ISSN 2345-1210 | Volume 10

Japanese Journal of Gastroenterology and Hepatology

Research Open Access

Evaluating Multiple Metabolic Indicators with Machine Learning Technology in Building a Risk StratificationModel for Gastric Intestinal Metaplasia Respect to Young, Mid-age and Elderly Individuals

Chieh Lee[1#], Jen Yu[2#], Chia-Yu Lai[3], Ta-Sen Yeh[4], Ming-Ling Chang[2,5] and Tsung-Hsing Chen[2,5*]

[1]Department of Information and Management, National Sun Yat-senUniversity, Kaohsiung, Taiwan
[2]Department of Gastroenterology and Hepatology, Linkou Medical Center,Chang Gung Memorial Hospital, Taoyuan, Taiwan
[3]Department of Management Information Systems, National PingtungUniversity of Science and Technology, Taiwan
[4]Department of General Surgery, Chang Gung Memorial Hospital, ChangGung University College of Medicine, Taoyuan City, Taiwan
[5]College of Medicine, Chang Gung University, Taoyuan, Taiwan

[#]Co-first author Chieh Lee, and Jen Yu

[*]Corresponding Author:
Tsung-Hsing Chen, Department of Gastroenterology and Hepatology, Linkou Medical Center,Chang Gung Memorial Hospital, Taoyuan, Taiwan and College of Medicine, Chang Gung University, Taoyuan, Taiwan

Received: 11 Mar 2025
Accepted: 16 Mar 2025
Published: 21 Mar 2025
J Short Name: JJGH

Keywords: Gastric Intestinal Metaplasia; Atherogenic Index of Plasma; Triglyceride-Glucose Index; Metabolic Indicators

Citation: TH Chen. Evaluating Multiple Metabolic Indicators with Machine Learning Technology in Building a Risk StratificationModel for Gastric Intestinal Metaplasia Respect to Young, Mid-age and Elderly Individuals. J Gastro Hepato. 2025; V10(14): 1-7

1. Abstract

1.1. Background

Metabolic syndrome is highly associated with gastric cancer (GC) formation, although the reliability of individual indices for predicting intestinal metaplasia (IM) risk remains inconsistent.

1.2. Objective

This retrospective cohort study applied univariate and multivariate analysis and machine learning to analyze the relationships between multiple metabolic indicators and IM.

1.3. Methods

The metabolic syndromes are used to predict IM. Multivariate analysis and machine learning are implemented to evaluate the predictability of metabolic indicators for IM.

1.4. Results

Our multivariateanalysis found that the accuracy associated with specific metabolic indicators of IM can vary according to age and gender. The AUC of elder individuals (>60%) was significantly higher compared to middle-aged individuals (54.7%, males; 57.1%, females).The MetS shows potential for young females to predict the IM with AUC equals 64.7%. To build a risk stratification model of IM, we implement a machine learning (ML) algorithm, XGBoost, to build the classification model. In the ML model, unlike in the multivariate analysis, the visceral adiposity index (VAI) and the Atherogenic Index of Plasma (API) become the top fivepredictors, revealing that the AIP and VAI are nonlinearly associated with IM. Our ML model's overall accuracy is 82.9% for males and 83.8% for females. By conducting a comprehensive multivariateanalysis of multiple metabolic indicators, our study reveals that significance varies according to gender and age. We also find that ML is a promising tool for future metabolic indicator-based classification models for IM with respect to different genders and age groups.

2. Introduction

Metabolic syndrome (MetS) is a complex medical condition characterized by a cluster of interrelated conditions, including obesity, hypertension, elevated blood sugar levels, and abnormal lipid profiles. Extensive research has demonstrated strong associations between MetS and chronic diseases, such as cardiovascular disorders and diabetes [1]. Moreover, MetS has emerged as a valuable predictor for postoperative complications, cancer recurrence, and increased overall mortality rates among patients with gastric cancer(GC) [2]. While the relationship between MetS and GC is well studied, the relationship between MetS and gastric precancerous lesions remains unclear.Some recent studies have suggested the Triglyceride-Glucose Index (TyG) as a novel serum biomarker with predictive potential for gastric carcinogenesis [3]. Furthermore, gastric intestinal metaplasia (IM) has been firmly established as a precancerous lesion in the development of GC, and the severity of gastric IM is closely associated with the risk of GC development [4]. Amato et al.[5] propose a visceral adiposity index (VAI) and waist circumference as risk indicators of metabolic-related disease.MetS is often noted as an age-dependent syndrome. Furthermore, age, gender, and lifestyle risk factors are strongly associated with [6,7].Found that elder, male, non-white individuals exhibit a higher risk of GIM [8]. Found that, compared with other ethnic groups, Western women with MetS have a higher GC risk. These results show that MetS and IM's relationship might vary betweengenders and age groups. Other than typical MetS, this study also included the predictive capabilities of two metabolic indices: the Atherogenic Index of Plasma (AIP) and TyG. AIP, which integrates arterial lipids and blood sugar levels, offers valuable insights into the risk of atherosclerosis and cardiovascular diseases. Additionally, TyG is a reflective marker of insulin resistance and may be closely correlated with MetS.Previous studies show evidence of the potential of implementing MetS in predicting GC [9]. However, a comprehensive examination of the predictability of MetS in IM is still needed. Compared to the traditional statistical method, machine learning exhibits greater power in predicting GC [10,11]. In this study, we first use the uni- and multi-variate statistical method to assess the predictability of MetS for IM. By comprehensively examining the relationship between MetS and gastric IM across different gender and age groups, we aim to characterize the association between MetS and IM risk. We then applied a ML modeling approach to develop an efficient IM risk stratification model. Such ML-based IM risk stratification model enables the outpatient unit to identify the potential IM individuals and suggest them for further examination.

3. Methods

We performed univariate and multivariate analyses to characterize the relationship between metabolic indicators and IM. In this stage, we aim to illustrate whether MetS's impact on IM differs

between genders and age groups. We then implement ML technology to build a risk stratification model. In addition, Python packages were used to build a database and conduct statistical and ML modeling approaches. This section presents our database inclusion and exclusion criteria and our statistical methods and measurements of the relationships between the metabolic indicators and IM. Lastly, we present our machine learning model and its performance indices. This study was approved by the Institutional Review Board (IRB) under protocol number 202300866B0.

4. Data Collection and Preprocessing

From 2010 to 2014, 59,143 subjects were enrolled in this retrospective cohort study at Chang Gung Memorial Hospital, Linkou, Taiwan,1,355 of whom had undergone endoscopic biopsy and underwent further analysis. After eliminating cases with incomplete blood test data, the sample size was narrowed to 10,380 subjects. The analysis then segregated these individuals into 2,088 subjects with IM and 8,292 subjects without IM, as illustrated in Figure 1. This section comprehensively outlines the study design and methodology used to investigate IM and the associations with metabolic indicators.We collected data related to the metabolic indicators from 10,380 individuals. The metabolic indicators encompassed a broader spectrum of factors, including pre-prandial blood glucose (AC), postprandial blood glucose (PC), total cholesterol (TC), triglycerides (TG), high-density lipoprotein (HDL), low-density lipoprotein (LDL), very-low-density lipoprotein (VLDL), the AIP, and the TyG index. Several studies, such as those by Khaw et al. [12] and Tseng et al. [13], have shown that hemoglobin A1C (HbA1C), representing blood sugar (glucose) level, is a significant risk factor for IM,after adjusting for various factors including fasting blood sugar, supporting the hypothesis that HbA1C and fasting glucose may be putative risk factors. In addition, previous studies by Liu et al. [14] and van der Poorten et al. [15] indicate that predictors of liver-related diseases are associated with metabolic syndromes, including cardiovascular diseases and stroke. The studies focused on subjects with non-alcoholic fatty liver disease (NAFLD), showing that visceral fat is an independent predictive factor positively associated with serum triglycerides, HDL, LDL, and interleukin-6 (IL-6), lipid-related parameters. These factors are known to be directly related to inflammatory processes, thereby validating their integration into analytical models. Similarly, the AIP and TyG indexes, as reported by Cheong et al.[16], influence disease, and are independent predictors of disease incidence and mortality rates, thus justifying their suitability for inclusion in the predictive indicators studied here. We used the following equations to calculate the AIP, TyG, and VAI:

$$AIP = \log\left(\frac{\text{Triglyceride}}{\text{HDL}}\right)$$

$$TyG = \frac{\log\left(\frac{\text{Triglyceride}}{2.2}\right) \times \text{Glucose(AC)}}{2}$$

$$VAI = \begin{cases} \frac{\text{Waist Circumference(cm)}}{39.68 + (1.88 \times \text{BMI})} \times \frac{\text{Triglyceride}}{1.03} \times \frac{1.31}{HDL} & \textit{Males} \\ \frac{\text{Waist Circumference(cm)}}{39.58 + (1.89 \times \text{BMI})} \times \frac{\text{Triglyceride}}{0.81} \times \frac{1.52}{HDL} & \textit{Females} \end{cases}$$

4.1. The Exclusion Criteria for This Study at The Subjects' Annual Health Checkup Were As Follows:

1. Uncooperative, unwilling, or individuals with impaired consciousness.
2. Individuals with conditions such as pregnancy or systemic diseases that may affect anesthesia and safety.
3. Individuals who have experienced bleeding or ischemic stroke within the last six months.
4. Those with cardiovascular or pulmonary diseases pose a risk during the checkup.
5. Individuals with abnormal liver function, bilirubin, or platelet levels.
6. Individuals with abnormal thyroid function or poorly controlled diabetes.
7. People who have undergone major surgery in the last six months.
8. Individuals with drug or alcohol addiction.
9. Individuals with severe ankylosing spondylitis or expected airway difficulties.
10. Obesity with severe obstructive sleep apnea or a BMI (Body Mass Index) greater than 35.
11. Abnormal potassium levels (K < 3.0 or K > 5.0).
12. Nail polish should ideally be completely removed or at least one fingernail on each hand for anesthesia safety. If not removed, anesthesia should be canceled.

The descriptive statistics of the patient cohort. Table 1 shows that the age, BMI, and blood pressure presented similar mean and standard deviation values between the control group (without IM) and the IM group. However, the TC and LDL values exhibited notably different mean values between the control and IM groups. Thus, we further implemented univariate and multivariate analyses to investigate the impact of metabolic indicators on IM classification.

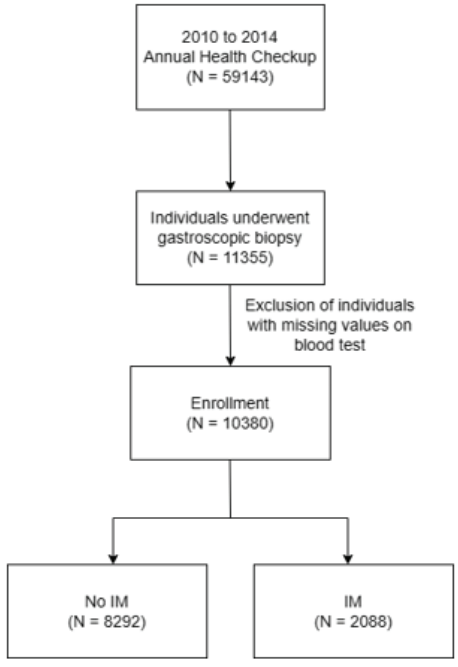

Figure 1: Flow chart of study cohort selection.

Table 1: Individual's basic information.

Variable	Control N=8292(95%CI)	IM N=2088(95%CI)
gender(male/female)	4732/ 3560	1447/ 641
Age	51.33±10.93	52.9±9.2
Waistline	85.6±12.9	86.0±8.2
BMI	24.3±3.45	24.3±2.9
Glucose (AC)	101.1±22.2	100.9±19.2
Glucose (PC)	110.1±41.3	109.4±34.5
Diastolic blood pressure (DBP)	81.7±11.8	82.1±10.4
Systolic blood pressure (SBP	133.8±19.7	133.9±17.4
Triglyceride (TG)	136.6±96.3	142.0±103.5
Total cholesterol (TC)	199.6±36.5	203.4±35.3
HDL	49.7±13.2	48.2±11.9
low-density lipoprotein (LDL)	122.9±32.9	126.5±33.0
Very-low-density (VLDL)	27.0±17.6	28.0±17.3

AC: Fasting Blood Glucose; **PC**: Postprandial Blood Glucose; **HDL**: High-Density Lipoprotein.

5. Statistical Methods

Our statistical methodology adopted a heuristic approach to analyze the incidence of IM. We analyzed the relationship between MetS and IM by utilizing multivariate stepwise logistic regression. Furthermore, to avoid the bias and confounding that might exist in our database, we also use the Propensity Score Matching (PSM) analysis to match our sampling from IM and non-IM groups. To determine the optimal cut-off value of the index for predicting IM, we employed the receiver operating characteristic (ROC) curve analysis [17]. The cut-off value was identified as the point on the ROC curve that maximized the Youden index within the area prioritizing sensitivity. We assessed the Youden index for various potential cut-off values, and the one yielding the highest value was considered the optimal cut-off point.

5.1. The IM Risk Stratification Model with Machine Learning Method

Our dataset comprises an extensive array of clinical variables, encompassing metabolic indices such as LDL, HDL, and triglycerides and qualitative evaluations like fatty liver and body mass index. Given the diverse nature of the dataset, XGBoost (Extreme Gradient Boosting) algorithm emerges as the optimal choice due to its capacity to handle large and heterogeneous datasets while mitigating the risk of overfitting.There are several performanceindices used in evaluating machinelearning methods. Those performance indices includeaccuracy, precision, F1-score, sensitivity and specificity. Accuracy refers to the number of correct predictions over the number of total samples. Precision is also known as positive predictive value. Precision refers to the total correct positive prediction over the total number of correct production samples. The F1 score, which is calculated from precision and recall, evaluates the effectiveness of the model. Where the best F1-score is one and worst is zero.

6. Results

The basic information for all individualsof all different ages and gender groups is listed in the online appendix Table A1-A2. To compare the impact of metabolic indicators between different genders and age groups, we repeated the univariate and multivariate analyses for males and females in three different age groups. Those three age groups were young (<45 years), middle-aged (45 to 70 years), and elderly (>70 years) subjects. In Table 1 we eliminate the metabolic indicators that are not significant for any group. From Table 1, we find that The LDL value was identified as the only significant indicator for both themiddle-aged male and female. For the middle-age female group, the UA value was also a significant indicator, indicating that the UA level is important forfemales but not for males. In addition, the TC is an important predictor of IM for females and young males. The AUC of using the metabolic indicators to predict for the young males was 57.2%, and for the young females was 64.7%. As shown in Table 1, for middle-aged individuals, bloodsugar (HbA1c), LDL, and fatty liver remained the common risk indicatoracross gender. The TG level and blood pressure were significant only for males. Note that the AUC of middle-aged males was 54.7%, similar to the overall male result. However, for females, the AUC of the metabolic indicator exhibited a higher value than the overall female result (57.1% vs. 55.5%). Interestingly, while the female group enjoys higher AUC, there are more significant indicators for middle-aged males compared with females. The AUC for elderly individuals was 62.8% for elderly males and 65.5% for elderly females. Hence, the AUCs of elderly individuals were over 60% regardless of gender. This result indicates that the metabolic indicators have a higher discriminability in classifying non-IM and IM elder individuals. Note that the VAI value was significant in predicting IM in elderly males, while SBP remained the most pronounced metabolic indicator across all age and gender groups. The metabolic indicators showed the highest discriminability in the elderly groups. The AUCs of all groups expect the middle-aged male were higher than the overall result, indicating that the impacts of the metabolic indicators vary by age group. We conclude that a single multivariate model cannot identify the correct MetS predictor for all patients. The MetS can serve as a very effect predictor for elder individuals with AUC equals 65.5% for female and 62.8% for male.As shown in Table 1, the AUC of the multivariate model ranges from 0.547 to 0.655. While the AUC shows that the metabolic indicators can assistphysiciansin identifyingpotential IM patients, the accuracy of the current model is not satisfactory. The multivariable analysis indicates that the MetS predictors of IM vary with gender and age. Thus, we build a more accurate machine learning (ML) based model to better understand how to stratify potential IM patients. We implementedan XGBoost to build our ML risk stratification model. The XGBoost is a decision-tree-based model with anembeddedbootstrapped sampling method. In predicting GIM risk, the objective function of the XGBoost model is essential for balancing accuracy and complexity, ensuring that the model is both well-fitted to the data and capable of generalizing to new cases. This balance is particularly critical given the diverse clinical and metabolic data used in this study, which vary in their influence on GIM risk. Table 2 summarizes the performance metrics of the XGBoost model across different demographics. The results demonstrate that while the model maintains high accuracy and specificity across various groups, there is variability in sensitivity and F1 scores, particularly when comparing different age groups and genders. For example, the model shows higher overall accuracy and precision for females than males but struggles with sensitivity in elderly populations.Compared with the multivariate model, the ML-based model exhibits higher specificity, higher than 90%, except for the older female. While the specificity is high, the sensitivity (recall) is not as promising. This result indicates that it is very accurate once the ML identifies a patient as a high-risk patient of IM. However, the ML model mightnot be able to catch all the IM patients. In other words, the ML model might incur a high false negative. To avoid this, the practitioner can increase the threshold value of the model and lower the false negative rate. We also find that, unlike in the multivariate analysis, separating patients by their age and gender will not increase the accuracy of the ML model. By comparing the significant/important factors in Table 1 and Table 3 below, we find that the ML and multivariate analysis havedifferent important risk factors. Note that the metabolic indices such as AIP and VAI are not pronounced in the multivariate analysis, but they are important indicators in the ML model. This result shows that the AIP and VAI have an indirect effect on the IM prediction, which might be ignored by traditional statistical analysis. Except for the AIP and VAI, other important metabolic indicators are very similar in multivariate analysis and in the ML-based model. It is worth noting that BMI and FL both remain significant factors across all age and gender groups. This shows that body and visceral fats contribute to IM for all individuals. The VAI has an additional impact on males, but not females.

7. Discussion

Previous studies have reported inconsistent findings regarding the relationships between lipids and various types of cancers, including GC [18]. Similarly, TC has been found to exhibit diverse associations with gastric neoplasms, including negative [19,20], positive [21], and no correlation [22]. Meanwhile, elevated levels of LDL cholesterol (LDL-C) and low levels of HDL cholesterol (HDL-C) have been linked to increased inflammation, and certain genes associated with the LDL receptor have been shown to be involved in regulating tumors, including GC [23]. High LDL levels have been associated with the development of GC [24,25] via suppression of the host immune system [26]. Despite the cholesterol level, Dyslipidemia is another component of MetS that is significantly related to GC. We consider the applications of AIP and TyGin other medical contexts, including GC [27], with a particular emphasis on gastric precancerous lesions [3]. MetS is characterized by chronic low-grade inflammation with elevated levels of pro-inflammatory cytokines. This inflammatory environment may contribute to changes in gastric mucosa, as chronic inflammation has been recognized as a precursor to cancer initiation and progression [28]. Both AIP and TyG are indicative of insulin resistance. Insulin resistance can lead to hyperinsulinemia and increased levels of insulin-like growth factors, which have been implicated in gastric carcinogenesis [29,30]. AIP may promote endothelial dysfunction and atherosclerosis, and has also been used to predict colon cancer [27]. While GC highly

Table 2: Summary of Multivariate Analysis Results.

	Young				Middle-Aged				Elder			
	Female		Male		Female		Male		Female		Male	
	AUC =0.647		AUC = 0.572		AUC = 0.571		AUC = 0.547		AUC =0.655		AUC =0.628	
	OR	P value	OR	P value	OR	P value	OR	P value	OR	P value	OR	P value
UA					0.931 (0.885-0.979)	0.005						
SBP	0.97 (0.955-0.985)	0.000					0.997 (0.994-0.999)	0.023	0.984 (0.971-0.996)	0.011	0.987 (0.977-0.997)	0.01
DBP	1.06 (1.037-1.084)	0.000										
HbA1c					1.073 (1.052-1.095)	0	1.019 (1.000-1.039)	0.05				
AC					0.995 (0.992-0.998)	0.002						
TC	1.008 (1.004-1.013)	0	0.991 (0.985-0.998)	0.011	0.993 (0.989-0.997)	0.001			1.01 (1.003-1.017)	0.004		
BMI							0.857 (0.803-0.915)	0.000	0.493 (0.336-0.722)	0.000		
Waistline									1.035 (1.001-1.070)	0.047		
TG	1.015 (1.004-1.026)	0.008										
HDL							0.989 (0.985-0.994)	0.000				
LDL			1.016 (1.008-1.023)	0.000	1.011 (1.007-1.016)	0.000	1.002 (1.001-1.004)	0.001			1.007 (1.001-1.013)	0.016
FL					1.226 (1.109-1.356)	0.000	1.095 (1.019-1.177)	0.014	1.657 (1.148-2.391)	0.007		
VAI											0.944 (0.903-0.987)	0.001

UA: Uric Acid; **DBP**: Diastolic blood pressure; **SBP**: Systolic blood pressure; **HbA1c**: Glycated Hemoglobin; **AC**: Fasting Blood Glucose; **PC**: Postprandial Blood Glucose; **TG**: Triglyceride; **TC**: total Cholesterol; **HDL**: High-Density Lipoprotein; **LDL**: Low-Density Lipoprotein; **VLDL**: Very-Low-Density Lipoprotein; **AIP**: Atherogenic Index of Plasma; **TyG**: Triglyceride-Glucose.

correlates with metabolism [31], limited research has explored the association between AIP and early GC. According to our study's results, AIP is significantly associated with the gastric IM in our ML model. Meanwhile, the TyG and VAI indeice has been reported as a novel predictive biomarker for gastric carcinogenesis [3], and it is also an important factor in our ML model. In our multivariate analysis and ML model both shows that MetS except PC, significantly impact IM. FL, HDL and LDL are significant, consistent with prior findings across genders and ages [32,33]. The expression of the VLDL receptor, which plays a significant role in TG metabolism, has been reported to be associated with the differentiation of gastrointestinal cancer [34]. While VLDL is a pronounced impact in the ML model. Herein, we hypothesize that gastric mucosal metaplasia may follow a similar mechanism. According to our investigation, LDL plays a significant role in IM individuals, aside from TyG [3] and AIP; hence, using LDL-C as a predictor for gastric precancerous lesions in the general population may facilitate early intervention and improve patient outcomes. Regular monitoring of LDL-C in individuals with or without MetS may help to identify those at higher risk, enabling timely endoscopic evaluations and preventive measures. The ML methods such as XGBoost can assist physicians in building an effective IM prediction model with MetS.

8. Conclusion

We obtained annual healthcare check-up data from 59,143 individuals at Chang Gung Memorial Hospital. Following data preprocessing, a dataset comprising 1,355 individuals who had undergone endoscopic biopsies was constructed. This study aims to investigate the relationship between metabolic syndrome (MetS) and intestinal metaplasia (IM) through univariate and multivariate analyses. After establishing the predictive capability of MetS for IM, we employed the

Table 3: The summary of Machine Learning Model efficiency indices.

		Accuracy	Precision	F1 score	Sensitivity	Specificity
Male	All	0.829 (0.824-0.835)	0.872 (0.863-0.880)	0.819 (0.814-0.824)	0.773 (0.768-0.778)	0.886 (0.877-0.895)
	Young	0.728 (0.690-0.766)	0.858 (0.811-0.905)	0.665 (0.607-0.722)	0.549 (0.477-0.620)	0.907 (0.872-0.943)
	Middle-Age	0.763 (0.754-0.772)	0.897 (0.879-0.916)	0.714 (0.702-0.727)	0.594 (0.579-0.609)	0.932 (0.918-0.945)
	Elder	0.614 (0.544-0.685)	0.697 (0.599-0.794)	0.487 (0.363-0.612)	0.393 (0.262-0.524)	0.836 (0.775-0.897)
Female	All	0.838 (0.827-0.849)	0.971 (0.963-0.979)	0.811 (0.795-0.828)	0.698 (0.673-0.722)	0.979 (0.973-0.985)
	Young	0.629 (0.534-0.724)	0.686 (0.517-0.855)	0.439 (0.235-0.643)	0.346 (0.164-0.528)	0.912 (0.886-0.938)
	Middle-Age	0.805 (0.784-0.826)	0.951 (0.941-0.962)	0.766 (0.736-0.796)	0.642 (0.603-0.682)	0.967 (0.961-0.974)
	Elder	0.602 (0.525-0.679)	0.667 (0.448-0.885)	0.416 (0.253-0.578)	0.313 (0.178-0.449)	0.890 (0.833-0.948)

Table 4: Summary of the importance of metabolic indicators in the ML model.

	All		Young		Middle-Aged		Elder	
	Female	Male	Female	Male	Female	Male	Female	Male
UA	0.033		0.058					0.048
DBP					0.033		0.065	
AC		0.03	0.05	0.041		0.029		
TC							0.07	
BMI	0.279	0.325	0.118	0.222	0.263	0.32	0.111	0.089
Waistline	0.033	0.027				0.029		
HDL			0.062					
VLDL	0.055	0.036			0.044	0.035	0.061	
AIP				0.043	0.033			0.054
FL	0.28	0.269	0.207	0.231	0.298	0.262	0.09	0.23
VAI				0.045				0.048

Online Appendix

Appendix A1Descriptive Statistical Analysis

As shown in Table A1, the univariate analysis revealed that all metabolic indicators, except for PC, significantly impacted the classification of IM patients. However, the multivariate analysis showed that only HDL and LDL were significant indicators of IM. The overall area under the ROC curve (AUC) was 55.5%, showing that the MetS may discriminate non-IM from IM individuals. To further investigate the discriminability of the metabolic indicators, we applied the univariate and multivariate analyses to the different age groups.

Table A1: The overall results for metabolic indicators.

		Univariate			Multivariate	
Variable	Control N=8292 (95%CI)	IM N=2088 (95%CI)	P value	OR	P value	Cut-off point
UA	5.94±1.49	6.12±1.31	0			
HbA1c	29.5±2.9	29.9±2.3	0			
AC	101.1±22.2	100.9±19.2	0.07			
PC	110.1±41.3	110.7±37.0	0.3			
TC	199.6±36.5	203.4±35.0	0			
TG	136.6±96.3	140.0±95.4	0.022			
HDL	49.7±13.2	48.2±11.9	0		0	0.511
LDL	122.9±32.9	127.3±33.1	0		0	0.511
VLDL	27.0±17.6	28.1±17.5	0			
AIP	0.38±0.31	0.41±0.3	0	1.287 (1.056-1.568)		
TyG	0.16±0.52	0.2±0.5	0			

UA: Uric Acid; **HbA1c:** Glycated Hemoglobin; **AC**: Fasting Blood Glucose; **PC**: Postprandial Blood Glucose; **TG**: Triglyceride; **TC**: total Cholesterol; **HDL**: High-Density Lipoprotein; **LDL**: Low-Density Lipoprotein; **VLDL**: Very-Low-Density Lipoprotein; **AIP**: Atherogenic Index of Plasma; **TyG**: Triglyceride-Glucose

Appendix A2

The results of multivariate analysis for the different gender groups are presented in Table A2. As shown in Table A2, the metabolic indicators exhibited different impacts in the different gender groups. However, LDL, BMI and fatty liver remained the most pronounced indicators. Furthermore, systolic blood pressure was a significant indicator of IM for males but not females. The detailed information is summarized in online Appendix A3. The AUC equals 54.9% the male and was 56.1% for female. The AUCs for male were lower than the overall result, while the female's AUC is higher than the overall AUC. This result indicates that the predictability of the MetS differs between different gender. The cut-off points are 0.502 for male and 0.504 for female.

Table A2: The metabolic indicators in males vs. females.

	Male					**Female**				
	Univariate			Multivariate		Univariate			Multivariate	
Variable	Control N=4732 (95%CI)	IM N=1447 (95%CI)	P value	OR	P value	Control N=3560 (95%CI)	IM N=641 (95%CI)	P value	OR	P value
UA	6.6±1.36	6.6±1.34	0.98			5.06±1.2	5.07±1.2	0.848		
SBP	136.7±18.3	135.7±18.0	0.056	0.997 (0.995-0.999)	0.017	129.8±20.8	130.4±20.0	0.534		
DBP	84.6±11.5	84.0±11.4	0.095			77.8±11.1	78.0±10.4	0.709		
HbA1c	30.1±2.6	30.2±2.50	0.170			28.75±3.0	29.2±2.9	0.001	1.050 (1.032-1.069)	0.000
AC	102.9±23.6	102.2±20.4	0.140			98.7±20.0	98.4±18.3	0.767		
PC	110.7±44.1	110.2±36.3	0.580			109.4±37.2	110.0±34.9	0.684		
TG	154.9±108.3	157.1±120.4	0.330			112.3±70.5	112.5±63.8	0.942	1.003 (1.001-1.004)	
BMI	58/1836 /1698/1140	24/607 /505/311	0.006	0.873 (0.824-0.924)	0.000	177/2053 /817/513	42/356 /161/82	0.439	0.901 (0.845-0.962)	0.002
Waistline	88.4±8.7	87.9±8.5	0.093			82.0±16.3	81.7±9.3	0.699		
FL	683/ 544/3505	271/ 159/1017	0.109	1.094 (1.025-1.168)	0.007	482/ 208/2870	91/ 50/500	0.058	1.234 (1.127-1.350)	0.000
TC	199.2±36.8	203.0±36.4	0.000			200.1±36.2	204.7±37.1	0.004		0.013
HDL	45.1±10.9	44.9±10.8	0.0.553	0.996 (0.992-1.000)	0.037	55.8±13.5	56.5±13.9	0.206		
LDL	123.6±33.4	127.5±33.4	0.0000	1.004 (1.003-1.005)	0.0000	121.9±32.2	125.7±34.5	0.006	1.003 (1.002-1.005)	0.000
VLDL	30.5±19.3	30.7±19.7	0.6			22.4±13.8	22.4±12.5	0.991		
AIP	0.48±0.3	0.49±0.3	0.32	1.290 (1.008-1.649)		0.26±0.3	0.26±0.3	0.740		
TyG	0.27±0.53	0.28±0.52	0.38			0.01±0.46	0.02±0.46	0.846		

UA: Uric Acid; **SBP**: Systolic Blood Pressure; **DBP**: Diastolic Blood Pressure; **HbA1c**: Glycated Hemoglobin; **AC**: Fasting Blood Glucose; **PC**: Postprandial Blood Glucose; **TG**: Triglyceride; **BMI**: Body Mass Index (underweight/healthy/ overweight/obesity); **FL**: Fatty Liver (Mild/Moderate/None); **TC**: total Cholesterol **HDL**: High-Density Lipoprotein; **LDL**: Low-Density Lipoprotein; **VLDL**: Very-Low-Density Lipoprotein; **AIP**: Atherogenic Index of Plasma; **TyG**: Triglyceride-Glucose.

machine learning algorithm, XGBoost, to develop an efficient IM risk stratification model. The resulting XGBoost-based IM risk startmodel has the potential for seamless integration into healthcare centers, assisting medical professionals in identifying individuals who may require surveillance endoscopy.Our study provides a comprehensive evaluation of metabolic indicators, highlighting variations in their significance based on age and gender. Among these indicators, low-density lipoprotein (LDL) emerged as the only consistent predictor, offering valuable insights for clinical risk assessment. Additionally, the visceral adiposity index (VAI) and atherogenic index of plasma (AIP) were identified as key predictors of IM in the artificial intelligence model. Compared to traditional multivariate analysis, in which AUCs ranges from 0.547 to 0.655, our machine learning-based IM risk stratification model demonstrated superior predictive performance, achieving an accuracy of 83.8% for females and 78.4% for males. Furthermore, the model exhibited high specificity, exceeding 90% across all age groups except for elderly females, for whom it maintained a high accuracy of 89.0%.

References

1. Handelsman Y, Butler J, Bakris GL, DeFronzo RA, Fonarow GC. Early intervention and intensive management of patients with diabetes, cardiorenal, and metabolic diseases. J Diabetes Complications. 2023; 37(2):108389.
2. Huang Z, Zhou J, Chen L, Zhang Y: Metabolic Syndrome and Clinical Outcomes of Patients with Gastric Cancer: A Meta-Analysis. Horm Metab Res. 2023; 55(5):333-342.
3. Kim YM, Kim JH, Park JS, Baik SJ, Chun J, Youn YH. Association

between triglyceride-glucose index and gastric carcinogenesis: a health checkup cohort study. Gastric Cancer. 2022; 25(1):33-41.

4. Lee JWJ, Zhu F, Srivastava S, Tsao SK, Khor C, Ho KY, Fock KM. Severity of gastric intestinal metaplasia predicts the risk of gastric cancer: a prospective multicentre cohort study (GCEP). Gut. 2022; 71(5):854-863.
5. Amato MC, Pizzolanti G, Torregrossa V, Misiano G, Milano S. Visceral adiposity index (VAI) is predictive of an altered adipokine profile in patients with type 2 diabetes. PLoS One. 2014; 9(3):e91969.
6. Ogbera AO. Prevalence and gender distribution of the metabolic syndrome. Diabetology & metabolic syndrome. 2010; 2:1-5.
7. Tan B, Anderson EG, Parker GG. Platform Pricing and Investment to Drive Third-Party Value Creation in Two-Sided Networks. Information Systems Research. 2020; 31(1):217-239.
8. Mariani M, Sassano M, Boccia S. Metabolic syndrome and gastric cancer risk: a systematic review and meta-analysis. European Journal of Cancer Prevention. 2021; 30(3):239-250.
9. Cai C, Li H, Zhang L, Li J, Duan S, Fang Z, Li C, Chen H. Machine learning identification of nutrient intake variations across age groups in metabolic syndrome and healthy populations. Nutrients. 2024; 16(11):1659.
10. Tran TT, Lee J, Gunathilake M, Kim J, Kim S-Y, Cho H, Kim J. A comparison of machine learning models and Cox proportional hazards models regarding their ability to predict the risk of gastrointestinal cancer based on metabolic syndrome and its components. Frontiers in Oncology. 2023; 13:1049787.
11. Vo D, Ghosh P, Sahoo D. Artificial intelligence-guided discovery of gastric cancer continuum. Gastric Cancer. 2023; 26(2):286-297.
12. Khaw KT, Wareham N, Bingham S, Luben R, Welch A. Preliminary communication: glycated hemoglobin, diabetes, and incident colorectal cancer in men and women: a prospective analysis from the European prospective investigation into cancer-Norfolk study. Cancer Epidemiol Biomarkers Prev. 2004; 13(6):915-919.
13. Tseng PH, Lee YC, Chiu HM, Chen CC, Liao WC, Tu CH, Yang WS. Association of diabetes and HbA1c levels with gastrointestinal manifestations. Diabetes Care. 2012; 35(5):1053-1060.
14. Liu B, Wen P, Gu X, Weng R, Liu S. Elevated serum triglyceride predicts recurrence of colorectal polyps in patients with advanced adenomas. Lipids Health Dis. 2020; 19(1):211.
15. van der Poorten D, Milner KL, Hui J, Hodge A, Trenell MI, Kench JG, London R. a key mediator of steatohepatitis in metabolic liver disease. Hepatology. 2008; 48(2):449-457.
16. Cheong CY, Yap P, Yap KB, Ng TP: Associations of Inflammatory, Metabolic, Malnutrition, and Frailty Indexes with Multimorbidity Incidence and Progression, and Mortality Impact: Singapore Longitudinal Aging Study. Gerontology. 2023; 69(4):416-427.
17. Ravi S, Larochelle H. Optimization as a model for few-shot learning. In: International conference on learning representations. 2016; 2016.
18. Hashmi S, Wang Y, Suman DS, Parhar RS, Collison K, Conca W, Al-Mohanna F. Human cancer: is it linked to dysfunctional lipid metabolism? Biochim Biophys Acta. 2015; 1850(2):352-364.
19. Asano K, Kubo M, Yonemoto K, Doi Y, Ninomiya T, Tanizaki Y, Arima H. Impact of serum total cholesterol on the incidence of gastric cancer in a population-based prospective study: the Hisayama study. Int J Cancer. 2008; 122(4):909-914.
20. Törnberg SA, Carstensen JM, Holm LE. Risk of stomach cancer in association with serum cholesterol and beta-lipoprotein. Acta Oncol. 1988; 27(1):39-42.
21. Huang YK, Kang WM, Ma ZQ, Liu YQ, Zhou L. Body mass index, serum total cholesterol, and risk of gastric high-grade dysplasia: A case-control study among Chinese adults. Medicine (Baltimore). 2016; 95(35):e4730.
22. Knekt P, Reunanen A, Aromaa A, Heliövaara M, Hakulinen T, Hakama M. Serum cholesterol and risk of cancer in a cohort of 39,000 men and women. J Clin Epidemiol. 1988; 41(6):519-530.
23. Caruso MG, Notarnicola M, Cavallini A, Di Leo A. 3-Hydroxy-3-methylglutaryl coenzyme A reductase activity and low-density lipoprotein receptor expression in diffuse-type and intestinal-type human gastric cancer. J Gastroenterol. 2002; 37(7):504-508.
24. Kim HY: Metabolic syndrome is associated with gastric dysplasia. Eur J Gastroenterol Hepatol. 2011; 23(10):871-875.
25. Jung MK, Jeon SW, Cho CM, Tak WY, Kweon YO, Kim SK. Hyperglycaemia, hypercholesterolaemia and the risk for developing gastric dysplasia. Dig Liver Dis. 2008; 40(5):361-365.
26. Bigler RD, Khoo M, Lund-Katz S, Scerbo L, Esfahani M. Identification of low density lipoprotein as a regulator of Fc receptor-mediated phagocytosis. Proc Natl Acad Sci U S A. 1990; 87(13):4981-4985.
27. Han M, Wang H, Yang S, Zhu S, Zhao G, Shi H, Li P: Triglyceride glucose index and Atherogenic index of plasma for predicting colorectal neoplasms in patients without cardiovascular diseases. Front Oncol. 2022; 12:1031259.
28. Karra S, Gurushankari B, Rajalekshmy MR, Elamurugan TP, Mahalakshmy T. Diagnostic Utility of NLR, PLR and MLR in Early Diagnosis of Gastric Cancer: an Analytical Cross-Sectional Study. J Gastrointest Cancer. 2023.
29. Guevara-Aguirre J, Peña G, Acosta W, Pazmiño G. Cancer in growth hormone excess and growth hormone deficit. Endocr Relat Cancer. 2023; 30(10).
30. Wang K, Yu Y, Wang W, Jiang Y, Li Y, Jiang X, Qiao Y, Chen L. Targeting the E3 ligase NEDD4 as a novel therapeutic strategy for IGF1 signal pathway-driven gastric cancer. Oncogene. 2023; 42(14):1072-1087.
31. Huang D, Shin WK, De la Torre K, Lee HW. Association between metabolic syndrome and gastric cancer risk: results from the Health Examinees Study. Gastric Cancer. 2023; 26(4):481-492.
32. Nam SY, Jeong J, Jeon SW. Constant Association between Low High-Density Lipoprotein Cholesterol and Gastric Cancer Regardless of Site. J Obes Metab Syndr. 2023; 32(2):141-150.
33. Pih GY, Gong EJ, Choi JY, Kim MJ, Ahn JY, Choe J, Bae SE. Associations of Serum Lipid Level with Gastric Cancer Risk, Pathology, and Prognosis. Cancer Res Treat. 2021; 53(2):445-456.
34. Chen T, Wu F, Chen FM, Tian J, Qu S. Variations of very low-density lipoprotein receptor subtype expression in gastrointestinal adenocarcinoma cells with various differentiations. World J Gastroenterol. 2005; 11(18):2817-2821.