




OPEN ACCESS

# Diagnostic accuracy of tongue coating in identifying acute appendicitis: a prospective cohort study

Hideki Mori ,<sup>1</sup> Kazumi Yamasaki,<sup>2</sup> Yusuke Saishoji,<sup>1</sup> Yuichi Torisu,<sup>1</sup> Takahiro Mori,<sup>1</sup> Yuki Nagai,<sup>1</sup> Yasumori Izumi<sup>1</sup>

Handling editor Shammi L Ramlakhan

<sup>1</sup>Department of General Internal Medicine, NHO Nagasaki Medical Center, Omura, Nagasaki, Japan

<sup>2</sup>Clinical Research Center, NHO Nagasaki Medical Center, Omura, Nagasaki, Japan

## Correspondence to

Dr Hideki Mori;  
mori.hideki.fu@mail.hosp.go.jp

Received 6 May 2024

Accepted 20 March 2025

## ABSTRACT

**Background** Acute appendicitis requires timely diagnosis. The diagnostic efficacy of tongue examination in making this diagnosis has not been established. This study investigates whether the Tongue Coating Index (TCI), a validated measure of tongue coating, can aid in diagnosing acute appendicitis.

**Methods** We conducted a prospective cohort study (1 September 2018–31 December 2020) at a single Japanese hospital. Adults ( $\geq 20$  years) with suspected acute appendicitis, presenting to either the emergency department or general outpatient clinic, were enrolled. Tongue images were taken at presentation; two independent examiners—unrelated to clinical care and blinded to patient data—later scored these images using the TCI. A composite reference standard (clinical findings, imaging, histopathology, follow-up) was used to confirm appendicitis. We compared the TCI's diagnostic performance with the Alvarado score and its components using C-index, area under the curve (AUC), sensitivity and specificity.

**Results** Of 145 included patients, 69 (47.6%) were diagnosed with acute appendicitis. The TCI demonstrated comparable discriminative ability (C-index AUC 0.62; 95% CI, 0.53 to 0.71) to that of the Alvarado score (0.66; 95% CI, 0.57 to 0.75). Of Alvarado score components, migration of pain had an AUC of 0.63 (95% CI, 0.55 to 0.71), anorexia 0.58 (95% CI, 0.50 to 0.66) and tenderness in the right lower quadrant 0.55 (95% CI, 0.50 to 0.60). At a cut-off of 3, the TCI demonstrated high sensitivity of 96% (95% CI, 88% to 98%) but low specificity of 21% (95% CI, 13% to 32%). Conversely, at a cut-off of 10, the TCI showed increased specificity of 83% (95% CI, 73% to 90%) but reduced sensitivity of 29% (95% CI, 20% to 41%).

**Conclusion** The TCI showed comparable diagnostic performance to the Alvarado score and its individual components. TCI may potentially serve as an additional non-invasive indicator for diagnosing or ruling out acute appendicitis. Further research is essential to validate its clinical utility.

## INTRODUCTION

Acute appendicitis is one of the most common abdominal emergencies,<sup>1</sup> with delayed treatment often resulting in serious complications.<sup>2 3</sup> Therefore, early and accurate diagnosis is crucial. However, diagnosing appendicitis based solely on symptoms or physical findings remains challenging.<sup>4</sup> Blood tests and imaging modalities, such as abdominal CT and ultrasonography (US), play a significant role in diagnosis,<sup>5</sup> it is important to optimise their

## WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The Alvarado score is widely used to diagnose acute appendicitis, but its accuracy is moderate and it relies on blood tests. In particular, in settings where laboratory or imaging studies are not readily available, a non-invasive diagnostic approach is valuable. Tongue coating has been reported as a simple indicator of intra-abdominal inflammation. Therefore, if proven useful for diagnosing acute appendicitis, it could serve as a rapid diagnostic adjunct.

## WHAT THIS STUDY ADDS

⇒ In this prospective study of adults with suspected appendicitis, the Tongue Coating Index (TCI) demonstrated diagnostic performance comparable to the Alvarado score, suggesting that the degree of tongue coating may assist in clinical decision-making. While further validation is required, the TCI could offer a rapid, accessible diagnostic adjunct.

## HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE, OR POLICY

⇒ TCI could be integrated into diagnostic protocols and might be most useful in settings where traditional diagnostic methods are less accessible, potentially informing decisions about further investigations.

use due to concerns regarding radiation exposure, cost and time. Furthermore, these tools may not easily be available in low-resource settings<sup>6</sup> and require additional expense and often referrals to higher-level hospitals. In addition, consequently, a diagnostic strategy that may help in early evaluation may provide guidance for further investigation.

Several clinical prediction rules have been developed to aid in diagnosing acute appendicitis.<sup>7</sup> The Alvarado score, published in 1986, is a widely used tool that combines clinical findings and blood tests.<sup>8</sup> While useful, the score's reliance on blood tests reduces its practicality for rapid diagnosis in situations where blood sampling is unavailable. Moreover, its overall diagnostic accuracy has been reported as moderate, underscoring the need for more efficient diagnostic tools.<sup>9 10</sup>

The use of tongue coating as a diagnostic feature for acute appendicitis was initially described over 70 years ago.<sup>11</sup> Tongue examination offers several advantages: it is simple, non-invasive, requires no



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY. Published by BMJ Group.

**To cite:** Mori H, Yamasaki K, Saishoji Y, et al. *Emerg Med J* Epub ahead of print: [please include Day Month Year]. doi:10.1136/emmermed-2024-214210



specialised equipment and places minimal burden on the patient. These factors make tongue inspection a potentially valuable tool for diagnosing appendicitis, particularly in low-resource environments.

In this work, we aimed to evaluate the diagnostic utility of tongue examination in acute appendicitis by examining the performance of the Tongue Coating Index (TCI), a previously validated measure for assessing tongue coating.<sup>12</sup>

## METHODS

### Ethical considerations

This is a prospective cohort study that is reported in accordance with the Standards for Reporting of Diagnostic Accuracy Studies guidelines.<sup>13</sup>

This study adheres to the Declaration of Helsinki. Before enrolment, each patient was informed about the study through written documentation and provided written informed consent.

### Data source and participants

This study was conducted at the National Hospital Organization Nagasaki Medical Center, a tertiary medical institution in Nagasaki, Japan, and included adult patients aged  $\geq 20$  years who visited the emergency department or outpatient clinic of the institution between 1 September 2018 and 31 December 2020 with suspected appendicitis. In the Japanese healthcare system, general outpatient clinics in acute care hospitals often accommodate patients with acute conditions, such as suspected appendicitis. Suspected appendicitis was determined by the attending physicians based on the patients' symptoms and physical examination findings. Patients who had previously undergone an appendectomy, those incapable of making decisions due to dementia or mental illness, individuals unable to follow instructions or deemed uncooperative and cases where tongue observation was difficult due to underlying conditions were excluded. Enrolment followed a consecutive sampling approach. Diagnostic imaging studies, such as CT, US and MRI were performed based on the treating physician's clinical judgement. Furthermore, the following parameters were recorded by the attending physician at the time of the patient's consultation: age, sex, body mass index, smoking history, vital signs (blood pressure, pulse rate, temperature, oxygen saturation) and the Alvarado score.<sup>8</sup>

### Outcomes

The reference standard for the diagnosis of acute appendicitis was based on a combination of clinical findings, imaging studies when deemed clinically appropriate, surgical findings and histopathological findings, and follow-up assessments. The diagnosis was adjudicated by the attending physicians involved in the patients' clinical care, following standard clinical coding workflows. In non-surgical cases, the diagnosis was established by the attending physicians' assessment, based on clinical evaluation and imaging reports if performed. The imaging criteria for diagnosing acute appendicitis were as follows:<sup>14–16</sup>

- ▶ US: Non-compressible appendix with a diameter  $\geq 6$  mm, periappendiceal fluid or increased echogenicity of surrounding fat.
- ▶ CT: Enlarged appendix ( $\geq 6$  mm in diameter), periappendiceal fat stranding or appendicolith.
- ▶ MRI: Enlarged appendix ( $\geq 6$  mm in diameter), periappendiceal oedema or adjacent tissue inflammation

To reduce verification bias, all patients were re-evaluated by their attending physicians through follow-up phone calls or

medical record reviews 7 days after the initial consultation to confirm their clinical status.

### Predictors

To quantitatively assess tongue coating, images were captured using a Nikon COOLPIX camera (resolution, 16.05 megapixels; focal-length range, 4.3–21.5 mm; aperture values, f/2.8 to f/8.0; Nikon, Tokyo, Japan). This camera was selected for its capacity to provide high-resolution images suitable for detailed analysis; images were captured before treatment initiation and saved for subsequent analysis. Photographs were obtained under prespecified conditions to maintain standardisation across the study.

The TCI, a reliable and valid method for the quantitative assessment of tongue coating, served as the index test.<sup>12</sup> The TCI divides the tongue surface into nine sections, with the coating in each section being evaluated on a 3-point scale: a score of 0 indicates no coating, 1 indicates partial coating and 2 indicates thick coating. The scores for each section are summed, leading to a cumulative total score ranging from 0 to 18. The TCI was independently assessed by two raters who were not involved in the clinical care of the patients and who evaluated the images later, separate from the patients' clinical visit. Both raters were blinded to the patients' reference standard diagnoses to ensure that the TCI assessments were not influenced by knowledge of the final clinical outcome. For the analysis, the average TCI score from the two raters was used.

### Sample size estimation

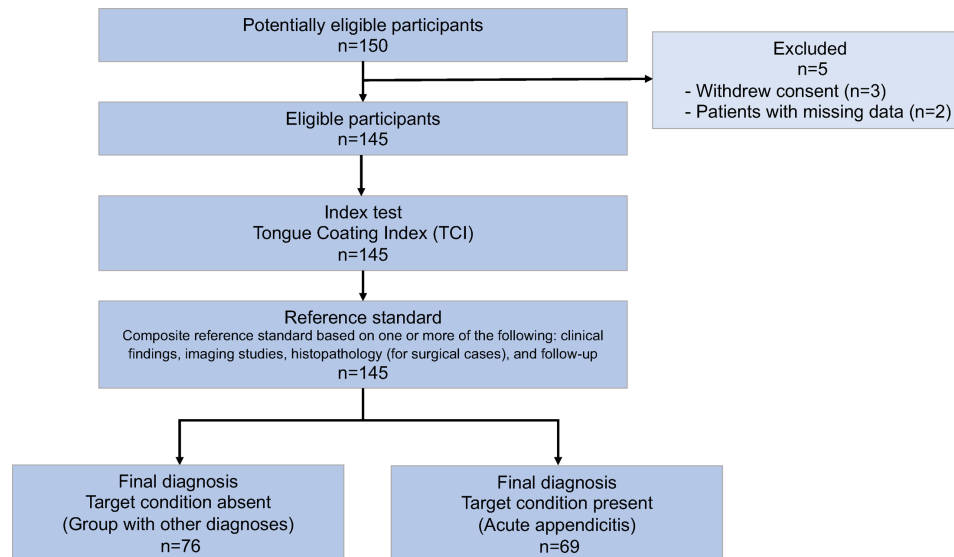
Following the methodology described by Hajian-Tilaki for sample size calculation in diagnostic accuracy studies,<sup>17</sup> we aimed for a sensitivity of 80% with a 5% margin of error and an estimated prevalence of 40% for acute appendicitis among the suspected cases. Given the clinical importance of high sensitivity in order to be able to rule out acute appendicitis (ie, high negative predictive value, NPV), we prioritised this metric among diagnostic performance parameters. Based on these criteria, the sample size calculation indicated that a minimum of 138 participants would be necessary to obtain reliable estimates of both sensitivity and specificity.

### Statistical analyses

Only complete cases with no missing data were included in the statistical analysis. Descriptive statistics of the patient clinical characteristics were calculated for the overall, appendicitis and non-appendicitis groups. Summary statistics were calculated after verifying data distribution. For continuous variables with a normal distribution, means and SD were computed; medians and IQRs were used for continuous variables without a normal distribution. For categorical variables, frequencies and percentages (%) were calculated. For comparisons between two groups, the Mann-Whitney U test was used for continuous variables without a normal distribution. For binary variables, Pearson's  $\chi^2$  test was applied.

The interexaminer reliability of the TCI was assessed using Bland-Altman plots. Diagnostic performance was evaluated for each examiner using the area under the curve (C-index AUC) analysis. Calibration curves were also generated to assess the consistency of diagnostic performance across examiners.

The discriminative ability of the overall Alvarado score and each score construct variable, as well as the TCI for diagnosing appendicitis was evaluated using the C-index AUC along with the corresponding 95% CI, through receiver operating characteristic (ROC) analysis.



**Figure 1** Standards for Reporting of Diagnostic Accuracy Studies flow diagram of participant enrolment and diagnostic evaluation process. This flow diagram illustrates the enrolment and classification of study participants over a 28-month period. A total of 150 patients were initially enrolled, with 5 excluded due to non-consent or missing data, resulting in 145 patients included in the analysis. These patients were categorised into two groups: those diagnosed with acute appendicitis (n=69) and those with other diagnoses (n=76). The reference standard was a composite reference standard, including clinical findings, imaging studies and follow-up for all patients. For surgical cases, histopathological findings were also included. The index test, Tongue Coating Index, was measured for all included patients using tongue images that were captured at the time of enrolment and evaluated at a later date.

A two-sided  $p$  value  $< 0.05$  was considered statistically significant. All statistical analyses were performed with EZR V.1.61 (Jichi Medical University, Saitama, Japan).

### Patient and public involvement

Patients or members of the public were not involved in the design, conduct or reporting of this clinical study.

## RESULTS

### Participant enrolment flow and imaging studies

Between 1 September 2018 and 31 December 2020, 150 patients were enrolled. After excluding three patients who did not consent and two patients with missing data, analyses were performed on 145 patients (figure 1).

### Patient characteristics

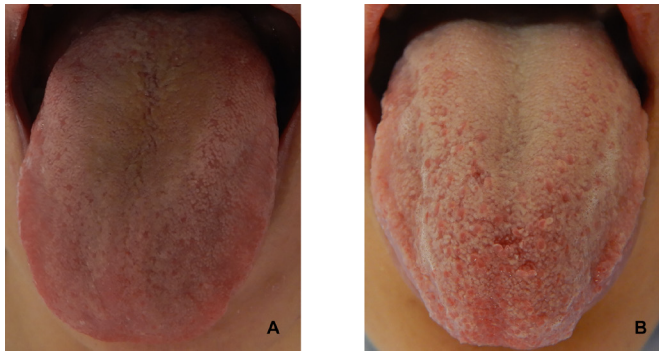
Patient characteristics and imaging results for the entire group and according to final diagnosis are shown in table 1. Among the 145 patients analysed, the median age was 43 years (IQR, 35–56) years, with 79 (55%) female patients. (table 1) Alongside the migration of pain (33% vs 59%), anorexia (63% vs 78%) and tenderness in the right lower quadrant (RLQ) (86% vs 96%), the median TCI was significantly higher in the acute appendicitis group (8 (IQR, 5–11)) than in the group with other diagnoses (6 (IQR, 4–9)) ( $p < 0.05$ ).

Figure 2 shows representative tongue images illustrating the difference in tongue coating between a patient with acute appendicitis (high TCI) and a patient with an alternative diagnosis (low TCI).

**Table 1** Clinical characteristics of patients

	Overall (n=145)	Group with other diagnoses (n=76)	Acute appendicitis (n=69)	P value
Age, median (IQR), years	43 (35–56)	42 (29–50)	47 (35–70)	0.008
Female sex, no. (%)	79 (55)	43 (57)	36 (52)	0.590
Migration of pain, no. (%)	66 (46)	25 (33)	41 (59)	0.001
Anorexia, no. (%)	102 (70)	48 (63)	54 (78)	0.047
Nausea, no. (%)	66 (46)	35 (46)	31 (45)	0.890
Tenderness in the RLQ, no. (%)	131 (90)	65 (86)	66 (96)	0.039
Rebound pain, no. (%)	65 (45)	31 (41)	34 (49)	0.30
Elevated temperature, no. (%)	46 (32)	24 (32)	22 (32)	0.92
Leucocytosis, no. (%)	80 (56)	37 (50)	43 (62)	0.14
Alvarado score, median (IQR)	6 (4–7)	5 (4–7)	7 (5–8)	$< 0.001$
TCI, median (IQR)	7 (4–10)	6 (4–9)	8 (5–11)	0.025
US performed, n (%)	132 (91)	70 (92)	62 (90)	
CT scans performed, n (%)	142 (97)	73 (96)	69 (100)	
MRI performed, n (%)	1 (0.7)	1 (1.3)	0 (0)	
No imaging, n (%)	2 (1.4)	2 (2.6)	0 (0)	

RLQ, right lower quadrant; TCI, Tongue Coating Index; US, ultrasonography.



**Figure 2** Comparative analysis of tongue coating thickness. (A) Patient with a diagnosis other than appendicitis, exhibiting a Tongue Coating Index (TCI) score of 1, indicating a thinner coating. (B) Patient diagnosed with acute appendicitis, characterised by a markedly thicker and unevenly distributed coating, with a TCI score of 13.5.

The Mann-Whitney U test was used for continuous variables with non-normal distributions, while Pearson's  $\chi^2$  test was used for categorical variables.

The most common other diagnoses were diverticulitis (25 cases, 17.2%) and acute enteritis (23 cases, 15.9%). Other conditions included caecal inflammation, pelvic inflammatory diseases (4 cases each, 2.8%), small bowel obstruction, acute pyelonephritis, ureteral stones (3 cases each, 2.1%), mesenteric lymphadenitis and functional gastrointestinal disorders (2 cases each, 1.4%). Less common diagnoses (1 case each, 0.7%) included acute cholecystitis, ovarian haemorrhage, pneumonia, menstrual pain and ruptured aneurysm.

### Discriminative ability of the TCI

The discriminative ability of the individual components of the Alvarado score, the overall score and the TCI for diagnosing acute appendicitis is summarised in table 2. Most C-index AUC values were within a similar range, with Migration of Pain, the composite Alvarado score and the TCI demonstrating statistically significant differences from 0.50. The TCI had a C-index AUC of 0.62 (95% CI, 0.53 to 0.71), which was comparable to individual variables of the Alvarado score, such as migration of pain to the RLQ (AUC 0.63 (95% CI, 0.55 to 0.71)), and to the composite Alvarado score (AUC 0.66 (95% CI, 0.57 to 0.75)).

**Table 2** Discriminative ability of components of the Alvarado score, total Alvarado score and TCI

Variables	C-index AUC (95% CI)	P value
Migration of pain	0.63 (0.55 to 0.71)	<0.001
Anorexia	0.58 (0.50 to 0.66)	NS
Nausea	0.49 (0.41 to 0.58)	NS
Tenderness in the RLQ	0.55 (0.50 to 0.60)	NS
Rebound pain	0.54 (0.46 to 0.62)	NS
Elevated temperature	0.50 (0.43 to 0.58)	NS
Leucocytosis	0.56 (0.48 to 0.64)	NS
Alvarado score	0.66 (0.57 to 0.75)	<0.001
TCI	0.62 (0.53 to 0.71)	0.003

The AUC was calculated using the ROC curve method. The p values were calculated using the z-test for the null hypothesis that AUC=0.5.

AUC, area under the curve; NS, not significant ( $p \geq 0.05$ ); RLQ, right lower quadrant; ROC, receiver operating characteristic; TCI, Tongue Coating Index.

### Diagnostic accuracy of the Alvarado score and the TCI in predicting acute appendicitis

Table 3 summarises the diagnostic accuracy of the Alvarado score and TCI for predicting acute appendicitis. As there were no established thresholds for the TCI, the cut-off values were set exploratively in this study to provide potential thresholds depending on the clinical objective. For the goal of ruling out acute appendicitis, a TCI cut-off of 3 showed high sensitivity of 96% (95% CI, 88% to 98%) and NPV of 84% (95% CI, 62% to 94%), indicating its potential usefulness in identifying patients unlikely to have the condition. Conversely, a cut-off of 10 achieved higher specificity of 83% (95% CI, 73% to 90%) and PPV of 61% (95% CI, 44% to 75%), which could be helpful for confirming the diagnosis in patients with a higher pretest probability.

### Interexaminer reliability of the TCI

The agreement between the two examiners for TCI was evaluated using the Bland-Altman plot (figure 3A), which revealed variability between examiners in the TCI scores assigned. While differences in scoring were evident, a consistent relationship emerged: when one examiner assigned a higher TCI score, the other examiner tended to do the same, and similarly for lower scores.

Figure 3B presents the C-index AUC values, reflecting the correlation of the TCI scores assigned by each examiner with the clinical diagnosis of acute appendicitis. Examiner 1 achieved a C-index AUC of 0.59 (95% CI, 0.48 to 0.67), while Examiner 2 reached a slightly higher C-index AUC of 0.63 (95% CI, 0.53 to 0.71). These results indicate that both examiners showed moderate performance in scoring the TCI in a way that aligned with the presence or absence of acute appendicitis.

Calibration curves (figure 3C) were also generated to assess the accuracy of predicted probabilities relative to the observed outcomes. Both examiners demonstrated well-calibrated predictions, with their predicted probabilities closely matching the actual outcomes. This consistency reinforces the reliability of their diagnostic predictions and suggests that the TCI scoring, while variable between raters, maintained a generally accurate reflection of clinical outcomes.

### DISCUSSION

This study suggests that the degree of tongue coating may be useful in the diagnosis of acute appendicitis, with performance comparable to the Alvarado score and its component of migration of pain. While tongue examination alone may not suffice for definitive diagnosis, its integration with other clinical signs could contribute to the development of new predictive models for acute appendicitis. Such models might enhance diagnostic accuracy, particularly in resource-limited settings where advanced imaging and laboratory tests are not readily available.

In previous studies, such as the systematic review by Kularatna *et al*,<sup>7</sup> the diagnostic performance of the Alvarado score (using a cut-off of 7) varied widely, with reported sensitivities ranging from 67.65% to 96.3% and specificities from 58.18% to 89.39%. This broad range suggests that the diagnostic performance of the Alvarado score is highly dependent on the underlying patient population. In studies reporting the upper limits of diagnostic accuracy, patients with more pronounced clinical features may have been enrolled, resulting in higher Alvarado scores. In contrast, our study enrolled patients with suspected appendicitis, many of whom exhibited intermediate Alvarado scores (4–7) that pose diagnostic challenges. Consequently,

**Table 3** Diagnostic accuracy of the TCI and the Alvarado score in predicting acute appendicitis

	TP	FN	FP	TN	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Alvarado score $\geq 4$	65	4	58	18	94% (86% to 98%)	24% (16% to 34%)	53% (44% to 62%)	82% (61% to 93%)
Alvarado score $\geq 5$	60	9	47	29	87% (77% to 93%)	38% (28% to 49%)	56% (47% to 65%)	76% (61% to 87%)
Alvarado score $\geq 7$	41	28	28	48	59% (48% to 70%)	63% (52% to 73%)	59% (48% to 70%)	63% (52% to 73%)
Alvarado score $\geq 9$	10	59	6	70	15% (8% to 25%)	92% (84% to 96%)	63% (38% to 82%)	54% (48% to 63%)
TCI $\geq 3$	66	3	60	16	96% (88% to 98%)	21% (13% to 32%)	52% (44% to 61%)	84% (62% to 94%)
TCI $\geq 10$	20	49	13	63	29% (20% to 41%)	83% (73% to 90%)	61% (44% to 75%)	56% (47% to 65%)

FN, false negative; FP, false positive; NPV, negative predictive value; PPV, positive predictive value; TCI, Tongue Coating Index; TN, true negative; TP, true positive.

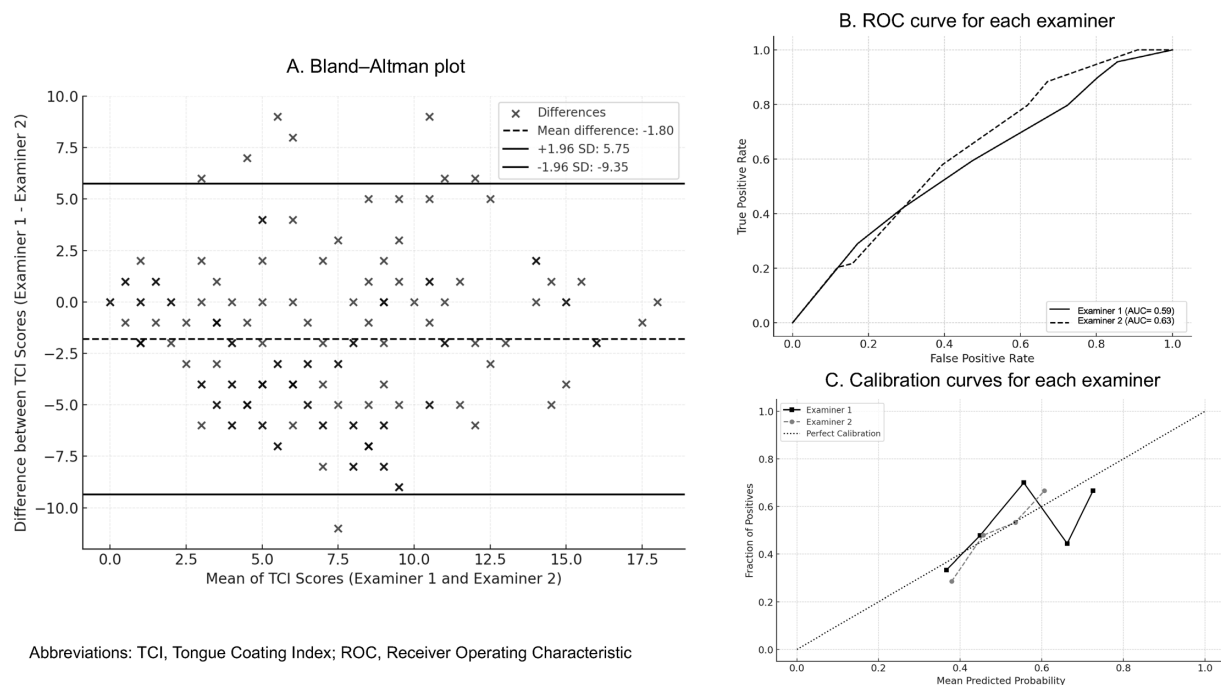
when using a cut-off of 7, our study demonstrated a sensitivity of 59% (95% CI, 48% to 70%) and a specificity of 63% (95% CI, 52% to 73%), which may have contributed to an overall lower AUC (0.62–0.68).

In our study, individual components of the Alvarado score demonstrated modest diagnostic performance. For instance, the migration of pain showed an AUC of 0.63 (95% CI, 0.55 to 0.71), anorexia 0.58 (95% CI, 0.50 to 0.66), and tenderness in the right lower quadrant 0.55 (95% CI, 0.50 to 0.60). These findings are in line with previous studies, which have similarly reported limited discriminative ability for individual clinical signs when used in isolation.<sup>4</sup> This supports the notion that

a composite score, such as the Alvarado score, is necessary to achieve improved diagnostic accuracy for acute appendicitis.

### Physiological implications of tongue coating

Tongue coating is thought to mirror the body's physiological and pathological status, with studies indicating its potential association with gastritis, enteritis, diabetes, ischaemic heart diseases and cancer.<sup>18–20</sup> Changes in the gut microbiota have been hypothesised to be a mechanism,<sup>21–22</sup> which may also apply to appendicitis. Although Chinese medicine has documented appendicitis diagnoses using tongue examinations, the



Abbreviations: TCI, Tongue Coating Index; ROC, Receiver Operating Characteristic

**Figure 3** Interexaminer reliability and diagnostic performance of the TCI. (A) Bland-Altman plot for the TCI. The solid lines represent the limits of agreement (mean difference  $\pm 1.96$  SD) between the TCI scores assigned by two examiners, highlighting the range of variability in their assessments. (B) ROC curve for each examiner. The C-index AUC for Examiner 1 was 0.59 (95% CI, 0.48 to 0.67), while Examiner 2 achieved a C-index AUC of 0.63 (95% CI, 0.53 to 0.71). Both values indicate moderate discriminative ability, with Examiner 2 performing slightly better in distinguishing between cases with and without acute appendicitis. (C) Calibration curves for each examiner. The calibration curves illustrate the predicted probabilities versus observed outcomes for appendicitis for each examiner. Examiner 1 is represented by black squares with a solid line, while Examiner 2 is shown with grey circles and a dashed line. The dotted line represents perfect calibration, where predicted probabilities match observed outcomes exactly. The calibration curves illustrate the agreement between predicted probabilities and the actual occurrence of appendicitis for each examiner. AUC, area under the curve.

practical application of such studies has been limited because of the requirement of specialised equipment.<sup>23 24</sup> Bailey and Clain also highlighted the importance of tongue coating in the diagnosis of acute appendicitis.<sup>11</sup> Although their statement appears to be rooted in personal observations, our study supports the hypothesis.

This study possesses several strengths, notably its focus on a population with an Alvarado score primarily between 4 and 7, where the clinical diagnosis of acute appendicitis is particularly challenging. Additionally, the prospective collection of data with minimal missing values significantly bolsters the reliability of the findings. Recent advancements in diagnostic imaging, such as Point-of-Care Ultrasound (POCUS), have demonstrated high specificity and reduced radiation exposure compared with CT.<sup>14</sup> However, POCUS requires operator expertise and equipment availability, which may limit its application in certain settings. In contrast, the TCI offers a simpler, non-technical alternative, potentially complementing tools like POCUS in resource-limited environments.

The study also enrolled patients based on clinical suspicion of acute appendicitis, a widely accepted approach in diagnostic accuracy research.<sup>25</sup> This methodology reflects real-world clinical practice, where healthcare providers must assess patients suspected of appendicitis. By adopting this approach, we ensure the external validity of the study and mirror the typical patient population encountered in everyday clinical settings.

However, some limitations must be acknowledged. The study was conducted at a single institution, which may introduce spectrum bias and limit the generalisability of our findings.

Cases where acute appendicitis was not initially suspected or were treated as other conditions (eg, with antibiotics)—may not have been included. While these might represent a small proportion, clinically significant cases of acute appendicitis are likely to return for follow-up at our institution due to the healthcare structure in our region, minimising the risk of missing severe cases. Additionally, mild cases of appendicitis that could resolve spontaneously may also have been excluded, potentially introducing selection bias. However, such mild cases are generally not able to be included in most studies and are unlikely to influence the overall findings of the study.

While the TCI has been recognised as a reliable and valid indicator,<sup>12</sup> our study observed variability between examiners, as shown in the Bland-Altman plots. However, the examiners demonstrated consistent scoring trends. The calibration curves and C-index AUC corroborate the overall accuracy and reliability of the TCI scores. Future research should focus on developing standardised and objective methods for measuring the TCI, potentially incorporating automated image analysis techniques. This could minimise interexaminer variability and enhance diagnostic precision.

Although histopathological assessment is traditionally the gold standard for diagnosing acute appendicitis, our study used a clinical diagnosis. Recent reports have highlighted a minimal difference in outcomes between surgical and conservative treatments for uncomplicated acute appendicitis,<sup>26 27</sup> which has led to an increase in non-surgical management.

Implementing histopathology as the gold standard could reduce the number of definitively confirmed cases, thereby challenging the feasibility of the present study by limiting the sample size. Notably, all patients diagnosed with acute appendicitis in this study underwent CT, given its established reliability in diagnosing this condition.<sup>4 28 29</sup>

## Conclusion

The degree of tongue coating holds potential value as a useful, non-invasive diagnostic tool for acute appendicitis. Further research is necessary to validate these findings with external data and to explore the potential development of new decision aids for acute appendicitis, incorporating tongue examination as a component.

**Acknowledgements** We express our gratitude to the Scientific Research Works Peer Support Group (SRWS-PSG) for their valuable discussions. We also thank Editage ([www.editage.com](http://www.editage.com)) for their assistance with English language editing.

**Contributors** HM conceived and designed the study, organised and conducted the data analysis, contributed to the writing of the paper and oversaw the entire study. KY contributed to the study design and data analysis. YS, YT, TM and YN contributed to the study design, data analysis and interpretation, revised the paper and approved the final version. YI critically revised the paper and approved the final version. All authors read drafts of the manuscript and approved the final version. HM is the guarantor for the paper.

**Funding** Canon Foundation for Scientific Research Grant

**Competing interests** The authors, HM and KY, wish to disclose a pending patent application directly related to the content of this manuscript. This disclosure is made to ensure transparency and uphold the integrity of the research process. The details of the patent application are maintained with the relevant authorities and can be disclosed upon request for purposes of academic inquiry and verification.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Consent obtained directly from patient(s).

**Ethics approval** This study was approved by the Institutional Ethics Committee of the NHO Nagasaki Medical Center (approval no.: 30047).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request. The analysed dataset is not publicly available due to specific reasons such as privacy and ethical restrictions but can be obtained from the corresponding author upon reasonable request.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

## ORCID iD

Hideki Mori <http://orcid.org/0000-0001-9417-8165>

## REFERENCES

- Humes DJ, Simpson J. Acute appendicitis. *BMJ* 2006;333:530–4.
- Bickell NA, Aufses AH Jr, Rojas M, *et al*. How time affects the risk of rupture in appendicitis. *J Am Coll Surg* 2006;202:401–6.
- Lewis SRR, Mahony PJ, Simpson J. Appendicitis. *BMJ* 2011;343:bmj.d5976.
- Andersson REB. Meta-analysis of the clinical and laboratory diagnosis of appendicitis. *Br J Surg* 2004;91:28–37.
- Eng KA, Abadeh A, Ligocki C, *et al*. Acute Appendicitis: A Meta-Analysis of the Diagnostic Accuracy of US, CT, and MRI as Second-Line Imaging Tests after an Initial US. *Radiology* 2018;288:717–27.
- Fleming KA, Horton S, Wilson ML, *et al*. The Lancet Commission on diagnostics: transforming access to diagnostics. *Lancet* 2021;398:1997–2050.
- Kularatna M, Lauti M, Haran C, *et al*. Clinical Prediction Rules for Appendicitis in Adults: Which Is Best? *World J Surg* 2017;41:1769–81.
- Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med* 1986;15:557–64.
- Howell JM, Eddy OL, Lukens TW, *et al*. Clinical policy: Critical issues in the evaluation and management of emergency department patients with suspected appendicitis. *Ann Emerg Med* 2010;55:71–116.
- Malik AA, Wani NA. Continuing diagnostic challenge of acute appendicitis: evaluation through modified Alvarado score. *Aust N Z J Surg* 1998;68:504–5.
- Bailey H, Clain A. *Demonstrations of physical signs in clinical surgery*. 1954.
- Shimizu T, Ueda T, Sakurai K. New method for evaluation of tongue-coating status. *J Oral Rehabil* 2007;34:442–7.
- Bossuyt PM, Reitsma JB, Bruns DE, *et al*. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;351:h5527.

- 14 Boyle MJ, Lin-Martore M, Graglia S. Point-of-care ultrasound in the assessment of appendicitis. *Emerg Med J* 2023;40:528–31.
- 15 Terasawa T, Blackmore CC, Bent S, *et al*. Systematic review: computed tomography and ultrasonography to detect acute appendicitis in adults and adolescents. *Ann Intern Med* 2004;141:537–46.
- 16 D'Souza N, Hicks G, Beable R, *et al*. Magnetic resonance imaging (MRI) for diagnosis of acute appendicitis. *Cochrane Database Syst Rev* 2021;12:CD012028.
- 17 Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform* 2014;48:193–204.
- 18 Yin FG, Tian DL, Wang CH. The relationship between fibergastroscopic picture and tongue inspection. *J Tradit Chin Med* 1983;3:49–54.
- 19 Jiang B, Liang X, Chen Y, *et al*. Integrating next-generation sequencing and traditional tongue diagnosis to determine tongue coating microbiome. *Sci Rep* 2012;2:936.
- 20 Han S, Chen Y, Hu J, *et al*. Tongue images and tongue coating microbiome in patients with colorectal cancer. *Microb Pathog* 2014;77:1–6.
- 21 Rauthan K, Joshi S, Kumar L, *et al*. Functional annotation of uncharacterized proteins from *Fusobacterium nucleatum*: identification of virulence factors. *Genomics Inform* 2023;21:e21.
- 22 Aiyoshi T, Kakihara T, Watanabe E, *et al*. A comprehensive microbial analysis of pediatric patients with acute appendicitis. *J Microbiol Immunol Infect* 2023;56:695–704.
- 23 Pang B, Zhang D, Wang K. Tongue image analysis for appendicitis diagnosis. *Inf Sci (Ny)* 2005;175:160–76.
- 24 Zhang D, Pang B, Li N, *et al*. Computerized diagnosis from tongue appearance using quantitative feature classification. *Am J Chin Med* 2005;33:859–66.
- 25 O'Sullivan JW, Banerjee A, Heneghan C, *et al*. Verification bias. *BMJ EBM* 2018;23:54–5.
- 26 Salminen P, Paajanen H, Rautio T, *et al*. Antibiotic Therapy vs Appendectomy for Treatment of Uncomplicated Acute Appendicitis: The APPAC Randomized Clinical Trial. *JAMA* 2015;313:2340–8.
- 27 Sallinen V, Akl EA, You JJ, *et al*. Meta-analysis of antibiotics versus appendectomy for non-perforated acute appendicitis. *Br J Surg* 2016;103:656–67.
- 28 Hlibczuk V, Dattaro JA, Jin Z, *et al*. Diagnostic accuracy of noncontrast computed tomography for appendicitis in adults: a systematic review. *Ann Emerg Med* 2010;55:51–9.
- 29 Rud B, Vejborg TS, Rapoport ED, *et al*. Computed tomography for diagnosis of acute appendicitis in adults. *Cochrane Database Syst Rev* 2019;2019:CD009977.