

## Article

# Construction of Tongue Image-Based Machine Learning Model for Screening Patients with Gastric Precancerous Lesions

Changzheng Ma <sup>1,†</sup>, Peng Zhang <sup>1,†</sup>, Shiyu Du <sup>2</sup>, Yan Li <sup>3</sup> and Shao Li <sup>1,\*</sup> 

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

<sup>2</sup> Department of Gastroenterology, China-Japan Friendship Hospital, Chaoyang District, Beijing 100029, China

<sup>3</sup> Department of Traditional Chinese Medicine, Yijishan Hospital of Wannan Medical College, Wuhu 241000, China

\* Correspondence: shaoli@tsinghua.edu.cn

† These authors contributed equally to this work.

**Abstract:** Screening patients with precancerous lesions of gastric cancer (PLGC) is important for gastric cancer prevention. The accuracy and convenience of PLGC screening could be improved with the use of machine learning methodologies to uncover and integrate valuable characteristics of noninvasive medical images related to PLGC. In this study, we therefore focused on tongue images and for the first time constructed a tongue image-based PLGC screening deep learning model (AITongue). The AITongue model uncovered potential associations between tongue image characteristics and PLGC, and integrated canonical risk factors, including age, sex, and Hp infection. Five-fold cross validation analysis on an independent cohort of 1995 patients revealed the AITongue model could screen PLGC individuals with an AUC of 0.75, 10.3% higher than that of the model with only including 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 developed a smartphone-based app screening system to enhance the application convenience of the AITongue model in the natural population from high-risk areas of gastric cancer in China. Collectively, our study has demonstrated the value of tongue image characteristics in PLGC screening and risk prediction.

**Keywords:** precancerous lesions of gastric cancer; tongue image; deep learning; disease screening; risk prediction; artificial intelligence



**Citation:** Ma, C.; Zhang, P.; Du, S.; Li, Y.; Li, S. Construction of Tongue Image-Based Machine Learning Model for Screening Patients with Gastric Precancerous Lesions. *J. Pers. Med.* **2023**, *13*, 271. <https://doi.org/10.3390/jpm13020271>

Academic Editors: Dorota Bielińska-Waż and Piotr Wąż

Received: 28 December 2022

Revised: 25 January 2023

Accepted: 29 January 2023

Published: 31 January 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Gastric cancer is the second leading cause of cancer death 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], suffer 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 great contribution to facilitating the early prevention of gastric cancer.

Current screening methods suffer from some challenges, including invasiveness and relatively low accuracy, which limits their applications in population screening. On the one hand, although gastroscopy and biopsy are the gold standards for gastric disease diagnosis [7], these methods remain inefficient and unfeasible for gastric disease screening [8]. As previous studies indicated, approximately half of the patients screened with gastroscopy are non-atrophic gastritis, and the early diagnosis rate of gastric cancer remains less than 20% [1,9]. On the other 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], has been limited for risk screening in natural populations due to the high sensitivity and specificity thresholds required [13].

In addition, it is not cost-effective to use either serum pepsinogen test screening or endoscopy as they difficulty in their practical application [14]. Screening high-risk groups for gastroscopy could triage patients and effectively improve the utilization efficiency of medical resources. Thus, considering the requirements of large-scale screening, screening methods with a high cost-effectiveness ratio and high accuracy are urgently needed to enhance their popularization [15].

As non-invasive indicators, tongue image characteristics have been used for the surveillance of a broad spectrum of diseases, inspired by the diagnosis experience in traditional Chinese medicine (TCM) [16–20]. Tongue image characteristics, including shape, color, and tongue coating, are believed to reflect the health condition, or the severity and progress of disease, especially for digestive diseases as the tongue is anatomically connected to the digestive system organs. For example, recent studies have indicated that tongue image characteristics show correlations with gastroscopic observations and could be used to predict gastric mucosal health [21,22]. In addition, it was revealed that tongue surface and color characteristics could be used as indicators to assist in gastric cancer diagnosis [23,24]. 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 [17,25–27]. From the pathologic and etiologic perspectives, the distribution of microorganisms on the tongue coating has also been found to be related to gastric diseases, which helps uncover non-invasive microbial markers for gastric disease risk screening [28–30]. The above studies demonstrate the great potential of tongue image characteristics in assisting 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 [31–37]. For example, some studies have applied deep learning to predict diverse cancer types, including prostate cancer and rectal cancer, based on medical images [33,38]. Using tongue images, some studies have applied deep learning techniques to identify risk features in tongue images for the detection of diseases such as stomach cancer and diabetes [23,39]. Therefore, deep learning techniques could be a pivotal tool to uncover the risk characteristics from tongue images, and further constructed a machine learning-based screening model.

Therefore, to improve the efficiency of screening patients with PLGC, particularly in natural populations, this study aimed to build a machine learning-based PLGC screening model which introduces tongue image information on the basis of existing risk indicators. In detail, we firstly explored the tongue image characteristics of patients with PLGC and integrated them with canonical screening indicators to develop a PLGC screening model called AITongue. We then evaluated its screening effect by external validation in an independent cohort and finally explored its potential value as a risk predictor of PLGC in a follow-up cohort. To our best knowledges, the AITongue model we have developed should be the first tongue image-based machine learning model for PLGC screening and risk prediction. We believe that our study will pave the way to addressing the urgent need for non-invasive PLGC screening in clinical practice.

## 2. Materials and Methods

### 2.1. Patient Enrollment, and Data Collection

Patients were enrolled in this study at the China-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 the Institution Review Board of Tsinghua University (protocol code 20200069). Inclusion criteria: At least 18 years of age, clear language skills, no barriers in communication and willingness to accept clinical investigation and sign informed consent. Exclusion criteria: The presence of heart, cerebrovascular, liver, kidney, hematopoietic system diseases.

## 2.2. Gastroscopy and Histological Examination

Using video endoscopes (Olympus Corp), upper gastroscopic examinations were performed by two gastroenterologists. Tissue samples for biopsy were reviewed blindly by the two 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].

## 2.3. Data Pre-Processing and Data Structuring

As the pivotal step for data pre-processing, a deep-learning model was constructed to identify and segment tongue bodies in raw images while excluding face and background information. Here, we trained the tongue body recognition and locating model using the YOLOv5 model and 180 tongue images that were labeled by TCM physicians with a square frame using “labelImg” software [43]. The YOLOv5 model is a common deep learning model for target detection which can accurately identify and locate the position of specific objects after training. Furthermore, using this model, we carried out tongue body recognition and cutting on tongue images, reshaped the images to  $224 \times 224$ , and formed a pre-processed tongue body image dataset. In this way, we could segment the tongue from the complex background to reduce the impact of the background on classification and improve accuracy. Python (3.7.0) and PyTorch were used for the tongue image preprocessing. Using this model, the tongue images were detected and cut into tongue body images for subsequent analysis.

Additional clinicopathological characteristics of the enrolled patients were obtained from electronic medical records. The obtained characteristics included basic 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  $\leq 50$  years based on the median of age distribution. Multiple interpolation methods were used to fill in the missing data. Tongue labels (fissure, etc.) were assigned by physicians.

## 2.4. PLGC Screening Model Construction

The PLGC screening model was constructed following two main steps: image classification with a deep learning model, and data integration with a logistic regression model.

Firstly, the image classification model was constructed with the ResNet50 deep learning model [44]. The ResNet50 model has a wide range of applications and good performance in the field of image classification as it can introduce the residual blocks. In our study, the residual blocks of the ResNet50 model are structured as two bottlenecks (BTNK), designated as BTNK 1 and BTNK 2. Their structure diagram is shown in Supplementary Figure S1, where CONV is the convolution block, BN is the batch normalization block, and Relu is an activation function in the bottleneck. After the ResNet50 module, tongue images were classified into two categories: high-risk and low-risk.

A logistic regression model was then used to predict the PLGC screening results by integrating the tongue image classification results and the clinicopathological indicators. Logistic regression models have good performance in the integration of a small number of variables and robust prediction of classification tasks, resulting in their wide application in disease classification and risk prediction research.

## 2.5. 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 both univariate and multivariate analyses were included in the model. The significance

of each factor adjusted for gender and age was calculated in the multivariate analysis. Binary logistic regression was used to construct the screening models. Chi-square tests were applied to calculate the significance of the independent variables for PLGC. Pearson's correlation coefficient was applied to evaluate the correlation between the independent variables. Accuracy, sensitivity, specificity, recall, precision, receiver operating characteristic (ROC) curve, and area under the curve (AUC) were used as evaluation metrics to evaluate model performance.

AUC-ROC curves are performance measures for classification problems under various thresholds. ROC is a probability curve, and AUC represents the degree or measure of separability. The horizontal coordinate of the ROC curve is the false positive rate (FPR), and the vertical coordinate is the true positive rate (TPR). The calculation formula is as follows.

$$TPR = \frac{TP}{TP+FN}$$

$$FPR = \frac{FP}{FP+TN}$$

TP, FP, TN, and FN represent true positives, false positives, true negatives, and false negatives, respectively. In the classification task, the model represents the prediction and ground truth. The higher the AUC, the better the classification performance of the model.

### 3. Results

#### 3.1. The Overall Design of Our Study

In our study, a total of three cohorts of patients were enrolled with undergoing gastroscopy and pathology. These included a development cohort, validation cohort, and follow-up cohort. Here, two categories, including PLGC and non-PLGC, were derived for each patient based on pathological diagnosis (Table 1). In detail, we developed the PLGC screening model and performed an internal cross-validation on the development cohort, which consisted of 325 patients, including 55 PLGC and 270 non-PLGC patients. We then performed external validation on the validation cohort, which had a total of 1995 patients, including 171 PLGC and 1824 non-PLGC patients. It should be noted that we also evaluated the risk prediction value of the PLGC screening model on the follow-up cohort, in which only non-PLGC patients were enrolled at the baseline timepoint, and were further classified as Pro or non-Pro according to the pathological lesions at the endpoint after a mean follow-up time of 22 months (Figure 1).

#### 3.2. Construction of AITongue Model with Integrating Tongue Image Characteristics

After preprocessing the tongue images with deep learning (Figure 2a), a PLGC screening model, which we named the AITongue model, was constructed in the development cohort by integrating the tongue image and clinicopathological characteristics, as shown in Figure 2b. Here, the tongue images were classified into two categories: high-risk and low-risk. The AITongue model took the categorized results, and canonical gastric cancer risk indicators (age, gender, and Hp infection) as the input, and the PLGC prediction results as the output.

**Table 1.** Basic information of three cohorts.

Cohorts	Groups	Counts	Age (Mean ± SD)	Gender (Male/Female)
Development cohort	PLGC	55	58.6 ± 10.5	31/24
	non-PLGC	270	48.0 ± 13.6	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 deviation.

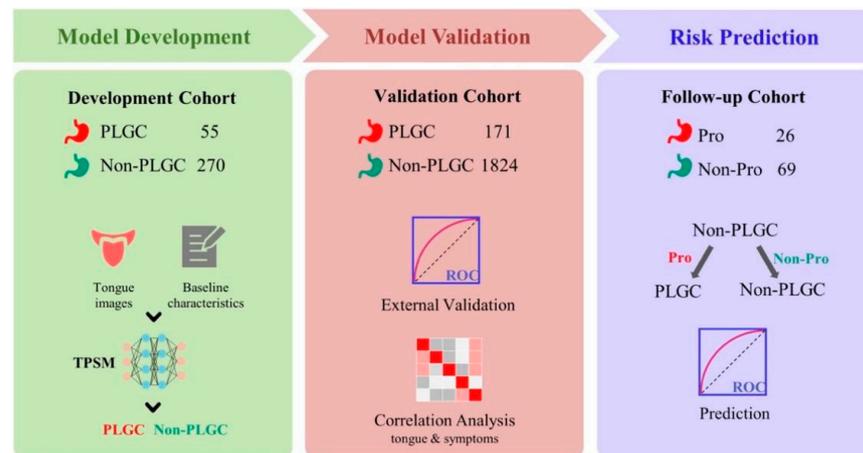


Figure 1. Outline of the workflow for our study.

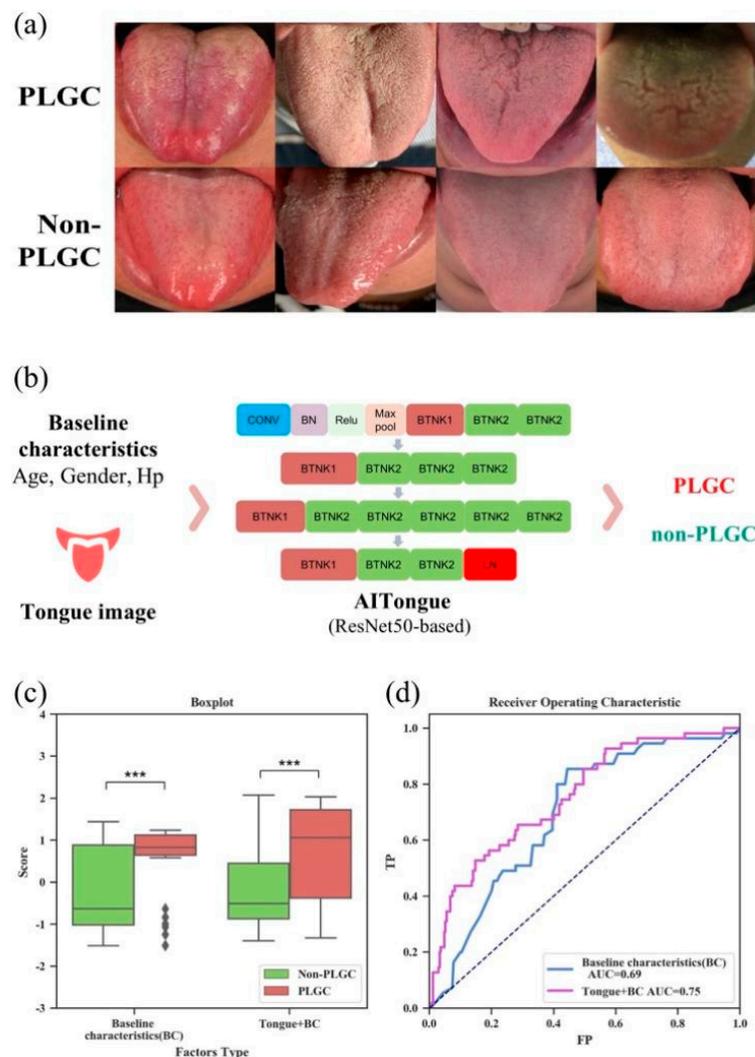


Figure 2. Construction of the AITongue model and results for PLGC screening based on tongue images. (a) Example of tongue images of PLGC and non-PLGC patients. (b). ResNet50-based deep learning screening model AITongue. (c). Boxplot of classification score comparisons with and without the inclusion of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

We then further investigated and interpreted the tongue image characteristics with PLGC screening potential. Through performing correlation analysis between image risk classification (high vs. low-risk) obtained by the deep-learning model and TDL labels generated by TCM experts, we found that five of the TDLs were statistically significant ( $p < 0.05$ ), namely greasy, fissured, dark, coating (yellow), and coating (thick). This indicated, to some extent, the medical significance of the risk features found in tongue images and suggests that there may also be some value of TDLs for PLGC screening (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
	Counts	Ratio	Counts	Ratio	
Coating (Yellow)	16	0.17	16	0.07	$9.6 \times 10^{-3}$
Greasy	28	0.30	34	0.15	$2.6 \times 10^{-3}$
Fissured	19	0.20	26	0.11	$4.8 \times 10^{-2}$
Coating (Thick)	14	0.15	15	0.06	$2.6 \times 10^{-2}$
Dark	12	0.13	7	0.03	$1.6 \times 10^{-3}$

p value refer to the comparison between high risk and low risk groups by Pearson’s chi squared test.

### 3.3. External Validation of PLGC Screening

We then validated the performance in PLGC screening of the AITongue model in the independent validation cohort. To enhance the robustness and application value of our model, the five representative TDLs that included greasy, fissured, dark, coating (yellow), and coating (thick), rather than the whole image characteristics, were selected as inputs for the AITongue model. Of note, it was found that these five TDLs, along with gender and age, showed significant correlations with PLGC in both univariate and multivariate analyses (Table 3, Table S2), supporting their value as input parameters for AITongue.

It was found that the AITongue model showed a comparably discriminative performance in the independent validation cohort compared with that of the development cohort. Here, the AITongue model exhibited an accuracy of 0.64 and AUC of 0.75 in distinguishing PLGC from non-PLGC patients. In contrast, the model with only including baseline indicators (age, sex, Hp) showed an accuracy of 0.53 and AUC of 0.68 (Figure 3). Thus, we could conclude that the discriminative performance between PLGC and non-PLGC has been significantly enhanced by 10.3% (0.68 vs. 0.75,  $p < 0.01$ , Figure 3) by introducing tongue image characteristics. The results furtherly validated the effectiveness of tongue image characteristics for PLGC screening.

In addition, we also focused on clinical symptom characteristics and investigated their screening value for PLGC and confounding effects for the AITongue model [45].

First, we investigated the screening value of symptom characteristics by analyzing their associations with PLGC. As a result, three symptom characteristics (xerostomia, bitter taste, belching) showed significant correlations with PLGC in both univariate and multivariate analyses, whereas the others, including stomach pain, bloating, chilliness, and loose stools, did not show significant correlations (Supplementary Table S2).

Further, we incorporated these three symptom characteristics into the AITongue model to evaluate the enhancement of introducing symptom characteristics for PLGC screening. The validation cohort was used as training data to construct a logistic regression model and a five-fold cross-validation was performed. It showed a small improvement in the discrimination between PLGC and non-PLGC after introducing symptom characteristics (0.76 vs. 0.73, Figure 4). These results indicate that the introduction of clinical symptom characteristics could improve the screening efficiency of PLGC, with a slightly lower effect than tongue image characteristics. In addition, there was a low correlation between tongue image and symptom characteristics (Supplementary Figure S2), which indicates that tongue image characteristics might be independent factors from clinical symptom characteristics in terms of PLGC screening.

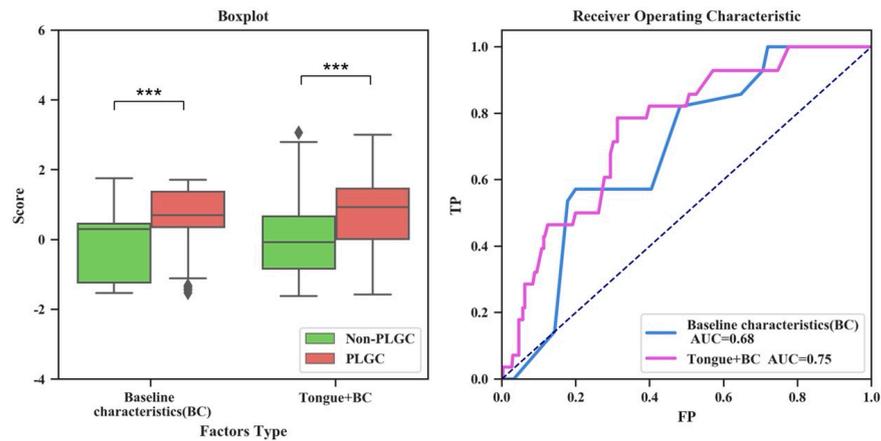
**Table 3.** Univariate and multivariate analyses of baseline factors and TDLs in PLGC screening.

Variable	Total (N = 1995) No. (%)	Non-PLGC (N = 1824) No. (%)	PLGC (N = 171) No. (%)	p Value *	Adjusted OR (95% CI) †	p Value #
Baseline Factors						
Age, Years				$3.7 \times 10^{-15}$		
>50	1067 (0.53)	926 (0.51)	141 (0.82)		4.14 ([2.59,6.61])	$2.9 \times 10^{-9}$
≤50	928 (0.47)	898 (0.49)	30 (0.18)			
Gender				$1.3 \times 10^{-7}$		
Male	785 (0.39)	685 (0.38)	100 (0.58)		1.47 ([1.00,2.14])	$4.8 \times 10^{-2}$
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				$1.9 \times 10^{-10}$		
Yellow	327 (0.16)	269 (0.15)	58 (0.34)		3.66 ([2.38,5.62])	$3.4 \times 10^{-9}$
White	1668 (0.84)	1555 (0.85)	113 (0.66)			
Fissure				$3.8 \times 10^{-9}$		
Yes	187 (0.09)	149 (0.08)	38 (0.22)		2.26 ([1.37,3.73])	$1.4 \times 10^{-3}$
No	1808 (0.91)	1675 (0.92)	133 (0.78)			
Greasy				$3.3 \times 10^{-4}$		
Yes	422 (0.21)	367 (0.20)	55 (0.32)		2.23 ([1.44,3.46])	$3.3 \times 10^{-4}$
No	1573 (0.79)	1457 (0.80)	116 (0.68)			
Coating				$1.3 \times 10^{-3}$		
Thick	239 (0.12)	205 (0.11)	34 (0.20)		2.18 ([1.29,3.67])	$3.6 \times 10^{-3}$
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])	$6.7 \times 10^{-3}$
No	1683 (0.84)	1549 (0.85)	134 (0.78)			

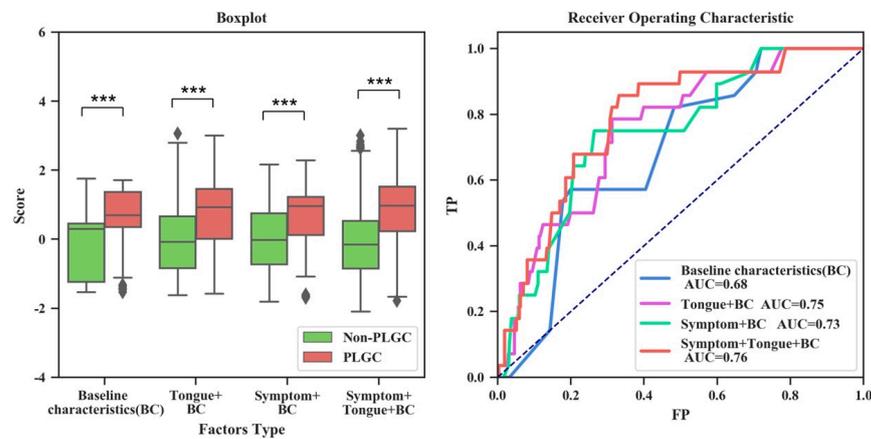
p value \* refer to the univariate analysis. p value # refer to the multivariate analysis (with adjustment for gender and age). † Variables without significance ( $p > 0.05$ ) are not shown.

### 3.4. Evaluation of the Validity of Tongue Image Characteristics for Risk Prediction of PLGC

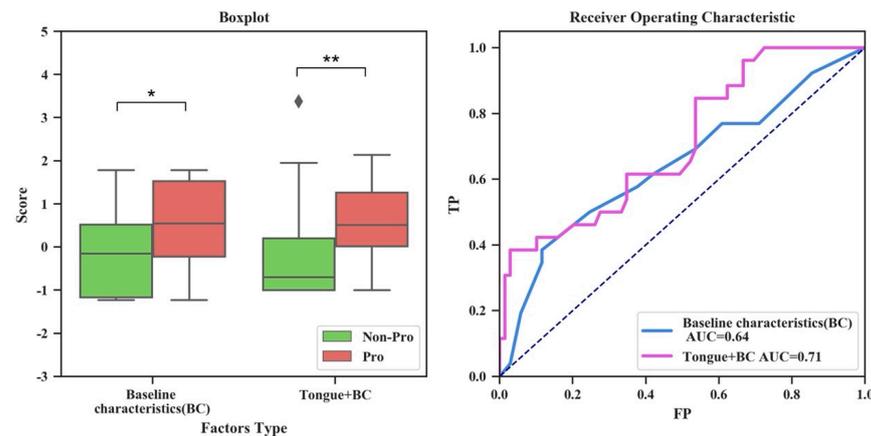
PLGC risk prediction is pivotal for gastric cancer early prevention. Thus, we further explored the value of the AITongue model in predicting the risk of PLGC. Here, we enrolled a cohort of non-PLGC patients and conducted a long-term follow-up surveillance, in which patients were divided into progressive (Pro) and non-progressive (non-Pro) groups, respectively, according to endpoint pathological diagnosis (Table 1). Using the AITongue model to score the PLGC risk for each patient in the follow-up cohort, we found that risk scores from the Pro group were significantly higher than those from the Reg group. The AUC value was 0.71, which showed a significant increase (10.94%, 0.71 vs. 0.64,  $p < 0.01$ ) compared with that derived from the model only including baseline indicators (Figure 5). In addition, we performed a univariate analysis of the TDLs for risk prediction of PLGC (Supplementary Table S3). It was found that the TDLs showed limited value for PLGC risk prediction, although they have an enhanced effect on PLGC screening. Therefore, tongue image characteristics are potentially valuable in PLGC risk prediction.



**Figure 3.** Boxplot (left) and ROC curve (right) comparisons of screening scores with and without the inclusion of tongue image characteristics for PLGC screening. (\*\*\*:  $p < 0.001$ ).



**Figure 4.** Boxplot (left) and ROC curve (right) comparisons of the screening scores with and without the inclusion of symptoms factors for PLGC screening. (\*\*\*:  $p < 0.001$ ).



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

**4. Discussion**

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 [46,47]. In this study, PLGC was analyzed using 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 [48,49].

Tongue diagnosis is an important part of the four diagnoses in TCM. In TCM theory, the characteristics of tongue images are quantified into various categories for the diagnosis of diseases [50,51]. In this study, we not only found that helpful feature information for PLGC screening could be extracted through a deep learning model, but also found that these features were related to some categories of tongue diagnosis in traditional Chinese medicine. This suggests that some categories of Chinese tongue diagnosis have interpretable morphological characteristics for PLGC screening and for tongue images of high-risk categories.

Our proposed method has better performance than another study of screening of PLGC. Wang et al. developed a model with non-invasive indicators for PLGC screening based on 290 patients with gastritis, and the AUC was 0.728 (95% CI: 0.651–0.793), whereas the AUC of our method was 0.76 [52].

It is necessary to adopt cost-effective methods to conduct the large-scale screening for gastric cancer risk in the natural population. Even though gastroscopy and pathological tests are the gold standard for the diagnosis of gastric diseases, they are not suitable for the natural population. The method we developed introduces tongue image information on the basis of conventional and invasive indicators, which improves the accuracy, reduces the difficulty of operation and improve its feasibility for application.

The study has some limitations. The data source was biased compared with 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 (Supplementary Figure S3). In further studies, more samples would be collected from natural populations to reduce bias, and larger external validation should be conducted.

## 5. Conclusions

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. It was then externally validated in an independent validation cohort and used to evaluate the capability for risk prediction of PLGC in a follow-up cohort. Our study demonstrates the value of tongue image characteristics in PLGC screening and its potential for risk prediction.

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 use in large-scale applications.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jpm13020271/s1>. Figure S1: The residual module of the ResNet50 model; Figure S2: Correlation analysis of TDLs and symptoms; Figure S3: The operation interface of the App; Table S1: Univariate and multivariate analysis of symptoms factors in PLGC screening; Table S2: Univariate analysis of tongue diagnostic and symptoms factors in PLGC screening; Table S3: The univariate analysis of tongue diagnostic labels in risk prediction of PLGC.

**Author Contributions:** Conceptualization, C.M. and S.L.; methodology, C.M. and P.Z.; software, C.M.; validation, C.M. and P.Z.; formal analysis, C.M.; data curation, C.M. and S.D.; writing—original draft preparation, C.M.; writing—review and editing, P.Z.; visualization, C.M.; supervision, S.L. and Y.L.; project administration, S.L. and S.D.; funding acquisition, S.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China, China (62061160369 and 81225025); and the Beijing National Research Center for Information Science and Technology, China (BNR2019TD01020 and BNR2019RC01012).

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and was approved by the Human Ethics Committee of Institution Review Board of Tsinghua University (protocol code 20200069).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Not applicable.

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

**Conflicts of Interest:** A patent application submitted by the authors is related to certain aspects of the work described in this manuscript.

## References

1. Zong, L.; Abe, M.; Seto, Y.; Ji, J. The challenge of screening for early gastric cancer in China. *Lancet* **2016**, *388*, 2606. [[CrossRef](#)] [[PubMed](#)]
2. Schlemper, R.J.; Riddell, R.H.; Kato, Y.; Borchard, F.; Cooper, H.S.; Dawsey, S.M.; Dixon, M.F.; Fenoglio-Preiser, C.M.; Fléjou, J.F.; Geboes, K.; et al. The Vienna classification of gastrointestinal epithelial neoplasia. *Gut* **2000**, *47*, 251–255. [[CrossRef](#)] [[PubMed](#)]
3. Song, H.; Ekhedden, I.G.; 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. [[CrossRef](#)] [[PubMed](#)]
4. de Vries, A.C.; van Grieken, N.C.; Looman, C.W.; Casparie, M.K.; de Vries, E.; Meijer, G.A.; Kuipers, E.J. Gastric cancer risk in patients with premalignant gastric lesions: A nationwide cohort study in the Netherlands. *Gastroenterology* **2008**, *134*, 945–952. [[CrossRef](#)]
5. Piazuolo, M.B.; Bravo, L.E.; Mera, R.M.; Camargo, M.C.; Bravo, J.C.; Delgado, A.G.; Washington, M.K.; Rosero, A.; Garcia, L.S.; Realpe, J.L.; et al. The Colombian Chemoprevention Trial: 20-Year Follow-Up of a Cohort of Patients with Gastric Precancerous Lesions. *Gastroenterology* **2021**, *160*, 1106–1117.e1103. [[CrossRef](#)]
6. Rugge, M.; Meggio, A.; Pravadelli, C.; Barbareschi, M.; Fassan, M.; Gentilini, M.; Zorzi, M.; Pretis, G.; Graham, D.Y.; Genta, R.M. 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*, 11–17. [[CrossRef](#)]
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*, 63. [[CrossRef](#)]
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. [[CrossRef](#)]
10. Tu, H.; Sun, L.; Dong, X.; Gong, Y.; Xu, Q.; Jing, J.; Bostick, R.M.; 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*, 704–715. [[CrossRef](#)]
11. Huang, S.; Guo, Y.; Li, Z.W.; Shui, G.; Tian, H.; Li, B.W.; Kadeerhan, G.; Li, Z.X.; 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*, e2114186. [[CrossRef](#)]
12. Huang, K.K.; Ramnarayanan, K.; Zhu, F.; Srivastava, S.; Xu, C.; Tan, A.L.K.; 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*, 137–150.e135. [[CrossRef](#)]
13. Cubiella, J.; Perez Aisa, A.; Cuatrecasas, M.; Diez Redondo, P.; Fernandez Esparrach, G.; Marin-Gabriel, J.C.; Moreira, L.; Nunez, H.; Pardo Lopez, M.L.; Rodriguez de Santiago, E.; et al. Gastric cancer screening in low incidence populations: Position statement of AEG, SEED and SEAP. *Gastroenterol. Hepatol.* **2021**, *44*, 67–86. [[CrossRef](#)]
14. Pimentel-Nunes, P.; Libanio, D.; Marcos-Pinto, R.; Areia, M.; Leja, M.; Esposito, G.; Garrido, M.; Kikuste, I.; Megraud, F.; Matsiyak-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*, 365–388.
15. Afrash, M.R.; Shafiee, M.; Kazemi-Arpanahi, H. Establishing machine learning models to predict the early risk of gastric cancer based on lifestyle factors. *BMC Gastroenterol.* **2023**, *23*, 6. [[CrossRef](#)]
16. Jiang, T.; Guo, X.J.; Tu, L.P.; Lu, Z.; Cui, J.; Ma, X.X.; Hu, X.J.; Yao, X.H.; Cui, L.T.; Li, Y.Z.; et al. Application of computer tongue image analysis technology in the diagnosis of NAFLD. *Comput. Biol. Med.* **2021**, *135*, 104622. [[CrossRef](#)]

17. 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. [[CrossRef](#)]
18. 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. [[CrossRef](#)]
19. 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. [[CrossRef](#)]
20. Zhang, B.; Kumar, B.V.; Zhang, D. Detecting diabetes mellitus and nonproliferative diabetic retinopathy using tongue color, texture, and geometry features. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 491–501. [[CrossRef](#)]
21. Shang, Z.; Du, Z.G.; Guan, B.; Ji, X.Y.; Chen, L.C.; Wang, Y.J.; Ma, Y. Correlation analysis between characteristics under gastroscopie and image information of tongue in patients with chronic gastritis. *J. Tradit. Chin. Med.* **2022**, *42*, 102–107.
22. 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. [[CrossRef](#)]
23. Gholami, E.; Tabbakh, S.R.K.; kheirabadi, M. Increasing the accuracy in the diagnosis of stomach cancer based on color and lint features of tongue. *Biomed. Signal Process. Control* **2021**, *69*, 102782. [[CrossRef](#)]
24. Zhu, X.; Ma, Y.; Guo, D.; Men, J.; Xue, C.; Cao, X.; Zhang, Z. A Framework to Predict Gastric Cancer Based on Tongue Features and Deep Learning. *Micromachines* **2023**, *14*, 53. [[CrossRef](#)]
25. 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. [[CrossRef](#)]
26. 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. [[CrossRef](#)]
27. 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*, 100493. [[CrossRef](#)]
28. 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. Genom.* **2022**, *15*, 60. [[CrossRef](#)]
29. 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*, 496–509. [[CrossRef](#)]
30. 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. [[CrossRef](#)]
31. 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*, 24–29. [[CrossRef](#)]
32. Zhou, W.; Yang, K.; Zeng, J.; Lai, X.; Wang, X.; Ji, C.; Li, Y.; Zhang, P.; Li, S. FordNet: Recommending traditional Chinese medicine formula via deep neural network integrating phenotype and molecule. *Pharmacol. Res.* **2021**, *173*. [[CrossRef](#)]
33. Skrede, O.J.; De Raedt, S.; Kleppe, A.; Hveem, T.S.; Liestol, K.; Maddison, J.; Askautrud, H.A.; Pradhan, M.; Nesheim, J.A.; Albrechtsen, F.; et al. Deep learning for prediction of colorectal cancer outcome: A discovery and validation study. *Lancet* **2020**, *395*, 350–360. [[CrossRef](#)]
34. Zhou, H.; Liu, Z.; Li, T.; Chen, Y.; Huang, W.; Zhang, Z. Classification of precancerous lesions based on fusion of multiple hierarchical features. *Comput. Biol. Med.* **2023**, *229*, 107301.
35. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.; van Ginneken, B.; Sanchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)]
36. Yaqoob, M.M.; Nazir, M.; Yousafzai, A.; Khan, M.A.; Shaikh, A.A.; Algarni, A.D.; Elmannai, H. Modified Artificial Bee Colony Based Feature Optimized Federated Learning for Heart Disease Diagnosis in Healthcare. *Appl. Sci.* **2022**, *12*, 12080. [[CrossRef](#)]
37. van der Laak, J.; Litjens, G.; Ciompi, F. Deep learning in histopathology: The path to the clinic. *Nat. Med.* **2021**, *27*, 775–784. [[CrossRef](#)]
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*, 233–241. [[CrossRef](#)]
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. [[CrossRef](#)]
40. Dixon, M.F.; Genta, R.M.; Yardley, J.H.; 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*, 1161–1181. [[CrossRef](#)]
41. You, W.C.; Blot, W.J.; Li, J.Y.; Chang, Y.S.; Jin, M.L.; Kneller, R.; Zhang, L.; Han, Z.X.; Zeng, X.R.; Liu, W.D.; et al. Precancerous gastric lesions in a population at high risk of stomach cancer. *Cancer Res.* **1993**, *53*, 1317–1321. [[PubMed](#)]
42. Zhang, L.; Blot, W.J.; You, W.C.; Chang, Y.S.; Kneller, R.W.; Jin, M.L.; Li, J.Y.; Zhao, L.; Liu, W.D.; Zhang, J.S.; et al. Helicobacter pylori antibodies in relation to precancerous gastric lesions in a high-risk Chinese population. *Cancer Epidemiol. Biomark. Prev.* **1996**, *5*, 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. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
45. Li, S.; Wang, R.; Zhang, Y.; Zhang, X.; Layon, A.J.; 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*, 937–947. [[CrossRef](#)] [[PubMed](#)]
46. 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*, 1576–1587. [[CrossRef](#)]
47. Li, S.; Lu, A.P.; Zhang, L.; Li, Y.D. Anti-Helicobacter pylori immunoglobulin G (IgG) and IgA antibody responses and the value of clinical presentations in diagnosis of H. pylori infection in patients with precancerous lesions. *World J. Gastroenterol.* **2003**, *9*, 755–758. [[CrossRef](#)]
48. Redeen, S.; Petersson, F.; Jonsson, K.A.; Borch, K. Relationship of gastroscopic features to histological findings in gastritis and Helicobacter pylori infection in a general population sample. *Endoscopy* **2003**, *35*, 946–950.
49. Su, S.; Lu, A.; Li, S.; Jia, W. Evidence-Based ZHENG: A Traditional Chinese Medicine Syndrome. *Evid. Based Complement Altern. Med.* **2012**, *2012*, 246538. [[CrossRef](#)]
50. Kanawong, R.; Ajayi, T.; Ma, T.; Xu, D.; Li, S.; Duan, Y. Automated tongue feature extraction for ZHENG classification in traditional Chinese medicine. *Evid.-Based Complement. Altern. Med.* **2012**, *2012*, 912852. [[CrossRef](#)]
51. Kanawong, R.; Xu, W.; Xu, D.; Li, S.; Ma, T.; Duan, Y. An automatic tongue detection and segmentation framework for computer-aided tongue image analysis. *Int. J. Funct. Inform. Pers. Med.* **2012**, *4*, 56–68. [[CrossRef](#)]
52. Wang, P.; Shi, B.; Wen, Y.; Tang, X. Construction of risk prediction model for precancerous lesions of gastric cancer combined with disease and syndrome. *Chin. J. Integr. Tradit. Chin. West. Med.* **2018**, *38*, 773–778. (in Chinese).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.