TỰ ĐỘNG HÓA XỬ LÝ HỒ SƠ VÀ TÀI LIỆU TRONG QUẢN LÝ ĐÀO TẠO DỰA TRÊN NHẬN DẠNG KÝ TỰ QUANG HỌC, NHẬN DẠNG CHỮ VIẾT TAY VÀ MÔ HÌNH NGÔN NGỮ LỚN

Nguyễn Hữu Toàn[1*], Đặng Hải Đăng[1], Lưu Tiến Trung[1], Võ Thị Linh Trà[1]
[*]Email: toannh.elc@hou.edu.vn. ORCID: 0009-0006-3341-6346

Ngày tòa soạn nhận được bài báo: 18/05/2026
Ngày phản biện đánh giá: 02/07/2026
Ngày bài báo được duyệt đăng: 20/07/2026
DOI: 10.59266/houjs.2026.1360

Tóm tắt: *Nghiên cứu này đề xuất một quy trình tự động hóa xử lý hồ sơ, tài liệu nhằm đáp ứng yêu cầu chuyển đổi số trong quản trị đại học, đặc biệt ở tầng nghiệp vụ. Trên cơ sở tổng quan các nghiên cứu về nhận dạng ký tự quang học (OCR), nhận dạng chữ viết tay (HTR), Document AI và xử lý tài liệu có cấu trúc, nghiên cứu phân tích đặc thù nghiệp vụ nhập liệu, lưu trữ và đối chiếu dữ liệu đối với văn bằng, chứng chỉ, bảng điểm và danh sách thi trong môi trường giáo dục đại học. Từ đó, nhóm tác giả xây dựng một mô hình xử lý theo kiến trúc mô-đun, tích hợp nhận dạng văn bản, mô hình ngôn ngữ lớn (LLM), cơ chế hậu xử lý và khâu đối chiếu có giám sát của con người nhằm bảo đảm khả năng vận hành theo quy trình khép kín từ tiếp nhận tài liệu đến chuẩn hóa dữ liệu đầu ra. Kết quả thử nghiệm trên dữ liệu nghiệp vụ thực tế cho thấy mô hình có tiềm năng giảm đáng kể khối lượng nhập liệu thủ công, đồng thời nâng cao tính nhất quán, khả năng truy xuất và hiệu quả xử lý hồ sơ. Đóng góp chính của nghiên cứu không nằm ở việc đề xuất thuật toán nhận dạng mới, mà ở việc thiết kế và thử nghiệm một qui trình xử lý (pipeline) tích hợp đầu cuối (end-to-end) phù hợp với điều kiện quản lý đào tạo tại cơ sở giáo dục đại học Việt Nam.*

Từ khóa: *mô hình ngôn ngữ lớn (LLM), nhận dạng ký tự quang học (OCR), nhận dạng chữ viết tay (HTR), quản lý đào tạo, xử lý hồ sơ, chuyển đổi số*

I. Đặt vấn đề

1.1. Bối cảnh nghiên cứu

Trong những năm gần đây, chuyển đổi số trong giáo dục đại học thường được nhấn mạnh ở các hệ thống trên lớp giao diện như cổng thông tin sinh viên, hệ thống quản lý đào tạo trực tuyến, diễn đàn học tập và các nền tảng học tập (LMS). Tuy nhiên, ở tầng vận hành, nhiều quy trình cốt lõi vẫn phụ thuộc vào tài liệu giấy: bảng điểm được in và lưu trữ vật lý, danh sách thi kết thúc học phần với phần điểm số do giảng viên ghi tay, văn bằng được cấp và lưu ở dạng bản cứng hoặc bản sao có công chứng. Điều này tạo ra một khoảng cách đáng kể giữa hiển thị số hóa

[1] Trường Đại học Mở Hà Nội, Hà Nội, Việt Nam

trên các trạng thái vật lý của các tài liệu liên quan.

Khoảng cách đó thể hiện rõ nhất ở khâu nhập liệu và đối chiếu dữ liệu. Giảng viên hoàn thành chấm thi trên phiếu giấy, sau đó cán bộ quản lý hoạt động đào tạo phải nhập lại điểm vào hệ thống; các thông tin trên văn bằng / chứng chỉ được nhập tay và lưu trữ trên các bảng Excel rồi mới cập nhật sang hệ thống quản lý. Quy trình nhiều bước này không chỉ tốn thời gian và nhân lực mà còn gia tăng nguy cơ sai sót, khiến dữ liệu khó đạt mức chính xác và nhất quán cần thiết cho phân tích, kiểm định và ra quyết định (Barchard và Pace, 2011).

1.2. Tổng quan nghiên cứu về OCR/HTR và trí tuệ nhân tạo xử lý tài liệu (Document AI)

Sự phát triển của các mô hình học sâu, đặc biệt là kiến trúc CNN, RNN và Transformer, đã giúp hệ thống nhận dạng chữ viết tay đạt độ chính xác cao trên nhiều bộ dữ liệu và ngôn ngữ khác nhau (Retsinas và cộng sự, 2022). Bên cạnh đó, sự xuất hiện của các mô hình thị giác-ngôn ngữ như TrOCR, Donut và LayoutLMv3 đã mở ra hướng tiếp cận mới trong bài toán hiểu tài liệu có cấu trúc, không chỉ nhận dạng ký tự mà còn trích xuất và tổ chức lại thông tin theo ngữ cảnh (Kim và cộng sự, 2022; Li và cộng sự, 2022; Huang và cộng sự, 2022).

Nhiều nghiên cứu hiện nay tập trung vào việc cải tiến kiến trúc mô hình nhận dạng chữ viết tay, đặc biệt là các mô hình kết hợp giữa CNN và RNN, cùng với hàm mất mát CTC. Đây là hướng tiếp cận kinh điển và vẫn giữ vai trò nền tảng trong phần lớn hệ thống HTR hiện đại. Một trong những công trình có ảnh hưởng lớn nhất là nghiên cứu của Graves và cộng sự (2006), trong đó nhóm tác giả giới thiệu phương pháp CTC, cho phép mô hình RNN xử lý các chuỗi đầu vào không được phân đoạn sẵn.

Một nhánh nghiên cứu khác tập trung vào tự động chấm bài, đặc biệt là chấm câu trả lời ngắn của sinh viên dựa trên nội dung ngữ nghĩa. Hướng này không trực tiếp xử lý chữ viết tay, nhưng liên quan chặt chẽ đến việc đánh giá nội dung sau khi văn bản đã được số hóa. Tổng quan toàn diện nhất về lĩnh vực này được trình bày trong nghiên cứu của Burrows và cộng sự (2015), mô tả các xu hướng, phương pháp và thách thức trong chấm bài tự động.

Một hướng nghiên cứu quan trọng khác tập trung vào số hóa các tài liệu lịch sử, tài liệu hành chính và những kho lưu trữ quy mô lớn. Các công trình trong lĩnh vực này thường xoay quanh việc phân tích bố cục tài liệu, xây dựng bộ dữ liệu gán nhãn (ground truth) và thiết kế các quy trình OCR/HTR hoàn chỉnh để xử lý lượng tài liệu đa dạng và phức tạp. Trong số đó, Aletheia là một ví dụ tiêu biểu. Công cụ này, được mô tả chi tiết trong nghiên cứu của Clausner và cộng sự (2011), hỗ trợ mạnh mẽ việc tạo dữ liệu gán nhãn và quản lý bố cục tài liệu. Nhờ tính linh hoạt và độ chính xác cao, Aletheia đã trở thành lựa chọn phổ biến trong nhiều dự án số hóa quy mô lớn trên thế giới.

Mặc dù các hệ thống OCR/HTR đã đạt những tiến bộ đáng kể, việc ứng dụng các công nghệ này trong bối cảnh quản lý đào tạo tại Việt Nam vẫn còn hạn chế. Các tài liệu đặc thù như danh sách thi có điểm ghi tay, văn bằng và chứng chỉ với trường viết tay không đồng nhất, hay bản quét có dấu xác nhận chồng lấn, đặt ra những

thách thức riêng mà các hệ thống xử lý tài liệu thông dụng chưa giải quyết triệt để.

Khoảng trống nghiên cứu không chủ yếu nằm ở thuật toán nhận dạng ký tự đơn lẻ, mà ở sự thiếu vắng các pipeline xử lý đầu cuối tích hợp phân tích bố cục, nhận dạng nội dung, trích xuất thông tin có cấu trúc và kết nối với hệ thống quản lý đào tạo. Nghiên cứu này tập trung lấp khoảng trống đó thông qua việc thiết kế và thử nghiệm một quy trình tích hợp phù hợp với dữ liệu nghiệp vụ thực tế trong giáo dục đại học, thay vì đặt mục tiêu đề xuất một kiến trúc nhận dạng mới.

II. Cơ sở lý thuyết

2.1. Các công nghệ OCR / HTR thời kỳ đầu

Trong bối cảnh lịch sử phát triển của OCR và HTR, các kiến trúc dựa trên CNN, RNN và hàm mất mát CTC đã đóng vai trò như nền tảng kỹ thuật chủ đạo trong hơn một thập niên, giúp hệ thống đạt được độ chính xác cao trên nhiều bộ dữ liệu và ngôn ngữ khác nhau (Graves và cộng sự, 2006; Shi và cộng sự, 2017). Các mô hình kinh điển như CRNN tận dụng CNN để trích xuất đặc trưng không gian từ ảnh văn bản, RNN (thường là LSTM hoặc BiLSTM) để mô hình hóa quan hệ tuần tự giữa các bước trong chuỗi và CTC để cho phép huấn luyện trên các chuỗi chưa được căn chỉnh tường minh ở mức ký tự (Graves và cộng sự, 2006; Shi và cộng sự, 2017). Tuy nhiên, khi yêu cầu về khả năng xử lý chuỗi dài, tài liệu nhiều dòng và tích hợp ngữ cảnh ngôn ngữ ngày càng tăng, các hạn chế cố hữu của kiến trúc tuần tự dựa trên RNN dần bộc lộ, tạo điều kiện cho sự xuất hiện của các mô hình mới dựa trên cơ chế chú ý (attention) và Transformer, trong đó cơ chế chú ý (self-attention) dần thay thế vai trò trung tâm của RNN trong việc mô hình hóa phụ thuộc dài hạn (Vaswani và cộng sự, 2017)

2.2. Attention trong kiến trúc Transformer

Cơ chế attention được giới thiệu trong công trình của Vaswani và cộng sự (2017) và trở thành nền tảng của hầu hết các mô hình ngôn ngữ và thị giác hiện đại. Về mặt tính toán, mỗi vector đầu vào được chiếu qua ba ma trận trọng số để tạo ra ba vector tương ứng là Query (Q), Key (K) và Value (V); khi xét trên toàn bộ chuỗi đầu vào gồm nhiều token, ba tập biểu diễn này được tổ chức thành các ma trận Q, K, V. Điểm attention được tính theo công thức chuẩn hóa:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

trong đó d_k là số chiều của vector khóa, được dùng để điều chỉnh độ lớn của tích vô hướng nhằm tránh hiện tượng giá trị gradient bị bão hòa trước khi qua hàm softmax (Vaswani và cộng sự, 2017).

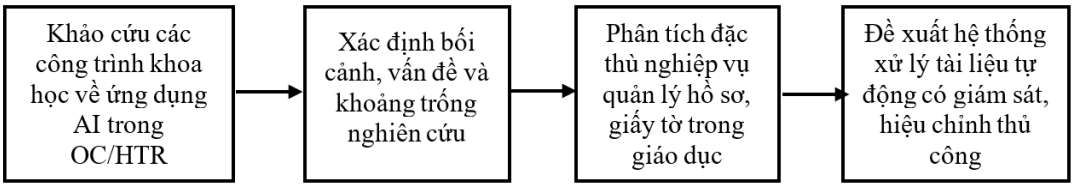
Trong phạm vi nghiên cứu này, cơ chế attention không được phân tích như một thành phần nội tại có thể can thiệp ở mức mô hình, mà được trình bày như cơ sở lý thuyết để lý giải khả năng mô hình ngôn ngữ lớn hỗ trợ trích xuất và tổ chức lại thông tin từ dữ liệu OCR/HTR. Bài báo khai thác mô hình ngôn ngữ lớn ở vai trò hỗ trợ diễn giải ngữ cảnh, chuẩn hóa trường dữ liệu và giảm sai lệch cục bộ sau bước nhận dạng và truy cập thông qua giao diện API. Vì vậy, phần cơ sở lý thuyết được giới hạn ở mức làm rõ nguyên lý liên quan, không đi sâu vào các chi tiết kiến trúc nội bộ của dịch vụ mô hình không

phải là đối tượng kiểm soát trực tiếp của nghiên cứu.

III. Phương pháp nghiên cứu

3.1. Quy trình nghiên cứu

Quy trình nghiên cứu được xây dựng theo trình tự từ nhận diện vấn đề thực tiễn, tổng quan các công trình về OCR, HTR và xử lý tài liệu bằng AI, đến đề xuất mô hình và triển khai thử nghiệm. Trên cơ sở phân tích nghiệp vụ quản lý đào tạo, nghiên cứu xác định các nhóm tài liệu trọng tâm và các điểm nghẽn trong nhập liệu, từ đó hình thành yêu cầu đối với một pipeline xử lý có khả năng vận hành thực tế. Chuỗi xử lý được thiết kế qua các bước: tiếp nhận tài liệu; phân loại biểu mẫu; nhận dạng văn bản in và chữ viết tay; xây dựng nhắc lệnh (prompt) theo từng loại biểu mẫu để hướng dẫn mô hình ngôn ngữ lớn trích xuất và tổ chức thông tin; hậu xử lý và chuẩn hóa dữ liệu; đối chiếu thủ công; đồng bộ vào hệ thống quản lý đào tạo.



Hình 1. Quy trình nghiên cứu

Bên cạnh tổng quan tài liệu và phân tích nghiệp vụ, nghiên cứu sử dụng phương pháp thiết kế-phát triển hệ thống gắn với thử nghiệm ứng dụng trên dữ liệu thực. Chuỗi xử lý AI được tổ chức theo hướng kết hợp nhận dạng tự động với kiểm soát của con người trước khi ghi nhận dữ liệu, nhằm bảo đảm độ tin cậy nghiệp vụ. Tổng quan tài liệu trong nghiên cứu này được thực hiện theo phương pháp khảo cứu có định hướng.

Trên cơ sở thiết kế đó, hệ thống được xây dựng thử nghiệm và triển khai trên dữ liệu thực tế để đo lường hiệu quả xử lý. Các chỉ báo sử dụng trong đánh giá bao gồm mức độ nhận dạng và điền đúng trường thông tin (được xác định thông qua đối chiếu với dữ liệu nguồn), thời gian xử lý, và chi phí ước tính. Đây là đánh giá theo tiếp cận ứng dụng và phản ánh hiệu quả vận hành trên dữ liệu triển khai cụ thể; trong khi các chỉ số học thuật như Character Error Rate (CER) và Word Error Rate (WER) được ghi nhận là hướng đánh giá cần bổ sung khi có bộ dữ liệu gán nhãn chuẩn hóa hơn.

IV. Kết quả và bàn luận

4.1. Phân tích nghiệp vụ xử lý OCR / HTR

Mô hình xử lý OCR/HTR được phân tích như một chuỗi các khối chức năng chính sau:

Tiếp nhận dữ liệu đầu vào: thu nhận ảnh chụp, bản scan hoặc tệp PDF của hồ sơ, giấy tờ.

Phân loại tài liệu: nhận diện nhóm biểu mẫu như bảng điểm, văn bằng hoặc chứng chỉ để lựa chọn hướng xử lý phù hợp.

Lựa chọn prompt xử lý: định hướng quá trình nhận dạng và trích xuất thông tin theo từng loại tài liệu.

Nhận dạng OCR/HTR: trong quy trình triển khai thực tế, hệ thống sử dụng Gemini API kết hợp langchain-google-genai để đồng thời nhận dạng văn bản và

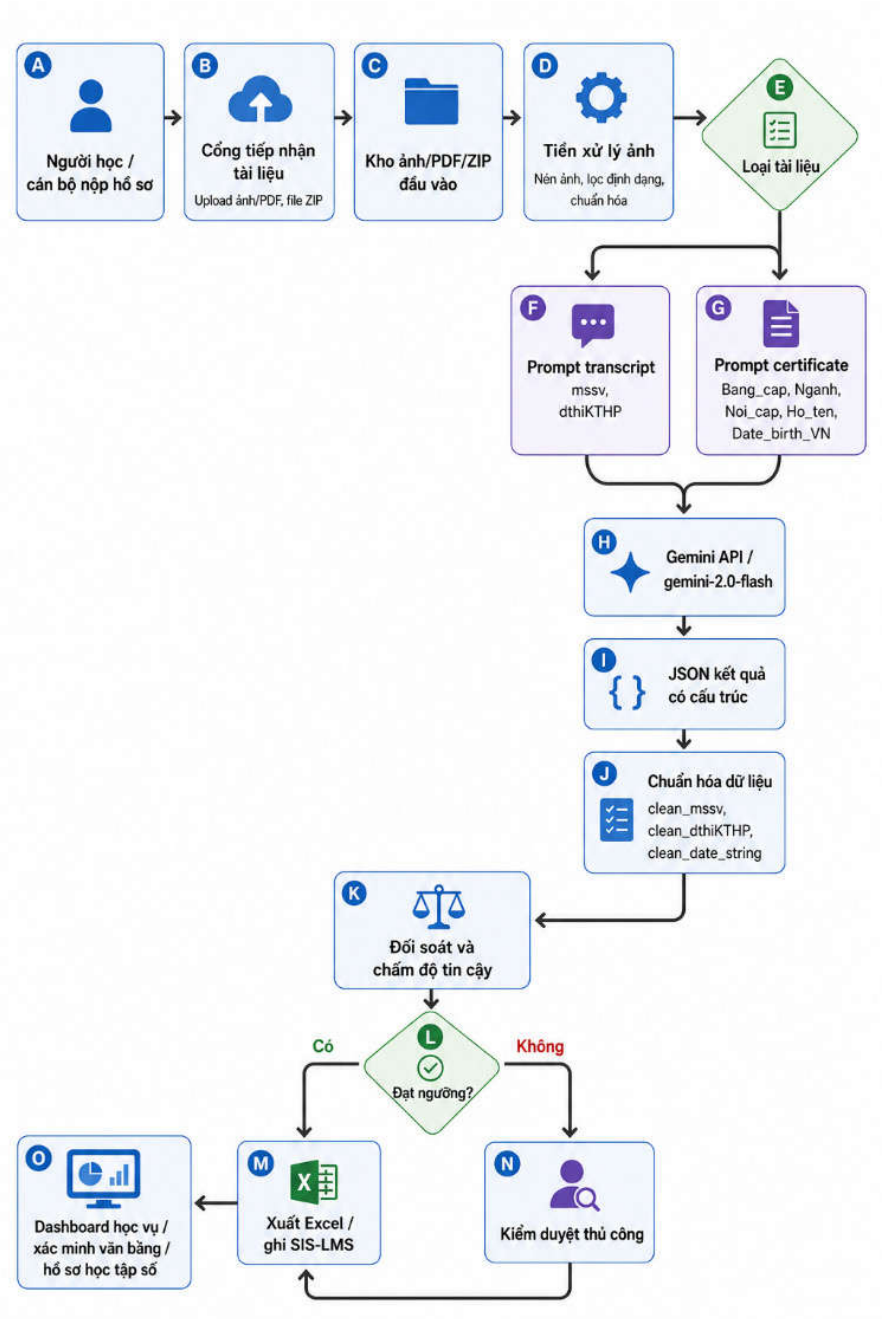
trích xuất có cấu trúc các trường thông tin; Google Vision được sử dụng trong một số trường hợp thử nghiệm bổ sung để so sánh kết quả nhận dạng thuần túy trước khi đưa qua bước suy luận ngữ cảnh.

Hậu xử lý và chuẩn hóa: hiệu chỉnh định dạng, đối sánh ngữ cảnh và tổ chức lại dữ liệu theo các trường thông tin cần thiết.

Hỗ trợ suy luận ngữ nghĩa: dùng mô hình ngôn ngữ lớn để hỗ trợ diễn giải kết quả nhận dạng và giảm sai lệch cục bộ.

Kiểm tra thủ công: rà soát và xác nhận dữ liệu trước khi tích hợp vào hệ thống quản lý đào tạo.

Cách phân tách này giúp làm rõ mối quan hệ giữa đầu vào, quá trình xử lý và đầu ra của mô hình đề xuất.

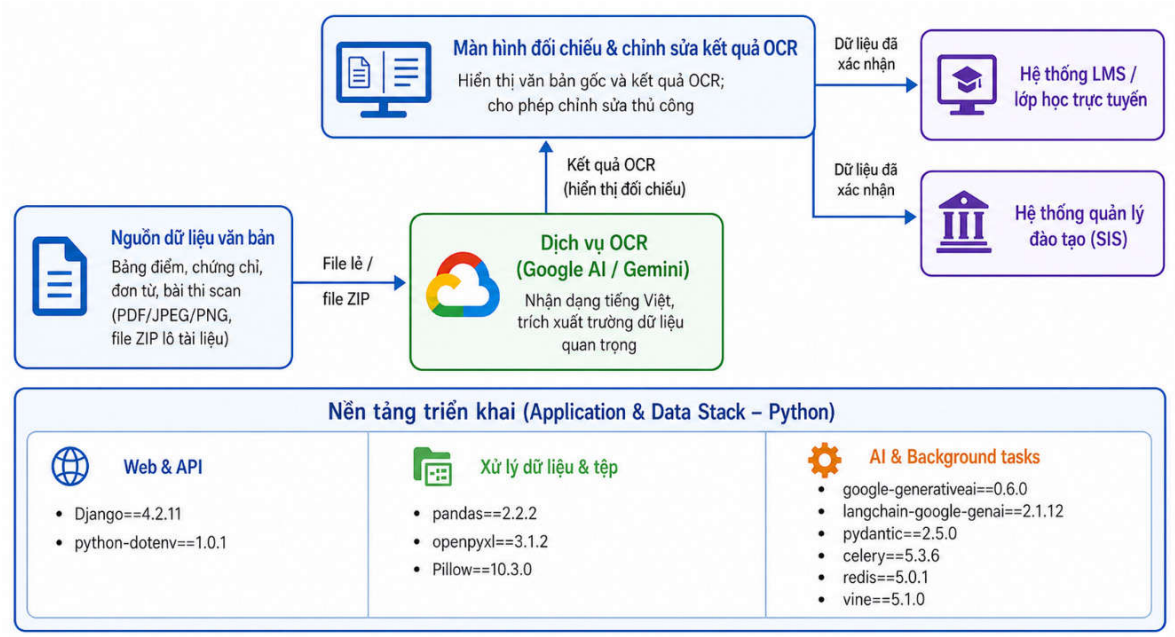


Hình 2. Mô hình phân tích nghiệp vụ xử lý OCR / HTR

4.2. Kết quả phát triển hệ thống

Hệ thống được xây dựng với bốn nhóm chức năng nhận dạng chính: văn bằng/chứng chỉ (loại chỉ có chữ in; loại kết hợp chữ in và chữ viết tay), căn cước công dân, giấy phép lái xe, và danh sách thi có điểm thi kết thúc học phần do giảng viên ghi tay. Công nghệ lõi là Gemini API với khả năng xử lý đa phương thức, đảm nhiệm đồng thời vai trò nhận dạng văn bản và trích xuất thông tin có cấu trúc từ ảnh tài liệu.

Hệ thống web được phát triển bằng Django; các tệp văn bằng và hồ sơ tiếng Việt (ảnh riêng lẻ hoặc lô ZIP chứa ảnh/ PDF) được nạp lên và đẩy vào hàng đợi xử lý nền thông qua Celery với Redis làm trình môi giới thông điệp (message broker). Đối với mỗi tài liệu, hệ thống tự động phân loại biểu mẫu và lựa chọn nhắc lệnh phù hợp trước khi gọi Gemini API, nhằm hướng dẫn mô hình trích xuất đúng các trường thông tin theo cấu trúc được định nghĩa trước bằng Pydantic.



Hình 3. Kiến trúc hệ thống xử lý OCR/ HTR kèm nền tảng phát triển

Trong lớp xử lý nền, các tác vụ sử dụng Pillow để tiền xử lý ảnh, Pandas và Openpyxl để chuẩn hóa, tổng hợp dữ liệu, sau đó gọi API Gemini thông qua thư viện Google-generativeai kết hợp Langchain-Google-genai để nhận dạng và trích xuất có cấu trúc các trường thông tin quan trọng của văn bằng, dữ liệu được ràng buộc cấu trúc dữ liệu (schema) bằng Pydantic. Kết quả OCR / HTR trước khi ghi nhận chính thức được đưa lên một màn hình đối chiếu, nơi nhân viên nhập liệu có thể quan sát song song văn bản gốc và kết quả nhận dạng, chỉnh sửa thủ công những lỗi còn lại và xác nhận.

Chỉ sau khi được xác nhận, dữ liệu đã hiệu chỉnh mới được đồng bộ tự động vào hệ thống LMS (hệ thống quản lý học tập) và hệ thống thông tin đào tạo (SIS), phục vụ cập nhật điểm, điều kiện tốt nghiệp và lưu trữ hồ sơ văn bằng. Toàn bộ mô hình ứng dụng vận hành trên nền tảng Python với stack Django 4.2.11, Celery 5.3.6, Redis 5.0.1,

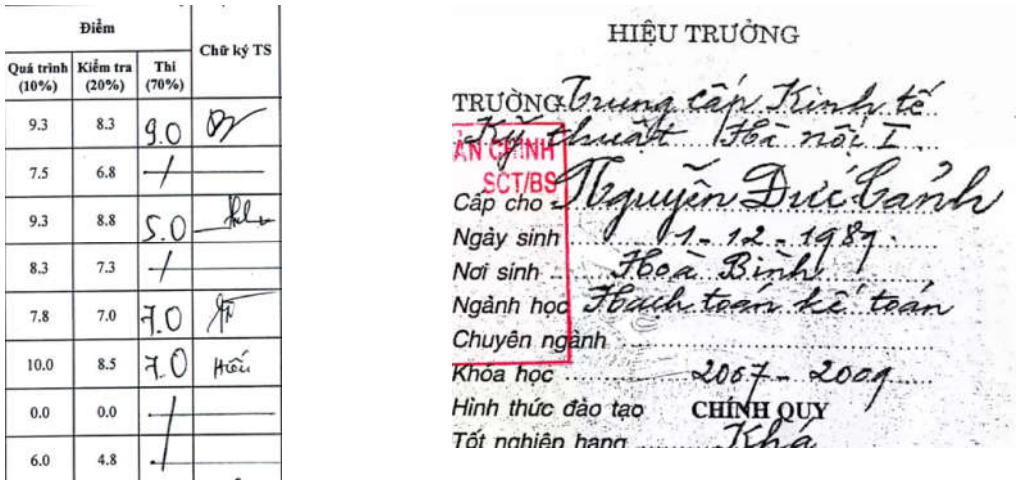
pandas 2.2.2, openpyxl 3.1.2, Pillow 10.3.0, google-generativeai 0.6.0, langchain-google-genai 2.1.12, pydantic 2.5.0 và python-dotenv 1.0.1 cho cấu hình bảo mật, cho phép xử lý lô tài liệu quy mô lớn mà vẫn đảm bảo độ tin cậy và khả năng kiểm soát của con người trong quy trình xác thực hồ sơ, tài liệu.

4.3. Đánh giá và thử nghiệm

4.3.1. Độ chính xác nhận dạng

Bảng 1. Độ chính xác nhận dạng theo từng loại hồ sơ, tài liệu

Loại hồ sơ / giấy tờ	Số lượng	Tỉ lệ % chính xác	Lỗi nhận dạng gặp phải
Văn bằng / chứng chỉ (chữ in)	102	100%	Không có
Văn bằng / chứng chỉ (viết tay)	112	99.1%	Dấu thanh không rõ
Căn cước công dân	88	100%	Không có
Danh sách thi	215	99.5%	Ký tự giảng viên tự quy ước
Bằng lái xe	118	100%	
Tổng số	**635**	**> 99%**	



Hình 4. Hai trường hợp nhận dạng chưa chính xác trên bảng điểm, văn bằng

Trong quá trình nhận dạng văn bằng, hệ thống gặp khó khăn chủ yếu khi xử lý các mẫu có chữ viết tay, được quét từ bản sao, hoặc lẫn dấu xác nhận sao y khiến đặc trưng thị giác bị nhiễu. Đối với nhận dạng bảng điểm và danh sách thi, sai sót thường phát sinh do giảng viên sử dụng các nét gạch không tuân thủ quy ước, làm giảm tính nhất quán của dữ liệu đầu vào. Cần lưu ý rằng kết quả độ chính xác được báo cáo phản ánh hiệu quả xử lý ở cấp trường thông tin sau khâu đối chiếu nghiệp vụ, không phải chỉ số lỗi ký tự (CER) hay lỗi từ (WER) đo trên bộ dữ liệu gán nhãn độc lập. Vì vậy, các kết quả này cần được hiểu trong giới hạn của nghiên cứu ứng dụng và là cơ sở để tiếp tục thiết kế đánh giá chặt chẽ hơn ở nghiên cứu tiếp theo.

4.3.2. Chi phí và thời gian xử lý

Kết quả thử nghiệm cho thấy toàn bộ quy trình xử lý 635 ảnh hồ sơ, giấy tờ được hoàn thành trong khoảng 35 phút, bao gồm thời gian tiếp nhận dữ liệu, xếp hàng tác vụ, nhận dạng, hậu xử lý và đối chiếu kết quả. Ở cấu hình thử nghiệm hiện tại, hệ thống xử lý theo lô với quy mô tối đa khoảng 100 ảnh cho mỗi tệp nén tại một thời điểm, qua đó bảo đảm sự ổn định của hàng đợi xử lý và khả năng kiểm soát

chất lượng đầu ra trong môi trường vận hành thực tế.

Xét theo cấu hình thử nghiệm, khối lượng xử lý đối với bảng điểm được ước lượng ở mức khoảng 300 - 320 token cho mỗi ảnh, với đầu ra trung bình khoảng 10 token; trong khi đó, đối với văn bằng và chứng chỉ có đầu ra gồm khoảng 5 trường thông tin cơ bản như họ tên, ngày sinh, tên văn bằng hoặc chứng chỉ, hạng tốt nghiệp và năm cấp bằng, khối lượng xử lý được ước lượng vào khoảng 350 token cho mỗi ảnh, với đầu ra trung bình khoảng 25 - 40 token. Theo mức giá tham khảo của mô hình Google Gemini 2.5 Flash, khoảng 0,30 USD cho 1 triệu token đầu vào và 2,50 USD cho 1 triệu token đầu ra, chi phí thực tế phụ thuộc vào cơ cấu token và cấu hình gọi mô hình. Trên cơ sở ước lượng của đợt thử nghiệm gồm 635 ảnh, tổng khối lượng xử lý vào khoảng 2 triệu token, trong đó token đầu vào chiếm trên 90%, nên tổng chi phí tính toán chưa đến 1 USD. Cách tính này cho thấy mô hình có khả năng vận hành với chi phí thấp trong bối cảnh xử lý hồ sơ, giấy tờ ở quy mô nghiệp vụ.

Kết quả này cho thấy mô hình đề xuất khả thi về mặt thời gian xử lý và chi phí trong bối cảnh nghiệp vụ. Tuy nhiên, nghiên cứu chưa thực hiện đối sánh có kiểm soát với các hệ thống OCR/HTR khác trên cùng bộ dữ liệu chuẩn hóa. Do đó, kết quả hiện mang giá trị bằng chứng triển khai ứng dụng và là tiền đề cho so sánh hệ thống trong nghiên cứu tiếp theo.

V. Kết luận

5.1. Kết luận chung

Nghiên cứu đã đề xuất và thử nghiệm một quy trình tự động hóa xử lý hồ sơ, tài liệu trong quản lý đào tạo dựa trên sự kết hợp giữa OCR/HTR, mô hình ngôn ngữ lớn và khâu hậu kiểm có giám sát. Trên cơ sở phân tích đặc thù nghiệp vụ trong môi trường giáo dục đại học, mô hình được thiết kế theo hướng xử lý khép kín từ tiếp nhận tài liệu, phân loại biểu mẫu, nhận dạng nội dung, hậu xử lý và chuẩn hóa, đối chiếu thủ công đến đồng bộ dữ liệu vào hệ thống quản lý đào tạo. Kết quả thử nghiệm bước đầu cho thấy hệ thống đạt hiệu quả cao trên nhiều nhóm hồ sơ phổ biến, góp phần rút ngắn thời gian xử lý, giảm khối lượng nhập liệu thủ công và nâng cao tính nhất quán của dữ liệu quản trị.

5.2. Hạn chế và hướng phát triển

Bên cạnh các kết quả đạt được, nghiên cứu vẫn còn một số hạn chế. Thứ nhất, quy mô mẫu thử nghiệm chưa đủ lớn để phản ánh đầy đủ mức độ đa dạng của tài liệu thực tế. Thứ hai, việc đánh giá chủ yếu dựa trên hiệu quả vận hành ở cấp ứng dụng, chưa triển khai các chỉ số học thuật chuẩn như CER, WER hay phép so sánh đối chứng trên bộ dữ liệu gán nhãn độc lập. Thứ ba, chưa thực hiện so sánh hệ thống với các công cụ OCR/HTR khác trên cùng điều kiện thử nghiệm. Trong thời gian tới, nghiên cứu cần mở rộng bộ dữ liệu kiểm thử, xây dựng dữ liệu chuẩn (ground-truth) chuẩn hóa, bổ sung CER/WER và thực hiện so sánh đối chứng nhằm nâng cao tính học thuật và sức thuyết phục của kết quả.

5.3. Đóng góp của nghiên cứu

Đóng góp của nghiên cứu được thể hiện trên ba phương diện chính. Về mặt lý luận, bài báo góp phần hệ thống hóa cơ

sở nghiên cứu về OCR, HTR và xử lý tài liệu bằng AI trong bối cảnh giáo dục đại học. Về mặt phương pháp, nghiên cứu đề xuất và thử nghiệm một quy trình xử lý (pipeline) tích hợp đầu cuối (end-to-end), kết hợp nhận dạng tự động, hậu kiểm bán tự động và kiểm soát nghiệp vụ của con người, phù hợp với đặc thù dữ liệu tại Việt Nam. Về mặt thực tiễn, hệ thống đã được triển khai và cho thấy tiềm năng hỗ trợ tự động hóa nhập liệu, chuẩn hóa hồ sơ và kết nối với hệ thống quản lý đào tạo trong môi trường nghiệp vụ thực tế, tạo tiền đề cho các bước phát triển tiếp theo hướng đến hệ sinh thái quản trị đại học thông minh.

Tài liệu tham khảo

Barchard, K. A., và Pace, L. A. (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. *Computers in Human Behavior, 27*(5), 1834-1839. https://doi.org/10.1016/j.chb.2011.04.004

Burrows, S., Gurevych, I., và Stein, B. (2015). The Eras and Trends of Automatic Short Answer Grading. *International Journal of Artificial Intelligence in Education, 25*(1), 60-117. https://doi.org/10.1007/s40593-014-0026-8

Clausner, C., Pletschacher, S., và Antonacopoulos, A. (2011). Aletheia - An advanced document layout and text ground-truthing system for production environments. In *2011 International Conference on Document Analysis and Recognition (ICDAR)* (pp. 48-52). IEEE. https://www.primaresearch.org/www/assets/papers/ICDAR2011_Clausner_Aletheia.pdf

Gemini Team Google. (2023). *Gemini: A family of highly capable multimodal models* (arXiv:2312.11805). arXiv. https://doi.org/10.48550/arXiv.2312.11805

Gemini Team Google. (2024). *Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context*. arXiv:2403.05530. https://doi.org/10.48550/arXiv.2403.05530

Graves, A., Fernández, S., Gomez, F., và Schmidhuber, J. (2006). Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd International Conference on Machine Learning* (pp. 369-376). https://doi.org/10.1145/1143844.1143891

Huang, Y., Lv, T., Cui, L., Lu, Y., và Wei, F. (2022). LayoutLMv3: Pre-training for document AI with unified text and image masking. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22)* (pp. 4083-4091). ACM. https://doi.org/10.1145/3503161.3548112

Kim, G., Hong, T., Yim, M., Nam, J., Park, J., Yim, J., Hwang, W., Yun, S., Han, D., và Park, S. (2022). OCR-Free document understanding transformer. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, và T. Hassner (Eds.), *Computer Vision - ECCV 2022* (Lecture Notes in Computer Science, Vol. 13688, pp. 498-517). Springer. https://doi.org/10.1007/978-3-031-19815-1_29

Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., Zhang, C., Li, Z., và Wei, F. (2023). TrOCR: Transformer-based optical character recognition with pre-trained models. *Proceedings of the AAAI Conference on Artificial Intelligence, 37*(11), 13094-13102. https://doi.org/10.1609/aaai.v37i11.26538

Retsinas, G., Sfikas, G., Gatos, B., và Nikou, C. (2022). Best practices for a handwritten text recognition system. In *Document Analysis Systems (DAS 2022)* (Lecture Notes in Computer Science, Vol. 13237, pp. 247-259). https://doi.org/10.1007/978-3-031-06555-2_17

Shi, B., Bai, X., và Yao, C. (2017). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(11), 2298-2304. https://doi.org/10.1109/TPAMI.2016.2646371

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., và Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (Vol. 30). https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

AUTOMATING DOCUMENT PROCESSING IN EDUCATIONAL RECORDS MANAGEMENT USING OCR, HTR, AND LARGE LANGUAGE MODELS

Nguyen Huu Toan[1], Dang Hai Dang[1], Luu Tien Trung[1], Vo Thi Linh Tra[1]

Abstract: *This study proposes an automated document processing pipeline to address digital transformation requirements at the operational level of higher education institutions. Building on a review of research in optical character recognition (OCR), handwritten text recognition (HTR), Document AI, and structured document processing, the study analyzes the specific workflow of data entry, storage, and verification for academic credentials, transcripts, and examination lists. A modular system architecture is proposed, integrating text recognition, large language model (LLM)-based information extraction, post-processing, and human-in-the-loop verification within a closed-loop pipeline from document ingestion to structured data output. The main contribution lies not in a novel recognition algorithm, but in the design and end-to-end integration of a pipeline suited to the data characteristics of Vietnamese higher education administration. Pilot results on 635 real-world documents demonstrate the system's potential for reducing manual data entry workload, improving data consistency, and enabling seamless integration with learning management and student information systems.*

Keywords: *large language model (LLM), optical character recognition (OCR), handwritten text recognition (HTR), records management, digital transformation*

[1] Hanoi Open University, Hanoi, Vietnam