Open Access Full Text Article

Integrating Machine Learning and the SHapley Additive exPlanations (SHAP) Framework to Predict Lymph Node Metastasis in Gastric Cancer Patients Based on Inflammation Indices and Peripheral Lymphocyte Subpopulations

Ziyu Zhu^{1,*}, Cong Wang^{1,*}, Lei Shi², Mengya Li³, Jiaqi Li³, Shiyin Liang³, Zhidong Yin¹, Yingwei Xue¹

¹Department of Gastroenterological Surgery, Harbin Medical University Cancer Hospital, Harbin, People's Republic of China; ²Department of Oncology, Beidahuang Industry Group General Hospital, Harbin, People's Republic of China; ³Key Laboratory of Preservation of Genetic Resources and Disease Control in China, Harbin Medical University, Harbin, People's Republic of China

*These authors contributed equally to this work

Correspondence: Yingwei Xue; Zhidong Yin, Email xueyingwei@hrbmu.edu.cn; 289065199@qq.com

Background: The prediction of lymph node metastasis in gastric cancer, a pivotal determinant affecting treatment approaches and prognosis, continues to pose a significant challenge in terms of accuracy.

Methods: In this study, we employed a combination of machine learning methods and the SHapley Additive exPlanations (SHAP) framework to develop an integrated predictive model. This model utilizes the preoperatively obtainable parameter of the inflammatory index, aiming to enhance the accuracy of predicting lymph node metastasis in gastric cancer patients.

Results: Lymph node metastasis stands as an independent prognostic risk factor for gastric cancer patients. Among various models, XGBoost emerges as the optimal machine learning model. In the training set, the XGBoost model exhibited the highest AUC value of 0.705. In the test set, XGBoost demonstrated the highest AUC of 0.695, and the lowest Brier score of 0.218. Notably, in terms of feature importance, PLR emerged as the most significant factor influencing lymph node metastasis in gastric cancer patients. Through the screening of differentially expressed genes, we ultimately identified the prognostic value of six genes: IGFN1, CLEC11A, STC2, TFEC, MUC5AC, and ANOS1, in predicting survival.

Conclusion: The XGBoost model can predict lymph node metastasis (LNM) in gastric cancer patients based on the inflammation index and peripheral lymphocyte subgroups. Combined with SHAP, it provides a more intuitive reflection of the impact of different variables on LNM. PLR emerges as the most crucial risk factor for lymph node metastasis in the inflammation index among gastric cancer patients.

Keywords: Machine Learning, SHAP Framework, Lymph Node Metastasis, Gastric Cancer, Inflammation Indices, Peripheral Lymphocyte Subpopulations

Introduction

Gastric cancer poses a significant public health challenge, standing as the fifth most frequently diagnosed cancer globally and the fourth leading cause of cancer-related deaths.¹ Due to the atypical nature of early gastric cancer symptoms, a significant number of patients are diagnosed with advanced-stage gastric cancer.² The primary route of metastasis in gastric cancer involves the spread of cancer cells through the lymphatic system, significantly impacting the prognosis of this disease.^{3,4} Currently, computer tomography scanning remains the primary tool for preoperative lymph node diagnosis

in cancer,⁵ with accuracy improving with advancements in equipment technology.⁶ While the accuracy of PET-CT in detecting perigastric lymph nodes and determining the N-stage is superior to enhanced CT, the sensitivity of PET-CT in detecting lymph nodes with a diameter of 3mm is less than satisfactory.⁷ At the same time, CT scans are unable to distinguish between the enlargement of lymph nodes due to inflammation and that caused by cancer. Therefore, accurately predicting the risk of lymph node metastasis in gastric cancer patients is of paramount importance for guiding treatment strategies. Current research primarily focuses on the prediction of lymph node metastasis in early gastric cancer, as these findings significantly influence the choice of surgical interventions, such as early endoscopic resection, local excision, or standard surgery.⁸ However, the predictive significance for advanced gastric cancer remains unclear, and existing recommendations and treatment strategies largely rely on the preferences of surgeons and advancements in neoadjuvant therapy, rather than on comprehensive predictive models. Therefore, conducting research on lymph node metastasis prediction in advanced gastric cancer is particularly essential for improving surgical decision-making and enhancing patient outcomes.

Gastric cancer is the result of complex interactions between environmental, host genetic, and microbial factors. There is substantial evidence supporting the association between chronic inflammation and cancer development.⁹ This association is particularly relevant in gastrointestinal cancers, where microbial pathogens are responsible for chronic inflammation, which may act as a triggering factor for these cancers.¹⁰ Inflammatory indices, such as the Neutrophil-to-Lymphocyte Ratio (NLR) and Platelet-to-Lymphocyte Ratio (PLR), have emerged as potential biomarkers reflecting the systemic inflammatory response in cancer patients. Due to their potential associations with tumor progression and metastasis, these indices have garnered significant attention in gastric cancer research.^{11,12} This study delves deep into the intricate interplay between inflammation and the progression of gastric cancer, with a particular emphasis on the role of Platelet-to-Lymphocyte Ratio (PLR). PLR, calculated from routine blood tests, has garnered significant attention as a potential non-invasive, cost-effective, and easily accessible biomarker.^{13–15} Reportedly, T lymphocytes play a crucial role in controlling and eliminating cancer,^{16–18} but their contribution to lymph node metastasis in gastric cancer patients is not yet clear. However, their practicality in predicting lymph node metastasis in gastric cancer patients remains a subject that requires further investigation. In this context, our study aims to explore the relationship between the inflammatory index and lymphocyte subpopulations in gastric cancer patients and lymph node metastasis using machine learning techniques. By harnessing the power of machine learning and integrating data from diverse patient populations, our research seeks to unveil the potential of the inflammatory index and lymphocyte subpopulations as valuable tools in the clinical management of gastric cancer. Ultimately, this may contribute to the development of more effective treatment strategies.

Artificial intelligence (AI) is an evolving scientific discipline that mimics, enhances, and broadens human intelligence's theoretical foundations, methods, technologies, and practical applications. Within the context of modern medical equipment and instruments, AI predominantly acquires data using machines, subsequently refining, and analyzing the data to yield qualitative or quantitative solutions.¹⁹ For example, machine learning has been employed in predicting peritoneal metastasis in gastric cancer patients.²⁰ Using machine learning, cancer-specific mortality can be predicted for patients with primary non-metastatic invasive breast cancer.²¹ Predicting preoperative early lymph node metastasis in patients using a machine learning model.⁸ However, current research only includes studies related to predicting early gastric cancer lymph node metastasis and does not encompass studies on predicting lymph node metastasis in both early and advanced stage gastric cancer. As a result, this research fills this void in scientific literature. This study aims to predict lymph node metastasis in gastric cancer patients by combining machine learning methods and transcriptome analysis. This approach not only helps improve the accuracy of predictive models but also contributes to a deeper understanding of biomarkers related to lymph node metastasis.

Methods

This study involved a retrospective analysis conducted at Harbin Medical University Cancer Hospital, encompassing a cohort of 1010 gastric cancer patients (Cohort I) diagnosed between 2014 and 2016. The primary focus was to utilize machine learning methods in predicting lymph node metastasis based on inflammation indices and peripheral lymphocyte subpopulations.

Patient Selection

Using the data derived from these patients, we divided the cohort into a training set and a test set, with 707 patients allocated to the training set for model optimization and 303 patients to the test set for model validation. This division maintained a ratio of 7:3.

Data Collection for Gene Expression Analysis

Additionally, we gathered transcriptomic information from 269 gastric cancer patients for further analysis. An internal validation was further conducted with 190 gastric cancer patients(CohortII)diagnosed in 2017 at Harbin Medical University Cancer Hospital.

Outcome Definition

The primary outcome of this study was the identification of perigastric lymph node metastasis. Positive lymph node metastasis was confirmed through pathological analysis, determined by the presence of one or more positive lymph nodes.

Feature Selection and Data Preprocessing

Missing values for variables with a missing rate below 30% were imputed using the KNNImputer technique.²² Because the range of diverse features varies significantly, and certain algorithms necessitate quantification for data normalization, One-Hot encoding is employed to manage multiple classes of variables.²³ The LASSO analysis identified the entry variables with the corresponding coefficients based on the best lambda values. We selected features related to lymph node metastasis in gastric cancer to build an ML model.

Model Development

We developed six machine learning models based on clinical data to predict lymph node metastasis in gastric cancer patients. These algorithms include logistic regression (LR), decision tree (DT), support vector machines (SVM), k-nearest neighbors (KNN), random forest (RF), and extreme gradient boosting (XGB). LR is a type of machine learning algorithm predominantly utilized for solving binary problems, often employed to predict the likelihood of an event's occurrence.²⁴ DT represent a foundational and notable algorithm within the realm of machine learning.²⁵ SVM, functioning in a multi-dimensional space, is commonly used to categorize items with multiple attributes into two separate groups, focusing on binary classification.²⁶ KNN stands as a remarkably simple and widely acknowledged method in data mining classification strategies, known for its effectiveness in statistical pattern recognition.²⁷ RF minimizes training variance, enhancing consistency and the ability to apply knowledge to new data.²⁸ XGBoost can efficiently tackle real-world large-scale problems while conserving minimal computational resources.²⁹ Patients were randomly divided into training and test sets in a 7:3 ratio. These machine learning models were constructed using the training set and subsequently assessed with the test set. Internal validation was instrumental in evaluating the models' performance on their respective training data (Figure 1).

The Interpretability of Optimal ML Model

Because it's challenging for machines to explain why machine learning algorithms can provide accurate predictions for specific patient cohorts, we introduced SHAP values in our research.³⁰ SHAP, introduced by Lundberg and Lee,³¹ is a unified framework for explaining machine learning predictions and a novel approach for interpreting various black-box machine learning models. The core idea behind SHAP is that the contribution of each feature to a model's prediction can be decomposed into a series of independent Shapley values. These Shapley values reveal the relative impact of each feature on each prediction, enabling us to understand why the model makes a particular prediction. Through SHAP analysis, we gain a clearer understanding of the model's prediction logic, which is instrumental in enhancing the model's transparency and credibility.



Figure I Analysis flow for the development and evaluation of models.

Internal Validation

In addition to model development, we also conducted internal validation using an independent dataset from our center. Internal validation in machine learning is particularly crucial in clinical practice. It not only accurately assesses the model's performance on training data but also promptly identifies and mitigates the risk of overfitting. This is vital for ensuring the reliability and effectiveness of the model, as in the healthcare sector, the precision and generalizability of a model directly impact patient diagnosis, treatment outcomes, and safety. Through internal validation, we can better fine-tune and optimize our models to adapt to the complex and varied clinical data, thereby enhancing the accuracy and efficiency of medical decision-making.

Construction of GC Database

We collected clinical information and preserved tumor samples from 269 gastric cancer patients who had undergone radical gastrectomy at Harbin Medical University (HMU) Cancer Hospital, establishing the HMU-GC cohort. As of December 2021, this dataset was last revised. All samples were acquired with the patients' written informed consent, and the study was granted approval by the Institutional Review Committee of Harbin Medical University Cancer Hospital. Novogene Biotech Co. Ltd. in Beijing, PR China, conducted the mRNA sequencing, RNA isolation, and library construction. The data have been deposited in the Gene Expression Omnibus (GEO) repository under the accessions GSE184336 and GSE179252.

Bioinformatics Analyses

We divided patients into two groups based on the median PLR from the sequencing database and conducted differential gene analysis. We performed differential gene analysis based on the presence or absence of lymph node metastasis in patients. We also conducted differential gene analysis between cancer tissue and adjacent normal tissue. Subsequently, we identified the intersection of differentially expressed genes in these three groups. The differential gene analysis was conducted using the limma software package. We utilized the R software package "glmnet" to integrate survival time, survival status, and gene expression data for regression analysis using the lasso-Cox method. Additionally, we

implemented 10-fold cross-validation to obtain the optimal model. The risk score formula is constructed as follows: Risk score = (\sum coefficientx * expression of signature genex), where "genex" refers to the identified genes. The regression coefficients were derived from Cox proportional hazards analysis. Gastric cancer patients were stratified into high-risk and low-risk groups based on the median cut-off value of the risk score. This analysis aimed to provide annotations related to the molecular function (MF), cellular component, and biological processes (BPs) associated with these distinct genes. For the most recent gene annotations for the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, we utilized the KEGGrestAPI (https://www.kegg.jp/kegg/rest/keggapi.html). Furthermore, we employed the "cluster profile" package for gene set enrichment analysis (GSEA) with the goal of investigating the enriched pathways within high-risk groups and gaining insights into the associated mechanisms. In this analysis, we utilized the hallmark as the reference genome. We selected and showcased the five pathways most significantly associated with the risk score. The ESTIMATE algorithm was employed to compute the immune score, stromal score, and tumor purity, which are determined by the proportion of immune and stromal cells. Additionally, the assessment of immune cell infiltration in individual tumor samples was conducted utilizing both the CIBERSORT algorithm.

Statistical Analysis

The analysis and calculations were performed using R software version 4.2.3 and Python version 3.9.

We represented categorical variables using frequencies and percentages (%) and examined distinctions utilizing either the chi-squared test or Fisher's exact test. Continuous variables were depicted by displaying both the median and mean values, in addition to detailing the interquartile range (IQR) and standard deviation (SD). Multivariate analyses were used Cox proportional hazards regression models. A comprehensive assessment of the ML models' discrimination was performed using multiple evaluation indices, which included Precision-Recall Curve, Area Under the Decision Curve (AUDC),and the area under the receiver operating characteristic curve (AUC). The Brier score was used to evaluate model calibration. P<0.05 was deemed statistically significantThe Kaplan-Meier method and Log rank test were then adopted to the comparison of the survival curves of the two groups of patients. Statistical analyses were conducted using bilateral tests, and a significance level of <0.05 was employed to determine statistical significance.

Results

The Relationship Between Clinicopathological Characteristics and OS

Our study initiated with a univariate Cox analysis, aiming to evaluate the influence of lymph node status on patients' overall survival (OS) independently of other potential factors. The outcomes of the univariate Cox regression analysis indicated a significant association between lymph node status and OS. Specifically, lymph node-positive patients exhibited reduced survival rates, and this association was statistically significant (p < 0.001). To gain a more comprehensive insight into the influence of lymph node status on OS, a multivariate Cox regression analysis was conducted. This comprehensive analysis considered several potential factors, encompassing age, gender, tumor stage, pathological characteristics, hematological parameters, and more. The outcomes of this multifactorial analysis reaffirmed that even after adjusting for these factors, lymph node status remained notably associated with OS, displaying a Hazard Ratio of 2.541 (with a 95% confidence interval between 1.766 and 3.657) (p < 0.001, Table 1) This implies that in the presence of multiple contributing elements, lymph node status sustains its substantial predictive value, signaling its relevance to the survival duration of gastric cancer patients.

Collectively, our findings, derived from both univariate and multivariate Cox regression analyses, lead to the conclusion that the relationship between lymph node status and patient survival remains consistent, regardless of the other contributing factors. This emphasizes the stand-alone importance of lymph nodes in the assessment of gastric cancer prognosis.

Baseline Characteristics of Gastric Cancer Patients

The internal cohort consists of 1010 gastric cancer patients, including 581 with lymph node-positive and 429 with lymph node-negative status. The internal validation cohort comprises 190 gastric cancer patients, with 108 being lymph node-positive and 82 lymph node-negative (Table S1)

Characteristics	Univariable Analysis HR (95% CI)		Multivariable Analysis HR (95% CI)	p Value
Sex				
Male	l (reference)		l (reference)	
Female	0.968(0.757-1.239)	0.797	NA	
AGE	1.024(1.012-1.036)	0.000	1.014(1.001-1.027)	0.040
pT				
ТІ	l (reference)		l (reference)	
Т2	2.099(1.013-4.348)	0.046	1.527(0.726-3.211)	0.264
ТЗ	7.755(4.550-13.219)	0.000	3.401(1.883-6.142)	0.000
T4	17.741(10.427-30.186)	0.000	4.785(2.547-8.991)	0.000
N0/N+	, ,			
NO	l (reference)		l (reference)	
N+	6.341(4.595-8.749)	0.000	2.541(1.766-3.657)	0.000
μM				
MO	l (reference)		l (reference)	
MI	6.458(4.427-9.421)	0.000	1.968(1.229–3.151)	0.005
MESENCHYME	••••••(
Medullary type	l (reference)		l (reference)	
Intermedius type	0.886(0.613–1.280)	0.518	NA	
Hard type	0.939(0.593-1.487)	0.789	NA	
Lauren	0.757(0.575 1.107)	0.707		
Intestinal	l (reference)		l (reference)	
Mixed	1555(1175-2057)	0.002		0 305
Diffuse	1.555(1.175 2.057)	0.002	0.915(0.626-1.337)	0.505
INF	1.005(1.200 2.207)	0.000	0.715(0.020 1.337)	0.010
INFa	l (reference)		l (reference)	
INFb	1 146(0.861 - 1.527)	0 350	1 137(0 834–1 548)	0417
INFc	1.922(1.436-2.572)	0.000	1319(0.935-1.861)	0.115
HFR2	1.722(1.100 2.072)	0.000	1.517(0.755 1.661)	0.115
0	l (reference)		l (reference)	
	0 942(0 724-1 225)	0.655	NA	
2	1 110(0.793 - 1.554)	0.543	NA	
3	1.184(0.729–1.923)	0.494	NA	
Chemotherapy	1.101(0.727 1.725)	0.171		
Yes	l (reference)		l (reference)	
No	0.872(0.698-1.088)	0.224	NA	
Histology	0.072(0.070 1.000)	0.221		
Well-differentiation	l (reference)		l (reference)	
Medium-differentiation	2 339(1 226_4 462)	0.010	1 325(0 673-2 608)	0.415
Poorly-differentiation	2.687(1.387_5.206)	0.003	1.325(0.075 2.000)	0.414
Poorly cohesive carcinoma	3 514(1 832_6 739)	0.000	1.5+0(0.00+-2.703) 1 512(0 726-3 147)	0.269
Papillary carcinoma	5.314(1.032-0.737) 5.294(2.015-13.911)	0.000	4 34(507_ 334)	0.207
Mucinous carcinomas	2 561(1 150-5 700)	0.001	0.686(0.287_1.638)	0.396
Location	2.301(1.130-3.700)	0.021	0.000(0.207-1.000)	0.570
Lower	l (reference)		l (reference)	
Middle	1 109(0 792-1 555)	0 547	86(0.837_1.482)	0 3 3 8
Lipper	1.107(0.772-1.333)	0.001	1 356(0.037-1.002)	0.336
Entire	6 690(4 733 9 464)	0.001	2 997(2 012 4 444)	0.075
	0.070(7.7)070(0.0)	0.000	2.777(2.012-7.704)	0.000
	l (reference)		l (reference)	
les No		0.000		0.000
	5.042(4.015-6.354)	0.000	1.800(1.313-2.469)	0.000

Table I The Relationship Between Clinicopathological Characteristics and OS

(Continued)

Characteristics	Univariable Analysis HR (95% CI)	p Value	Multivariable Analysis HR (95% CI)	p Value
Tumor size	1.021(1.018-1.025)	0.000	1.002(0.997-1.008)	0.389
PNI	0.948(0.932-0.965)	0.000	1.021(0.995-1.047)	0.110
LMR	0.843(0.787-0.902)	0.000	0.923(0.850-1.004)	0.061
NLR	1.139(1.091–1.190)	0.000	0.959(0.874-1.053)	0.379
PLR	1.003(1.002-1.004)	0.000	0.998(0.996-1.000)	0.107
DIR	2.318(1.616-3.324)	0.000	1.532(0.956-2.455)	0.077
INI	0.976(0.964–0.987)	0.000	1.006(0.993-1.018)	0.382
GLR	1.151(1.092-1.213)	0.000	1.072(0.983-1.169)	0.114
FLR	1.298(1.223-1.377)	0.000	1.162(1.011–1.336)	0.035
BMI	0.944(0.913-0.977)	0.001	0.966(0.931-1.003)	0.072
CD3	0.996(0.985-1.007)	0.450	NA	
CD4	0.976(0.964–0.989)	0.000	1.002(0.987-1.016)	0.837
CD8	1.023(1.010-1.036)	0.001	1.022(1.009–1.036)	0.001
CD4/CD8	0.892(0.793-1.004)	0.059	NA	
Double positive T cells	0.917(0.767-1.095)	0.338	NA	
CD19	0.978(0.954-1.004)	0.091	NA	
IgA	1.123(1.040-1.212)	0.003	1.078(0.988-1.177)	0.092
lgG	0.994(0.964-1.026)	0.723	NA	
lgM	0.789(0.635-0.981)	0.033	0.789(0.643-0.968)	0.023
TRF	0.886(0.744-1.056)	0.176	NA	
KAP	1.005(0.971-1.040)	0.798	NA	
LAM	1.050(1.016-1.086)	0.004	1.054(1.012-1.097)	0.010
K/L	0.913(0.743-1.122)	0.388	NA	

Table	1	(Continued)
Iable	. (Continueu).

Selected Variables

The LASSO method involves selecting variables that demonstrate the lowest mean squared error (MSE) across 10 crossvalidation folds. A predictive ML model is constructed by employing the features identified with the minimum mean square error (MSE) in LASSO. Ultimately, the ML models were trained using seven variables: PLR, BMI, PNI, CD4, DIR, FLR and INI. (Figure 2) In our analysis, we established distinct thresholds for each feature using the maximal Youden Index, dividing them into high and low groups. We then explored the influence of these features on survival and mortality outcomes. Notably, patients in the low groups of PLR, DIR, and FLR exhibited significantly improved survival, while those in the high groups of BMI, PNI, CD4, and INI demonstrated better survival rates. Extending our investigation through inter-group comparisons, we verified the consistent distribution patterns of these 7 features between the survival and mortality groups (Figure S*1).

The Predictive Ability of Different ML Models

Our machine learning model was crafted utilizing data from our facility, encompassing a total of 1010 gastric cancer patients. These individuals were allocated randomly into two groups, with 707 forming the training set and 303 comprising the test set, maintaining a ratio of 7:3 (Table S2). In the training set, the XGBoost (XGB) model exhibited the highest average accuracy and the highest AUC value of 0.705 (Figure 3A and B). In the test set, the XGB model demonstrated the highest AUC value of 0.695, the highest average precision of 0.765, and the highest AUDC value of 0.249 (Figure 3C–E). The calibration curve for this model (depicted in red) closely matches the ideal calibration curve (in black), underscoring its reliability. Additionally, the XGBoost model achieved the lowest Brier score of 0.218 (Figure 3F), underscoring its calibration capability. Through comprehensive comparative analysis, the results indicate that the XGBoost (XGB) model outperforms the other five models.



Figure 2 Misclassification error of different quantitative variables revealed by the LASSO regression model. (A) The red dot represents the misclassification error value, gray line represents the standard error (SE), and left and right vertical dashed lines represent the optimal value under the minimum criterion and I-SE criterion, respectively, and "lambda" is the tuning parameter. (B) Variation in coefficient values (Coefficients) corresponding to the variables with the lambda value of the tuning parameter.



Figure 3 The performance comparison of different ML models in train and test sets. (A) The AUC comparison of different ML models in train set (10-fold cross validation). (B) The average accuracy comparison of different ML models in train set. (C) ROC of Various Models in the Validation Set. (D) Precision-Recall Curves of Various Models in the Validation Set. (E) Area Under the Decision Curve (AUDC) of Various Models in the Validation Set. (F) Calibration Curves of Various Models in the Validation Set.

Visualization of Feature Importance

In Figure 4A, we present the relative importance of various features used in predicting gastric cancer lymph node metastasis. The order of these features is determined based on the average absolute SHAP values, aiding in the identification of the most influential features for model predictions. The graph provides a clear visualization of the



Figure 4 The XGB model's interpretation. (A): The importance ranking of the top 7 variables according to the mean (|SHAP value|); (B): The importance ranking of the top 7 risk factors with stability and interpretation using the optimal model. The higher SHAP value of a feature is given, the higher risk of death the patient would have. The red part in feature value represents higher value. (C and D): The interpretation of model prediction results with the two samples (the values of each variable are normalized values).

significant role of PLR in predicting lymph node metastasis and offers an intuitive means to understand which factors play a pivotal role in predicting lymph node metastasis. This visualization deepens our understanding of the model's decision-making process. Additionally, we explored the predictive capabilities of each feature in relation to lymph node metastasis (Figure S*2). With Figure 4B we can systematically observe how SHAP values change for different features as their values vary. For instance, this figure illustrates that higher PLR levels might be associated with an increased risk of lymph node metastasis, while higher PNI levels could relate to a reduced risk. The dynamic view of SHAP is an interactive tool that enables us to explore the feature importance and its impact on predictions in greater depth.

Interpretation of Personalized Predictions

SHAP is an additive interpretability model inspired by Shapley values. For each prediction sample, the model generates a prediction value, and the SHAP value represents the numerical allocation of each feature within that sample. We have obtained the optimal prediction model and the required optimal set of metrics to investigate the impact of each feature on the results and the relationship between feature value magnitude and the risk of severity. In Figure 4C, for the positive samples, the values of PLR are relatively larger, displayed in red, indicating that PLR has a positive impact on the outcome. The red bar for FLR is the widest, highlighting its substantial influence on the outcome. As shown in Figure 4D, for the negative samples, the values of FLR are relatively smaller, shown in blue. This suggests that FLR helps to reduce the SHAP value of the sample, thereby having a negative impact on the outcome. Notably, the blue bar for PLR is the widest, indicating its significant impact on the outcome.

Subgroup Analysis Findings

In our study, we categorized different locations of gastric cancer into four subgroups: gastric fundus, gastric body, gastric antrum, and whole stomach. We also classified them based on the depth of infiltration into early-stage gastric cancer and advanced-stage gastric cancer. We placed a strong focus on investigating the performance of the XGBoost (XGB) model within these six subgroups. The area under the curve (AUC) values for the ROC curves of each subgroup were 0.750, 0.725, 0.764, 0.952, 0.730, and 0.725, respectively (Figure 5). This analysis aimed to provide a deeper understanding of the model's performance across different subgroups. By examining ROC curves, we could assess the model's classification performance. It's evident that the XGB model consistently demonstrated favorable predictive capabilities in various subgroups. Additionally, the SHAP analysis enabled us to comprehend how each feature influenced the model's predictions within each subgroup and whether these influences varied among the subgroups. The feature importance rankings across the six subgroups were generally similar. In each subgroup, PLR consistently emerged as the most significant contributor to lymph node metastasis. Moreover, a higher PLR value was associated with an increased likelihood of lymph node metastasis (Figure 6). In our research, we employed the XGB model to calculate a risk score for each patient, then examined how these scores influenced lymph node metastasis in diverse subgroups across various cohorts. Our observations revealed a notable correlation, with most subgroups showing a higher likelihood of lymph node metastasis at increased risk scores (Figure S*3).

Significance of PLR as a Prognostic Indicator in Gastric Cancer

Regardless of whether in the high PLR group or the low PLR group, patients with lymph node positivity exhibited a poorer prognosis compared to those with lymph node negativity (p<0.001 and p<0.001) (Figure S*4A and B) Within the lymph node-positive patients, those with low PLR levels demonstrated significantly better survival compared to those with high PLR levels (p=0.001) (Figure S*4C). However, in lymph node-negative patients, we did not observe a significant difference in survival between gastric cancer patients with high and low PLR levels (Figure S*4D). These results suggest that PLR may have a more significant predictive value in patients with lymph node metastasis. Based on the association between PLR and lymph node status, we designed a novel scoring system. Specifically, patients



Figure 5 AUC Values of the XGB Model Across Different Subgroups: Early Gastric Cancer (A), Advanced Gastric Cancer (B), Cardia Gastric Cancer (C), Body Gastric Cancer (D), Antral Gastric Cancer (E), and Total Gastric Cancer (F).



Figure 6 Explanation and Importance Ranking of Features in the XGB Model Across Different Subgroups: Early Gastric Cancer (A), Advanced Gastric Cancer (B), Cardia Gastric Cancer (C), Body Gastric Cancer (D), Antral Gastric Cancer (E), and Total Gastric Cancer (F).

were categorized into four subgroups: high PLR and lymph node metastasis positive (score of 2), low PLR and lymph node metastasis negative (score of 0), high PLR and lymph node metastasis negative, and low PLR and lymph node metastasis positive (score of 1). Through this approach, we established a concise and efficient scoring system. To validate the effectiveness of this novel scoring system, we conducted detailed survival analysis. The K-M curves clearly demonstrated survival differences among different scoring groups, confirming that patients with a score of 0 had better survival outcomes than those with scores of 1 and 2 (P<0.001, <u>Figure S*4E</u>). Our study reveals that PLR levels are a crucial indicator for predicting the occurrence of lymph node metastasis and are prognostic factors influencing the outcomes of gastric cancer patients. Therefore, we further explored the transcriptome to investigate the association between PLR levels and lymph node metastasis, providing insights into potential molecular mechanisms regulating the survival of gastric cancer patients.

Establishment of Risk Scores

During the feature selection phase, we identified a significant correlation between PLR and lymph node metastasis in gastric cancer patients. We aim to investigate at the transcriptomic level how PLR influences lymph node metastasis through specific biological behaviors, ultimately impacting patient prognosis. We conducted differential gene analysis among the group with different PLR, group with and without lymph node metastasis, group with tumor and adjacent tissue. The Venn diagram analysis revealed the gene overlap among different groups. Through comparison, we identified a total of 53 common genes (Figure S*5A), all of which exhibited significant differential expression in these key categories. The LASSO method was further employed to analyze these genes, significantly minimizing the potential overfitting issues (Figure S*5B and C). Cox Proportional-Hazards analysis confirmed six prognosis related genes (IGFN1, CLEC11A, STC2, TFEC, MUC5AC, ANOS1) to adhere to the proportional hazards assumption, subsequently used in building a risk score model. Among these six signature genes, CLEC11A was identified as a protective gene, while the remaining five were associated with increased risk. The overall prognostic differences are significant, as indicated by the logtest (1.63659410552773e-10), sctest (1.75595344427975e-10), and wald test (2.74492820234409e-10). The C-index is 0.680938731637 (Figure S*5D).

The Prognostic Value of Risk Scores

In the HMU-GC cohort, all patients were assigned a risk score, and based on these scores, they were categorized into high and low-risk groups. It became evident that patients in the high-risk group exhibited poorer prognoses. A Kaplan-Meier (KM) survival curve was plotted, and the results of the Log rank test showed statistical significance (p<0.05). We analyzed the relationship between different risk scores and patients' follow-up duration, events, and the expression

changes of various genes. It can be observed that as the risk score increases, the patients' survival rate significantly decreases (Figure S*5E).

Enrichment Analysis of Risk Score

In the HMU-GC cohort, we conducted LIMMA analysis on both the high-risk score group and the low-risk score group to identify differentially expressed genes between the two groups. We discovered that 169 genes were significantly upregulated, while 14 genes were significantly downregulated (Figure S*6A). Subsequent analyses will focus on these genes with significant alterations to explore their potential roles in the process of lymph node metastasis. To comprehend the functional implications of these genes across various biological dimensions, we conducted a Gene Ontology (GO) enrichment analysis. The figure below illustrates the results of GO enrichment analysis, covering biological processes (BP), cellular components (CC), and molecular functions (MF). In terms of biological processes, significant correlations were found with the negative regulation of cell activation, leukocyte cell-cell adhesion, and the regulation of leukocyte proliferation. Regarding cellular components, there were significant links to the collagen-containing extracellular matrix, collagen trimer, and endoplasmic reticulum lumen. In the realm of molecular functions, genes involved in glycosaminoglycan binding, heparin binding, and those functioning as structural constituents of the extracellular matrix were identified (Figure S*6B). Based on GSEA enrichment analysis using HALLMARK gene sets on patients from the training cohort with high and low risk score expression based on HMU-GC, we aimed to elucidate the potential mechanism of PLR promoting lymph node metastasis. The findings suggest that PLR is primarily involved in processes such as allograft rejection, epithelial-mesenchymal transition, and Interferon Gamma Response. Additionally, it plays a significant role in inflammatory response and Interferon AlphaResponse. These biological behaviors collectively contribute positively to the progression of lymph node metastasis (Figure S*6C).

Correlation Between Gene Expression and Immune Cell Infiltration in the TME of GC

CIBERSORT packages were employed to analyze the correlation between different risk score GC patients in the HMU-GC cohort and the infiltration levels of various immune cells in the immune tumor microenvironment (TME). CIBERSORT analysis revealed that patients with high-risk score had higher levels of T_cells_CD4_memory_activated, T_cells_follicular_helper, Macrophages_M0, and Macrophages_M1. Conversely, these patients displayed decreased levels of Plasma_cells and Mast_cells_resting (Figure S*7A). Furthermore, ESTIMATE analysis indicates that the high-risk group exhibited significantly higher stromal and Immune Score compared to the low-risk group. Additionally, the analysis pointed out that tumor purity was notably lower in the high-risk group than in the low-risk group, suggesting a more complex interaction within their tumor microenvironment (Figure S7*B).

Development of Nomograms to Predict Individual Survival Outcomes

First, we performed variable selection to ensure that only the most significant variables were included in the nomogram. This selection process was based on their p-values and hazard ratios (HR) from both single-factor and multiple-factor Cox regression models. Only variables that were statistically significant were considered. We demonstrated that pT, pN, CA199, CA125, and Risk Score had a significant impact on survival. Subsequently, we utilized these selected variables to construct the nomogram for exploring the one-year, three-year, and five-year survival rates of gastric cancer patients (Figure S*8A). Furthermore, we conducted a study on the three-year and five-year Decision Curve Analysis (DCA) relative to predictions made by the Nomogram, as opposed to the Risk Score alone. The results demonstrated that, for both the three-year and five-year DCA, the Nomogram's predictive capability exceeded that of the Risk Score (Figure S*8B and C). The calibration curve demonstrated good consistency (Figure S*8D) We also explored the predictive capabilities of the nomogram. The AUC value of 0.678 for the 1-year prediction indicates moderate accuracy, while the 3-year prediction demonstrates an improved AUC value of 0.740, surpassing the 1-year forecast. Notably, the 5-year predictive capability achieves an AUC of 0.750, signifying that the nomogram's accuracy peaks over this longer period when compared against the other time points. Overall, the nomogram is more adept at forecasting long-term outcomes (3 and 5 years) than shorter-term results (1 year) (Figure S*8E).

Discussion

Gastric cancer is the fifth most common malignancy worldwide and the fourth leading cause of common cancer-related deaths.¹ The presence of metastatic lymph nodes significantly contributes to the unfavorable prognosis of individuals diagnosed with gastric cancer.^{32,33} The inflammation index has also been confirmed to be associated with poor prognosis in gastric cancer. For example, after neoadjuvant chemotherapy, the platelet-to-lymphocyte ratio is negatively correlated with the prognosis of gastric cancer patients.³⁴ The preoperative NLR can serve as a prognostic factor for patients with gastric cancer, with a high NLR specifically associated with a poor prognosis in GC patients.³⁵ The peripheral blood T-cell subpopulation has been confirmed as a prognostic factor for gastric cancer.³⁶ Therefore, we aim to predict lymph node metastasis in gastric cancer patients using the inflammation index and peripheral lymphocyte subpopulations can be obtained through routine preoperative blood tests, enabling physicians to assess patients' conditions and prognosis more conveniently and guide them in selecting appropriate treatment strategies.

Machine learning can handle vast amounts of patient data, enabling comprehensive analysis of patients' medical conditions.³⁷ Machine learning models can automatically extract significant features from the data without the need for manual feature selection or design, aiding in the discovery of hidden predictive factors.³⁸ Additionally, machine learning models can address complex nonlinear relationships, thus providing a more accurate capture of the multidimensional characteristics of gastric cancer patients. While machine learning may not be as intuitive as traditional models in terms of interpretability and understanding, we have employed SHAP to provide a visual and intuitive explanation of the models used in this study. Currently, numerous studies have been conducted to predict the lymph node metastasis in gastric cancer patients.³⁹⁻⁴³ However, current research in this field tends to fall into one of three categories: either solely focusing on early-stage gastric cancer, utilizing isolated inflammation indices for prediction, or creating novel scoring systems that lack confirmation from other researchers regarding their suitability for lymph node metastasis prediction in gastric cancer patients. Most lymph node metastasis predictions rely on nomogram charts, which, while providing an intuitive means to forecast event probabilities or scores, are often limited by their reliance on linear or simple nonlinear models. The factors influencing lymph node metastasis in gastric cancer patients are typically complex and may not adhere to linear relationships. These limitations underscore the current state of research in this area. There is no prior research reporting the relationship between peripheral lymphocyte subpopulations and lymph node metastasis in gastric cancer patients. This study employs machine learning to predict lymph node metastasis in gastric cancer patients using various inflammation indices and peripheral lymphocyte subpopulations, effectively addressing limitations identified in previous research.

In this study, our aim is to utilize inflammatory indices and peripheral lymphocyte subpopulations to predict lymph node metastasis in gastric cancer patients. We conducted a retrospective analysis based on data from our center and constructed machine learning models. Notably, a significant gap exists in the literature concerning the predictive value of lymph node metastasis in advanced gastric cancer. Consequently, our research not only addresses this gap by incorporating machine learning techniques but also reinforces the importance of predictive models for lymphatic spread among patients with advanced gastric cancer. Subgroup analyses and internal cohort validation were also performed to examine the predictive capabilities of the XGB model across different cohorts. Comparisons among these subgroups revealed that the XGB model consistently demonstrated robust predictive performance. In our model, we analyzed the factors influencing lymph node metastasis in gastric cancer patients and discovered that PLR is significantly associated with lymph node metastasis in these patients. This suggests that PLR could serve as a predictive marker for assessing the risk of lymph node metastasis in patients with gastric cancer. This finding is consistent with previous research. Studies by Kwon et al⁴⁴ and Stefan et al⁴⁵ have shown that a higher PLR is associated with poorer prognosis across various types of cancer. The possible biological mechanism is that platelets promote the growth and metastasis of tumor cells within the tumor microenvironment, while lymphocytes play a crucial role in tumor immune surveillance.⁴⁶ Therefore, a higher PLR reflects a state of immune suppression and promotion of tumor metastasis within the tumor microenvironment.And our results also indicate that PLR levels play a crucial role in influencing the prognosis of gastric cancer patients. Subsequently, through transcriptomic analysis and the selection of differentially expressed genes, we identified six genes

—IGFN1, CLEC11A, STC2, TFEC, MUC5AC, and ANOS1—that demonstrated significant predictive value for survival. Although the results of transcriptome studies and the construction of prediction models are in two independent parts, the connection between them cannot be ignored. The genes identified in the transcriptomic analysis not only reveal the underlying biological mechanisms driving lymph node metastasis, but also provide key features to improve the accuracy of the prediction models. This comprehensive approach, combining gene expression profiling with machine learning models, enriches our understanding of gastric cancer progression and highlights the clinical relevance of these markers. We constructed a prognostic model using these six genes and conducted functional enrichment analysis. The analysis revealed enrichment in pathways such as Allograft Rejection, Epithelial Mesenchymal Transition, Interferon Gamma Response, Inflammatory Response, and Interferon Alpha Response. This provides evidence supporting the notion that these pathways may play a role in promoting lymph node metastasis by PLR, thereby influencing the prognosis. This comprehensive approach not only underscores the importance of PLR in gastric cancer prognosis but also highlights the potential prognostic significance of these identified genes at the transcriptomic level. These findings offer new insights into the molecular mechanisms of gastric cancer and provide potential targets for the development of future targeted therapy strategies.

While our study represents the first attempt to predict lymph node metastasis in gastric cancer patients using a machine learning model based on inflammatory indices and peripheral lymphocyte subgroups, acknowledging certain known limitations is crucial. Despite the valuable insights provided by SHAP regarding feature importance, interpreting complex machine learning models, such as XGBoost, may pose challenges. Although our bioinformatics analysis offers a comprehensive risk assessment tool and introduces new molecular markers for predicting gastric cancer lymph node metastasis, we acknowledge the need for further prospective validation and functional experiments to determine the clinical relevance of these molecular markers and elucidate their roles in disease progression. Despite these limitations, we believe that utilizing machine learning and the SHAP framework for predicting gastric cancer lymph node metastasis holds significant clinical application prospects.

In summary, in the context of predicting lymph node metastasis in gastric cancer patients based on machine learning utilizing inflammation indices and peripheral lymphocyte subpopulations, factors such as PLR,BMI, PNI, CD4, DIR, FLR, and INI emerge as significant predictive features. Among these, PLR emerges as the most crucial.(2–5) With ongoing improvements and validation, these predictive models have the potential to assist healthcare professionals in making more informed decisions regarding lymph node metastasis, ultimately enhancing patient treatment outcomes.

Conclusions

The ML model, especially XGBoost, can more accurately predict lymph node metastasis in gastric cancer patients. The combination of XGBoost and SHAP intuitively reflects the impact of different variables on LNM, with PLR being the most crucial risk factor for LNM among inflammatory indices. Additionally, CD4 serves as a good indicator for predicting LNM.

Data Sharing Statement

All data can be obtained from the corresponding author upon request.

Ethics Approval and Consent to Participate

This retrospective study received approval from the institutional review boards at each participating center and adhered to the principles of the 1964 helsinki Declaration and its subsequent revisions. Informed consent was obtained from all patients before treatment. The investigators obtained informed consent from each participant of the study.

Consent for Publication

All authors made significant contributions to the work reported, whether in the conception, study design, execution, acquisition of data, analysis, and interpretation, or in all these areas. Ziyu Zhu and Cong Wang, as the first authors, participated in drafting, revising, and critically reviewing the article. Zhidong Yin and Yingwei Xue, as corresponding authors, were involved in guiding the research and reviewing the manuscript. All authors have given final approval of the

version to be published, agreed on the journal to which the article has been submitted, and are accountable for all aspects of the work.

Author Contributions

Ziyu Zhu and Cong Wang proposed and designed this project, and their contributions to this research are equal. All authors made substantial contributions to data acquisition, data analysis, and interpretation. They collectively agreed to submit it to the current journal, approved the final version for publication, and agreed to be accountable for all aspects of the work.

Funding

This study was supported by Nn10 program of Harbin Medical University Cancer Hospital, China (No. Nn10 PY 2017-03).

Disclosure

The authors declare that they have no competing interests.

References

- 1. Sung H, Ferlay J, Siegel RL. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *Ca a Cancer J Clinicians*. 2021;71(3):209–249. doi:10.3322/caac.21660
- Rice TW, Gress DM, Patil DT, Hofstetter WL, Kelsen DP, Blackstone EH. Cancer of the esophagus and esophagogastric junction-Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *Ca a Cancer J Clinicians*. 2017;67(4):304–317. doi:10.3322/ caac.21399
- 3. Zhou YX, Yang LP, Wang ZX, et al. Lymph node staging systems in patients with gastric cancer treated with D2 resection plus adjuvant chemotherapy. J Cancer. 2018;9(4):660–666. doi:10.7150/jca.22016
- 4. Fukagawa T, Katai H, Mizusawa J, et al. A prospective multi-institutional validity study to evaluate the accuracy of clinical diagnosis of pathological stage III gastric cancer (JCOG1302A). *Gastric Cancer*. 2018;21(1):68–73. doi:10.1007/s10120-017-0701-1
- 5. Kwee RM, Kwee TC. Imaging in assessing lymph node status in gastric cancer. Gastric Cancer. 2009;12(1):6–22. doi:10.1007/s10120-008-0492-5
- 6. Luo M, Lv Y, Guo X, Song H, Su G, Chen B. Value and impact factors of multidetector computed tomography in diagnosis of preoperative lymph node metastasis in gastric cancer: a PRISMA-compliant systematic review and meta-analysis. *Medicine*. 2017;96(33):e7769. doi:10.1097/ MD.0000000000007769
- 7. Jiang M, Wang X, Shan X, et al. Value of multi-slice spiral computed tomography in the diagnosis of metastatic lymph nodes and N-stage of gastric cancer. *J Int Med Res.* 2019;47(1):281–292. doi:10.1177/0300060518800611
- 8. Zhu H, Wang G, Zheng J, et al. Preoperative prediction for lymph node metastasis in early gastric cancer by interpretable machine learning models: a multicenter study. *Surgery*. 2022;171(6):1543–1551. doi:10.1016/j.surg.2021.12.015
- 9. Multhoff G, Molls M, Radons J. Chronic inflammation in cancer development. Front Immunol. 2011;2:98. doi:10.3389/fimmu.2011.00098
- 10. Jaroenlapnopparat A, Bhatia K, Coban SJD. Inflammation and gastric cancer. Diseases. 2022;10:35.
- 11. Zhang X, Zhao W, Yu Y, et al. Clinicopathological and prognostic significance of platelet-lymphocyte ratio (PLR) in gastric cancer: an updated meta-analysis. *World J Surg Oncol.* 2020;18(1):191. doi:10.1186/s12957-020-01952-2
- 12. Lian L, Xia YY, Zhou C, et al. Application of platelet/lymphocyte and neutrophil/lymphocyte ratios in early diagnosis and prognostic prediction in patients with resectable gastric cancer. *Cancer Biomarker*. 2015;15(6):899–907. doi:10.3233/CBM-150534
- 13. Li P, Li H, Ding S, Zhou J, PLR, NLR. LMR and MWR as diagnostic and prognostic markers for laryngeal carcinoma. *Am J Transl Res.* 2022;14 (5):3017–3027.
- 14. Hu C, Bai Y, Li J, et al. Prognostic value of systemic inflammatory factors NLR, LMR, PLR and LDH in penile cancer. *BMC Urol.* 2020;20(1):57. doi:10.1186/s12894-020-00628-z
- 15. Kim SG, Eom BW, Yoon H, Kim YW, Ryu KW. Prognostic Value of Preoperative Systemic Inflammatory Parameters in Advanced Gastric Cancer. *J Clin Med.* 2022;11:5318.
- 16. Choi HS, Ha SY, Kim HM, et al. The prognostic effects of tumor infiltrating regulatory T cells and myeloid derived suppressor cells assessed by multicolor flow cytometry in gastric cancer patients. *Oncotarget*. 2016;7(7):7940–7951. doi:10.18632/oncotarget.6958
- 17. Hirschhorn-Cymerman D, Budhu S, Kitano S, et al. Induction of tumoricidal function in CD4+ T cells is associated with concomitant memory and terminally differentiated phenotype. J Exp Med. 2012;209(11):2113–2126. doi:10.1084/jem.20120532
- 18. Perez-Diez A, Joncker NT, Choi K, et al. CD4 cells can be more efficient at tumor rejection than CD8 cells. *Blood*. 2007;109(12):5346–5354. doi:10.1182/blood-2006-10-051318
- 19. Zhou CM, Wang Y, Yang -J-J, Zhu Y. Predicting postoperative gastric cancer prognosis based on inflammatory factors and machine learning technology. *BMC Med Inform Decis Mak.* 2023;23(1):53. doi:10.1186/s12911-023-02150-2
- Zhou C, Wang Y, Ji MH, Tong J, Yang JJ, Xia H. Predicting Peritoneal Metastasis of Gastric Cancer Patients Based on Machine Learning. *Cancer Control*. 2020;27(1):1073274820968900. doi:10.1177/1073274820968900
- 21. Zhou CM, Xue Q, Wang Y, Tong J, Ji M, Yang JJ. Machine learning to predict the cancer-specific mortality of patients with primary non-metastatic invasive breast cancer. *Surg Today*. 2021;51(5):756–763. doi:10.1007/s00595-020-02170-9
- 22. Rodríguez-Pérez R, Bajorath J. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *J Comput Molecul Des*. 2020;34(10):1013–1026. doi:10.1007/s10822-020-00314-0

- 23. Okada S, Ohzeki M, Taguchi S. Efficient partition of integer optimization problems with one-hot encoding. *Sci Rep.* 2019;9(1):13036. doi:10.1038/ s41598-019-49539-6
- 24. Nick TG, Campbell KM. Logistic regression. Methods Mol Biology. 2007;404:273-301.
- 25. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak.* 2019;19(1):281. doi:10.1186/s12911-019-1004-8
- 26. Noble WS. What is a support vector machine? Nat Biotechnol. 2006;24(12):1565–1567. doi:10.1038/nbt1206-1565
- 27. Salvador-Meneses J, Ruiz-Chavez Z, Garcia-Rodriguez J. Compressed kNN: k-Nearest Neighbors with Data Compression. *Entropy.* 2019;21 (3):234. doi:10.3390/e21030234
- 28. Jiang H, Mao H, Lu H, et al. Machine learning-based models to support decision-making in emergency department triage for patients with suspected cardiovascular disease. Int J Med Inform. 2021;145:104326. doi:10.1016/j.ijmedinf.2020.104326
- Chen T, Guestrin C, XGBoost, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.
- 30. Wang K, Tian J, Zheng C, et al. Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP. *Comput Biol Med.* 2021;137:104813. doi:10.1016/j.compbiomed.2021.104813
- 31. Lundberg SM, Lee S-IJAINIPS. A unified approach to interpreting model predictions. Adv Neural Information Process Syst. 2017;30:4765-4774.
- 32. Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. *Lancet.* 2020;396(10251):635–648. doi:10.1016/S0140-6736(20) 31288-5
- 33. Isik A, Okan I, Firat D, Yilmaz B, Akcakaya A, Sahin M. A new prognostic strategy for gastric carcinoma: albumin level and metastatic lymph node ratio. *Minerva chirurgica*. 2014;69(3):147–153.
- 34. Gong W, Zhao L, Dong Z, et al. After neoadjuvant chemotherapy platelet/lymphocyte ratios negatively correlate with prognosis in gastric cancer patients. J Clin Lab Analy. 2018;32(5):e22364. doi:10.1002/jcla.22364
- 35. Yu L, Lv CY, Yuan AH, Chen W, Wu AW. Significance of the preoperative neutrophil-to-lymphocyte ratio in the prognosis of patients with gastric cancer. *World J Gastroenterol*. 2015;21(20):6280–6286. doi:10.3748/wjg.v21.i20.6280
- 36. Ohwada S, Iino Y, Nakamura S, et al. Peripheral blood T cell subsets as a prognostic factor in gastric cancer. Jpn J Clin Oncol. 1994;24(1):7-11.
- 37. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artif Intell Med. 2001;23(1):89–109. doi:10.1016/ S0933-3657(01)00077-X
- 38. 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
- 39. Yin XY, Pang T, Liu Y, et al. Development and validation of a nomogram for preoperative prediction of lymph node metastasis in early gastric cancer. *World J Surg Oncol.* 2020;18(1):2. doi:10.1186/s12957-019-1778-2
- 40. Kang WZ, Xiong JP, Li Y, et al. A New Scoring System to Predict Lymph Node Metastasis and Prognosis After Surgery for Gastric Cancer. Front Oncol. 2022;12:809931. doi:10.3389/fonc.2022.809931
- 41. Guo CG, Zhao DB, Liu Q, et al. A nomogram to predict lymph node metastasis in patients with early gastric cancer. *Oncotarget*. 2017;8 (7):12203-12210. doi:10.18632/oncotarget.14660
- 42. Yang T, Martinez-Useros J, Liu J, et al. A retrospective analysis based on multiple machine learning models to predict lymph node metastasis in early gastric cancer. *Front Oncol.* 2022;12:1023110. doi:10.3389/fonc.2022.1023110
- 43. Wang H, Gong H, Tang A, Cui Y. Neutrophil/lymphocyte ratio predicts lymph node metastasis in patients with gastric cancer. *Am J Transl Res.* 2023;15(2):1412–1420.
- 44. Kwon HC, Kim SH, Oh SY, et al. Clinical significance of preoperative neutrophil-lymphocyte versus platelet-lymphocyte ratio in patients with operable colorectal cancer. *Biomarkers*. 2012;17(3):216–222. doi:10.3109/1354750X.2012.656705
- 45. Diem S, Schmid S, Krapf M, et al. Neutrophil-to-Lymphocyte ratio (NLR) and Platelet-to-Lymphocyte ratio (PLR) as prognostic markers in patients with non-small cell lung cancer (NSCLC) treated with nivolumab. *Lung Cancer*. 2017;111:176–181. doi:10.1016/j.lungcan.2017.07.024
- 46. Zhu Y, Wei Y, Zhang R, et al. Elevated Platelet Count Appears to Be Causally Associated with Increased Risk of Lung Cancer: a Mendelian Randomization Analysis. *Cancer Epidemiol Biomarker Prevention*. 2019;28(5):935–942. doi:10.1158/1055-9965.EPI-18-0356

Journal of Inflammation Research

Dovepress

Publish your work in this journal

The Journal of Inflammation Research is an international, peer-reviewed open-access journal that welcomes laboratory and clinical findings on the molecular basis, cell biology and pharmacology of inflammation including original research, reviews, symposium reports, hypothesis formation and commentaries on: acute/chronic inflammation; mediators of inflammation; cellular processes; molecular mechanisms; pharmacology and novel anti-inflammatory drugs; clinical conditions involving inflammation. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

Submit your manuscript here: https://www.dovepress.com/journal-of-inflammation-research-journal