

HIỆU QUẢ ỨNG DỤNG CÔNG NGHỆ NHẬN DẠNG KÝ TỰ QUANG HỌC TRONG TỰ ĐỘNG ĐỒNG BỘ VÀ NHẬN ĐỊNH KẾT QUẢ ĐO CHỨC NĂNG HÔ HẤP TẠI BỆNH VIỆN ĐẠI HỌC Y HÀ NỘI

Bùi Mỹ Hạnh^{1,2,✉}, Vương Thị Ngân², Nguyễn Tất Hậu²

¹Trường Đại học Y Hà Nội

²Bệnh viện Đại học Y Hà Nội

Nghiên cứu đánh giá hiệu quả ứng dụng công nghệ nhận dạng ký tự quang học trong tự động đồng bộ và nhận định kết quả đo chức năng hô hấp (CNHH). 40 chuyên gia thực hiện phân tích định tính và định lượng ứng dụng hệ thống trên 11.778 bản đo chức năng hô hấp tại Bệnh viện Đại học Y Hà Nội trong giai đoạn từ 04/2025 đến 09/2025. Kết quả cho thấy 100% người dùng đánh giá cao tính thân thiện, khả năng tích hợp, tính nhanh chóng, độ chính xác và tính sẵn sàng của hệ thống. Thời gian xử lý mỗi bản đo giảm 24 lần, tiết kiệm 4.554 giờ làm việc, tương ứng giảm 204,2 - 536,7 triệu đồng chi phí nhân lực, vận hành; với 2,14 tỷ điểm dữ liệu được lưu trữ. Hệ thống chứng minh hiệu quả vượt trội trong tự động hóa quy trình đo - phân tích - lưu trữ, góp phần chuẩn hóa dữ liệu, giảm sai sót và nâng cao hiệu quả quản lý, khai thác dữ liệu y tế.

Từ khóa: Đo chức năng hô hấp, tự động, nhận dạng ký tự quang học, học sâu, trí tuệ nhân tạo, hiệu quả.

I. ĐẶT VẤN ĐỀ

Đo chức năng hô hấp (CNHH) có vai trò quan trọng trong phát hiện sớm, phân loại giai đoạn, tiên lượng và theo dõi đáp ứng điều trị các bệnh lý hô hấp.¹ Quy trình nhập liệu và nhận định kết quả đo CNHH truyền thống phụ thuộc nhiều vào thao tác thủ công, thiếu nhất quán, dễ xảy ra sai sót, mất nhiều thời gian và đánh mất dữ liệu. Bên cạnh đó, dữ liệu đo CNHH hiện nay thường ở dạng non-DICOM, chủ yếu được lưu trữ dưới dạng file ảnh hoặc PDF, khiến việc đồng bộ và trích xuất dữ liệu từ phần mềm đo sang hệ thống bệnh viện gặp nhiều khó khăn. Nhiều cơ sở khám chữa bệnh chưa có điều kiện đầu tư máy đo có phần mềm kết nối tương thích với HIS do chi phí cao, dẫn đến tình trạng dữ liệu chưa được số hóa, không

thể tích hợp và khai thác phục vụ chẩn đoán, nghiên cứu hay quản lý bệnh mạn tính. Vậy làm thế nào để ứng dụng công nghệ trong tự động đồng bộ, nhận định kết quả đo CNHH nhanh chóng, chính xác và tin cậy?

Sự phát triển của công nghệ nhận dạng ký tự quang học kết hợp mô hình học sâu (Optical Character Recognition-Deep learning, OCR-DL), đã đánh dấu bước tiến quan trọng trong tự động hóa xử lý và phân tích dữ liệu y học.² Trên thế giới, một vài nghiên cứu đã phát triển và ứng dụng các công cụ hỗ trợ đảm bảo và kiểm soát chất lượng, chẩn đoán tự động, lưu trữ và chia sẻ dữ liệu đo CNHH trong hệ thống quản lý y tế.³⁻⁵ Tuy nhiên, các nghiên cứu mới tập trung vào việc tự động phân tích, dự đoán, chưa giải quyết triệt để bài toán trích xuất-chuẩn hóa-đồng bộ-liên thông dữ liệu đo CNHH từ ảnh/PDF sang HIS/EMR, đặc biệt ở các quốc gia có hạ tầng y tế không đồng nhất. Tại Việt Nam, từ năm 2018, Bệnh viện Đại học Y Hà Nội đã phát

Tác giả liên hệ: Bùi Mỹ Hạnh

Trường Đại học Y Hà Nội

Email: buimyhhanh@hmu.edu.vn

Ngày nhận: 30/10/2025

Ngày được chấp nhận: 11/12/2025

triển hệ thống tự động hóa nhận định kết quả đo CNHH bằng ngôn ngữ Javascript đảm bảo rút ngắn quá trình trả lời kết quả 3 - 4 lần với độ chính xác lên tới 100% nếu không có những sai sót do người nhập liệu đầu vào cũng như trực trặc về công nghệ thông tin.⁶ Tuy nhiên, mô hình hiện có vẫn phụ thuộc hoàn toàn vào việc nhập tay các thông số đo, chưa thể trích xuất trực tiếp dữ liệu từ file ảnh hoặc PDF chưa đồng bộ được các thông tin kỹ thuật của phép đo và chưa thiết lập được liên kết trực tiếp giữa dữ liệu hình ảnh, kết quả đo và bệnh án điện tử. Vì vậy, để giải quyết các khoảng trống trên, nghiên cứu tiến hành với mục tiêu: Khảo sát mức độ hài lòng của người dùng và đánh giá hiệu quả (dữ liệu, thời gian, chi phí, độ chính xác) ứng dụng công nghệ nhận dạng ký tự quang học trong tự động đồng bộ và nhận định kết quả đo CNHH tại Bệnh viện Đại học Y Hà Nội.

II. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP

1. Đối tượng

- **Đánh giá định tính:** 08 bác sĩ, kỹ thuật viên phụ trách đo CNHH; 10 bác sĩ điều trị, 14 điều dưỡng chăm sóc người bệnh hô hấp; 08 chuyên viên nghiên cứu khoa học tập trung phản ánh trải nghiệm và mức độ hài lòng của người dùng.

- **Đánh giá định lượng:** Chọn mẫu toàn bộ 11.778 bản đo CNHH trong giai đoạn nghiên cứu; chỉ các bản đo đạt tiêu chuẩn kỹ thuật theo ATS/ERS và đã được bác sĩ kiểm duyệt mới được đưa vào phân tích. 03 bác sĩ chuyên khoa có trên 05 năm kinh nghiệm trong lĩnh vực đo và phân tích kết quả CNHH ghi nhận và so sánh thời gian thao tác ở các cấp độ chuyển đổi số, đồng thời đối chiếu độ chính xác của hệ thống TTNT với bản đọc chuẩn, đánh giá liên tục phát hiện và hiệu chỉnh kịp thời các sai số trong quá trình ứng dụng.

Tiêu chuẩn loại trừ

- (1) Phiếu mờ, nhòe, nghiêng hoặc sai mẫu;
- (2) Tập ảnh/PDF lỗi hoặc không mở được;
- (3) Bản đo không đạt chuẩn kỹ thuật;
- (4) Phiếu thiếu dữ liệu thiết yếu.

2. Phương pháp

Thiết kế nghiên cứu

Mô tả cắt ngang.

Thời gian và địa điểm

Từ tháng 4/2025 đến tháng 9/2025 tại Khoa Thăm dò chức năng, Bệnh viện Đại học Y Hà Nội.

Quy trình nghiên cứu

- **Lựa chọn công cụ đánh giá:** Việc lựa chọn các biến số và phương pháp đánh giá đáp ứng yêu cầu của Luật Khoa học, Công nghệ và Đổi mới sáng tạo 2025 (số 93/2025/QH15), Quyết định 2244/QĐ-TTg năm 2025 về Bộ tiêu chí đánh giá hiệu quả khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số, đồng thời tích hợp khung đánh giá quốc tế ELICIT (Evaluation in Life Cycle of Information Technology).⁷⁻⁹ Khung này được tùy chỉnh phù hợp với đặc điểm hệ thống y tế và cơ sở hạ tầng Việt Nam, bảo đảm đánh giá toàn diện các đổi mới y tế tích hợp với bệnh án điện tử, bao quát bốn pha của vòng đời công nghệ thông tin: lập kế hoạch - phát triển - triển khai - vận hành và ba cấp độ xã hội - người dùng - công nghệ.

- **Đánh giá định tính:** Phỏng vấn, thu thập, phân tích phản hồi của 40 chuyên gia về mức độ hài lòng của người dùng trong quá trình triển khai và vận hành hệ thống theo thang điểm Likert 5 mức.

- **Đánh giá định lượng:** 03 bác sĩ chuyên khoa tính toán, đánh giá độc lập các chỉ số hiệu quả về thời gian, dữ liệu và tỷ lệ lỗi qua 4 giai đoạn trưởng thành chuyển đổi số:

- Thủ công hoàn toàn: nhập tay thông tin người bệnh, các thông số đo từ file ảnh/PDF, phân tích kết quả và viết kết luận.

- Điện tử hóa cơ bản: tích hợp HIS một số trường thông tin người bệnh.

- Tích hợp hệ thống tự động hóa nhận định kết quả CNHH.

- Ứng dụng AI OCR trích xuất trực tiếp dữ liệu từ file ảnh/PDF; tự động nhận dạng, đồng bộ các thông số kết quả, đối chiếu quy chuẩn lâm sàng, trích xuất kết luận, đồng bộ trực tiếp HIS/EMR.

Mô tả kỹ thuật OCR-DL:

Hệ thống sử dụng mô hình OCR-DL dựa trên CNN để nhận dạng văn bản và thông số đo từ ảnh/PDF. Thuật toán tự động phát hiện vùng chứa thông tin thực (ROI - Region of Interest) bằng cơ chế phân vùng và hộp giới hạn, sau đó trích xuất các trường dữ liệu tương ứng. Các biến số được ánh xạ sang mã LOINC và chuẩn hóa theo cấu trúc FHIR Observation, trước khi

đồng bộ vào HIS/EMR thông qua pipeline HL7/FHIR.

- *Kiểm định, xác thực:* nhóm nghiên cứu so sánh và kiểm định hai lần độ nhất quán giữa các nhóm đánh giá, nhằm đảm bảo tính tin cậy, khách quan và độ chính xác của kết quả.

Biến số

- *Biến số định tính:* Bộ câu hỏi phỏng vấn theo thang Likert 5 mức độ (1-Rất không đồng ý, 2-Không đồng ý, 3- Bình thường, 4- Đồng ý, 5- Rất đồng ý) thuộc 5 khía cạnh định tính:

- (1) thân thiện.
- (2) tích hợp.
- (3) nhanh chóng.
- (4) chính xác.
- (5) sẵn sàng.

Bảng 1. Bảng khảo sát người dùng về hệ thống ứng dụng TTNT tự động đồng bộ và nhận định kết quả đo CNHH

STT	Nội dung đánh giá	Thang điểm
1	Giao diện thân thiện, bố cục rõ ràng, thao tác đơn giản, dễ sử dụng.	
2	Hệ thống tích hợp tốt với HIS/EMR, hoạt động ổn định và không làm gián đoạn quy trình chuyên môn.	
3	Hệ thống nhận diện, phân tích, hiển thị kết quả đo CNHH nhanh chóng, rút ngắn thời gian làm việc.	
4	Kết quả chính xác, đáng tin cậy, hỗ trợ hiệu quả cho bác sĩ lâm sàng.	
5	Hệ thống đạt mức hoàn thiện cao, sẵn sàng chia sẻ, triển khai tại các bệnh viện vệ tinh và cơ sở y tế liên kết.	

- Biến số định lượng được chia thành 4 nhóm:

(1) *Biến số điểm dữ liệu (chi tiết tại Phụ lục 1):*

- + Thông tin người bệnh
- + Thông tin dịch vụ kỹ thuật
- + Truy xuất kết quả
- + Tiền sử và yếu tố nguy cơ
- + Chỉ số CNHH
- + Kết luận CNHH

+ Chuẩn hóa và kết nối

+ Thông số điểm ảnh

(2) *Biến số thời gian:*

- + Tải dữ liệu lên HIS
- + Nhận diện, hiển thị thông số
- + Nhận định, hiển thị kết luận
- + Tạo phiếu báo cáo kết quả
- + In và trả kết quả

(3) Biến số chi phí:

- + Tiết kiệm do giảm thời gian làm việc
- + Tiết kiệm do giảm in ấn, lưu trữ, tài liệu

hóa

- + Tổng chi phí tiết kiệm

(4) Các lỗi sai gặp phải:

- + Tải nhầm file lên HIS
- + Không nhận diện, hiển thị thông số
- + Nhận diện, hiển thị sai thông số
- + Không trích xuất, hiển thị kết luận
- + Trích xuất, hiển thị sai kết luận
- + Sai chính tả khi viết kết luận
- + Không tạo phiếu báo cáo kết quả

Xử lý và phân tích số liệu

Dữ liệu được xử lý bằng các phương pháp thống kê mô tả sử dụng phần mềm SPSS 20.0. Các biến thời gian được trình bày dưới dạng giá trị trung bình $\pm$ độ lệch chuẩn (mean $\pm$ SD), các biến tỷ lệ lỗi và độ chính xác được mô tả bằng tỷ lệ phần trăm. Độ chính xác được tính theo bản đồ, trong đó một bản đồ chỉ được tính là chính xác khi toàn bộ các trường dữ liệu được

hệ thống trích xuất trùng khớp tuyệt đối với bản đọc chuẩn; bất kỳ sai lệch nào ở một trường dữ liệu đều được phân loại là không chính xác.

3. Đạo đức nghiên cứu

Nghiên cứu đã được sự chấp thuận của Hội đồng đạo đức Trường Đại học Y Hà Nội (IRB00003121, Quyết định 563/GCN-HĐĐĐNCYSH-ĐHYHN ngày 30/9/2021). Nghiên cứu tuân thủ các quy định về đạo đức y sinh học và bảo mật thông tin y tế, mọi dữ liệu người bệnh đều được mã hóa, ẩn danh và chỉ sử dụng cho mục đích nghiên cứu khoa học. Việc quản lý và lưu trữ dữ liệu tuân theo Điều 10 Thông tư 46/2018/TT-BYT, Nghị định 53/2022/NĐ-CP và tiêu chuẩn quốc tế ISO/IEC 27799:2016 về an ninh thông tin y tế.

III. KẾT QUẢ**1. Giao diện làm việc**

Giao diện làm việc được thiết kế sẵn sàng cho việc nhập liệu, phân tích, lưu trữ và trích xuất kết quả đo CNHH, được tích hợp trực tiếp trên hệ thống HIS và bệnh án điện tử như hình 1.

The screenshot displays the OCR-DL system interface. It includes a sidebar with navigation options like 'Thông tin bệnh nhân', 'Kết quả', and 'Bệnh án'. The main area shows patient information (Name: Nguyễn Văn A, ID: 123456789) and a table of test results. The table has columns for 'STT', 'NỘI DUNG', and 'KẾT QUẢ'. Below this, there's a section for 'KẾT QUẢ' with a table showing various test results and their percentages. At the bottom, there's a 'TỔNG KẾT' section with a table summarizing the overall results.

STT	NỘI DUNG	KẾT QUẢ
1	Chẩn đoán lâm sàng	100
2	Chẩn đoán hình ảnh	97
3	Chẩn đoán phân tử	100

Chỉ số	% so với tham chiếu	Chỉ số	% so với tham chiếu
Đúng tích vùng (SWC)	99	Lưu lượng dòng (PEV)	99
Đúng tích vùng thời gian (FVC)	99	Lưu lượng thời gian 25% FVC	99
Thời gian tích vùng và trung gian đầu tiên (FEV1)	97	Lưu lượng thời gian 50% FVC	99
Chỉ số Gensler (FVC/FEV1)	99	Lưu lượng thời gian 75% FVC	99
Chỉ số Tiffeneau (FVC/FEV1)	99	Lưu lượng thời gian 25-75% FVC	99

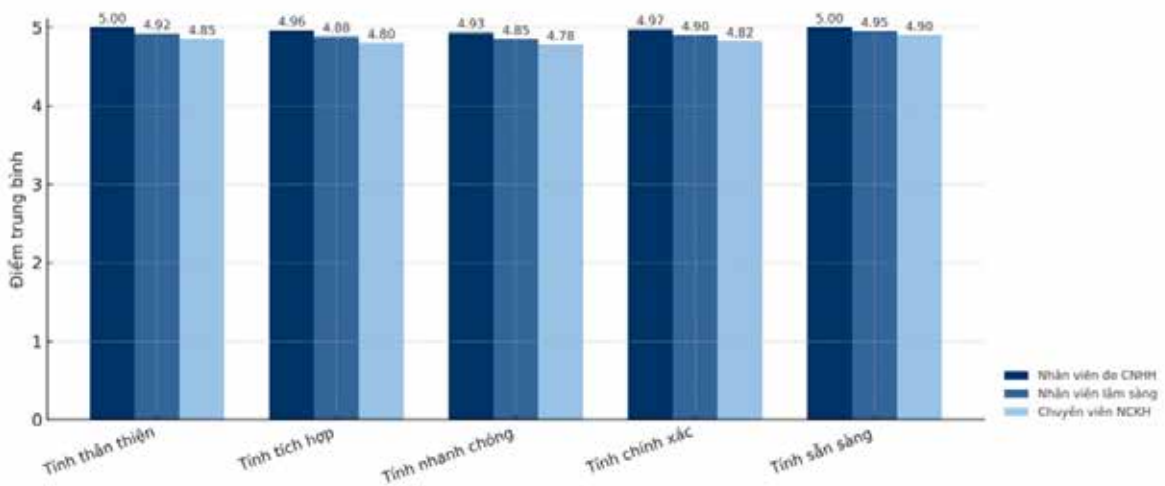
Chỉ số	% so với tham chiếu	Chỉ số	% so với tham chiếu
FEV1 sau thuốc	99	FEV1 sau thuốc	99
% thay đổi	99	% thay đổi	99
Chỉ số Gensler	99	Chỉ số Gensler	99

Hình 1. Hình minh họa giao diện người dùng hệ thống OCR-DL tự động đồng bộ và nhận định kết quả đo CNHH

Đánh giá khảo sát người dùng về hệ thống tự động đồng bộ và nhận định kết quả đo CNHH

Khảo sát được thực hiện trên 40 người dùng thuộc 03 nhóm đối tượng: 08 nhân viên thăm dò CNHH (20,0%), 24 nhân viên lâm sàng (60,0%), và 08 chuyên viên nghiên cứu khoa học (20,0%), độ tuổi trung bình dao động từ 26 đến 54. Biểu đồ 1 minh họa mức độ hài lòng rất cao của người dùng đối với hệ thống ở cả 5 khía cạnh, với điểm trung bình dao động từ

4,78 đến 5,0 trên thang 5 điểm. Trong đó, “Tính thân thiện” và “Tính sẵn sàng” đạt mức cao nhất (5,0 điểm ở nhóm nhân viên đo CNHH, 4,95 ở nhóm lâm sàng và 4,90 ở nhóm nghiên cứu khoa học). Các tiêu chí “Tính tích hợp”, “Tính nhanh chóng” và “Tính chính xác” đều đạt trên 4,8 điểm, cho thấy hệ thống hoạt động ổn định, dễ sử dụng, đảm bảo tốc độ xử lý và độ chính xác cao, đồng thời có khả năng thích ứng tốt trong môi trường lâm sàng thực tế.



Biểu đồ 1. Kết quả khảo sát người dùng về hệ thống (n = 40)

3. Đánh giá hiệu quả ứng dụng TTNT tự động đồng bộ và nhận định kết quả đo CNHH

Hiệu quả về xây dựng bộ dữ liệu

Bảng 2 trình bày hiệu quả dữ liệu qua từng giai đoạn chuyển đổi số.

Bảng 2. Số điểm dữ liệu được tích hợp, lưu trữ (n = 11.778)

Loại dữ liệu/bản đo	Thủ công (1)	Thủ công +HIS (2)	Thủ công +HIS+tự động (3)	Thủ công +HIS+tự động + TTNT (4)
Thông tin người bệnh	0	12	12	12
Thông tin dịch vụ kỹ thuật	0	0	11	11
Truy xuất kết quả	0	0	4	4
Tiền sử, yếu tố nguy cơ	0	0	14	14
Chỉ số CNHH	0	0	20	20

Loại dữ liệu/bản đo	Thủ công (1)	Thủ công +HIS (2)	Thủ công +HIS+tự động (3)	Thủ công +HIS+tự động + TTNT (4)
Kết luận CNHH	0	0	3	3
Dữ liệu chuẩn hóa, kết nối	0	0	0	2.560
Thông số điểm ảnh	0	0	0	179.200
Điểm dữ liệu được lưu trữ/bản đo (n = 1)	0	11	63	181.824
Tổng dữ liệu được lưu trữ (n = 11.778)	0	129.558	742.014	2,14 tỷ

Bảng 2 cho thấy số lượng và loại dữ liệu được lưu trữ tăng dần qua các giai đoạn phát triển hệ thống. Ở giai đoạn thủ công hoàn toàn, không có dữ liệu nào được số hóa hoặc lưu lại trên hệ thống. Khi tích hợp HIS, ghi nhận 12 biến thông tin người bệnh và dịch vụ kỹ thuật, tổng cộng 129.558 điểm dữ liệu được lưu trữ/năm. Ở giai đoạn tự động hóa, dữ liệu được mở rộng với 63 biến, bao gồm các thông số đo và kết luận CNHH, đạt 742.014 điểm dữ liệu/năm.

Khi ứng dụng OCR-DL, mỗi bản đo CNHH tạo ra 181.824 điểm dữ liệu, bổ sung dữ liệu chuẩn hóa, dữ liệu kết nối và các thông số điểm ảnh. Với 11.778 bản đo/năm, tổng số điểm dữ liệu được lưu trữ đạt xấp xỉ 2,14 tỷ điểm dữ liệu mỗi năm.

Hiệu quả về thời gian

Thời gian nhập liệu, nhận định kết quả, tạo phiếu trả lời CNHH qua các giai đoạn được thể hiện ở bảng 3.

Bảng 3. Hiệu quả về thời gian khi ứng dụng hệ thống (n = 11.778)

Thời gian trung bình/ bản đo (phút)	Thủ công (1)	Thủ công + HIS (2)	Thủ công + HIS + tự động (3)	Thủ công + HIS + tự động + TTNT (4)
Tải dữ liệu lên HIS	(-)	2,1 ± 0,4	0,6 ± 0,2	0,3 ± 0,1
Nhận diện, hiển thị thông số	10,4 ± 1,6	5,3 ± 0,7	3,2 ± 0,6	0,2 ± 0,1
Nhận định, hiển thị kết luận	5,2 ± 1,1	5,1 ± 1,0	0,3 ± 0,1	0,2 ± 0,1
Tạo phiếu báo cáo kết quả	5,3 ± 1,0	2,1 ± 0,4	0,6 ± 0,2	0,3 ± 0,1
In và trả kết quả	3,2 ± 0,5	3,1 ± 0,5	(-)	(-)
Thời gian trung bình/ bản đo	24,2 ± 3,8	17,7 ± 2,9	4,7 ± 1,0	1,0 ± 0,2
Tổng (n = 11.778)	285.027	208.470	55.357	11.778
Tiết kiệm 1-4		273.249		
Tiết kiệm 2-4		196.692		
Tiết kiệm 3-4		43.579		

(-): Không thực hiện

Ở giai đoạn thủ công, tổng thời gian trung bình xử lý một bản đo là $24,2 \pm 3,8$ phút, giảm còn $17,7 \pm 2,7$ phút khi tích hợp HIS, $4,7 \pm 1,0$ phút ở giai đoạn tự động hóa, và rút ngắn 24 lần chỉ còn $1,0 \pm 0,2$ phút khi ứng dụng OCR-DL. Tổng thời gian xử lý toàn bộ 11.778 bản đo giảm từ 285.027 phút xuống 11.778 phút,

tiết kiệm 4.554 giờ làm việc. Ở giai đoạn 3 và 4, quy trình đã được tích hợp giữa thiết bị đo, phân tích và bệnh án điện tử, cho phép tự động đồng bộ, hiển thị và lưu trữ trên HIS/EMR, loại bỏ hoàn toàn bước in và trả kết quả thủ công.

Hiệu quả về chi phí

Bảng 4. Hiệu quả về chi phí khi ứng dụng hệ thống (n = 11.778)

Nội dung	Giá trị	Công thức tính
Tổng số lượt đo CNHH	11.778	Số liệu thống kê thực tế tại bệnh viện
Tổng thời gian tiết kiệm (giờ)	4.554	$(24,2 - 1,0) \times 11.778 / 60$
Hệ số lương cơ bản	2,34 - 6,92	Theo quy định Nhà nước
Mức lương cơ bản 2024	2,35 triệu VNĐ	Theo quy định Nhà nước
Chi phí tiết kiệm do giảm thời gian làm việc	$\approx 157,0 - 462,5$ triệu VNĐ	Thời gian tiết kiệm ÷ giờ công/tháng $\times$ hệ số lương $\times$ lương cơ bản
Chi phí in ấn	23,6 - 41,2 triệu VNĐ	Trung bình 2.000 - 3.500 VNĐ/ca, bao gồm giấy, mực và khấu hao máy in
Chi phí nhân lực lưu trữ hồ sơ	17,7 - 23,6 triệu VNĐ	Trung bình 1.500 - 2.000 VNĐ/ca, tương đương 2 - 3 phút thao tác/ca
Chi phí kho lưu trữ và vật tư bảo quản	5,9 - 9,4 triệu VNĐ	500 - 800 VNĐ/phiếu, tính theo diện tích và vật tư bảo quản.
Chi phí tiết kiệm do giảm in ấn, lưu trữ, tài liệu hóa	$\approx 47,2 - 74,2$ triệu VNĐ	Chi phí in ấn + Chi phí nhân lực lưu trữ hồ sơ + Chi phí kho lưu trữ và vật tư bảo quản
Tổng chi phí tiết kiệm (n = 11.778)	$\approx 204,2 - 536,7$ triệu VNĐ	Chi phí tiết kiệm thời gian + chi phí tiết kiệm do giảm in ấn, lưu trữ, tài liệu hóa

Bảng 4 cho thấy hiệu quả kinh tế khi ứng dụng OCR-DL tự động đồng bộ và nhận định kết quả đo CNHH. Tổng chi phí tiết kiệm đạt 204,2 - 536,7 triệu VNĐ/6 tháng, trong đó chi phí tiết kiệm do giảm thời gian làm việc chiếm tỷ trọng

lớn hơn (157 - 462,5 triệu VNĐ, tương đương 77-86%) so với chi phí tiết kiệm do giảm in ấn, lưu trữ và tài liệu hóa (47,2 - 74,2 triệu VNĐ).

Hiệu quả về độ chính xác

Bảng 5. Tỷ lệ lỗi và độ chính xác tổng thể (n = 11.778)

Tên lỗi	Thủ công (1)	Thủ công +HIS (2)	Thủ công +HIS+tự động (3)	Thủ công +HIS+tự động +TTNT (4)
Tải nhầm file lên HIS	(-)	3%	2%	0%
Không nhận diện, hiển thị thông số	(-)	(-)	(-)	0%
Nhận diện, hiển thị sai thông số	8%	5%	1%	0%
Không trích xuất, hiển thị kết luận	(-)	(-)	0%	0%
Trích xuất, hiển thị sai kết luận	5%	5%	0%	0%
Sai chính tả khi viết kết luận	10%	8%	0%	0%
Không tạo phiếu báo cáo kết quả	5%	4%	0%	0%
Tổng số lỗi sai	28%	25%	3%	0%
Độ chính xác tổng thể	72%	75%	97%	100%

(-): Không thực hiện

Ở giai đoạn thủ công, tổng tỷ lệ lỗi là 28%, giảm còn 25% khi tích hợp HIS, 3% ở giai đoạn tự động hóa và 0% khi ứng dụng TTNT. Các lỗi thường gặp ở giai đoạn đầu như nhập sai thông số, sai chính tả khi viết kết luận được loại bỏ hoàn toàn ở giai đoạn 4. Độ chính xác tổng thể của hệ thống đạt 100%.

IV. BÀN LUẬN

Nghiên cứu đã phát triển công nghệ OCR-DL cho phép tự động chuyển đổi dữ liệu từ file ảnh/PDF kết quả đo CNHH thành thông tin có cấu trúc. Khác với nhiều mô hình quốc tế chủ yếu tập trung vào thuật toán dự đoán rối loạn thông khí mà chưa giải quyết được bài toán dữ liệu, hệ thống có ưu điểm nổi bật ở khả năng xử lý trực tiếp dữ liệu non-DICOM, dạng dữ liệu phổ biến nhưng khó tích hợp trong thực hành lâm sàng.³⁻⁵ Hệ thống tự động nhận dạng, trích xuất và chuẩn hóa dữ liệu đo CNHH, đồng thời đồng bộ và lưu trữ trực tiếp vào HIS/EMR, khép kín toàn bộ chu trình dữ liệu. Với khả năng tương thích cao với nhiều định dạng kết quả

khác nhau, dữ liệu được đồng bộ ngay khi hình ảnh được tải lên, không cần thao tác nhập tay từ nhân viên y tế. Giải pháp có thể tích hợp vào phần mềm nội viện, không đòi hỏi hạ tầng công nghệ phức tạp, dễ dàng mở rộng đến các cơ sở y tế tuyến dưới.

Kết quả khảo sát cho thấy mức độ hài lòng cao, thể hiện sự chấp nhận và tin tưởng mạnh mẽ của người dùng đối với hệ thống. Phần lớn các nghiên cứu quốc tế đều tập trung vào hiệu năng thuật toán do đó ít báo cáo kết quả trải nghiệm người dùng.³⁻⁵ Trái lại, nghiên cứu của chúng tôi khảo sát 40 người dùng qua 5 khía cạnh định tính, điểm trung bình 4,78 - 5,0/5 cung cấp bằng chứng sự hài lòng của người dùng thực tế mà nhiều nghiên cứu trước đây còn thiếu. 03 nhóm tham gia đánh giá được bố trí hợp lý, đại diện cho toàn bộ các khâu trong quy trình khám, điều trị và quản lý, phản ánh toàn diện trải nghiệm sử dụng và mức độ phù hợp của hệ thống ở cả góc độ chuyên môn, vận hành và kỹ thuật.

Nghiên cứu cho thấy ở giai đoạn thủ công

hoàn toàn, toàn bộ kết quả đo CNHH được ghi chép và xử lý bằng tay, không có dữ liệu nào được số hóa hay lưu lại trên hệ thống, dẫn đến không tận dụng được giá trị dữ liệu thứ cấp cho nghiên cứu. Khi tích hợp HIS, quy trình bắt đầu điện tử hóa một phần, lưu trữ các trường hành chính cơ bản như mã hồ sơ, mã bệnh án, tuổi, giới... song vẫn hạn chế về độ sâu và tính liên thông dữ liệu. Đến giai đoạn tự động hóa nhận định kết quả đo CNHH, hệ thống bắt đầu thu nhận dữ liệu đa lớp, bao gồm thông tin hành chính, chỉ số đo và kết luận lâm sàng, giúp tăng số lượng điểm dữ liệu lưu trữ trong năm từ 129.558 lên 742.014, đồng thời chuẩn hóa kết quả đo và tự động sinh nhận định. Khi tích hợp OCR-DL, toàn bộ dữ liệu đầu vào phi cấu trúc được chuyển đổi thành thông tin có cấu trúc, mỗi biến được mã hóa theo 40 thuộc tính LOINC, liên thông tạo thành 2.560 mối tương quan dữ liệu. Với cỡ mẫu 11.778 bản đo, hệ thống ghi nhận và lưu trữ hơn 2,14 tỷ điểm dữ liệu/năm, tăng gần 2.900 lần so với giai đoạn tự động hóa, phản ánh năng lực mở rộng vượt trội về chiều sâu và chất lượng dữ liệu. Mô hình góp phần hiện thực hóa Nghị quyết 57-NQ/TW về khoa học, đổi mới sáng tạo và chuyển đổi số, cụ thể hóa định hướng “dữ liệu là tài nguyên số”, hình thành kho dữ liệu lớn, có cấu trúc về CNHH, tạo tiền đề mở rộng ứng dụng TTNT trong quản lý bệnh mạn tính và phát triển y tế thông minh.¹⁰

Thế mạnh nổi bật của hệ thống là khả năng rút ngắn đáng kể thời gian thao tác và tiết kiệm chi phí vận hành. Kết quả nghiên cứu tại Bệnh viện Đại học Y Hà Nội cho thấy thời gian xử lý trung bình mỗi bản đo giảm từ $24,2 \pm 3,8$ phút xuống còn $1,0 \pm 0,2$ phút, nhanh hơn khoảng 24 lần, tương ứng tiết kiệm 4.554 giờ làm việc cho 11.778 lượt đo. Từ đó, chi phí tiết kiệm trực tiếp ước tính đạt 204,2 - 536,7 triệu đồng/năm. Tương tự nghiên cứu của Topalovic và cộng sự (2019) chứng minh hiệu quả của TTNT trong

diễn giải kết quả đo CNHH rút ngắn thời gian đọc và ra quyết định từ vài phút xuống còn 0,2 giây.³ Das và cộng sự (2020) cũng ghi nhận mô hình học sâu theo tiêu chuẩn ATS/ERS giúp chuẩn hóa quy trình đánh giá, rút ngắn 90% thời gian kiểm định chất lượng phép đo.⁴

Độ chính xác tổng thể của hệ thống đạt 100% thể hiện sự tin cậy gần như tuyệt đối với kết quả thu được trong toàn bộ chu trình vận hành đầy đủ các khâu (nhập liệu-phân tích-nhận định-đồng bộ-lưu trữ). Các lỗi thường gặp trước đây như nhận diện sai thông số, sai chính tả khi gõ kết luận đã được loại bỏ hoàn toàn. Tương tự, Topalovic và cộng sự (2019) báo cáo hệ thống TTNT đạt độ chính xác 100% trong nhận dạng mẫu giãn đồ, vượt xa kết quả của bác sĩ chuyên khoa hô hấp ($44,6 \pm 8,7\%$; $k = 0,35$, $p < 0,001$).³ Das và cộng sự (2020) ghi nhận mô hình học sâu đạt độ chính xác tổng thể 92%, độ nhạy 92% và độ đặc hiệu 96% trong kiểm định chất lượng phép đo.⁴

Bằng cách tích hợp trực tiếp trên hệ thống HIS/EMR, toàn bộ dữ liệu CNHH được chuẩn hóa, số hóa và lưu trữ có hệ thống, thuận tiện cho tra cứu, truy xuất, chia sẻ và nghiên cứu, đồng thời hình thành kho dữ liệu lớn có khả năng mở rộng và nâng cấp ứng dụng trong tương lai.

HẠN CHẾ

Nghiên cứu thực hiện tại một trung tâm và sử dụng một cấu hình thiết bị đo KOKO, nên khả năng khái quát hóa sang các dòng máy khác cần được thẩm định thêm. Bên cạnh đó, nghiên cứu chủ yếu đánh giá hiệu quả vận hành trước mắt, chưa phân tích tác động lâm sàng như thay đổi quy trình chăm sóc hoặc ảnh hưởng đến kết quả điều trị. Các nghiên cứu đa trung tâm và thử nghiệm trên nhiều cấu hình thiết bị khác sẽ cần thiết để đánh giá toàn diện giá trị ứng dụng của hệ thống trong thực hành lâm sàng rộng rãi.

V. KẾT LUẬN

Nghiên cứu đánh giá hiệu quả ứng dụng CNTT trong tự động đồng bộ và nhận định kết quả đo CNHH. 100% người dùng đánh giá cao hệ thống ở cả 5 khía cạnh: tính thân thiện, khả năng tích hợp, tính nhanh chóng, độ chính xác và tính sẵn sàng. Thời gian xử lý mỗi bản đo giảm 24 lần, tiết kiệm 4.554 giờ làm việc, tương ứng giảm 204,2 - 536,7 triệu đồng chi phí nhân lực, vận hành; với 2,14 tỷ điểm dữ liệu được lưu trữ.

TÀI LIỆU THAM KHẢO

1. Stanojevic S, Graham BL, Cooper BG, et al. Official ERS technical standard: Global Lung Function Initiative reference values for static lung volumes in individuals of European ancestry. *Eur Respir J.* 2022; 60(1): 2101499. doi:10.1183/13993003.01499-2021.
2. Osagie E, Ayo-Ogbor S. Challenges in the Design of Optical Character Recognition for Medical Image Modalities. *International Journal of Multidisciplinary Research Configuration.* 07/15 2022; 4: 109-112.
3. Topalovic M, Das N, Burgel PR, Daenen M, Derom E, Haenebalcke C, et al. Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. *Eur Respir J.* 2019; 53(4): 1801660. doi:10.1183/13993003.01660-2018.
4. Das N, Topalovic M, Janssens W. Deep-learning based automated quality control of spirometry data according to ATS/ERS standards. *Eur Respir J.* 2020; 56(6): 2000603. doi:10.1183/13993003.00603-2020.
5. Giri PC, Chowdhury AM, Bedoya A, Chen H, Lee HS, Lee P, Henriquez C, MacIntyre NR and Huang Y-CT. Application of Machine Learning in Pulmonary Function Assessment Where Are We Now and Where Are We Going? *Front. Physiol.* 2021; 12: 678540. doi: 10.3389/fphys.2021.678540.
6. Bùi Mỹ Hạnh. Nghiên cứu xây dựng và triển khai hệ thống hỗ trợ chẩn đoán tự động lưu trữ và trả lời kết quả đo chức năng hô hấp cho hệ thống quản lý thông tin y tế. *Tạp chí Y học Việt Nam.* 2020; 497: 26-35.
7. Quốc hội. Luật số 93/2025/QH15: Luật Khoa học, công nghệ và đổi mới sáng tạo. 2025.
8. Thủ tướng chính phủ. Quyết định 2244/QĐ-TTg về Bộ tiêu chí đánh giá hiệu quả khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số đối với phát triển kinh tế-xã hội. 13/10/2025.
9. Kukhareva PV, Weir C, Del Fiol G, et al. Evaluation in Life Cycle of Information Technology (ELICIT) framework: Supporting the innovation life cycle from business case assessment to summative evaluation. *Journal of biomedical informatics.* Mar 2022; 127: 104014. doi:10.1016/j.jbi.2022.104014.
10. Bộ Chính trị, Nghị quyết số 57-NQ/TW của Bộ Chính trị về đột phá phát triển khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số quốc gia. 22/12/2024.

Summary

EFFECTIVENESS OF OPTICAL CHARACTER RECOGNITION IN AUTOMATED SYNCHRONIZATION AND INTERPRETATION OF PULMONARY FUNCTION TEST RESULTS AT HANOI MEDICAL UNIVERSITY HOSPITAL

This study evaluated the effectiveness of applying optical character recognition in the automated synchronization and interpretation of pulmonary function test (PFT) results. A total of 40 specialists conducted both qualitative and quantitative analyses of the system using 11,778 PFT reports collected at Hanoi Medical University Hospital from April 2025 to September 2025. Results showed that 100% of users rated the system highly across five dimensions: user-friendliness, integration capability, processing speed, accuracy, and operational readiness. The average processing time per report was reduced 24-fold, saving 4,554 working hours and an estimated 204.2 - 536.7 million VND in labor and operational costs, while the system stored more than 2.14 billion data points. The OCR-DL system demonstrated outstanding performance in automating the entire workflow - from data acquisition and analysis to storage - contributing to data standardization, error reduction, and enhanced efficiency in medical data management and utilization.

Keywords: Pulmonary function test, automation, optical character recognition, deep learning, artificial intelligence, efficiency.