Cancer Biomarkers 41 (2024) 25–40
DOI 10.3233/CBM-240023
IOS Press

Machine learning identifies a 5-serum cytokine panel for the early detection of chronic atrophy gastritis patients

Fangmei An[a,1], Yan Ge[b,1], Wei Ye[b], Lin Ji[a], Ke Chen[a], Yunfei Wang[a], Xiaoxue Zhang[a], Shengrong Dong[a], Yao Shen[a], Jiamin Zhao[a], Xiaojuan Gao[a], Simon Junankar[c], Robin Barry Chan[b], Dimitris Christodoulou[b], Wen Wen[b], Peihua Lu[d,*] and Qiang Zhan[a,*]

[a]*Department of Gastroenterology, Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi People's Hospital, Wuxi Medical Center, Nanjing Medical University, National Clinical Research Center for Digestive Diseases (Xi 'an) Jiangsu Branch Wuxi, Jiangsu, China*

[b]*AliveX Biotech, Shanghai, China*

[c]*AliveX Biotech, Darlington, Australia*

[d]*Department of Medical Oncology, Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi People's Hospital, Wuxi Medical Center, Nanjing Medical University, Wuxi, Jiangsu, China*

Received 3 March 2024
Accepted 19 July 2024

Abstract.

BACKGROUND: Chronic atrophy gastritis (CAG) is a high-risk pre-cancerous lesion for gastric cancer (GC). The early and accurate detection and discrimination of CAG from benign forms of gastritis (e.g. chronic superficial gastritis, CSG) is critical for optimal management of GC. However, accurate non-invasive methods for the diagnosis of CAG are currently lacking. Cytokines cause inflammation and drive cancer transformation in GC, but their utility as a diagnostic for CAG is poorly characterized.

METHODS: Blood samples were collected, and 40 cytokines were quantified using a multiplexed immunoassay from 247 patients undergoing screening via endoscopy. Patients were divided into discovery and validation sets. Each cytokine importance was ranked using the feature selection algorithm Boruta. The cytokines with the highest feature importance were selected for machine learning (ML), using the LightGBM algorithm.

RESULTS: Five serum cytokines (IL-10, TNF-α, Eotaxin, IP-10 and SDF-1a) that could discriminate between CAG and CSG patients were identified and used for predictive model construction. This model was robust and could identify CAG patients with high performance (AUC $= 0.88$, Accuracy $= 0.78$). This compared favorably to the conventional approach using the PGI/PGII ratio (AUC $= 0.59$).

CONCLUSION: Using state-of-the-art ML and a blood-based immunoassay, we developed an improved non-invasive screening method for the detection of precancerous GC lesions.

FUNDING: Supported in part by grants from: Jiangsu Science and Technology Project (no. BK20211039); Top Talent Support Program for young and middle-aged people of Wuxi Health Committee (BJ2023008); Medical Key Discipline Program of Wuxi Health Commission (ZDXK2021010), Wuxi Science and Technology Bureau Project (no. N20201004); Scientific Research Program of Wuxi Health Commission (Z202208, J202104).

Keywords: Chronic Atrophy Gastritis (CAG), Chronic Superficial Gastritis (CSG), Screening, Cytokines, Machine Learning (ML) Algorithms

[1]Contributed equally to this paper.

*Corresponding authors: Peihua Lu, Department of Medical Oncology, Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi People's Hospital, Wuxi Medical Center, Nanjing Medical University, Wuxi, Jiangsu, China. E-mail: lphty1_1@163.com. Qiang Zhan, Department of Gastroenterology, Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi People's Hospital, Wuxi Medical Center, Nanjing Medical University, National Clinical Research Center for Digestive Diseases (Xi'an) Jiangsu Branch Wuxi, Jiangsu, China. E-mail: qiangzhan@njmu.edu.cn.

1. Introduction

Gastric cancer (GC) is the fifth most common malignant tumor in the world, both in terms of incidence and mortality [1]. In China, the incidence of GC is extremely high, only lung cancer and colorectal cancer rank higher [2]. A report on cancer incidence in 2020 stated that there were approximately 479,000 new cases of GC (ranked second) and 374,000 deaths related to GC (ranked third) in China [2]. GC is thought to develop in a progressive manner, and Correa et al. proposed the following gastric cancer progression cascade (from benign to malignant stages): Chronic Superficial Gastritis (CSG) – Gastric Atrophy (GA) – Gastric Intestinal Metaplasia (GIM) – Dysplasia – GC. This progressive cascade is currently recognized as the main mode of formation of intestinal type GC[3]. CSG is the more benign, chronic mucosal inflammation stage often caused by *Helicobacter pylori* (*H. pylori*), at this stage the mucosal structure destruction is typically reversible. The GA and GIM phenotypes are collectively known as Chronic Atrophy Gastritis (CAG) [4]. CAG consists primarily of precancerous lesions that can develop into GC, and these are thus the gastritis states that need to be identified in an accurate and timely manner to improve patient outcome. In 2020, the American Society of Gastroenterology conducted a meta-analysis of the medical history of GIM patients; it was found that 1 in 100 GIM patients developed GC during the 5-year follow-up period [5]. Another study showed that the incidence rate of GC in the 5 years following CAG diagnosis was 0.6% [6]. *H. pylori* has been listed as a human carcinogen by international cancer research institutes, and is known to promote the development of GC [7]. In recent years, many studies have suggested that antibiotic eradication of *H. pylori* reduces the risk of GC if enacted prior to the development of CAG, unfortunately beyond this point, the malignant progression to GC can no longer be reversed [8]. Thus, the diagnosis of CAG patients is extremely important so that these patients can be monitored more closely to allow for the early detection of those patients that progress to GC.

The gold standard method for CAG diagnosis is endoscopy (gastroscopy) combined with gastric biopsy, a relatively invasive procedure that can cause anxiety and discomfort for the patient. As a result, a large proportion of the population are reluctant to undergo this type of examination. This delays the identification of CAG patients who would benefit from close monitoring for progression to GC. In addition, endoscopic examination is expensive and consumes significant hospital resources. There is thus a clear clinical need for a non-invasive, scalable, and accurate diagnostic method for the early detection of CAG patients. Existing approaches such as measuring the ratio of serum concentrations of pepsinogen I to II (PGI and PGII), or the commercialized GastroPanel, an ELISA-based approach that measures serum gastrin-17 (G17), PGI, PGII, and *H. pylori* antibodies [9], are the most widely used non-invasive approach for the diagnosis of CAG [10]. However, these are not accurate enough to replace endoscopy for the diagnosis of CAG patients [11]. Therefore, the gap for an accurate diagnostic method to detect CAG patients remains.

In the last couple of decades, cytokines have been implicated in multiple aspects of tumor development and progression [12], including GC [13]. Chronic inflammation in the stomach, often caused by *H. pylori* infection is a known risk factor for GC. Cytokines, particularly pro-inflammatory ones such as TNF-α, IL-1, and IL-6 are produced during chronic inflammation and can contribute to the development of cancer [14,15]. Studies have shown that TNF-α and IL-1 can recruit immune cells and induce inflammation [16,17], which can have both positive and negative effects on tumor progression. In addition, cytokines secreted by immune cells and tumor cells within the tumor microenvironment can influence tumor growth, invasion, and metastasis. For example, cytokines TGF-β and VEGF can promote angiogenesis necessary for tumor growth and metastasis [18,19]. Specific to GC, the expression of cytokine/chemokine/growth factor markers is higher in the gastric tumor stroma compared to the normal gastric stroma [20]. Particularly, circulating cytokines play a pivotal role in the pathogenesis of the cancer [21]. A serum cytokine biomarker panel has been proposed for GC diagnosis [22], another study reported that increased circulating levels of IL-8 may indicate increased risk of GC [23] and serum IL-17A level increases significantly in *H. pylori*-infected patients with premalignant gastric lesions [24]. Furthermore, studies have shown that levels of circulating cytokines, IL-6, VEGF, and TNF-α, are associated with an increased risk of colorectal cancer [25]. Therefore, we hypothesize that the measurement of circulating cytokines has the potential to become a non-invasive and accurate method for CAG diagnosis.

In this, a diagnostic performance study, blood samples were collected from 247 patient prior to a routine endoscopy procedure [26]. The patients were divided into two groups: CSG and CAG, according to the gastric mucosa histopathological grading – with CSG as

Table 1
The characteristics of the clinical features between the CAG and CSG

Variable	CSG ($n = 157$)	CAG ($n = 90$)
Age, years (mean ± SD)	56.35 ± 7.98	58.98 ± 8.08
Sex, (n)		
Male	64	39
Female	93	51
H. Pylori (n)		
Positive	75	43
Negative	63	47
NA	19	–
Standard Gastritis Markers (ng/mL)		
PGI (mean ± SD)	180.41 ± 92.25	175.74 ± 87.48
PGII (mean ± SD)	15.27 ± 7.36	16.09 ± 8.63
PGI/PGII (mean ± SD)	11.94 ± 2.34	11.25 ± 2.23*

CSG: Chronic Superficial Gastritis, CAG: Chronic Atrophy Gastritis, SD: standard deviation, NA: Not Available, $^*P < 0.05$, compared with CSG group.

the "control" group. We quantified 40 cytokines in each serum sample, using the Meso Scale Discovery (MSD) enzyme linked immunoassay platform. We identified those cytokines that exhibited the highest CAG prediction capacity utilizing the Boruta algorithm [27,28]. We then selected these cytokines for the establishment of a predictive machine-learning (ML) model, based on the LightGBM algorithm, to accurately identify CAG from non-CAG patients. LightGBM is an ensemble model of decision trees used for classification and regression prediction, shown to have prediction precision and model stability [29]. Taken together, we successfully developed a scalable, non-invasive lab-based assay in combination with a ML predictive model for the early detection of CAG versus CSG patients.

2. Materials and methods

2.1. Ethics statement

All studies involving human participants were reviewed and approved by Institutional Ethics Committee of Nanjing Medical University (KY21069). The patients provided their written informed consent prior to enrollment.

2.2. Study population inclusion and exclusion criteria

The target population, aged between 40 and 69 years, were invited to participate in the study from January to December of 2021. In China, the incidence of GC rises rapidly after the age of 40 [30]. Additionally, as this study was a community-based GC screening project, individuals over the age of 70 were excluded as they have a higher risk of complications associated with undergoing gastroscopy and biopsy. Individuals were excluded from the study for any of the following reasons: history of gastric surgery, including endoscopic mucosal resection or submucosal dissection, coagulation disorders or severe cardiovascular or cerebrovascular diseases, liver, kidney, or psychiatric disorders. In addition, patients were excluded if taking proton pump inhibitors within 2 weeks because these can influence PG Levels in the serologic tests. In addition, patients who routinely take an antiplatelet drug such as aspirin were excluded, as these drugs can cause gastrointestinal bleeding during the endoscopic biopsy procedure.

2.3. Sample size calculation

We based our sample size calculation on obtaining a sensitivity and specificity of 0.7. From prior experience approximately 30–40% of patients screened at our hospital have CAG. Using this information, we calculated that for a sensitivity of 0.7 and with 35% positivity rate we would require 230 patients for this study [31].

2.4. Patient characteristics

A total of 247 samples, including 157 CSG patients (the control group for this study) and 90 CAG patients, met the inclusion criteria and were included in our study. The overall male to female ratio was 103:144. The median age was 58. The clinical characteristics and cytokines levels of the two cohorts are summarized in Tables 1 and 2 respectively.

2.5. Human blood sample collection and serum preparation

Blood samples were obtained immediately prior to the endoscopy procedure for 247 patients as part of a community-based screening program for gastric cancer between January and December of 2021. The blood

Table 2
The cytokines concentration between the CAG and CSG. The average concentration (pg/ml) $\pm$ the standard deviation for the 37 cytokines. The adjusted p-value between CAG vs CSG patients were calculated for all these features, with the significantly different features marked with bold

Cytokine	CSG ($n = 157$)	CAG ($n = 90$)	Adjusted_P_value
GM-CSF	1.19 $\pm$ 12.07	0.12 $\pm$ 0.11	0.573
IFN-γ	14.65 $\pm$ 27.0	13.81 $\pm$ 10.84	0.822
IL-10	0.51 $\pm$ 0.83	0.37 $\pm$ 0.25	0.526
IL-17A	1.71 $\pm$ 1.36	1.62 $\pm$ 1.08	0.728
IL-1β	0.19 $\pm$ 0.31	0.15 $\pm$ 0.19	0.427
IL-4	0.12 $\pm$ 0.78	0.05 $\pm$ 0.05	0.595
IL-5	0.7 $\pm$ 0.75	0.57 $\pm$ 0.41	0.463
IL-6	1.12 $\pm$ 0.61	1.28 $\pm$ 1.12	0.444
IL-8	4.09 $\pm$ 10.16	3.05 $\pm$ 4.95	0.565
ENA-78	519.08 $\pm$ 924.45	532.03 $\pm$ 732.69	0.935
Eotaxin-2	1041.79 $\pm$ 949.26	898.57 $\pm$ 371.24	0.4
IL-12p70	2.53 $\pm$ 3.41	1.78 $\pm$ 1.79	0.337
IL-13	9.4 $\pm$ 8.2	8.41 $\pm$ 6.43	0.522
IL-1RA	185.2 $\pm$ 350.63	173.76 $\pm$ 186.0	0.843
IL-2Ra	961.68 $\pm$ 305.48	1017.43 $\pm$ 318.22	0.385
IL-33	1.88 $\pm$ 2.44	1.47 $\pm$ 1.49	0.397
TNF-α	5.06 $\pm$ 4.67	3.06 $\pm$ 2.16	**0.004**
VEGF-A	54.41 $\pm$ 95.34	39.72 $\pm$ 53.04	0.372
Eotaxin	206.25 $\pm$ 72.75	262.93 $\pm$ 88.93	**5.76e-06**
Eotaxin-3	94.34 $\pm$ 838.24	30.81 $\pm$ 88.01	0.629
IP-10	459.47 $\pm$ 211.84	519.96 $\pm$ 244.58	0.32
MCP-1	86.29 $\pm$ 25.39	88.38 $\pm$ 23.54	0.648
MCP-4	72.96 $\pm$ 52.06	84.0 $\pm$ 56.69	0.507
MDC	955.84 $\pm$ 334.25	984.48 $\pm$ 302.8	0.644
MIP-1α	15.89 $\pm$ 7.23	16.16 $\pm$ 6.4	0.858
MIP-1β	74.52 $\pm$ 41.71	83.15 $\pm$ 46.71	0.509
MIP-3α	22.11 $\pm$ 21.15	29.43 $\pm$ 60.61	0.426
TARC	68.36 $\pm$ 76.05	76.72 $\pm$ 102.51	0.641
G-CSF	6.78 $\pm$ 2.37	7.23 $\pm$ 2.82	0.337
IL-12/IL-23p40	191.91 $\pm$ 75.54	207.86 $\pm$ 110.22	0.354
IL-15	3.13 $\pm$ 0.86	3.44 $\pm$ 0.8	0.055
IL-16	352.02 $\pm$ 881.33	402.35 $\pm$ 1045.45	0.796
IL-7	3.26 $\pm$ 4.53	3.24 $\pm$ 4.48	0.971
SDF-1a	1624.61 $\pm$ 608.45	1844.76 $\pm$ 520.9	0.055
TNF-β	0.51 $\pm$ 0.27	0.47 $\pm$ 0.28	0.491
MIF	92429.07 $\pm$ 116540.46	71956.35 $\pm$ 86443.54	0.424
MIP-5	13841.92 $\pm$ 7452.53	12392.17 $\pm$ 4733.04	0.523

was collected into two vacutainers (5 ml into each) containing K2 Ethylenediamine tetraacetic acid (EDTA), one sample was used for multiplex cytokine concentration detection by Meso Scale Discovery (MSD) and the other for pepsinogen detection by enzyme-linked immunosorbent assay (ELISA). The blood samples were mixed by gentle inversion and kept at room temperature (RT) for no more than 2 hours, they were then centrifuged at 3,000 $\times$ g for 10 min at 4°C, The upper one-third of the serum supernatant was collected. The serum samples were aliquoted and stored at -80°C until sample analysis.

2.6. Biopsy collection and histological reporting

All participants underwent a biopsy procedure during the gastroscopy procedure. Two biopsies were taken, one from the antrum and one from the small curvature of the stomach for histology examination. We did not do the full 5-point biopsy according to the OLGIM (Operative Link on Gastric intestinal metaplasia assessment) evaluation system because a severity grading of atrophy was not required for this study.

2.7. Diagnosis of the CSG and CAG by histological examination of biopsy samples

Chronic Superficial Gastritis (CSG) was characterized by the gastric mucosa having lymphocytes and plasma cell infiltration without changes in atrophy or intestinal metaplasia by pathological analysis. Gastric Atrophy (GA) refers to the reduction of intrinsic glands in the gastric cavity, the thinning of gastric mucosa, and

the shallowing of gastric pits by pathological analysis. Gastric Intestinal Metaplasia (GIM) is the replacement of gastric mucosa epithelial cells by intestinal mucosa epithelial cells (goblet cells, pan cells and absorptive cells) by pathological analysis [32]. In this study, we collectively referred to GA and GIM as Chronic Atrophic Gastritis (CAG) [33].

There are two types of CAG: a gastric antrum predominant type in patients infected with *H.pylori*, and an autoimmune type limited to the gastric body and fundus [34]. Due to the rare occurrence of the autoimmune type, a distinct subgroup of patients with specific biomarkers gastrin G17 and pepsinogen [35], none of the autoimmune type were included in this study.

2.8. Serum detection of Pepsinogen I (PGI) and Pepsinogen II (PGII)

Serum expressions of Pepsinogen I (PGI) and Pepsinogen II (PGII) were determined using Enzyme-linked immunosorbent assay (ELISA). A total of 5 ml of blood was collected from each of the 247 patients, and the serum was separated from the blood by centrifugation at 3000 $\times$ g for 10 minutes at 4°C. The PGI and PGII concentrations (in nanograms per liter) were measured using an ELISA kit (Biohit Co., Ltd., Finland) immediately according to the manufacturer's instructions. Briefly, both standard and serum samples were added to the ELISA plates (100 μl/well), followed by a 2-hour incubation at 37°C. After that, detection solution A was added (100 μl/well) and incubated at 37°C for 1 hour. The plates were washed 3 times, and detection solution B was added (100 μl/well) and incubated at 37°C for 1 hour, followed by 5 washes. Tetramethylbenzidine (TMB) substrate was added (90 μl/well), and the plates were incubated at 37°C for 15–25 minutes. Finally, the reaction was terminated by adding termination solution (50 μl/well), and the optical density (OD) value was read out at 450 nm.

2.9. Cytokine concentration quantification by multiplex Meso Scale Discovery (MSD) enzyme linked immunoassay

The following 40 cytokines: Tumor necrosis factor-α (TNF-α). Tumor necrosis factor-β (TNF-β), Eotaxin, Eotaxin-3, interferon-inducible protein-10 (IP-10), Stromal cell-derived factor 1α (SDF-1α), Macrophage migration inhibitory factor (MIF), Macrophage Inflammatory Protein-5 (MIP-5), Interferon gamma (IFN-γ), Macrophage Inflammatory Protein-3α (MIP-3α), MCP-4, Eotaxin-2, Interleukin-12p70 (IL-12p70), Macrophage-derived chemokine (MDC), Macrophage inflammatory protein-1β (MIP-1β), Interleukin-13(IL-13), Interleukin-15(IL-15), Interleukin 17A(IL-17A), Interleukin-1RA (IL-1RA), Interleukin-1β (IL-1β), Interleukin-2 (IL-2), Interleukin-4 (IL-4), Interleukin-5 (IL-5), Interleukin-6 (IL-6), Interleukin-8 (IL-8), Interleukin-10 (IL-10), Interleukin 12 (IL-12/IL-23p40), Interleukin-15 (IL-15), Interleukin 16 (IL-16), Interleukin-1α (IL-1α), Interleukin7 (IL-7), Interleukin-1RA (IL-1RA), Interleukin-2Ra (IL-2Ra), Interleukin-3 (IL-3), Interleukin-33 (IL-33), Granulocyte-macrophage colony-stimulating factor (GM-CSF), Epithelial-neutrophil activating peptide (ENA-78), and Vascular endothelial growth factor A (VEGF-A) were measured using the Meso Scale Discovery platform (MSD) according to the manufacturer's instructions (product insert of K15067, MSD, Rockville, MD).

MSD is an indirect binding quantitative electrochemiluminescence (ECL) method designed to detect multiple antibodies in human serum simultaneously. In the MSD platform, biological reagents (antigens or antibodies) are coated on the carbon electrode at the bottom of a micro titer plate. Bound samples are then incubated with a ECL labeled (SULFO-TAG) detection antibody. Following application of an electrical current across the carbon electrode, the SULFO-TAG reacts with the Ru(bpy)32+ reagent and Tripropylamine (TPA) catalyst in the detection buffer and resulting in the emission of light that can be quantitated at 620 nm with a CCD camera.

2.10. Data pre-processing and statistics

Three cytokines, IL-2, IL-3 and IL-1α were not detected in 42.5%, 69.2% and 45.6% of patients, and therefore, were excluded in subsequent machine learning analyses. For the remaining 37 cytokines, the average missing rate is 0.52% (the median is 0). The few missing values for these cytokine concentrations were filled in using the lower limit of detection (LLOD) of the MSD assay divided by 10, reflecting the fact that the concentration of these cytokines is too low to be detected. Student t-test was used to assess the statistical significance of cytokine levels between the different groups. An overall adjusted *p-value* of 0.05 using Benjamini-Hochberg for multiple testing correction was considered as the significance cut-off for all statistical analyses. All statistical analyses and machine learning algorithms for this study were performed using R version 4.1.0 and Python version 3.8.8 [36]. In

order to reduce overfitting in the downstream analyses from the feature selection and machine learning, we randomly divided the whole dataset into a training dataset (Dataset1) and a validation dataset (Dataset2) for five times, while maintaining the proportion of CAG and CSG for each division (Supplementary Table 1). The Dataset2 here is considered as a validation dataset since it is a dataset the algorithm did not include in the training process and was only used for validation. In detail, the Dataset1 includes 125 CSG and 72 CAG, while Dataset2 includes 32 CSG and 18 CAG.

2.11. Feature selection

To reduce over-fitting and improve the generalizability of the model feature selection was employed [37]. The machine learning model then only utilizes the selected features as its input.

Feature selection was performed using the Boruta algorithm [27,28] and other redundant or non-important features were discarded. Boruta was used to select the features that correlated with the dependent variable. Boruta feature selection was performed on each of the 5 training datasets and the common features from all five data splits were then used by the ML algorithm. The Boruta method is a wrapping algorithm that is based on the random forest (RF) method. It shuffles the original real features to construct shadow features, then joins the real features and shadow features into a feature matrix for training. Finally, the feature importance score of shadow features is used as the reference base to select the feature set truly related to the dependent variable from the real features (see Supplementary Method for more detail).

The Boruta program that we used in this work was Python package Boruta (version 0.3), the default parameters were selected for execution.

2.12. Machine Learning Algorithms: LightGBM

LightGBM is a well-known gradient boosting framework that uses a tree-based learning algorithm [38,39, 40]. It can be regarded as an improved version of Gradient Boosting Decision Tree (GBDT), which uses the negative gradient of the loss function of the current Decision Tree (DT) as an approximation of the residual and fits a new DT recursively. It is designed to be distributed and efficient by incorporating two new technologies, gradient based one side sampling (GOSS) and exclusive feature bunching (EFB) [41], both of which greatly improve the efficiency and ensure the accuracy of classification.

In this study, the features are modeled using LightGBM and the parameters are continuously optimized by means of a grid search features (see Supplementary Methods for more detail). LightGBM program (version 3.3.5) is implemented through a Python package, Scikit-learn [36].

2.13. Model evaluation

Model evaluation is a crucial step in machine learning to assess the performance of a predictive model. There are various metrics that can be used to evaluate the model's performance, and here we mainly use, area under the curve (AUC), area under the precision recall curve (PR-AUC) and accuracy. AUC is the area under the receiver operating characteristic curve (ROC curve) and is a metric that evaluates the model's ability to distinguish between positive and negative classes. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at different threshold values. PR-AUC, the area under the precision-recall curve, is a metric that evaluates the model's ability to identify positive class instances while minimizing false positives. The PR-AUC is calculated as the area under the precision-recall curve. Precision is the fraction of true positives (TP) among the samples predicted as positive (TP + false positives), and recall (or sensitivity) is the fraction of true positives among all the actual positives (TP + false negatives). Accuracy is a metric that measures the ratio of correctly predicted observations (true positive + true negative) to the total number of observations. From the design of these different metrics, AUC and PR-AUC focus on the detection of the positive samples, while the accuracy metric gives the same "importance/weight" on both the positive samples and negative samples. In this study, our aim is to accurately identify the positive samples (CAGs), therefore, we evaluated the predictive models primarily based on AUC and PR-AUC.

2.14. The subgroup analysis of logistic regression model

Hypertension, diabetes, coronary heart disease, stroke, and chronic hepatitis B are common chronic diseases. We examined whether these common chronic diseases were confounding factors in this study. A subgroup analysis using a logistic regression model was performed on these five confounding factors using the R

Table 3
The feature importance based on the Boruta feature selection method, performed five independent times. Five cytokines ranked consistently high, across all five feature selection runs

No.1	Rank	No.2	Rank	No.3	Rank	No.4	Rank	No.5	Rank
TNF-α	1	**TNF-α**	1	**TNF-α**	1	**TNF-α**	1	**TNF-α**	1
Eotaxin	1	**Eotaxin**	1	**Eotaxin**	1	**Eotaxin**	1	**Eotaxin**	1
IL-10	1	**IL-10**	1	**IL-10**	1	**IL-10**	1	**IL-10**	1
IP-10	1	**IP-10**	1	**IP-10**	1	**IP-10**	1	**IP-10**	1
SDF-1a	1	**SDF-1a**	1	**SDF-1a**	1	**SDF-1a**	1	**SDF-1a**	1
IL-12p70	1	MIF	1	IL-12p70	1	MCP-4	1	MIF	1
IL-1RA	1	IL-12p70	1	MCP-4	1	IL-15	1		
MIP-3α	1	MCP-4	1						
IL-15	1	MIP-3α	1						
		PGI/PGII	1						

software (R version 4.2.3), and then the differential effects of the 5-cytokine panel diagnosis were compared in each subgroup dataset. By examining the interaction in the subgroup regression of the generalized linear model (p for interaction), we assessed whether these confounding factors had an impact on the 5-cytokine panel diagnostic effects.

3. Results

3.1. Serum cytokine measurements

For all 247 patients, we collected blood samples and extracted serum. We measured the concentrations of 40 key cytokines in the serum using MSD. The raw cytokine concentration data were cleaned, as described in the Materials and Methods section, leaving data on 37 cytokines for further analysis (Table 2).

3.2. Machine learning based biomarker discovery study design

To optimally utilize the information from our patient cohort and develop a predictive model for CAG patients, we designed the process outlined in Fig. 1. Starting with the cytokine concentration matrix composed of 247 patient samples (including 157 CSGs and 90 CAGs) and 37 cytokine features, we first performed feature analysis in order to identify the most relevant features for our downstream analyses. We calculated the statistical significance for all features among the two groups, to determine the difference of these features between CSG and CAG patients (the adjusted p-value is shown in Table 2). Only two cytokines TNF-a and Eotaxin were found to be significantly different between the groups. Then, we utilized the Boruta algorithm [27,28] to select the features (all cytokine features plus the clinical features) with the highest capacity for discrimination between CSG and CAG patients.

Once we had identified five key features, we developed a machine learning model based on the selected features by utilizing the LGBM model [41]. We divided the whole dataset (a total of 247 samples) into training set (Dataset 1) and test set (Dataset 2), and we repeated this split randomly five times to avoid over-fitting. We used Dataset1 for training with 5-fold cross-validation and Dataset2 for validation (i.e., validation in a dataset the algorithm has never seen).

3.3. Feature Selection enables us to reduce the input features for our subsequent Machine Learning modeling

Clinical features (PGI/PGII ratio, age and gender, Table 1) and cytokine features (Table 2) were first analyzed using the Boruta method [27,28]. For this feature selection procedure, we exclusively used the training set. For every one of the five independent training sets (Supplementary Table 1), we utilized Boruta for feature selection, and calculated the feature importance, ranking their capacity to distinguish CAG vs CSG patients. In turn, we selected the features which ranked first (i.e. rank of "#1") for each separate training (see Fig. 2A and Table 3 for details, see Supplementary Table 2 for all features ranking). Subsequently, we took the intersection of these five separate feature selection attempts, which yielded a total of 5 features: IL-10, TNF- α, Eotaxin, IP-10 and SDF-1a. These features (cytokine levels) were retained as the key features that would be used for the follow-up modeling (see Table 3, in bold). We also did feature selection using another method Lasso [42], these 5 features also rank very high by Lasso (see Supplementary Table 3).

Furthermore, we compared the differences of each feature in the two groups through statistical testing (student's t-test), and we verified that four of these five

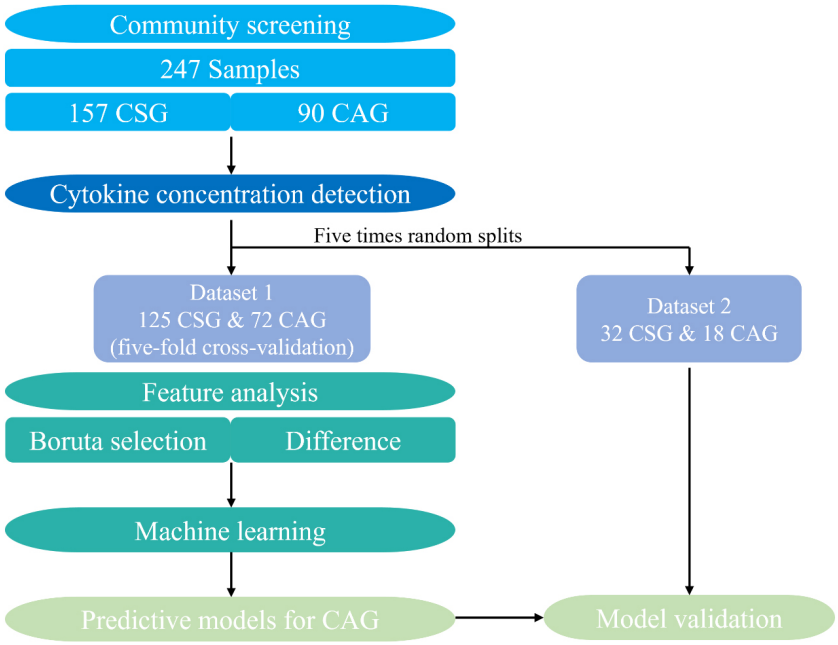

Fig. 1. Schematic showing the details of the ML-based biomarker discovery pipeline following the serum cytokine measurements.

cytokine features, TNF-α, Eotaxin, IP-10 and SDF-1a, demonstrate statistically significant differences between the two groups ($P < 0.05$) (Supplementary Fig. 1), although two of them, IP-10 and SDF-1a, showed no significance after correction for multiple testing (see adjusted P-value in Table 2). In addition, we analyzed the correlation between different features among patients (see Fig. 2B and Supplementary Table 4) and found that some features showed strong correlations with each other. This indicating the redundancy of some features and the benefits of the feature selection performed here. The correlation plot also shows the independent predictive power of the five features we selected here as they did not correlate with each other (Fig. 2B), again shows the validity of the feature selection process. In addition, the 5 selected cytokines showed no significant correlation with age or gender demonstrating these clinical features would not confound the results.

In order to assess whether common health conditions could confound the 5-cytokine panel diagnostic ability, we conducted a subgroup analysis using a logistic regression model on common confounding factors. These results demonstrated that hypertension, diabetes, coronary heart disease, stroke, and chronic hepatitis B all had insignificant interaction values with these cytokines, with the one exception for hypertension with IP-10 levels ($p = 0.009$). Therefore, the 5-cytokine diagnostic panels efficiency is largely unaffected by the above confounding factors (see Supplementary Table 5).

3.4. *Building an accurate predictive model for CAG patients, using LGBM*

After reducing the number of the features to those that show significant predictive capacity for CAG patients, we moved on to develop a predictive model for CAG patients. For our predictive model, considering the relatively small sample size, the light gradient-boosting machine (LGBM) classifier was selected [41]. To avoid overfitting – a typical challenge in machine learning approaches – we randomly divided the whole dataset into training set (Dataset 1) and validation set (Dataset 2) for five times, as specified in the Material and Methods section. For each division, there is a training set as well as its corresponding validation set (Supplementary Table 1). We further performed five-fold-cross-validation on the training set to avoid overfitting to the maximum degree. The LGBM model included the five selected features, namely IL-10, TNF- α, Eotaxin, IP-10 and SDF-1a, as input candidate features. The classifier with the best model performance was selected by continuously iterating the optimization parameters for the training set of each division. Out of all 5 different independent runs, the training set of division 1 (sheet of "Training1" in Supplementary Table 1) is

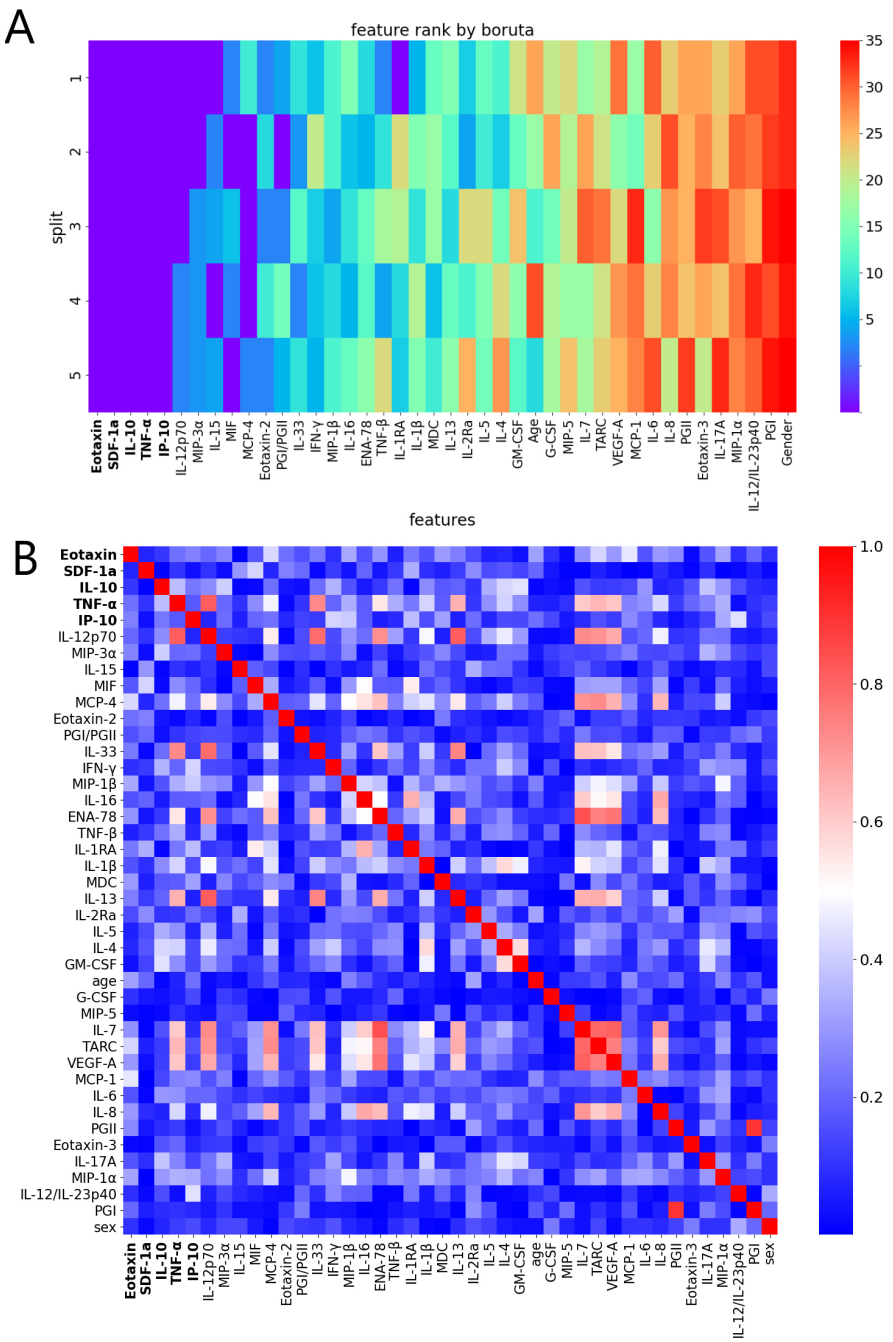

Fig. 2. Feature importance and correlation analysis. (A). Heatmap of feature importance based on the Boruta feature selection method (scores range from 1 to 35), performed five independent times through five random splits of the data. Showing the results for 37 cytokines, Gender, Age, PGI, PGII, and PGI/PGII ratio. The five cytokines with the highest ranking are displayed in bold. (B). Heatmap of absolute values of Pearson correlation coefficients of features among samples. The five selected features show low correlation with each other.

shown in Fig. 3; the performance of the other divisions (Training2 to Training5) are provided in Supplementary Fig. 2, and demonstrate similar performance. In summary, the performance of our model was quite high, with AUC = 0.85, and PRAUC = 0.81. Furthermore, we also constructed the predictive models using other 6 machine learning classifiers, i.e., Logistic Regression (LR), Decision Tree (DT), Support vector Machine

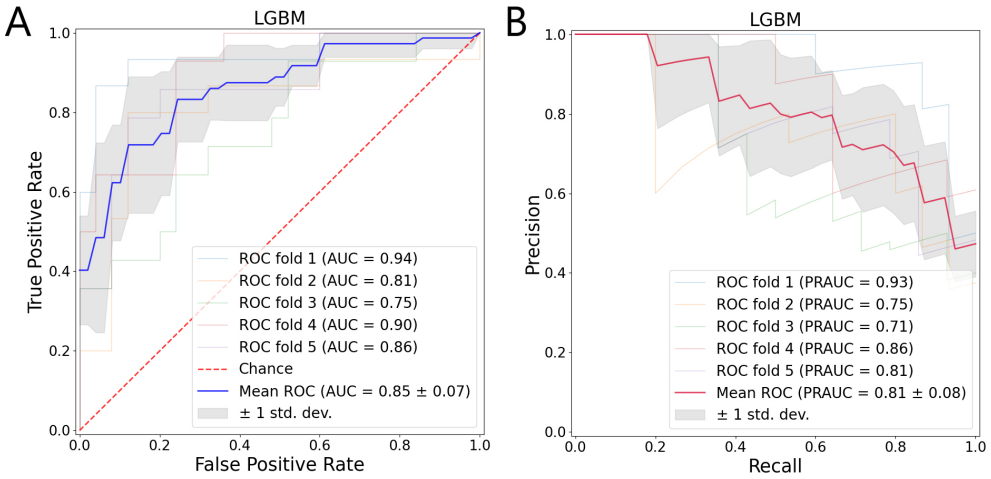

Fig. 3. The training performance of the five cytokine features derived LGBM modelling based on Training1 of division 1. **(A).** The mean AUC curve is shown in blue. The light gray shade shows the standard deviation stemming from the five-fold-cross validation results. **(B).** The PRAUC mean is shown in red. The light gray shade shows the standard deviation stemming from the five-fold-cross validation results. AUC: Area Under the Curve, PRAUC: Precision Recall Area Under the Curve.

Table 4
Robustness and stability of the constructed five cytokine based LGBM model among the five different divisions by interrogating the validation set performance

Division	AUC	PRAUC	Accuracy
1	0.88	0.76	0.78
2	0.83	0.78	0.68
3	0.81	0.69	0.72
4	0.79	0.63	0.74
5	0.85	0.77	0.82
Variance	0.00121	0.00432	0.00292

AUC: Area Under the Curve, PRAUC: Precision Recall Area Under the Curve.

(SVM), Neural Network (NN), Naive Bayes (NB), K Nearest Neighbors (KNN), and compared their performances with the model constructed by LGBM. Overall, the LGBM model's performance is better than other 6 machine learning models (see Supplementary Table 6 and 7).

3.5. Validation

The final LGBM model was validated in its corresponding validation dataset ("Validation1" in Supplementary Table 1). The obtained AUC = 0.88, PRAUC = 0.76, Accuracy = 0.78, Sensitivity = 0.78, Specificity = 0.78 show that our model's capacity to accurately predict CAG patients is high. The validation results are shown in Fig. 4.

In the same way as the training sets (division 1), we also evaluated the performance of each of the other 4 training division derived models on their respective validation set (Supplementary Fig. 3). We have summarized the validation performance of the different training set derived model in Table 4. It clearly shows that the models' performance across all different data partitions stays high and shows robustness. As the main goal of our work here is to accurately identify CAG patients, the metrics, "AUC" and "PRAUC" are most relevant as they focus on the true positive – the CAG patients (as opposed to using the "accuracy" metric (TP+TN)/total samples). By inspecting the variance of the models' performance (AUC, PRAUC and Accuracy) among the five divisions, we can clearly see the limited variance, which indicates the model is very stable and robust against changes in the training datasets. Therefore, we conclude that, the derived model is valid and can be generalized.

To further demonstrate that the five features selected by the Boruta method led to the best prediction algorithm, we compared the performance of our 5-cytokine based ML model with a ML model that includes the complete set of 37 cytokines we measured (Supplementary Fig. 4) as well as a 40-feature ML model which contains all available features (37 cytokines + 3 clinical features) (Supplementary Fig. 5). In addition, we also built a model based only on the significant cytokines, (TNF- α, Eotaxin; adjusted P value < 0.05) (Table 2), with the aim to see whether our feature selection in combination with ML outperforms this simple combination (Supplementary Fig. 6). In summary, our results demonstrate that the performance of our concise 5-feature (cytokine) predictive model is comparable with the performance of the much more complex multi-feature models, and better than the simple statistical

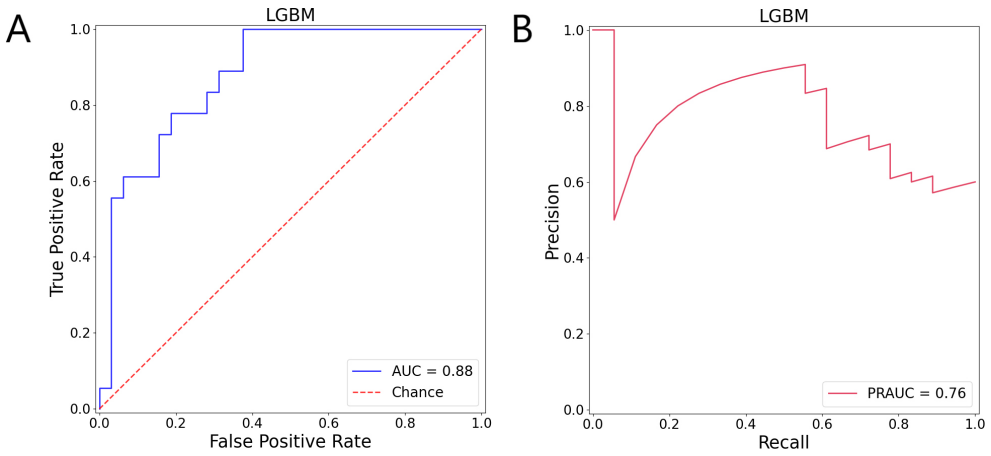

Fig. 4. The performance of the five cytokines based LGBM model in the independent, validation dataset shows high predictive capacity. (A). Area under the curve analysis. (B). Precision recall area under the curve analysis. AUC: Area Under the Curve, PRAUC: Precision Recall Area Under the Curve.

test derived marker model, which demonstrates that the five selected features suffice to empower our predictive model to accurately distinguish CAG patients.

We next examined how our model compared to white-light gastroscopy, a slightly less invasive procedure than the gold-standard gastroscopy/biopsy procedure, as it does not require the collection of biopsies. We found that clinician diagnosis of CAG using white-light gastroscopy (Supplementary Table 1) had a lower accuracy than the 5-cytokine ML model (0.67 vs 0.78) and a dramatically lower sensitivity (0.4 vs 0.78). This sensitivity is similar to a previously published study that indicated white-light gastroscopy had a sensitivity of 0.41 [43]. In contrast white-light gastroscopy had a slightly higher specificity than the 5-cytokine ML model (0.83 vs 0.78). This demonstrates that the 5-cytokine ML model outperforms white-light gastroscopy on accuracy and sensitivity in addition to being a less invasive procedure for the patients.

3.6. *Evaluation of the predictive power of PGI/PGII in our dataset*

As previous approaches to identify CAG patients have utilized PGI and PGII measurements, we wanted to investigate how our approach compared with these more established approaches. We found the PGI/PGII ratio to be significantly different between the two sample groups in this study using a standard student t-test ($P = 0.023$) (Fig. 5A).

First, we employed the criteria that has been used in a number of publications [44], (PG I $\leqslant$ 70 ng/mL and a PG I/PG II ratio of $\leqslant$ 3) to interrogate our full dataset, we found only 3 samples met this criterion (Sample IDs: 1907649, 1908664, 1907503, see detail in sheet "Complete_data" of the Supplementary Table 1). Only one of these was a true positive, CAG (1908664), and other two were false positives, CSG (1907649, 1907503), this gives rise to a Sensitivity = 0.01 and Accuracy = 0.63 (of note, only the PG I/PG II ratio influenced the outcome in this analysis). This performance is dramatically lower than the 5-cytokine approach that we have developed. We next sought to determine whether the utility of PGI/PGII could be improved with the aid of machine learning. To properly evaluate the discriminatory capacity of PGI/PGII on CAG detection, we constructed a model based on logistic regression to predict CAG patients, using the PGI/PGII ratio. We performed five-fold cross validation by employing the entire sample set, which leads to a model with a mean AUC of 0.59 (Fig. 5B). This result strongly suggests that a model based solely on the PGI/PGII information fails to accurately discriminate CAG patients from CSG patients. Taken together, these results demonstrate that with or without machine learning, the common approach for CAG identification utilizing PGI and PGII concentrations cannot achieve the performance of our cytokine-based model.

In addition, we tested whether the addition of the remaining clinical features (age and gender) to the PGI/PGII ratio could improve the prediction performance. We therefore trained a LGBM model using these three features and evaluated the performance in its capacity to discriminate CAG vs CSG patients. These results confirm once more that the clinical features alone do not have strong predictive capacity (see Supplemen-

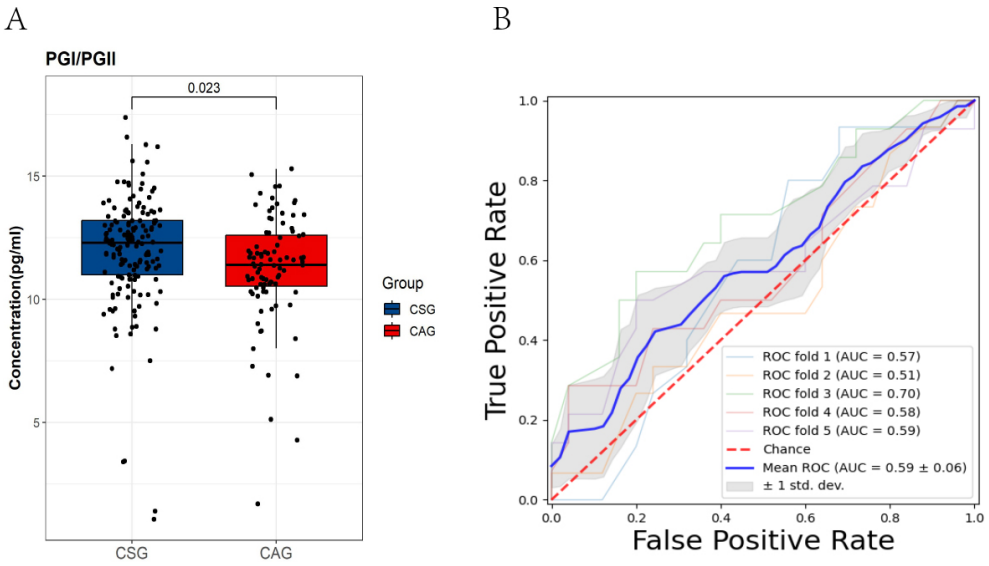

Fig. 5. (A). The difference of PGI/PGII between CSG and CAG patients is significant (p-value $= 0.023$, using student's t-test). (B). The performance of PGI/PGII in classifying CAG with the logistic regression: the classifier does not have strong discrimination capacity. CSG: Chronic Superficial Gastritis; CAG: Chronic Atrophic Gastritis; PGI: pepsinogen I; PGII: pepsinogen II.

tary Fig. 7), and is in line with our previous findings, which demonstrated that their feature importance did not typically rank highly (Fig. 2A).

4. Discussion

Detecting potential GC patients as early as possible is of paramount importance for optimal treatment outcomes. If the guidelines for screening, early diagnosis, and treatment of GC, for example in China were followed [45], there would be approximately 300 million people whom would require a gastroscopy. To reduce the burden on both the patient and the healthcare system, an effective and efficient method to screen for the high-risk individuals is critical. With only these high-risk patients required to undergo the more invasive gastroscopy. In this study, using a non-invasive blood sampling approach, we utilized the concentrations of just five circulating cytokines to build a Machine Learning based predictive model, able to discriminate CAG patients with a high confidence.

Circulating cytokines play a pivotal role in the pathogenesis of the cancer [21]. In this study, the concentration of 40 circulating cytokines were detected using an MSD assay. These cytokine features as well as PGI, PGII, age, and sex were ranked using the Boruta algorithm. This led to the identification of 5 key cytokines, namely IL-10, TNF-α, Eotaxin, IP-10 and SDF-1α that were able to discriminate CSG from CAG. Out of these five cytokines, two had significant differences between CSG and CAG patients (Eotaxin and TNF-alpha). It has been observed that CAG patients tend to have increased expression of a number of cytokines and chemokines [46]. Our results are consistent with many other works: TNF-α is an inflammatory cytokine which may play a role in the development of GC [47]. The chemokine (CXCL) family plays an important role in inflammation, with chemokine family members such as CXCL10 (IP-10), CXCL12 (SDF-1α) hypothesized to play a very important role in the pathogenesis of GC and have also been suggested as markers for the development of GC [48]. Eotaxin, a C-C motif chemokine is a chemoattractant primarily for eosinophils but also other immune cell types [49]. Although it has been previously identified as a potential biomarker for GC in a small patient cohort [50], and also in a number of other cancer types, there is currently limited understanding for its direct role in cancer development or progression [49]. While in contrast to our findings a previous study indicated that the level of anti-inflammatory cytokines such as IL-10 were increased in the blood of GC patients and could be used as a potential biomarker for the diagnosis of GC [51]. This could indicate that IL-10 levels fluctuate during the different stages of GC development with them dropping during CAG formation before rising again when patients progress to GC. Of note our machine learning approach did not identify pro-inflammatory cytokines such as IL-6 and IL-1β that are known to associate with several other malig-

nancies as well as other common conditions such as hypertension [52,53].

As cytokines are involved in many biological processes, single cytokine concentration measurements in the context of cancer may be influenced by processes apart from cancer development. For example, the cytokine concentration may differ due to different inflammatory conditions or illnesses. Thus, to get a robust prediction, the screening of a large number of cytokines is necessary to identify those that correlate most closely with disease. These can then be used to develop the machine learning model. However, traditional cytokine measurement methods such as ELISA do not scale well for analyzing multiple cytokines. In contrast, MSD is an immunoassay-based platform that is designed to detect multiple cytokines in human serum simultaneously. MSD is more efficient and sensitive than ELISA. The power of MSD Multiplex has previously been utilized to analyze immune responses to SARS-COV-2 [54] and to identify other circulating cytokine biomarkers [55, 56]. In this study, we measured the concentration of 40 cytokines. From these 40 cytokines we identified 5 cytokines that could discriminate CAG from CSG. These cytokines were chosen in an unbiased manner using the Boruta algorithm. The analysis showed that predictions based on these 5 cytokines is robust (Table 4), which indicates that they may also play an important role in gastric cancer development. This would be an interesting avenue for future research.

CAG is widely recognized as one of the main precursors to intestinal-type GC, once a patient is diagnosed with CAG, there is no proven treatment including *H. pylori* eradication that can effectively reverse disease progression [57]. Thus, the early diagnosis and management of CAG patients are important to allow for the closer monitoring of patients most likely to progress to GC. Previously, one of the most frequently used non-invasive clinical biomarkers for early CAG patient detection, were the values of PG I and the PG I/PG II ratio. In particular, a value of PG I $\leqslant$ 70 ng/mL and a PG I/PG II ratio of $\leqslant$ 3 has been frequently used in the CAG diagnosis [58]. In the current study, we found that the ratio of PGI/PGII was significantly different between CAG and CSG patients. However, the logistic regression analysis we performed clearly showed that the PGI/PGII ratio is insufficient to distinguish CAG from CSG patients (Fig. 5B). This is supported by a meta-analysis that PGI/PGII diagnosis values for sensitivity, specificity and diagnostic odds ratio are much lower in CAG [44]. This strongly suggests that using the values of PGI and the PGI/PGII ratio alone would not have enough accuracy for the diagnosis of CAG [11]. In contrast, our cytokines-based machine learning method can accurately discriminate CAG from CSG (AUC on validation dataset was 0.88), which has the potential to replace invasive endoscopy method as the reliable CAG detection method in the future. It is important to note that the further validation of this 5-cytokine panel and machine learning algorithm approach on additional patient cohorts would be critical before this approach could be adopted in the clinic. Future studies could also apply the multiplexed cytokine-based machine learning method described in this study as a framework for identifying predictive biomarkers for other cancer types.

Machine learning is increasingly gaining traction in the medical domain, demonstrating highly promising results [59]. Specifically, in recent years, LGBM models have been widely employed for calculating individual patient's risk of developing cancer [60,61,62]. These models combine conventional machine learning techniques with deep learning technology, allowing for the extraction of knowledge regarding cancer molecular mechanisms from diverse signals, such as multi-omics data encompassing genomics, proteomics, epigenetics, and transcriptomics. Consequently, this enables more precise cancer risk prediction. A recent study highlighted the efficacy of an LGBM model in accurately diagnosing ovarian cancer, thereby enhancing the overall effectiveness of cancer prediction [63]. Furthermore, LGBM models find utility in analyzing the pathogenesis and predict the treatment outcomes of liver cancer, where such models were utilized in helping doctors better identify high-risk patients, provide more effective treatment plans, and predict the development of the disease more accurately [64]. It has been demonstrated that LightGBM exhibits superior performance compared to other algorithms in terms of prediction precision, model stability, as well as computing efficiency through a series of benchmark tests [29]. Our results further confirm that LightGBM is the best algorithm for predictive model generation. In line with these studies, our result here also confirms the validity of the machine learning approach in cancer risk prediction, in particular, the early CAG detection along the GC development.

4.1. Limitations of the study

It is important to note that this was a single center study and that further validation of this 5-cytokine panel machine learning algorithm in additional patient cohorts would be critical before this approach could be adopted in the clinic. Furthermore, while the incorporation of

5 cytokines in this model rather than a single cytokine measurement does increase the specificity of this test it is possible that some other cancer types or diseases may lead to a similar cytokine profile and thus lead to false positive results.

As there were no autoimmune type CAG patients included in this study, further studies would need to be conducted to demonstrate whether this 5 cytokine panel has any utility in this patient subgroup.

Finally, while most components of the GastroPanel, the alternate serum screening method, were compared to the 5-cytokine panel model, a direct comparison with the GastroPanel was not included in this study, we thus cannot definitively state that this new system is superior to the GastroPanel.

5. Conclusions

In this study, we developed a model constructed from the serum concentration of 5 cytokines that can predict whether a patient has CAG with high accuracy. This method is an improvement over the traditional methods utilizing PGI and PGII concentrations. We believe our results can pave the way for the introduction of a 5-cytokine lab assay into standard clinical practice, enabling the earlier, more accurate diagnosis of patients. While at the same time reducing the burden (in terms of both time and cost) on the health system and patient.

Conflict of interest

The authors have no conflict of interest to declare.

Supplementary data

The supplementary files are available to download from http://dx.doi.org/10.3233/CBM-240023.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray, Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, *CA: a Cancer Journal For Clinicians* **71** (2021), 209–249.

[2] W. Cao, H.-D. Chen, Y.-W. Yu, N. Li and W.-Q. Chen, Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020, *Chinese medical journal* **134** (2021), 783–791.

[3] P. Correa, Human gastric carcinogenesis: a multistep and multifactorial process – first American Cancer Society award lecture on cancer epidemiology and prevention, *Cancer research* **52** (1992), 6735–6740.

[4] M. Banks, D. Graham, M. Jansen, T. Gotoda, S. Coda, M. Di Pietro, N. Uedo, P. Bhandari, D.M. Pritchard and E.J. Kuipers, British Society of Gastroenterology guidelines on the diagnosis and management of patients at risk of gastric adenocarcinoma, *Gut* **68** (2019), 1545–1575.

[5] A.J. Gawron, S.C. Shah, O. Altayar, P. Davitkov, D. Morgan, K. Turner and R.A. Mustafa, AGA technical review on gastric intestinal metaplasia – natural history and clinical outcomes, *Gastroenterology* **158** (2020), 705–731. e705.

[6] A.C. de Vries, N.C.T. van Grieken, C.W.N. Looman, M.K. Casparie, E. de Vries, G.A. Meijer and E.J. Kuipers, Gastric cancer risk in patients with premalignant gastric lesions: a nationwide cohort study in the Netherlands, *Gastroenterology* **134** (2008), 945–952.

[7] M. Amieva and R.M. Peek, Pathobiology of Helicobacter pylori-Induced Gastric Cancer, *Gastroenterology* **150** (2016), 64–78.

[8] H.-N. Chen, Z. Wang, X. Li and Z.-G. Zhou, Helicobacter pylori eradication cannot reduce the risk of gastric cancer in patients with intestinal metaplasia and dysplasia: evidence from a meta-analysis, *Gastric Cancer: Official Journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association* **19** (2016), 166–175.

[9] K. Syrjänen, A Panel of Serum Biomarkers (GastroPanel® in Non-invasive Diagnosis of Atrophic Gastritis. Systematic Review and Meta-analysis, *Anticancer Research* **36** (2016), 5133–5144.

[10] N. Chapelle, P. Petryszyn, J. Blin, M. Leroy, C. Le BerreâĂŘScoul, I. Jirka, M. Neunlist, D. Moussata, D. Lamarque and R. Olivier, A panel of stomachâĂŘspecific biomarkers (GastroPanel® for the diagnosis of atrophic gastritis: A prospective, multicenter study in a low gastric cancer incidence area, *Helicobacter* **25** (2020), e12727.

[11] C. Grad, A. Pop, E. Gaborean, S. Grad and D. Dumitrascu, Value of GastroPanel in the diagnosis of atrophic gastritis, *Experimental and Therapeutic Medicine* **22** (2021), 1–8.

[12] D. Hanahan and R.A. Weinberg, Hallmarks of Cancer: The Next Generation, *Cell* **144** (2011), 646–674.

[13] K.A. Bockerstett and R.J. DiPaolo, Regulation of Gastric Carcinogenesis by Inflammatory Cytokines, *Cellular and Molecular Gastroenterology and Hepatology* **4** (2017), 47–53.

[14] S. Kuninaka, T. Yano, H. Yokoyama, Y. Fukuyama, Y. Terazaki, T. Uehara, T. Kanematsu, H. Asoh and Y. Ichinose, DIRECT INFLUENCES OF PRO-INFLAMMATORY CYTOKINES (IL-1β, TNF-α, IL-6) ON THE PROLIFERATION AND CELL-SURFACE ANTIGEN EXPRESSION OF CANCER CELLS, *Cytokine* **12** (2000), 8–11.

[15] M. Niiya, K. Niiya, T. Kiguchi, M. Shibakura, N. Asaumi, K. Shinagawa, F. Ishimaru, K. Kiura, K. Ikeda, H. Ueoka and M. Tanimoto, Induction of TNF-α, uPA, IL-8 and MCP-1 by doxorubicin in human lung carcinoma cells, *Cancer Chemotherapy and Pharmacology* **52** (2003), 391–398.

[16] C. Liu, D. Chu, K. Kalantar-Zadeh, J. George, H.A. Young and G. Liu, Cytokines: From Clinical Significance to Quantification, *Advanced Science* **8** (2021), 2004433.

[17] H. Zhao, L. Wu, G. Yan, Y. Chen, M. Zhou, Y. Wu and Y. Li, Inflammation and tumor progression: signaling pathways and targeted intervention, *Signal Transduction and Targeted Therapy* **6** (2021), 263.

[18] S.B. Jakowlew, Transforming growth factor-β in cancer and metastasis, *Cancer and Metastasis Reviews* **25** (2006), 435–457.

[19] F. Khodabakhsh, P. Merikhian, M.R. Eisavand and L. Farahmand, Crosstalk between MUC1 and VEGF in angiogenesis

and metastasis: a review highlighting roles of the MUC1 with an emphasis on metastatic and angiogenic signaling, *Cancer Cell International* **21** (2021), 200.

[20] U.M. Raja, G. Gopal, S. Shirley, A.S. Ramakrishnan and T. Rajkumar, Immunohistochemical expression and localization of cytokines/chemokines/growth factors in gastric cancer, *Cytokine* **89** (2017), 82–90.

[21] J. Song, A. Li, Y. Qian, B. Liu, L. Lv, D. Ye, X. Sun and Y. Mao, Genetically Predicted Circulating Levels of Cytokines and the Risk of Cancer, *Frontiers In Immunology* **13** (2022), 886144.

[22] D. Wu, P. Zhang, J. Ma, J. Xu, L. Yang, W. Xu, H. Que, M. Chen and H. Xu, Serum biomarker panels for the diagnosis of gastric cancer, *Cancer Medicine* **8** (2019), 1576–1583.

[23] M. Epplein, Y.-B. Xiang, Q. Cai, R.M. Peek, H. Li, P. Correa, J. Gao, J. Wu, A. Michel, M. Pawlita, W. Zheng and X.-O. Shu, Circulating cytokines and gastric cancer risk, *Cancer Causes & Control* **24** (2013), 2245–2250.

[24] C. Della Bella, S. D'Elios, S. Coletta, M. Benagiano, A. Azzurri, F. Cianchi, M. de Bernard and M.M. D'Elios, Increased IL-17A Serum Levels and Gastric Th17 Cells in Helicobacter pylori-Infected Patients with Gastric Premalignant Lesions, in: *Cancers*, 2023.

[25] Í. Coşkun, Í. í–ztopuz and í.F. í–zkan, Determination of IL-6, TNF-Î± and VEGF levels in the serums of patients with colorectal cancer, *Cellular and Molecular Biology* **63** (2017), 97–101.

[26] S.G. Baker and R. Etzioni, Prediagnostic evaluation of multicancer detection tests: Design and analysis considerations, *J Natl Cancer Inst* (2024).

[27] M.B. Kursa, A. Jankowski and W.R. Rudnicki, Boruta – a system for feature selection, *Fundamenta Informaticae* **101** (2010), 271–285.

[28] M.B. Kursa and W.R. Rudnicki, Feature selection with the Boruta package, *Journal of Statistical Software* **36** (2010), 1–13.

[29] J. Yan, Y. Xu, Q. Cheng, S. Jiang, Q. Wang, Y. Xiao, C. Ma, J. Yan and X. Wang, LightGBM: accelerated genomically designed crop breeding through ensemble learning, *Genome Biology* **22** (2021), 271.

[30] S.M. Wang, R.S. Zheng, S.W. Zhang, H.M. Zeng, R. Chen, K.X. Sun, X.Y. Gu, W.W. Wei and J. He, Epidemiological characteristics of gastric cancer in China, 2015, *Zhonghua Liu Xing Bing Xue Za Zhi* **40** (2019), 1517–1521.

[31] A. Negida, N.K. Fahim and Y. Negida, Sample Size Calculation Guide – Part 4: How to Calculate the Sample Size for a Diagnostic Test Accuracy Study based on Sensitivity, Specificity, and the Area Under the ROC Curve, *Adv J Emerg Med* **3** (2019), e33.

[32] E. Yakirevich and M.B. Resnick, Pathology of gastric cancer and its precursor lesions, *Gastroenterology Clinics* **42** (2013), 261–284.

[33] M. Banks, D. Graham, M. Jansen, T. Gotoda, S. Coda, M.d. Pietro, N. Uedo, P. Bhandari, D.M. Pritchard, E.J. Kuipers, M. Rodriguez-Justo, M.R. Novelli, K. Ragunath, N. Shepherd and M. Dinis-Ribeiro, British Society of Gastroenterology guidelines on the diagnosis and management of patients at risk of gastric adenocarcinoma, *Gut* **68** (2019), 1545–1575.

[34] Y. Li, R. Xia, B. Zhang and C. Li, Chronic Atrophic Gastritis: A Review, **37** (2018), 241–259.

[35] R. Magris, V. De Re, S. Maiero, M. Fornasarig, G. Guarnieri, L. Caggiari, C. Mazzon, G. Zanette, A. Steffan, V. Canzonieri and R. Cannizzaro, Low Pepsinogen I/II Ratio and High Gastrin-17 Levels Typify Chronic Atrophic Autoimmune Gastritis Patients With Gastric Neuroendocrine Tumors, *Clin Transl Gastroenterol* **11** (2020), e00238.

[36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss and V. Dubourg, Scikit-learn: Machine learning in Python, *the Journal of machine Learning research* **12** (2011), 2825–2830.

[37] J. Li, K. Cheng, S. Wang, F. Morstatter, R.P. Trevino, J. Tang and H. Liu, Feature selection: A data perspective, *ACM computing surveys (CSUR)* **50** (2017), 1–45.

[38] A. Alsharkawi, M. Al-Fetyani, M. Dawas, H. Saadeh and M. Alyaman, Poverty classification using machine learning: The case of jordan, *Sustainability* **13** (2021), 1412.

[39] L. Li, Y. Lin, D. Yu, Z. Liu, Y. Gao and J. Qiao, A multi-organ fusion and lightgbm based radiomics algorithm for high-risk esophageal varices prediction in cirrhotic patients, *IEEE Access* **9** (2021), 15041–15052.

[40] X. Sun, M. Liu and Z. Sima, A novel cryptocurrency price trend forecasting model based on LightGBM, *Finance Research Letters* **32** (2020), 101084.

[41] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye and T.-Y. Liu, Lightgbm: A highly efficient gradient boosting decision tree, *Advances in Neural Information Processing Systems* **30** (2017).

[42] R. Tibshirani, Regression Shrinkage and Selection Via the Lasso, *Journal of the Royal Statistical Society: Series B (Methodological)* **58** (1996), 267–288.

[43] X. Yu, J. Chen, L. Zheng, J. Song, R. Lin and X. Hou, Quantitative Diagnosis of Atrophic Gastritis by Probe-Based Confocal Laser Endomicroscopy, *Biomed Res Int* **2020** (2020), 9847591.

[44] M. Dinis-Ribeiro, G. Yamaki, K. Miki, A. Costa-Pereira, M. Matsukawa and M. Kurihara, Meta-analysis on the validity of pepsinogen test for gastric carcinoma, dysplasia or chronic atrophic gastritis screening, *Journal of Medical Screening* **11** (2004), 141–147.

[45] Z. Wang, W. Han, F. Xue, Y. Zhao, P. Wu, Y. Chen, C. Yang, W. Gu and J. Jiang, Nationwide gastric cancer prevention in China, 2021–2035: a decision analysis on effect, affordability and cost-effectiveness optimisation, *Gut* **71** (2022), 2391–2400.

[46] J.G. Fox and T.C. Wang, Inflammation, atrophy, and gastric cancer, *The Journal of Clinical Investigation* **117** (2007), 60–69.

[47] M. Gholamalizadeh, S. Mirzaei Dahka, H. Sedigh Ebrahim-Saraie, M.E. Akbari, A. Pourtaheri, S. Rastgoo, A. Hajipour, A. Shafaghi, S. Doaei and N. Kalantari, The Role of Tumor Necrosis Factor-α (TNF-α) Polymorphisms in Gastric Cancer: a Meta-Analysis, *Journal of Gastrointestinal Cancer* **53** (2022), 756–769.

[48] X. Chen, R. Chen, R. Jin and Z. Huang, The role of CXCL chemokine family in the development and progression of gastric cancer, *International Journal of Clinical and Experimental Pathology* **13** (2020), 484–492.

[49] M. Zajkowska and B. Mroczko, From Allergy to Cancer – Clinical Usefulness of Eotaxins, *Cancers* **13** (2021), 128.

[50] Ü. Koç, E. Çetinkaya, E.B. Bostanci, A.S. Kemik, M. Tez, İ. Gömceli and M. Akoğlu, Diagnostic Significance of Serum Eotaxin-1 Level in Gastric Cancer Patients, *Disease Markers* **35** (2013), 274515.

[51] N. Sánchez-Zauco, J. Torres, A. Gómez, M. Camorlinga-Ponce, L. Muñoz-Pérez, R. Herrera-Goepfert, R. Medrano-Guzmán, S. Giono-Cerezo and C. Maldonado-Bernal, Circulating blood levels of IL-6, IFN-γ, and IL-10 as potential diagnostic biomarkers in gastric cancer: a controlled study, *BMC Cancer* **17** (2017), 384.

[52] E. Caiazzo, M. Sharma, A.O.M. Rezig, M.I. Morsy, M. Czesnikiewicz-Guzik, A. Ialenti, J. Sulicka-Grodzicka, P. Pellicori, S.H. Crouch, A.E. Schutte, D. Bruzzese, P. Maffia and T.J. Guzik, Circulating cytokines and risk of developing hypertension: A systematic review and meta-analysis, *Pharmacol Res* **200** (2024), 107050.

[53] V. Procházka, L. Lacina, K. Smetana, M. Svoboda, K. Skřivanová, M. Beňovská, J. Jarkovský, L. Křen and Z. Kala, Serum concentrations of proinflammatory biomarker interleukin-6 (IL-6) as a predictor of postoperative complications after elective colorectal surgery, *World Journal of Surgical Oncology* **21** (2023), 384.

[54] L. Chu, K. Vrbicky, D. Montefiori, W. Huang, B. Nestorova, Y. Chang, A. Carfi, D.K. Edwards, J. Oestreicher and H. Legault, Immune response to SARS-CoV-2 after a booster of mRNA-1273: an open-label phase 2 trial, *Nature Medicine* **28** (2022), 1042–1049.

[55] R.S. Eid, J.A. Chaiton, S.E. Lieblich, T.S. Bodnar, J. Weinberg and L.A. Galea, Early and late effects of maternal experience on hippocampal neurogenesis, microglia, and the circulating cytokine milieu, *Neurobiology of Aging* **78** (2019), 1–17.

[56] G. Sigal, M. Segal, A. Mathew, L. Jarlsberg, M. Wang, S. Barbero, N. Small, K. Haynesworth, J. Davis and M. Weiner, Biomarkers of tuberculosis severity and treatment effect: a directed screen of 70 host markers in a randomized clinical trial, *EBioMedicine* **25** (2017), 112–121.

[57] H. Yang, X. Zhou and B. Hu, The 'reversibility' of chronic atrophic gastritis after the eradication of Helicobacter pylori, *Postgraduate Medicine* **134** (2022), 474–479.

[58] E.-J. Cho, H.-K. Kim, T.-D. Jeong, D.-H. Ko, S.E. Bae, J.-S. Lee, W. Lee, J.W. Choe, S. Chun, H.-Y. Jung and W.-K. Min, Method evaluation of pepsinogen I/II assay based on chemiluminescent immunoassays and comparison with other test methods, *Clinica Chimica Acta; International Journal of Clinical Chemistry* **452** (2016), 149–154.

[59] M.K. Gould, B.Z. Huang, M.C. Tammemagi, Y. Kinar and R. Shiff, Machine Learning for Early Lung Cancer Identification Using Routine Clinical and Laboratory Data, *American Journal of Respiratory and Critical Care Medicine* **204** (2021), 445–453.

[60] B. Ma, J. Pan, X. Hou, C. Li, T. Xiong, Y. Gong and F. Song, The Construction of Polygenic Risk Scores for Breast Cancer Based on LightGBM and Multiple Omics Data, (2021).

[61] I. Tsai, W.-C. Shen, C.-L. Lee, H.-D. Wang and C.-Y. Lin, Machine learning in prediction of bladder cancer on clinical laboratory data, *Diagnostics* **12** (2022), 203.

[62] D. Wang, Y. Zhang and Y. Zhao, LightGBM: an effective miRNA classification method in breast cancer patients, in: *Proceedings of the 2017 international conference on computational biology and bioinformatics*, 2017, pp. 7–11.

[63] M.M. Ahamad, S. Aktar, M.J. Uddin, T. Rahman, S.A. Alyami, S. Al-Ashhab, H.F. Akhdar, A. Azad and M.A. Moni, Early-Stage Detection of Ovarian Cancer Based on Clinical Data Using Machine Learning Approaches, *Journal of Personalized Medicine* **12** (2022).

[64] B. Qiu, X.H. Su, X. Qin and Q. Wang, Application of machine learning techniques in real-world research to predict the risk of liver metastasis in rectal cancer, *Frontiers In Oncology* **12** (2022), 1065468.