SO SÁNH CÁC GIẢI PHÁP OCR TRONG NHẬN DẠNG VĂN BẰNG TIẾNG VIỆT VÀ ĐỀ XUẤT ỨNG DỤNG THỰC TIỄN

Đặng Hải Đăng[1*], Lưu Tiến Trung[1]
[*]Tác giả liên hệ, Email: dangdh@hou.edu.vn

Ngày tòa soạn nhận được bài báo: 02/06/2025
Ngày phản biện đánh giá: 08/09/2025
Ngày bài báo được duyệt đăng: 15/10/2025

DOI: 10.59266/houjs.2025.720

Tóm tắt: *Công nghệ nhận dạng ký tự quang học (OCR) là một công cụ hiện đại để trích xuất văn bản từ hình ảnh, đem lại hiệu quả đáng kể trong việc tự động hóa nhận dạng văn bằng. Nghiên cứu này thử nghiệm bốn phương pháp OCR phổ biến và đưa ra các khuyến nghị triển khai thực tế, đó là DeepSeek, Meta AI's LLaMA-3.2-11B-Vision-Instruct-Turbo, EasyOCR và API Gemini của Google, trong việc nhận dạng văn bản từ văn bằng tiếng Việt. Sau khi thực hiện xây dựng, cài đặt và triển khai các giải pháp tập trung vào ba tiêu chí: độ chính xác, tính ổn định và khả năng thực tiễn cho thấy DeepSeek không hỗ trợ ngôn ngữ tiếng Việt, EasyOCR yêu cầu xác định tọa độ thủ công, trong khi API Gemini và LLaMA đạt hiệu suất vượt trội với độ chính xác cao cho văn bằng, chứng chỉ có sử dụng chữ viết tay, trong đó LLaMA là nền tảng mã nguồn mở, không phụ thuộc nhà cung cấp. Nghiên cứu cho thấy triển khai OCR sử dụng LLaMA là giải pháp tối ưu nhất tính cả trên phương diện thời gian, công sức và nguồn lực tài chính.*

Từ khóa: *nhận dạng ký tự quang học (OCR), nhận dạng văn bản, tiếng Việt, số hóa tài liệu, xử lý hình ảnh*

I. Đặt vấn đề

Nhận dạng ký tự quang học (OCR) là công nghệ chuyển đổi hình ảnh văn bản thành dữ liệu để máy có thể đọc được, giúp cho việc số hóa tài liệu được triển khai nhanh chóng, chính xác. Trong lĩnh vực giáo dục, OCR đã được đưa vào để chấm điểm trắc nghiệm khách quan trong hầu hết các kỳ thi từ tuyển sinh các cấp học, thi hết học phần cho tới tốt nghiệp trung học phổ thông hỗ trợ cho việc chấm, trả điểm được xử lý nhanh chóng, chính xác. Riêng đối với tài liệu học thuật như bằng tốt nghiệp, chứng chỉ, bảng điểm

[1] Viện Đào tạo và Phát triển Học tập Suốt đời, Trường Đại học Mở Hà Nội

thường chứa cả văn bản chữ in và chữ viết tay gây khó khăn cho các hệ thống OCR do đặc điểm ngôn ngữ tiếng Việt với các dấu thanh và chữ viết tay đa dạng [1]. Việc tự động hóa nhận dạng văn bản (NDVB) không chỉ giúp rút ngắn thời gian và công sức mà còn giúp tối ưu hóa quy trình hành chính trong các tổ chức giáo dục và cơ quan nhà nước [2].

Tuy nhiên, việc áp dụng OCR cho văn bằng tiếng Việt đặt ra thách thức riêng. Một mặt, ngôn ngữ tiếng Việt có hệ thống dấu thanh phức tạp, dễ gây sai lệch trong nhận diện. Mặt khác, nhiều văn bằng chứa thông tin viết tay với nét chữ đa dạng, chất lượng scan không đồng nhất, dẫn tới khó khăn trong xử lý. Các phương pháp OCR truyền thống như Tesseract hay EasyOCR cho thấy hiệu quả nhất định, nhưng chưa được đánh giá toàn diện trong trường hợp văn bằng tiếng Việt có chữ viết tay.

Trước khoảng trống đó, nghiên cứu này lựa chọn triển khai bốn giải pháp OCR hiện đại—DeepSeek, EasyOCR, LLaMA3.2 và Gemini API—nhằm so sánh theo ba tiêu chí: độ chính xác, tính ổn định, và khả năng ứng dụng thực tiễn, từ đó đưa ra khuyến nghị lựa chọn giải pháp tối ưu cho các cơ sở giáo dục và đơn vị hành chính tại Việt Nam.

Mục tiêu chính là xác định phương pháp nào có độ chính xác, tính ổn định và tính thực tiễn cao nhất khi trích xuất các trường thông tin như họ tên, ngày sinh, ngày cấp và số văn bằng. Câu hỏi trọng tâm của nghiên cứu là: Làm thế nào để ứng dụng OCR một cách hiệu quả trong nhận dạng văn bằng tiếng Việt, đặc biệt khi xử lý các văn bản chữ viết tay và chữ in, đồng thời đảm bảo tính thực tiễn trong triển khai thực tế? Nghiên cứu này không chỉ góp phần đánh giá và lựa chọn các mô hình OCR hiện đại mà còn cung cấp khuyến nghị cho các tổ chức giáo dục và nhà phát triển công nghệ trong việc lựa chọn và tối ưu hóa công cụ OCR trong tác nghiệp.

II. Tổng quan tài liệu

2.1. Công nghệ OCR và ứng dụng trong số hóa tài liệu

Công nghệ OCR đã phát triển qua nhiều thế hệ, từ các phương pháp dựa trên quy tắc (rule-based) đến các mô hình học sâu (deep learning) và Transformer [3]. Những tiến bộ này đã cải thiện độ chính xác của OCR trong việc NDVB từ hình ảnh, đặc biệt trong các lĩnh vực như số hóa hồ sơ và xác thực danh tính [4]. Tuy nhiên, hiệu suất của OCR bị chi phối bởi nhiều yếu tố như chất lượng hình ảnh, định dạng văn bản và khả năng hỗ trợ ngôn ngữ của mô hình [5].

2.2. Thách thức trong NDVB tiếng Việt

Tiếng Việt có hệ thống dấu thanh phức tạp (ví dụ: á, ả, ã), đòi hỏi các mô hình OCR phải có khả năng nhận diện chính xác các ký tự đặc biệt này [1]. Ngoài ra, sự đa dạng trong phong cách chữ viết tay, đặc biệt trong các văn bằng học thuật, làm tăng độ phức tạp của quá trình nhận dạng [6]. Nhìn chung, các mô

hình OCR truyền thống thường gặp khó khăn khi xử lý chữ viết tay do sự biến đổi về hình dạng ký tự và chất lượng hình ảnh [7].

2.3. Các phương pháp OCR hiện đại

Các mô hình OCR dựa trên học sâu, chẳng hạn như Tesseract, EasyOCR và các API thương mại như Google Cloud Vision, đã được thương mại hóa và sử dụng rộng rãi [8]. Gần đây, các mô hình thị giác-ngôn ngữ (vision-language models) như LLaMA và DeepSeek đã được ứng dụng để NDVB từ hình ảnh, cho độ chính xác cao [9]. Để áp dụng các mô hình này cho tiếng Việt và các ngôn ngữ ít phổ biến khác cần phải được đánh giá toàn diện trong các nghiên cứu tiếp theo [10].

2.4. Ứng dụng OCR trong nhận dạng văn bằng

Việc nhận dạng văn bằng bằng OCR mang lại lợi ích lớn trong tự động hóa quy trình xác minh học thuật, giảm thiểu sai sót thủ công [2]. Mặc dù vậy, các nghiên cứu về OCR trong nhận dạng văn bằng tiếng Việt còn hạn chế, đặc biệt với các văn bằng, chứng chỉ chứa chữ viết tay [1]. Nghiên cứu này tiếp nối các công trình trước, tập trung vào việc đánh giá và lựa chọn các công cụ OCR hiện đại trong bối cảnh văn bằng, chứng chỉ phải xử lý là tiếng Việt.

III. Phương pháp nghiên cứu

3.1. Dữ liệu

Nghiên cứu được thực hiện trên hai tập dữ liệu:

- Tập 1: Gồm 74 hình ảnh văn bằng, chứng chỉ tiếng Việt chữ in, chứa các trường thông tin tiêu chuẩn như họ tên, ngày sinh, ngày cấp, số văn bằng và cơ quan cấp. Các hình ảnh được thu thập từ các văn bằng đại học và chứng chỉ nghề, với độ phân giải từ 300 đến 600 DPI.

- Tập 2: Gồm 63 hình ảnh văn bằng, chứng chỉ tiếng Việt có chứa chữ viết tay, bao gồm các trường thông tin tương tự như Tập 1.

Các văn bằng, chứng chỉ này có sự đa dạng về phong cách viết tay và chất lượng hình ảnh (được scan từ rõ nét đến nhòe nhẹ, có cả bản gốc và bản công chứng).

Cả hai tập dữ liệu được chuẩn hóa để đảm bảo tính nhất quán trong định dạng và độ phân giải, đồng thời được kiểm tra thủ công để xác định văn bản gốc làm cơ sở so sánh.

3.2. Lựa chọn phương pháp triển khai

Bốn phương pháp OCR được lựa chọn để triển khai, bao gồm:

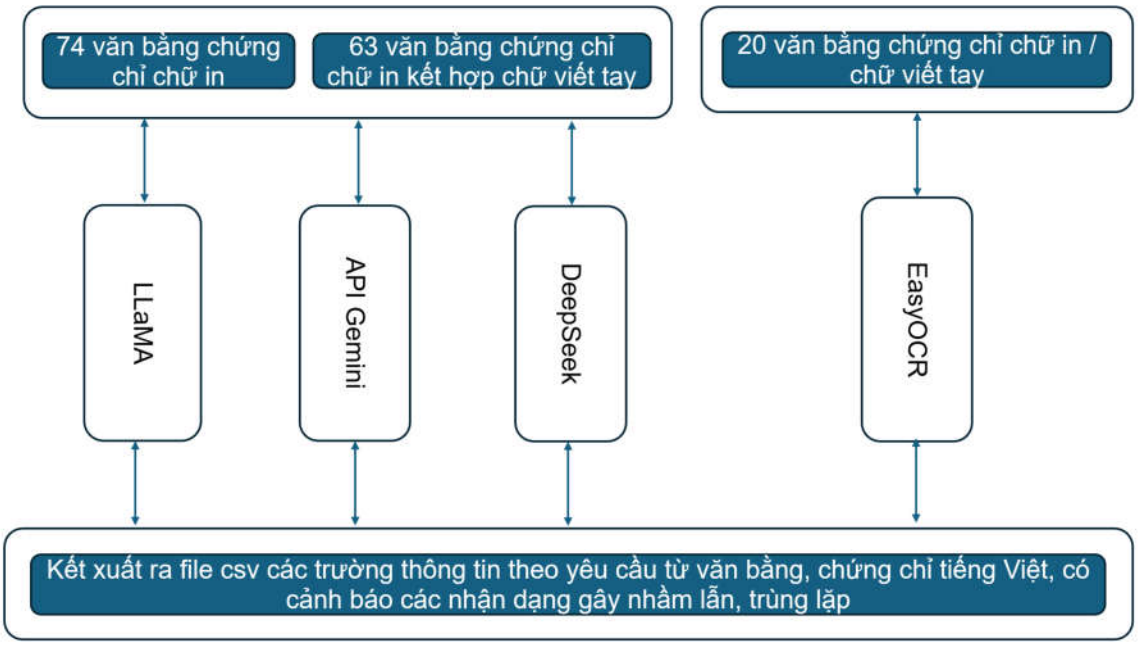
1. DeepSeek: Mô hình thị giác-ngôn ngữ sử dụng lời nhắc (prompt) để xử lý hình ảnh. DeepSeek được thực hiện bằng cách tải lên hình ảnh văn bằng thông qua API, với kỳ vọng NDVB tiếng Việt như tuyên bố của nhà cung cấp.

2. LLaMA-3.2-11B-Vision-Instruct-Turbo: Mô hình thị giác-ngôn ngữ mã nguồn mở trên nền tảng Hugging Face, được huấn luyện để NDVB từ hình ảnh. Mô hình này được tinh chỉnh trên dữ liệu văn bằng để cải thiện hiệu suất.

3. EasyOCR: Thư viện Python mã nguồn mở, hỗ trợ NDVB đa ngôn ngữ, bao gồm tiếng Việt. EasyOCR yêu cầu xác định tọa độ thủ công cho các trường thông tin cần nhận dạng.

4. API Gemini của Google: Đây là một API thương mại hỗ trợ NDVB tiếng Việt, xử lý theo lô thông qua lời nhắc có khả năng xử lý cả chữ in và chữ viết tay.

3.3. Triển khai thực tế:



Hình 1. Mô hình triển khai thực tế

Triển khai thực tế được thiết kế theo hai hướng:

- Đối với sử dụng các API bên ngoài như DeepSeek /LLaMA / Google API Gemini:. Kết quả được ghi nhận và so sánh với văn bản gốc trên cả 2 tập ảnh, xử lý theo lô với 74 hình ảnh văn bằng chữ in và 63 hình ảnh chữ in / chữ viết tay. Đầu ra được so sánh với văn bản gốc để đánh giá độ chính xác.

- Đối với thư viện EasyOCR: Các trường thông tin (họ tên, ngày sinh, v.v.) được xác định thủ công bằng tọa độ trên hình ảnh, sau đó sử dụng EasyOCR để đọc và trích xuất văn bản.

3.4. Phương pháp đánh giá

Nghiên cứu có ba tiêu chí chính để đánh giá các phương pháp OCR:

- Độ chính xác: Tỷ lệ phần trăm văn bản được nhận dạng đúng so với tài liệu gốc, tính trên tất cả các trường thông tin.

- Tính ổn định: Mức độ nhất quán của kết quả qua nhiều lần chạy, đặc biệt quan trọng với các mô hình như LLaMA.

- Tính thực tiễn: Mức độ dễ triển khai trong thực tế, bao gồm yêu cầu xử lý thủ công, thời gian xử lý và khả năng hỗ trợ ngôn ngữ tiếng Việt.

IV. Phân tích hiệu suất các phương pháp

4.1. EasyOCR

Phương pháp này đạt độ chính xác 100% trên cả hai tập dữ liệu (mỗi tập chỉ lấy 10 file) khi các trường thông tin được xác định thủ công bằng tọa độ. Tuy nhiên, yêu cầu xác định tọa độ thủ công làm tăng thời gian xử lý, khiến EasyOCR không phù hợp cho các ứng dụng cần xử lý số lượng lớn tài liệu. Tính ổn định của EasyOCR cao, với kết quả nhất quán qua các lần chạy. Tuy nhiên, nghiên cứu chưa tìm được cách xử lý tự động căn tọa độ cho các trường dữ liệu. EasyOCR có đặc thù là thư viện cài đặt trên nền tảng Python, bộ cài đặt nhẹ và tối ưu chỉ riêng cho OCR. Tuy nhiên để mô hình có thể tối ưu thì phải tìm được cách tự động hóa nhận dạng các trường dữ liệu và tinh chỉnh được

4.2. DeepSeek

DeepSeek không thể NDVB tiếng Việt do thiếu hỗ trợ ngôn ngữ này trong API. Kết quả trả về là các chuỗi ký tự ngẫu nhiên hoặc không chính xác, với độ chính xác coi như bằng 0% trên cả hai tập dữ liệu. Phương pháp này không khả thi cho nhận dạng văn bằng tiếng Việt và không được đánh giá thêm về tính ổn định hoặc tính thực tiễn.

4.3. LLaMA-3.2-11B-Vision-Instruct -Turbo LLaMA-3.2

LlaMA cho thấy hiệu suất ổn định, có độ chính xác lần lượt là 100 % và 98% trên cả hai tập dữ liệu. Các lỗi chủ yếu liên quan đến việc nhận diện sai dấu thanh (ví dụ: “Nguyễn” bị nhận thành “Nguyen”) và khó khăn trong việc xử lý chữ viết tay. Qua ba lần chạy, các kết quả giống nhau cho thấy tính nhất quán trong xử lý của mô hình.

4.4. API Gemini của Google API Gemini

Đầu vào: Các file định dnagj ảnh thông dụng được xử lý nén

- Tập 1 (chữ in): Đạt độ chính xác 100% trên 74 hình ảnh, nhận diện đúng tất cả các trường thông tin, bao gồm dấu thanh và định dạng phức tạp.

- Tập 2 (chữ viết tay): Đạt độ chính xác 98% trên 63 hình ảnh, với sai sót chủ yếu ở các đoạn chữ viết tay khó đọc (ví dụ: chữ viết mờ hoặc dấu thanh không rõ ràng). Gemini có tính ổn định cao, có kết quả nhất quán qua các lần chạy và tính thực tiễn vượt trội nhờ khả năng xử lý theo lô và không yêu cầu thao tác thủ công.

V. Kết quả và thảo luận

5.1 Kết quả

Đầu ra của các dữ liệu chiết xuất được thiết lập dưới định dạng file csv, sẵn sàng để kết nối và làm đầu vào cho các hệ thống tác nghiệp quản lý đào tạo liên quan đến quản lý chứng chỉ, văn bằng thông qua các hàm API.

ĐT&PT học tập suốt đời	Nguyễn Thùy Chi	01/06/1988	BẰNG THẠC SĨ	QUẢN LÝ CÔNG
ĐT&PT học tập suốt đời	NGUYỄN THUỲ CHI	01/06/1988	BẰNG TỐT NGHIỆP Đ	Quản trị nhân lực
ĐT&PT học tập suốt đời	Trần Thị Hà Mi	24/07/1997	BẰNG TỐT NGHIỆP C	Báo chí
ĐT&PT học tập suốt đời	Ông Bùi Tuấn Quang	29/12/1998	BẰNG CỬ NHÂN	Báo chí
ĐT&PT học tập suốt đời	Nguyễn Thành Tuyên	03/05/1977	BẰNG TỐT NGHIỆP	CAO CẤP LÝ LUẬN CHÍNH TRỊ
ĐT&PT học tập suốt đời	Đào Thị Khánh Huyền	03/12/1993	BẰNG CỬ NHÂN	Tiếng Hàn Quốc
ĐT&PT học tập suốt đời	Bà Hoàng Thị Thanh T	17/10/2002	BẰNG CỬ NHÂN	Quản lý thông tin
ĐT&PT học tập suốt đời	Hà Anh Dũng	26/12/2001	BẰNG TỐT NGHIỆP C	Kỹ thuật chế biến món ăn
ĐT&PT học tập suốt đời	Trương Thu Minh	18/01/1978	CAO ĐẲNG	Văn - Gd. công dân
ĐT&PT học tập suốt đời	TRƯƠNG THU MINH	18/01/1978	BẰNG CỬ NHÂN	Quản lý Giáo dục
ĐT&PT học tập suốt đời	NGUYỄN ĐOÀN PHƯC	11/08/2002	BẰNG CỬ NHÂN	CHÍNH TRỊ HỌC
ĐT&PT học tập suốt đời	Nguyễn Thị Lan	25/11/1989	BẰNG TỐT NGHIỆP C	QUẢN TRỊ KINH DOANH
ĐT&PT học tập suốt đời	Nguyễn Thị Huyền	26/06/1997	BẰNG CỬ NHÂN	NGÔN NGỮ ANH
ĐT&PT học tập suốt đời	Ông Không Tuấn Minl	17/10/1988	BẰNG CỬ NHÂN	CÔNG NGHỆ THỐNG TIN
ĐT&PT học tập suốt đời	KHÔNG TUẤN MÌNH	17/10/1988	BẰNG TỐT NGHIỆP T	Thú y

Hình 2. Đầu ra csv có cảnh báo các trường hợp phát hiện bất thường

Kết quả triển khai các mô hình cho thấy EasyOCR, mặc dù đạt độ chính xác cao, lại không thực tiễn cho các ứng dụng quy mô lớn do yêu cầu xác định tọa độ thủ công, dẫn đến thời gian xử lý lâu và chi phí nhân lực cao. DeepSeek hoàn toàn không phù hợp do thiếu hỗ trợ ngôn ngữ tiếng Việt, một hạn chế chung ở nhiều mô hình chưa được tối ưu hóa cho các ngôn ngữ ít phổ biến [11]. Trong khi đó API Gemini của Google tỏ ra là phương pháp hiệu quả cho nhận dạng văn bằng tiếng Việt, đặc biệt với văn bản chữ in. Hiệu suất vượt trội của Gemini có thể được lý giải bởi kiến trúc mô hình tiên tiến, khả năng hỗ trợ ngôn ngữ tiếng Việt mạnh mẽ và khả năng xử lý hình ảnh chất lượng thấp [12]. Độ chính xác của LLaMA-3.2 cho thấy tiềm năng trong việc NDVB từ hình ảnh. Hơn thế nữa, LlaMA là hệ thống mã nguồn mở, có khả năng cài đặt trên mạng nội bộ, không phụ thuộc vào nhà cung cấp cấp giải pháp, do vậy, chủ động được về mặt công nghệ và chi phí triển khai.

Trên cơ sở thực hiện trên mô hình API Gemini trên dịch vụ điện toán đám mây của Google cho thấy chi phí thực hiện trên 1000 ảnh chỉ khoảng 1 USD, không cần yêu cầu về hạ tầng triển khai do sử dụng mô hình SaaS (Software as a Service). Mô hình hoàn toàn phù hợp với các tổ chức muốn triển khai nhanh, không cần đầu tư phức tạp về hạ tầng và có số lượng dữ liệu không quá đồ sộ.

5.2. Thách thức và cơ hội

Nhận dạng văn bằng tiếng Việt đặt ra các thách thức đặc thù, bao gồm sự phức tạp của dấu thanh và sự đa dạng trong chữ viết tay. Các mô hình như DeepSeek cần được đào tạo thêm trên dữ liệu tiếng Việt để cải thiện hiệu suất. Ngược lại, các giải pháp như LlaMA, Gemini và EasyOCR cho thấy tiềm năng lớn trong việc hỗ trợ tự động hóa xử lý văn bằng và chứng chỉ, đặc biệt trong các tổ chức giáo dục và cơ quan hành chính.

5.3. Hạn chế và phương hướng phát triển tương lai

Nghiên cứu tồn tại một vài điểm chưa hoàn thiện. Thứ nhất, quy mô dữ liệu

chỉ có 137 hình ảnh còn nhỏ so với thực tế ứng dụng quy mô lớn. Thứ hai, các văn bằng được sử dụng có định dạng tương đối đơn giản, chưa bao gồm các trường hợp phức tạp như văn bằng nhiều trang hoặc có bố cục không chuẩn hay định dạng song ngữ. Thứ ba, nghiên cứu chưa được thực trong điều kiện hình ảnh chất lượng thấp hơn (ví dụ: hình ảnh bị nhiễu hoặc ánh sáng kém).

Hướng phát triển tương lai nên tập trung vào:

1. Mở rộng tập dữ liệu gốc, thêm các văn bằng có định dạng phức tạp hơn. Huấn luyện thêm cho các mô hình với tập dữ liệu viết tay lớn hơn.

2. Tối ưu hóa mô hình EasyOCR thông qua huấn luyện trên dữ liệu văn bằng lớn hơn. Phát triển các phương pháp tự động hóa xác định tọa độ trong EasyOCR để giảm thiểu thao tác thủ công.

3. Đánh giá hiệu suất OCR trong các điều kiện thực tế, chẳng hạn như hình ảnh chất lượng thấp hoặc văn bằng đa ngôn ngữ.

4. Bổ sung đánh giá theo các metrics chuẩn dành cho OCR như: Precision, Recall, F1, CER

VI. Kết luận

Nghiên cứu đã thực hiện so sánh bốn phương pháp OCR trong bối cảnh nhận dạng văn bằng tiếng Việt, qua hai tập dữ liệu gồm chữ in và chữ viết tay. Kết quả cho thấy:

- DeepSeek chưa khả thi vì chưa hỗ trợ tiếng Việt.
- EasyOCR đạt độ chính xác cao ở mẫu thử nhỏ nhưng thiếu tính thực tiễn ở quy mô lớn.
- Gemini API chứng minh ưu thế nhờ khả năng xử lý theo lô, kết quả ổn định, độ chính xác cao cho cả chữ in và chữ viết tay.
- LLaMA3.2 đạt hiệu quả tương đương Gemini và vượt trội ở khía cạnh mã nguồn mở, có thể triển khai trong hạ tầng nội bộ, phù hợp với yêu cầu cao về chủ động công nghệ và bảo mật dữ liệu.

Trên cơ sở đó, bài báo khẳng định LLaMA3.2 là giải pháp tối ưu để ứng dụng OCR trong xử lý văn bằng tiếng Việt. Đồng thời, Gemini là lựa chọn hợp lý cho các đơn vị cần triển khai nhanh ở quy mô dịch vụ đám mây.

Những phát hiện này giúp cho các tổ chức giáo dục và cơ quan hành chính lựa chọn và triển khai phương pháp OCR phù hợp. Mô hình như LLaMA được lựa chọn do tính mở, khả năng phát triển nhanh, khả năng tích hợp dễ dàng vào các hệ thống sẵn có.

Tài liệu tham khảo:

[1]. Le, A., Lam, T., & Nguyen, D. (2025). A survey on Vietnamese document analysis and recognition: Challenges and future directions (arXiv:2506.05061). https://arxiv.org/pdf/2506.05061

[2]. Nguyen, T. H., Dinh, V. S., & Nguyen, P. L. (2022, May). *Vietnamese OCR: Research and applications*. BKAI Workshop. https://bkai.ai/wp-content/uploads/2022/05/Nguyen-Phi-Le_BKAI-workshop_OCR-Project_v3.pdf

[3]. Smith, R. (2007). An overview of the Tesseract OCR engine. In Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR) (Vol. 2, pp. 629–633). IEEE. https://ieeexplore.ieee.org/document/4376991

[4]. Zacharias, E., & Teuchler, M. (2020). Image processing based scene-text detection and recognition with Tesseract. arXiv. https://arxiv.org/abs/2004.08079

[5]. Hegghammer, T. (2022). OCR with Tesseract, Amazon Textract, and Google Document AI: A benchmarking experiment. *Journal of Computational Social Science, 5*, 861–882. https://doi.org/10.1007/s42001-021-00149-1

[6]. Memon, J., Sami, M., Khan, R. A., & Uddin, M. (2020). Handwritten optical character recognition (OCR): A comprehensive systematic literature review (SLR). IEEE Access, 8, 142647–142668 https://doi.org/10.1109/ACCESS.2020.3012542

[7]. Nguyen, N. H., Vo, D. T. D., & Nguyen, K. V. (2022). UIT-HWDB: Using transferring method to construct a novel benchmark for evaluating unconstrained handwriting image recognition in Vietnamese. arXiv. https://arxiv.org/abs/2211.05407

[8]. Patel, D., & Patel, A. (2022). Comparative study of optical character recognition using different machine learning techniques. In Proceedings of the International Conference on Intelligent Systems and Data Science (pp. 447–456). Springer. https://doi.org/10.1007/978-981-19-9512-5_38

[9]. Raj, A., Sharma, S., Singh, J., & Singh, A. (2023). Revolutionizing data entry: An in-depth study of optical character recognition technology and its future potential. International Journal for Research in Applied Science & Engineering Technology, 11(2). https://doi.org/10.22214/IJRASET.2023.49108

[10]. Nguyen, Q. D., Le, D. A., Phan, N. M., & Zelinka, I. (2020). An in-depth analysis of OCR errors for unconstrained Vietnamese handwriting. In *Future Data and Security Engineering* (pp. 448–461). Springer. https://doi.org/10.1007/978-3-030-63924-2_26

[11]. Agarwal, M., & Anastasopoulos, A. (2024). A concise survey of OCR for low-resource languages. In Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024) (pp. 88–102). Association for Computational Linguistics. https://aclanthology.org/2024.americasnlp-1.10/

[12]. Google. (2025). Gemini Developer API https://ai.google.dev/gemini-api/docs

COMPARISON OF OCR SOLUTIONS FOR VIETNAMESE CERTIFICATE RECOGNITION AND PRACTICAL APPLICATION PROPOSALS

Dang Hai Dang[2], Luu Tien Trung[2]

Abstract: *Optical Character Recognition (OCR) technology is a modern tool for extracting text from images, offering significant efficiency in automating diploma recognition. This study examines four popular OCR methods and provides practical implementation recommendations: DeepSeek, Meta AI's LLaMA-3.2-11B-Vision-Instruct-Turbo, EasyOCR, and Google's Gemini API in recognizing text from Vietnamese diplomas. After building, installing, and deploying these solutions—evaluated on three criteria: accuracy, stability, and practical applicability—the results show that DeepSeek does not support Vietnamese language, EasyOCR requires manual coordinate input, while Gemini API and LLaMA deliver outstanding performance with high accuracy, especially on diplomas and certificates containing handwritten text. Among them, LLaMA stands out as an open-source platform that is vendor-independent. The research concludes that LLaMA's implementation of OCR is the most optimal in terms of time, effort, and financial resources.*

Keywords: *optical Character Recognition (OCR), text recognition, Vietnamese language, document digitzation, image processing*

---

[2] Institute for Training and Lifelong Learning Development, Hanoi Open University