



Prediction Model of Ocular Metastases in Gastric Adenocarcinoma: Machine Learning-Based Development and Interpretation Study

Technology in Cancer Research & Treatment
Volume 23: 1-15
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/15330338231219352
journals.sagepub.com/home/tct



Jie Zou, MD¹, Yan-Kun Shen, MD¹ , Shi-Nan Wu, MD^{1,2}, Hong Wei, MD¹, Qing-Jian Li, MD², San Hua Xu, MD¹, Qian Ling, MD¹, Min Kang, MD¹, Zhao-Lin Liu, MD³, Hui Huang, MD¹, Xu Chen, MD⁴, Yi-Xin Wang, MD⁵, Xu-Lin Liao, MD⁶, Gang Tan, MD, PhD³, and Yi Shao, MD, PhD^{1,*} 

Abstract

Background: Although gastric adenocarcinoma (GA) related ocular metastasis (OM) is rare, its occurrence indicates a more severe disease. We aimed to utilize machine learning (ML) to analyze the risk factors of GA-related OM and predict its risks. **Methods:** This is a retrospective cohort study. The clinical data of 3532 GA patients were collected and randomly classified into training and validation sets in a ratio of 7:3. Those with or without OM were classified into OM and non-OM (NOM) groups. Univariate and multivariate logistic regression analyses and least absolute shrinkage and selection operator were conducted. We integrated the variables identified through feature importance ranking and further refined the selection process using forward sequential feature selection based on random forest (RF) algorithm before incorporating them into the ML model. We applied six ML algorithms to construct the predictive GA model. The area under the receiver operating characteristic (ROC) curve indicated the model's predictive ability. Also, we established a network risk calculator based on the best performance model. We used Shapley additive interpretation (SHAP) to identify risk factors and to confirm the interpretability of the black box model. We have de-identified all patient details. **Results:** The ML model, consisting of 13 variables, achieved an optimal predictive performance using the gradient boosting machine (GBM) model, with an impressive area under the curve (AUC) of 0.997 in the test set. Utilizing the SHAP method, we identified crucial factors for OM in GA patients, including LDL, CA724, CEA, AFP, CA125, Hb, CA153, and Ca²⁺. Additionally, we validated the model's reliability through an analysis of two patient cases and developed a functional online web prediction calculator based on the GBM model. **Conclusion:** We used the ML method to establish a risk prediction model for GA-related OM and showed that GBM performed best among the six ML models. The model may identify patients with GA-related OM to provide early and timely treatment.

¹ Department of Ophthalmology, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Jiangxi Branch of National Clinical Research Center for Ocular Disease, Nanchang, Jiangxi, People's Republic of China

² Fujian Provincial Key Laboratory of Ophthalmology and Visual Science, Eye Institute of Xiamen University, School of Medicine, Xiamen University, Xiamen, Fujian, People's Republic of China

³ Department of Ophthalmology, the First Affiliated Hospital of University of South China, Hunan Branch of National Clinical Research Center for Ocular Disease, Hengyan, Hunan Province, People's Republic of China

⁴ Department of Ophthalmology and Visual Sciences, Maastricht University, Maastricht, Limburg Province, Netherlands

⁵ School of Optometry and Vision Sciences, Cardiff University, Cardiff, UK

⁶ Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, People's Republic of China

*Current affiliation: Department of Ophthalmology, Eye & ENT Hospital of Fudan University, Shanghai, China

Jie Zou, Yan-Kun Shen, and Shi-Nan Wu have contributed equally to this work.

Corresponding Authors:

Yi Shao, Department of Ophthalmology, The First Affiliated Hospital of Nanchang University, Nanchang, Jiangxi 330006, P.R. China.
Email: freebee99@163.com

Gang Tan, Department of Ophthalmology, The First Affiliated Hospital of University of South China, Hengyan, Hunan 421001, P.R. China.
Email: tangang99@hotmail.com



Keywords

ocular metastases in gastric adenocarcinoma, machine learning, gradient boosting machine, Shapley additive explanations, risk factor

Received: November 30, 2022; Revised: October 10, 2023; Accepted: November 8, 2023.

Introduction

Globally, gastric adenocarcinoma (GA) is the fifth most common cause of cancer-related death in men and the fourth most common in women.¹ It most often occurs in the distal pyloric region. The disease is asymptomatic, and in areas without an effective GA screening system, most cases are diagnosed at an advanced period,² when the prognosis is generally poor if metastases have formed.³ Hematoxylin and eosin (HE) staining and characteristic immunohistochemistry of GA are shown in Figure 1. GA usually progresses slowly but ultimately develops into multiple metastases, most commonly occurring in the liver and distant lymph nodes, while ocular metastases (OM) are rare.⁴ A case report has shown cases of iris metastasis three years after the diagnosis of gastric cancer,⁵ and a case of eye metastasis eight months after tumor diagnosis, when the patient was receiving systemic chemotherapy, has been reported.⁶ OM of GA mainly manifests as eyelid tumefaction, eye pain and even blindness, with poor quality of life.⁷ To relieve patients' suffering, early monitoring of eye diseases in GA patients is of great significance.

Artificial intelligence (AI) shows excellent potential in multiple medical fields.⁸ Machine learning (ML) is a division of AI and is superior to other techniques in developing predictive

models.⁹ It "learns" from data without direct programming, which signifies that the expression of a particular task advances with incremental data and variables. Recently, ML has shown exciting promise for diagnostic applications. For example, the accuracy of an AI diagnosis system in examining upper gastrointestinal cancer is reportedly over 91.7%.¹⁰ In addition, researchers have proposed that deep learning models can classify gastric diffuse adenocarcinoma with high specificity and help pathologists diagnose the potential of workflow systems.¹¹ ML has also made progress in the treatment of GA and its complications. Researchers have developed a powerful ML method to forecast anastomotic leakage in patients with GA undergoing gastrectomy in real-time, which can guide surgeons' intraoperative decision-making.¹² Moreover, machine learning (ML) can help to judge the prognosis of patients with GA. For example, lymphatic vessel invasion and perineural invasion are associated with poor prognosis of GA. ML-based CT texture analysis may predict lymphatic invasion and perineural invasion of tubular GA.¹³

Though numerous previous studies have demonstrated the strengths of AI in the diagnosis, treatment, and prognosis of GA, there is still no model to forecast the risks of OM in patients with GA. This research aims to identify potential

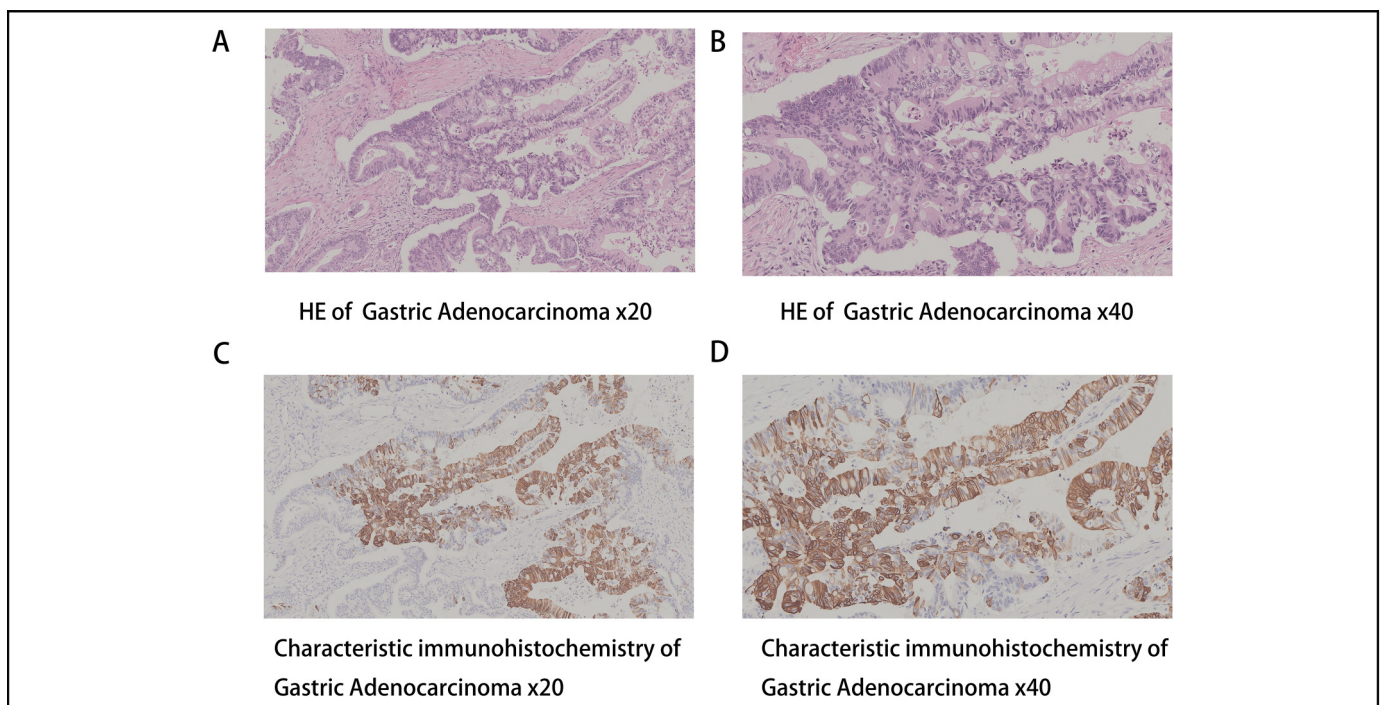


Figure 1. He and characteristic immunohistochemistry of gastric adenocarcinoma. Notes: Immunohistochemical staining showed CK7 as the immunomarker in (C) and CK20 as the immunomarker in (D). Abbreviation: HE, Hematoxylin eosin staining.

biomarkers of GA OM and construct risk prediction models via demographic and serological indicators in a large sample of GA patients. In addition, by comparing ML models, an optimal model was chosen, and a web calculator was developed to realize personalized prediction of GA eye metastasis, thereby further improving the prognosis of GA patients.

Methods

Study Design

This retrospective cohort study included 3532 patients diagnosed with GA from June 2003 to May 2019. This study was approved by the Medical Research Ethics Committee of the First Affiliated Hospital of Nanchang University, and the approved number was cdyfy20170411. In this study, other types of gastric cancer were excluded by pathological examination, including signet ring cell carcinoma, adenosquamous carcinoma, myeloid carcinoma, and undifferentiated cell carcinoma. Patients with unknown information were culled. Patients were randomly classified into a training set and a validation set at a ratio of 7:3. OM of GA can be divided into choroidal metastasis and retinal metastasis. During choroidal metastasis, fundus fluorescein angiography showed that the tumor components dominated by cancer cells portrayed diffuse fluorescence consistent with the size of the tumor in the arterial phase and early venous phase due to choroidal fluorescence shielding. Optical coherence tomography portrayed choroidal uplift lesions with subretinal effusion. B-ultrasound portrayed a flat bulge tumor, and Doppler ultrasound showed blood flow in the tumor. When retinal metastasis occurred, high-resolution thin-layer scanning manifested calcification. When the tumor invaded the outside of the eyeball, it spread along the optic nerve and invaded the brain, showing optic nerve thickening and orbital or intracranial mass. Magnetic resonance imaging (MRI) showed irregular soft tissue masses in the eyeball. T1WI showed a slightly higher signal, and T2WI showed an unevenly low signal than normal vitreous. When accompanied by a retinal detachment, MRI showed the nature of subretinal effusion, manifested as water signal change and equal signal with the vitreous. All the patients underwent imaging diagnosis to exclude the OM and were classified as non-OM (NOM). We use synthetic minority oversampling technique (SMOTE) on the valid test set to generalize the OM group sample size¹⁴ (Figure 2). SMOTE has a good balance performance index, it can reduce the overfitting of the model, and improve the accuracy, sensitivity, and specificity of the test set.¹⁵ All participants understood the purpose and content of our study and signed informed consent. The researchers deleted the subjects' private information for this study. The reporting of this study conformed to STROBE guidelines.¹⁶

Data Collection

Clinical indicators were gathered from patient documents, including (a) demographic data: gender, age, histological grade, and treatment measures and (b) diagnostic serological

test results: Hb, calcium, alkaline phosphatase (ALP), total cholesterol (TC), triglyceride (TG), high-density lipoprotein (HDL), LDL, apolipoprotein A1 (ApoA1), Apolipoprotein B (ApoB), AFP, CA724, CA125, CA153, CA199, CEA, cytokeratin-19 fragment (CYFRA21-1), and lipoprotein (a) (Lp (a)).

Statistical Analysis

Statistical analysis was conducted using Python (version 3.8, Python Software Foundation) and R software (version 4.0.2). In Python, the training set data were used to build the model, and the validation set data were used to validate and evaluate the model. An independent samples t-test was used for normally distributed continuous numerical data; for non-normally distributed continuous data, a Mann–Whitney U test was applied. For categorical count data, a chi-square test was used. Univariate and multivariate logistic regression were used to determine the risk factors for OM in patients with GA. The selection of ML feature variables is based on the feature importance of the random forest (RF) algorithm and forward sequential feature selection. The Python programming language (Version 3.8) was also used to develop and evaluate ML models and design network calculators. Shapley additive interpretation (SHAP) was implemented for model interpretation using the Python SHAP package. A p-value < 0.05 was considered statistically significant, and all logistic regression analyses were performed using a 95% confidence interval (CI).

Data Preprocessing and Feature Engineering

The criteria for selecting feature variables in subsequent ML algorithms were: feature importance ranking and order positive selection based on RF algorithm.¹⁶ On this basis, all the feature variables were included in the feature importance ranking and screening. Then, these variables were input into the hierarchical clustering algorithm to remove the characteristic variables with multicollinearity. Then, we reordered the pre-selected variables. Finally, the optimal ML feature variable was determined according to when the AUC of the receiver operating characteristic (ROC) reached a stable value. We selected the first 13 feature variables for subsequent ML algorithm development, because after the 14th iteration, AUC did not improve significantly (Figure 4A). Subsequently, the feature variables of ML model construction include LDL, CEA, CA724, CA125, TC, Ca²⁺, HDL, AFP, CA153, CA199, TG, Hb, and ALP. We used the SHAP package to establish a risk factor variable importance ranking for patients with GA. For each patient, the SHAP model can produce a predicted value, and the total or mean of the absolute Shapley values of all samples is the integrated significance score of the feature. In addition, the SHAP method demonstrates the positive or negative influence of each eigenvalue on the prediction results, similar to the coefficient value in logistic regression. A positive SHAP value signifies that the corresponding characteristic has a higher probability of recurrence risk, and a negative value signifies a lower risk.¹⁷

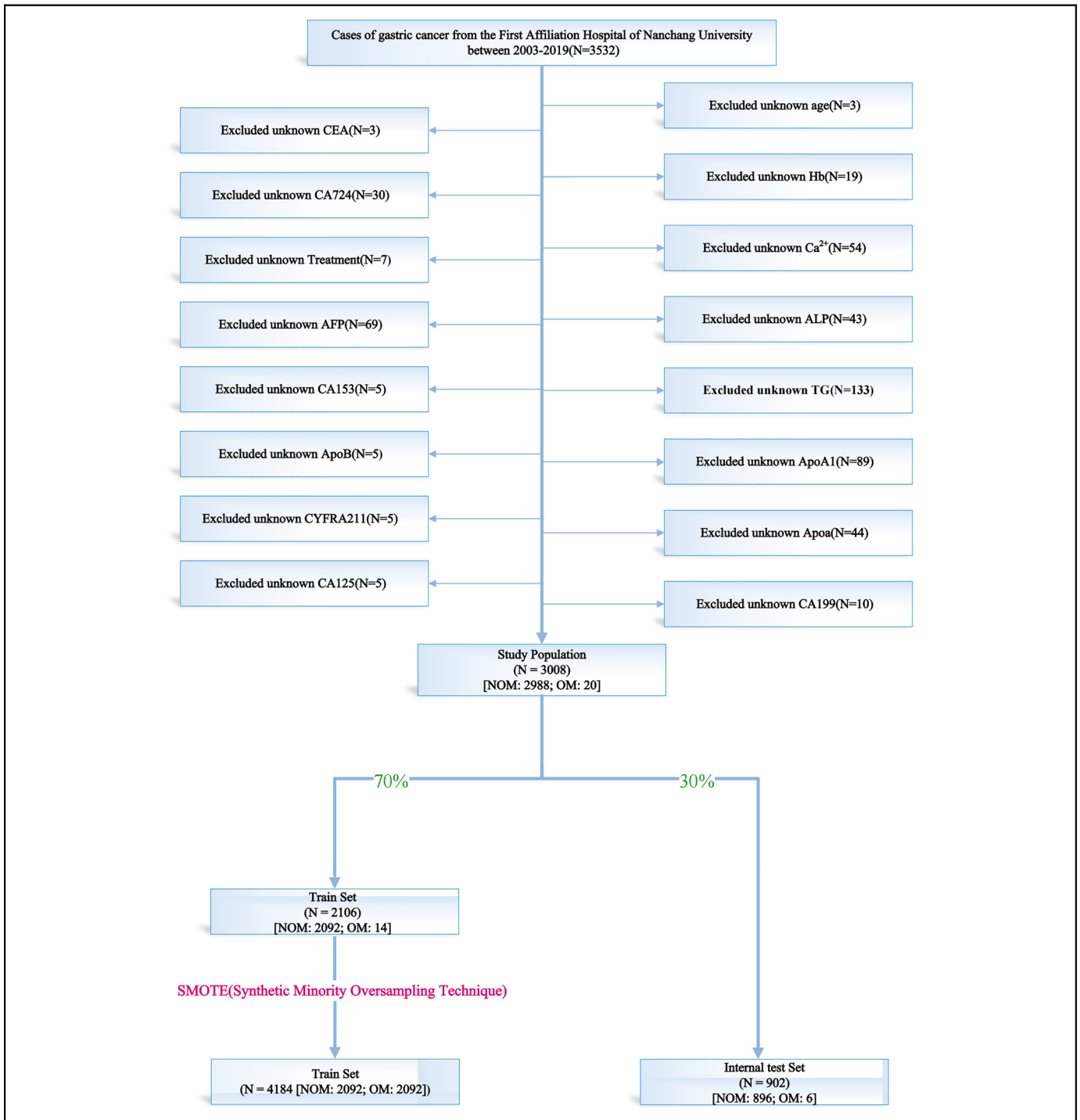


Figure 2. Summary of patients inclusion. Abbreviations: HB, Hemoglobin; Ca^{2+} , calcium; ALP, alkaline phosphatase; TG, triglyceride; HDL, high-density lipoprotein; LDL, low-density lipoprotein; ApoA1, apolipoprotein A1; ApoB, apolipoprotein B; AFP, alpha fetoprotein; CA724, carbohydrate antigen-724; CEA, carcinoembryonic antigen; CA125, carbohydrate antigen-125; CA153, carbohydrate antigen-153; CA199, carbohydrate antigen-199; CYFRA21-1, cytokeratin 19 fragment antigen21-1; Lp (a), lipoprotein (a); NOM, no ocular metastasis; OM, ocular metastasis.

Model Establishment

All algorithm models were built using scikit-learn (Version 0.24.2). In this research, we utilized six ML algorithms: multi-layer perceptron (MLP)¹⁸; adaboost (AB) model¹⁹; bagging

classification (BAG) model²⁰; logistic regression (LR)²¹; gradient boosting machine (GBM)²²; and extreme gradient boosting (XGB) model.²³ The ML algorithm was trained and modulated to forecast OM in patients with GA. The model's hyperparameters were adjusted using the random search approach in

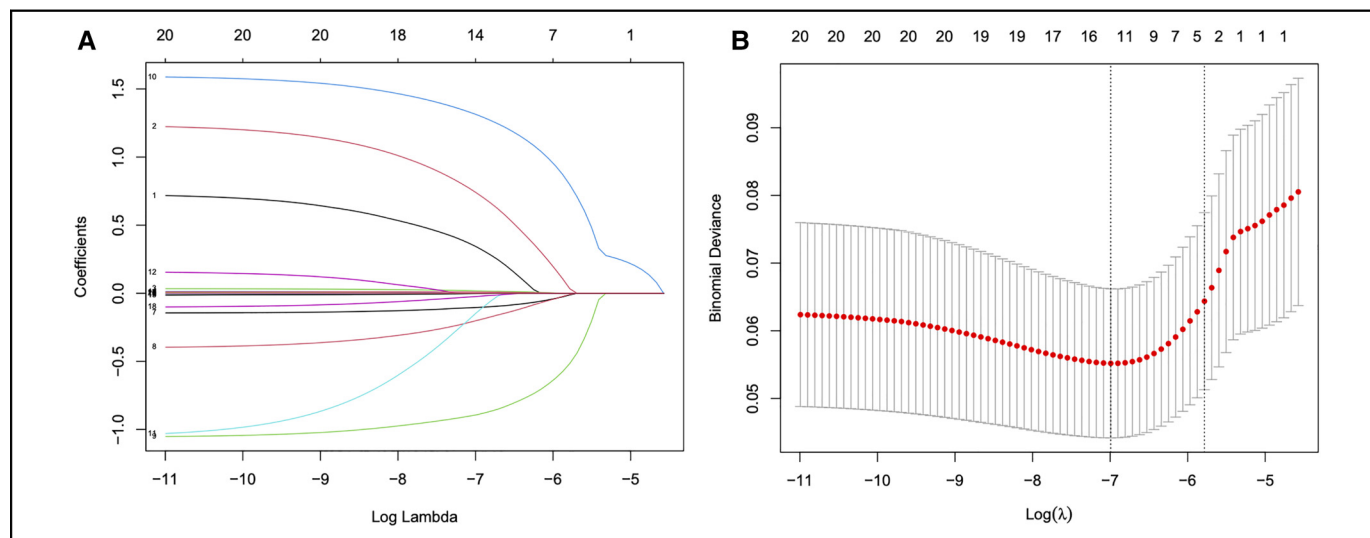


Figure 3. (A) Plot for LASSO regression coefficients. (B) Cross validation plot. Abbreviations: LASSO, least absolute shrinkage and selection operator.

scikit-learn. By comparing the performance of different ML models on the training and testing datasets, metrics such as AUC value, precision–recall (PR) curve, confusion matrix, sensitivity, specificity, and F1 score were evaluated. Finally, we chose the most excellent manifesting model to construct a web calculator.

Results

Demographic Baseline Data

After the screening, the data of 3008 GA patients were included. Among them, during the study period, 20 patients had OM, and 2988 patients had NOM. There were significant differences in pathological type, LDL, CA724, and CEA between the OM and NOM groups ($p < 0.05$). The demographic and clinicopathological features of the above patients are detailed in Table 1.

Univariate Analysis, Multivariate Logistic Regression, and Least Absolute Shrinkage and Selection Operator (LASSO) Regression

By establishing a univariate logistic regression model, we screened variables with $p < 0.05$ in univariate analysis for multivariate logistic regression analysis to determine the risk factors for OM in patients with GA. In univariate logistic regression, LDL was the risk factor for postoperative recurrence. Multivariate logistic regression analysis also showed that LDL was the independent risk factors for OM of GA (Table 2). LASSO regression was then used to screen out the pathological type, TG, TC, HDL, and LDL, and these were included in the six ML model characteristic variables (Figure 3).

Model Performance

We established six different ML models of MLP, AB, BAG, LR, GBM, and XGB to evaluate the risk probability and related accuracy of OM in patients with GA. The GBM model performed best in the training results, with an AUC of 0.997, an accuracy of 0.989, a sensitivity of 0.556, and a specificity of 0.995 (Table 3). The GBM algorithm for predicting OM in GA patients had an AUC value of 1 in the training set and a performance of 0.950 in the internal test set (Figure 4B). According to the PR curve results of the training set and the test set, the AUC of the GBM model was 1 in the training set (Figure 5A) and 0.747 in the test set (Figure 5B). In addition, we draw the confusion matrix of the model, and all predictions were correct in the training set after SMOTE processing (Figure 6A). In the original distribution of the test set, there were 744 correct predictions and 8 incorrect predictions, with an accuracy rate of 0.989 (Figure 6B). We conducted ROC analysis on imbalanced data using various ML methods. After using SMOTE to balance the data, the test set performed better without significant overfitting (Figure 6C; Table 3). The best ML model GBM was used and internal five-fold cross-validation was performed. The average AUC value was 0.99 and the standard deviation was 0 (Figure 6D). For the above six ML models, we also draw radar maps to evaluate the performance of different models. Compared with other ML models, GBM had the best value in the evaluation of F1-score, accuracy and AUC value (Figure 6E).

Importance of Characteristic Variables

We used the SHAP library based on GBM to establish a risk factor model for OM in patients with GA (Figure 7). The results of variable importance ranking showed that LDL, CA724, CEA, AFP, CA125, Hb, CA153, and Ca^{2+} were

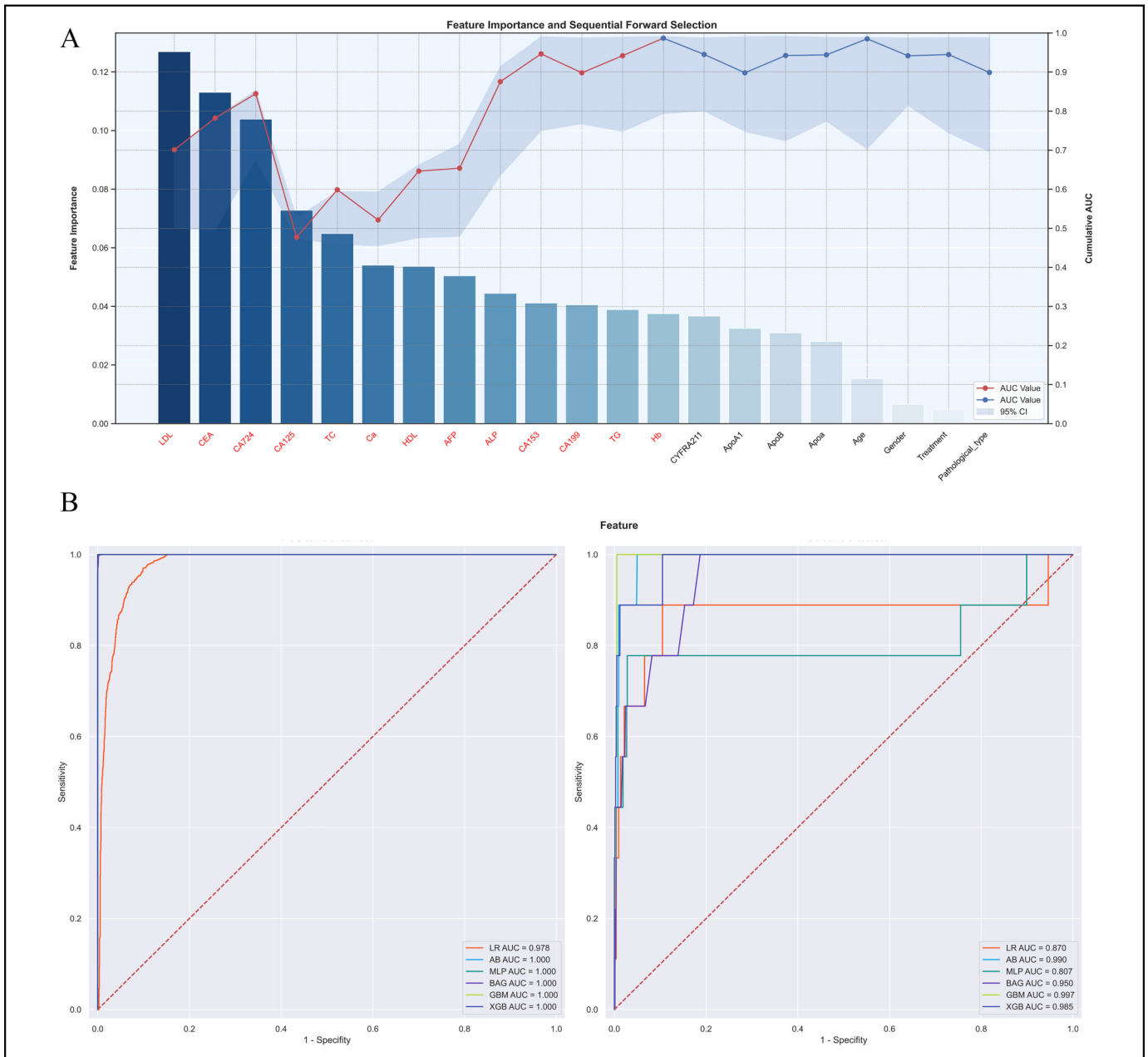


Figure 4. Machine learning feature variable extraction and ROC curves for both the training and testing sets. Notes: (A) displayed feature importance ranking and forward sequential feature selection based on random forests, with features selected highlighted in red font. (B) depicted the ROC curve results for the training and testing sets under different ML algorithms. Notably, the GBM model achieves an AUC value of 1 in the training set and 0.997 in the testing set. Abbreviations: AUC, the area under the curve; AB, adaptive boosting; MLP, multilayer perceptron; BAG, bootstrapped aggregating; LR, logistic regression; GBM, gradient boosting machine; XGB, extreme gradient boost.

important factors affecting OM in GA patients. In Figure 7A, the red highlighted factor is an important risk factor for OM, and the blue highlighted factor is a protective factor against OM. Correspondingly, in the samples used in this study, Figure 7B had a violin plot depicting the importance ranking of each sample feature. In addition, according to the SHAP value, we selected two subjects, including members of the OM group and the NOM group. In the OM group, TC=0.95,

CA199 = 22.99, Ca²⁺ = 2.25, AFP = 3.42, LDL = 3.20, HDL = 1.62 were important risk factors for recurrence. On the contrary, CA724 = 0.93, CEA = 1.74 and other factors were important protective factors for preventing recurrence (Figure 7C). In NOM samples, TC = 0.82, AFP = 5.19, HDL = 1.59 were important risk factors for recurrence, while LDL = 1.58, CA724 = 3.52, CEA = 2.93 were important protective factors for recurrence (Figure 7D).

Table 1. Demographic and Clinicopathological Characteristics of Patients.

Variables	Total (n = 3008)	0 (n = 2988)	1 (n = 20)	P value
Gender, n (%)				0.119
Female	1021 (34)	1018 (34)	3 (15)	
Male	1987 (66)	1970 (66)	17 (85)	
Treatment, n (%)				0.079
Palliative surgery	91 (3)	91 (3)	0 (0)	
Radical gastrectomy	591 (20)	590 (20)	1 (5)	
Chemotherapy	299 (10)	299 (10)	0 (0)	
Support treatment	186 (6)	186 (6)	0 (0)	
Comprehensive treatment	1841 (61)	1822 (61)	19 (95)	
Pathological type, n (%)				<0.001*
Highly differentiated	63 (2)	63 (2)	0 (0)	
Moderately differentiated	1061 (35)	1061 (36)	0 (0)	
Poorly differentiated	1599 (53)	1580 (53)	19 (95)	
Other types	285 (9)	284 (10)	1 (5)	
Age, Median (Q1,Q3)	57 (47, 65)	56 (47, 65)	60 (51.75, 71.25)	0.064
Hb, Median (Q1,Q3)	118 (103, 129)	118 (103, 129)	122.5 (113, 133)	0.305
Ca ²⁺ , Median (Q1,Q3)	2.25 (2.09, 2.38)	2.24 (2.09, 2.38)	2.32 (2.15, 2.4)	0.376
ALP, Median (Q1,Q3)	82 (63, 115)	82 (64, 115)	72 (60, 115.75)	0.567
TG, Median (Q1,Q3)	4.52 (3.71, 5.43)	4.52 (3.71, 5.43)	4.78 (3.87, 5.74)	0.464
TC, Median (Q1,Q3)	1.25 (0.86, 1.96)	1.25 (0.86, 1.96)	1.19 (0.84, 1.85)	0.809
HDL, Median (Q1,Q3)	1.33 (1.07, 1.77)	1.33 (1.07, 1.77)	1.46 (1.36, 1.65)	0.083
LDL, Median (Q1,Q3)	2.75 (2.1, 3.45)	2.74 (2.1, 3.41)	5.4 (4.54, 6.38)	<0.001*
ApoA1, Median (Q1,Q3)	1.5 (1.29, 1.92)	1.5 (1.29, 1.92)	1.43 (1.35, 1.6)	0.259
ApoB, Median (Q1,Q3)	0.93 (0.74, 1.26)	0.93 (0.74, 1.26)	1 (0.82, 1.33)	0.453
AFP, Median (Q1,Q3)	3.25 (2.05, 4.5)	3.25 (2.05, 4.49)	3.16 (2, 5.46)	0.711
CA724, Median (Q1,Q3)	5 (2.32, 10.14)	5 (2.32, 9.95)	16.01 (14.56, 18.67)	<0.001*
CEA, Median (Q1,Q3)	2.34 (1.35, 4.4)	2.33 (1.34, 4.27)	13.22 (12.35, 16.23)	<0.001*
CA199, Median (Q1,Q3)	10.44 (6.55, 25.94)	10.43 (6.55, 26.03)	10.89 (7.32, 19.88)	0.999
CA125, Median (Q1,Q3)	13.63 (8.44, 31.51)	13.63 (8.44, 31.49)	15.13 (10.53, 55.55)	0.586
CYFRA211, Median (Q1,Q3)	3.58 (2.43, 5.49)	3.59 (2.43, 5.56)	3.42 (2.19, 4.23)	0.111
CA153, Median (Q1,Q3)	12.15 (8, 16.59)	12.2 (8, 16.63)	10.35 (7.15, 13.43)	0.259
Apoa, Median (Q1,Q3)	182 (118, 262)	182 (117, 262)	191.5 (151.75, 230.75)	0.404

* $p < 0.05$.

Abbreviations: HB, Hemoglobin; Ca²⁺, calcium; ALP, alkaline phosphatase; TC, total cholesterol; TG, triglyceride; HDL, high-density lipoprotein; LDL, low-density lipoprotein; ApoA1, apolipoprotein A1; ApoB, apolipoprotein B; AFP, alpha fetoprotein; CA724, carbohydrate antigen-724; CEA, carcinoembryonic antigen; CA125, carbohydrate antigen-125; CA153, carbohydrate antigen-153; CA199, carbohydrate antigen-199; CYFRA21-1, cytokeratin 19 fragment antigen-21-1; Lp (a), lipoprotein (a).

Web Page Calculator

Based on the GBM model, an optimal predictive performance ML algorithm, we developed the above web predictor to forecast the risk of OM in patients with GA. A probability prediction for the OM of GA was made using variable settings in the sidebar of the website (<https://weicancer2-nnnjun76tn7vtkya5glyua.streamlit.app/> Figure 8).

Discussion

For the first time, the study used a variety of ML algorithms to predict OM in patients with GA, obtained a GBM model that can be used to predict OM of GA clinically, and explained the model. The GBM model is a commonly used algorithm with sound recognition and fast classification speed. It can make effective and accurate predictions for linear and nonlinear characteristic variables.²⁴ Subsequently, we designed a network

risk calculator based on the GBM model to estimate the probability risk of OM in patients with GA and help clinicians develop targeted preventive measures to elevate the prognosis and life quality.

With the development of diagnosis and treatment technology, the occurrence and mortality of gastric cancer have declined. Despite this, gastric cancer maintains the most common cause of death globally, this is closely related to its asymptomatic course and delayed diagnosis.²⁵ Approximately 90-97% of gastric cancer is adenocarcinoma, which can be histologically divided into intestinal and diffuse.²⁶ Gastric cancer cells can migrate to the liver, lungs, adrenal glands, lymph nodes, peritoneum, bones, brain and eyes through blood vessels and lymphatic vessels.²⁷ The process of GA metastasis is closely related to tumor phenotype and follows multiple steps, the most important of which are proteolytic activity, migration, adhesion, proliferation, and neovascularization.²⁸ The eye has the most plentiful blood flow of any organ in the

Table 2. Single Factor Analysis and Multifactor Logistic Regression.

Characteristics	Category	Univariate analysis		Multivariate analysis	
		OR (95% CI)	P value	OR (95% CI)	P value
Gender	Female	Ref.	Ref.	Ref.	Ref.
	Male	2.928 (0.856-10.015)	0.087	\	\
Pathological type	Highly differentiated	Ref	Ref	Ref	Ref
	Moderately differentiated	1 (0-Inf)	1	\	\
	Poorly differentiated	27 933 574.298 (0-Inf)	0.996	\	\
	Other types	8 179 215.602 (0-Inf)	0.997	\	\
Age	\	1.036 (0.999-1.074)	0.057	\	\
LDL	\	1.473 (1.295-1.675)	<0.001*	1.53 (1.326-1.764)	<0.001*
ApoA1	\	0.43 (0.157-1.176)	0.1	\	\
CYFRA211	\	0.829 (0.667-1.032)	0.093	\	\
TG	\	1.08 (0.852-1.369)	0.526	\	\
Hb	\	1.009 (0.988-1.03)	0.42	\	\
Ca ²⁺	\	1.033 (0.23-4.635)	0.966	\	\
ALP	\	0.999 (0.995-1.004)	0.805	\	\
CEA	\	1.002 (0.996-1.008)	0.512	\	\
TC	\	0.95 (0.68-1.327)	0.764	\	\
HDL	\	1.082 (0.823-1.422)	0.571	\	\
CA724	\	1.007 (0.999-1.015)	0.102	\	\
CA153	\	0.992 (0.956-1.029)	0.656	\	\
ApoB	\	1.246 (0.843-1.843)	0.27	\	\
AFP	\	0.995 (0.93-1.064)	0.883	\	\
CA199	\	0.997 (0.99-1.005)	0.466	\	\
CA125	\	1 (0.997-1.003)	0.979	\	\
Apo(a)	\	1.001 (0.998-1.003)	0.643	\	\

**p* < 0.05.

Abbreviations: Hb, Hemoglobin; Ca²⁺, calcium; ALP, alkaline phosphatase; TC, total cholesterol ; TG, triglyceride; HDL, high-density lipoprotein; LDL, low-density lipoprotein; ApoA1, apolipoprotein A1; ApoB, apolipoprotein B; AFP, alpha fetoprotein; CA724, carbohydrate antigen-724; CEA, carcinoembryonic antigen; CA125, carbohydrate antigen-125; CA153, carbohydrate antigen-153; CA199, carbohydrate antigen-199; CYFRA21-1, cytokeratin 19 fragment antigen21-1; Lp (a), lipoprotein (a).

Table 3. Comparison of indicators of six machine learning methods in test set.

Training	Model	F1	AUC	Accuracy	Sensitivity	Specificity
IBT	AB	0.525	0.592	0.891	0.556	0.895
IBT	LR	0.597	0.779	0.989	0.111	1.000
IBT	BAG	0.497	0.731	0.988	0.000	1.000
IBT	MLP	0.497	0.624	0.988	0.000	1.000
IBT	GBM	0.764	0.971	0.991	0.444	0.997
IBT	XGB	0.679	0.929	0.991	0.222	1.000
BT	AB	0.747	0.99	0.989	0.444	0.996
BT	LR	0.582	0.87	0.927	0.778	0.929
BT	BAG	0.658	0.95	0.968	0.667	0.972
BT	MLP	0.696	0.807	0.984	0.444	0.991
BT	GBM	0.775	0.997	0.989	0.556	0.995
BT	XGB	0.756	0.985	0.976	0.556	0.996

Abbreviations: ML, machine learning; AUC, area under the curve; AB, adaptive boosting; LR, logistic regression; XGB, extreme gradient boosting; BAG, bootstrapped aggregating; MLP, multilayer perceptron; GBM, gradient boosting machine; IBT: Imbalanced Training, BT: Balanced Training.

Note: Comparing the evaluation metrics on the test set between training with and without SMOTE technology for balancing the training set, it can be observed that balancing the training set with SMOTE leads to a more significant improvement in the model's performance on the test set. Furthermore, it further reduces the occurrence of overfitting.

body. According to the "seed and soil" theory proposed by Paget, the colonization of metastatic tumors primarily relies on an appropriate microenvironment and sufficient blood

flow.²⁹ In addition, Duke Elder and Perkins first suggested that the short posterior ciliary artery, with multiple branches and abundant peripheral blood vessels, facilitates the transport

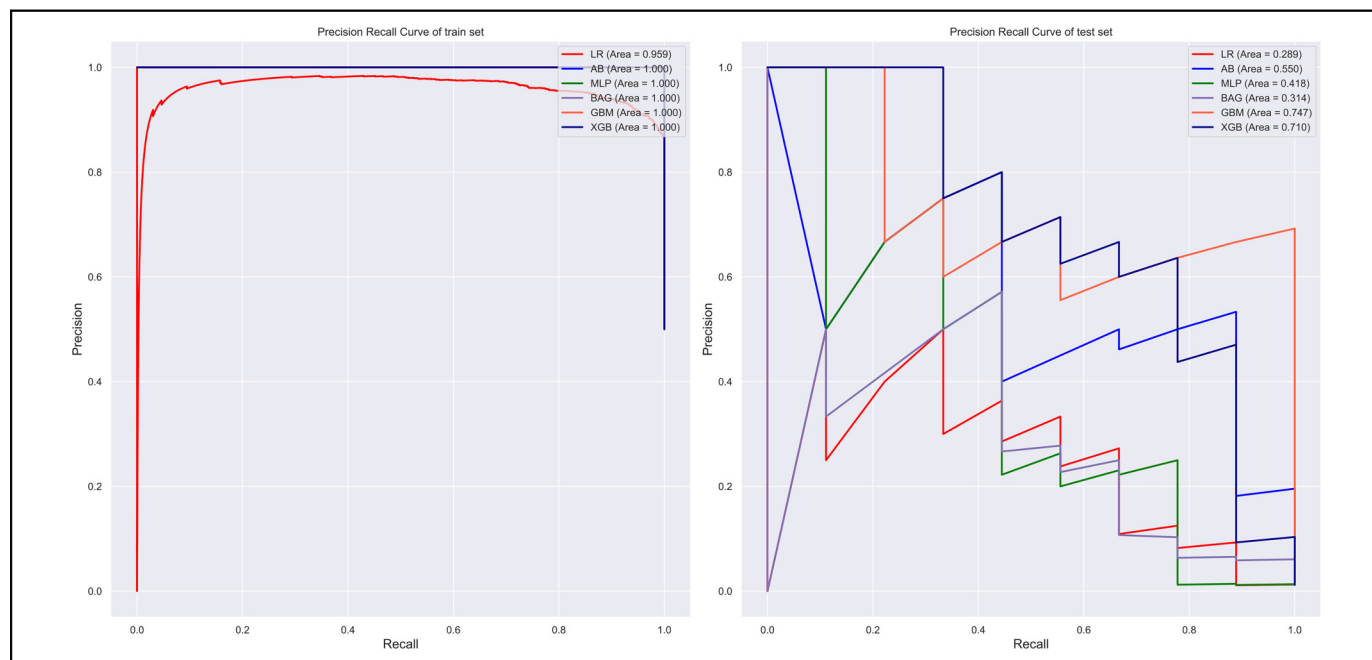


Figure 5. Precision–Recall curves for six different ML algorithms on both the training and testing sets. Notes: (A) displayed the Precision-Recall (PR) curve for the training set, GBM performed the best in the training set with an AUC of 1. (B) showed the PR curve for the testing set, GBM performed the best in the testing set with an AUC of 0.747. Abbreviations: AB, adaptive boosting; MLP, multilayer perceptron; BAG, bootstrapped aggregating; LR, logistic regression; GBM, gradient boosting machine; XGB, extreme gradient boost.

of tumor emboli to the posterior uvea.³⁰ This explains why the metastases impact the choroid more than the ciliary body or sclera.³¹ The clinical manifestations of OM of GA include blurred vision, visual loss, flashes and floaters, metamorphopsia, and diplopia, among which metamorphopsia and vision loss are the most apparent.³² When gastric cancer patients develop OM, the average survival time reduces from 25.4 months at initial diagnosis to only 3.3 months.³³ Therefore, early examination of OM is highly significant for the prognosis of patients with gastric cancer but remains challenging clinically. The application of imaging technology is limited by high expense and low sensitivity and specificity. Thus, low-cost and expedient serum tumor markers may be extensively utilized.

A meta-analysis has shown that a high-fat diet positively correlated with gastric cancer incidence.³⁴ In addition, evidence increasingly shows that hyperlipidemia is positively correlated with the metastasis of many types of cancer, including esophageal and gastric cancers.^{35,36} Hypercholesterolemia is dyslipidemia characterized by elevated levels of LDL cholesterol in the plasma. LDL is readily oxidized by reactive oxygen species, thereby enhancing lipid peroxidation in tissues.³⁷ In addition, evidence shows that oxidized low-density lipoprotein (ox-LDL) is a common pathogenic element in cancer metastasis.³⁸ For example, one study showed that an increase in oxLDL was linked with the development of cancer by binding lectin-like oxLDL receptor (LOX).³⁹ Recent studies have shown that oxLDL may up-regulate the expression of vascular endothelial growth factor-C and increase the secretion of

gastric cancer cells through the LOX-1/nuclear factor kappa-B signaling pathway. In addition, the knockout of LOX-1 in HGC-27 gastric cancer cells with interference fragments and specific inhibitors can restrain the above functions of oxLDL.⁴⁰ The present study is consistent with the above findings, since LDL in patients with OM was significantly higher than in the NOM group. Our results show that LDL may be utilized as an independent risk factor for the early prediction of OM. Our study also shows GA patients may have malignant metastasis by absorbing cholesterol. Low HDL levels are positively correlated with increased cancer risk. In patients with rectal cancer, HDL levels were positively correlated with increased overall survival.⁴¹ At the same time, preoperative serum HDL levels in patients with gallbladder cancer are also closely related to distant metastasis.⁴² This is consistent with our study that HDL level is an indicator of reducing OM in GA patients.

CA724 is a molecular glycoprotein, mainly distributed in the stomach, breast, pancreas, and ovary. It is a tumor marker that is mainly used to detect GA and various gastrointestinal tumors.⁴³ CA724 was considered to be associated with *Helicobacter pylori* infection in China.⁴⁴ Another study suggested that the early stage of CA724 was related to vascular invasion of GA and a decrease in CA724 may lead to a better prognosis.⁴⁵ These findings suggest that CA724 can be used to evaluate distant metastasis in GA patients. CA125 is a glycoprotein that was detected as an epithelial ovarian cancer antigen in 1983 and binds to the monoclonal antibody OC125.⁴⁶ CA125 can also be detected in the serum of patients with gastrointestinal cancer. Most studies on the relationship between CA 125

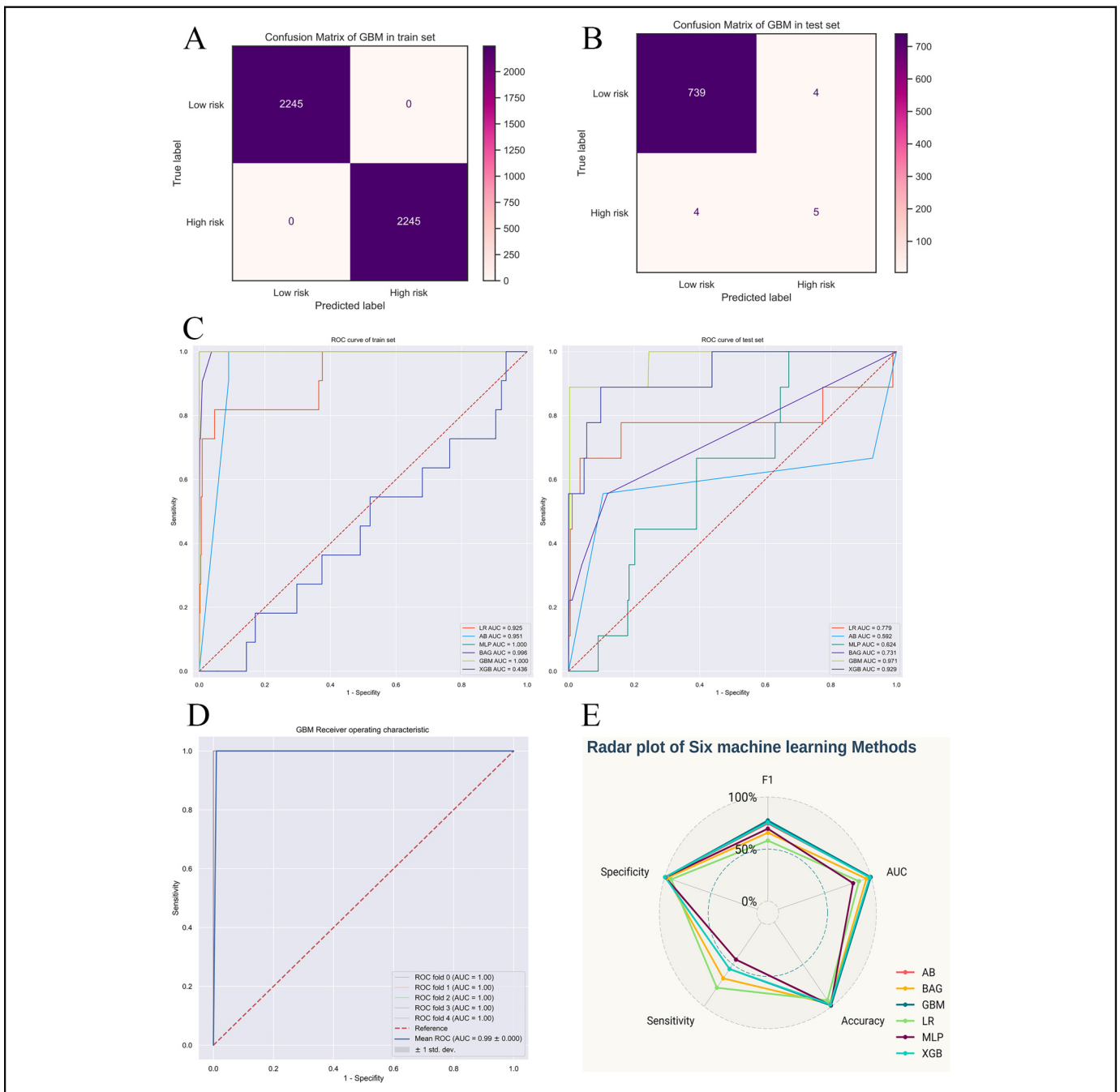


Figure 6. The confusion matrices for the training and testing sets, the five-fold cross-validated ROC curve for the best ML model GBM, and a comparative radar chart of different ML methods under various evaluation metrics. Notes: (A, B) displayed the confusion matrices for the training and testing sets. In the testing set's confusion matrix, the accuracy was observed to be 0.989. (C) showed the test set performed better without significant overfitting after using SMOTE to balance the data. (D) showed the five-fold cross-validated ROC curve results for the best ML model GBM, with an AUC of 0.99 ± 0.00 . (E) presented a radar chart visualization of different ML algorithms under various evaluation metrics. In this chart, the GBM algorithm exhibits the best performance in terms of F1-score, accuracy and AUC value. Abbreviations: GBM, gradient boosting machine; ROC, receiver operating characteristic; AUC, the area under the ROC curve.

and GA have shown that elevated serum CA 125 is associated with peritoneal metastasis.⁴⁷ CA125 usually plays a key role in the shedding and adhesion of GA cells from primary lesions to adjacent tissues, and is closely related to the invasion and metastasis of tumors.⁴⁸ In our study, the results also suggested

that carbohydrate antigen family was a factor affecting OM in GA patients. CEA is the most commonly used tumor marker for GA.⁴⁹ In one study, the level of CEA in peritoneal fluid has been shown to be a reliable marker for early metastasis of GA.⁵⁰ In addition to the level of CEA in peritoneal fluid,

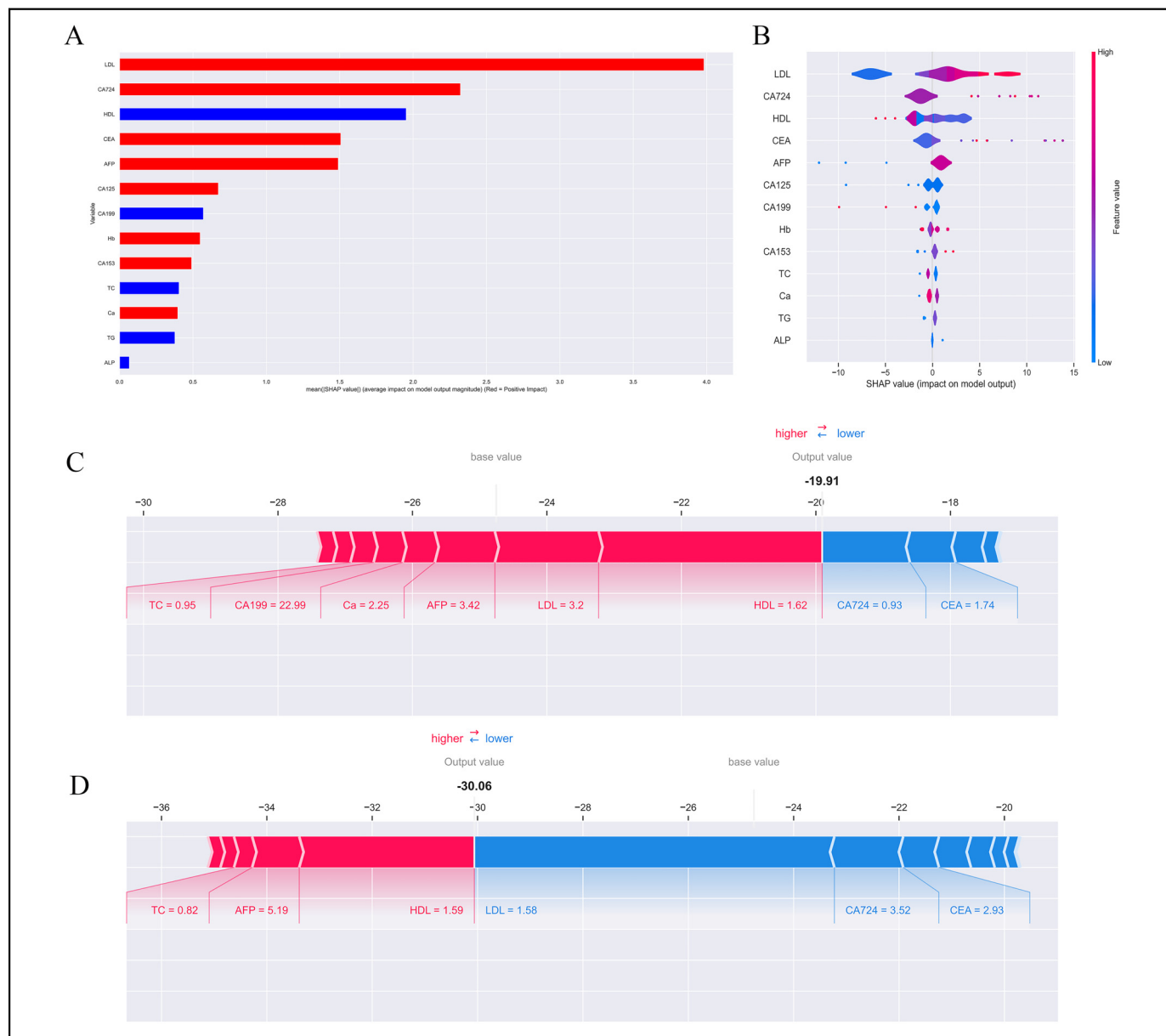


Figure 7. SHAP summary plot and SHAP explanation of two patients. Notes: (A)(B) displayed the importance ranking of feature variables based on the GBM model. Red represented variables that are risk factors for postoperative recurrence in hepatocellular carcinoma, while blue represented variables that act as protective factors. (C) presented a high-risk SHAP interpretation model for postoperative recurrence in GA patients with OM. (D) showed a low-risk SHAP interpretation model for postoperative recurrence in the same patient group. Abbreviation: SHAP, Shapley additive explanations; GBM, gradient boosting machine; Hb, Hemoglobin; Ca²⁺, calcium; ALP, alkaline phosphatase; TC, total cholesterol; TG, triglyceride; HDL, high-density lipoprotein; LDL, low-density lipoprotein; AFP, alpha fetoprotein; CA724, carbohydrate antigen-724; CEA, carcinoembryonic antigen; CA125, carbohydrate antigen-125; CA153, carbohydrate antigen-153; CA199, carbohydrate antigen-199.

serum CEA has been shown to be associated with hematogenous and lymphatic metastasis of GA. The results of meta-analysis also showed that pre-treatment serum CEA concentration >5 ng / mL was significantly associated with GA metastasis.⁵¹ In addition, the combination of CEA with CA724, CA199, and CA125 can improve the diagnostic ability of GA metastasis.⁵² AFP is an oncoprotein with a glycoprotein structure. In clinical practice, it may be elevated in many organs, especially in

GA.⁵³ Studies have found that compared with AFP-negative GA, AFP-positive GA has higher proliferation activity, lower apoptosis, and abundant neovascularization, more prone to liver and lymph node metastasis, and poor prognosis.⁵⁴ He et al's GA study on high AFP showed that serum AFP level was a prognostic factor for overall survival, which was highly significant.⁵⁵ In our study, abnormal elevation of AFP was also associated with OM in GA patients.

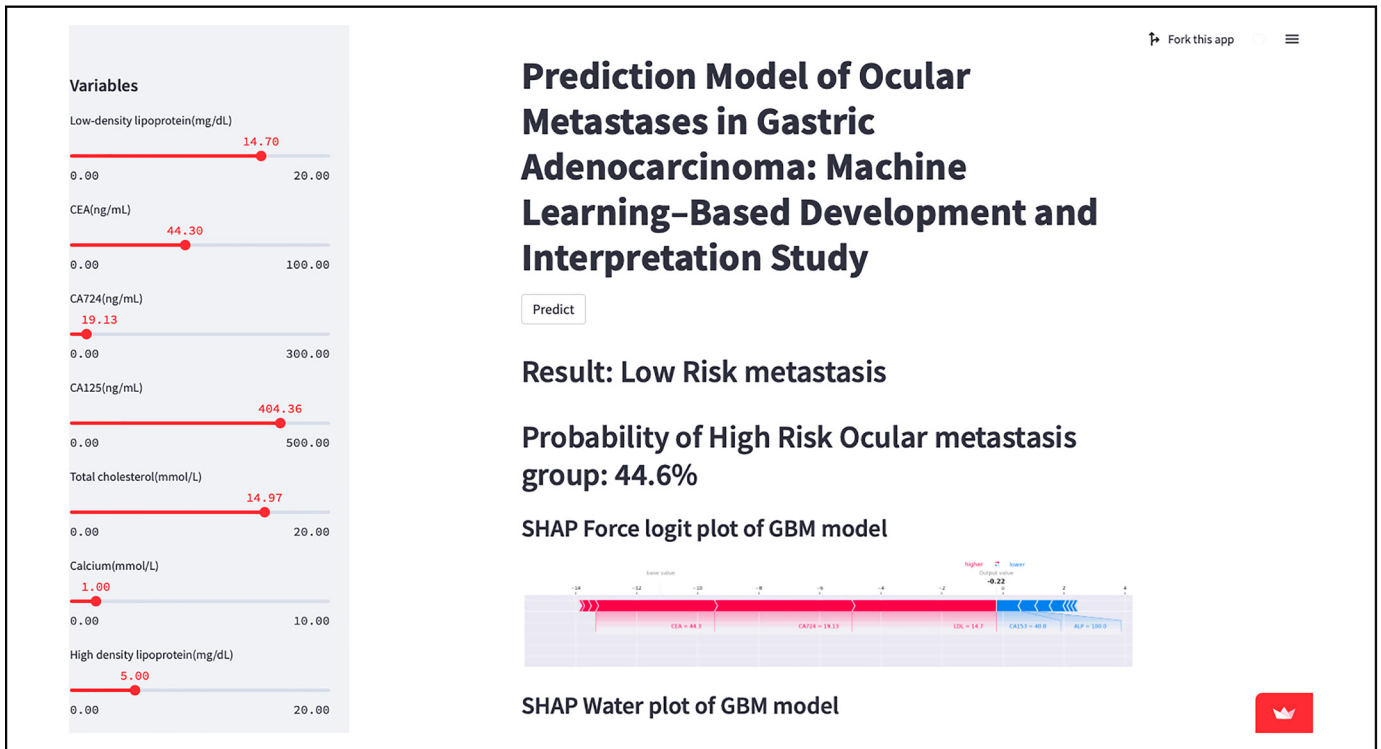


Figure 8. Web calculator for predicting OM of gastric adenocarcinoma. Note: <https://weicancer2-nnnjun76tn7vtkya5glyua.streamlit.app/>.

ML is an emerging medical field with the capacity to manage large, complicated and diverse data. It is the future of biomedical investigation and can promote global health care.⁵⁶ Using ML and survival analysis, Li et al found that differentially expressed mRNAs had potential diagnostic and prognostic value for gastric cancer.⁵⁷ In ophthalmology, a simple linear model has been utilized to predict advanced age-related macular degeneration.⁵⁸ The RF algorithm has been used to identify features that best predict the progression of geographic atrophy in age-related macular degeneration and to discover prognostic indicators of visual outcomes after intravitreal anti-vascular endothelial growth factor treatment.⁵⁹ Unlike traditional statistical models, ML models are powerful but are more complex and challenging to interpret due to their black box characteristics, limiting further clinical application. Therefore, we introduced SHAP to explain the GBM model. Based on the concept of Shapley value in game theory, SHAP can be rationalized as an extension of the unknowable interpretation method of the local interpretable model.¹⁷ The SHAP method can explain the GBM model and its predictions, producing a feature value for a single prediction from any GBM model and revealing its black-box properties.⁶⁰ We compared six different ML models to forecast the risk probability of OM in patients with GA and compared the F1 score, sensitivity, specificity, AUC, accuracy, and other indicators. Finally, the best-performing model was identified. Based on this, we developed a web page calculator which clinicians may use to enter the patient's indicators and thus obtain their personalized prediction probability of OM, which can assist with targeted and precise preventive measures.

However, our study has some limitations. First, this research is single-center, it is difficult to conduct external validation, and the performance of ML may vary depending on the characteristics of patients in different regions. Second, this is a retrospective study, which we need to verify further through follow-up studies. Third, the database only records the original diagnosis of the patient and no follow-up data for further analysis. Also, there is a significant imbalance between the cohort of OM positive and negative patients using training and test sets. In our following study, we will obtain large samples from multiple centers to verify the robustness and repeatability of the model.

Conclusion

We used the ML method to establish a risk prediction model for OM in patients with GA and showed that the GBM model performed best among the six ML models. Establishing this prediction model is a step toward an auto-loading diagnostic system, which may help clinicians make an accurate diagnosis and apply appropriate prevention measures for OM in patients with GA.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Jie Zou, Yan-Kun Shen, and Shi-Nan Wu. Dr Yi Shao and Dr Gang Tan were the guarantors of integrity of the entire study. The first draft of the manuscript was written by Yan-Kun Shen, Shi-Nan Wu, Hong

Wei, and Qing-Jian Li and all authors commented on previous versions of the manuscript. The statistical analysis was performed by San Hua Xu, Qian Ling, and Min Kang. Clinical data were collected by Zhao-Lin Liu, Hui Huang, and Xu Chen. Literature research was performed by Yi-Xin Wang and Xu-Lin Liao. All authors read and approved the final manuscript.

Availability of Data and Materials

The datasets used and/or analyzed during the present study are available from the corresponding author on reasonable request.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical Approval and Consent to Participate

The study methods and protocols were approved by the Medical Ethics Committee of the First Affiliated Hospital of Nanchang University (Nanchang, China) and followed the principles of the Declaration of Helsinki. All subjects were notified of the objectives and content of the study and latent risks, and then provided written informed consent to participate.



Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Major (Key) R&D Program of Jiangxi Province, Jiangxi Province Double Thousand Plan Science and Technology Innovation High-end Talent Project (2022), Excellent Talents Development Project of Jiangxi Province, National Natural Science Foundation of China (grant number 20181 bbg70004, 20203BBG73059, 2022103, 2022, 20192BCBL23020, 82160195).

Foundation Item

National Natural Science Foundation (No: 82160195); Jiangxi Province Double Thousand Plan Science and Technology Innovation High-end Talent Project (2022); Major (Key) R&D Program of Jiangxi Province (No: 2022103; 20181 bbg70004; 20203BBG73059); Excellent Talents Development Project of Jiangxi Province (No.: 20192BCBL23020)

ORCID iDs

Yan-Kun Shen  <https://orcid.org/0000-0002-0837-0992>
Yi Shao  <https://orcid.org/0000-0003-1571-2433>

Supplemental Material

Supplemental material for this article is available online.

References

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136(5):E359-E386. Epub 2014 Oct 9. PMID: 25220842.
2. Subhash VV, Yeo MS, Tan WL, Yong WP. Strategies and advancements in harnessing the immune system for gastric cancer immunotherapy. *J Immunol Res*. 2015;2015(2015):1-14. doi: 10.1155/2015/308574
3. Nienhüser H, Schmidt T. Gastric cancer lymph node resection-the more the merrier? *Transl Gastroenterol Hepatol*. 2018;3(1):1, PMID: 29441366; PMCID: PMC5803030.
4. Kawai S, Nishida T, Hayashi Y, et al. Choroidal and cutaneous metastasis from gastric adenocarcinoma. *World J Gastroenterol*. 2013;19(9):1485-1488. PMID: 23538460; PMCID: PMC3602510.
5. Shields CL, Kaliki S, Crabtree GS, et al. Iris metastasis from systemic cancer in 104 patients: The 2014 Jerry A. Shields lecture. *Cornea*. 2015;34(1):42-48. PMID: 25343701.
6. Celebi AR, Kilavuzoglu AE, Altiparmak UE, Cosar CB, Ozkiris A. Iris metastasis of gastric adenocarcinoma. *World J Surg Oncol*. 2016;14(71). doi:10.1186/s12957-016-0840-6. PMID: 26957317; PMCID: PMC4782568.
7. Tsuruta Y, Maeda Y, Kitaguchi Y, et al. A case of endonasal endoscopic surgery for intraorbital metastasis of gastric ring cell carcinoma. *Ear Nose Throat J*. 2022;101(1):NP24.
8. Quer G, Arnaout R, Henne M, Arnaout R. Machine learning and the future of cardiovascular care: JACC state-of-the-art review. *J Am Coll Cardiol*. 2021;77(3):300-313. PMID: 33478654; PMCID: PMC7839163.
9. Shung DL, Au B, Taylor RA, et al. Validation of a machine learning model that outperforms clinical risk scoring systems for upper gastrointestinal bleeding. *Gastroenterology*. 2020;158(1):160-167. Epub 2019 Sep 25. PMID: 31562847; PMCID: PMC7004228.
10. Luo H, Xu G, Li C, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: A multicentre, case-control, diagnostic study. *Lancet Oncol*. 2019;20(12):1645-1654. Epub 2019 Oct 4. PMID: 31591062.
11. Kanavati F, Tsuneki M. A deep learning model for gastric diffuse-type adenocarcinoma classification in whole slide images. *Sci Rep*. 2021;11(1):20486. PMID: 34650155; PMCID: PMC8516929.
12. Shao S, Liu L, Zhao Y, Mu L, Lu Q, Qin J. Application of machine learning for predicting anastomotic leakage in patients with gastric adenocarcinoma who received total or proximal gastrectomy. *J Pers Med*. 2021;11(8):748. PMID: 34442391; PMCID: PMC8400241.
13. Yardımcı AH, Koçak B, Turan Bektaş Ç, et al. Tubular gastric adenocarcinoma: Machine learning-based CT texture analysis for predicting lymphovascular and perineural invasion. *Diagn Interv Radiol*. 2020;26(6):515-522. PMID: 32990246; PMCID: PMC7664741.
14. Xu Z, Shen D, Kou Y, Nie T. A synthetic minority oversampling technique based on Gaussian mixture model filtering for imbalanced data classification. *IEEE Trans Neural Netw Learn Syst*. 2022;22(2022-08):1-14.
15. Karim M, Saad Missen MM, Umer M, et al. Comprehension of polarity of articles by citation sentiment analysis using TF-IDF and ML classifiers. *PeerJ Comput Sci*. 2022;8:e1107. PMID: 37346319; PMCID: PMC10280177.
16. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE initiative. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Ann Intern Med*. 2007;147(8):573-577. Erratum in: *Ann Intern Med*. 2008 Jan 15;148(2):168. PMID: 17938396.

17. Rodríguez-Pérez R, Bajorath J. Interpretation of compound activity predictions from Complex machine learning models using local approximations and Shapley values. *J Med Chem.* 2020; 63(16):8761-8777. Epub 2019 Sep 26. PMID: 31512867.
18. Kalafi EY, Nor NAM, Taib NA, Ganggayah MD, Town C, Dhillon SK. Machine learning and deep learning approaches in breast cancer survival prediction using clinical data. *Folia Biol (Praha).* 2019;65(5-6):212-220. PMID: 32362304.
19. Zhong LK, Xie CL, Jiang S, et al. Prioritizing susceptible genes for thyroid cancer based on gene interaction network. *Front Cell Dev Biol.* 2021;9(241):740267.
20. Liu C, Zhao Z, Gu X, et al. Establishment and verification of a bagged-trees-based model for prediction of sentinel lymph node metastasis for early breast cancer patients. *Front Oncol.* 2019;(9):282.
21. Zhou Z, Huang H, Liang Y. Cancer classification and biomarker selection via a penalized logsum network-based logistic regression model. *Technol Health Care.* 2021;29(S1):287-295. PMID: 33682765; PMCID: PMC8150479.
22. Bibault JE, Chang DT, Xing L. Development and validation of a model to predict survival in colorectal cancer using a gradient-boosted machine. *Gut.* 2021;70(5):884-889. Epub 2020 Sep 4. PMID: 32887732.
23. Perez BC, Bink MCAM, Svenson KL, Churchill GA, Calus MPL. Prediction performance of linear models and gradient boosting machine on complex phenotypes in outbred mice. *G3 (Bethesda).* 2022;12(4):jkac039. PMID: 35166767; PMCID: PMC8982369.
24. Castano-Duque L, Vaughan M, Lindsay J, Barnett K, Rajasekaran K. Gradient boosting and bayesian network machine learning models predict aflatoxin and fumonisin contamination of maize in Illinois - first USA case study. *Front Microbiol.* 2022;(13):1039947. PMID: 36439814; PMCID: PMC9684211.
25. Hoshi H. Management of gastric adenocarcinoma for general surgeons. *Surg Clin North Am.* 2020;100(3):523-534. Epub 2020 Mar 18. PMID: 32402298.
26. Takeda T, Ueyama H, Fu KI, Murata S, Nagahara A. Minute gastric adenocarcinoma of the fundic-gland type with submucosal invasion. *Endoscopy.* 2022;54(9):E468-E469. Epub 2021 Sep 27. PMID: 34571564.
27. Ichikawa J, Matsumoto S, Shimoji T, et al. Intraneural metastasis of gastric carcinoma leads to sciatic nerve palsy. *BMC Cancer.* 2012;(12):313.
28. Tkacz M, Tarnowski M, Staniszewska M, Pawlik A. Role of prometastatic factors in gastric cancer development. *Postepy Hig Med Dosw.* 2016;(70):1367-1377.
29. Qu Z, Liu J, Zhu L, Zhou Q. A comprehensive understanding of choroidal metastasis from lung cancer. *Onco Targets Ther.* 2021;(14):4451-4465.
30. Duke-Elder S. Diseases of the uveal tract: Muscular tumours: Leiomyoma. *Syst Ophthalmol.* 1966;10(9):799-805.
31. Konstantinidis L, Damato B. Intraocular metastases--A review. *Asia Pac J Ophthalmol.* 2017;6(2):208-214. PMID: 28399345.
32. Yang CJ, Tsai YM, Tsai MJ, Chang HL, Huang MS. The effect of chemotherapy with cisplatin and pemetrexed for choroidal metastasis of non-squamous cell carcinoma. *Cancer Chemother Pharmacol.* 2014;73(1):199-205. Epub 2013 Nov 8. PMID: 24202667.
33. Amemiya T, Hayashida H, Dake Y. Metastatic orbital tumors in Japan: A review of the literature. *Ophthalmic Epidemiol.* 2002; 9(1):35-47.
34. Han J, Jiang Y, Liu X, et al. Dietary fat intake and risk of gastric cancer: A meta-analysis of observational studies. *PLoS One.* 2015;10(9):e0138580. PMID: 26402223; PMCID: PMC4581710.
35. Sako A, Kitayama J, Kaisaki S, Nagawa H. Hyperlipidemia is a risk factor for lymphatic metastasis in superficial esophageal carcinoma. *Cancer Lett.* 2004;208(1):43-49. PMID: 15105044.
36. Jung JI, Cho HJ, Jung YJ, et al. High-fat diet-induced obesity increases lymphangiogenesis and lymph node metastasis in the B16F10 melanoma allograft model: Roles of adipocytes and M2-macrophages. *Int J Cancer.* 2015;136(2):258-270. Epub 2014 May 31. PMID: 24844408.
37. Levitan I, Volkov S, Subbaiah PV. Oxidized LDL: Diversity, patterns of recognition, and pathophysiology. *Antioxid Redox Signal.* 2010;13(1):39-75. PMID: 19888833; PMCID: PMC2877120.
38. Schnoeller TJ, Jentzmik F, Schrader AJ, Steinestel J. Influence of serum cholesterol level and statin treatment on prostate cancer aggressiveness. *Oncotarget.* 2017;8(29):47110-47120. PMID: 28445145; PMCID: PMC5564548.
39. Chen M, Zhang J, Sampieri K, et al. An aberrant SREBP-dependent lipogenic program promotes metastatic prostate cancer. *Nat Genet.* 2018;50(2):206-218. Epub 2018 Jan 15. PMID: 29335545; PMCID: PMC6714980.
40. Yuan B, Fu J, Yu WL, et al. Prognostic value of serum high-density lipoprotein cholesterol in patients with gallbladder cancer. *Rev Esp Enferm Dig.* 2019;111(11):839-845. PMID: 31595756.
41. Zheng HC, Zheng YS, Xia P, et al. The pathobiological behaviors and prognosis associated with Japanese gastric adenocarcinomas of pure WHO histological subtypes. *Histol Histopathol.* 2010; 25(4):445-452. PMID: 20183797.
42. Zhou X, Yao K, Zhang L, et al. Identification of differentiation-related proteins in gastric adenocarcinoma tissues by proteomics. *Technol Cancer Res Treat.* 2016;15(5):697-706. PMID: 27624754.
43. Chen Y, Yang YC, Tang LY, et al. Risk factors and their diagnostic values for ocular metastases in gastric adenocarcinoma. *Cancer Manag Res.* 2021;(13):5835-5843.
44. Hu PJ, Chen MY, Wu MS, et al. Clinical evaluation of CA72-4 for screening gastric cancer in A healthy population: A multicenter retrospective study. *Cancers (Basel).* 2019;11(5):733. PMID: 31137895; PMCID: PMC6562516.
45. Sun Z, Zhang N. Clinical evaluation of CEA, CA19-9, CA72-4 and CA125 in gastric cancer patients with neoadjuvant chemotherapy. *World J Surg Oncol.* 2014;(12):397.
46. Lordick F, Janjigian YY. Clinical impact of tumour biology in the management of gastroesophageal cancer. *Nat Rev Clin Oncol.* 2016;13(6):348-360. Epub 2016 Mar 1. PMID: 26925958; PMCID: PMC5521012.
47. Ieni A, Barresi V, Rigoli L, Caruso RA, Tuccari G. HER2 Status in premalignant, early, and advanced neoplastic lesions of the stomach. *Dis Markers.* 2015;(2015):1-10.
48. Kim DH, Yun HY, Ryu DH, et al. Preoperative CA 125 is significant indicator of curative resection in gastric cancer patients. *World J Gastroenterol.* 2015;21(4):1216-1221. PMID: 25632195; PMCID: PMC4306166.

49. Sisik A, Kaya M, Bas G, Basak F, Alimoglu O. CEA And CA 19-9 are still valuable markers for the prognosis of colorectal and gastric cancer patients. *Asian Pac J Cancer Prev*. 2013;14(7):4289-4294. PMID: 23991991.
50. Yamamoto M, Baba H, Toh Y, Okamura T, Maehara Y. Peritoneal lavage CEA/CA125 is a prognostic factor for gastric cancer patients. *J Cancer Res Clin Oncol*. 2007;133(7):471-476. Epub 2007 Jan 17. PMID: 17226046.
51. Nurczyk K, Nowak N, Carlson R, Skoczylas T, Wallner G. Pre-therapeutic molecular biomarkers of pathological response to neoadjuvant chemotherapy in gastric and esophago-gastric junction adenocarcinoma: A systematic review and meta-analysis. *Adv Med Sci*. 2023;68(1):138-146. Epub 2023 Mar 20. PMID: 36944288.
52. Yang AP, Liu J, Lei HY, Zhang QW, Zhao L, Yang GH. CA72-4 Combined with CEA, CA125 and CA19-9 improves the sensitivity for the early diagnosis of gastric cancer. *Clin Chim Acta*. 2014;(437):183-186.
53. Isonishi S, Ogura A, Kiyokawa T, et al. Alpha-fetoprotein (AFP)-producing ovarian tumor in an elderly woman. *Int J Clin Oncol*. 2009;14(1):70-73. Epub 2009 Feb 20. PMID: 19225928.
54. Kanda T, Yoshida H, Mamada Y, et al. Resection of liver metastases from an alpha-fetoprotein-producing gastric cancer. *J Nippon Med Sch*. 2005;72(1):66-70. PMID: 15834210.
55. He R, Yang Q, Dong X, et al. Clinicopathologic and prognostic characteristics of alpha-fetoprotein-producing gastric cancer. *Oncotarget*. 2017;8(14):23817-23830. PMID: 28423604; PMCID: PMC5410346.
56. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. Edocctor: Machine learning and the future of medicine. *J Intern Med*. 2018;284(6):603-619. Epub 2018 Sep 3. PMID: 30102808.
57. Li Q, Liu X, Gu J, Zhu J, Wei Z, Huang H. Screening lncRNAs with diagnostic and prognostic value for human stomach adenocarcinoma based on machine learning and mRNA-lncRNA co-expression network analysis. *Mol Genet Genomic Med*. 2020;8(11):e1512. Epub 2020 Oct 1. PMID: 33002344; PMCID: PMC7667366.
58. Schmidt-Erfurth U, Waldstein SM, Klmscha S, et al. Prediction of individual disease conversion in early AMD using artificial intelligence. *Invest Ophthalmol Vis Sci*. 2018;59(8):3199-3208. PMID: 29971444.
59. Bogunovic H, Montuoro A, Baratsits M, et al. Machine learning of the progression of intermediate age-related macular degeneration based on OCT imaging. *Invest Ophthalmol Vis Sci*. 2017;58(6):BIO141-BIO150. PMID: 28658477.
60. Rodríguez-Pérez R, Bajorath J. Interpretation of machine learning models using shapley values: Application to compound potency and multi-target activity predictions. *J Comput Aided Mol Des*. 2020;34(10):1013-1026. Epub 2020 May 2. PMID: 32361862; PMCID: PMC7449951.