

Construction of tongue image-based machine learning model for screening patients with gastric precancerous lesions

Author List: Changzheng Ma, Peng Zhang, Shao Li

Author list

Changzheng Ma¹, Peng Zhang¹, Shao Li¹

1. Institute of TCM-X/MOE Key Laboratory of Bioinformatics, Bioinformatics Division, BNRist/Department of Automation, Tsinghua University, 100084 Beijing, China

Corresponding authors

Shao Li

shaoli@tsinghua.edu.cn

86-13693316285

Tsinghua University, 100084 Beijing, China

Abstract

Screening patients with precancerous lesions of gastric cancer (PLGC) is important for gastric cancer prevention. It could improve the accuracy and convenience of PLGC screening to uncover and integrate valuable characteristics of noninvasive medical images involving in PLGC, by applying machine learning methodologies. In this study, based on unbiasedly uncovering potential associations between tongue image characteristics and PLGC and integrating gastric cancer-related canonical risk factors, including age, sex, Hp infection, we focused on tongue images and constructed a tongue image-based PLGC screening deep learning model (AITongue). Then, validation analysis on an independent cohort of 1,995 patients revealed the AITongue model could screen PLGC individuals with an AUC of 0.75, 10.3% higher than that of the model constructed with gastric cancer-related canonical risk factors. Of note, we investigated the value of the AITongue model in predicting PLGC risk by establishing a prospective PLGC follow-up cohort, reaching an AUC of 0.71. In addition, we have developed a smartphone-based App screening system to enhance the application convenience of the AITongue model in the natural population. Collectively, our study has demonstrated the value of tongue image characteristics in PLGC screening and risk prediction.

Construction of tongue image-based machine learning model for screening patients with gastric precancerous lesions

Abstract

Screening patients with precancerous lesions of gastric cancer (PLGC) is important for gastric cancer prevention. It could improve the accuracy and convenience of PLGC screening to uncover and integrate valuable characteristics of noninvasive medical images involving in PLGC, by applying machine learning methodologies. In this study, based on unbiasedly uncovering potential associations between tongue image characteristics and PLGC and integrating gastric cancer-related canonical risk factors, including age, sex, Hp infection, we focused on tongue images and constructed a tongue image-based PLGC screening deep learning model (AITongue). Then, validation analysis on an independent cohort of 1,995 patients revealed the AITongue model could screen PLGC individuals with an AUC of 0.75, 10.3% higher than that of the model constructed with gastric cancer-related canonical risk factors. Of note, we investigated the value of the AITongue model in predicting PLGC risk by establishing a prospective PLGC follow-up cohort, reaching an AUC of 0.71. In addition, we have developed a smartphone-based App screening system to enhance the application convenience of the AITongue model in the natural population. Collectively, our study has demonstrated the value of tongue image characteristics in PLGC screening and risk prediction.

Trial Registration:

ChiCTR2100044006

Keywords:

precancerous lesions of gastric cancer, tongue image, deep learning, disease screening, risk prediction, artificial intelligence

Introduction

Gastric cancer is the second leading cause of cancer death in men and women in China, and more than 80% of patients are diagnosed at an advanced stage [1]. Patients with precancerous lesions of gastric cancer (PLGC), including intestinal metaplasia and dysplasia [2, 3], suffered a higher risk of gastric tumorigenesis, with an annual incidence of 0.25%-6% [4-6]. Screening and conducting reasonable health surveillance for patients with PLGC in the natural population would make a great contribution to facilitating the early prevention of gastric cancer.

Current screening methods suffered from some challenges, including invasive and relatively low accuracy, which limited their applications in population screening. On the one hand, although gastroscopy and biopsy are the gold standards for gastric disease diagnosis [7], it remains inefficient and

unreasonable use with respect to gastric disease screening [8], resulting in approximately half of the patients screened with gastroscopy being non-atrophic gastritis and the early diagnosis rate of gastric cancer is less than 20% [1, 9]. On another hand, the application of serum markers that are commonly used as screening factors in various gastric cancer risk assessment methods, such as pepsinogen I/II and gastrin-17 [10-12], were limited by high sensitivity and specificity threshold for serological tests for risk screening in natural populations [13]. In addition, it remains cost-effective for both serum pepsinogen test screening and endoscopy, presenting difficulty for their practical application [14]. Thus, non-invasive and highly accurate screening models for PLGC patients are urgently needed.

As non-invasive indicators, tongue image characteristics were used for the surveillance of a broad spectrum of diseases, inspired by the diagnosis experience in traditional Chinese medicine [15-21]. Tongue image characteristics, including shape, color, and tongue coating, are believed to reflect the state of health and the progress and severity of disease, especially for digestive diseases as the tongue is physically connected to the digestive system. For example, massive studies indicated that tongue image characteristics have been shown to correlate with gastroscopy findings and can be used to predict gastric mucosal health [22, 23]. In addition, it was revealed that tongue surface and color characteristics as indicators can improve the accuracy of gastric cancer diagnosis [24]. From the pathologic perspective, the distribution of microorganisms on tongue coating has also been suggested to be related to the distribution in gastric tissue, which can be used as a marker for gastric disease risk screening [25-27]. Moreover, morphological markers based on tongue images are considered to be valuable for risk screening for other diseases, such as diabetes, fatty liver disease, and CoVID-19 [16, 28-31]. The above studies indicated that tongue image characteristics can assist in disease screening. Therefore, uncovering the risk characteristics of tongue images is potentially valuable for constructing PLGC screening models.

Recently, deep learning techniques are widely used in building biomedical image-based disease screening and prediction models [32-37]. For example, some studies have applied deep learning to predict cancer, such as prostate cancer and rectal cancer, based on medical images [34, 38]. For tongue images, some studies have used deep learning techniques to identify risk features in tongue images for the detection of diseases such as stomach cancer and diabetes [24, 39]. So deep learning techniques could be utilized to uncover tongue risk characteristics.

Therefore, to improve the efficiency of screening patients with PLGC, particularly in natural populations, in this study, we explored the tongue image characteristics of patients with PLGC and integrated them with traditional screening indicators to develop a PLGC screening model-AITongue. Then, we evaluated its screening effect by external validation in an independent dataset and explored its potential value as a risk predictor of PLGC in a follow-up

dataset. We believed that our study would pave the way to address the urgent need for non-invasive PLGC screening in clinical practice.

Methods

Patient enrollment, and data collection

Gastritis patients were enrolled in this study at cChina-Japan Friendship Hospital and Yijishan Hospital of Wannan Medical College from 2015 to 2022. The experimental protocol was established according to the ethical guidelines of the "Declaration of Helsinki" and was approved by the Human Ethics Committee of Institution Review Board of Tsinghua University and the Chinese clinical trial registry. Inclusion criteria: Above 18 years of age with clear consciousness, clear language skills, and no barriers in communication and were willing to accept clinical investigation and sign informed consent. Exclusion criteria: Combined with heart, cerebrovascular, liver, kidney, hematopoietic system, and other primary diseases.

Gastroscopy and histological examination

Using video endoscopes (Olympus Corp), upper gastroscopic examinations were performed by 2 gastroenterologists. Tissue samples for biopsy were reviewed blindly by 2 pathologists according to the criteria proposed by the Updated Sydney System and the Chinese Association of Gastric Cancer [40, 41]. The results of each biopsy were reported as normal, superficial gastritis, chronic atrophic gastritis, intestinal metaplasia, intraepithelial neoplasia, or gastric cancer, and each participant was assigned a global diagnosis based on the most severe gastric histologic finding among any biopsy. *Helicobacter pylori* (Hp) infection status was determined by enzyme-linked immunosorbent assay for plasma IgG [42].

Data pre-processing and structuring

Python (3.7.0) and PyTorch was used for tongue image preprocessing. We trained a tongue detection model using the YOLOv5 deep learning model [43]. With the model tongue images were detected and cut into tongue body images for subsequent analysis.

Basic information and clinical symptom characteristics were obtained from electronic medical records. The content included baseline information (gender, age) and symptom characteristics (xerostomia, bitter taste, gastric distention, stomach pain, etc.). All the above indicators were structured as two-category labeled data. Among them, age was divided into >50 and ≤50 years based on the differentiation effect. Multiple interpolation methods were used to fill in the missing data. Tongue labels (fissure, etc.) were labeled by physicians.

Screening model construction

The main body of the image classification part of AITongue model is the ResNet50 model [44]. ResNet50 is a commonly used model in deep learning and

is heavily used in research on images. Since this model introduces the residual block, it has a wide range of applications and good performance in the field of image classification. The residual blocks in the ResNet50 model are structured as two bottlenecks (BTNK), which are BTNK 1, BTNK 2 and their structure diagram is in Appendix 1. CONV is the convolution block, BN is the batch normalization block, and Relu is the activation function in the bottleneck.

The tongue image was reshaped as 224*224*3 and input to the model. Then, after the ResNet50 module, an MLP and a sigmoid module were used to classify the tongue images into two categories: high-risk and low-risk. Next, the image category labels were fed into a logistic regression model along with labels such as basic information. Finally, the model output the predicted scores for PLGC.

Statistical Analysis

All analysis procedures were performed using Python (3.7.0) and the sklearn package. Tongue diagnostic labels (TDL) and clinical symptoms with statistical significance ($p < 0.05$) by univariate and multivariate analyses were included in the model. Binary logistic regression was used to construct screening models. Chi-square tests were applied to calculate the significance of independent variables for PLGC. Pearson's correlation coefficient was applied to evaluate the correlation between independent variables. Accuracy, sensitivity, specificity, recall, precision, ROC curve, and AUC were used as evaluation metrics to evaluate model performance.

Results

The overall design of our study

All patients underwent gastroscopy and pathology and were divided into three cohorts: development cohort, validation cohort, and follow-up cohort for model development, external validation, and risk prediction, respectively, where two categories, including PLGC and non-PLGC, were classified for patients based on pathological diagnosis (Table 1). In detail, we developed AITongue model for PLGC screening and did internal validation on the development cohort, which had a total of 325 patients, including 55 PLGC patients and 270 non-PLGC ones. Then, we performed external validation on the validation cohort, which had a total of 1995 patients, including 171 PLGC patients and 1824 non-PLGC ones. Of note, we evaluated AITongue model for risk prediction of PLGC on the follow-up cohort, in which only non-PLGC patients were enrolled in the baseline, and patients with PLGC and non-PLGC in the endpoint were classified as Pro and non-Pro, respectively, after a mean follow-up time of 22 months (Figure 1).

Table 1. Basic information of 3 Cohorts.

Cohort	Class	Num	Age (Mean \pm SD)	Gender (Male/Female)
development	PLGC	55	58.6 \pm 10.48	31/24

cohort	non-PLGC	270	48.0±13.61	135/135
validation cohort	PLGC	171	58.5±9.3	100/71
	non-PLGC	1824	49.9±13.1	685/1139
follow-up cohort	Pro	26	51.7±8.7	14/12
	non-Pro	69	46.1±11.9	24/45

SD: Squared Difference

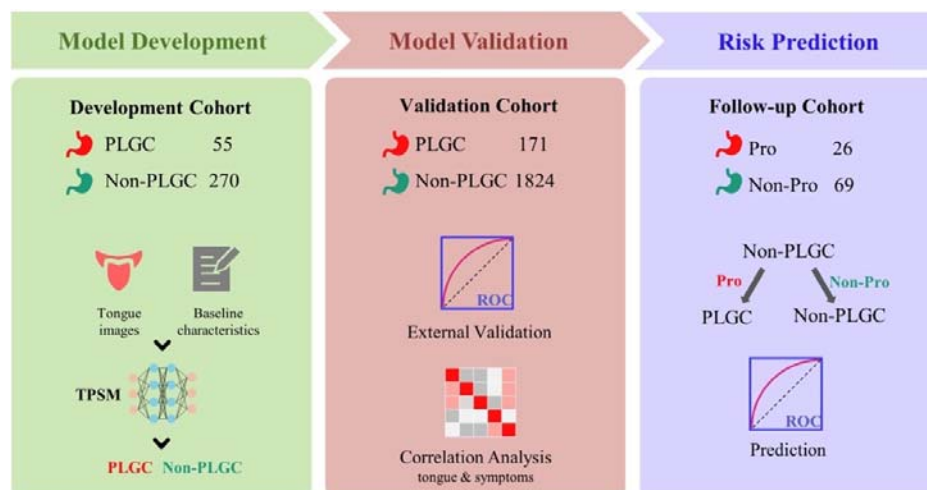


Figure 1. The study outline of work flow.

Construction of AITongue model integrating tongue image characteristics

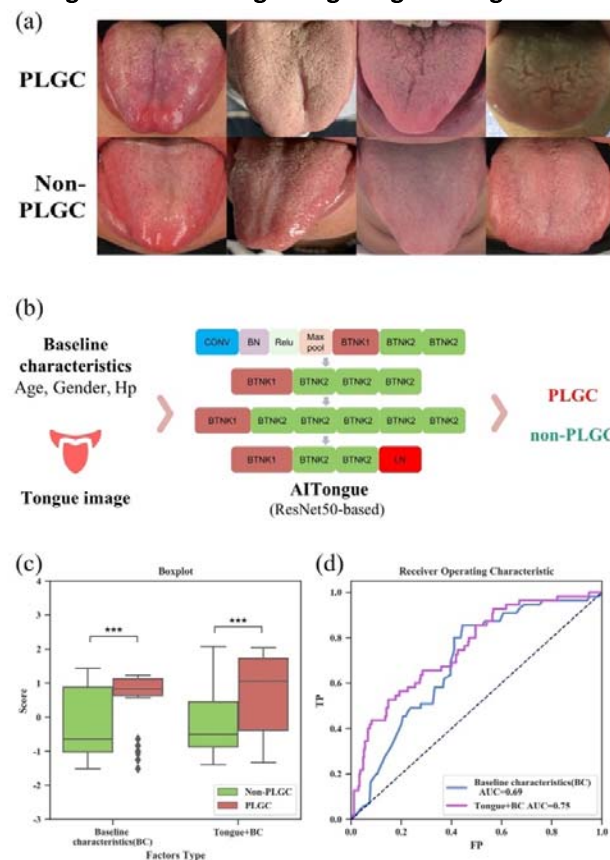


Figure 2. **Construction of AITongue model and results for PLGC screening based on tongue images.** A. Example of tongue images of PLGC and non-PLGC. B. ResNet50-based deep learning screening model- AITongue. C. Boxplot of classification score comparisons with and without the inclusion of tongue images for PLGC screening. D. ROC curves and AUC comparisons for PLGC screening (***: $p < 0.001$)

AITongue model was constructed based on deep learning and was shown in Figure 2b. The model took the preprocessed tongue images as input and output the risk score of the images for PLGC screening. The preprocessed tongue images were shown in Figure 2a. And then, the risk scores of the images were input to a logistic regression model together with baseline labels, and finally, the PLGC classification results were output.

The inclusion of tongue image characteristics improved the PLGC screening effect by 8.7%. The development cohort was used as training data, and the screening model was validated internally by a five-fold cross-validation method. The boxplot showed that the classification scoring results with the introduction of tongue image were more discriminative for PLGC and non-PLGC (Figure 2c). The screening model with baseline (age, sex, Hp) as input had an accuracy of 0.60, a sensitivity of 0.80, a specificity of 0.56, and an AUC of 0.69. With the introduction of tongue images, the accuracy was 0.69, sensitivity was 0.71, specificity was 0.69, and AUC was 0.75. In terms of AUC, the screening effect was

improved by 8.7% (Figure 2d). Therefore, it indicated that the introduction of tongue image characteristics contributed to the efficiency of PLGC screening. The medical significance of risk features in the tongue images was interpreted and a risk typing analysis of the tongue images was performed. Traditional Chinese medicine (TCM) practitioners labeled 325 tongue images based on the tongue diagnosis method. The above screening model classified 94 and 233 tongue images as high-risk and low-risk, respectively. We found that 5 of the tongue diagnosis labels were statistically significant ($p < 0.05$), namely greasy, fissure, dark, coating (yellow), and coating (thick), which explained, to some extent, the medical significance of the risk features in tongue images and suggested that there may be some value of TDL for PLGC screening as well (Table 2).

Table 2. TDL analysis of high and low-risk tongue images classified by AITongue model.

Characteristics	High Risk (94)		Low Risk (233)		P value
	Num	Rate	Num	Rate	
Coating (Yellow)	16	0.17	16	0.07	9.6e-3
Greasy	28	0.30	34	0.15	2.6e-3
Fissure	19	0.20	26	0.11	4.8e-2
Coating (Thick)	14	0.15	15	0.06	2.6e-2
Dark	12	0.13	7	0.03	1.6e-3

External validation for PLGC screening

To validate the AITongue model and analyze the interpretability of tongue image characteristics in the AITongue model, the tongue images in the validation cohort were transformed into interpretable TDL by TCM tongue diagnosis, to directly quantify the role of different features of tongue images in PLGC screening [45, 46].

TDL that were statistically significant for PLGC were screened as indicators before validation. Gender, age, and 5 TDL showed significance in univariate and multivariate analyses of the validation cohort (Table 3). TDL such as teeth marks that did not show significance were in Appendix 2. The 5 significant characteristics were greasy, fissure, dark, coating (yellow), and coating (thick), which were included in the screening model to validate the effect of introducing tongue image characteristics to enhance PLGC screening.

Table 3. Univariate and multivariate analysis of baseline factors and TDL in PLGC screening.

VARIABLE	TOTAL (N=1995)	NON-PLGC (N=1824)	PLGC (N=171)	P VALUE*	ADJUSTED OR (95% CI) †	P VALUE
	NO. (%)	NO. (%)	NO. (%)			

Baseline Factors						
Age, years		3.7e-15				
>50	1067(0.53)	926(0.51)	141(0.82)	<0.001	4.14 ([2.59,6.61])	2.9e-9
<=50	928(0.47)	898(0.49)	30(0.18)			
Gender		1.3e-7				
Male	785(0.39)	685(0.38)	100(0.58)		1.47([1.00,2.14])	4.8e-2
Female	1210(0.61)	1139(0.62)	71(0.42)			
Hp		0.71				
Yes	174(0.24)	142(0.24)	32(0.22)		1.15([0.73,1.81])	0.55
No	556(0.76)	445(0.76)	111(0.78)			
Tongue Diagnostic Labels						
Coating		1.9e-10				
Yellow	327(0.16)	269(0.15)	58(0.34)		3.66([2.38,5.62])	3.4e-9
White	1668(0.84)	1555(0.85)	113(0.66)			
Fissure		3.8e-9				
Yes	187(0.09)	149(0.08)	38(0.22)		2.26([1.37, 3.73])	1.4e-3
No	1808(0.91)	1675(0.92)	133(0.78)			
Greasy		3.3e-4				
Yes	422(0.21)	367(0.20)	55(0.32)		2.23([1.44,3.46])	3.3e-4
No	1573(0.79)	1457(0.80)	116(0.68)			
Coating		1.3e-3				
Thick	239(0.12)	205(0.11)	34(0.20)		2.18([1.29,3.67])	3.6e-3
Thin	1756(0.88)	1619(0.89)	137(0.80)			
Dark		0.03				
Yes	312(0.16)	275(0.15)	37(0.22)		1.93([1.20,3.11])	6.7e-3
No	1683(0.84)	1549(0.85)	134(0.78)			

*P values refer to the comparison between PLGC and non-PLGC groups in the univariate analysis.

†For variables not significant (p>0.05) in the logistic regression model, multivariable data are not shown.

The effectiveness of PLGC screening was improved by 10.3% in the validation cohort. We used a validation cohort to validate the effectiveness of tongue image characteristics for PLGC screening by logistic regression. The boxplot showed that the classification scoring of introducing tongue image characteristics was more discriminative for PLGC and non-PLGC (Figure 4). The accuracy of the baseline screening model with baseline (age, sex, Hp) as input was 0.53, sensitivity was 0.76, specificity was 0.51, and AUC was 0.68. With the introduction of TDL, the accuracy was 0.64, sensitivity was 0.72, specificity was 0.63, and AUC was 0.75. In terms of AUC, the screening effect was significantly improved by 10.3% (p<0.01) (Figure 3). The results illustrated that the introduction of TDL helped to improve the screening efficiency of PLGC and validated the effectiveness of tongue image characteristics for PLGC screening.

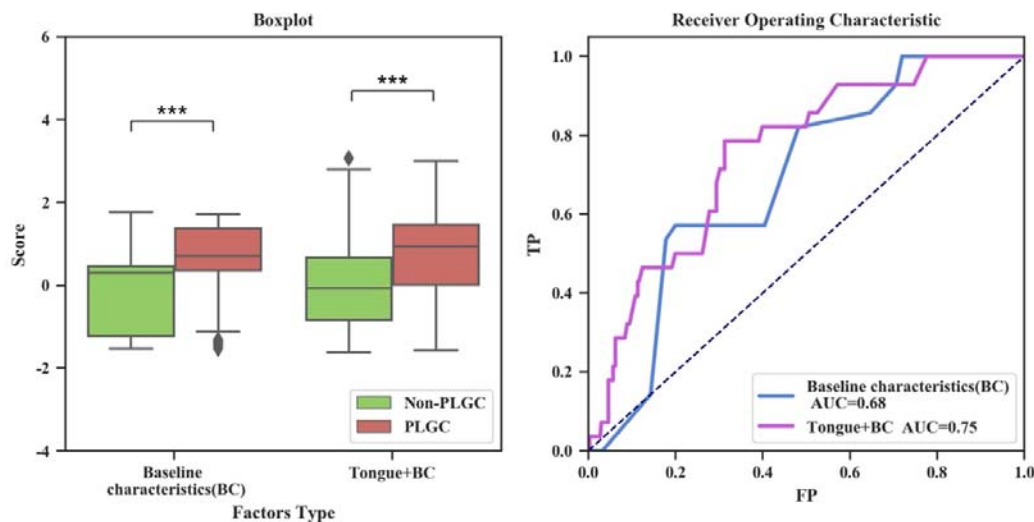


Figure 3. Boxplot of screening score and ROC curves comparisons with and without the inclusion of tongue image characteristics for PLGC screening. (***: $p < 0.001$)

In addition, the clinical symptom characteristics of patients were counted in the validation cohort to assess their correlation with tongue image characteristics and their effect on screening.

First, we investigated the correlation between clinical symptoms and PLGC. The 3 symptom factors (xerostomia, bitter taste, belching) showed significance in univariate and multivariate analysis and they were included in the screening model to validate the effect of introducing clinical symptoms characteristics to enhance PLGC screening (Appendix 2). Factors that did not show significance such as stomach pain, bloating, chilliness, and loose stools were in Appendix 3.

Further, we incorporated symptom characteristics into the screening model to evaluate the enhancement effect of introducing symptom characteristics for PLGC screening. The validation cohort was used as training data to construct a logistic regression model and performed five-fold cross-validation. The boxplot showed that the classification scoring result after introducing symptom characteristics has a small improvement in the discrimination between PLGC and non-PLGC (Figure 4). The AUC of the baseline screening model with baseline (age, sex, Hp) as input was 0.68. The AUC with tongue label and baseline as input was 0.75. The AUC with clinical symptoms and baseline as input was 0.73. The AUC with TDL, clinical symptoms, and baseline as input was 0.76 (Figure 5). The results indicated that the introduction of clinical symptom characteristics improved the screening efficiency of PLGC, but its effect was slightly lower than that of tongue image characteristics. In addition, there was a low consistency between tongue image characteristics and the appearance of gastric symptoms (Appendix 4). Therefore, we found that the introduction of symptom characteristics on top of tongue image characteristics had a small improvement in the effect of PLGC screening.

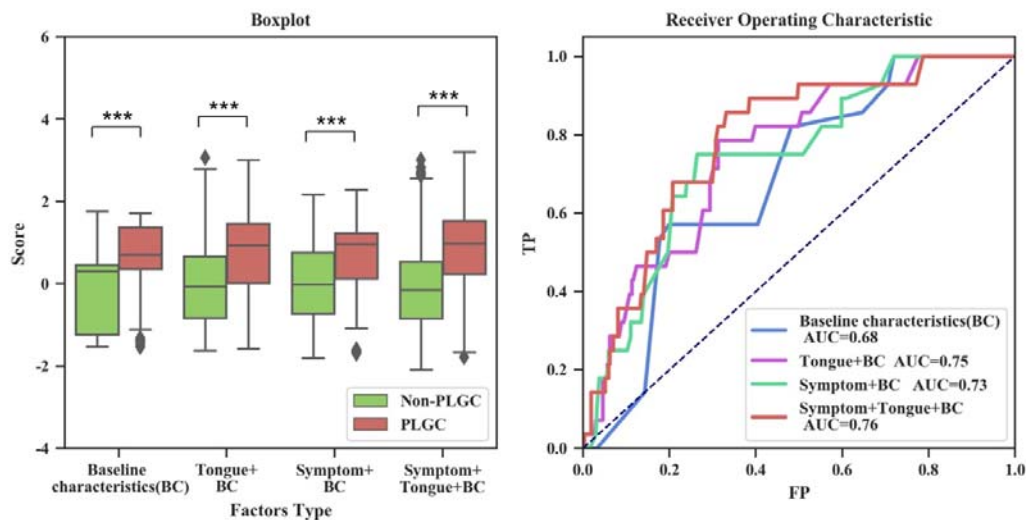


Figure 4. Boxplot of screening score and ROC curves comparisons with and without the inclusion of symptoms factors for PLGC screening. (***: $p < 0.001$)

Evaluation of the validity of tongue image characteristics for risk prediction of PLGC

We further resolved the ability of the model for risk predicting of PLGC by establishing the follow-up cohort, in which the initial pathological test was non-PLGC. Patients were divided into progressive (Pro) and non-progressive (non-Pro) groups with a second test as PLGC and non-PLGC (Table 1). Based on the previous logistic regression PLGC screening model based on the validation cohort, all patients in the Follow-up Cohort were scored for PLGC risk. The introduction of tongue image characteristics compared to baseline metrics showed a significant increase ($p < 0.01$) in the differentiation of risk scores between the two groups, with an AUC increase of 10.9% (0.64 to 0.71) (Figure 5). In addition, we performed a univariate analysis of TDL for risk prediction of PLGC, and the results were in Appendix 5. Therefore, tongue image characteristics were potentially valuable for enhancing the predictive ability of PLGC risk, and further research was needed.

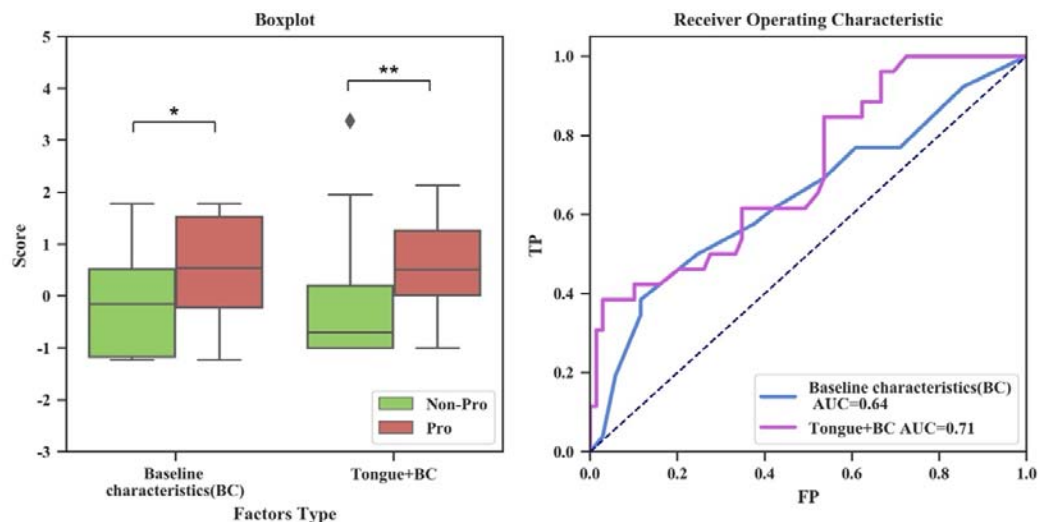


Figure 5. Boxplot of screening score and ROC curves comparisons with and without the inclusion of tongue image characteristics for risk prediction of PLGC. (*: $p < 0.01$, **: $p < 0.05$)

Discussion

Screening patients with PLGC is important for the prevention and treatment of gastric cancer. In this study, we analyzed the tongue image characteristics associated with PLGC and based on this, constructed a PLGC screening model on a development cohort, then externally validated it in an independent validation cohort and evaluated the ability of risk prediction of PLGC in a follow-up cohort. Our study demonstrated the value of tongue image characteristics in PLGC screening and its potential for risk prediction.

We found that *H. pylori* infection was weakly correlated with PLGC and non-PLGC, although Hp infection is the most prominent risk factor for GC. Similar results have been found in other studies on the prediction of gastric cancer risk[47]. In this study, PLGC was analyzed with symptoms. We found only a small proportion of symptoms correlated with PLGC and their screening efficiency was not high, which is consistent with the findings of other studies [48].

Our proposed method has better performance than another study in screening of PLGC. Wang et al. developed a model with non-invasive indicators for PLGC screening including 290 patients with gastritis and the AUC was 0.728 (95% CI [0.651-0.793]) while our method was 0.76[49].

The study has limits. The data source was biased compared to the natural population. Due to the need for accurate information on the stage of gastritis, the data for establishing the system all came from patients with gastric disease, which had a certain deviation compared with the natural population. We have developed a smartphone-based App screening system to enhance the application convenience of the AITongue model in the natural population (Appendix 6). In further studies, more samples would be collected from natural populations to reduce bias and larger external validation should be done.

The screening model constructed in this study could improve the accuracy of PLGC screening. Tongue image characteristics were validated for their value in PLGC screening and risk prediction, which may drive tongue image characteristics as a new risk indicator in the future. By extracting tongue image characteristics through deep learning techniques, this study proposes a new approach for non-invasive PLGC screening and shows the possibility of its being used in large-scale applications.

Declarations

Ethics approval and consent to participate

The experimental protocol was established, according to the ethical guidelines of the Helsinki Declaration and was approved by the Human Ethics Committee of Institution Review Board of Tsinghua University.

Consent for publication

Not applicable.

Competing Interests

The authors declare that they have no competing interests.

Funding

Funding for this study was provided by the National Natural Science Foundation of China, China [81225025 and 62061160369]; and the Beijing National Research Center for Information Science and Technology, China [BNR2019TD01020 and BNR2019RC01012].

Authors' Contributions

Changzheng Ma contributed to the study design, data collection, data analyses, and writing. Peng Zhang, Xinxing Lai, Aidi Tan, and Xin Wang contributed to the study design and writing. Chaofan Ji, Qingrui Zhang, Shiyu Du, and Yan Li contributed to the data collection. Shao Li is the corresponding author.

Acknowledgments

Thanks to the doctors and nurses of China-Japan Friendship Hospital and Yijishan Hospital of Wannan Medical College for their support of data collection.

References

1. Zong L, Abe M, Seto Y, Ji J: **The challenge of screening for early gastric cancer in China.** *Lancet* 2016, **388**(10060):2606.
2. Schlemper RJ, Riddell RH, Kato Y, Borchard F, Cooper HS, Dawsey SM, Dixon MF, Fenoglio-Preiser CM, Fléjou JF, Geboes K *et al*: **The Vienna classification of gastrointestinal epithelial neoplasia.** *Gut* 2000, **47**(2):251-255.

3. Song H, Ekheden IG, Zheng Z, Ericsson J, Nyren O, Ye W: **Incidence of gastric cancer among patients with gastric precancerous lesions: observational cohort study in a low risk Western population.** *BMJ* 2015, **351**:h3867.
4. de Vries AC, van Grieken NC, Looman CW, Casparie MK, de Vries E, Meijer GA, Kuipers EJ: **Gastric cancer risk in patients with premalignant gastric lesions: a nationwide cohort study in the Netherlands.** *Gastroenterology* 2008, **134**(4):945-952.
5. Piazuolo MB, Bravo LE, Mera RM, Camargo MC, Bravo JC, Delgado AG, Washington MK, Rosero A, Garcia LS, Realpe JL *et al*: **The Colombian Chemoprevention Trial: 20-Year Follow-Up of a Cohort of Patients With Gastric Precancerous Lesions.** *Gastroenterology* 2021, **160**(4):1106-1117 e1103.
6. Rugge M, Meggio A, Pravadelli C, Barbareschi M, Fassan M, Gentilini M, Zorzi M, Pretis G, Graham DY, Genta RM: **Gastritis staging in the endoscopic follow-up for the secondary prevention of gastric cancer: a 5-year prospective study of 1755 patients.** *Gut* 2019, **68**(1):11-17.
7. Yan H, Li M, Cao L, Chen H, Lai H, Guan Q, Chen H, Zhou W, Zheng B, Guo Z *et al*: **A robust qualitative transcriptional signature for the correct pathological diagnosis of gastric cancer.** *J Transl Med* 2019, **17**(1):63.
8. Endoscopy CSoD: **Consensus on screening and endoscopic diagnosis and treatment of early gastric cancer in China (Changsha, 2014).** *Zhonghua Xiao Hua Nei Jing Za Zhi* 2014, **31**:361-377.
9. Du Y, Bai Y, Xie P, Fang J, Wang X, Hou X, Tian D, Wang C, Liu Y, Sha W *et al*: **Chronic gastritis in China: a national multi-center survey.** *BMC Gastroenterol* 2014, **14**:21.
10. Tu H, Sun L, Dong X, Gong Y, Xu Q, Jing J, Bostick RM, Wu X, Yuan Y: **A Serological Biopsy Using Five Stomach-Specific Circulating Biomarkers for Gastric Cancer Risk Assessment: A Multi-Phase Study.** *Am J Gastroenterol* 2017, **112**(5):704-715.
11. Huang S, Guo Y, Li ZW, Shui G, Tian H, Li BW, Kadeerhan G, Li ZX, Li X, Zhang Y *et al*: **Identification and Validation of Plasma Metabolomic Signatures in Precancerous Gastric Lesions That Progress to Cancer.** *JAMA Netw Open* 2021, **4**(6):e2114186.
12. Huang KK, Ramnarayanan K, Zhu F, Srivastava S, Xu C, Tan ALK, Lee M, Tay S, Das K, Xing M *et al*: **Genomic and Epigenomic Profiling of High-Risk Intestinal Metaplasia Reveals Molecular Determinants of Progression to Gastric Cancer.** *Cancer Cell* 2018, **33**(1):137-150 e135.
13. Cubiella J, Perez Aisa A, Cuatrecasas M, Diez Redondo P, Fernandez Esparrach G, Marin-Gabriel JC, Moreira L, Nunez H, Pardo Lopez ML, Rodriguez de Santiago E *et al*: **Gastric cancer screening in low incidence populations: Position statement of AEG, SEED and SEAP.** *Gastroenterol Hepatol* 2021, **44**(1):67-86.

14. Pimentel-Nunes P, Libanio D, Marcos-Pinto R, Areia M, Leja M, Esposito G, Garrido M, Kikuste I, Megraud F, Matysiak-Budnik T *et al*: **Management of epithelial precancerous conditions and lesions in the stomach (MAPS II): European Society of Gastrointestinal Endoscopy (ESGE), European Helicobacter and Microbiota Study Group (EHMSG), European Society of Pathology (ESP), and Sociedade Portuguesa de Endoscopia Digestiva (SPED) guideline update 2019.** *Endoscopy* 2019, **51**(4):365-388.
15. Jiang T, Guo XJ, Tu LP, Lu Z, Cui J, Ma XX, Hu XJ, Yao XH, Cui LT, Li YZ *et al*: **Application of computer tongue image analysis technology in the diagnosis of NAFLD.** *Comput Biol Med* 2021, **135**:104622.
16. Li J, Huang J, Jiang T, Tu L, Cui L, Cui J, Ma X, Yao X, Shi Y, Wang S *et al*: **A multi-step approach for tongue image classification in patients with diabetes.** *Comput Biol Med* 2022, **149**:105935.
17. Zhuang Q, Gan S, Zhang L: **Human-computer interaction based health diagnostics using ResNet34 for tongue image classification.** *Comput Methods Programs Biomed* 2022, **226**:107096.
18. Hu Y, Wen G, Luo M, Yang P, Dai D, Yu Z, Wang C, Hall W: **Fully-channel regional attention network for disease-location recognition with tongue images.** *Artif Intell Med* 2021, **118**:102110.
19. Zhang B, Kumar BV, Zhang D: **Detecting diabetes mellitus and nonproliferative diabetic retinopathy using tongue color, texture, and geometry features.** *IEEE Trans Biomed Eng* 2014, **61**(2):491-501.
20. Kanawong R, Obafemi-Ajayi T, Ma T, Xu D, Li S, Duan Y: **Automated Tongue Feature Extraction for ZHENG Classification in Traditional Chinese Medicine.** *Evid Based Complement Alternat Med* 2012, **2012**:912852.
21. Wentao X, Kanawong R, Xu D, Shao L, Tao M, Guixu Z, Ye D: **An automatic tongue detection and segmentation framework for computer-aided tongue image analysis.** In: *2011 IEEE 13th International Conference on e-Health Networking, Applications and Services: 13-15 June 2011 2011*; 2011: 189-192.
22. Shang Z, Du ZG, Guan B, Ji XY, Chen LC, Wang YJ, Ma Y: **Correlation analysis between characteristics under gastroscopy and image information of tongue in patients with chronic gastriti.** *J Tradit Chin Med* 2022, **42**(1):102-107.
23. Kainuma M, Furusyo N, Urita Y, Nagata M, Ihara T, Oji T, Nakaguchi T, Namiki T, Hayashi J: **The association between objective tongue color and endoscopic findings: results from the Kyushu and Okinawa population study (KOPS).** *BMC Complement Altern Med* 2015, **15**:372.
24. Gholami EaKT, Seyed and Kheirabadi, Maryam: **Increasing the accuracy in the diagnosis of stomach cancer based on color and lint features of tongue.** *Biomedical Signal Processing and Control* 2021, **69**:102782.

25. Cui J, Hou S, Liu B, Yang M, Wei L, Du S, Li S: **Species composition and overall diversity are significantly correlated between the tongue coating and gastric fluid microbiomes in gastritis patients.** *BMC Med Genomics* 2022, **15**(1):60.
26. Cui J, Cui H, Yang M, Du S, Li J, Li Y, Liu L, Zhang X, Li S: **Tongue coating microbiome as a potential biomarker for gastritis including precancerous cascade.** *Protein Cell* 2019, **10**(7):496-509.
27. Xu J, Xiang C, Zhang C, Xu B, Wu J, Wang R, Yang Y, Shi L, Zhang J, Zhan Z: **Microbial biomarkers of common tongue coatings in patients with gastric cancer.** *Microb Pathog* 2019, **127**:97-105.
28. Li J, Yuan P, Hu X, Huang J, Cui L, Cui J, Ma X, Jiang T, Yao X, Li J *et al*: **A tongue features fusion approach to predicting prediabetes and diabetes with machine learning.** *J Biomed Inform* 2021, **115**:103693.
29. Lu C, Zhu H, Zhao D, Zhang J, Yang K, Lv Y, Peng M, Xu X, Huang J, Shao Z *et al*: **Oral-Gut Microbiome Analysis in Patients With Metabolic-Associated Fatty Liver Disease Having Different Tongue Image Feature.** *Front Cell Infect Microbiol* 2022, **12**:787143.
30. Pang W, Zhang D, Zhang J, Li N, Zheng W, Wang H, Liu C, Yang F, Pang B: **Tongue features of patients with coronavirus disease 2019: a retrospective cross-sectional study.** *Integr Med Res* 2020, **9**(3):100493.
31. Li S, Wang R, Zhang Y, Zhang X, Layon AJ, Li Y, Chen M: **Symptom combinations associated with outcome and therapeutic effects in a cohort of cases with SARS.** *Am J Chin Med* 2006, **34**(6):937-947.
32. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J: **A guide to deep learning in healthcare.** *Nat Med* 2019, **25**(1):24-29.
33. Greener JG, Kandathil SM, Moffat L, Jones DT: **A guide to machine learning for biologists.** *Nat Rev Mol Cell Biol* 2022, **23**(1):40-55.
34. Skrede OJ, De Raedt S, Kleppe A, Hveem TS, Liestol K, Maddison J, Askautrud HA, Pradhan M, Nesheim JA, Albrechtsen F *et al*: **Deep learning for prediction of colorectal cancer outcome: a discovery and validation study.** *Lancet* 2020, **395**(10221):350-360.
35. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak J, van Ginneken B, Sanchez CI: **A survey on deep learning in medical image analysis.** *Med Image Anal* 2017, **42**:60-88.
36. van der Laak J, Litjens G, Ciompi F: **Deep learning in histopathology: the path to the clinic.** *Nat Med* 2021, **27**(5):775-784.
37. Lei Y, Li S, Liu Z, Wan F, Tian T, Li S, Zhao D, Zeng J: **A deep-learning framework for multi-level peptide-protein interaction prediction.** *Nat Commun* 2021, **12**(1):5465.
38. Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, van der Laak J, Hulsbergen-van de Kaa C, Litjens G: **Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study.** *Lancet Oncol* 2020, **21**(2):233-241.

39. Li J, Chen Q, Hu X, Yuan P, Cui L, Tu L, Cui J, Huang J, Jiang T, Ma X *et al*: **Establishment of noninvasive diabetes risk prediction model based on tongue features and machine learning techniques.** *Int J Med Inform* 2021, **149**:104429.
40. Dixon MF, Genta RM, Yardley JH, Correa P: **Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994.** *Am J Surg Pathol* 1996, **20**(10):1161-1181.
41. You WC, Blot WJ, Li JY, Chang YS, Jin ML, Kneller R, Zhang L, Han ZX, Zeng XR, Liu WD *et al*: **Precancerous gastric lesions in a population at high risk of stomach cancer.** *Cancer Res* 1993, **53**(6):1317-1321.
42. Zhang L, Blot WJ, You WC, Chang YS, Kneller RW, Jin ML, Li JY, Zhao L, Liu WD, Zhang JS *et al*: **Helicobacter pylori antibodies in relation to precancerous gastric lesions in a high-risk Chinese population.** *Cancer Epidemiol Biomarkers Prev* 1996, **5**(8):627-630.
43. Redmon J, Divvala S, Girshick R, Farhadi A: **You Only Look Once: Unified, Real-Time Object Detection.** *Proc Cvpr Ieee* 2016:779-788.
44. He K, Zhang X, Ren S, Sun J: **Deep Residual Learning for Image Recognition.** *IEEE* 2016.
45. Su SB, Lu A, Li S, Jia W: **Evidence-Based ZHENG: A Traditional Chinese Medicine Syndrome.** *Evid Based Complement Alternat Med* 2012, **2012**:246538.
46. Li S: **Mapping ancient remedies: Applying a network approach to traditional Chinese medicine.** *Science* 2015, **350**:S72-S74.
47. Cai Q, Zhu C, Yuan Y, Feng Q, Feng Y, Hao Y, Li J, Zhang K, Ye G, Ye L *et al*: **Development and validation of a prediction rule for estimating gastric cancer risk in the Chinese high-risk population: a nationwide multicentre study.** *Gut* 2019, **68**(9):1576-1587.
48. Redeen S, Petersson F, Jonsson KA, Borch K: **Relationship of gastroscopic features to histological findings in gastritis and Helicobacter pylori infection in a general population sample.** *Endoscopy* 2003, **35**(11):946-950.
49. Wang P, Shi B, Wen Y, Tang X: **Construction of risk prediction model for precancerous lesions of gastric cancer combined with disease and syndrome(in Chinese).** *Chinese Journal of integrated traditional Chinese and Western Medicine* 2018, **38**(7):773 - 778.