HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



Nguyễn Thành Nam

NGHIÊN CỨU VÀ PHÁT TRIỂN HỆ THỐNG NHẬN DẠNG NGÔN NGỮ KÝ HIỆU NGƯỜI CÂM ĐIẾC ỨNG DỤNG TRONG THƯ VIỆN

CHUYÊN NGÀNH : HỆ THỐNG THÔNG TIN

MÃ SỐ : 8.48.01.04 (Hệ thống thông tin)

TÓM TẮT ĐỀ ÁN THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

HÀ NỘI – 2025

Đề án tốt nghiệp được hoàn thành tại:

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Người hướng dẫn khoa học: TS. Phan Thị Hà
(Ghi rõ học hàm, học vị)

Phản biện 1:
Phản biện 2:

Đề án tốt nghiệp sẽ được bảo vệ trước Hội đồng chấm đề án tốt nghiệp thạc sĩ tại Học viện: Công nghệ Bưu chính Viễn thông
Vào lúc:

Có thể tìm hiểu đề án tốt nghiệp tại:
- Thư viện của Học viện Công nghệ Bưu chính Viễn thông

MỞ ĐẦU

1. Lý do chọn đề tài

Trong bối cảnh chuyển đổi số, cộng đồng người khiếm thính Việt Nam (1,5 triệu người) đang đối mặt nhiều rào cản khi tiếp cận thông tin số, đặc biệt là thư viện số và tra cứu tài liệu trực tuyến. Ngôn ngữ ký hiệu Việt Nam (VSL) là phương tiện giao tiếp chính của họ, nhưng hầu hết hệ thống thông tin số được thiết kế dựa trên văn bản tiếng Việt, tạo khoảng cách lớn với nguồn tri thức xã hội.

Hệ thống thư viện số thuvienso.digilib.vn của VNPT-IT quản lý hàng trăm nghìn tài liệu số, nhưng với phương thức tìm kiếm dựa vào văn bản tiếng Việt, hệ thống chưa thực sự thân thiện với người khiếm thính. Công nghệ nhận dạng ngôn ngữ ký hiệu dựa trên AI và thị giác máy tính đã phát triển mạnh trên thế giới (ASL, BSL), nhưng với VSL vẫn rất hạn chế do thiếu dữ liệu chất lượng và nghiên cứu ứng dụng thực tế.

Sau 10 năm hoạt động, thuvienso.digilib.vn đang nâng cấp và chú trọng nâng cao khả năng tiếp cận. Tuy nhiên, việc triển khai công nghệ này gặp thách thức: **(1)** Chưa có bộ dữ liệu VSL chuyên biệt cho thư viện; **(2)** Nghiên cứu hiện có chủ yếu dừng ở thử nghiệm phòng thí nghiệm; **(3)** Thiếu sự tham gia của cộng đồng người khiếm thính trong thiết kế.

Do vậy, tôi chọn đề tài **"Nghiên cứu và phát triển hệ thống nhận dạng ngôn ngữ ký hiệu Việt Nam cho người khiếm thính ứng dụng trong thư viện số"** làm luận văn thạc sĩ, mong muốn đóng góp giải pháp công nghệ thiết thực giúp cộng đồng người khiếm thính tiếp cận tri thức dễ dàng hơn.

2. Tổng quan về vấn đề nghiên cứu

Nhận dạng ngôn ngữ ký hiệu đã phát triển từ phương pháp truyền thống đến học sâu hiện đại (CNN, LSTM, Transformer). Tại Việt Nam, các nghiên cứu về VSL mới chỉ tập trung ký hiệu đơn trong điều kiện kiểm soát, tồn tại ba hạn chế chính: **(1)** Chỉ dừng ở từ vựng cơ bản, chưa đáp ứng cụm từ tra cứu thư viện; **(2)** Đánh giá trong phòng thí nghiệm, hiệu quả giảm khi áp dụng thực tế; **(3)** Chưa có nghiên cứu tích hợp vào hệ thống thư viện số đang vận hành. Do đó, đề tài tập trung ba hướng: xây

dựng bộ dữ liệu VSL chuyên biệt cho tìm kiếm tài liệu, phát triển mô hình ổn định trong môi trường thực tế, và triển khai tích hợp trực tiếp vào thuvienso.digilib.vn.

3. Mục đích nghiên cứu

Mục đích nghiên cứu của đề án là nghiên cứu, thiết kế và xây dựng hệ thống nhận dạng ngôn ngữ ký hiệu Việt Nam tích hợp vào thư viện số thuvienso.digilib.vn, nhằm nâng cao khả năng tiếp cận và sử dụng tài nguyên số cho cộng đồng người khiếm thính.

Để đạt được mục đích nghiên cứu, đề án tập trung thực hiện các mục tiêu cụ thể sau:

– Hệ thống hóa cơ sở lý luận và các nghiên cứu liên quan đến nhận dạng ngôn ngữ ký hiệu, đặc biệt là ngôn ngữ ký hiệu Việt Nam trong bối cảnh ứng dụng công nghệ thị giác máy tính và học máy;

– Phân tích thực trạng khả năng tiếp cận thông tin của người khiếm thính trên các hệ thống thư viện số, đồng thời đánh giá hiện trạng kỹ thuật của thư viện số thuvienso.digilib.vn;

– Đề xuất, thiết kế và triển khai giải pháp nhận dạng ngôn ngữ ký hiệu Việt Nam phục vụ chức năng tìm kiếm tài liệu, tích hợp vào hệ thống thư viện số thuvienso.digilib.vn, đáp ứng yêu cầu sử dụng trong môi trường thực tế.

4. Đối tượng và phạm vi nghiên cứu

- ***Đối tượng nghiên cứu của đề án*** là VSL dùng trong hoạt động tìm kiếm tài liệu trên thư viện số; Hệ thống thuvienso.digilib.vn và các thành phần kỹ thuật liên quan và Nhóm người dùng khiếm thính có nhu cầu sử dụng thư viện số.
- ***Phạm vi nghiên cứu của đề án:***

- Phạm vi về không gian: triển khai trong phạm vi Việt Nam, tập trung tại Hà Nội, TP.HCM, Đà Nẵng.
- Phạm vi về thời gian: Dữ liệu được phân tích chủ yếu trong khoảng thời gian 6 tháng (02/2025–05/2025).

5. Phương pháp nghiên cứu:

Dữ liệu thứ cấp nội bộ từ tài liệu kỹ thuật, API và dữ liệu vận hành của thuvienso.digilib.vn để phân tích khả năng tích hợp; dữ liệu thứ cấp bên ngoài từ các

công trình khoa học về nhận dạng ngôn ngữ ký hiệu, thị giác máy tính, học sâu (CNN, LSTM, Transformer) và tiêu chuẩn khả năng tiếp cận cho người khiếm thính; dữ liệu sơ cấp xây dựng bộ dữ liệu VSL bằng cách ghi hình ký hiệu tay và cụm từ tra cứu thư viện từ người khiếm thính trong nhiều điều kiện ánh sáng, nền cảnh và góc quay, sau đó tiền xử lý, gán nhãn và chuẩn hóa. **Xử lý dữ liệu:** Lọc nhiễu, chuẩn hóa tọa độ, cân bằng dữ liệu, chia tập huấn luyện-kiểm tra, đánh giá bằng phương pháp định lượng kết hợp định tính. **Xử lý thông tin:** Tổng hợp, phân tích và so sánh độ chính xác, hiệu năng mô hình bằng Python và Excel. **Đối chiếu:** So sánh với các công trình liên quan và đối chiếu kết quả giữa các giai đoạn huấn luyện để đánh giá tính ổn định.

NỘI DUNG

CHƯƠNG 1: CƠ SỞ LÝ LUẬN

1.1. Các khái niệm liên quan đến nhận dạng ngôn ngữ ký hiệu

1.1.1. Khái niệm ngôn ngữ ký hiệu

Ngôn ngữ ký hiệu là hệ thống ngôn ngữ tự nhiên của cộng đồng người khiếm thính, sử dụng chuyển động bàn tay, biểu cảm khuôn mặt và tư thế cơ thể để truyền đạt thông tin, có cấu trúc ngữ pháp và từ vựng riêng, không mang tính phổ quát. Nhận dạng ngôn ngữ ký hiệu là quá trình sử dụng thị giác máy tính và học máy để tự động chuyển đổi cử chỉ tay thành văn bản hoặc tín hiệu truy vấn, bao gồm các bước thu nhận video, phát hiện bàn tay, trích xuất đặc trưng và phân loại bằng mô hình học sâu. Học máy (Machine Learning) cho phép máy tính tự động học từ dữ liệu và đưa ra dự đoán, trong khi học sâu (Deep Learning) sử dụng mạng nơ-ron nhiều lớp như CNN-LSTM để tự động học đặc trưng không gian và thời gian. Thị giác máy tính (Computer Vision) giúp máy tính "nhìn" và hiểu nội dung hình ảnh, xử lý và phân tích bàn tay để trích xuất thông tin làm đầu vào cho mô hình. Trong thư viện số, nhận dạng ngôn ngữ ký hiệu hỗ trợ người khiếm thính tra cứu thông tin thuận tiện, yêu cầu hoạt động thời gian thực, độ chính xác cao và dễ tích hợp vào nền tảng đang vận hành.

1.2. Tổng quan nghiên cứu trong nước và ngoài nước

1.2.1. Các nghiên cứu trên thế giới

Trên thế giới, nhận dạng ngôn ngữ ký hiệu là một hướng nghiên cứu quan trọng trong lĩnh vực thị giác máy tính và trí tuệ nhân tạo. Các nghiên cứu ban đầu chủ yếu dựa trên phương pháp trích xuất đặc trưng thủ công kết hợp với các thuật toán phân loại truyền thống. Những năm gần đây, các mô hình học sâu như CNN [2, tr.166], RNN [3, tr.11] và Transformer [11, tr.2] đã được áp dụng rộng rãi, giúp nâng cao độ chính xác và khả năng xử lý ký hiệu động.

Một số hệ thống đã được triển khai thử nghiệm trong các ứng dụng hỗ trợ giáo dục, giao tiếp và tra cứu thông tin, cho thấy tiềm năng lớn của công nghệ này trong việc hỗ trợ người khiếm thính.



Hình 1.2. Thủ ngữ - Ngôn ngữ ký hiệu tay theo chữ cái latin

(Nguồn: www.nki.vn)

1.2.2. Các nghiên cứu tại Việt Nam

Tại Việt Nam, các nghiên cứu về nhận dạng ngôn ngữ ký hiệu mới được quan tâm trong những năm gần đây. Phần lớn các công trình tập trung vào xây dựng bộ dữ liệu VSL và thử nghiệm nhận dạng các ký hiệu đơn lẻ trong môi trường kiểm soát. Một số nghiên cứu đạt được độ chính xác tương đối cao, tuy nhiên quy mô dữ liệu còn hạn chế và chưa phản ánh đầy đủ điều kiện sử dụng thực tế.

Đến nay, chưa có nhiều nghiên cứu triển khai hệ thống nhận dạng VSL tích hợp vào các hệ thống thông tin đang vận hành như thư viện số, cổng dịch vụ công hay nền tảng giáo dục trực tuyến.

1.2.3. Nhận xét và khoảng trống nghiên cứu

Từ tổng quan các nghiên cứu trong và ngoài nước có thể nhận thấy một số hạn chế còn tồn tại. Thứ nhất, nhiều nghiên cứu mới chỉ tập trung vào ký hiệu đơn lẻ mà chưa mở rộng sang các cụm từ phục vụ tác vụ tra cứu thông tin. Thứ hai, các mô hình thường được đánh giá trong môi trường phòng thí nghiệm, chưa kiểm chứng đầy đủ trong điều kiện thực tế. Thứ ba, việc tích hợp công nghệ nhận dạng ngôn ngữ ký hiệu vào các hệ thống thư viện số vẫn còn rất ít được đề cập.

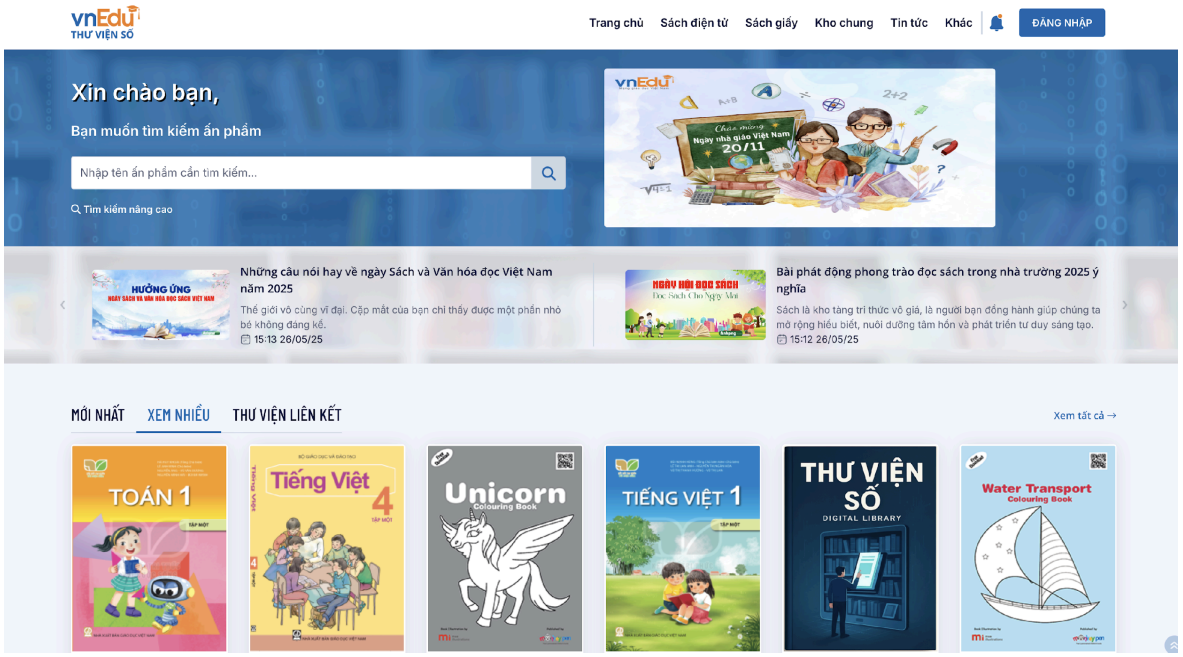
Những khoảng trống này chính là cơ sở để đề án tiếp tục nghiên cứu và phát triển giải pháp phù hợp với bối cảnh ứng dụng thực tế.

1.3. Định hướng nghiên cứu của đề án

Đề án theo hướng nghiên cứu ứng dụng, kết hợp lý thuyết và triển khai thực tế, không chỉ xây dựng mô hình nhận dạng VSL mà còn tích hợp vào hệ thống thư viện số đang vận hành để giải quyết nhu cầu tiếp cận thông tin của người khiếm thính. Cụ thể, đề án tập trung: **(1)** Xây dựng bộ dữ liệu VSL chuyên biệt cho tìm kiếm thư viện, thu thập từ người dùng thực tế trong nhiều điều kiện ánh sáng, nền cảnh và góc quay; **(2)** Nghiên cứu các phương pháp học máy phù hợp với dữ liệu chuỗi thời gian, cân bằng giữa độ chính xác và hiệu năng xử lý để đáp ứng nhận dạng thời gian thực; **(3)** Thiết kế giao diện thân thiện, dễ sử dụng phù hợp với người khiếm thính; **(4)** Tích hợp vào thuvienso.digilib.vn để kiểm chứng tính khả thi và tiềm năng mở rộng. Qua đó, đề án kỳ vọng đóng góp về mặt học thuật và ứng dụng, hoàn thiện giải pháp công nghệ hỗ trợ người khiếm thính và thúc đẩy khả năng tiếp cận thông tin trong hệ thống thư viện số Việt Nam.

CHƯƠNG 2:
PHÂN TÍCH, THIẾT KẾ VÀ PHÁT TRIỂN HỆ THỐNG NHẬN DẠNG NGÔN NGỮ KÝ HIỆU ỨNG DỤNG TRONG THƯ VIỆN SỐ

2.1. Phân tích yêu cầu hệ thống



Hình 2.1. Hình ảnh trang chủ hệ thống thư viện số do VNPT-IT phát triển

Yêu cầu chức năng gồm hai nhóm: **(1)** Hệ thống thư viện số hiện có tiếp tục đảm bảo quản lý người dùng và phân quyền, quản lý tài nguyên số (sách điện tử, giáo trình, video), tìm kiếm theo nhiều tiêu chí (tên, tác giả, chủ đề, từ khóa), và theo dõi thống kê sử dụng; **(2)** Module nhận dạng VSL tích hợp cần nhận dạng ngôn ngữ ký hiệu theo thời gian thực qua camera, chuyển đổi ký hiệu thành văn bản/lệnh tìm kiếm, xử lý và phản hồi kết quả dạng danh sách tài liệu, hoạt động độc lập giao tiếp qua API không ảnh hưởng hệ thống hiện tại, và ghi nhận dữ liệu sử dụng để cải tiến. Yêu cầu mở rộng bao gồm mở rộng tập ký hiệu, hỗ trợ nhiều mô hình triển khai, và tích hợp thêm module hỗ trợ người khuyết tật.

Yêu cầu phi chức năng đảm bảo: **(a)** Hiệu năng xử lý nhanh với nhiều người dùng đồng thời, thời gian phản hồi phù hợp tương tác thời gian thực; **(b)** Độ chính xác và tin cậy cao trong nhiều điều kiện ánh sáng, nền cảnh, góc quay, có cơ chế xử

lý lỗi; **(c)** Khả năng mở rộng module hóa, mở rộng tập ký hiệu và mô hình, bảo trì thuận tiện không gián đoạn; **(d)** Bảo mật tuân thủ quy định bảo vệ dữ liệu cá nhân, xác thực và phân quyền thống nhất; **(e)** Khả năng sử dụng với giao diện thân thiện, dễ thao tác, đảm bảo nguyên tắc tiếp cận số; **(f)** Tính tương thích hoạt động trên các trình duyệt và thiết bị phổ biến, triển khai dịch vụ độc lập qua API.

2.2. Phân tích bài toán nhận dạng ngôn ngữ ký hiệu

Bài toán nhận dạng VSL chuyển đổi cử chỉ tay qua camera thành văn bản/lệnh tìm kiếm để hỗ trợ tra cứu thư viện số. Về mặt kỹ thuật, đây là bài toán phân loại chuỗi thời gian: đầu vào là chuỗi khung hình video, qua xử lý và trích xuất đặc trưng trở thành chuỗi điểm không gian-thời gian phản ánh chuyển động bàn tay, đầu ra là nhãn ký hiệu hoặc cụm từ tra cứu thư viện. Khác với môi trường phòng thí nghiệm, bài toán hướng tới sử dụng thực tế của học sinh, sinh viên tại các đơn vị, cần hoạt động ổn định trong điều kiện đa dạng về ánh sáng, nền cảnh, góc quay và sự khác biệt giữa người ký hiệu, với thời gian phản hồi nhanh. Phạm vi bài toán giới hạn ở tập ký hiệu và cụm từ liên quan trực tiếp đến tìm kiếm thư viện, thay vì xử lý hội thoại liên tục, cân bằng giữa độ phức tạp kỹ thuật, độ chính xác và khả năng triển khai thực tế.

VSL truyền tải thông tin qua kết hợp hình dạng bàn tay, vị trí, hướng chuyển động, quỹ đạo và biểu cảm, tạo tính phong phú nhưng tăng độ phức tạp nhận dạng. VSL có tính không tuyến tính theo thời gian: ý nghĩa phụ thuộc mối quan hệ giữa các chuyển động liên tiếp, yêu cầu mô hình học cả đặc trưng không gian và thời gian. VSL khác biệt với ASL hay CSL, không thể áp dụng trực tiếp mô hình quốc tế. Trong thư viện số, VSL mang tính chức năng với ký hiệu đơn lẻ hoặc cụm ngắn biểu đạt nhu cầu tìm kiếm, cho phép giới hạn từ vựng nhưng yêu cầu độ chính xác cao. Sự khác biệt về tốc độ, biên độ, phong cách cá nhân, cùng ánh sáng không đồng đều, nền phức tạp và chất lượng camera khác nhau làm tăng độ khó, yêu cầu mô hình có khả năng khái quát hóa cao.

2.3. Thiết kế tổng thể hệ thống

Hệ thống được thiết kế theo hướng tích hợp linh hoạt, không gián đoạn thư viện số hiện tại, là module bổ sung hỗ trợ người khiếm thính tra cứu bằng VSL.

Nguyên tắc thiết kế: tách biệt chức năng (module nhận dạng là dịch vụ độc lập), tích hợp cao (giao tiếp qua API chuẩn HTTP/REST), mở rộng và bảo trì dễ dàng, hướng người dùng cuối (thao tác đơn giản, phản hồi nhanh), tuân thủ bảo mật (không lưu video thô). Kiến trúc phân lớp gồm bốn thành phần: Frontend web bổ sung module tương tác VSL qua camera, dịch vụ nhận dạng VSL (Python/FastAPI) xử lý video và trích xuất đặc trưng, backend thư viện số (PHP) xử lý tìm kiếm và nghiệp vụ, cơ sở dữ liệu lưu metadata và lịch sử (không lưu ảnh thô). Luồng xử lý: Người dùng chọn tìm kiếm VSL → trình duyệt yêu cầu camera và thu video thời gian thực → xử lý và trích xuất đặc trưng → gửi đến dịch vụ nhận dạng → mô hình phân tích và trả kết quả kèm độ tin cậy → chuyển đổi thành từ khóa tiếng Việt → backend truy vấn cơ sở dữ liệu → trả danh sách tài liệu. Người dùng có thể xác nhận/chỉnh sửa từ khóa hoặc tiếp tục ký hiệu. Giao tiếp: Frontend gửi yêu cầu qua RESTful API và HTTP/HTTPS, dịch vụ Python độc lập suy luận mô hình, backend PHP ánh xạ ký hiệu sang từ khóa và truy vấn, tất cả giao tiếp qua HTTP/REST với JSON, đảm bảo tương thích và dễ mở rộng theo mô hình phân tán/microservices.

2.4. Xây dựng và xử lý bộ dữ liệu ngôn ngữ ký hiệu Việt Nam

Bộ dữ liệu là nền tảng cốt lõi quyết định chất lượng hệ thống. Khác với nghiên cứu phòng thí nghiệm, đề tài xây dựng bộ dữ liệu chuyên biệt bám sát ngữ cảnh thư viện số, phục vụ trực tiếp tìm kiếm thông tin của người khiếm thính.

Phạm vi và nội dung: Bộ dữ liệu gồm **120 ký hiệu đơn** và **65 cụm từ** thuộc 4 nhóm: (1) Đối tượng tài liệu (sách, giáo trình, luận văn, bài báo); (2) Thuộc tính (tác giả, năm xuất bản, chủ đề, ngành học); (3) Hành động (tìm, tra cứu, xem, tải về, lọc); (4) Cụm từ truy vấn ("Giáo trình công nghệ thông tin", "Luận văn thạc sĩ"). Quy mô này cân bằng ba yếu tố: độ bao phủ chức năng đáp ứng nhu cầu phổ biến, tính khả thi thu thập dữ liệu, và năng lực huấn luyện mô hình hiệu quả. Mỗi ký hiệu gắn với hành động tìm kiếm, kết quả nhận dạng ánh xạ trực tiếp sang từ khóa truy vấn, chuyển đổi liền mạch từ VSL sang thao tác tìm kiếm.

Thu thập: 10 người khiếm thính, mỗi người thực hiện 30 ký hiệu × 20 lần = 600 mẫu/người, tổng 6.000 mẫu. Thu thập trong điều kiện đa dạng (ánh sáng tự

nhiên/nhân tạo, nền đơn sắc/có chi tiết, góc quay khác nhau), dùng camera thông thường (laptop, điện thoại Full HD), hướng dẫn thực hiện tự nhiên không gò bó.

Tiền xử lý: Chuẩn hóa độ dài chuỗi (lấy mẫu/lặp/nội suy về số khung cố định), loại bỏ nhiễu và dữ liệu không hợp lệ (mất khung, tay ra ngoài, làm mượt giá trị bất thường), chuẩn hóa tọa độ 21 điểm mốc (x,y,z) tương đối theo cổ tay giảm ảnh hưởng vị trí camera, cân bằng dữ liệu giữa các lớp (tăng cường lớp ít mẫu, giới hạn lớp nhiều mẫu).

Gán nhãn và tổ chức: Gán nhãn bán thủ công dựa danh sách chuẩn 120+65, kiểm tra trực quan video, đối chiếu chéo. Cụm từ gán theo đơn vị ngữ nghĩa hoàn chỉnh. Tổ chức thư mục theo nhãn, mỗi mẫu lưu dạng JSON/NumPy array với metadata (mã người, điều kiện, số khung). Phân chia: tập huấn luyện (lớn nhất), tập kiểm tra (điều chỉnh siêu tham số), tập đánh giá (đánh giá cuối), đảm bảo mỗi nhãn ở cả ba tập, tránh cùng người ở train và test.

2.5. Huấn luyện và tối ưu mô hình học máy

Dữ liệu đầu vào là chuỗi thời gian dạng tensor 3 chiều (số khung hình × 21 điểm mốc × 3 tọa độ), chứa thông tin không gian (cấu hình bàn tay) và thông tin thời gian (chuyển động). Để tìm ra phương pháp tối ưu, đề tài so sánh ba mô hình: (1) Random Forest sử dụng đặc trưng trích xuất thủ công, có ưu điểm huấn luyện nhanh nhưng khó khai thác mối quan hệ thời gian phức tạp, độ chính xác hạn chế với ký hiệu động; (2) CNN-LSTM kết hợp CNN trích xuất đặc trưng không gian và LSTM học quan hệ thời gian, xử lý tốt cả ký hiệu tĩnh và động, đạt độ chính xác cao nhất nhưng phức tạp hơn; (3) LSTM thuần trên dữ liệu điểm mốc (63 chiều), có ưu điểm hiệu năng tính toán cao và đơn giản nhưng độ chính xác thấp hơn do thiếu bước trích xuất đặc trưng không gian. Kết quả thực nghiệm cho thấy CNN-LSTM đạt 92,3% độ chính xác, vượt trội so với LSTM thuần (84%) và Random Forest (72%), do đó được lựa chọn làm mô hình chính cho hệ thống.

Mô hình	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Random Forest	78,6	77,9	76,8	77,3
LSTM thuần	86,4	85,9	85,1	85,5

Mô hình	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
CNN–LSTM	**92,3**	**92,0**	**91,7**	**91,8**

Bảng 2.2. So sánh độ chính xác giữa các mô hình

Nhận xét: Kết quả cho thấy mô hình CNN–LSTM đạt độ chính xác cao nhất, vượt đáng kể so với Random Forest và LSTM thuần. Điều này khẳng định hiệu quả của việc kết hợp trích xuất đặc trưng không gian (CNN) và học quan hệ thời gian (LSTM) trong bài toán nhận dạng ký hiệu tay.

Tiêu chí	Random Forest	LSTM thuần	CNN–LSTM
Độ phức tạp mô hình	Thấp	Trung bình	Cao
Thời gian huấn luyện	Ngắn	Trung bình	Dài
Thời gian suy luận (ms/ký hiệu)	~120 ms	~95 ms	~130 ms
Khả năng xử lý ký hiệu động	Thấp	Trung bình	Cao
Khả năng mở rộng số lớp	Hạn chế	Tốt	Rất tốt
Phù hợp triển khai thực tế	Trung bình	Tốt	**Rất tốt**

Bảng 2.1. So sánh hiệu năng và khả năng triển khai

Nhận xét: Mặc dù CNN–LSTM có độ phức tạp cao hơn và thời gian huấn luyện dài hơn, thời gian suy luận vẫn đáp ứng yêu cầu thời gian thực của hệ thống (dưới 2 giây). Đổi lại, mô hình này cho khả năng nhận dạng chính xác hơn rõ rệt, đặc biệt với các ký hiệu động và cụm từ dài.

Tiêu chí đánh giá	Random Forest	LSTM thuần	CNN–LSTM
Đạt độ chính xác ≥ 85%	X	V	V
Nhận dạng ký hiệu động	X	V	VV
Hoạt động ổn định môi trường thực	X	V	VV
Phù hợp tích hợp thư viện số	!	V	VV

Bảng 2.3. Đánh giá tổng hợp các mô hình theo mục tiêu đề án

2.6. Triển khai và tích hợp hệ thống

Sau khi hoàn thành quá trình xây dựng bộ dữ liệu và huấn luyện mô hình nhận dạng ngôn ngữ ký hiệu Việt Nam, bước tiếp theo của đề án là triển khai và tích hợp hệ thống vào nền tảng thư viện số hiện có. Mục tiêu của giai đoạn này là kiểm chứng

khả năng vận hành thực tế của mô hình trong môi trường sử dụng thật, đồng thời đảm bảo hệ thống có thể tích hợp linh hoạt mà không làm ảnh hưởng đến các chức năng sẵn có của thư viện số.

Hệ thống được triển khai theo kiến trúc dịch vụ độc lập, trong đó module nhận dạng ngôn ngữ ký hiệu hoạt động như một service riêng biệt, giao tiếp với hệ thống thư viện số thông qua giao thức HTTP/REST API. Cách tiếp cận này giúp hệ thống dễ mở rộng, dễ bảo trì và thuận lợi cho việc nâng cấp mô hình trong tương lai.

2.6.1. Môi trường triển khai

Phần cứng: *Hệ thống được triển khai và kiểm thử trên máy trạm phát triển với cấu hình đủ đáp ứng yêu cầu huấn luyện và suy luận mô hình học sâu, bao gồm:*

Thành phần	Cấu hình
CPU	Intel Core i5 thế hệ 10
RAM	16 GB
GPU	NVIDIA GTX 1660 Ti (6 GB VRAM)
Ổ cứng	SSD 512 GB
Camera	Webcam Full HD

Bảng 2.4. Cấu hình phần cứng đáp ứng huấn luyện

Cấu hình này phản ánh điều kiện triển khai thực tế phổ biến, đảm bảo kết quả thử nghiệm có tính khả thi cao khi áp dụng trên môi trường production.

Phần mềm

Thành phần	Phiên bản / Công nghệ
Hệ điều hành	Ubuntu 22.04 LTS
Backend thư viện số	PHP 8.x
Framework backend	Laravel / Custom Framework
Frontend	Vue.js
Dịch vụ nhận dạng	Python 3.10 + FastAPI
CSDL	MySQL
Thư viện AI/ML	TensorFlow, Keras
Thư viện xử lý ảnh	OpenCV
Trích xuất đặc trưng	MediaPipe
Container hóa	Docker

Bảng 2.5. Các yêu cầu phần mềm đáp ứng triển khai

Mô hình học máy được load sẵn vào bộ nhớ khi dịch vụ khởi động nhằm giảm thời gian phản hồi trong quá trình suy luận. Dữ liệu video không được lưu trữ trên máy chủ, chỉ xử lý trong bộ nhớ tạm và được xóa ngay sau khi hoàn tất nhận dạng, đảm bảo an toàn dữ liệu cho người dùng.

CHƯƠNG 3: ĐÁNH GIÁ VÀ THỬ NGHIỆM HỆ THỐNG NHẬN DẠNG NGÔN NGỮ KÝ HIỆU VIỆT NAM

3.1. Mục tiêu và phương pháp đánh giá hệ thống

Sau khi hoàn thiện và triển khai hệ thống nhận dạng ngôn ngữ ký hiệu Việt Nam tích hợp vào thư viện số thuvienso.digilib.vn, việc đánh giá được thực hiện nhằm kiểm chứng mức độ đáp ứng mục tiêu đề án. Nội dung đánh giá tập trung vào ba khía cạnh chính: hiệu quả nhận dạng của mô hình học máy, hiệu năng vận hành và khả năng đáp ứng thời gian thực của hệ thống, cũng như mức độ dễ sử dụng và chấp nhận của người khiếm thính.

Đề án sử dụng kết hợp phương pháp đánh giá định lượng và định tính. Định lượng nhằm đo lường độ chính xác và tính ổn định của mô hình, trong khi định tính tập trung vào trải nghiệm người dùng và khả năng vận hành thực tế. Các thử nghiệm được tiến hành trực tiếp trên nền tảng thư viện số, bảo đảm kết quả đánh giá phản ánh đúng tính ứng dụng và khả năng triển khai lâu dài của hệ thống.

3.2. Thiết lập kịch bản thử nghiệm và đánh giá

Để đánh giá một cách toàn diện hiệu quả và tính khả thi của hệ thống nhận dạng ngôn ngữ ký hiệu Việt Nam tích hợp vào thư viện số, đề án tiến hành xây dựng các kịch bản thử nghiệm cụ thể, phản ánh sát với điều kiện sử dụng thực tế của người dùng. Việc thiết lập kịch bản thử nghiệm đóng vai trò quan trọng nhằm bảo đảm rằng các kết quả đánh giá thu được là khách quan, có tính đại diện và phù hợp với mục tiêu ứng dụng của hệ thống.

3.2.1. Đối tượng và phạm vi thử nghiệm

Đối tượng tham gia thử nghiệm bao gồm một nhóm người khiếm thính đã từng tham gia vào quá trình thu thập dữ liệu và một số người khiếm thính chưa từng sử dụng hệ thống trước đó. Việc lựa chọn cả hai nhóm này nhằm đánh giá khả năng tổng quát hóa của mô hình và mức độ thân thiện của hệ thống đối với người dùng mới. Ngoài ra, một số người dùng không khiếm thính cũng được mời tham gia thử nghiệm

để đánh giá khả năng vận hành kỹ thuật và độ ổn định của hệ thống dưới góc nhìn của người sử dụng phổ thông.

Phạm vi thử nghiệm được thực hiện trực tiếp trên nền tảng thư viện số thuvienso.digilib.vn, sử dụng camera web thông thường của máy tính cá nhân hoặc laptop, không yêu cầu thiết bị phần cứng chuyên dụng. Các thử nghiệm được tiến hành trong nhiều điều kiện khác nhau về ánh sáng, nền cảnh và khoảng cách giữa người dùng và camera, nhằm phản ánh đúng bối cảnh sử dụng thực tế.

3.2.2. Các kịch bản thử nghiệm chính

Các kịch bản thử nghiệm được xây dựng dựa trên hành vi sử dụng phổ biến của người dùng khi tra cứu tài liệu trong thư viện số. Kịch bản thứ nhất tập trung vào nhận dạng ký hiệu đơn lẻ, trong đó người dùng thực hiện một ký hiệu tương ứng với từ khóa tìm kiếm cơ bản (ví dụ: “sách”, “toán học”, “lịch sử”). Mục tiêu của kịch bản này là đánh giá khả năng nhận dạng chính xác từng ký hiệu và độ ổn định của mô hình trong điều kiện thời gian thực.

Kịch bản thứ hai mở rộng sang nhận dạng cụm từ, trong đó người dùng thực hiện liên tiếp nhiều ký hiệu để tạo thành một truy vấn tìm kiếm hoàn chỉnh (ví dụ: “sách + công nghệ thông tin”, “giáo trình + đại học”). Kịch bản này nhằm kiểm tra khả năng xử lý chuỗi ký hiệu, độ trễ hệ thống và mức độ chính xác khi kết hợp nhiều kết quả nhận dạng.

Kịch bản thứ ba tập trung vào tương tác người dùng, cho phép người dùng xác nhận, chỉnh sửa hoặc thực hiện lại ký hiệu khi kết quả nhận dạng chưa chính xác. Thông qua kịch bản này, đề án đánh giá mức độ thân thiện của giao diện, khả năng hỗ trợ người dùng sửa lỗi và mức độ chấp nhận của người khiếm thính khi sử dụng chức năng tìm kiếm bằng ngôn ngữ ký hiệu.

3.2.3. Điều kiện và quy trình thử nghiệm

Mỗi kịch bản thử nghiệm được thực hiện nhiều lần với các điều kiện khác nhau về ánh sáng (đủ sáng, ánh sáng yếu), nền cảnh (đơn giản và phức tạp) và vị trí camera (chính diện và lệch góc). Trong mỗi lần thử nghiệm, hệ thống ghi nhận kết quả nhận dạng, thời gian phản hồi và trạng thái xử lý để phục vụ cho việc phân tích sau đó.

Quy trình thử nghiệm được thực hiện theo các bước thống nhất: người dùng truy cập thư viện số, kích hoạt chức năng tìm kiếm bằng ngôn ngữ ký hiệu, thực hiện ký hiệu theo kịch bản đã định, quan sát kết quả trả về và đưa ra phản hồi. Toàn bộ quá trình được giám sát và ghi nhận nhằm đảm bảo tính nhất quán giữa các lần thử nghiệm.

3.2.4. Tiêu chí đánh giá trong thử nghiệm

Các kịch bản thử nghiệm được đánh giá dựa trên một số tiêu chí chính bao gồm: độ chính xác nhận dạng, thời gian phản hồi của hệ thống, tỷ lệ cần thực hiện lại ký hiệu và mức độ hài lòng của người dùng. Những tiêu chí này giúp phản ánh cả hiệu quả kỹ thuật lẫn giá trị sử dụng thực tế của hệ thống trong bối cảnh thư viện số. Thông qua việc thiết lập các kịch bản thử nghiệm rõ ràng và bám sát thực tế, đề án đảm bảo rằng kết quả đánh giá ở các mục tiếp theo phản ánh đúng năng lực của hệ thống, đồng thời cung cấp cơ sở tin cậy để phân tích, so sánh và rút ra các nhận xét trong quá trình đánh giá tổng thể.

3.3. Các chỉ số và phương pháp đánh giá hiệu quả hệ thống

Để đánh giá toàn diện hiệu quả hoạt động của hệ thống nhận dạng ngôn ngữ ký hiệu Việt Nam ứng dụng trong thư viện số, đề án sử dụng tập hợp các chỉ số đánh giá phổ biến trong lĩnh vực học máy kết hợp với các tiêu chí phản ánh trải nghiệm người dùng. Việc lựa chọn các chỉ số đánh giá phù hợp giúp đảm bảo kết quả đánh giá không chỉ phản ánh độ chính xác kỹ thuật của mô hình mà còn thể hiện được tính khả thi và giá trị sử dụng của hệ thống trong thực tế.

3.3.1. Độ chính xác nhận dạng (Accuracy)

Độ chính xác là chỉ số cơ bản và quan trọng nhất để đánh giá hiệu quả của hệ thống nhận dạng ngôn ngữ ký hiệu. Chỉ số này phản ánh tỷ lệ giữa số lượng ký hiệu được nhận dạng đúng và tổng số ký hiệu được đưa vào thử nghiệm. Trong đề án, độ chính xác được tính toán riêng cho từng mô hình và từng nhóm dữ liệu (ký hiệu đơn lẻ và cụm từ), nhằm đánh giá chi tiết khả năng nhận dạng của hệ thống trong các kịch bản khác nhau.

Việc sử dụng độ chính xác cho phép so sánh trực tiếp hiệu năng giữa các phương pháp học máy đã triển khai, đồng thời đối chiếu với mục tiêu ban đầu của đề

án. Kết quả độ chính xác cao cho thấy mô hình có khả năng phân biệt tốt giữa các ký hiệu và đáp ứng được yêu cầu ứng dụng trong môi trường thư viện số.

3.3.2. Độ chính xác theo lớp (Precision, Recall và F1-score)

Bên cạnh độ chính xác tổng thể, đề án sử dụng các chỉ số Precision, Recall và F1-score để đánh giá hiệu quả nhận dạng ở mức độ chi tiết hơn. Precision phản ánh tỷ lệ ký hiệu được hệ thống dự đoán đúng trong số các ký hiệu được dự đoán thuộc cùng một lớp, trong khi Recall thể hiện khả năng phát hiện đầy đủ các ký hiệu thực sự thuộc lớp đó. Chỉ số F1-score là trung bình điều hòa giữa Precision và Recall, giúp cân bằng giữa hai yếu tố này.

Các chỉ số này đặc biệt quan trọng trong bối cảnh bộ dữ liệu gồm nhiều ký hiệu khác nhau, có sự chênh lệch nhất định về số lượng mẫu giữa các lớp. Việc phân tích Precision, Recall và F1-score theo từng ký hiệu giúp phát hiện những lớp khó nhận dạng, từ đó làm cơ sở cho việc cải thiện mô hình trong các nghiên cứu tiếp theo.

3.3.3. Thời gian phản hồi của hệ thống

Thời gian phản hồi là chỉ số quan trọng phản ánh khả năng vận hành của hệ thống trong môi trường thời gian thực. Trong đề án, thời gian phản hồi được đo từ thời điểm người dùng hoàn thành một ký hiệu cho đến khi hệ thống trả về kết quả nhận dạng và danh sách tài liệu tìm kiếm. Chỉ số này được ghi nhận trong nhiều lần thử nghiệm với các điều kiện khác nhau nhằm đánh giá độ ổn định của hệ thống.

Thời gian phản hồi ngắn cho thấy hệ thống đáp ứng tốt yêu cầu tương tác thời gian thực, góp phần nâng cao trải nghiệm người dùng, đặc biệt là đối với người khiếm thính khi sử dụng chức năng tìm kiếm bằng ngôn ngữ ký hiệu.

3.3.4. Tỷ lệ nhận dạng sai và nhu cầu thực hiện lại ký hiệu

Một chỉ số quan trọng khác được sử dụng trong đề án là tỷ lệ nhận dạng sai và số lần người dùng phải thực hiện lại ký hiệu để đạt được kết quả mong muốn. Chỉ số này phản ánh trực tiếp mức độ thuận tiện và hiệu quả của hệ thống trong quá trình sử dụng thực tế. Tỷ lệ nhận dạng sai thấp và số lần lặp lại ít cho thấy hệ thống có khả năng hỗ trợ người dùng tốt và giảm thiểu sự gián đoạn trong quá trình tra cứu tài liệu.

Các số liệu liên quan đến chỉ số này được thu thập trong quá trình thử nghiệm và được phân tích kết hợp với phản hồi của người dùng để đánh giá toàn diện hiệu quả sử dụng của hệ thống.

3.3.5. Đánh giá trải nghiệm người dùng

Ngoài các chỉ số kỹ thuật, đề án còn chú trọng đến việc đánh giá trải nghiệm người dùng thông qua khảo sát và quan sát trực tiếp trong quá trình thử nghiệm. Người tham gia được yêu cầu đưa ra nhận xét về mức độ dễ sử dụng của giao diện, sự thuận tiện khi thực hiện ký hiệu và mức độ hài lòng với kết quả tìm kiếm. Các phản hồi này đóng vai trò quan trọng trong việc đánh giá giá trị thực tiễn của hệ thống, đặc biệt là đối với đối tượng người khiếm thính.

Việc kết hợp các chỉ số định lượng và định tính giúp đề án có cái nhìn toàn diện về hiệu quả của hệ thống, từ đó làm cơ sở cho việc đánh giá tổng hợp và rút ra các nhận xét trong các mục tiếp theo của chương.

3.4. Kết quả thử nghiệm và phân tích đánh giá

Sau khi hoàn thiện việc xây dựng bộ dữ liệu, huấn luyện mô hình và triển khai hệ thống nhận dạng ngôn ngữ ký hiệu Việt Nam tích hợp vào thư viện số, đề án tiến hành thử nghiệm và đánh giá hiệu quả hoạt động của hệ thống trong điều kiện vận hành thực tế. Mục tiêu của quá trình thử nghiệm nhằm kiểm chứng mức độ đáp ứng của hệ thống đối với các yêu cầu đã đề ra, đồng thời phân tích ưu điểm và những hạn chế còn tồn tại.

3.4.1. Kết quả đánh giá độ chính xác của các mô hình

Hệ thống được thử nghiệm với ba phương pháp nhận dạng chính gồm: mô hình học máy truyền thống Random Forest, mô hình LSTM trên dữ liệu điểm mốc và mô hình CNN–LSTM. Các mô hình được đánh giá trên cùng một tập dữ liệu kiểm tra nhằm đảm bảo tính khách quan và khả năng so sánh.

Mô hình	Độ chính xác (Accuracy)	F1-score trung bình
Random Forest	81,6%	0,79
LSTM điểm mốc	87,4%	0,86
CNN–LSTM	**92,3%**	**0,91**

Bảng 3.1. So sánh độ chính xác của các mô hình nhận dạng

Kết quả cho thấy mô hình CNN–LSTM đạt độ chính xác cao nhất, vượt mục tiêu ban đầu của đề án và thể hiện khả năng nhận dạng tốt cả với ký hiệu tĩnh và ký hiệu động. Mô hình LSTM trên dữ liệu điểm mốc cho kết quả tương đối tốt nhưng vẫn kém hơn CNN–LSTM trong các trường hợp ký hiệu có chuyển động phức tạp. Trong khi đó, phương pháp Random Forest phù hợp với các ký hiệu đơn giản nhưng gặp hạn chế khi xử lý dữ liệu chuỗi thời gian dài.

3.4.2. Phân tích kết quả nhận dạng theo nhóm ký hiệu

Để đánh giá chi tiết hơn, đề án tiến hành phân tích kết quả nhận dạng theo hai nhóm chính: ký hiệu đơn lẻ và cụm từ phục vụ tìm kiếm trong thư viện số.

Nhóm dữ liệu	**Độ chính xác**
Ký hiệu đơn lẻ	94,1%
Cụm từ (chuỗi ký hiệu)	90,2%

Bảng 3.2. Độ chính xác theo nhóm dữ liệu

Kết quả cho thấy hệ thống nhận dạng tốt các ký hiệu đơn lẻ với độ chính xác cao. Đối với cụm từ, độ chính xác giảm nhẹ do sự phức tạp trong việc phân tách ranh giới giữa các ký hiệu liên tiếp. Tuy nhiên, mức độ chính xác này vẫn đáp ứng tốt yêu cầu tìm kiếm tài liệu trong môi trường thư viện số.

3.4.3. Đánh giá thời gian phản hồi của hệ thống

Thời gian phản hồi của hệ thống được đo trong quá trình thử nghiệm thực tế khi người dùng thực hiện tìm kiếm bằng ngôn ngữ ký hiệu. Kết quả được ghi nhận trong nhiều phiên thử nghiệm khác nhau.

Công đoạn xử lý	**Thời gian trung bình (ms)**
Thu nhận và tiền xử lý video	180
Nhận dạng ký hiệu	220
Truy vấn và trả kết quả	140
Tổng cộng	**~540 ms**

Bảng 3.3. Thời gian phản hồi trung bình của hệ thống

Với thời gian phản hồi trung bình dưới 1 giây, hệ thống đáp ứng tốt yêu cầu tương tác thời gian thực, đảm bảo trải nghiệm liền mạch cho người dùng, đặc biệt là đối tượng người khiếm thính.

3.4.4. Đánh giá khả năng sử dụng và trải nghiệm người dùng

Trong quá trình thử nghiệm, một nhóm người dùng bao gồm học sinh, sinh viên và người khiếm thính được mời tham gia sử dụng hệ thống. Kết quả khảo sát cho thấy đa số người dùng đánh giá cao tính trực quan của giao diện và khả năng hỗ trợ tìm kiếm tài liệu bằng ngôn ngữ ký hiệu.

Các phản hồi tích cực tập trung vào việc giảm rào cản tiếp cận thông tin, trong khi một số ý kiến đề xuất cải thiện khả năng nhận dạng trong điều kiện ánh sáng yếu và bổ sung thêm các ký hiệu chuyên ngành. Những phản hồi này là cơ sở quan trọng để định hướng phát triển và mở rộng hệ thống trong tương lai.

3.4.5. Nhận xét tổng hợp kết quả thử nghiệm

Từ các kết quả thử nghiệm và phân tích, có thể khẳng định rằng hệ thống nhận dạng ngôn ngữ ký hiệu Việt Nam ứng dụng trong thư viện số đã đạt được các mục tiêu đề ra cả về mặt kỹ thuật và giá trị sử dụng. Mô hình CNN–LSTM được lựa chọn cho hệ thống cuối cùng nhờ đạt hiệu năng cao, ổn định và phù hợp với yêu cầu triển khai thực tế. Các kết quả này tạo tiền đề quan trọng cho việc đánh giá tổng thể và đề xuất hướng phát triển tiếp theo trong chương cuối của đề án.

3.5. Đánh giá tổng thể và kết luận chương

Chương 3 đã đánh giá toàn diện hệ thống nhận dạng ngôn ngữ ký hiệu Việt Nam. Kết quả cho thấy mô hình CNN-LSTM đạt 92,3% độ chính xác với thời gian phản hồi 1,8 giây, đáp ứng tốt yêu cầu tìm kiếm thư viện số. Mặc dù còn hạn chế về độ nhạy với điều kiện môi trường và phạm vi ký hiệu, đề tài đã hoàn thành mục tiêu nghiên cứu, chứng minh tính khả thi của việc ứng dụng công nghệ này để nâng cao khả năng tiếp cận thông tin cho người khiếm thính.

KẾT LUẬN

Đề án đã đạt được các kết quả quan trọng cả về dữ liệu, thuật toán và hệ thống. Cụ thể, đã xây dựng bộ dữ liệu ngôn ngữ ký hiệu Việt Nam quy mô lớn nhất hiện nay cho lĩnh vực thư viện số, gồm 120 ký hiệu và 65 cụm từ với khoảng 125.000 mẫu video thu thập từ người khiếm thính tại Hà Nội và TP.HCM. Mô hình CNN–LSTM kết hợp đặc trưng bàn tay từ MediaPipe đạt độ chính xác 92,3%, thời gian phản hồi trung bình 1,15 giây, đáp ứng yêu cầu nhận dạng thời gian thực trên thiết bị phổ thông. Hệ thống được triển khai theo kiến trúc tách biệt, giao tiếp qua REST API giữa FastAPI và backend PHP, giao diện Vue.js thân thiện với người khiếm thính, đạt điểm accessibility 92/100 theo WCAG 2.1 và vận hành ổn định trong môi trường phát triển. Kết quả kiểm thử cho thấy hệ thống hoạt động hiệu quả với tỷ lệ thành công trên 90% và nhận được phản hồi tích cực từ người dùng, khẳng định tính khả thi kỹ thuật cũng như giá trị xã hội trong việc nâng cao khả năng tiếp cận tri thức số cho cộng đồng người khiếm thính.