

Pre-endoscopic screening of precancerous lesions in gastric cancer using deep learning

1 Lan Wang^{1#}, Qian Zhang^{1#}, Peng Zhang¹, Bowen Wu¹, Shiyu Du², Kaiqiang Tang³, ShaoLi^{1*}

2 ¹ Institute for TCM-X, MOE Key Laboratory of Bioinformatics, Bioinformatics Division, BNRIST,
3 Department of Automation, Tsinghua University, Beijing, China

4 ² Department of Gastroenterology, China-Japan Friendship Hospital, Chaoyang District, Beijing,
5 China

6 ³Department of Control and Systems Engineering, Nanjing University, Nanjing, China

7 # These authors contributed equally to this work.

8 * Correspondence:

9 Shao Li, MD

10 Institute for TCM-X, MOE Key Laboratory of Bioinformatics, Bioinformatics Division, BNRIST,
11 Department of Automation, Tsinghua University, Beijing, China

12 Email: shaoli@mail.tsinghua.edu.cn

13 Financial Support:

14 National Natural Science Foundation of China (62061160369).

15 National Natural Science Foundation of China (Grant Nos. T2341008).

16 **Keywords:** deep learning, tongue images, inquiry information, precancerous lesions of gastric
17 cancer, pre-endoscopic screening.

18 Abstract

19 **Objective:** Given the high cost of endoscopy in gastric cancer (GC) screening, there is an urgent
20 need to explore cost-effective methods for the large-scale prediction of precancerous lesions of
21 gastric cancer (PLGC). We aim to construct a hierarchical artificial intelligence-based multimodal
22 non-invasive method for pre-endoscopic risk screening, to provide tailored recommendations for
23 endoscopy.

24 **Design:** From December 2022 to December 2023, a large-scale screening study was conducted in
25 Fujian, China. Based on traditional Chinese medicine theory, we simultaneously collected tongue
26 images and inquiry information from 1034 participants, considering the potential of these data for
27 PLGC screening. Then, we introduced inquiry information for the first time, forming a multimodality
28 artificial intelligence model to integrate tongue images and inquiry information for pre-endoscopic
29 screening. Moreover, we validated this approach in another independent external validation cohort,
30 comprising 143 participants from the China-Japan Friendship Hospital.

31 **Results:** A multimodality artificial intelligence-assisted pre-endoscopic screening model based on
32 tongue images and inquiry information (AITonguequiry) was constructed, adopting a hierarchical
33 prediction strategy, achieving tailored endoscopic recommendations. Validation analysis revealed
34 that the area under the curve (AUC) values of AITonguequiry were 0.74 for PLGC (95% confidence
35 interval (CI) 0.71 to 0.76, $p < 0.05$) and 0.82 for high-risk PLGC (95% CI 0.82 to 0.83, $p < 0.05$),

36 which were significantly and robustly better than those of the independent use of either tongue
37 images or inquiry information alone, and also demonstrated superior performance compared to
38 existing screening methodologies. In the independent external verification, the AUC values were
39 0.69 for PLGC and 0.76 for high-risk PLGC.

40 **Conclusion:** Our AITonguequiry artificial intelligence model, for the first time, incorporates inquiry
41 information and tongue images, leading to a higher precision and finer-grained pre-endoscopic
42 screening of PLGC. This enhances patient screening efficiency and alleviates patient burden.

43 1 Introduction

44 According to recent surveys, gastric cancer is the fourth leading cause of cancer-related deaths
45 worldwide and the second in China [1]. There are approximately 480,000 new cases and 370,000
46 deaths of gastric cancer in China each year, accounting for half of the cases in the world [2]. Gastric
47 cancer is thought to develop from precancerous lesions of gastric cancer (PLGC) (e.g., chronic
48 atrophic gastritis, intestinal metaplasia, or gastric epithelial dysplasia), and the graded screening and
49 diagnosis of natural populations for PLGC is essential to reduce gastric cancer mortality [3-5].
50 However, the screening and diagnosis of gastric diseases still rely on gastroscopy, but its application
51 is greatly limited because of its invasiveness, high cost, and the need for professional endoscopists
52 [6]. Meanwhile, endoscopic screening is not suitable for large-scale natural populations, especially in
53 rural China. Alternatively, the application of serum markers that are commonly used as screening
54 factors in various gastric cancer risk assessment methods, such as pepsinogen I/II and gastrin-17, has
55 been limited for risk screening in natural populations due to its invasiveness, and the high sensitivity
56 and specificity thresholds [7, 8]. Therefore, there is an urgent need for affordable and non-invasive
57 screening methods suitable for large-scale natural populations to improve diagnostic efficiency and
58 reduce the incidence of gastric cancer.

59 Early studies have indicated that non-invasive features, including imaging characteristics and clinical
60 phenotypic information, have the potential to predict the occurrence and progression of PLGC. The
61 theory of Traditional Chinese medicine (TCM) suggests that the tongue's shape, color, size, and
62 coating characteristics detected using tongue images reflect health status and disease
63 severity/progression, and recent studies have shown the potential for tongue surface and color
64 characteristics to assist in the diagnosis of PLGC [9, 10]. For digestive diseases, tongue images
65 characteristics have been found to correlate with gastroscopic observations and predict gastric
66 mucosal health [11, 12]. Additionally, interrogating inquiry information (e.g., living habits, dietary
67 preferences, and physical symptoms) is crucial in understanding the disease and medical history [13].
68 In this regard, recent studies have built risk prediction models for PLGC before endoscopy using
69 demographics and clinical risk factors, including *H. pylori* infection, sex, age, race/ethnicity, and
70 smoking status [14]. However, the integration of tongue images and inquiry information for high-
71 precision endoscopic screening of PLGC to facilitate precise endoscopic recommendations has not
72 been studied.

73 With the rapid development of artificial intelligence (AI) technology, machine learning algorithms
74 based on deep neural networks can accurately analyze diagnostic clinical images, identify therapeutic
75 targets, and process large datasets, which can play a role in screening and diagnosis of a variety of
76 diseases [15-18]. Pan et al. reviewed studies related to AI methods for lung cancer risk prediction,
77 diagnosis, prognosis, and treatment response monitoring [19]. Wang et al. construct an AI-based
78 model of two-dimensional shear wave elastography of the liver and spleen to precisely assess the risk
79 of GEV and high-risk gastroesophageal varices [20]. Ma et al. constructed the deep learning model

80 for screening precancerous lesions of gastric cancer based on tongue images [21]. Li et al. found that
81 both tongue images and the tongue-coating microbiome can be used as tools for the diagnosis of
82 gastric cancer [22]. However, the value of these features in pre-endoscopic PLGC risk screening
83 remains uncertain, and integrating these features based on AI to achieve a more refined PLGC risk
84 screening still poses significant challenges.

85 In this study, we integrated tongue image data and inquiry data to construct an AI-based multimodal
86 model to assist in pre-endoscopic screening of PLGC. We evaluated the performance of predicting
87 PLGC and high-risk PLGC in a cohort of patients diagnosed with chronic gastritis in Fujian, China,
88 and assessed the superiority of multimodal fusion over single modality. Additionally, we validated
89 the model's performance in another independent cohort.

90 **2 Materials and methods**

91 **2.1 Design and overview**

92 This research recruited a cohort of patients diagnosed with chronic gastritis from Fujian, China, and
93 the recruitment period spanned from December 2022 to December 2023. Those who volunteered to
94 participate were included in this study. This study was approved by the ethics committee of the
95 Fujian Medical University (Approval number 58 in 2020).

96 **2.2 Patient enrollment**

97 One thousand and thirty-four potentially eligible patients were enrolled in this study. The inclusion
98 criteria were as follows: 1) age between 18 and 70 years; 2) have the gastroscopy examination results
99 saved within the past three months, or will undergo gastroscopy examination in the coming three
100 months; 3) no previous diagnosis of cancer; 4) resides locally and is willing to cooperate with
101 doctor's follow-up; and 5) written informed consent. Patients were excluded for the following
102 reasons: 1) cancer patients; 2) contraindications for endoscopic examination; 3) pregnant or women
103 planning pregnancy, as well as lactating women; 4) cardiovascular, pulmonary, renal, endocrine,
104 neurological, and hematological disorders; 5) mental disorder; and 6). unwilling to participate or poor
105 compliance.

106 **2.3 Tongue images and inquiry information**

107 It is recommended that patients adhere to a standardized procedure to acquire high-quality tongue
108 images. Patients are advised to present themselves in natural light conditions during the morning,
109 ensuring an empty stomach. Patients should protrude their tongue from the oral cavity, with
110 particular attention to positioning the tip slightly downward and flattening the surface to ensure the
111 entire tongue body is adequately visualized.

112 Simultaneously, the healthcare practitioner will engage in a comprehensive traditional Chinese
113 medicine consultation with patients. This consultation encompasses an exploration of demographic
114 details, such as gender, along with an assessment of pertinent lifestyle factors, including a history of
115 smoking and alcohol consumption. Additionally, an inquiry into the patient's family medical history,
116 dietary habits, and an evaluation of physical symptoms will be conducted. The physical symptoms
117 evaluation involves an assessment of potential discomfort in the stomach and mouth, the patient's
118 mental state, and their bowel and urinary habits.

119 **2.4 Endoscopic evaluation**

120 Two independent gastroenterology experts, each of whom had carried out more than 1000
121 endoscopies, performed esophagogastroduodenoscopy (EGD) on all patients. The biopsy results were
122 reported as normal, superficial gastritis, chronic atrophic gastritis, intestinal metaplasia, or
123 intraepithelial neoplasia, and a diagnosis was assigned to each participant based on the most severe
124 histological finding in the biopsy. The *Helicobacter pylori* (Hp) infection status was determined by
125 enzyme-linked immunosorbent assay of plasma IgG [23]. The procedure was conducted up to 3
126 months before or after the acquisition of images and traditional Chinese medicine inquiry, and the
127 operators were unaware of the results of the tongue examination and inquiry information.

128 **2.5 Single-modality deep Learning Risk Prediction Models**

129 **2.5.1 Tongue images deep learning risk prediction (Single-Tongue) model**

130 This section proposes Single-Tongue, a new diagnostic approach based on single-modality deep
131 learning using tongue images to predict PLGC and high-risk PLGC. We applied the Segment
132 Anything Models to segment tongue images to extract the features of the effective area and avoid the
133 influence of irrelevant edge noise information.

134 All patients were randomly divided into training and validation cohorts. The training cohort was
135 utilized to train a deep neural network designed for this study. The performance of the trained model
136 was evaluated through its application to the validation cohort. In order to extract the features from the
137 tongue images, we employed a pre-trained ResNet framework that had been previously trained on the
138 ImageNet dataset [24]. Distinct from convolutional neural networks (CNNs), ResNet tackles the
139 issues of vanishing gradients and network degradation by introducing direct skip connections within
140 the network, which retain a certain proportion of the output from the previous network layer. Data
141 augmentation techniques such as random cropping, flipping, and rotation were applied to all image
142 data to mitigate overfitting. The images were passed through ResNet during training, specifically
143 through the bottleneck and residual units. After passing through 12 bottleneck layers, an adaptive
144 average pooling operator was used to obtain image features, which were then flattened to a size of
145 2048×1 . The final classification results were generated through a softmax layer. In the single-
146 modality experiment, two binary classification networks were trained to determine the presence or
147 absence of PLGC or high-risk PLGC.

148 **2.5.2 Inquiry information deep learning risk prediction (Single-Inquiry) model**

149 The inquiry information encompasses variables such as sex, age, and the individuals' history of
150 smoking and alcohol consumption. Furthermore, it incorporates the family medical history, dietary
151 habits, and physical symptoms of the patients, including discomfort in the stomach and mouth and
152 their mental state. We employed a segregation approach by categorizing the features into numerical
153 and factor types to enhance the effectiveness of utilizing the inquiry information. The numerical
154 features were subjected to min-max normalization, scaling them between 0 and 1. On the other hand,
155 the factor features were transformed into numeric vectors using keyword-based encoding techniques.

156 After feature filtering and mapping, the inquiry information was input into a multilayer perceptron to
157 obtain corresponding feature vectors. Then, two binary classification networks were trained to
158 determine the presence or absence of PLGC or high-risk PLGC.

159 **2.6 Multimodality deep learning risk prediction (AITonguequiry) model**

160 Medical data are frequently multimodal. For instance, both tongue images and inquiry information
161 encompass details associated with PLGC. Consequently, in this section, we integrated these two
162 modalities with an attentional mechanism.

163 In the multimodality experiment, similar to the single-modality experiment, we trained two binary
164 classification networks to determine the presence or absence of PLGC or high-risk PLGC. We
165 employed the dropout method to eliminate a certain proportion of model parameters. Subsequently,
166 we utilized the feature embedding method to align the feature vectors of tongue images and inquiry
167 information for comprehensive patient information utilization.

168 **2.7 Statistical analysis**

169 The prediction results were validated by quantitative indexes, including sensitivity, specificity,
170 positive predictive value and negative predictive value. The chi-squared test and t test were used to
171 determine whether there was any significant difference in patient characteristics. The area under the
172 receiver operating characteristic (ROC) curve (AUC) was used to estimate the probability that the
173 model would produce a correct prediction. The DeLong test was used to test whether there was a
174 significant difference in risk prediction between AITonguequiry and other methods.

175 **3 Results**

176 **3.1 Patient characteristics**

177 In this research, a cohort of 1034 participants was recruited in Fujian, China, and the recruitment
178 period spanned from December 2022 to December 2023. Among these patients, NPLGC was
179 documented in 855 (82.61%) patients, and PLGC was documented in 180 (17.39%) patients. Among
180 PLGC, low-risk PLGC and high-risk PLGC account for 346(65.90%) and 179(34.10%), respectively.
181 After randomization of these patients, 828 patients were assigned to the training cohort. The other
182 207 patients composed the validation cohort.

183 The baseline characteristics of the study population are summarized in Table 1 and Table 2. Overall,
184 the average age of patients with PLGC was 62, and the standard deviation was 7. In terms of gender,
185 the proportion of males with PLGC (180[34.29%]), and females with PLGC (345[65.71%]). Between
186 the training and validation cohorts, there were no significant differences in any of the baseline
187 characteristics ($p>0.05$) or in the distribution of patients between NPLGC and PLGC.

188 **3.2 Construction of the AITonguequiry model**

189 We build a deep learning risk prediction model based on tongue images and inquiry information
190 (AITonguequiry) from 1034 patients to evaluate their potential in the grading screening and
191 diagnosis of PLGC. The AITonguequiry flow chart is shown in Figure 1. As shown in Figure 1a, the
192 study cohort and the diagnostic model were designed to assess the risk of PLGC and high-risk PLGC
193 based on tongue images and inquiry information. The detailed multimodality model is shown in
194 Figure 1b. The patients were categorized into two groups: Non-PLGC (NPLGC) and PLGC, and the
195 PLGC group was further divided into low-risk PLGC (chronic atrophic gastritis) and high-risk PLGC
196 (intestinal metaplasia or gastric epithelial dysplasia) stages. We advocate for patients predicted with
197 PLGC to undergo endoscopic examination, with a concurrently recommending prompt endoscopic
198 examination for those predicted with high-risk PLGC.

199 **3.3 Comparison of the AITonguequiry model, single modality models and baseline** 200 **characteristics**

201 We chose Single-Tongue, Single-Inquiry and baseline characteristics to identify the presence or
202 absence of PLGC and high-risk PLGC.

203 The selection of baseline characteristics is based on individuals aged ≥ 45 who meet any of the
204 following criteria, which are indicative of a high-risk profile for gastric cancer: 1) Long-term
205 residence in high-incidence areas of gastric cancer; 2) Hp infection; 3) History of chronic atrophic
206 gastritis, gastric ulcer, gastric polyp, residual stomach after surgery, hypertrophic gastritis, pernicious
207 anemia, or other precancerous diseases of the stomach; 4) First-degree relatives with a history of
208 gastric cancer; 5) Presence of other high-risk factors for gastric cancer such as high salt intake,
209 pickled diet, smoking, and heavy alcohol consumption [25-28]. Since our screening is conducted in
210 high-risk areas, we ignore the first criterion.

211 In identifying the presence or absence of PLGC, AITonguequiry demonstrated statistically higher
212 AUCs than Single-Tongue, Single-Inquiry and baseline characteristics ($p < 0.05$) (Figure 2a, Table
213 3). Impressively, AITonguequiry had an AUC of 0.736 for the diagnosis of PLGC, which was higher
214 than other methods (Figure 2a, all $p < 0.05$). The sensitivity and specificity analyses also
215 demonstrated that AITonguequiry universally outperformed the Single-Tongue, Single-Inquiry and
216 baseline characteristics for assessing PLGC and high-risk PLGC (Table 3).

217 In identifying the presence or absence of high-risk PLGC, AITonguequiry demonstrated statistically
218 higher AUCs than Single-Tongue, Single-Inquiry and baseline characteristics ($p < 0.05$) (Figure 2b,
219 Table 3). Impressively, AITonguequiry had an AUC of 0.816 for the diagnosis of high-risk PLGC,
220 which was higher than other methods (Figure 2b, all $p < 0.05$). The sensitivity and specificity analyses
221 also demonstrated that AITonguequiry universally outperformed the Single-Tongue, Single-Inquiry
222 and baseline characteristics for assessing PLGC and high-risk PLGC (Table 3).

223 **3.4 Evaluation of the diagnostic robustness of the AITonguequiry model**

224 All patients were randomly divided into training and validation cohorts. Simultaneously, all the
225 images of each patient were allocated to the cohort corresponding to that patient, ensuring that there
226 was no simultaneous presence of different images from the same patient in both cohorts. In either the
227 validation cohort, the results in the three ROC curves always overlapped each other (Figure 3), and
228 no significant differences were found (all $p > 0.05$, Table 4). These results revealed that
229 AITonguequiry demonstrated robust and consistent performances regardless of the data from which
230 medical centers, as long as the number of enrolled patients in different training cohorts was fairly
231 constant.

232 **3.5 Independent external validation of the AITonguequiry model**

233 Moreover, 143 participants were recruited for independent external validation from the China-Japan
234 Friendship Hospital. In the independent validation cohort, the AUC of PLGC was 0.69, and the AUC
235 of high-risk PLGC was 0.77 (Figure 4, Table 5). These results demonstrate the effectiveness of the
236 AITonguequiry model in independent external validation.

237

238

239 4 Discussion

240 Given the significant burden of GC in China and globally, it becomes imperative to adopt cost-
241 effective approaches for large-scale screening of PLGC risk prediction in the natural population.
242 While gastroscopy and pathological tests serve as the gold standard for diagnosing gastric diseases,
243 they are not suitable for widespread application in the natural population. To address this pressing
244 issue, there is an urgent need to develop noninvasive and effective screening and diagnostic methods
245 to provide tailored recommendations for endoscopy. Deep neural network technology offers a
246 promising avenue for advancing healthcare systems by providing heightened accuracy and
247 computational power, thereby playing an increasingly vital role in disease risk prediction. In our
248 study, a multimodality AI-assisted pre-endoscopic screening model based on tongue images and
249 inquiry information (AITonguequiry) was constructed for the first time, adopting a hierarchical
250 prediction strategy, achieving tailored endoscopic recommendations.

251 In this study, the diagnostic accuracy of AITonguequiry was significantly better than that of Single-
252 Tongue or Single-Inquiry in assessing PLGC and high-risk PLGC. In the evaluation of PLGC, the
253 AUC in the validation cohort was 0.74 (Figure 2a, Table 3). Thus, AITonguequiry was effective for
254 the assessment of PLGC. In the evaluation of high-risk PLGC, the AUC in the verification cohort
255 was 0.82, showing that AITonguequiry was effective for the assessment of high-risk PLGC (Figure
256 2b, Table 3). The above analysis revealed that the multimodality models were effective in
257 discriminating patients with PLGC from participants without PLGC, and could also effectively
258 differentiate patients with high-risk PLGC from those with low-risk PLGC. Thus, the fusion model of
259 tongue images and inquiry information further improved the diagnostic value (Figure 2, Table 3).

260 The risk prediction of PLGC using tongue images and inquiry information can serve as an effective,
261 noninvasive auxiliary diagnostic method that can support primary healthcare systems worldwide.
262 With the recent advancements in deep neural networks, significant progress has been made in
263 standardizing tongue images and inquiry information risk prediction, especially in TCM. Many
264 results have been achieved in tongue image preprocessing, tongue detection, segmentation, feature
265 extraction and tongue analysis [28]. Xu et al. developed a multi-task joint learning model to segment
266 and classify tongue images using deep neural networks, which can optimally extract tongue image
267 features [29]. Li et al. pioneered the use of both tongue images and the tongue-coating microbiome as
268 diagnostic tools for gastric cancer [22]. Similarly, Ma et al. constructed the first deep learning model
269 for screening precancerous lesions of gastric cancer based on tongue images [21]. To our knowledge,
270 AITonguequiry is the first simplified, novel, AI-based image processing tool that can accurately and
271 noninvasively identify PLGC and high-risk PLGC. Moreover, our AITonguequiry multimodal
272 model, for the first time, incorporates inquiry information and employs a hierarchical prediction
273 strategy, resulting in more refined endoscopic recommendations. We advocate for patients predicted
274 with PLGC to undergo endoscopic examination, with a concurrently recommending prompt
275 endoscopic examination for those predicted with high-risk PLGC. With the AITonguequiry, the
276 operator only needs to carry out the daily data acquisition workflow to analyze the critical
277 information automatically, making this model very convenient for clinical applications, which
278 enhances patient screening efficiency and alleviates patient burden.

279 According to the principles of TCM, tongue manifestation serves as a crucial indication of
280 "collaterals" [30]. It plays a significant role in various disorders such as gastric diseases [21],
281 rheumatic conditions [31], hepatic disorders [32], and tumors [10], among others. Tongue color,
282 thickness, moisture, and other changes, as well as the rich network of vessels in the tongue area [30],
283 to a certain extent, can reflect the pathological changes in the spleen, stomach, internal organs, joints,

284 and many other parts. Tongue diagnosis and inquiry are integral components of the four diagnostic
285 methods in TCM. The characteristics of the tongue's shape, color, size, and coating, as detected
286 through tongue images, can reflect an individual's health status and the severity/progression of
287 diseases [33]. In addition, inquiry information, including demographics, behavior, and physical
288 symptoms, plays a crucial role in understanding the disease and medical history. In this study, we
289 discovered that useful features for predicting PLGC can be extracted from tongue images using a
290 deep learning model. This finding indicates that the evaluation of human health based on tongue
291 characteristics, as proposed by TCM theory, has scientific grounds. Furthermore, we found that the
292 deep-learning model can extract informative features for PLGC prediction from inquiry information,
293 suggesting that the evaluation of human health based on demographic, behavioral, and physical
294 symptom information in inquiries, as guided by TCM theory, is scientifically supported. We believe
295 that with the widespread application of AI and deep learning methods, tongue images and inquiry
296 information can potentially become cost-effective, non-invasive and acceptable approaches for
297 predicting and screening PLGC, which will also lead to significant socio-economic impacts.

298 Nevertheless, our research has some limitations. One limitation of our study is the restricted number
299 of patients included. For future research, it will be essential to involve a larger screening population
300 to improve the training of the deep learning model. Additionally, there is potential to expand by
301 creating a semantic dataset of tongue images and establishing a multimodal large language model for
302 PLGC prediction, offering personalized predictions and detailed explanations for clinicians and
303 patients. These areas merit further exploration in the future.

304 In conclusion, this study introduces a hierarchical AI-based multimodal non-invasive method for pre-
305 endoscopic risk screening. All these findings indicate that AITonguequiry is a non-invasive method
306 for predicting and assessing PLGC and high-risk PLGC, showcasing its strong performance in graded
307 screening and diagnosis. Our method has a good potential for widespread clinical use, and further
308 studies in larger patient populations are needed. We will further promote the application of
309 AITonguequiry in PLGC risk prediction and provide tailored recommendations for endoscopy to
310 improve the diagnosis rate, especially high-risk PLGC, in the natural population. Moreover, this
311 study provides scientific support for the theory of tongue images and inquiry information diagnosis in
312 TCM.

313 **5 References**

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 71(3):209-249. doi: 10.3322/caac.21660.
2. Zong L, Abe M, Seto Y, Ji J. (2016) The challenge of screening for early gastric cancer in China. *Lancet.* 388(10060):2606. doi: 10.1016/S0140-6736(16)32226-7.
3. Rawla P, Barsouk A. (2019) Epidemiology of gastric cancer: global trends, risk factors and prevention. *Prz Gastroenterol.* 14(1):26-38. doi: 10.5114/pg.2018.80001.
4. de Vries AC, van Grieken NC, Looman CW, Casparie MK, de Vries E, Meijer GA, et al. (2008) Gastric cancer risk in patients with premalignant gastric lesions: a nationwide cohort study in the Netherlands. *Gastroenterology.* 134(4):945-52. doi: 10.1053/j.gastro.2008.01.071.
5. Thrift AP, El-Serag HB. (2020) Burden of Gastric Cancer. *Clin Gastroenterol Hepatol.* 18(3):534-542. doi: 10.1016/j.cgh.2019.07.045.
6. Zhang P, Yang M, Zhang Y, Xiao S, Lai X, Tan A, Du S, Li S. (2019) Dissecting the Single-Cell Transcriptome Network Underlying Gastric Premalignant Lesions and Early Gastric Cancer. *Cell Rep.* 27(6): 1934-1947. e5. doi: 10.1016/j.celrep. 2019. 04. 052.
7. Huang S, Guo Y, Li ZW, Shui G, Tian H, Li BW, et al. (2021) Identification and Validation of Plasma Metabolomic Signatures in Precancerous Gastric Lesions That Progress to Cancer. *JAMA Netw Open.* 4(6): e2114186. doi: 10.1001/ jamanetworkopen. 2021.14186.
8. Cubiella J, Pérez Aisa Á, Cuatrecasas M, Díez Redondo P, Fernández Esparrach G, Marín-Gabriel JC, et al. (2021) Gastric cancer screening in low incidence populations: Position statement of AEG, SEED and SEAP. *Gastroenterol Hepatol.* 44(1):67-86. Spanish. doi:10.1016/j.gastrohep.2020.08.004.
9. Gholami E, Tabbakh S, kheirabadi M. (2021) Increasing the accuracy in the diagnosis of stomach cancer based on color and lint features of tongue. *Biomed. Signal Process. Control.* 69, 102782. doi:10.1016/j.bspc.2021.102782
10. Zhu X, Ma Y, Guo D, Men J, Xue C, Cao X, et al. (2022) A Framework to Predict Gastric Cancer Based on Tongue Features and Deep Learning. *Micromachines (Basel).* 14(1):53. doi: 10.3390/mi14010053.
11. Shang Z, Du ZG, Guan B, Ji XY, Chen LC, Wang YJ, Ma Y. (2022) Correlation analysis between characteristics under gastroscopy and image information of tongue in patients with chronic gastriti. *J Tradit Chin Med.* 42(1):102-107. doi: 10.19852/ j.cnki. jtc.2022.01.006.
12. Kainuma M, Furusyo N, Urita Y, Nagata M, Ihara T, Oji T, et al. (2015) The association between objective tongue color and endoscopic findings: results from the Kyushu and Okinawa population study (KOPS). *BMC Complement Altern Med.* 15:372. doi: 10.1186/s12906-015-0904-0.

13. Hou C, Cui Y, Xu Y, Wang Y, Hao Y. (2022) TCM Syndrome Recognition Model of Type 2 Diabetes Mellitus in Shanghai Based on TCM Inquiry Information. *Evid Based Complement Alternat Med.* 2022:2843218. doi: 10.1155/2022/2843218.
14. Tan MC, Sen A, Kligman E, Othman MO, Liu Y, El-Serag HB, et al. (2023) Validation of a pre-endoscopy risk score for predicting the presence of gastric intestinal metaplasia in a U.S. population. *Gastrointest Endosc.* 98(4):569-576.e1. doi: 10.1016/j.gie.2023.05.048.
15. Takenaka K, Ohtsuka K, Fujii T, Negi M, Suzuki K, Shimizu H, et al. (2020) Development and Validation of a Deep Neural Network for Accurate Evaluation of Endoscopic Images From Patients With Ulcerative Colitis. *Gastroenterology.* 158(8):2150-2157. doi: 10.1053/j.gastro.2020.02.012.
16. Yu G, Sun K, Xu C, Shi XH, Wu C, Xie T, Meng RQ, Meng XH, Wang KS, Xiao HM, Deng HW. (2021) Accurate recognition of colorectal cancer with semi-supervised deep learning on pathological images. *Nat Commun.* 12(1):6311. doi: 10.1038/s41467-021-26643-8.
17. Cheung CY, Xu D, Cheng CY, et al. (2021) A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre. *Nat Biomed Eng.* 5:498–508. doi: 10.1038/s41551-020-00626-4.
18. Zhang P, Wang B, Li S. (2023) Network-based cancer precision prevention with artificial intelligence and multi-omics. *Sci Bull (Beijing).* 68(12):1219-1222. doi: 10.1016/j.scib.2023.05.023.
19. Pan F, Feng L, Liu B, Hu Y, Wang Q. (2023) Application of radiomics in diagnosis and treatment of lung cancer. *Front Pharmacol.* 14:1295511. doi: 10.3389/fphar.2023.1295511.
20. Wang L, He R, et.al. (2023) Deep learning radiomics for assessment of gastroesophageal varices in people with compensated advanced chronic liver disease. *arXiv preprint arXiv:2306.07505.*
21. Ma C, Zhang P, Du S, Li Y, Li S. (2023) Construction of Tongue Image-Based Machine Learning Model for Screening Patients with Gastric Precancerous Lesions. *J Pers Med.* 13(2):271. doi: 10.3390/jpm13020271.
22. Li Y. and Lin Y. et.al. (2023) Development of a tongue image-based machine learning tool for the diagnosis of gastric cancer: a prospective multicentre clinical cohort study. *eClinicalMedicine.* 57:101834. doi: 10.1016/j.eclinm.2023.101834.
23. Li S, Lu AP, Zhang L, Li YD. (2003) 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* , 9(4), 755–758. doi:10.3748/wjg.v9.i4.755
24. He K, Zhang X, Ren S, et al. (2016) Deep residual learning for image recognition. *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* 770-778. doi: 10.1109/CVPR.2016.90.
25. He J, Chen WQ, Li ZS, et al. (2022) China guideline for the screening, early detection and early treatment of esophageal cancer (2022, Beijing). *Zhong hua Zhong Liu Za Zhi.* 44(6):491-522. Chinese. doi: 10.3760/cma.j.cn112152-20220517-00348.

26. Zhang Q, Yang M, Zhang P, Wu B, Wei X, Li S. (2023) Deciphering gastric inflammation-induced tumorigenesis through multi-omics data and AI methods. *Cancer Biol Med.* j.issn.2095-3941. 2023. doi: 10.20892/j.issn.2095-3941.2023.0129.
27. Thrift AP, Wenker TN, El-Serag HB. (2023) Global burden of gastric cancer: epidemiological trends, risk factors, screening and prevention. *Nat Rev Clin Oncol.* 20(5):338-349. doi: 10.1038/s41571-023-00747-0.
28. Li J, Zhang Z, Zhu X, Zhao Y, Ma Y, Zang J, Li B, Cao X, Xue C. (2022) Automatic Classification Framework of Tongue Feature Based on Convolutional Neural Networks. *Micromachines (Basel).* 13(4):501. doi: 10.3390/mi13040501.
29. Tania MH, Lwin K, Hossain MA. (2019) Advances in automated tongue diagnosis techniques. *Integr Med Res.* 8(1):42-56. doi: 10.1016/j.imr.2018.03.001.
30. 李梢 (2002). 王永炎院士从“络”辨治痹病学术思想举隅. *北京中医药大学学报.* 25(1):43-45.
31. Gualtierotti R, Marzano A V, Spadari F, et al. (2018). Main oral manifestations in immune-mediated and inflammatory rheumatic diseases. *Journal of Clinical Medicine,* 8(1): 21. doi: 10.3390/jcm8010021.
32. Yao Y, Habib M, Bajwa HF, et al. (2024). Sympathetic circuits regulating hepatic glucose metabolism: where we stand. *Physiological reviews,* 104(1), 85–101. doi: 10.1152/physrev.00005.2023.
33. Anastasi JK, Chang M, Quinn J, Capili B. (2014) Tongue Inspection in TCM: Observations in a Study Sample of Patients Living with HIV. *Med Acupunct.* 26(1):15-22. doi: 10.1089/acu.2013.1011.

314

315 **Author contributions**

316 Study conception and design: Shao Li, Peng Zhang, Lan Wang, Qian Zhang and Kaiqiang Tang;

317 Data collection and analysis: Shao Li, Lan Wang, Qian Zhang, Bowen Wu, Shiyu Du and Kaiqiang
318 Tang;

319 The first draft of the manuscript: Lan Wang, Kaiqiang Tang, Qian Zhang;

320 Commented on previous versions of the manuscript: Shao Li, Peng Zhang.

321 All authors read and approved the final manuscript.

322

323 **Compliance with Ethical Requirements**

324 (1) Conflict of Interest (CoI) statements

325 Lan Wang, Peng Zhang, Qian Zhang, Bowen Wu, Shiyu Du, Kaiqiang Tang and Shao Li declared no
326 conflicts of interest related to this study.

327 (2) Informed Consent in Studies with Human Subjects

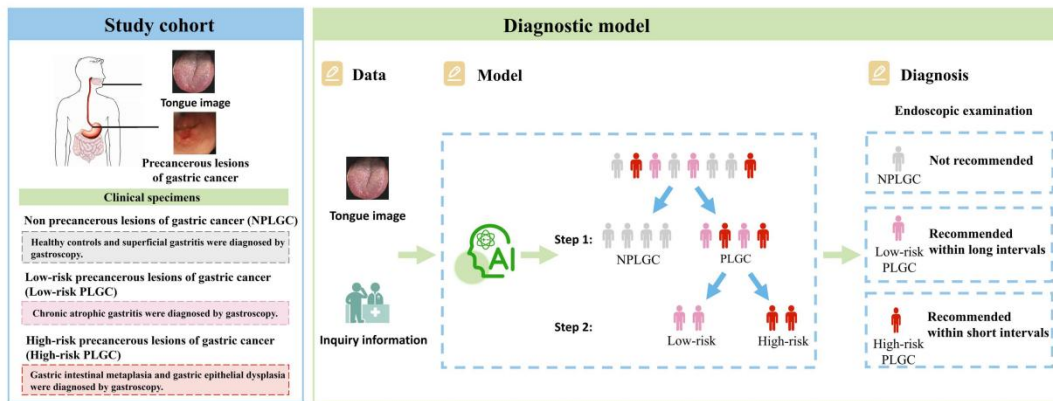
328 All procedures followed were in accordance with the ethical standards of the responsible committee
329 on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as
330 revised in 2008 (5). Informed consent was obtained from all patients for being included in the study.

331 (3) Data availability

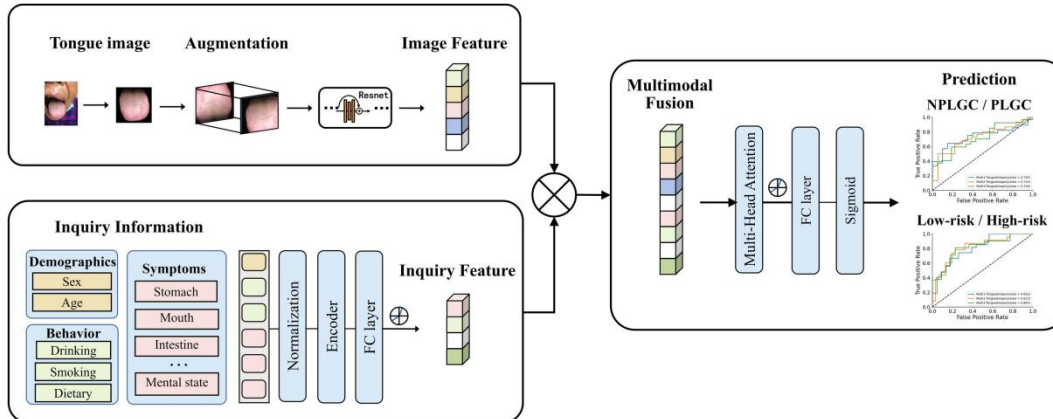
332 The data of individual deidentified participants will not be shared, but it is available upon request via
333 email: shaoli@mail.tsinghua.edu.cn.

334 **Figure 1.** AITonguequiry flow chart. (a) The study cohort and the diagnostic model. (b) A deep
 335 learning-based multimodality classification model was developed to assess the risk of PLGC and high-
 336 risk PLGC based on the tongue images and inquiry information.

a



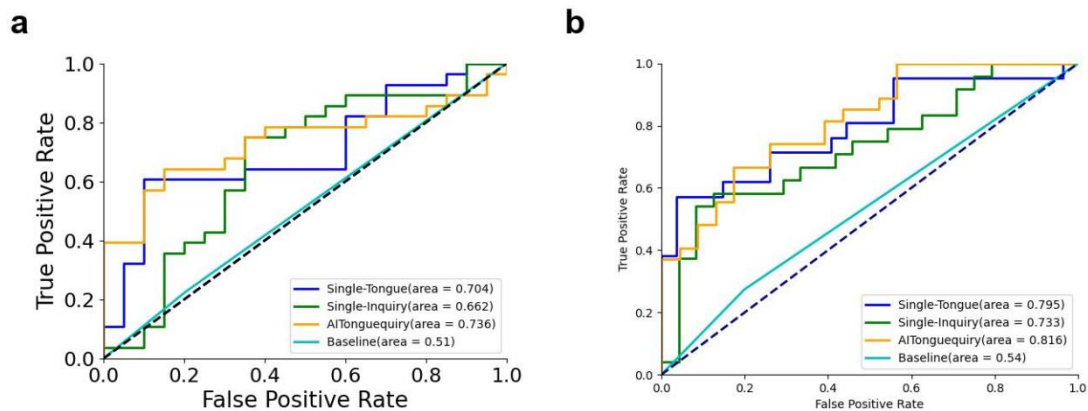
b



337

338 PLGC, precursor lesions of gastric cancer; NPLGC, Non-precursor lesions of gastric cancer.

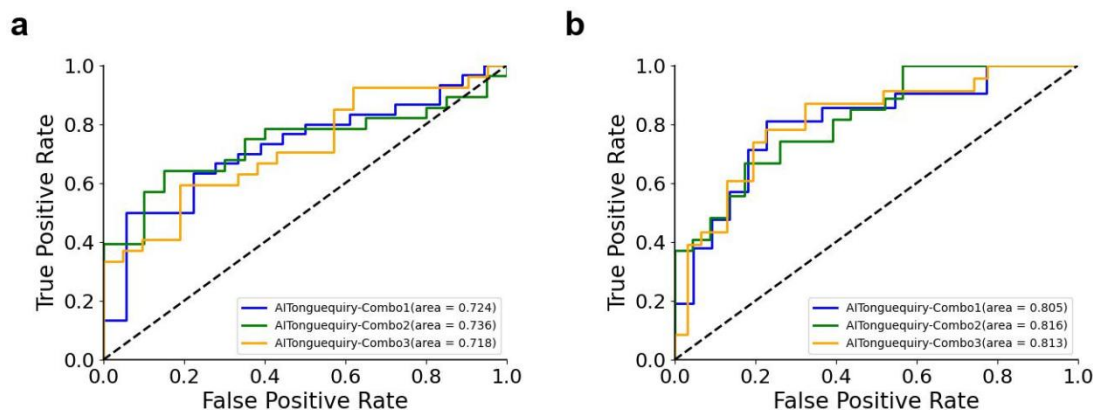
339 **Figure 2.** Comparison of ROC curves between different methods for classifying the presence or
340 absence of PLGC and high-risk PLGC in the validation cohorts. (a) Presence or absence of
341 NPLGC/PLGC in the validation cohorts. (b) Presence or absence of low-risk PLGC/high-risk PLGC
342 in validation cohorts.



343

344 PLGC, precursor lesions of gastric cancer; NPLGC, Non-precursor lesions of gastric cancer; Single-
345 Tongue, single-modality deep learning risk prediction with tongue images; Single-Inquiry, single-
346 modality deep learning risk prediction with inquiry information; AITonguequiry, multimodality deep
347 learning risk prediction with tongue images and inquiry information.

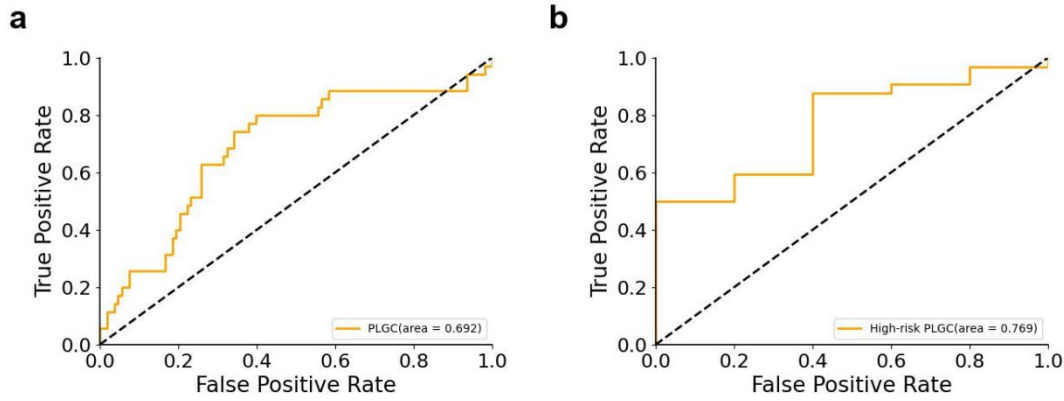
348 **Figure 3.** Comparison of receiver operating characteristic (ROC) curves among different combinations
349 using AITonguequiry. (a) Presence or absence of NPLGC/PLGC in the validation cohorts. (b) Presence
350 or absence of low-risk PLGC/high-risk PLGC in validation cohorts.



351

352 PLGC, precursor lesions of gastric cancer; NPLGC, Non-precursor lesions of gastric cancer;
353 AITonguequiry, multimodality deep learning risk prediction with tongue images and inquiry
354 information.

355 **Figure 4.** Receiver operating characteristic (ROC) curves during the external validation using Single-
356 Tongue. Presence or absence of NPLGC/PLGC in the external validation cohorts and presence or
357 absence of low-risk PLGC/high-risk PLGC in the external validation cohorts.



358
359 PLGC, precursor lesions of gastric cancer; NPLGC, Non-precursor lesions of gastric cancer.
360

361 **Table 1.** Baseline characteristics of the study cohort: NPLGC and PLGC.

Characteristic	Training and validation dataset ¹		p-value ²
	NPLGC, N = 509	PLGC, N = 525	
Age (Years)	59 (9)	61 (8)	<0.001
Sex (Female/Male)	343 /166	345 /180	0.57
Family history (Yes/No)	39/470	47/478	0.45
Drinking (Yes/No)	136/373	144/381	0.80
Smoking (Yes/No)	141/368	153/372	0.61
Hp (Positive/Negative/Unknown)	14/31/464	19/32/474	0.73

362 ¹Mean (SD); n (%)

363 ²Welch Two Sample t-test; Pearson's Chi-squared test; Fisher's exact test

364 PLGC, precursor lesions of gastric cancer; NPLGC, Non-precursor lesions of gastric cancer; Hp,
365 helicobacter pylori.

366 **Table 2.** Baseline characteristics of the study cohort: low-risk PLGC and high-risk PLGC.

Characteristic	Training and validation dataset ¹		p-value ²
	Low-risk PLGC, N = 346	High-risk PLGC, N = 179	
Age (Years)	61 (8)	62 (7)	0.45
Sex (Female/Male)	240/106	105/74	0.014
Family history (Yes/No)	27/319	20/159	0.20
Drinking (Yes/No)	77/269	67/112	<0.001
Smoking (Yes/No)	86/260	67/112	0.003
Hp (Positive/Negative/Unknown)	6/19/321	13/13/153	0.004

367 ¹Mean (SD); n (%)

368 ²Welch Two Sample t-test; Pearson's Chi-squared test; Fisher's exact test

369 PLGC, precursor lesions of gastric cancer; NPLGC, Non-precursor lesions of gastric cancer; Hp,
370 helicobacter pylori.

371 **Table 3.** Diagnostic performance of AITonguequiry for the assessment of NPLGC/PLGC and low-risk
 372 PLGC/high-risk PLGC in the validation cohorts.

		AUC	Specificity (%)	Sensitivity (%)	PPV (%)	NPV (%)
Step 1:	Single-Tongue	0.70*	54.55	86.67	46.43	90.00
NPLGC		(0.68 to 0.73)	(53.14 to 56.02)	(84.23 to 89.18)	(43.65 to 49.31)	(87.99 to 91.96)
/PLGC	Single-Inquiry	0.66*	62.50	68.75	78.57	50.00
		(0.64 to 0.69)	(59.63 to 65.58)	(67.13 to 70.32)	(76.33 to 80.96)	(46.58 to 53.42)
	AITonguequiry	0.74	66.67	73.33	78.57	60.00
		(0.71 to 0.76)	(63.94 to 69.43)	(71.59 to 75.04)	(76.16 to 80.87)	(56.66 to 63.27)
Step 2:						
Low-risk	Single-Tongue	0.79*	90.00	52.63	95.24	33.33
PLGC		(0.78 to 0.82)	(86.97 to 92.58)	(51.54 to 53.69)	(93.60 to 96.57)	(30.58 to 36.00)
/High-risk	Single-Inquiry	0.73*	64.00	65.22	62.50	66.67
PLGC		(0.71 to 0.75)	(61.91 to 66.18)	(63.03 to 67.39)	(59.50 to 65.60)	(63.80 to 69.50)
	AITonguequiry	0.82	66.67	78.26	66.67	78.26
		(0.80 to 0.83)	(64.61 to 68.66)	(76.09 to 80.44)	(63.79 to 69.44)	(75.52 to 80.85)

373

374 The AUC of AITonguequiry was statistically compared with the AUCs of Single-Tongue and Single-
 375 Inquiry (*P<0.05; **P<0.01).

376 AUC, area under the receiver operating characteristic curve; PLGC, precursor lesions of gastric cancer;
 377 NPLGC, Non-precursor lesions of gastric cancer; Single-Tongue, single-modality deep learning risk
 378 prediction with tongue images; Single-Inquiry, single-modality deep learning risk prediction with
 379 inquiry information; AITonguequiry, multimodality deep learning risk prediction with tongue images
 380 and inquiry information; PPV, positive predictive value; NPV, negative predictive value.

381 **Table 4.** Diagnostic robustness of AITonguequiry for the assessment of NPLGC/PLGC and low-risk
 382 PLGC/high-risk PLGC in the validation cohorts.

	AUC	Specificity (%)	Sensitivity (%)	PPV (%)	NPV (%)
Step1:					
NPLGC/PLGC					
Combo 1	0.74 (0.71 to 0.76)	66.67 (63.91 to 69.44)	73.33 (71.62 to 75.06)	78.57 (76.16 to 80.88)	60.00 (56.90 to 63.27)
Combo 2	0.72 (0.70 to 0.75)	54.55 (52.21 to 56.88)	76.92 (75.00 to 78.79)	66.67 (64.00 to 69.20)	66.67 (63.33 to 70.00)
Combo 3	0.72 (0.69 to 0.74)	60.00 (57.32 to 62.52)	67.86 (65.96 to 69.73)	70.37 (67.73 to 73.07)	57.14 (53.83 to 60.46)
Step2:					
Low-risk PLGC /High-risk PLGC					
Combo 1	0.82 (0.80 to 0.83)	66.67 (64.63 to 68.66)	78.26 (75.95 to 80.51)	66.67 (63.80 to 69.35)	78.26 (75.30 to 80.96)
Combo 2	0.81 (0.79 to 0.82)	73.91 (72.00 to 76.09)	75.00 (72.80 to 77.17)	71.43 (68.65 to 74.39)	77.27 (74.78 to 79.77)
Combo 3	0.81 (0.79 to 0.83)	86.36 (84.22 to 88.53)	62.50 (60.66 to 64.32)	86.96 (84.61 to 89.31)	61.29 (58.45 to 64.11)

383 Statistical values are presented with the 95% CIs when applicable.
 384 AUCs obtained with three different combinations of patients were statistically compared with each
 385 other in each classification and each cohort (*P<0.05; **P<0.01).
 386 AUC, area under the receiver operating characteristic curve; PLGC, precursor lesions of gastric cancer;
 387 Combo, combination of patients for training and validation cohorts; PPV, positive predictive value;
 388 NPV, negative predictive value.

389 **Table 5.** Diagnostic performance of Single-Tongue for the assessment of NPLGC/PLGC and low-risk
390 PLGC/high-risk PLGC in the external validation cohorts.

	AUC	Specificity (%)	Sensitivity (%)	PPV (%)	NPV (%)
Step 1:					
NPLGC/PLGC	0.69 (0.66 to 0.72)	80.56 (79.36 to 81.73)	40.00 (36.36 to 43.56)	40.00 (35.58 to 44.18)	80.56 (78.61 to 82.52)
Step 2:					
Low-risk PLGC	0.77	33.33	88.24	93.75	20.00
/High-risk PLGC	(0.74 to 0.79)	(26.49 to 40.00)	(87.61 to 88.90)	(92.60 to 94.85)	(15.19 to 24.81)

391

392 AUC, area under the receiver operating characteristic curve; PLGC, precursor lesions of gastric
393 cancer; NPLGC, Non-precursor lesions of gastric cancer; Single-Tongue, single-modality deep
394 learning risk prediction with tongue images; Single-Inquiry, single-modality deep learning risk
395 prediction with inquiry information; AITonguequiry, multimodality deep learning risk prediction with
396 tongue images and inquiry information; PPV, positive predictive value; NPV, negative predictive
397 value.