

ASSESSMENT OF AGREEMENT AMONG TRADITIONAL MEDICINE EXPERTS IN PATIENTS WITH ATROPHIC GASTRITIS

Tang Khanh Huy¹, Ly Vu Minh Chi¹, Tran Ngoc Dang², Nguyen Thi Huong Duong¹

¹ Faculty of Traditional Medicine, University of Medicine and Pharmacy at Ho Chi Minh City, 217 Hong Bang, Cho Lon Ward, Ho Chi Minh City, Vietnam

² Faculty of Public Health, University of Medicine and Pharmacy at Ho Chi Minh City, 217 Hong Bang, Cho Lon Ward, Ho Chi Minh City, Vietnam

Received 18/08/2025

Revised 15/09/2025; Accepted 24/09/2025

ABSTRACT

Objective: As a potential alternative to invasive treatments, tongue diagnosis has increasingly been used as a non-invasive diagnostic method for stomach diseases. However, this method has not yet been systematically applied in the diagnosis of atrophic gastritis (AG). To establish a standardized, consistent, and objective foundation for tongue diagnosis (She zhen) in patients with AG, it is essential to evaluate the degree of inter-expert agreement within an expert consensus panel in Traditional Medicine.

Subjects and Methods: A cross-sectional analytical study was conducted, collecting tongue image data from 120 patients with AG. Each case was independently assessed twice, and the findings were recorded accordingly. The prevalence of tongue features was analyzed, and both intra-rater and inter-rater agreement among experts were evaluated using Kappa statistics.

Results: The intra-rater agreement of individual experts yielded Kappa values ranging from 0.70 to 0.84, indicating substantial to almost perfect reliability ($p < 0.01$). Inter-rater agreement across the expert panel demonstrated only moderate reliability, with a Kappa value of 0.46 ($p < 0.01$). Among the examined features, tongue fissures showed the highest inter-rater consistency, with a Kappa value of 0.67, corresponding to substantial agreement.

Conclusion: Individual experts exhibited substantial to almost perfect intra-rater consistency in the assessment of tongue characteristics, supporting the potential for standardized diagnostic criteria in AG tongue diagnosis.

Keywords: Tongue diagnosis, concordance, Kappa statistic.

*Corresponding author

Email: huongduongdr2012@ump.edu.vn Phone: (+84) 982990747 <https://doi.org/10.52163/yhc.v66iCD15.3290>



ĐÁNH GIÁ SỰ ĐỒNG THUẬN GIỮA CÁC CHUYÊN GIA Y HỌC CỔ TRUYỀN TRONG CHẨN ĐOÁN LƯỠI TRÊN NGƯỜI BỆNH VIÊM TEO DẠ DÀY

Tăng Khánh Huy¹, Lý Vũ Minh Chi¹, Trần Ngọc Đăng², Nguyễn Thị Hương Dương¹

¹Khoa Y học cổ truyền, Đại học Y Dược TP. Hồ Chí Minh, 217 Hồng Bàng, Phường Chợ Lớn, TP. Hồ Chí Minh, Việt Nam

²Khoa Y tế công cộng, Đại học Y Dược TP. Hồ Chí Minh, 217 Hồng Bàng, Phường Chợ Lớn, TP. Hồ Chí Minh, Việt Nam

Ngày nhận bài: 18/08/2025

Ngày chỉnh sửa: 15/09/2025; Ngày duyệt đăng: 24/09/2025

TÓM TẮT

Mục tiêu: Sử dụng phương pháp phân tích đặc điểm lưỡi để chẩn đoán bệnh lý dạ dày thay thế các thủ thuật xâm lấn ngày càng phổ biến, nhưng vẫn chưa áp dụng trên viêm teo dạ dày (VTDD). Để xây dựng nền tảng mang tính chuẩn hoá, nhất quán, khách quan cao trong việc chẩn đoán lưỡi (thiệt chẩn) trên người bệnh VTDD được thực hiện bởi các chuyên gia Y học cổ truyền cần đánh giá mức độ đồng thuận trong hội đồng chuyên gia (HĐCG).

Đối tượng và phương pháp nghiên cứu: Nghiên cứu cắt ngang phân tích thực hiện thu thập dữ liệu hình ảnh lưỡi, tiến hành đọc và ghi nhận kết quả qua hai lần đánh giá trên 120 người bệnh VTDD, khảo sát tỉ lệ đặc điểm lưỡi, đánh giá mức độ đồng thuận trong nội bộ các chuyên gia và giữa các chuyên gia với nhau theo kết quả các đặc điểm lưỡi ghi nhận được bằng giá trị Kappa.

Kết quả: Giá trị Kappa mức độ đồng thuận nội bộ của mỗi chuyên gia dao động từ 0,70 đến 0,84 đạt mức độ tin cậy mạnh đến gần hoàn hảo ($p < 0,01$). Mức độ đồng thuận của HĐCG đạt mức độ tin cậy trung bình (giá trị Kappa là 0,46; $p < 0,01$). Đặc điểm đường nứt lưỡi có mức độ đồng thuận HĐCG cao nhất trong các đặc điểm, giá trị Kappa 0,67 đạt mức độ đồng thuận mạnh.

Kết luận: Các chuyên gia có mức độ đồng thuận nội bộ đạt mức mạnh đến gần hoàn hảo khi đánh giá các đặc điểm lưỡi, đây là cơ sở tiềm năng cho việc tiêu chuẩn hóa chẩn đoán lưỡi ở người bệnh VTDD.

Từ khóa: Thiệt chẩn, mức độ đồng thuận, giá trị Kappa.

1. ĐẶT VẤN ĐỀ

Viêm teo niêm mạc dạ dày mạn tính hay viêm teo dạ dày (VTDD) là một trong những bệnh đường tiêu hóa phổ biến [1], và là bước quan trọng trong tiến triển đến ung thư dạ dày [2]. VTDD diễn tiến âm thầm làm tăng nguy cơ ung thư dạ dày gấp 4,12 lần [3], ước tính mỗi năm có hàng triệu người trên thế giới có nguy cơ tử vong sớm do ung thư dạ dày từ viêm dạ dày mạn tính gây ra [4]. Sàng lọc giúp phát hiện ung thư dạ dày giai đoạn sớm [5] và điều trị giúp làm giảm tỉ lệ tử vong [6], nhưng đánh giá mô bệnh học bằng sinh thiết [7] và quan sát hình ảnh nội soi [1] đều là phương pháp xâm lấn, có thể gây khó chịu cho người bệnh và ít khả thi trong các điều kiện y tế hạn chế.

Những năm gần đây, Y học cổ truyền (YHCT) đã tiến hành các nghiên cứu sâu rộng về chẩn đoán và điều trị VTDD. Trong đó, thiệt chẩn (chẩn đoán bằng quan sát lưỡi) là thành phần đặc trưng và quan trọng trong chẩn đoán hội chứng YHCT [3]. Trong thực hành lâm sàng, các bác sĩ thực hiện quan sát các đặc điểm lưỡi như màu sắc, hình dạng và lượng nước bọt để đưa ra chẩn đoán bệnh, tuy nhiên, thường bị ảnh hưởng bởi đánh giá chủ quan. Nguyên nhân do khác biệt từ kiến thức cá nhân, kinh nghiệm, tư duy, kỹ năng chẩn đoán và nhận thức hoặc giải thích về màu sắc. Những người khác nhau có thể đưa ra những kết luận khác nhau trên cùng một lưỡi, thậm chí cùng một người có thể

*Tác giả liên hệ

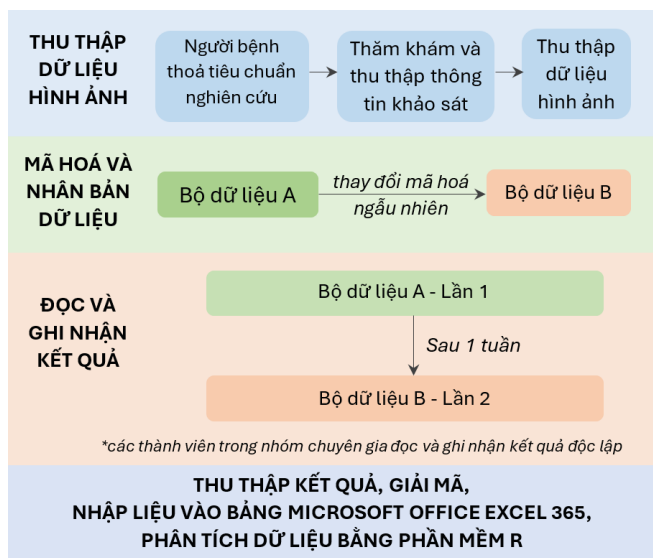
Email: huongduongdr2012@ump.edu.vn Điện thoại: (+84) 979446799 <https://doi.org/10.52163/yhc.v66iCD15.3290>

đưa ra những chẩn đoán khác nhau vào những thời điểm khác nhau. Nghiên cứu cho thấy mức độ đồng thuận nội bộ của mỗi bác sĩ đạt 61%, trong khi mức độ đồng thuận giữa các bác sĩ thấp tới 18,2% [8]. Sự mâu thuẫn này khiến việc chẩn đoán YHCT gặp nhiều hoài nghi và hạn chế, mặc dù đây là phương pháp chẩn đoán lưỡi thường quy, phổ biến và dễ thực hiện, đặc biệt khi tiến hành phát triển việc phân tích hình ảnh lưỡi để đánh giá bệnh lý VTDD. Hiện nay, chưa có nghiên cứu đánh giá về sự đồng thuận giữa các chuyên gia Y học cổ truyền trong chẩn đoán lưỡi trên người bệnh VTDD, vì vậy cần thực hiện nghiên cứu này.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Thiết kế nghiên cứu

Nghiên cứu cắt ngang phân tích được tiến hành từ tháng 11/2024 đến tháng 6/2025 (Hình 1).



Hình 1. Các bước nghiên cứu

Tiêu chuẩn chọn mẫu: Người bệnh có chẩn đoán VTDD dựa trên kết quả nội soi dạ dày thực quản theo phân loại Kimura-Takemoto; người bệnh gồm cả nam và nữ từ đủ 18 tuổi trở lên và đồng thuận tham gia nghiên cứu.

Tiêu chuẩn loại trừ: Người bệnh có một trong các tiêu chuẩn sau: có bệnh lý ảnh hưởng đến lưỡi như bệnh về gan, thận, bệnh lý nhiễm trùng, những người đang mang thai, hoặc suy giảm nhận thức theo thang đo MMSE; có các bệnh lý miệng hoặc lưỡi như viêm lưỡi, không thể thè lưỡi ổn định, những người có nguy cơ mắc bệnh khớp thái dương hàm, trật khớp; người đã cắt dạ dày, ung thư dạ dày, loét dạ dày, tá tràng; có sử dụng thuốc hoặc thức ăn có thể làm thay đổi đặc điểm lưỡi; hút thuốc lá, uống rượu; hoặc không tuân thủ các yêu cầu nghiên cứu, không hợp tác trong quá trình thu thập dữ liệu.

2.2. Cỡ mẫu

Rubio khuyến cáo số lượng tối thiểu là 3 chuyên gia và tối đa là 10 chuyên gia [9]. Nghiên cứu của chúng tôi có 5 chuyên gia.

Theo công thức ước lượng cỡ mẫu giá trị Kappa:

$$n = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2 \times [P_0(1-P_0)]}{(P_1 - P_0)^2}$$

P_0 : tỉ lệ đồng ý quan sát tương ứng với $\kappa_0 = 0,4$

P_1 : tỉ lệ đồng ý quan sát tương ứng với $\kappa_1 = 0,6$

Số người đánh giá: 5. Chọn $\alpha = 5\%$ và $1-\beta = 80\%$

Số mức phân loại từ 2 đến 5

Vì nghiên cứu có 10 biến đánh giá để đảm bảo cho Power đạt yêu cầu với những đặc điểm có nhiều mức đánh giá và độ lệch cao.

Dự phòng thiếu dữ liệu hoặc loại trừ 10%.

Cỡ mẫu đề xuất cho nghiên cứu là 120 người.

2.3. Phương pháp thu thập số liệu

Người bệnh được các bác sĩ chuyên khoa tiêu hoá tại Bệnh viện Đại học Y Dược TP.HCM cơ sở 1 và cơ sở 2 xác định chẩn đoán VTDD dựa trên nội soi theo thang đo Kimura-Takemoto. Mức độ phân độ này được sử dụng làm tiêu chuẩn lựa chọn và mô tả mẫu nghiên cứu. Các bác sĩ nội soi đã được tập huấn thống nhất phương pháp đọc kết quả và hoàn toàn được làm mù thông tin với kết quả chẩn đoán lưỡi.

Thành viên nhóm nghiên cứu mời người bệnh tham gia nghiên cứu. Người bệnh đồng ý sẽ được phỏng vấn bằng bộ câu hỏi khảo sát soạn sẵn và thu thập 03 hình ảnh và 01 đoạn phim lưỡi.

Hội đồng chuyên gia (HĐCG) gồm các bác sĩ chuyên khoa Y học cổ truyền được tập huấn và thống nhất các định nghĩa về đặc điểm lưỡi và quy trình thực hiện vào ngày 06/01/2025 tại Đại học Y Dược TP.HCM. Nội dung tập huấn được ban hành và gửi đến các thành viên hội đồng chuyên gia sử dụng trong suốt quá trình thực hiện nghiên cứu.

HĐCG tiến hành quan sát hình ảnh lưỡi và đoạn phim của người bệnh độc lập, được làm mù thông tin kết quả nội soi lần thông tin lâm sàng khác. Dữ liệu hình ảnh lưỡi được mã hoá và sắp xếp ngẫu nhiên trong 2 bộ dữ liệu độc lập và được đánh giá hai lần cách nhau một tuần.

2.4. Công cụ đo lường

Công cụ thu thập hình ảnh: Máy chụp hình kĩ thuật số SONY ILCE-6000/BAP2; thông số kỹ thuật: Power zoom Lens kit E PZ 16-50mm F3.5-5.6 OSS; độ phân giải hình ảnh thu được: 24 MP (6000 x 4000 điểm ảnh).

Đặc điểm lưỡi bao gồm 10 đặc điểm phân thành các đặc điểm về chất lưỡi (màu lưỡi, hình dạng lưỡi, độ nhuận), đặc điểm rêu lưỡi (màu rêu lưỡi, độ dày rêu

lưỡi, đặc điểm bề mặt) và các đặc điểm đặc biệt (dấu ấn răng, đường nứt lưỡi, điểm ứ huyết, gai lưỡi).

2.5. Phương pháp thống kê

Cohen's kappa là một thống kê mạnh mẽ hữu ích cho cả thử nghiệm độ tin cậy giữa các nhà đánh giá (cho hai người đánh giá) hoặc trong cùng một nhà đánh giá. Như với tất cả các thống kê tương quan, kappa là một giá trị chuẩn hóa và do đó được diễn giải giống nhau trong nhiều nghiên cứu. Chỉ số Cohen's kappa gọi tắt là Kappa được sử dụng để đánh giá sự đồng thuận nội bộ của mỗi chuyên gia trong hai lần đánh giá. Chỉ số Fleiss Kappa để đo lường mức độ đồng thuận giữa các nhà đánh giá được điều chỉnh từ Cohen's kappa cho 3 người đánh giá trở lên). Chỉ số Fleiss Kappa được sử dụng để đánh giá sự đồng thuận giữa các thành viên HĐCG với nhau cho mỗi lần đánh giá.

Giá trị Kappa và Fleiss Kappa được phân chia thành 5 mức đồng thuận theo Landis & Koch (1977), trong đó 0,0-0,20 thể hiện mức đồng thuận rất thấp, từ 0,21-0,40 là mức đồng thuận thấp, 0,41-0,60: đồng thuận trung bình, 0,61-0,80: đồng thuận mạnh, 0,81-1,0: đồng thuận gần hoàn hảo [10].

2.6. Đạo đức nghiên cứu

Nghiên cứu được sự chấp thuận của Hội đồng đạo đức trong nghiên cứu y sinh học Đại học Y dược Thành phố Hồ Chí Minh theo văn bản số 2841/HĐĐĐ-ĐHYD ngày 07/10/2024.

3. KẾT QUẢ

3.1. Đặc điểm mẫu nghiên cứu

Bảng 1. Đặc điểm về nhân khẩu học của mẫu nghiên cứu (n=120)

Đặc điểm	N = 120
Độ tuổi, Trung bình ± Độ lệch chuẩn, năm	45,33 ± 12,57
Giới tính, n%	
Nam	54 (45)
Nữ	66 (55)
Nghề nghiệp, n%	
Lao động chân tay	35 (29,17)
Lao động trí óc	47 (39,17)
Nghỉ việc do bệnh nặng/nghỉ hưu	14 (11,66)
Nội trợ/ Thất nghiệp	21 (17,50)
Khác	3 (2,5)
Trình độ học vấn, n%	
Dưới cấp I	10 (8,33)

Đặc điểm	N = 120
Tiểu học (Cấp I)	14 (11,67)
Trung học cơ sở (Cấp II)	29 (24,17)
Trung học phổ thông (Cấp III)	27 (22,50)
Đại học, Cao đẳng, Trung cấp	38 (31,67)
Sau đại học	2 (1,66)
BMI, n%	
Nhẹ	5 (4,17)
Trung bình	55 (45,83)
Thừa cân	31 (25,83)
Béo phì độ I	26 (21,67)
Béo phì độ II	3 (2,5)
Nhiễm <i>Helicobacter pylori</i> , n%	
Không	97 (80,83)
Có	15 (12,5)
Không rõ	8 (6,67)
Tăng huyết áp, n%	14 (11,67)
Đái tháo đường, n%	8 (6,67)
Rối loạn lipid máu, n%	19 (15,83)
Hút thuốc lá, n%	15 (12,5)
Tiền căn gia đình mắc VTDD, n%	0 (0)
Tiền căn gia đình mắc Ung thư dạ dày, n%	2 (1,67)
Phân độ Kimura-Takemoto, n%	
C1	50 (41,67)
C2	50 (41,67)
C3	15 (12,5)
O1	1 (0,83)
O2	1(0,83)
O3	3 (2,5)

Nhóm người bệnh đưa vào nghiên cứu có độ tuổi trung bình 45,33 ± 12,57; tỉ lệ nam : nữ là 1:1,2. Tỉ lệ phân độ Kimura-Takemoto tập trung chủ yếu ở C1 và C2.

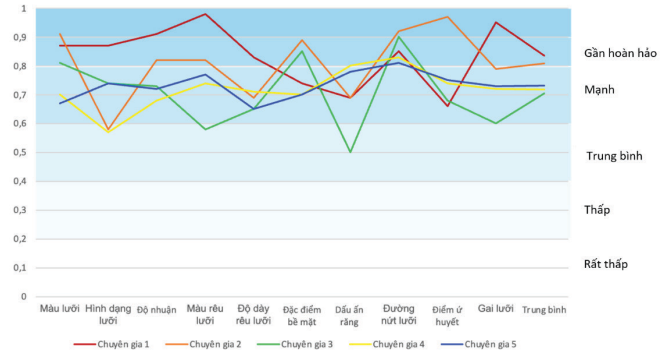
3.2. Đặc điểm hội đồng chuyên gia

Trong hội đồng 5 chuyên gia YHCT đưa vào nghiên cứu có 3 chuyên gia nam (60%) và 2 chuyên gia nữ (40%). Các chuyên gia có trình độ từ bác sĩ (1/5 – 20%), bác sĩ chuyên khoa I (1/5 – 20%), Thạc sĩ (1/5 – 20%) và Tiến sĩ (2/5 – 40%). Thời gian hoạt động chuyên môn trung bình 10,8 năm, chuyên gia có thời gian hoạt động ít nhất là 7 năm, cao nhất là 14 năm.

3.3. Đánh giá mức độ đồng thuận nội bộ chuyên gia

Có 2/5 chuyên gia có giá trị kappa trung bình đạt mức độ đồng thuận nội bộ gần hoàn hảo (chuyên gia 1: 0,84; chuyên gia 2: 0,81). Các chuyên gia còn lại có giá trị Kappa đạt mức độ đồng thuận mạnh (các giá trị Kappa lần lượt 0,70; 0,72; 0,73). Các số liệu này có ý nghĩa thống kê với $p < 0,01$.

Về các đặc điểm, chuyên gia 1 đánh giá đặc điểm màu rêu lưỡi có sự đồng thuận gần hoàn hảo với giá trị Kappa 0,98 (95%CI 0,95-1,02, $P_o = 0,99$). Trong khi đó, chuyên gia 3 đánh giá đặc điểm dấu ấn răng có sự kém đồng thuận nhất, có mức độ trung bình, với giá trị Kappa 0,5 (95%CI 0,35-0,66, $P_o = 0,75$).



Hình 2. So sánh mức độ đồng thuận nội bộ của mỗi chuyên gia theo thang điểm Kappa

Bảng 2. Mức độ đồng thuận nội bộ của mỗi chuyên gia

Đặc điểm	Màu lưỡi	Hình dạng lưỡi	Độ nhún	Màu rêu lưỡi	Độ dày rêu lưỡi	Đặc điểm bề mặt	Dấu ấn răng	Đường nứt lưỡi	Điểm ứ huyết	Gai lưỡi
Chuyên gia 1										
K	0,87	0,87	0,91	0,98	0,83	0,74	0,69	0,85	0,66	0,95
Khoảng tin cậy 95%	(0,80-0,94)	(0,79-0,95)	(0,84-0,97)	(0,95-1,02)	(0,73-0,92)	(0,39-1,1)	(0,56-0,82)	(0,75-0,94)	(0,43-0,89)	(0,88-1,01)
P_o	0,91	0,92	0,94	0,99	0,9	0,98	0,85	0,92	0,93	0,98
Chuyên gia 2										
K	0,90	0,58	0,82	0,82	0,69	0,88	0,69	0,92	0,97	0,79
Khoảng tin cậy 95%	(0,84-0,97)	(0,44-0,71)	(0,73-0,91)	(0,73-0,91)	(0,58-0,8)	(0,66-1,11)	(0,55-0,83)	(0,84-0,99)	(0,92-1,03)	(0,66-0,92)
P_o	0,93	0,76	0,89	0,89	0,8	0,99	0,87	0,96	0,99	0,92
Chuyên gia 3										
K	0,81	0,74	0,73	0,58	0,65	0,85	0,5	0,9	0,68	0,6
Khoảng tin cậy 95%	(0,72-0,89)	(0,64-0,85)	(0,62-0,84)	(0,45-0,7)	(0,52-0,77)	(0,64-1,06)	(0,35-0,66)	(0,82-0,98)	(0,54-0,82)	(0,46-0,74)
P_o	0,86	0,83	0,83	0,74	0,8	0,98	0,75	0,95	0,85	0,8
Chuyên gia 4										
K	0,70	0,57	0,68	0,74	0,71	0,7	0,8	0,83	0,74	0,72
Khoảng tin cậy 95%	(0,59-0,81)	(0,44-0,7)	(0,56-0,8)	(0,63-0,86)	(0,59-0,83)	(0,4-0,99)	(0,69-0,91)	(0,73-0,93)	(0,57-0,9)	(0,6-0,85)
P_o	0,82	0,74	0,81	0,87	0,83	0,97	0,9	0,92	0,92	0,87
Chuyên gia 5										
K	0,67	0,74	0,84	0,77	0,65	0,7	0,78	0,81	0,75	0,73
Khoảng tin cậy 95%	(0,55-0,78)	(0,63-0,84)	(0,6-0,83)	(0,67-0,87)	(0,54-0,77)	(0,4-0,99)	(0,66-0,89)	(0,71-0,92)	(0,59-0,91)	(0,6-0,86)
P_o	0,79	0,84	0,72	0,85	0,78	0,97	0,89	0,91	0,92	0,88

3.4. Đánh giá mức độ đồng thuận trong HĐCG

Giá trị Kappa về sự đồng thuận giữa 5 chuyên gia lần lượt qua hai lần đánh giá, có trị số trung bình $0,47 \pm 0,1$ và $0,44 \pm 0,13$. Kết quả này cho thấy sự đồng thuận giữa các chuyên gia đạt mức độ trung bình

(0,41-0,6) khi nhìn chung trên các đặc điểm. Kết quả cụ thể sau mỗi lần đánh giá các đặc điểm lưỡi được ghi nhận trong Bảng 3.

Bảng 3. Sự đồng thuận giữa các chuyên gia

Đặc điểm	Lần 1			Lần 2			Trung bình
	Fleiss Kappa	Khoảng tin cậy 95%	P _o	Fleiss Kappa	Khoảng tin cậy 95%	P _o	
Màu lưỡi	0,45	(0,38-0,52)	0,62	0,47	(0,4-0,54)	0,64	0,46
Hình dạng lưỡi	0,42	(0,35-0,49)	0,64	0,37	(0,29-0,44)	0,61	0,40
Độ nhuận	0,5	(0,43-0,57)	0,69	0,45	(0,38-0,53)	0,67	0,48
Màu rêu lưỡi	0,5	(0,43-0,57)	0,7	0,44	(0,36-0,51)	0,67	0,47
Độ dày rêu lưỡi	0,39	(0,32-0,46)	0,63	0,38	(0,31-0,45)	0,61	0,39
Đặc điểm bề mặt	0,58	(0,40-0,77)	0,96	0,54	(0,36-0,73)	0,96	0,57
Dấu ấn răng	0,55	(0,47-0,63)	0,78	0,49	(0,41-0,57)	0,75	0,52
Đường nứt lưỡi	0,65	(0,57-0,73)	0,83	0,69	(0,61-0,77)	0,85	0,67
Điểm ứ huyết	0,35	(0,25-0,45)	0,79	0,26	(0,15-0,35)	0,76	0,31
Gai lưỡi	0,33	(0,25-0,42)	0,69	0,25	(0,17-0,34)	0,65	0,29
Trung bình	0,47±0,1		0,73	0,44±0,13		0,72	0,46

Ghi chú: Giá trị trung bình được tính bằng trung bình cộng hệ số Fleiss Kappa của các đặc điểm lưỡi.

Trong đó, đặc điểm đường nứt lưỡi, đặc điểm bề mặt và dấu ấn răng có nhóm ba đặc điểm có độ đồng thuận từ trung bình đến mạnh ở cả hai lần đánh giá. Hai đặc điểm điểm ứ huyết và gai lưỡi các chuyên gia có sự đồng thuận chỉ đạt mức thấp và thấp hơn các đặc điểm còn lại được đưa vào nghiên cứu.

4. BÀN LUẬN

4.1. Mức độ đồng thuận nội bộ chuyên gia

Sự đồng thuận nội bộ của mỗi chuyên gia qua hai lần đánh giá có trị số trung bình từ thấp nhất là 0,70 đến 0,84. Mức độ đồng thuận từ mạnh đến gần hoàn hảo chứng tỏ mỗi chuyên gia có sự nhất quán và thống nhất trong việc đánh giá đặc điểm lưỡi. Mặc dù hoàn toàn mang tính yếu tố con người, quan điểm cá nhân và thời điểm, nhưng các kết quả cho thấy rằng việc đánh giá đặc điểm thiết chẩn trên người bệnh VTDD có độ tin cậy từ trung bình trở lên tương đương với từ 32% đến 81% dữ liệu đáng tin cậy. Kết quả này có tính khả quan hơn các nghiên cứu trước đó, như trong nghiên cứu của Minah Kim (2008) [8] achieved by a group of TCM practitioners. Method: Ten (10, độ đồng thuận nội bộ của các bác sĩ chỉ là 61%. Sự khác biệt nổi bật giữa hai nghiên cứu về số lượng mẫu và số lượng chuyên gia. Mặc dù nghiên cứu này có số

lượng mẫu cao hơn, nhưng chỉ có năm chuyên gia, các chuyên gia được tập huấn và các nội dung tập huấn cho phép được làm tài liệu tham khảo trong quá trình đánh giá. Điều này giúp loại bỏ các yếu tố ảnh hưởng bởi kiến thức cá nhân, vì vậy ghi nhận được mức độ đồng thuận nội bộ của mỗi người tốt hơn các nghiên cứu trước đó. Đây là nền tảng cơ bản nhất cho tính nhất quán trong việc đánh giá thăm khám bằng thiết chẩn.

4.2. Mức độ đồng thuận trong HĐCG

Khi nghiên cứu so sánh giá trị Kappa HĐCG lần lượt qua hai lần đánh giá, có trị số trung bình $0,47 \pm 0,1$ và $0,44 \pm 0,13$ tương ứng với mức độ đồng thuận giữa các chuyên gia đạt mức trung bình (0,41-0,6) trên tất cả các đặc điểm. Tỷ lệ đồng ý thô trung bình (P_o) lần lượt qua hai lần đánh giá là 73% và 72%, dao động từ 61% đến 96%. Điều này khẳng định cho mức độ đồng thuận giữa con người với nhau là yếu tố trở ngại lớn nhất trong chẩn đoán YHCT. Các yếu tố ảnh hưởng đến kết quả có thể kể đến như đánh giá cá nhân của mỗi người về màu sắc khác nhau, thời gian hành nghề dao động từ 7 – 14 năm có thể là một trong các yếu tố hạn chế kinh nghiệm hoặc phân tán kết quả của người đánh giá.

Nghiên cứu này có số lượng chuyên gia thấp hơn nghiên cứu Lun-chien Lo (2012) [11], thông thường số lượng ý kiến tham gia càng lớn sự đồng nhất càng giảm, tuy nhiên với các đặc điểm có số lựa chọn từ 2-5 lựa

chọn thì số lượng ý kiến lớn giúp câu trả lời cho khuynh hướng tập trung và định hướng hơn. Bên cạnh đó, thời lượng tập huấn của nghiên cứu 2012 kéo dài 2 năm, các chuyên gia được luyện tập với 1000 hình ảnh lưỡi. Vì những hạn chế về thời gian nghiên cứu, chúng tôi chỉ thực hiện được một buổi tập huấn và ban hành tài liệu quy định thống nhất về đánh giá các đặc điểm lưỡi hỗ trợ cho quá trình làm việc của chuyên gia, nhưng cách thức này vẫn thật sự có nhiều giới hạn.

Trong đó, đặc điểm đường nứt lưỡi, đặc điểm bề mặt và dấu ấn răng có nhóm ba đặc điểm có độ đồng thuận trung bình đến mạnh ở cả hai lần đánh giá bởi vì đây đều là các đặc điểm có định nghĩa rõ ràng, dễ quan sát và có tính thống nhất cao, ít ảnh hưởng bởi các kiến thức cá nhân. Hai đặc điểm điểm ứ huyết và gai lưỡi các chuyên gia có sự đồng thuận chỉ đạt mức thấp và thấp nhất trong các đặc điểm được đưa vào nghiên cứu, nguyên nhân có thể do đặc tính khó quan sát của hai yếu tố này trong thực tế.

Nhưng từ kết quả ghi nhận được, chúng ta ghi nhận rằng đây vẫn là một trong những phương thức đánh giá chẩn đoán lưỡi phù hợp và cơ bản nhất trong YHCT, đặc biệt ở những vùng nông thôn hoặc điều kiện kinh tế chưa phát triển để trang bị máy móc, ưu thế bởi tính linh hoạt và đơn giản. Mặc dù thế, việc tìm kiếm một phương thức đánh giá khắc phục được những hạn chế khi thực hiện bởi con người nhưng vẫn đảm bảo tính đồng thuận với phương thức đánh giá truyền thống này vẫn là xu hướng phát triển chung. Nghiên cứu này cung cấp nền tảng cơ bản đầu tiên giúp cho việc chẩn đoán lưỡi trên người bệnh VTDD trong YHCT có tính khả thi và áp dụng rộng rãi trong tương lai, đồng thời cung cấp các gợi ý đáng tin cậy để tạo điều kiện thuận lợi cho bác sĩ trong việc chẩn đoán và điều trị hiệu quả bệnh lý VTDD.

5. KẾT LUẬN

Các chuyên gia có mức độ đồng thuận nội bộ đạt mức mạnh đến gần hoàn hảo khi đánh giá các đặc điểm lưỡi. Mức độ đồng thuận trong HĐCG đạt mức độ trung bình đến mạnh khi đánh giá đặc điểm đường nứt lưỡi, đặc điểm bề mặt, và dấu ấn răng.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ kinh phí bởi Đại học Y Dược Thành phố Hồ Chí Minh theo hợp đồng số 275/2025/HĐ-ĐHYD ngày 28 tháng 4 năm 2025.

TÀI LIỆU THAM KHẢO

[1] Đào Trần Tiến, Vũ Trường Khanh, Phí Thị Thùy Ngân, và cộng sự. Nhận xét thói quen ăn uống, lâm sàng ở bệnh nhân viêm teo niêm mạc dạ dày. *Bản B của Tạp chí Khoa học và Công nghệ Việt Nam*. 2021;63(7):11-16. doi: 10.31276/VJST.63(7).11-16.

- [2] Marques-Silva L, Areia M, Elvas L, et al. Prevalence of gastric precancerous conditions: a systematic review and meta-analysis. *Eur J Gastroenterol Hepatol*. 2014;26(4):378–387. doi: 10.1097/MEG.0000000000000065.
- [3] Li J, Pan J, Xiao D, et al. Chronic atrophic gastritis and risk of incident upper gastrointestinal cancers: a systematic review and meta-analysis. *J Transl Med*. 2024;22(1):429. doi: 10.1186/s12967-023-04736-w.
- [4] Sipponen P, Maaroos HI. Chronic gastritis. *Scand J Gastroenterol*. 2015;50(6):657–667. doi: 10.3109/00365521.2015.1019918.
- [5] Abe T, Kodama M, Murakami K, et al. Impact of *Helicobacter pylori* CagA diversity on gastric mucosal damage: an immunohistochemical study of East-Asian-type CagA. *J Gastroenterol Hepatol*. 2011;26(4):688–693. doi:10.1111/j.1440-1746.2010.06565.x.
- [6] Sahara S, Sugimoto M, Vilaichone RK, et al. Role of *Helicobacter pylori* cagA EPIYA motif and vacA genotypes for the development of gastrointestinal diseases in Southeast Asian countries: a meta-analysis. *BMC Infect Dis*. 2012;12:223. doi: 10.1186/1471-2334-12-223.
- [7] Yao K, Uedo N, Kamada T, et al. Guidelines for endoscopic diagnosis of early gastric cancer. *Dig Endosc*. 2020;32(5):663–698. doi: 10.1111/den.13684.
- [8] Kim M, Cobbin D, Zaslowski C. Traditional Chinese Medicine Tongue Inspection: An Examination of the Inter- and Intrapractitioner Reliability for Specific Tongue Characteristics. *The Journal of Alternative and Complementary Medicine [Internet]*. 2008 [cited 2025 Aug 17];14(5):527–536. doi: 10.1089/acm.2007.0079.
- [9] Rubio DM, Berg-Weger M, Tebb SS, et al. Objectifying content validity: Conducting a content validity study in social work research. *Soc Work Res*. 2003;27(2):94–104. doi: 10.1093/swr/27.2.94.
- [10] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977 Mar;33(1):159-74. PMID: 843571.
- [11] Lo LC, Chen YF, Chen WJ, et al. The Study on the Agreement between Automatic Tongue Diagnosis System and Traditional Chinese Medicine Practitioners. *Evid Based Complement Alternat Med*. 2012;2012:505063. doi: 10.1155/2012/505063.

