, 294 (26.8%) had LNM. The ML models based on clinical features showed good predictive performance for LNM status, with a median bootstrapped AUC of 0.659 (range: 0.592, 0.715). The optimal model using the naive Bayes algorithm with feature selection by determination coefficient had the highest AUC of 0.715 (95% CI: 0.671, 0.763). In the external test set, the optimal ML model achieved an AUC of 0.752 (95% CI: 0.674, 0.829), which was superior to that of T stage (0.624, 95% CI: 0.547, 0.701).

Conclusions: ML models provide good LNM prediction value for stage T1-T2 ESCC patients, and the naive Bayes algorithm with feature selection by determination coefficient performed best.

KEYWORDS

esophageal squamous cell carcinoma, machine learning, lymph node metastasis, predictive model, stage T1-T2

Introduction

Esophageal cancer (EC) ranks seventh in annual incidence and sixth in mortality globally, with half of the cases occurring in China, and esophageal squamous cell carcinoma (ESCC) is the predominant histopathological type in Asian populations (1–4). Although esophagectomy with lymphadenectomy remains the gold standard (5–7), endoscopic mucosal resection (EMR) and endoscopic submucosal dissection (ESD) represent new treatment options for early ESCC (5, 8, 9). However, regional lymph node metastasis (LNM) is not uncommon in stage T1-T2 ESCC patients, with a reported occurrence rate ranging from 12.9% to 49.1% (10–14). Controversy still exists about the treatment of early ESCC patients without clinical nodal involvement. For patients with stage T1a ESCC, the decision to perform EMR/ESD is typically influenced by the depth of invasion (DOI), which is associated with the LNM risk (8, 9, 11, 15). Stage T1b-T2 patients with LNM have worse outcomes than those with negative lymph nodes (5–7, 11, 16, 17). Given that LNM negatively impacts survival and prognosis (11, 18, 19), the accurate identification of LNM is highly important to guide a surgeon's decision about the implementation of endoscopic procedures, surgery, and the subsequent treatment of early-stage ESCC.

Currently, several examination methods are used for preoperative lymph node staging for ESCC. The ability of computed tomography (CT) is unsatisfactory in identifying LNM, with a reported sensitivity, specificity, and accuracy of 39.7%, 77.3%, and 54.5%, respectively (20). Although positron emission tomography-CT (PET-CT) can reliably identify the metastatic lymph nodes that are not enlarged in size, its low sensitivity and high cost remain a concern (21). Endoscopic ultrasonography (EUS) and endobronchial ultrasound (EBUS) show excellent sensitivity but their specificity remains controversial (22, 23). Lymph node biopsy, including EBUS-guided transbronchial needle aspiration (EBUS-TBNA) and EUS-fine-needle aspiration (FNA), may confirm the status of lymph nodes; however, the invasive procedure and post-puncture hematoma limit their wider applications (23, 24). Thus, for preoperatively estimating LNM status, an efficient

and precise noninvasive diagnostic approach that is clinically relevant and generalizable is urgently needed.

Machine learning (ML) in artificial intelligence has emerged as a less costly and noninvasive approach to precision medicine in ESCC. In medical research, ML has proven to be an area of interest with many applications, where an acceptable generalization can be attained by using different algorithms and techniques to search an n-dimensional space for a set of medical samples (25). High-dimensional clinical features that are available before and after surgical extirpation of the primary tumor provide a deeper understanding of the LNM that is imperceptible to human eyes. Given the adverse effect of LNM on survival, any decision should be made after careful and accurate preoperative assessment. In this setting, optimizing negative predictive value (NPV), with a focus on minimizing false-negative results, is one of the major objectives of predictive models. ML algorithms can fit fairly complex multinomial interactions or nonlinear relationships, and the resulting predictive accuracy is impressive (25, 26). It has been shown that ML algorithms based on clinical features can identify LNM in other carcinomas (27, 28), without requiring access to and complex preprocessing of imaging data. The current methods for predicting LNM in early ESCC are mainly based on multivariate analysis of clinicopathological characteristics, and lymph node morphology on imaging. However, the literature on clinical feature-based ML prediction models for LNM of T1-2 ESCC is limited. The aim of this study was to develop and externally test ML predictive models for identifying LNM in early-T-stage patients by utilizing clinical features.

Methods

Study design and patients

The clinical variables of patients with early-T-stage ESCC were retrospectively collected from three centers (Nanchong Central Hospital, Affiliated Hospital of North Sichuan Medical College, and Suining Central Hospital) between January 2014 and December 2019. The study was registered in the Chinese Clinical Trial Registry

(ChiCTR2100051728) and approved by the relevant review boards (Nanchong Central Hospital: 2019-041, Affiliated Hospital of North Sichuan Medical College: 2020ER181-1, Suining Central Hospital: LLSNCH20200027). The informed consent requirement was waived since retrospective, deidentified data were used. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) was followed in the present study (29).

Patients with stage T1-T2 ESCC who had no clinical signs of nodal involvement (cN0) and underwent esophagectomy were identified. The inclusion criteria were defined as follows: (1) patients with primary ESCC; (2) patients aged ≥ 18 years; (3) McKeown or Ivor Lewis esophagectomy and lymphadenectomy were performed; and (4) pathologically confirmed stage T1 or T2. The following exclusion criteria were applied: (1) multiple primary tumors; (2) neoadjuvant therapy administered prior to surgery; (3) diagnosis of distant metastasis; (4) lymph node examination < 5 ; and (5) unknown lymph node dissection information. Figure 1 displays a flow chart of participants included and excluded from the overall study.

Predictor variables

The characteristics included clinical variables (sex, age, body mass index [BMI], history of surgery, tumor location, and preoperative comorbidities), preoperative hematologic indices (leukocytes, neutrophils, lymphocytes, erythrocytes, hemoglobin, aspartate, alanine, total protein, albumin, globulin, high-density lipoprotein, low-density lipoprotein, lactate, urea, creatinine and glucose), and pathological variables (endoscopic tumor length, tumor size, tumor differentiation and TNM8 T-stage).

Construction of ML models

Patients from Nanchong Central Hospital and Affiliated Hospital of North Sichuan Medical College were analyzed as the training set. Accounting for potential variation across different research institutes, patients from the Suining Central Hospital were designated as the external test set.

The data analysis involved six feature-selection methods, including random forest (RF), Boruta, least absolute shrinkage and selection operator (LASSO), determination coefficient (DC), relief and recursive feature elimination (RFE). The ML algorithms we evaluated included support vector machine (SVM), generalized boosted regression modeling (GBRM), k-nearest neighbors (KNN), naive Bayes (NB), RF and extreme gradient boosting machine (XGB). These feature selection methods and ML algorithms were common methods, which were introduced in previous studies (27, 30). A total of 36 ML models were developed using the six modeling algorithms and six feature selection techniques for predicting LNM. To validate performance, the bootstrap method was applied with 1,000 repetitions as described in previous literature (31). The area under the receiver operating characteristic (ROC) curve (AUC) was used to assess each model's overall performance, and a bootstrap resampling methodology was used to assess 95% confidence intervals [CIs]. The sensitivity, specificity, NPV and positive predictive value (PPV) were also evaluated. The best-performing model in the training set was chosen as the final model to predict LNM for the external test set. The AUC of T stage was calculated as a benchmark for the optimal prediction model. Furthermore, to assess the ability of the model to discriminate LNM in patients with different T stages, we conducted a performance evaluation of the optimal model separately for stage T1 and T2 patients in the external test set.

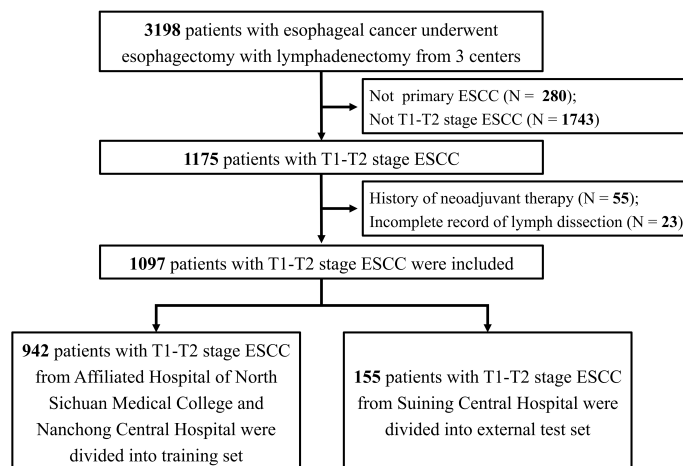


FIGURE 1

The flow chat for patient inclusion and exclusion. ESCC: esophageal squamous cell carcinoma.

Statistical analysis

R software version 3.63 was used for the statistical analysis and modeling process. The mean \pm standard deviation was used to represent quantitative variables, while the number and percentage were applied to represent categorical variables. The performance of the combined model was evaluated using ROC analysis and AUC calculation.

Results

Patient characteristics

In total, 1097 patients were included in our current study. Of these patients, the training set included 942 (85.9%) patients (median age, 65 [41-85] years), and the external test set included 155 (14.1%) patients (median age, 64 [40-80] years). The patient clinical characteristics are summarized in Table 1. A total of 294 patients (26.8%) had LNM by final histopathology, including 233 (24.7%) and 61 (39.4%) in the training and external test sets, respectively. The average endoscopic tumor length was 3.8 ± 2.0

cm and 3.7 ± 1.9 cm in the training and external test sets, respectively. The average tumor sizes were 2.8 ± 1.4 cm and 3.3 ± 1.6 cm in the training and external test sets, respectively. In stage T1 and T2 ESCC, the total LN metastasis rates were 16.4% (85/519) and 36.2% (209/578), respectively.

Predictive performance of machine learning models

Supervised ML models were trained using patient characteristics to identify patients with LNM. After data preprocessing, 6 discrete features and 20 continuous features in total, listed in Table 1 and Supplementary Table 1, were used for ML modeling. Figure 2 shows the AUC of each machine learning algorithm (columns) with each feature selection method (rows) in the form of heatmaps. The ML models based on clinical features showed good predictive performance for LNM status, with a median bootstrapped AUC of 0.659 (range: 0.592, 0.715). The NB model using feature selection by determination coefficient exhibited the highest AUC of 0.715 (95% CI: 0.671, 0.763) among all ML models.

TABLE 1 Clinical characteristics of patients with T1-T2 stage esophageal squamous cell carcinoma.

Characteristics	Training set	External test set
Age (year) (median [range])	65 (41-85)	64 (40-80)
Sex		
Male	665 (69.5%)	117 (75.5%)
Female	287 (30.5%)	38 (24.5%)
BMI	22.9 ± 4.1	21.7 ± 2.5
History of surgery		
Yes	290 (30.8%)	40 (25.8%)
No	652 (69.2%)	115 (74.2%)
Tumor location		
Upper	131 (13.9%)	21 (13.5%)
Middle	606 (64.3%)	78 (50.3%)
Lower	205 (21.8%)	56 (36.2%)
Preoperative complications		
Yes	525 (55.7%)	114 (73.5%)
No	417 (44.3%)	41 (26.5%)
Endoscopic tumor length (cm)	3.8 ± 2.0	3.7 ± 1.9
Tumor size (cm)	2.8 ± 1.4	3.3 ± 1.6
Tumor differentiation		
G1	322 (34.2%)	28 (18.1%)
G2	530 (56.3%)	95 (61.3%)
G3	90 (9.5%)	32 (20.6%)
T stage [†]		
T1a	154 (16.4%)	13 (8.4%)
T1b	294 (31.2%)	58 (37.4%)
T2	494 (52.4%)	84 (54.2%)

BMI, body mass index. [†]The 8th edition of the UICC and AJCC cancer staging system.

External test for the optimal machine learning model

Table 2 and Figure 3 show the results of the optimal model on the external test set. The AUC of the optimal model was 0.787 (95% CI: 0.674, 0.829), which outperformed that of T stage (AUC, 0.624, 95% CI: 0.547, 0.701). The sensitivity, specificity, NPV, and PPV of the optimal model were 78.7%, 63.8%, 82.2% and 58.5%, respectively, which were superior to those of T stage (67.2%, 54.3%, 71.8% and 48.8%, respectively). In both the training and external tests, the performance of the NB model was consistent. We also tested the NB model separately for stages T1 and T2, and the prediction performance was consistent and even better in stage T1. Figure 4 shows the relative distance of each patient from the decision threshold of the NB model, as determined by their classification probability. The predicted value of the NB model could obviously distinguish the different LNM outcomes of patients with stage T1-T2, stage T1, and stage T2 ESCC.

Discussion

In this multicenter study, with routinely available clinical data, we developed and validated thirty-six ML models for

predicting LNM in stage T1-T2 ESCC patients and revealed 2 main findings. First, the optimal NB model performed well in discriminating LNM with an AUC of 0.715 in the training set and demonstrated similar discrimination on the external test set (AUC 0.753). Second, the predicted value of the NB model could obviously distinguish the different lymph node statuses for patients with stage T1-T2, T1, and T2 ESCC. These novel findings suggest that a clinical feature-based ML model has the potential to be a more effective noninvasive method for identifying LNM in stage T1-T2 ESCC patients.

Lymph node (LN) status is the most important independent prognostic factor in ESCC (11, 32, 33). At present, preoperative assessment of LNM in patients with ESCC is primarily based on CT images using LN size criteria. However, several previous studies showed unsatisfactory discrimination (20, 34). Although EUS, PET-CT, and lymph node biopsy have shown varying degrees of recognition capacity (21–24), high cost and invasive procedures remain a concern. Thus, an efficient and precise noninvasive diagnostic approach that is clinically relevant and generalizable is urgently needed.

The present study showed the feasibility of the ML models for predicting LNM in stage T1-T2 ESCC patients. Traditional logistic regression (LR) analysis showed that tumor length, tumor size, tumor location, T1 substage, differentiation, lymphovascular invasion(LVI), depth of tumor invasion, and macroscopic type

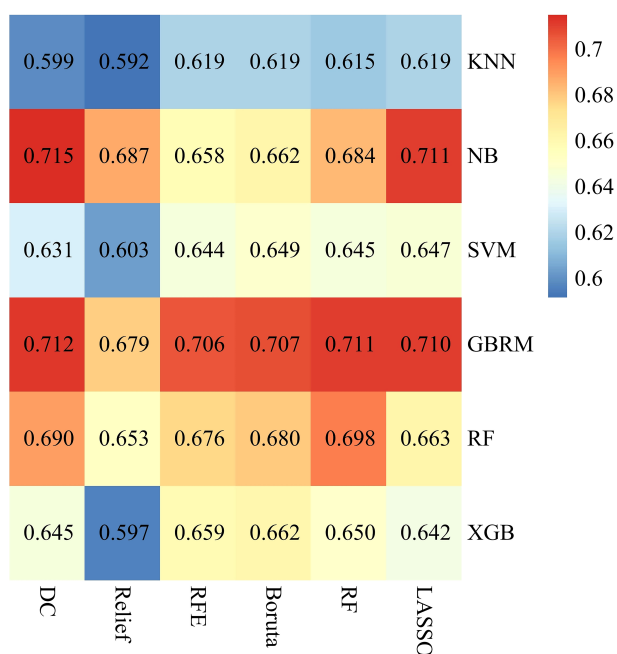


FIGURE 2 Performance of 36 machine learning models. This Heatmap showed the area under the curve of each machine learning algorithm (columns) with each feature selection method (rows). KNN, k-nearest neighbours; NB, naive-bayes; SVM, support vector machine; GBRM, generalized boosted regression modeling; RF, random forest; XGB, extreme gradient boosting machine; LASSO, least absolute shrinkage and selection operator; RFE, recursive feature elimination; DC, determination coefficient.

TABLE 2 Performance of the optimal machine learning model and the T stage.

Model/factor	External validation data	AUC (95% CI)	Sensitivity	Specificity	NPV	PPV
NB model	T1-T2 stage ESCC	0.752 (0.674-0.829)	0.787	0.638	0.822	0.585
T stage	T1-T2 stage ESCC	0.624 (0.547-0.701)	0.672	0.543	0.718	0.488
NB model	T1 stage ESCC	0.789 (0.669-0.901)	0.700	0.765	0.867	0.538
NB model	T2 stage ESCC	0.704 (0.590-0.818)	0.732	0.628	0.711	0.652

NB, Naive Bayes; ESCC, esophageal squamous cell carcinoma; AUC, the area under the curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

were associated with LNM occurrence (11, 14, 35–39). Multivariate analysis demonstrated that poor differentiation, LVI, depth of tumor invasion, T1 substage, high-density lipoprotein cholesterol (HDL-C) level, and preoperative alanine aminotransferase/aspartate aminotransferase ratio (LSR) were significant independent risk factors for LNM (11, 14, 35, 37–39). Hence, the twenty-six routinely available features extracted from patients, including clinical variables, hematological indicators and pathological variables, are feasible for predicting LNM. In most cases, prediction models have been developed based on input features considered significant by clinicians (14, 35, 40). Through human assumptions, this approach may limit the choice of input features and result in biasing. Selecting a large set of input features and using ML models to select those that perform the best may mitigate the issue to some extent (30). Some of these may not correlate with those deemed to be most important by clinical professionals and may highlight features that were previously not considered. Our models may prove useful in implementing personalized treatment stratification and close surveillance in the future by using extensive clinical features.

The literature on ML models for LNM prediction in early ESCC constructed using readily available clinical data is limited. An artificial neural network (ANN) based on clinical features was built for superficial esophageal squamous cell carcinoma (SESCC), and its ability to predict LNM was compared with that of a traditional LR model (41). The ANN model outperformed the LR model with respect to the AUC, specificity, PPV, and accuracy (41). It may be a valuable tool, especially for determining the need for additional treatment after ESD procedures. However, only one ML algorithm and feature selection method were used in their work. In our study, we compared multiple feature selection methods and ML algorithms and determined the optimal model using the bootstrap method. NB is a simple but powerful classification method widely used within ML technique. This probabilistic classifier has been proven to be highly professional and based on solid mathematical principles, with the advantages of fast predictions, adaptability to different numbers of datasets, and quickly updates as new training data becomes available (42). In addition, a concern demonstrated in their work is the lack of

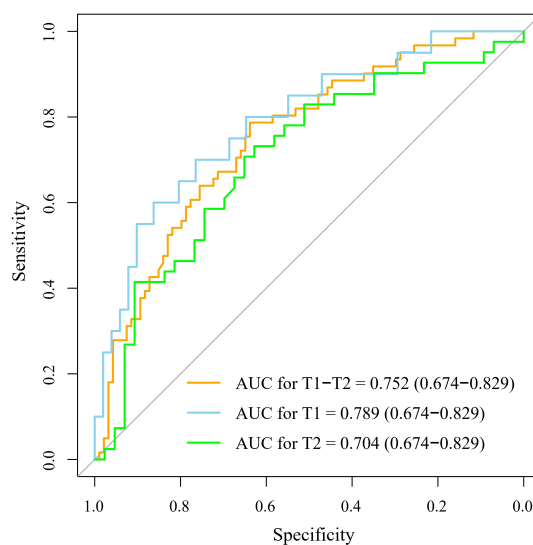


FIGURE 3

The receiver operator characteristic curve of the optimal machine learning model. The optimal machine learning model exhibited a good performance to predict the LNM for patients with T1-T2, T1, and T2 stage esophageal squamous cell carcinoma. The 95% confidence intervals were showed in the parentheses. AUC, the area under the curve.

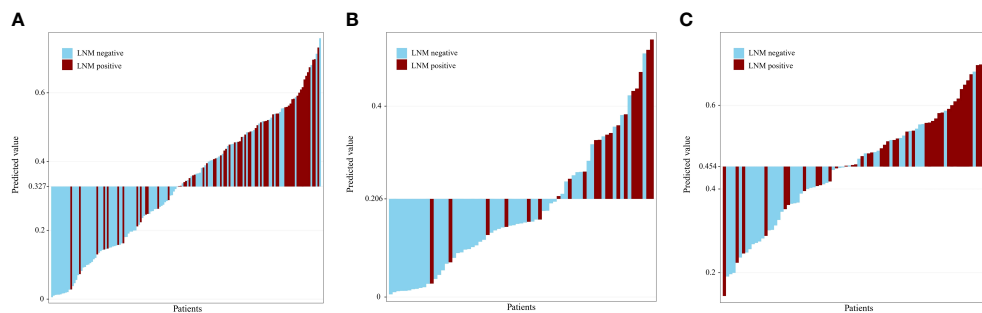


FIGURE 4

Predicted value for patients with T1-T2 (A), T1 (B), T2 (C) stage esophageal squamous cell carcinoma. The predicted value of the naive bayes model could obviously distinguish the different lymph node statuses for patients with T1-T2, T1, and T2 stage esophageal squamous cell carcinoma. LNM, lymph node metastasis.

external testing regarding the predictive performance of their models. In this study, we confirmed the performance of our optimal NB model in an independent external test dataset. The results were consistent in both the training and external test sets, suggesting that the model was not overfitted. Building on this work, using multiple ML algorithms with feature selection methods for our specific dataset is a robust approach to select the most suitable model.

T stage is an independent risk factor for LNM of ESCC (43). The rate of LNM differs substantially between different T stages. In this study, the LNM rates of stage T1 and T2 were 16.4% and 36.2%, respectively, which was consistent with the reported incidence (10, 35, 37, 44). To distinguish the ability of the model to discriminate LNM in patients with T1 and T2 stages, we tested the optimal model on different T stages of the external test set. The optimal NB model exhibited good performance in predicting LNM for patients with T1-T2, T1, and T2 stage ESCC. In addition, we constructed grayscale histograms to distinguish benign and malignant LNs in different T stages. The predicted value of the NB model could obviously distinguish the different lymph node statuses for patients with T1-T2, T1, and T2 stage ESCC. The current study showed that a stable classification model with a similar AUC value in stages T1-T2, T1, and T2 is useful to differentiate metastatic from nonmetastatic lymph nodes.

Other strengths of this study include its large sample size and multicenter design. Models were developed utilizing readily available clinical data without complicated preprocessing of imaging data. Various features were examined, including patient demographics, physical fitness, and tumor characteristics. In low-resource settings, this approach to modeling using local clinical datasets could be replicated across health systems to benefit the treatment stratification and subsequent surveillance of patients with early-stage ESCC at risk of LNM. Furthermore, to demonstrate that clinical

feature-based models have real benefits, they should be compared to more advanced models based on clinical features. Our methodology could provide the foundation for such models.

The limitations of the present study include three aspects. First, it is a retrospective design article. A prospective external test of the NB model in a large cohort population would be necessary for generalizing the findings. Second, the amount of dataset information is inadequate. Several features were omitted. Commonly, a larger amount of data will improve the confidence and performance of our model. Third, considering the clinical features alone in this study, the number of the models established was insufficient, and the AUCs were not high enough. An integrated radiomics analysis of the primary tumor and lymph nodes could potentially improve prediction performance.

Conclusions

In this study, we developed prediction models for identifying LNM in stage T1-T2 ESCC by comparing multiple ML algorithms and feature selection methods. These models achieved reasonable prediction performance. The optimal NB model demonstrated similar discrimination in the training and external test sets. Although advanced models may surpass this approach, the use of routinely available clinical data can be beneficial. Validated and externally tested, this robust and ready-to-use ML model sets the stage for future clinical trials involving the risk stratification of LNM for early ESCC.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by The Ethics Committee of Nanchong Central Hospital, Affiliated Hospital of North Sichuan Medical College, and Suining Central Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

DL and LZ: methodology, data collection, data analysis, and original draft. H-JY: methodology, statistical analysis, data visualization, and manuscript editing. YZ, XG, ST, HH, and HY: data collection and manuscript editing. CQ, JZ, and HG: data collection and manuscript editing. HZ and DT: conceptualization, project administration, and manuscript editing. All authors had access to the data and reviewed the manuscript.

References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, 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(3):209–49. doi: 10.3322/caac.21660
- Fitzmaurice C, Abate D, Abbasi N, Abbastabar H, Abd-Allah F, Abdel-Rahman O, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2017: A systematic analysis for the global burden of disease study. *JAMA Oncol* (2019) 5(12):1749–68. doi: 10.1001/jamaoncol.2019.2996
- Agrawal N, Jiao Y, Bettgowda C, Hutfless SM, Wang Y, David S, et al. Comparative genomic analysis of esophageal adenocarcinoma and squamous cell carcinoma. *Cancer Discov* (2012) 2(10):899–905. doi: 10.1158/2159-8290.CD-12-0189
- Wang W, He X, Zheng Z, Ma X, Hu X, Wu D, et al. Serum HOTAIR as a novel diagnostic biomarker for esophageal squamous cell carcinoma. *Mol Cancer* (2017) 16(1):75. doi: 10.1186/s12943-017-0643-6
- Ajani JA, D'Amico TA, Bentrem DJ, Chao J, Corvera C, Das P, et al. Esophageal and esophagogastric junction cancers, version 2.2019, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw* (2019) 17(7):855–83. doi: 10.6004/jnccn.2019.0033
- Naftoux P, Depypere L, Van Veer H, Coosemans W, Lerut T. Principles of esophageal cancer surgery, including surgical approaches and optimal node dissection (2-3-field). *Ann Cardiothorac Surg* (2017) 6(2):152–8. doi: 10.21037/acs.2017.03.04
- Wang H, Tang H, Fang Y, Tan L, Yin J, Shen Y, et al. Morbidity and mortality of patients who underwent minimally invasive esophagectomy after neoadjuvant chemoradiotherapy vs neoadjuvant chemotherapy for locally advanced esophageal squamous cell carcinoma: A randomized clinical trial. *JAMA Surg* (2021) 156(5):444–51. doi: 10.1001/jamasurg.2021.0133
- Berger A, Rahmi G, Perrod G, Pioche M, Canard J-M, Cesbron-Métivier E, et al. Long-term follow-up after endoscopic resection for superficial esophageal squamous cell carcinoma: a multicenter Western study. *Endoscopy* (2019) 51(4):298–306. doi: 10.1055/a-0732-5317
- Qi Z-P, Chen T, Li B, Ren Z, Yao L-Q, Shi Q, et al. Endoscopic submucosal dissection for early esophageal cancer in elderly patients with relative indications for endoscopic treatment. *Endoscopy* (2018) 50(9):839–45. doi: 10.1055/a-0577-2560
- Jia R, Luan Q, Wang J, Hou D, Zhao S. Analysis of predictors for lymph node metastasis in patients with superficial esophageal carcinoma. *Gastroenterol Res Pract* (2016) 2016:3797615. doi: 10.1155/2016/3797615

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2022.986358/full#supplementary-material>

- Ancona E, Rampado S, Cassaro M, Battaglia G, Ruol A, Castoro C, et al. Prediction of lymph node status in superficial esophageal carcinoma. *Ann Surg Oncol* (2008) 15(11):3278–88. doi: 10.1245/s10434-008-0065-1
- Chen J, Liu S, Pan J, Zheng X, Zhu K, Zhu J, et al. The pattern and prevalence of lymphatic spread in thoracic oesophageal squamous cell carcinoma. *Eur J Cardiothorac Surg* (2009) 36(3):480–6. doi: 10.1016/j.ejcts.2009.03.056
- Jiang K, Huang H, Chen W, Yan H, Wei Z, Wang X, et al. Risk factors for lymph node metastasis in T1 esophageal squamous cell carcinoma: A systematic review and meta-analysis. *World J Gastroenterol* (2021) 27(8):737–50. doi: 10.3748/wjg.v27.i8.737
- Tian D, Jiang K-Y, Huang H, Jian S-H, Zheng Y-B, Guo X-G, et al. Clinical nomogram for lymph node metastasis in pathological T1 esophageal squamous cell carcinoma: a multicenter retrospective study. *Ann Transl Med* (2020) 8(6):292. doi: 10.21037/atm.2020.02.185
- Aoyama J, Kawakubo H, Mayanagi S, Fukuda K, Irino T, Nakamura R, et al. Discrepancy between the clinical and final pathological findings of lymph node metastasis in superficial esophageal cancer. *Ann Surg Oncol* (2019) 26(9):2874–81. doi: 10.1245/s10434-019-07498-2
- Yang H, Liu H, Chen Y, Zhu C, Fang W, Yu Z, et al. Neoadjuvant chemoradiotherapy followed by surgery versus surgery alone for locally advanced squamous cell carcinoma of the esophagus (NEOCRTEC5010): A phase III multicenter, randomized, open-label clinical trial. *J Clin Oncol* (2018) 36(27):2796–803. doi: 10.1200/JCO.2018.79.1483
- Shapiro J, van Lanschot JJB, Hulshof MCCM, van Hagen P, van Berge Henegouwen MI, Wijnhoven BPL, et al. Neoadjuvant chemoradiotherapy plus surgery versus surgery alone for oesophageal or junctional cancer (CROSS): long-term results of a randomised controlled trial. *Lancet Oncol* (2015) 16(9):1090–8. doi: 10.1016/S1470-2045(15)00040-6
- Wang J, Wu N, Zheng Q-F, Yan S, Lv C, Li S-L, et al. Evaluation of the 7th edition of the TNM classification in patients with resected esophageal squamous cell carcinoma. *World J Gastroenterol* (2014) 20(48):18397–403. doi: 10.3748/wjg.v20.i48.18397
- Law S, Kwong DLW, Kwok K-F, Wong K-H, Chu K-M, Sham JST, et al. Improvement in treatment results and long-term survival of patients with esophageal cancer: impact of chemoradiation and change in treatment strategy. *Ann Surg* (2003) 238(3):339–48. doi: 10.1097/01.sla.0000086545.45918.ee
- Foley KG, Christian A, Fielding P, Lewis WG, Roberts SA. Accuracy of contemporary oesophageal cancer lymph node staging with radiological-pathological correlation. *Clin Radiol* (2017) 72(8):693.e1–e7. doi: 10.1016/j.crad.2017.02.022

21. Okada M, Murakami T, Kumano S, Kuwabara M, Shimono T, Hosono M, et al. Integrated FDG-PET/CT compared with intravenous contrast-enhanced CT for evaluation of metastatic regional lymph nodes in patients with resectable early stage esophageal cancer. *Ann Nucl Med* (2009) 23(1):73–80. doi: 10.1007/s12149-008-0209-1
22. Shan H-B, Zhang R, Li Y, Gao X-Y, Lin S-Y, Luo G-Y, et al. Application of endobronchial ultrasonography for the preoperative detecting recurrent laryngeal nerve lymph node metastasis of esophageal cancer. *PLoS One* (2015) 10(9): e0137400. doi: 10.1371/journal.pone.0137400
23. Fu X, Wang F, Su X, Luo G, Lin P, Rong T, et al. Endobronchial ultrasound improves evaluation of recurrent laryngeal nerve lymph nodes in esophageal squamous cell carcinoma patients. *Ann Surg Oncol* (2021) 28(7):3930–8. doi: 10.1245/s10434-020-09241-8
24. Vazquez-Sequeiros E, Norton ID, Clain JE, Wang KK, Affi A, Allen M, et al. Impact of EUS-guided fine-needle aspiration on lymph node staging in patients with esophageal carcinoma. *Gastrointest Endosc* (2001) 53(7):751–7. doi: 10.1067/mge.2001.112741
25. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J* (2015) 13:8–17. doi: 10.1016/j.csbj.2014.11.005
26. Deo RC. Machine learning in medicine. *Circulation* (2015) 132(20):1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593
27. Wu Y, Liu J, Han C, Liu X, Chong Y, Wang Z, et al. Preoperative prediction of lymph node metastasis in patients with early-T-Stage non-small cell lung cancer by machine learning algorithms. *Front Oncol* (2020) 10:743. doi: 10.3389/fonc.2020.00743
28. Farrokhan N, Holcomb AJ, Dimon E, Karadaghy O, Ward C, Whiteford E, et al. Development and validation of machine learning models for predicting occult nodal metastasis in early-stage oral cavity squamous cell carcinoma. *JAMA Netw Open* (2022) 5(4):e227226. doi: 10.1001/jamanetworkopen.2022.7226
29. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* (2015) 350:g7594. doi: 10.1136/bmj.g7594
30. Hindocha S, Charlton TG, Linton-Reid K, Hunter B, Chan C, Ahmed M, et al. A comparison of machine learning methods for predicting recurrence and death after curative-intent radiotherapy for non-small cell lung cancer: Development and validation of multivariable clinical prediction models. *EBioMedicine* (2022) 77:103911. doi: 10.1016/j.ebiom.2022.103911
31. Tian D, Shiiya H, Takahashi M, Terasaki Y, Urushiyama H, Shinozaki-Ushiku A, et al. Noninvasive monitoring of allograft rejection in a rat lung transplant model: Application of machine learning-based f-fluorodeoxyglucose positron emission tomography radiomics. *J Heart Lung Transplant Off Publ Int Soc Heart Transplant* (2022) 41(6):722–31. doi: 10.1016/j.healun.2022.03.010
32. Akutsu Y, Matsubara H. The significance of lymph node status as a prognostic factor for esophageal cancer. *Surg Today* (2011) 41(9):1190–5. doi: 10.1007/s00595-011-4542-y
33. Twine CP, Lewis WG, Morgan MA, Chan D, Clark GWB, Havard T, et al. The assessment of prognosis of surgically resected oesophageal cancer is dependent on the number of lymph nodes examined pathologically. *Histopathology* (2009) 55(1):46–52. doi: 10.1111/j.1365-2559.2009.03332.x
34. Betancourt Cuellar SL, Sabloff B, Carter BW, Benveniste MF, Correa AM, Maru DM, et al. Early clinical esophageal adenocarcinoma (cT1): Utility of CT in regional nodal metastasis detection and can the clinical accuracy be improved? *Eur J Radiol* (2017) 88:56–60. doi: 10.1016/j.ejrad.2017.01.001
35. Shen W, Shen Y, Tan L, Jin C, Xi Y. A nomogram for predicting lymph node metastasis in surgically resected T1 esophageal squamous cell carcinoma. *J Thorac Dis* (2018) 10(7):4178–85. doi: 10.21037/jtd.2018.06.51
36. Wu J, Chen Q-X, Shen D-J, Zhao Q. A prediction model for lymph node metastasis in T1 esophageal squamous cell carcinoma. *J Thorac Cardiovasc Surg* (2018) 155(4):1902–8. doi: 10.1016/j.jtcvs.2017.11.005
37. Zhou Y, Du J, Li H, Luo J, Chen L, Wang W. Clinicopathologic analysis of lymph node status in superficial esophageal squamous carcinoma. *World J Surg Oncol* (2016) 14(1):259. doi: 10.1186/s12957-016-1016-0
38. Wang S, Chen X, Fan J, Lu L. Prognostic significance of lymphovascular invasion for thoracic esophageal squamous cell carcinoma. *Ann Surg Oncol* (2016) 23(12):4101–9. doi: 10.1245/s10434-016-5416-8
39. Yachida T, Oda I, Abe S, Sekiguchi M, Nonaka S, Suzuki H, et al. Risk of lymph node metastasis in patients with the superficial spreading type of esophageal squamous cell carcinoma. *Digestion* (2020) 101(3):239–44. doi: 10.1159/000499017
40. Min B-H, Yang JW, Min YW, Baek S-Y, Kim S, Kim HK, et al. Nomogram for prediction of lymph node metastasis in patients with superficial esophageal squamous cell carcinoma. *J Gastroenterol Hepatol* (2020) 35(6):1009–15. doi: 10.1111/jgh.14915
41. Chen H, Zhou X, Tang X, Li S, Zhang G. Prediction of lymph node metastasis in superficial esophageal cancer using a pattern recognition neural network. *Cancer Manage Res* (2020) 12:12249–58. doi: 10.2147/CMAR.S270316
42. Mansour NA, Saleh AI, Badawy M, Ali HA. Accurate detection of covid-19 patients based on feature correlated naïve bayes (FCNB) classification strategy. *J Ambient Intell Humanized Computing* (2022) 13(1):41–73. doi: 10.1007/s12652-020-02883-2
43. Yun JK, Kim HR, Park SI, Kim Y-H. Risk prediction of occult lymph node metastasis in patients with clinical T1 through T2 N0 esophageal squamous cell carcinoma. *J Thorac Cardiovasc Surg* (2022) 164(1):265–75. doi: 10.1016/j.jtcvs.2021.10.033
44. Tian D, Huang H, Yang Y-S, Jiang K-Y, He X, Guo X-G, et al. Depth of invasion into the circular and longitudinal muscle layers in T2 esophageal squamous cell carcinoma does not affect prognosis or lymph node metastasis: A multicenter retrospective study. *World J Surg* (2020) 44(1):171–8. doi: 10.1007/s00268-019-05194-6