Nghiên cứu tổng quan

DỮ LIỆU LÂM SÀNG SẴN SÀNG CHO TRÍ TUỆ NHÂN TẠO TRONG Y HỌC CỔ TRUYỀN HÀN QUỐC

AI-Ready clinical data for Korean Traditional Medicine

Sanghun Lee[1,2,*], **Trần Nhật Minh**[3]

[1]*Phòng Dữ liệu Y học Hàn Quốc, Viện Y học Cổ truyền Hàn Quốc (KIOM), Daejeon, Hàn Quốc*

[2]*Khoa Khoa học Y học Hội tụ Hàn Quốc, Đại học Khoa học và Công nghệ (UST), Daejeon, Hàn Quốc*

[3]*Khoa Y học cổ truyền, Trường Đại học Y – Dược, Đại học Huế, TP. Huế, Việt Nam*

*Tác giả liên hệ | Corresponding author: ezhani00@gmail.com

TÓM TẮT

Mục tiêu: Tổng hợp các yêu cầu cốt lõi để xây dựng dữ liệu lâm sàng sẵn sàng cho trí tuệ nhân tạo trong y học cổ truyền. **Phương pháp nghiên cứu:** Tổng quan tường thuật các tài liệu về chuẩn hóa, quản trị và khai thác dữ liệu lâm sàng phục vụ trí tuệ nhân tạo. **Kết quả:** Dữ liệu sẵn sàng cho trí tuệ nhân tạo cần bảo đảm nguồn gốc rõ ràng, chất lượng và cấu trúc nhất quán, cùng với siêu dữ liệu đầy đủ. Trong y học cổ truyền, chuẩn hóa cần bắt đầu từ việc chuyển đổi các khái niệm lâm sàng định tính thành các biến số có thể ghi nhận và đo lường, đồng thời thống nhất thiết bị, đơn vị và quy trình thu thập. Chẩn đoán lưỡi cho thấy rõ những thách thức này do sự khác biệt giữa người đánh giá và điều kiện thu nhận dữ liệu. Việc xây dựng dữ liệu tham chiếu ở người khỏe mạnh và các bộ hình ảnh chuẩn có thể hỗ trợ đối chiếu, giáo dục và phát triển các hệ thống trí tuệ nhân tạo. Bên cạnh đó, khả năng liên thông và quản trị dữ liệu là điều kiện để dữ liệu được chia sẻ và tái sử dụng. **Kết luận:** Xây dựng dữ liệu sẵn sàng cho trí tuệ nhân tạo trong y học cổ truyền là quá trình chuẩn hóa xuyên suốt từ đo lường, dữ liệu và thuật ngữ đến quản trị và ứng dụng lâm sàng.

Từ khóa: Y học cổ truyền Hàn Quốc; chuẩn hóa dữ liệu; dữ liệu lâm sàng; trí tuệ nhân tạo.

ABSTRACT

Objective: To synthesize the core requirements for building artificial intelligence (AI)-ready clinical data in traditional medicine. **Methods:** A narrative review of the literature on the standardization, governance, and utilization of clinical data for AI applications. **Results:** AI-ready data require clear provenance, consistent quality and structure, and comprehensive metadata. In traditional medicine, standardization should begin by transforming qualitative clinical concepts into recordable and measurable variables, while harmonizing devices, units, and data collection procedures. Tongue diagnosis clearly illustrates these challenges because of inter-rater variability and differences in data acquisition conditions. Establishing reference data from healthy populations and standardized image sets may support comparison, education, and the development of AI systems. In addition, data interoperability and governance are essential for enabling data sharing and reuse. **Conclusion:** Building AI-ready data in traditional medicine requires comprehensive standardization, from measurement, data, and terminology to governance and clinical application.

Keywords: Korean traditional medicine; data standardization; clinical data; artificial intelligence.

Received: 05/10//2026 |	Accepted: 09/10/2026 |	Publication date: 01/12/2026 |

©The authors. This **open-access** work is licensed under a CC BY 4.0 License

I. ĐẶT VẤN ĐỀ

Trí tuệ nhân tạo (artificial intelligence, AI) đang chuyển nhanh từ giai đoạn phát triển thuật toán sang giai đoạn ứng dụng rộng rãi trong y tế. Các hệ thống AI hiện được nghiên cứu cho chẩn đoán hình ảnh, phân tầng nguy cơ, hỗ trợ quyết định và quản lý dữ liệu lâm sàng. Tuy nhiên, hiệu quả của một hệ thống AI không chỉ phụ thuộc vào mô hình mà còn phụ thuộc sâu sắc vào chất lượng, tính đại diện, khả năng truy xuất và tính nhất quán của dữ liệu được sử dụng để huấn luyện và đánh giá. Một phân tích trên Nature Medicine nhấn mạnh rằng các tiêu chuẩn đối với bộ dữ liệu y tế có vai trò quan trọng trong việc nâng cao tính minh bạch, đa dạng và khả năng sử dụng an toàn của AI trong y tế.[1]

Đối với y học cổ truyền Hàn Quốc, thách thức này càng được thể hiện rõ. Nhiều thông tin lâm sàng quan trọng được hình thành từ quá trình quan sát, hỏi bệnh, bắt mạch, xem lưỡi và tổng hợp các biểu hiện theo lý luận y học cổ truyền. Nếu các khái niệm này chỉ tồn tại dưới dạng nhận định chủ quan hoặc các trường dữ liệu không thống nhất, chúng khó có thể trở thành dữ liệu huấn luyện chất lượng cao cho AI. Vì vậy, chuyển đổi số trong y học cổ truyền không nên chỉ bắt đầu bằng việc lựa chọn một thuật toán AI, mà cần bắt đầu từ việc xác định cách đo lường, cách biểu diễn, chuẩn hóa và mô tả dữ liệu.

Trong những năm gần đây, các diễn đàn quốc tế cũng ngày càng quan tâm đến AI và y học cổ truyền. Tổ chức Y tế Thế giới đã tổ chức cuộc họp kỹ thuật toàn cầu về ứng dụng AI trong y học cổ truyền vào tháng 9 năm 2024, với sự tham gia của các chuyên gia từ nhiều khu vực và tập trung vào năng lực, chia sẻ tri thức, quản trị và đạo đức.[2] Bài viết này tập trung vào yêu cầu dữ liệu sẵn sàng cho AI (AI-ready data), chuẩn hóa dữ liệu lâm sàng và xây dựng hệ sinh thái liên thông, bền vững.

II. NỘI DUNG TỔNG QUAN

1. Từ AI dựa trên mô hình đến AI dựa trên dữ liệu

Một xu hướng quan trọng trong phát triển và ứng dụng AI hiện nay là sự chuyển dịch từ cách tiếp cận "model-centric AI" sang "data-centric AI". Trong cách tiếp cận truyền thống, phần lớn nguồn lực được tập trung vào lựa chọn thuật toán, kiến trúc mô hình và tối ưu tham số. Trong khi đó, cách tiếp cận dựa trên dữ liệu xem chất lượng, tính nhất quán và khả năng sử dụng của dữ liệu là một thành phần cốt lõi của hiệu quả AI. Đối với lĩnh vực y tế, điều này càng quan trọng vì sai sót dữ liệu có thể dẫn đến sai lệch mô hình và ảnh hưởng trực tiếp đến quyết định lâm sàng.[1]

Sự khác biệt về mức độ số hóa và chuẩn hóa dữ liệu giữa các lĩnh vực hình ảnh y tế cho thấy ảnh hưởng rõ rệt của chất lượng dữ liệu đầu vào đối với hiệu quả của AI. Trong chẩn đoán hình ảnh, dữ liệu hình ảnh thường đã có định dạng số và quy trình thu nhận tương đối chuẩn hóa. Ngược lại, trong giải phẫu bệnh, nhiều công đoạn như lấy mẫu mô, xử lý mô, nhuộm, tạo tiêu bản và chuyển đổi thành hình ảnh đều có thể tạo ra biến thiên và sai số. Điều này cho thấy cùng một công nghệ AI có thể đạt hiệu quả rất khác nhau tùy thuộc vào mức độ chuẩn hóa của dữ liệu đầu vào.

Trong Y học cổ truyền, vấn đề này càng rõ rệt vì nhiều khái niệm lâm sàng có thể được diễn giải khác nhau nếu thiếu định nghĩa thao

tác và phương pháp đo thống nhất. Vì vậy, chuyển đổi kiến thức lâm sàng thành các biến số định lượng là bước nền tảng để dữ liệu có thể được kiểm tra, so sánh, liên thông và sử dụng cho các ứng dụng AI.

2. Ba yêu cầu cốt lõi của dữ liệu sẵn sàng cho trí tuệ nhân tạo

Báo cáo khái quát yêu cầu đối với dữ liệu cho AI thành ba nhóm, có thể gọi là "3C": nguồn gốc và kiểm chứng rõ ràng (Clear lineage and validation); sạch và có cấu trúc (Clean and well-structured); và siêu dữ liệu, tài liệu mô tả đầy đủ (Comprehensive metadata and documentation).

Bảng 1. Ba yêu cầu cốt lõi của dữ liệu sẵn sàng cho trí tuệ nhân tạo

Yêu cầu	Nội dung cốt lõi	Ý nghĩa đối với y học cổ truyền
Clear lineage and validation	Truy xuất được nguồn gốc dữ liệu; phép đo phù hợp với tiêu chuẩn; có khả năng kiểm chứng	Biết dữ liệu được tạo ra bởi ai, bằng thiết bị nào, theo SOP nào và đã qua bước kiểm tra nào
Clean and well-structured	Thành phần dữ liệu được định nghĩa rõ, thống nhất và có cấu trúc	Thống nhất tên biến, kiểu dữ liệu, đơn vị, mã hóa và cách xử lý giá trị thiếu
Comprehensive metadata and documentation	Ghi nhận đầy đủ thông tin mô tả và bối cảnh dữ liệu	Cho phép người sử dụng và hệ thống AI hiểu được đối tượng, điều kiện đo, quy trình, nhãn và giới hạn của dữ liệu

3. Chuẩn hóa phép đo là nền tảng của dữ liệu lâm sàng y học cổ truyền

AI chất lượng cao cần AI-ready data; AI-ready data cần bộ dữ liệu chuẩn hóa và có khả năng liên thông; còn bộ dữ liệu chuẩn hóa lại phụ thuộc vào thiết bị đo và quy trình đo được chuẩn hóa. Vì vậy, "chuẩn hóa" phải được thực hiện từ lớp vật lý của phép đo trước khi tiến đến lớp dữ liệu và thuật toán.

Ba thành phần được nhấn mạnh là từ vựng, đơn vị và độ không đảm bảo phép đo. Từ vựng cần quy định cách gọi và định nghĩa các khái niệm. Đơn vị cần được thống nhất để các giá trị có thể so sánh. Độ không đảm bảo phép đo giúp người sử dụng hiểu giới hạn của phép đo thay vì coi một giá trị quan sát được là hoàn toàn chính xác. Cách tiếp cận này phù hợp với nguyên tắc chung của quản lý chất lượng dữ liệu và đánh giá chất lượng hệ thống AI.[3,4]

Một ví dụ rõ về vai trò của chuẩn hóa dữ liệu là sự phát triển của trí tuệ nhân tạo giọng nói. Để các hệ thống có thể nhận diện và xử lý giọng nói từ nhiều nguồn khác nhau, dữ liệu cần được thu thập theo những quy chuẩn thống nhất về nội dung, định dạng, chất lượng bản ghi và thông tin đi kèm. Khi các tiêu chuẩn này được thiết lập, dữ liệu từ nhiều nguồn có thể được tích hợp và sử dụng hiệu quả hơn cho việc phát triển các mô hình AI. Kinh nghiệm này cho thấy trong Y học cổ truyền, việc chuẩn hóa dữ liệu cũng cần được xem là một phần của quá trình xây dựng hệ thống AI, chứ không phải là bước bổ sung sau khi mô hình đã được phát triển. Điều này đặc biệt quan trọng đối với những dữ

liệu có tính chủ quan hoặc phụ thuộc vào người thu thập, như triệu chứng, tứ chẩn và hình ảnh lưỡi.

4. Chẩn đoán lưỡi: minh họa cho các điểm nghẽn của chuẩn hóa

Chẩn đoán lưỡi là một ví dụ điển hình cho những thách thức khi xây dựng dữ liệu lâm sàng Y học cổ truyền cho các ứng dụng trí tuệ nhân tạo. Trước hết, nhiều đặc điểm của lưỡi có bản chất liên tục nhưng trong thực hành lâm sàng thường được phân thành các nhóm hoặc mức độ rời rạc. Việc chuyển từ một phổ liên tục sang các nhãn phân loại có thể làm mất thông tin và tạo ra sự khác biệt giữa những người đánh giá. Đây là vấn đề quan trọng trong xây dựng dữ liệu huấn luyện AI, bởi sai số trong nhãn có thể trở thành nguồn sai lệch của mô hình.

Bên cạnh đó, đánh giá lưỡi có tính cá thể và phụ thuộc đáng kể vào người quan sát. Mặc dù một bác sĩ có thể tương đối nhất quán khi đánh giá lại cùng một hình ảnh, cách lựa chọn nhãn giữa các bác sĩ vẫn có thể khác nhau. Trong một nghiên cứu với 10 bác sĩ Y học Hàn Quốc, độ đồng thuận trung bình đạt kappa = 0,47, thuộc mức đồng thuận trung bình theo thang diễn giải được sử dụng trong nghiên cứu. Kết quả này cho thấy việc xây dựng dữ liệu nhãn chỉ dựa trên kinh nghiệm chuyên gia chưa đủ để bảo đảm tính nhất quán cao; cần có tiêu chí đánh giá rõ ràng và quy trình chuẩn hóa.

Một thách thức khác là sự biến thiên của điều kiện thu nhận dữ liệu. Các yếu tố trước khi chụp hoặc đo có thể ảnh hưởng đến đặc điểm hình ảnh của lưỡi. Ranc và cộng sự cho thấy lớp nước bọt trên bề mặt niêm mạc lưỡi có thể làm thay đổi các đặc tính bề mặt.[5] Do đó, tình trạng bề mặt lưỡi, việc súc miệng, thời điểm chụp và các yếu tố liên quan đến ăn uống có thể ảnh hưởng đến kết quả quan sát, trong đó tình trạng rêu lưỡi cũng có thể thay đổi theo thời điểm sau ăn. Nếu không quy định thống nhất thời điểm, điều kiện chuẩn bị và quy trình chụp, cùng một người có thể tạo ra những dữ liệu khác nhau ở các thời điểm khác nhau.

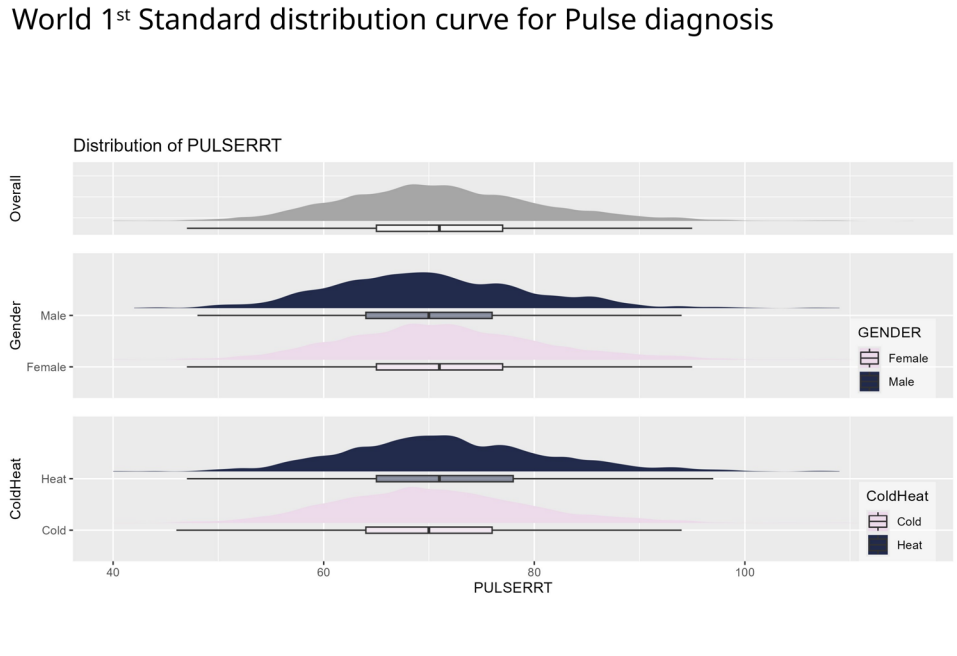
Những vấn đề này cho thấy chất lượng dữ liệu lưỡi không chỉ phụ thuộc vào số lượng mẫu mà còn phụ thuộc vào độ ổn định của phép đo, tính nhất quán của nhãn và mức độ chuẩn hóa của quy trình thu thập. Vì vậy, mở rộng kích thước bộ dữ liệu mà không kiểm soát các nguồn biến thiên trên chưa chắc làm tăng giá trị của dữ liệu và có thể làm gia tăng sai lệch trong các mô hình AI.

Bảng 2. Các điểm nghẽn chính trong chuẩn hóa dữ liệu chẩn đoán lưỡi

Điểm nghẽn	Biểu hiện	Hướng chuẩn hóa
Thang đo rời rạc	Đặc điểm liên tục bị chuyển thành nhóm/mức độ	Xác định định nghĩa thao tác, thang đo và ngưỡng phân loại
Tính cá thể và thiếu chuẩn tham chiếu	Khác biệt giữa các bác sĩ khi gán nhãn	Xây dựng bộ mẫu chuẩn, đào tạo và đánh giá đồng thuận
Thiếu quy trình chuẩn	Điều kiện trước đo và thời điểm đo khác nhau	Xây dựng SOP về thiết bị, ánh sáng, tư thế, thời điểm và điều kiện chuẩn bị

5. Từ dữ liệu cá thể đến dữ liệu tham chiếu chuẩn

Một hướng tiếp cận để giải quyết những điểm nghẽn trên là xây dựng dữ liệu tham chiếu ở người khỏe mạnh trên cơ sở chuẩn hóa thiết bị, quy trình thu nhận và cách ghi nhận dữ liệu. Một bộ dữ liệu khoảng 13.000 người khỏe mạnh đã được sử dụng để xây dựng các đường cong phân bố chuẩn hóa cho một số đặc điểm, trong đó có tốc độ mạch, màu lưỡi và lượng nước bọt. Các đường cong này cho phép chuyển những nhận định định tính thành vị trí tương đối của một cá thể trong quần thể tham chiếu. Chẳng hạn, thay vì chỉ mô tả "mạch mạnh" hay "mạch yếu", đặc điểm có thể được biểu diễn theo vị trí của cá thể trong phân bố của nhóm tham chiếu phù hợp (**Hình 1**). Tương tự, màu lưỡi hoặc lượng nước bọt có thể được diễn giải theo giới và nhóm tuổi khi có đủ dữ liệu tham chiếu.

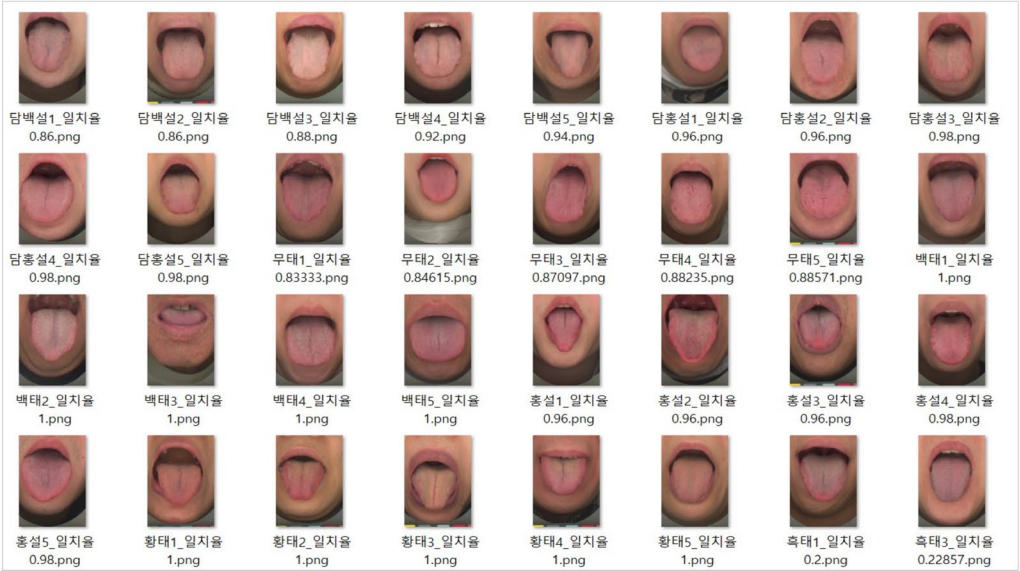


Hình 1. Đường cong phân bố tham chiếu của tốc độ mạch ở người khỏe mạnh

Bên cạnh dữ liệu định lượng, việc xây dựng các bộ hình ảnh chuẩn hóa cũng có ý nghĩa quan trọng đối với chẩn đoán lưỡi. Các hình ảnh được lựa chọn dựa trên mức độ nhất quán khi một bác sĩ đánh giá lặp lại và mức độ đồng thuận giữa nhiều bác sĩ, từ đó tạo thành các mẫu tham chiếu phục vụ đào tạo và góp phần giảm sự khác biệt trong cách hiểu và gán nhãn giữa các cơ sở (**Hình 2**).

Từ góc độ xây dựng dữ liệu, một bộ tham chiếu có giá trị không chỉ nằm ở số lượng hay chất lượng hình ảnh mà còn ở thông tin đi kèm và tính chuẩn hóa của quá trình thu nhận. Các thông tin về đối tượng, điều kiện đo, nhãn, mức độ đồng thuận và quy trình thu thập cần được lưu trữ đầy đủ. Khi đó, cùng một nguồn dữ liệu có thể phục vụ nhiều mục đích, từ đào tạo và nghiên cứu đến phát triển và đánh giá các mô hình trí tuệ nhân tạo.

World 1st Reference tongue diagnosis Standard for Education



Hình 2. Minh họa bộ hình ảnh chuẩn phục vụ giáo dục chẩn đoán lưỡi

6. Từ nhận định chủ quan đến hỗ trợ quyết định dựa trên dữ liệu

Trong thực hành truyền thống, bác sĩ có thể đưa ra nhận định như "mạch của người bệnh có vẻ yếu" dựa trên kinh nghiệm và quá trình thăm khám. Khi có dữ liệu tham chiếu được xây dựng và chuẩn hóa, cùng một thông tin có thể được đặt trong phân bố của quần thể chuẩn, chẳng hạn xác định vị trí của giá trị đo so với nhóm tham chiếu phù hợp về giới và tuổi. Cách tiếp cận này giúp chuyển một phần nhận định định tính thành thông tin có thể định lượng và so sánh.

Điều này không có nghĩa là thay thế phán đoán lâm sàng. Giá trị của dữ liệu tham chiếu nằm ở việc cung cấp một thước đo khách quan hơn để hỗ trợ giao tiếp, theo dõi và đánh giá diễn biến. Kết quả đo có thể được so sánh giữa các lần khám, giữa các cơ sở hoặc giữa các nhóm dân số khi điều kiện thu nhận dữ liệu được chuẩn hóa.

Để đạt được điều đó, dữ liệu của người bệnh cần được thu thập theo cùng một SOP với dữ liệu tham chiếu. Nếu thiết bị, điều kiện đo hoặc quy trình thu nhận khác nhau, sự khác biệt quan sát được có thể phản ánh sai khác về phương pháp thay vì thay đổi thực sự của người bệnh. Vì vậy, tiêu chuẩn hóa cần được thực hiện xuyên suốt, từ thiết bị và quy trình thu thập đến phần mềm, lưu trữ và sử dụng dữ liệu trong thực hành lâm sàng.

7. Khả năng liên thông và quản trị dữ liệu

Dữ liệu AI-ready không nên được xây dựng như những "kho dữ liệu đóng". Khả năng liên thông là điều kiện để dữ liệu có thể được chia sẻ, kết hợp và tái sử dụng. HL7 FHIR là một trong những chuẩn phổ biến cho biểu diễn và trao đổi thông tin sức khỏe, góp phần hỗ trợ liên thông dữ liệu.[6] Tuy nhiên, đối với YHCT, bên cạnh các chuẩn chung của y tế còn cần có lớp thuật ngữ và mô hình dữ liệu đặc thù để biểu diễn đầy đủ các khái niệm chuyên môn.

BioCroissant là một ví dụ về hướng mở rộng siêu dữ liệu cho lĩnh vực y sinh, trong đó có các nhóm thông tin về bản thể y sinh, bối cảnh lâm sàng, đặc điểm người bệnh, quản trị và pháp lý, nguồn gốc dữ liệu và khả năng so sánh giữa các hệ thống. Những thành phần này phù hợp với xu hướng xây dựng siêu dữ liệu có cấu trúc và máy có thể đọc được, qua đó giúp dữ liệu dễ tìm kiếm, dễ hiểu và thuận lợi hơn cho các quy trình phân tích và phát triển AI.[7]

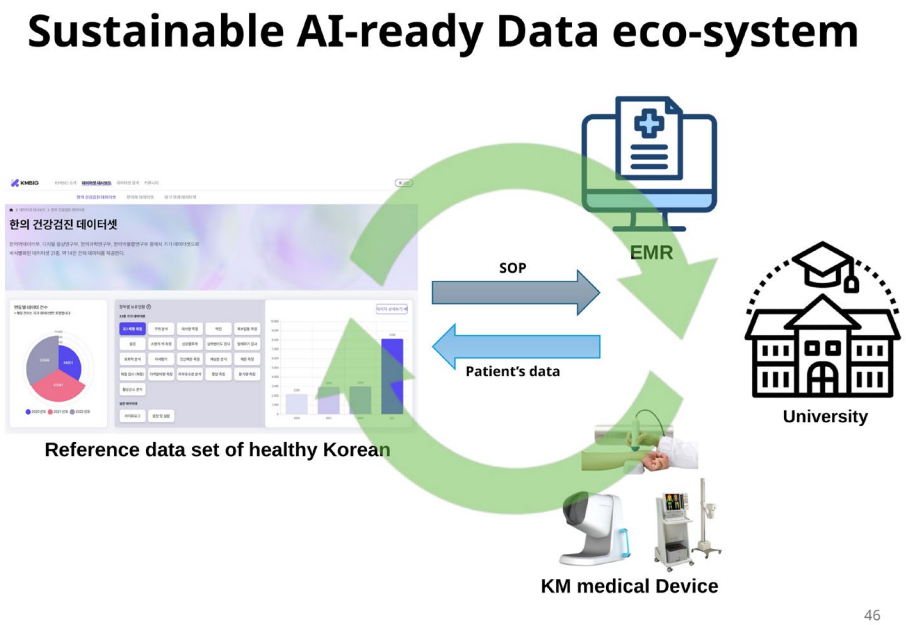
Quản trị dữ liệu cần được thiết kế song song với các yêu cầu kỹ thuật. Các tiêu chuẩn hiện hành đề cập đến nhiều khía cạnh như chất lượng dữ liệu, chất lượng mô hình AI, vòng đời dữ liệu, độ tin cậy, tính minh bạch và quản lý rủi ro.[3,4,8] Đối với dữ liệu lâm sàng, các thông tin về sự đồng thuận, quyền riêng tư, kiểm soát truy cập và mục đích sử dụng cần được gắn với dữ liệu ngay từ khi tạo lập.

8. Hướng tới hệ sinh thái dữ liệu AI-ready bền vững

Mục tiêu cuối cùng không phải chỉ xây dựng một bộ dữ liệu lớn mà là hình thành một hệ sinh thái trong đó dữ liệu được tạo ra, kiểm định, sử dụng và tái tạo liên tục. Để đạt được mục tiêu đó cần kết nối ít nhất năm thành phần: dữ liệu tham chiếu ở người khỏe mạnh; dữ liệu người bệnh; hồ sơ bệnh án điện tử; thiết bị y học cổ truyền; và các trường đại học, cơ sở đào tạo.

Trong mô hình này, SOP là cầu nối giữa cơ sở đào tạo, bệnh viện, hệ thống hồ sơ bệnh án điện tử và nhà sản xuất thiết bị. Dữ liệu mới có thể bổ sung vào bộ tham chiếu, trong khi bộ tham chiếu được cập nhật tiếp tục hỗ trợ giáo dục và nghiên cứu, tạo thành một vòng lặp dữ liệu bền vững.

Tính bền vững còn phụ thuộc vào khả năng mở rộng. Siêu dữ liệu có cấu trúc, phiên bản và nguồn gốc rõ ràng giúp giảm phụ thuộc vào thao tác thủ công. Croissant 1.1 là một ví dụ về hướng tiếp cận siêu dữ liệu máy đọc, hỗ trợ nguồn gốc và liên thông thuật ngữ.[7] Các nguyên tắc này có thể áp dụng cho hình ảnh, tín hiệu sinh học, triệu chứng, chẩn đoán, điều trị và kết quả trong y học cổ truyền.



Hình 3. Mô hình hệ sinh thái dữ liệu AI-ready bền vững

III. KẾT LUẬN

Dữ liệu lâm sàng sẵn sàng cho trí tuệ nhân tạo là nền tảng của quá trình chuyển đổi số y học cổ truyền. Việc xây dựng dữ liệu AI-ready cần bắt đầu từ chuẩn hóa phép đo và quy trình, tiếp tục bằng cấu trúc dữ liệu, siêu dữ liệu, khả năng truy xuất và liên thông, đồng thời tích hợp quản trị, đạo đức và quản lý rủi ro. Ví dụ từ chẩn đoán lưỡi cho thấy những hạn chế về thang đo, tính nhất quán giữa chuyên gia và điều kiện thu nhận có thể trở thành điểm nghẽn trước cả giai đoạn lựa chọn thuật toán. Ngược lại, dữ liệu tham chiếu chuẩn hóa, đường cong phân bố và bộ mẫu giáo dục có thể tạo nền tảng cho việc diễn giải khách quan hơn và phát triển các ứng dụng AI có khả năng tái sử dụng. Hướng đi bền vững là xây dựng một hệ sinh thái kết nối dữ liệu người khỏe mạnh, người bệnh, hồ sơ bệnh án điện tử, thiết bị và cơ sở đào tạo, trong đó các chuẩn dữ liệu và SOP được dùng xuyên suốt vòng đời dữ liệu. Đây là tiền đề để y học cổ truyền chuyển từ số hóa thông tin sang chuyển đổi số thực sự, trong đó dữ liệu có thể được hiểu, chia sẻ, kiểm chứng và sử dụng lại bởi cả con người và hệ thống AI.

TUYÊN BỐ CỦA TÁC GIẢ

Đạo đức nghiên cứu:

Nghiên cứu không tiến hành can thiệp trực tiếp trên đối tượng là con người, do đó không phát sinh các vấn đề đạo đức liên quan đến can thiệp trên người. Các thông tin và dữ liệu sử dụng trong nghiên cứu được bảo đảm tính bảo mật, chỉ phục vụ cho mục đích nghiên cứu và không tiết lộ thông tin nhận diện cá nhân.

Nguồn tài trợ:

Không có nguồn tài trợ

Xung đột lợi ích:

Các tác giả cam kết không có xung đột lợi ích trong nghiên cứu này.

Công bố trước đó (nếu có): Không

TÀI LIỆU THAM KHẢO

1. Arora A, Alderman JE, Palmer J, et al. The value of standards for health datasets in artificial intelligence-based applications. *Nat Med.* 2023;29(11):2929-2938. DOI:10.1038/s41591-023-02608-w.

2. World Health Organization. Kicking off the journey of artificial intelligence in traditional medicine. Published September 13, 2024.

3. International Organization for Standardization, International Electrotechnical Commission. ISO/IEC 5259-1:2024. *Artificial intelligence—Data quality for analytics and machine learning (ML)—Part 1: Overview, terminology, and examples.* Geneva: International Organization for Standardization; 2024.

4. International Organization for Standardization, International Electrotechnical Commission. ISO/IEC 25059:2023. *Software engineering—Systems and software Quality Requirements and Evaluation (SQuaRE)—Quality model for AI systems.* Geneva: International Organization for Standardization; 2023.

5. Ranc H, Elkhyat A, Servais C, Mac-Mary S, Launay B, Humbert P. Friction coefficient and wettability of oral mucosal tissue: changes induced by a salivary layer. *Colloids Surf A Physicochem Eng Asp.* 2006;276(1-3):155-161. DOI:10.1016/j.colsurfa.2005.10.033.

6. Health Level Seven International. Fast Healthcare Interoperability Resources (FHIR).

Health Level Seven International. Accessed September 27, 2026.

7. MLCommons. Croissant Format Specification, Version 1.1. Published January 29, 2026.

8. International Organization for Standardization, *International Electrotechnical Commission. ISO/IEC 23894:2023. Information technology—Artificial intelligence—Guidance on risk management*. Geneva: International Organization for Standardization; 2023.