

86635\_Auto\_Edited.docx

**Name of Journal:** *World Journal of Gastrointestinal Surgery*

**Manuscript NO:** 86635

**Manuscript Type:** ORIGINAL ARTICLE

*Retrospective Study*

**Predicting Lymph Node Metastasis in Colorectal Cancer: An Analysis of Influencing Factors to develop a Risk Model**

Lei YP *et al.* Lymph Node Metastasis in Colorectal Cancer

**Abstract**

**BACKGROUND**

Colorectal cancer (CRC) is a significant global health issue, and lymph node metastasis (LNM) is a crucial prognostic factor. Accurate prediction of LNM is essential for developing individualized treatment strategies for patients with CRC. However, the prediction of LNM is challenging and depends on various factors such as tumor histology, clinicopathological features, and molecular characteristics. The most reliable method to detect LNM is the histopathological examination of surgically resected specimens; however, this method is invasive, time-consuming, and subject to sampling errors and interobserver variability.

**AIM**

Analyze influencing factors and develop and validate a risk prediction model for LNM in CRC based on a large patient queue.

**METHODS**

This study retrospectively analyzed 300 patients who underwent CRC surgery at two Peking University Shenzhen hospitals between January and December 2021. A deep learning approach was used to extract features potentially associated with LNM from

primary tumor histological images while a logistic regression model was employed to predict LNM in CRC using machine-learning-derived features and clinicopathological variables as predictors.

## RESULTS

The prediction model constructed for LNM in CRC was based on a logistic regression framework that incorporated machine learning-extracted features and clinicopathological variables. The model achieved high accuracy (0.86), sensitivity (0.81), specificity (0.87), positive predictive value (0.66), negative predictive value (0.94), area under the curve for the receiver operating characteristic (0.91), and a low Brier score (0.10). The model showed good agreement between the observed and predicted probabilities of LNM across a range of risk thresholds, indicating good calibration and clinical utility.

## CONCLUSION

The present study successfully developed and validated a potent and effective risk-prediction model for LNM in patients with CRC. This model utilizes machine-learning-derived features extracted from primary tumor histology and clinicopathological variables, demonstrating superior performance and clinical applicability compared to existing models. The study provides new insights into the potential of deep learning to extract valuable information from tumor histology, in turn, improving the prediction of LNM in CRC and facilitate risk stratification and decision-making in clinical practice.

**Key Words:** Colorectal cancer; Lymph node metastasis; Machine learning; Risk prediction model; Clinicopathological factors; Individualized treatment strategies

Lei YP, Song QZ, Liu S, Xie JY, Lv GQ. Predicting Lymph Node Metastasis in Colorectal Cancer: An Analysis of Influencing Factors to develop a Risk Model. *World J Gastrointest Surg* 2023; In press

**Core Tip:** This study developed a robust risk prediction model for lymph node metastasis (LNM) in colorectal cancer (CRC) using machine learning and clinicopathological factors. The model achieved high accuracy, sensitivity, and specificity, demonstrating its superior performance compared to existing models. By leveraging deep learning to extract information from tumor histology, the model improves LNM prediction, facilitating individualized treatment strategies and clinical decision-making in CRC.

## **INTRODUCTION**

Colorectal cancer (CRC) is a significant global health issue, and lymph node metastasis (LNM) is a crucial prognostic factor in CRC patients. Accurate prediction of LNM is essential for developing individualized treatment strategies. A study analyzed data from 300 patients who underwent CRC surgery in two Peking University Shenzhen hospitals and constructed and validated a risk prediction model for LNM based on a logistic regression framework incorporating machine learning extracted features and clinicopathological variables. The model achieved high accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and area under the receiver operating characteristic curve in the validation set and demonstrated good calibration and clinical utility. This study provides new insight into the potential of deep learning in extracting valuable information from tumor histology to improve the prediction of LNM in CRC, facilitating risk stratification and decision-making in clinical practice. CRC has emerged as the dominant form of cancer worldwide and plays a substantial role in the mortality associated with cancer<sup>[1]</sup>. LNM is a crucial prognostic factor and determinant of the treatment strategy for CRC<sup>[2]</sup>. Patients with LNM have significantly worse survival outcomes than those without LNM and require more aggressive treatment modalities, such as adjuvant chemotherapy or radiotherapy<sup>[3]</sup>. Therefore, the accurate prediction of LNM is essential for developing individualized treatment strategies for patients with CRC and for improving patient outcomes.

However, the prediction of LNM is challenging and depends on various factors such as tumor histology, clinicopathological features, and molecular characteristics, including the type of surgical intervention. For instance, studies have indicated that old age and male sex are associated with a higher risk of lymph node metastasis<sup>[4]</sup>. Notably, tumors in the right colon exhibit an elevated risk of LNM compared with those in the left colon<sup>[5]</sup>. Additionally, larger and poorly differentiated tumors are more likely to metastasize to lymph nodes<sup>[6]</sup>. Research indicates that the extent of lymph node dissection during surgery may influence the likelihood of detecting lymph node metastasis<sup>[7]</sup>. Moreover, neoadjuvant therapy can potentially influence lymph node status during surgical procedures<sup>[8]</sup>. Nevertheless, these influencing factors have not been thoroughly evaluated or integrated into a risk prediction model for patients diagnosed with CRC.

The most reliable method to detect LNM is histopathological examination of surgically resected specimens<sup>[9]</sup>. However, this method is invasive, time-consuming, and subject to sampling errors and interobserver variability. Moreover, some patients may undergo unnecessary surgery or overtreatment owing to false-positive or false-negative results<sup>[10]</sup>. Therefore, there is an unmet need to develop noninvasive and reliable methods to predict LNM in CRC before surgery. This would aid in preoperative patient counseling, guide treatment decisions, and facilitate patient selection for clinical trials.

Several studies have attempted to identify the risk factors or biomarkers for LNM in CRC using conventional statistical methods or machine learning techniques<sup>[11-13]</sup>. However, most of these studies had limitations, such as small sample size, single-center design, lack of external validation, and low predictive performance<sup>[14-16]</sup>. Moreover, most of these studies have singularly focused on either tumor histology or clinicopathological factors, while ignoring the potential synergistic effects of combining multiple sources of information<sup>[17-19]</sup>.

In the present study, we analyzed the influencing factors, and developed and validated a risk prediction model for LNM in CRC using a large cohort of patients. Relevant

attributes were extracted from the histology of primary tumors and clinicopathological data using machine learning. The extracted attributes were then combined with known predictors to construct a logistic regression model. The efficacy of the model was extensively assessed, focusing on its applicability on an external cohort. The model was juxtaposed with existing predictive models, and its efficiency in stratifying risks and aiding clinical decision-making processes was explored.

## **MATERIALS AND METHODS**

### *Study population and data collection*

This study retrospectively analyzed 300 patients who underwent CRC surgery at Peking University Shenzhen Hospital between January and December 2021. The eligibility criteria for the study was as follows: histopathologically confirmed diagnosis of adenocarcinoma of the colon or rectum, diagnosis of stage I-III disease as per the 8th edition of the American Joint Committee on Cancer staging system, availability of primary tumor histology slides and clinicopathological data, and no prior history of neoadjuvant therapy, synchronous or metachronous malignancy, or hereditary colorectal cancer syndrome. The exclusion criteria was as follows: incomplete or missing data, and substandard quality of histology slides. Finally, 258 participants were included in this study. The participants were then stratified into two groups, with 188 randomly selected for the training set and the remaining 70 assigned to the external validation set. The ethical review boards of both hospitals approved the study protocol, and the requirement for informed consent was waived due to the retrospective design of the study.

Clinicopathological data for patients was collected from electronic medical records and pathology reports. Data included age, gender, tumor location, tumor size, tumor differentiation, depth of tumor invasion, lymphovascular invasion, perineural invasion, tumor budding, and lymph node status. Primary tumor histology slides were retrieved from the pathology archives and scanned using a digital slide scanner (Aperio AT2, Leica Biosystems, Wetzlar, Germany). The scanned images were stored on a secure

server and accessed using Image Viewer software (Aperio ImageScope, Leica Biosystems).

#### *Feature extraction from primary tumor histology*

A deep learning approach was used to extract features potentially associated with LNM from primary tumor histological images. The utilized method can be divided into three steps: generating deep learning embeddings of small patches of tumor tissue using a convolutional neural network (CNN), clustering the embeddings into groups using the k-means algorithm, and selecting the top clusters that added predictive value to the logistic regression model when combined with known baseline clinicopathological variables.

First, a CNN model pre-trained on a large dataset of colorectal polyps to generate embeddings of small patches of tumor tissue was used. The architecture of the CNN model comprises 50 convolutional layers, a global average pooling layer, and a fully connected layer with 512 units. It categorized colorectal polyps into four distinct classes. The output of the model served as the embedding vector for each patch. This model was tested using the training set and subsequently, validated using both the training and external validation sets. From each histological image, patches of  $224 \times 224$  pixels were extracted at 20X magnification, utilizing a sliding window approach with a stride of 112 pixels. We performed manual annotation and excluded patches containing < 50% tumor tissue. Each patch is then fed into the CNN model and produces a 512-dimensional embedding vector.

Second, the k-means algorithm was used to cluster the embeddings into groups based on similarity. Based on the elbow method, k value of 100 was selected as the number of clusters. Then, the cluster frequency was computed for each histological image as the proportion of patches belonging to each cluster. These cluster frequencies were used as feature vectors corresponding to each image.

Finally, a logistic regression model was incorporated to identify the top clusters that could augment the predictive value of the fundamental clinicopathological variables. LNM was the outcome variable in our analysis and six fundamental clinicopathological

variables (age, gender, tumor location, tumor size, tumor differentiation, and tumor invasion depth) served as primary predictors. Subsequently, each cluster frequency was appended as an auxiliary predictor and likelihood ratio tests were conducted to compare the predictive capacities of the models with and without each cluster frequency. By ordering the clusters based on their  $p$ -values, the top ten clusters with  $p < 0.05$  were selected. These top ten clusters were then utilized for machine learning to extract the features corresponding to each image.

#### *Construction and validation of risk prediction model*

The present study employed a logistic regression model for LNM in CRC, leveraging machine learning-derived features and clinicopathological variables as predictors. LNM was utilized as the outcome variable and 10 predictors i.e., age, gender, tumor position, size, differentiation, depth of invasion, lymphovascular invasion, perineural invasion, tumor budding, and frequency of the top ten clusters, were used. To optimize the model, variable selection was performed via backward elimination based on the Akaike Information Criterion. The developed model was then evaluated using a training cohort and validated using both training and external validation cohorts.

Model performance was evaluated using various metrics including accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), calibration plot, decision curve analysis, net reclassification improvement (NRI), integrated discrimination improvement (IDI), and Brier scores. We compared our model with three existing models for LNM prediction in CRC: Kikuchi's model that uses four clinicopathological variables (tumor size  $> 3$  cm, depth of submucosal invasion  $> 1$  m, positive lymphatic invasion, and positive venous invasion), Ueno's model that uses five clinicopathological variables (depth of submucosal invasion  $> 1000$   $\mu$ m, positive lymphatic invasion, positive venous invasion, high budding grade, and poor differentiation grade), and Krogue's model that uses six clinicopathological variables (age  $> 65$  years, male gender, right-sided tumor location,



tumor size > 4 cm, depth of submucosal invasion > 1000  $\mu\text{m}$ , and poor differentiation grade) and machine learning extracted features from primary tumor histology<sup>[20-22]</sup>.

#### *Statistical analysis*

Statistical analysis was performed using R software version 4.0.3. Descriptive statistics was employed to summarize the clinicopathological characteristics of the study population, with the chi-square or Fisher's exact test used to compare categorical variables and the student's t-test or Mann-Whitney U test used to compare continuous variables. Logistic regression models estimated predictor coefficients and odds ratios for LNM in patients with CRC. We employed a backward elimination process guided by the Akaike information criterion for predictor selection of the final model. Likelihood ratio tests were used to compare the nested models with and without each predictor. The discriminatory abilities of the models were assessed using receiver operating characteristic (ROC) and precision-recall curves. The quantitative measures of the discriminatory performance of the models was based on the AUROC and AUPRC.

Calibration plots assessed the correlation between the observed and model predicted LNM probabilities. Decision curve analysis appraised the clinical utility of the models by comparing the net benefits of using the models with alternative strategies (treating all or none) at various risk thresholds. The enhancement, owing to extracted features through machine learning, to the clinicopathological variables for LNM prediction was measured using the NRI and IDI.

The Brier score evaluated the overall performance of the predictive models and quantified the average squared discrepancy between the actual and predicted probabilities for LNM. Bootstrap resampling with 1,000 iterations was used to compute 95% confidence intervals for the AUROC, AUPRC, NRI, IDI, and Brier scores. Statistical significance was set at a threshold of  $p < 0.05$  for statistical significance in all evaluations.

## **RESULTS**

### *Clinicopathological characteristics of the study population*

The study cohort comprised of 258 patients who met the inclusion criteria. The training cohort comprised 188 patients: 141 (75%) with no LNM and 47 (25%) with LNM. The external validation cohort included 70 patients: 52 (74.3%) without LNM and 18 (25.7%) with LNM, as shown in Figure 1. Table 1 shows the clinicopathological characteristics of these patients, with all  $p$ -values greater than 0.05, suggesting no significant differences in various parameters between the two cohorts.

#### *Extraction of features from primary tumor histology*

In our study, a CNN model was employed that was pretrained on an extensive dataset of colorectal polyps to derive 512-dimensional embeddings for each tumor tissue patch. These embeddings were subsequently grouped into 100 clusters and the cluster frequency for each histological description was calculated. The top 10 clusters were selected based on the statistical significance of the likelihood ratio tests, with and without including each cluster frequency as an additional predictor. Representative patches of tumor tissues and detailed descriptions of each of the top 10 clusters are presented in Table 2.

#### *Construction and validation of risk prediction model*

The prediction model for LNM in CRC was based on a logistic regression framework incorporating machine learning-extracted features and clinicopathological variables. Backward elimination based on the Akaike Information Criterion led to the selection of 10 significant predictors: age, tumor location, tumor size, tumor differentiation, tumor invasion depth, lymphovascular invasion, perineural invasion, tumor budding, cluster frequency of cluster 1, and cluster frequency of cluster 2.

Therefore, the resulting logistic regression model integrates these 10 predictors and can be represented by the following mathematical formula:

$$\log p/1-p = -4.32 + 0.02 \times \text{age} - 0.65 \times \text{left colon} - 1.04 \times \text{rectum} + 0.17 \times \text{tumor size} - 0.48 \times \text{tumor differentiation} + 1.32 \times T1 + 2.12 \times T2 + 3.45 \times T3 + 4.67 \times T4a + 5.89 \times T4b + 1.23 \times \text{lymphovascular invasion}$$

vasion+1.01×perineural invasion+0.87×tumor budding+0.05×cluster frequency of cluster 1−0.04×cluster frequency of cluster 2 ( $p$  is the probability of LNM in CRC, and  $Tis$  is the depth of tumor invasion. Right colon is the reference for tumor location)

Table 3 presents the coefficients and odds ratios of the determinants incorporated into the final model. All predictors showed a positive statistically significant correlation with LNM, except tumor differentiation and cluster frequency in cluster 2.

#### *Performance of the risk prediction model*

To evaluate the effectiveness of the risk-prediction model, a comparative analysis was conducted with existing models for LNM prediction in CRC using a range of performance metrics. As shown in Table 4, these metrics included the accuracy, sensitivity, specificity, PPV, NPV, AUROC, AUPRC, NRI, IDI, and Brier score. The model achieved high accuracy (0.86), sensitivity (0.81), specificity (0.87), PPV (0.66), NPV (0.94), AUROC (0.91), AUPRC (0.77), NRI (0.28), and IDI (0.11) and a low Brier score (0.10) in the validation set. Furthermore, the current model outperformed the three existing models in terms of LNM prediction in the validation set, with higher scores for these metrics except for the Brier score. These results underscore the robust discriminative capacity of this model, allowing it to accurately distinguish between patients with and without LNM. In addition, the current model exhibited excellent calibration and resilience within an external cohort, emphasizing its potential clinical utility.

#### *Calibration and decision curve analysis of the risk prediction model*

The calibration and decision curve analysis of the predicted probabilities of LNM between our risk prediction model and existing models in the validation set was assessed, as shown in Figures 2 and 3. The current model demonstrated good agreement between the observed and predicted probabilities of LNM across a range of risk thresholds, with a higher net benefit across a wide range of risk thresholds, indicating good calibration and clinical utility. However, the existing models showed

some deviation from the ideal line with a lower net benefit, especially at low- and high-risk thresholds, indicating poor calibration and clinical utility.

#### *Risk stratification and treatment guidance of the risk prediction model*

The risk prediction model devised in this study facilitated the classification of patients into various risk strata based on the calculated probability of LNM: low (<10%), intermediate (10–30%), and high-risk (>30%). As shown in Table 5, of the 70 CRC patients, 27 (38.6%) were classified as low-risk, 26 (37.1%) as intermediate-risk, and 17 (24.3%) as high-risk. Among the 18 cases of LNM, 1 (5.6%) was classified as low-risk, 6 (33.3%) as intermediate-risk, and 11 (61.1%) as high-risk. Appropriate therapeutic strategies should be judiciously selected in accordance with the assessment of LNM risk in patients with CRC.

## **DISCUSSION**

In the present study, we constructed and validated a risk-prediction model for LNM in CRC based on a large cohort of patients. For this purpose, machine learning techniques were used to extract features from primary tumor histology and clinicopathological data and combine them with established predictors to build a logistic regression model. The features extracted from machine learning provided additional predictive value for LNM in CRC, thereby improving the accuracy and discriminating capacity of the logistic regression model. Additionally, the model demonstrated good calibration and robustness in the validation set. Furthermore, the model can effectively stratify patients into different risk groups and guide the selection of appropriate treatment options.

To avoid the toxicity of over-therapeutic procedures in patients classified as low risk by the current model, curative treatment measures, such as endoscopic resection or local excision that eliminate the need for lymphadenectomy or adjuvant therapy, are recommended. For patients presenting with intermediate-risk T1 CRC, the recommendation is surgical resection accompanied by lymphadenectomy as curative treatment, with the decision for adjuvant therapy contingent on other prognostic factors. In contrast, patients with high-risk T1 CRC are recommended to be

administered neoadjuvant therapy followed by surgical resection with lymphadenectomy for significantly minimizing the risk of recurrence.

Previous studies have shown that tumor histology contains rich information that reflects the biological behavior and molecular characteristics of tumor cells<sup>[23-25]</sup>. However, conventional histopathological evaluations are subjective, qualitative, and limited by human perception. Deep learning is a powerful tool that automatically learns complex and high-dimensional patterns from images and generates quantitative and objective features<sup>[26-28]</sup>. The current study used a pre-trained CNN model to generate the quantitative features of tumor histology. The top cluster frequencies were identified and added to the logistic regression model to enhance their predictive value when combined with baseline clinicopathological variables. These selected histological features may reflect the aggressiveness and invasiveness of the tumor cells and their interaction with the microenvironment, which may influence their ability to metastasize to the lymph nodes.

In this study, we opted for logistic regression as the optimum modeling method, driven by its simplicity, interpretability, and extensive usage in clinical practice<sup>[29-30]</sup>. To avoid the risks of overfitting and multicollinearity, backward elimination, based on the Akaike information criterion for the variable selection, was employed. The current model's ten predictors align with previous research that identified risk factors for LNM in CRC<sup>[31-32]</sup>. Furthermore, we discovered two clusters that were significantly associated with LNM, in line with the current understanding of disease stages based on tumor biological behavior. The positively correlated cluster 1 (characterized by poorly differentiated tumor cells with a high nuclear-cytoplasmic ratio, irregular glandular formation, and sparse stroma) represented a more malignant phenotype of tumor cells. In contrast, the negatively correlated cluster 2 (characterized by well-differentiated tumor cells with a low nuclear-cytoplasmic ratio, regular glandular formation, and abundant stroma) represents a more benign phenotype of tumor cells that are less likely to metastasize to the lymph nodes.

## **CONCLUSION**

The present study provides valuable insights into the determinants of LNM in patients with CRC and presents an innovative risk prediction model; however, it has certain limitations. First, the retrospective design of the study may have contributed to selection bias. Second, we could not include potentially relevant risk factors, such as genetic and lifestyle influences, because of data unavailability<sup>[33]</sup>. Third, model validation was performed exclusively within an independent cohort from the same institution, underscoring the need for future studies to validate the model across diverse populations and within a prospective framework. Finally, the limited size of the validation cohort could potentially affect the reliability of the validation findings. Despite these constraints, our model exhibited high efficacy and adeptness in distinguishing between patients with and without LNM, in addition to displaying sound calibration and robustness within an external cohort. This study forms a strong foundation for future efforts to develop and refine predictive models for LNM in patients with CRC.

In conclusion, a potent and effective risk-prediction model for LNM in patients with CRC was successfully developed and validated. The model utilized machine learning-derived features extracted from primary tumor histology and clinicopathological variables and demonstrated superior performance and clinical applicability compared to existing models. The study provided new insights into the potential of deep learning to extract valuable information from tumor histology, which can improve the prediction of LNM in CRC and facilitate risk stratification and decision-making in clinical practice. Further investigations are crucial to affirm the utility of the present model within larger and more heterogeneous cohorts and to probe the biological and molecular mechanisms underlying the features extracted using machine learning.

## **ARTICLE HIGHLIGHTS**

*Research background*

+ADw-html+AD4APA-p+AD4-Colorectal cancer (CRC) is a significant global health issue, and accurate prediction of lymph node metastasis (LNM) is crucial for individualized treatment strategies. However, predicting LNM is challenging due to various factors and limitations in the histopathological examination method. This study aimed to develop a risk prediction model for LNM in CRC by incorporating machine learning and clinicopathological factors. The model demonstrated high accuracy, sensitivity, and specificity, providing valuable insights into the potential of deep learning in improving LNM prediction and guiding clinical decision-making for CRC patients.+ADw-/p+AD4APA-/html+AD4-

### ***Research motivation***

+ADw-html+AD4APA-p+AD4-Accurate prediction of lymph node metastasis (LNM) in colorectal cancer (CRC) is crucial for improving patient outcomes and developing personalized treatment strategies. However, existing methods are invasive, time-consuming, and prone to errors. This study aimed to address these limitations by developing a risk prediction model using machine learning and clinicopathological factors. The motivation was to provide a more accurate and efficient approach for predicting LNM in CRC, enabling clinicians to make informed decisions regarding treatment and facilitating improved patient care.+ADw-/p+AD4APA-/html+AD4-

### ***Research objectives***

+ADw-html+AD4APA-p+AD4-The main objectives of this study were to analyze the factors influencing lymph node metastasis (LNM) in colorectal cancer (CRC), and to develop and validate a risk prediction model for LNM based on a large patient cohort. The study aimed to utilize machine learning techniques and clinicopathological factors to construct an accurate prediction model that outperforms existing models. The goal was to improve the prediction of LNM in CRC, enabling personalized treatment strategies and enhancing clinical decision-making. Additionally, the study sought to

explore the potential of deep learning in extracting valuable information from tumor histology for improved risk stratification.

### ***Research methods***

In this study, a retrospective analysis was conducted on 300 patients who underwent colorectal cancer (CRC) surgery at two Peking University Shenzhen hospitals between January and December 2021. The main approach involved the development of a risk prediction model for lymph node metastasis (LNM) in CRC. A deep learning method was utilized to extract features from primary tumor histological images that could be associated with LNM. Additionally, a logistic regression model was used, incorporating these machine-learning-derived features along with clinicopathological variables, as predictors for LNM in CRC. The performance of the prediction model was evaluated based on accuracy, sensitivity, specificity, positive predictive value, negative predictive value, receiver operating characteristic curve (ROC AUC), and Brier score to assess calibration and clinical utility.

### ***Research results***

The developed risk prediction model for lymph node metastasis (LNM) in colorectal cancer (CRC) achieved excellent performance. The model demonstrated high accuracy (0.86), sensitivity (0.81), specificity (0.87), positive predictive value (0.66), negative predictive value (0.94), and area under the curve for the receiver operating characteristic (0.91). Additionally, it exhibited a low Brier score of 0.10. The observed and predicted probabilities of LNM showed strong agreement across various risk thresholds, indicating reliable calibration and clinical utility. These findings highlight the effectiveness and clinical applicability of the model, which utilizes machine-learning-derived features from primary tumor histology and clinicopathological variables.



### *Research conclusions*

The study successfully developed and validated a powerful risk prediction model for lymph node metastasis (LNM) in colorectal cancer (CRC). The model, incorporating machine-learning-derived features from primary tumor histology and clinicopathological variables, displayed superior performance and clinical applicability compared to existing models. By leveraging deep learning techniques, valuable information was extracted from tumor histology, leading to improved LNM prediction. This development has significant implications for individualized treatment strategies and clinical decision-making in CRC, enabling better risk stratification. The findings highlight the potential of machine learning and deep learning in enhancing LNM prediction and improving patient outcomes in CRC management.

### *Research perspectives*

The successful development and validation of a potent risk prediction model for lymph node metastasis (LNM) in colorectal cancer (CRC) opens up promising research avenues. Further exploration can focus on refining the model by incorporating additional molecular characteristics and genetic data to enhance its predictive accuracy. Additionally, prospective studies can be conducted to validate the model's performance in larger and diverse patient populations. Furthermore, the integration of real-time image analysis techniques and artificial intelligence algorithms can streamline the prediction process, enabling faster and more accurate LNM assessment. These advancements have the potential to revolutionize clinical practice and optimize treatment strategies for CRC patients.

2%

SIMILARITY INDEX

PRIMARY SOURCES

1	<a href="http://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a> Internet	39 words — 1%
2	Christopher J. Dente, Michael J. Mina, Bryan C. Morse, Hannah Hensman et al. "Predicting the need for massive transfusion: Prospective validation of a smartphone-based clinical decision support tool", Surgery, 2021 Crossref	19 words — < 1%
3	<a href="http://doctorpenguin.com">doctorpenguin.com</a> Internet	18 words — < 1%

EXCLUDE QUOTES ON  
EXCLUDE BIBLIOGRAPHY ON

EXCLUDE SOURCES < 12 WORDS  
EXCLUDE MATCHES < 12 WORDS