Tạp chí Khoa học Đại học Cần Thơ
Phần A: Khoa học tự nhiên, Công nghệ và Môi trường
ISSN 1859-2333 | e-ISSN 2815-5599

DOI:10.22144/ctujos.2025.046

PHÁT HIỆN TƯƠNG ĐỒNG HÌNH ẢNH TRONG BÀI BÁO KHOA HỌC BẰNG PHƯƠNG PHÁP XỬ LÝ ẢNH KẾT HỢP MẠNG HỌC SÂU RESNET50

Trần Thanh Điện[1*], Trần Thị Trúc Linh[2], Lê Duy Anh[1], Nguyễn Thị Kim Quyên[1], Nguyễn Bạch Đan[3], Nguyễn Thanh Hải[4] và Nguyễn Thái Nghe[4]
[1]Nhà xuất bản Đại học Cần Thơ, Trường Đại học Cần Thơ, Việt Nam
[2]Viettel Store, Cần Thơ, Việt Nam
[3]Tạp chí Khoa học Trường Đại học Cần Thơ, Trường Đại học Cần Thơ, Việt Nam
[4]Trường Công nghệ Thông tin và Truyền thông, Trường Đại học Cần Thơ, Việt Nam
**Tác giả liên hệ (Corresponding author): thanhdien@ctu.edu.vn*

Thông tin chung (Article Information)

Nhận bài (Received): 20/12/2024
Sửa bài (Revised): 10/03/2025
Duyệt đăng (Accepted): 24/03/2025

Title: *Image similarity detection in scientific articles using image processing and deep learning Resnet50*

Author(s): *Tran Thanh Dien[1*], Tran Thi Truc Linh[2], Le Duy Anh[1], Nguyen Thi Kim Quyen[1], Nguyen Bach Dan[3], Nguyen Thanh Hai[4] and Nguyen Thai Nghe[4]*

Affiliation(s): *[1]Can Tho University Publishing House, Can Tho University, Viet Nam; [2]Viettel Store, Can Tho City, Viet Nam; [3]Can Tho University Journal of Science, Can Tho University, Viet Nam; [4]College of Information and Communication Technology, Can Tho University, Viet Nam*

TÓM TẮT

Nghiên cứu này đề xuất mô hình học sâu ResNet50 để phân loại hình ảnh trong bài báo khoa học, nhằm phát hiện tương đồng và cải thiện tìm kiếm hình ảnh. Mô hình sử dụng ResNet50 đã được huấn luyện trước, kết hợp với tập dữ liệu gồm 12.049 ảnh thuộc 11 lớp, trích xuất từ Tạp chí Khoa học Đại học Cần Thơ bằng PyMuPDF. Phương pháp Activation Map Visualization giúp làm nổi bật vùng dữ liệu huấn luyện thông qua sáu kênh đầu tiên của từng lớp khác nhau trên mô hình học sâu. Kết quả cho thấy phương pháp đề xuất đạt độ tin cậy trên 90% trong phát hiện tương đồng hình ảnh, đồng thời xác định được tác giả và năm xuất bản bài báo gốc. Mô hình ResNet50 cũng được so sánh với AlexNet và VGG16, cho thấy khả năng tổng quát hóa vượt trội trong bài toán nhận diện ảnh phức tạp. Nghiên cứu này đặt nền móng cho giải pháp phát hiện tương đồng hình ảnh các ấn phẩm khoa học.

Từ khóa: *Bài báo khoa học, học sâu, ResNet50, trích xuất đặc trưng, tương đồng hình ảnh*

ABSTRACT

This study proposes a deep learning-based model using ResNet50 for image classification in scientific articles to detect similarity and improve image similarity search. The model employs a pre-trained ResNet50 combined with a dataset of 12,049 images categorized into 11 classes extracted from the Can Tho University Journal of Science using PyMuPDF. The Activation Map Visualization method highlights training data regions through the first six channels of each different layer in the deep learning model. The results indicate that the proposed approach achieves exceeding 90% reliability in detecting image similarity and can identify the original article's author and publication year. ResNet50 is also compared with AlexNet and VGG16, demonstrating superior generalization capability for complex image recognition tasks. The result establishes a foundation for developing an image similarity detection system for scientific publications.

Keywords: *Deep learning, feature extraction, image similarity, ResNet50, scientific paper*

1. GIỚI THIỆU

Trong bối cảnh phát triển mạnh mẽ của công nghệ và mạng xã hội, việc xử lý và lưu trữ hình ảnh ngày càng trở nên phổ biến. Hình ảnh đóng vai trò quan trọng trong nhiều lĩnh vực như y tế, giáo dục và khoa học. Tuy nhiên, việc tìm kiếm và phát hiện sự trùng lặp hình ảnh trong các bài báo khoa học vẫn là một thách thức đáng kể. Trí tuệ nhân tạo (Artificial Intelligence - AI) là một lĩnh vực nghiên cứu nhằm giúp máy tính có thể tự động hóa các hành vi thông minh của con người. Các hệ thống AI hoạt động bằng cách thu thập dữ liệu gán nhãn, phân tích các mẫu và mối tương quan để đưa ra dự đoán chính xác. Đặc biệt, học máy (machine learning - ML), một nhánh của AI, có khả năng học tập và đưa ra quyết định dựa trên dữ liệu đầu vào mà không cần lập trình cụ thể (Russell & Norvig, 2021; McCorduck, 2004). Do đó, ML trở thành công cụ hữu ích trong việc phân tích và tính toán tương đồng hình ảnh.

Trí tuệ nhân tạo được chính thức trở thành một lĩnh vực nghiên cứu từ năm 1956 (McCorduck, 2004), trải qua nhiều giai đoạn phát triển mạnh mẽ. Trong đó, ML được xem là một trong những công nghệ cốt lõi, hỗ trợ nhiều ứng dụng trong thực tế. Thay vì lập trình tường minh cho từng tác vụ, ML sử dụng các thuật toán để phân tích dữ liệu, học hỏi từ các mẫu và đưa ra dự đoán hoặc quyết định. Trong lĩnh vực xử lý hình ảnh, các mô hình AI có thể phân loại, tìm kiếm và phát hiện các hình ảnh tương đồng từ một tập dữ liệu lớn. Một trong những phương pháp quan trọng trong xử lý ảnh là trích xuất đặc trưng (Feature Extraction) (Guyon & Elisseeff, 2006). Đây là quá trình trích xuất các điểm đặc trưng từ hình ảnh để biểu diễn dưới dạng vector số, giúp so sánh và xác định mức độ tương đồng giữa các hình ảnh. Phương pháp này có ý nghĩa quan trọng trong phát hiện trùng lặp hình ảnh trong bài báo khoa học và đảm bảo tính chính xác của nghiên cứu.

Học sâu (Deep learning - DL) sử dụng các mạng nơ-ron nhân tạo để phân tích dữ liệu ở nhiều cấp độ trừu tượng khác nhau. Một trong những mô hình nổi bật của deep learning trong xử lý ảnh là mạng nơ-ron tích chập (Convolutional Neural Network - CNN). CNN đã chứng minh hiệu quả vượt trội trong nhận diện, phân loại, phát hiện đối tượng và phân đoạn hình ảnh. Các mô hình CNN tiêu biểu bao gồm: AlexNet (2012), VGG16 (2014), GoogleNet Inception-V1 (2014), ResNet50 (2015), DenseNet (2016) (He et al., 2016).

Trong thời gian diễn ra đại dịch COVID-19, một nhóm nghiên cứu đã đề xuất một hệ thống nhận diện và cảnh báo khi phát hiện người không đeo hoặc đeo khẩu trang sai cách dựa trên các kỹ thuật học sâu AlexNet, GoogLeNet và VGG16 (Luu et al., 2022). Nghiên cứu này nhấn mạnh tầm quan trọng của việc đeo khẩu trang trong không gian công cộng để giảm nguy cơ lây nhiễm. Bằng cách sử dụng tập dữ liệu với 4.950 ảnh với 3 lớp (không đeo khẩu trang, đeo sai cách, đeo đúng cách), nghiên cứu đã cho kết quả dự đoán độ chính xác của hệ thống đạt trên 95%. Trong khi đó, một nghiên cứu khác đề xuất một phương pháp sử dụng kỹ thuật học máy để phân loại chất lượng tấm pin năng lượng mặt trời sử dụng ba mô hình gồm hồi quy logistic (Logistic Regression), máy vector hỗ trợ (Support Vector Machine) và mạng nơ-ron nhân tạo (Artificial Neural Network). Hình ảnh các tấm pin được thu thập bằng camera hồng ngoại trong phòng tối với 900 ảnh được chia thành 4 lớp, mỗi lớp biểu thị mức độ hư hỏng khác nhau. Kết quả cho thấy, máy vector hỗ trợ là mô hình tối ưu nhất cho bài toán phân loại chất lượng tấm pin với độ chính xác cao nhất đạt khoảng 97% (Luu et al., 2023).

Trong nghiên cứu này, ResNet50 được lựa chọn để phân tích, xử lý và tính toán tương đồng hình ảnh, đồng thời so sánh với hai mô hình học sâu phổ biến khác là AlexNet và VGG16 để so sánh hiệu suất phân loại của từng mô hình, trong đó sâu ResNet50 chứng minh được khả năng tổng quát hóa tốt nhất, là lựa chọn tối ưu cho các bài toán nhận diện và phân loại ảnh có độ phức tạp cao. ResNet50 là một phiên bản của kiến trúc ResNet với 50 lớp, bao gồm các lớp tích chập (convolutional layers) kết hợp với các Residual Blocks (khối trong mạng nơ-ron sâu giúp giải quyết các vấn đề khi mạng trở nên quá sâu gây suy giảm đạo hàm và giảm hiệu suất huấn luyện) để cải thiện khả năng học của mô hình, được giới thiệu lần đầu bởi nhóm nghiên cứu của Kaiming He, Xiangyu Zhang, Shaoqing Ren và Jian Sun trong bài báo "Deep Residual Learning for Image Recognition" tại hội nghị CVPR 2015 (He et al., 2016).

2. CÁC NGHIÊN CỨU LIÊN QUAN

Trong những năm gần đây, trí tuệ nhân tạo (AI) đã phát triển mạnh mẽ và được ứng dụng rộng rãi vào nhiều lĩnh vực trong đời sống. Theo Zhai et al. (2021), AI là kết quả của việc mô phỏng các quá trình trí tuệ của con người thông qua máy móc, đặc biệt là các hệ thống máy tính. Một số nghiên cứu tập trung vào xử lý hình ảnh và dữ liệu thông qua các mô hình tiên tiến. Việc ứng dụng ML đã được tối ưu

hóa, tạo ra bước tiến vượt bậc trong công nghệ nhờ khả năng linh hoạt và tự học của các thuật toán.

Sự phát triển của DL đã thúc đẩy nhiều nghiên cứu tập trung vào CNN, giúp nâng cao hiệu quả trong việc xác định và tính toán độ tương đồng hình ảnh. Một số nghiên cứu gần đây tập trung vào học tương phản (Contrastive Learning), nhằm tìm ra các cặp đặc trưng có tính tương đồng hoặc tương phản trong tập dữ liệu. Quy trình thực hiện bao gồm một số bước chính. Tăng cường dữ liệu (data augmentation) là một phần quan trọng, bởi nếu không đủ phức tạp, đạo hàm của một hàm mất mát (gradient) sẽ không đủ mạnh để mô hình học được các đặc trưng. Các phương pháp giúp tăng cường dữ liệu như Random Crop, Random Color Distortions và Random Gaussian Blur làm cho mô hình học sâu trở nên tổng quát hơn và chống overfitting. Kết quả cho thấy, việc kết hợp random crop và color distortion mang lại hiệu suất cao nhất.

Ngoài ra, nghiên cứu của Chechik et al. (2010) đã đề xuất thuật toán học trực tuyến để đo độ tương đồng hình ảnh dựa trên phương pháp OASIS. Mô hình này giúp xử lý hiệu quả tập dữ liệu quy mô lớn bằng cách chuyển vấn đề thành một bài toán tuyến tính đơn giản.

Bên cạnh đó, có các nghiên cứu liên quan đến tính toán tương đồng dựa trên đặc trưng. Hirematch and Puijari (2007) sử dụng PageRank để xây dựng đồ thị tương đồng hình ảnh, trong đó các ảnh được liên kết với nhau dựa trên mức độ tương đồng. Trong khi đó, Russakoff et al. (2004) sử dụng Gradient Vector Flow để trích xuất đặc trưng hình dạng, kết hợp với thông tin về kết cấu và màu sắc để tăng cường hiệu suất tìm kiếm hình ảnh.

Một số nghiên cứu mới hơn cũng đã đề xuất cách tiếp cận kết hợp giữa deep learning và truyền thống để tăng cường hiệu suất phân loại hình ảnh. He et al. (2016) đã giới thiệu mô hình ResNet, một bước đột phá trong mạng nơ-ron tích chập với kỹ thuật residual learning, giúp giải quyết vấn đề suy giảm gradient khi mạng học sâu hơn. ResNet50 với 50 lớp học đã chứng minh hiệu quả vượt trội trong các bài toán nhận diện hình ảnh. Szegedy et al. (2015) đề xuất kiến trúc Inception để cải thiện hiệu suất mô hình CNN bằng cách sử dụng nhiều kích thước bộ lọc trong cùng một tầng mạng, cho phép mô hình học được các đặc trưng đa cấp độ. Ngoài ra, nghiên cứu của Krizhevsky et al. (2012) về AlexNet đã chứng minh CNN có thể đạt độ chính xác cao trong các bài toán phân loại hình ảnh, mở đường cho sự phát triển của deep learning trong xử lý ảnh.

Gần đây, nghiên cứu của Tan and Le (2019) đã đề xuất mô hình EfficientNet, sử dụng kỹ thuật compound scaling để tối ưu hóa kích thước mô hình trong khi vẫn duy trì hiệu suất cao. EfficientNet đã đạt được kết quả tốt hơn so với ResNet50 trong nhiều tác vụ phân loại hình ảnh. Bên cạnh đó, các nghiên cứu tập trung vào học tự giám sát như SimCLR (Chen et al., 2020) đã cho thấy khả năng học đặc trưng hình ảnh mà không cần gán nhãn, mở ra hướng đi mới cho các bài toán phân loại hình ảnh với dữ liệu hạn chế.

Nghiên cứu tập trung vào việc xử lý hình ảnh tương đồng trong các bài báo khoa học, sử dụng phương pháp trích xuất đặc trưng từ hình ảnh để xác định mức độ tương đồng với độ chính xác cao. Việc phát triển một hệ thống tự động với khả năng phân tích hình ảnh có ý nghĩa quan trọng không chỉ trong lĩnh vực khoa học mà còn trong nhiều lĩnh vực khác như y tế, kinh doanh và công nghệ. Với sự hỗ trợ của các mô hình DL tiên tiến, nghiên cứu này hứa hẹn sẽ đóng góp đáng kể vào việc cải thiện độ chính xác trong nhận diện và tìm kiếm hình ảnh tương đồng.

3. MÔ HÌNH ĐỀ XUẤT

3.1. Sơ đồ tổng quát của nghiên cứu

Kiến trúc tổng quát xử lý ảnh tương đồng được biểu diễn như Hình 1. Kiến trúc bao gồm các bước như sau: Hệ thống tiếp nhận ảnh đầu vào, còn được gọi là ảnh mục tiêu, để thực hiện quá trình tìm kiếm ảnh tương đồng. Cụ thể như sau:

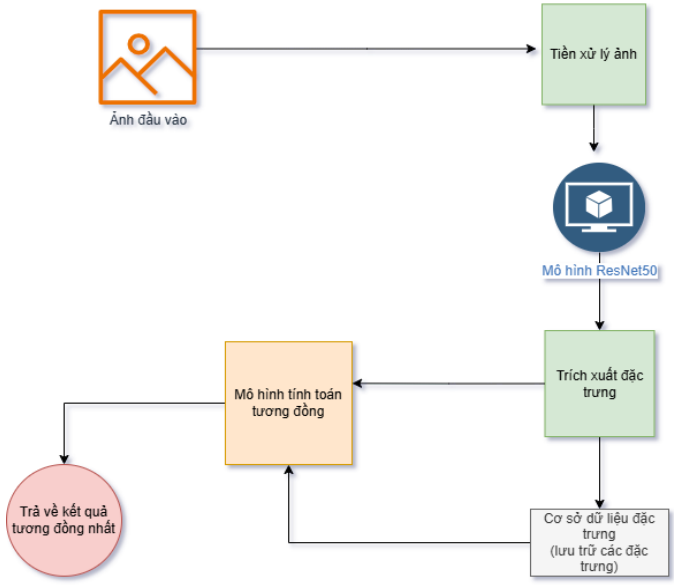
+ Tiền xử lý và huấn luyện mô hình: Sử dụng mô hình ResNet50 để huấn luyện trên tập dữ liệu phân loại hình ảnh. Quá trình bao gồm sử dụng mô hình ResNet50 đã được huấn luyện trước, điều chỉnh phù hợp với số lớp của tập dữ liệu, thực hiện huấn luyện và lưu lại mô hình sau khi tối ưu. Dữ liệu huấn luyện được chia thành tập train và test, đảm bảo tính độc lập giữa các mẫu.

+ Trích xuất ảnh: Thư viện PyMuPDF được sử dụng để trích xuất hình ảnh từ các bài báo khoa học, đảm bảo thu thập đầy đủ thông tin như tên file, số trang, DOI, tiêu đề bài báo, tạp chí, nguồn gốc,...

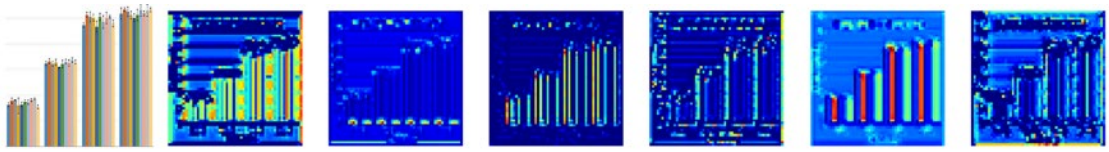
Tiếp theo, ảnh đầu vào được tiền xử lý nhằm chuẩn hóa dữ liệu đầu vào cho mô hình ResNet50. Cụ thể, ảnh được thay đổi kích thước về 224×224 pixel nhằm đảm bảo sự đồng nhất về kích thước và phù hợp với yêu cầu của mô hình. Sau đó, ảnh được chuyển đổi từ định dạng PIL (Python Imaging Library) hoặc NumPy array sang dạng tensor, một cấu trúc dữ liệu phổ biến trong PyTorch.

+ Tinh chỉnh (fine-tuning) ResNet50 cho bài toán tính toán tương đồng hình ảnh. Tinh chỉnh ResNet50 bằng cách giữ nguyên các lớp convolutional ban đầu và chỉ huấn luyện lại các lớp Fully Connected (FC) để phù hợp với bài toán. Lớp FC cuối được điều chỉnh với 11 đầu ra tương ứng với số lớp của tập dữ liệu.

Đồng thời, các giá trị pixel của ảnh được chuẩn hóa (normalized) từ phạm vi [0, 255] về [0, 1] và tiếp tục điều chỉnh dựa trên giá trị trung bình (mean) và độ lệch chuẩn (standard deviation - std) của tập dữ liệu huấn luyện. Quá trình này giúp đảm bảo tính nhất quán trong dữ liệu đầu vào khi đưa vào mô hình ResNet50. Sau khi tiền xử lý, mô hình ResNet50 được tải lên cùng với các trọng số đã được huấn luyện trước đó. Hình ảnh đầu vào sẽ được đưa qua mô hình để trích xuất đặc trưng. Các đặc trưng này sau đó được lưu trữ trong cơ sở dữ liệu cùng với thông tin liên quan, bao gồm đường dẫn ảnh gốc, đường dẫn lưu trữ tệp chứa đặc trưng và các đặc trưng được lưu dưới dạng tệp NumPy riêng biệt cho từng ảnh. Quá trình tính toán độ tương đồng giữa ảnh đầu vào và các ảnh trong cơ sở dữ liệu được thực hiện bằng cách truy xuất thông tin từ hệ thống. Cụ thể, hệ thống lấy danh sách dữ liệu gồm đường dẫn ảnh (image_path), đường dẫn tệp chứa đặc trưng (feature_path), và ID của bài báo tương ứng. Hệ thống sẽ trả về danh sách chứa các thông tin trên và tiếp tục thực hiện truy vấn để thu thập dữ liệu bài báo, bao gồm DOI, tiêu đề, tạp chí, năm xuất bản và danh sách tác giả. Độ tương đồng giữa ảnh đầu vào và các ảnh trong cơ sở dữ liệu được tính toán dựa trên hàm đo lường cosine similarity giữa vector đặc trưng của ảnh mục tiêu (target_vector) và vector đặc trưng của các ảnh đã lưu trữ. Cuối cùng, hệ thống sắp xếp các ảnh theo độ tương đồng giảm dần, lọc kết quả theo ngưỡng do người dùng thiết lập và hiển thị danh sách các ảnh có mức độ tương đồng cao nhất. Đồng thời, thông tin bài báo liên quan, bao gồm DOI, tiêu đề, tạp chí, năm xuất bản và danh sách tác giả, cũng được hiển thị nhằm cung cấp ngữ cảnh đầy đủ cho người dùng.



Hình 1. Kiến trúc tổng quát xử lý ảnh tương đồng



a) Hình gốc b) Hình 6 kênh của từng lớp trong trực quan hóa bản đồ đặc trưng

Hình 2. Ảnh gốc và trực quan hóa các đặc trưng tương đồng

3.2. Trực quan hóa đặc trưng của ResNet50

Việc trực quan hóa đặc trưng của ResNet50 giúp hiểu rõ hơn về cách mô hình trích xuất và xử lý thông tin từ ảnh đầu vào. Quá trình này cung cấp góc nhìn sâu hơn về cách ResNet50 nhận diện và phân biệt các đặc trưng hình ảnh, từ đó hỗ trợ đánh giá cũng như tối ưu hóa mô hình.

Một trong những kỹ thuật quan trọng được áp dụng là Activation Map Visualization (trực quan hóa bản đồ đặc trưng). Khi hình ảnh được đưa vào mạng ResNet50, nó sẽ trải qua nhiều lớp tích chập (convolutional layers). Mỗi lớp có nhiệm vụ trích xuất các đặc trưng ở các cấp độ trừu tượng khác nhau, hình thành nên bản đồ đặc trưng (feature maps). Những bản đồ này giúp xác định các vùng mà mô hình tập trung vào, bao gồm cạnh, kết cấu, hình dạng và các chi tiết phức tạp hơn trong ảnh.

Trong nghiên cứu này, các bản đồ đặc trưng từ bốn lớp chính của ResNet50 (layer1, layer2, layer3, layer4) được trích xuất và trực quan hóa. Hình 2 minh họa ảnh gốc (a) và cách ResNet50 nhận diện đặc trưng trên ảnh đầu vào (b) bằng cách chọn tối đa 6 kênh đầu tiên từ từng lớp khác nhau. Các giá trị trong bản đồ đặc trưng được chuẩn hóa nhằm cải thiện độ rõ nét và được lưu lại dưới dạng hình ảnh trực quan.

Phương pháp trực quan hóa này giúp làm rõ các đặc trưng mà mô hình đã học được, ngay cả khi bằng mắt thường, ảnh gốc có thể không có sự tương đồng rõ ràng. Điều này giúp người quan sát dễ dàng phân tích và đánh giá quá trình học của mô hình. Lớp đầu tiên trích xuất các đặc trưng cơ bản như cạnh và góc. Các lớp sâu hơn phát hiện các đặc trưng phức tạp hơn, liên quan đến hình dạng và cấu trúc tổng thể của đối tượng. Giới hạn hiển thị 6 kênh đầu tiên giúp đơn giản hóa việc quan sát, giảm bớt sự phức tạp trong hình ảnh trực quan nhưng vẫn giữ được những thông tin quan trọng nhất mà mô hình đã học được. Phương pháp này không chỉ giúp phân tích quá trình học của mô hình mà còn hỗ trợ điều chỉnh và tối ưu hóa mạng nơ-ron, nâng cao hiệu suất trong các bài toán nhận diện và phân loại hình ảnh.

4. KẾT QUẢ THỰC NGHIỆM

4.1. Thu thập dữ liệu

Tập dữ liệu hình ảnh được trích xuất từ các bài báo khoa học bằng ngôn ngữ tiếng Anh của Tạp chí Khoa học Đại học Cần Thơ được sử dụng trong nghiên cứu. Tuy nhiên, do số lượng ảnh thu được từ nguồn này còn hạn chế, tập dữ liệu bổ sung từ Kaggle đã được sử dụng nhằm đảm bảo quá trình huấn luyện mô hình đạt hiệu quả cao hơn.

Bảng 1. Bảng dữ liệu ảnh dùng phân lớp

Class (Lớp)	Số ảnh
Biểu đồ cột (Bar chart)	1.038
Biểu đồ miền (Area Chart)	285
Ảnh chụp (Photograph)	75
Sơ đồ (Diagram)	1.339
Lưu đồ (Flowchart)	1.229
Đồ thị (Graph)	1.053
Biểu đồ tăng trưởng (Growth chart)	844
Ảnh dạng văn bản (Text)	2.984
Biểu đồ tròn (Pie chart)	1.327
Bảng dữ liệu (Table)	1.601
Hình ảnh thí nghiệm (Experiment image)	274
Tổng cộng	**12.049**

Tổng thể, dữ liệu hình ảnh được phân thành 11 lớp, bao gồm 12.049 ảnh. Trong đó, 8 lớp ảnh được lấy từ tập dữ liệu đồ thị trên Kaggle, trong khi 3 lớp ảnh được trích xuất từ các bài báo khoa học. Do tập dữ liệu chứa hình ảnh với kích thước không đồng nhất, việc chuẩn hóa kích thước là cần thiết để đảm bảo tính nhất quán trước khi đưa vào quá trình huấn luyện mô hình. Mô tả chi tiết về các lớp ảnh có trong tập dữ liệu được trình bày như Bảng 1.

Tập dữ liệu bao gồm các hình ảnh đã được gán nhãn, được phân thành 11 lớp khác nhau, mỗi lớp đại diện cho một loại đối tượng hoặc chủ đề cụ thể. Để đảm bảo tính hiệu quả trong quá trình huấn luyện mô hình, tập dữ liệu được chia thành hai phần theo tỷ lệ 8:2, trong đó 80% dữ liệu được sử dụng cho huấn luyện (training) và 20% còn lại dành cho kiểm tra (testing). Việc lựa chọn tỷ lệ này nhằm đảm bảo mô hình có đủ dữ liệu để học và tổng quát hóa tốt các đặc trưng của tập huấn luyện, đồng thời giữ lại một phần dữ liệu chưa từng thấy để đánh giá độ chính xác và khả năng dự đoán. Điều này giúp kiểm chứng tính tổng quát của mô hình khi áp dụng vào dữ liệu mới, đảm bảo rằng mô hình không bị quá khớp (overfitting) với tập huấn luyện.

Nghiên cứu này được thực hiện trên nền tảng ngôn ngữ lập trình Python, sử dụng các thư viện quan trọng phục vụ cho việc triển khai và huấn luyện mô hình học máy (Machine Learning - ML). Toàn bộ quá trình được thực hiện trên Google Colab, một môi trường điện toán đám mây hỗ trợ GPU, giúp tăng tốc độ huấn luyện mô hình. Google Colab cung cấp sẵn các thư viện như PyTorch, Torchvision và cho phép tải mô hình sau khi huấn luyện về máy cục bộ. Nghiên cứu sử dụng Python phiên bản 3.8 trở lên để phát triển mã nguồn và triển khai các thư viện

hỗ trợ cần thiết cho quá trình huấn luyện và đánh giá mô hình.

Trong nghiên cứu này, hiệu suất của các mô hình ResNet50, AlexNet và VGG16 được đánh giá thông qua các độ đo phổ biến trong bài toán phân loại, bao gồm Precision (Độ chính xác), Recall (Độ nhạy) và độ chính xác F1-score. Những độ đo này giúp đánh giá mức độ chính xác và khả năng tổng quát hóa của mô hình trên tập dữ liệu kiểm tra

4.2. Kết quả thực nghiệm

Mô hình học sâu đã trở thành công cụ quan trọng trong nhiều bài toán xử lý hình ảnh, đặc biệt là nhận diện và phân loại ảnh. Trong nghiên cứu này, mô hình ResNet50 được huấn luyện trên tập dữ liệu đã mô tả, sau đó đánh giá kết quả thông qua các thông số như độ chính xác (Precision), độ nhạy (Recall) và F1-score. Để đánh giá hiệu quả của mô hình ResNet50, hai mô hình CNN phổ biến khác là AlexNet và VGG16 cũng được sử dụng để so sánh. Việc so sánh dựa trên các thông số Precision, Recall và F1-score nhằm đánh giá khả năng tổng quát hóa và độ chính xác của từng mô hình.

Mô hình ResNet50: ResNet50 là mô hình CNN sử dụng Residual Blocks để giải quyết vấn đề vanishing gradient. Thực nghiệm cho thấy ResNet50 hội tụ sau 10 epoch. Kết quả huấn luyện ghi nhận các độ đo Precision, Recall và F1-score đạt mức khá cao (Bảng 2).

Biểu đồ Train và Test Loss (bên trái): Loss trên tập Train giảm qua các epoch, trong 2 epoch đầu đã giảm đáng kể. Điều này cho thấy mô hình học tốt trên tập huấn luyện. Loss trên tập Test giảm chậm hơn so với tập Train, có thể do dữ liệu phức tạp hoặc mô hình cần thời gian để tổng quát hóa trên tập Test. Sau đó, Test Loss hội tụ nhanh và dao động trong khoảng 0,2 - 0,25, cho thấy mô hình bắt đầu hội tụ.

Biểu đồ trực quan hóa hiệu suất của mô hình được thể hiện ở Hình 3. Biểu đồ Train và Test Loss (bên trái): Loss trên tập Train giảm qua các epoch, trong 2 epoch đầu đã giảm đáng kể. Điều này cho thấy mô hình học tốt trên tập huấn luyện. Loss trên tập Test giảm chậm hơn so với tập Train, có thể do dữ liệu phức tạp hoặc mô hình cần thời gian để tổng quát hóa trên tập Test. Sau đó, Test Loss hội tụ nhanh và dao động trong khoảng 0,2 - 0,25, cho thấy mô hình đã bắt đầu hội tụ.

Biểu đồ Train và Test F1-score (bên phải): F1-score trên tập Train tăng nhanh và đạt giá trị cao gần 1 từ khoảng epoch 3, sau đó ổn định ở mức khoảng 0,97. F1-score trên tập Test tăng trong 3 epoch đầu và dao động nhẹ sau đó. Giá trị Test F1-score đạt cao nhất tại epoch 10 với 0,9004, cho thấy kết quả khá tốt.

Bảng 2. Độ đo đạt được sau 10 epoch của mô hình ResNet50

Precision	Recall	F1-score
0,8795	0,9275	0,9004

Như vậy, mô hình ResNet50 hội tụ từ 8 đến 10 epoch, khi Loss và F1-score đều ổn định. Mô hình đạt hiệu suất cao nhất tại epoch 10, khi Test F1-score đạt 0,9004, cho thấy khả năng tổng quát hóa tốt nhất tại thời điểm này. Về độ lỗi, Train Loss giảm liên tục và hội tụ ở mức rất thấp (0,015 - 0,020) từ epoch 7 trở đi, trong khi Test Loss duy trì trong khoảng 0,14 - 0,19. Về độ chính xác, Train F1-score đạt mức cao khoảng 0,97, còn Test F1-score tăng ổn định và đạt đỉnh ở epoch 10 với 0,9004.

Ngoài mô hình ResNet50, hai mô hình VGG16 và AlexNet cũng được thử nghiệm để so sánh và đánh giá.

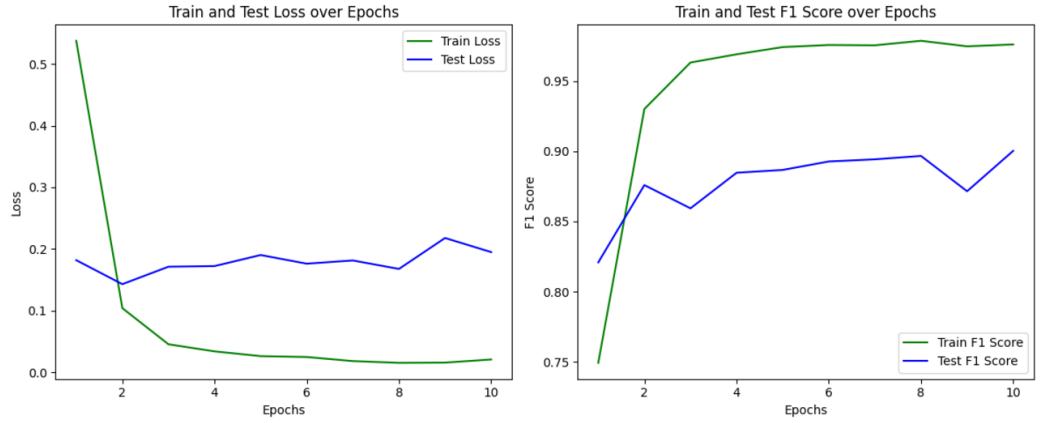
Mô hình AlexNet: Biểu đồ trực quan hóa hiệu suất của mô hình AlexNet trong quá trình huấn luyện được thể hiện ở Hình 4.

Biểu đồ Train và Test Loss (bên trái): Loss trên tập Train giảm đều qua các epoch, đạt 0,0380 ở epoch 10. Loss trên tập Test giảm dần trong vài epoch đầu, nhưng sau đó dao động trong khoảng 0,3-0,35 từ epoch 6 đến epoch 10, cho thấy mô hình bị overfitting.

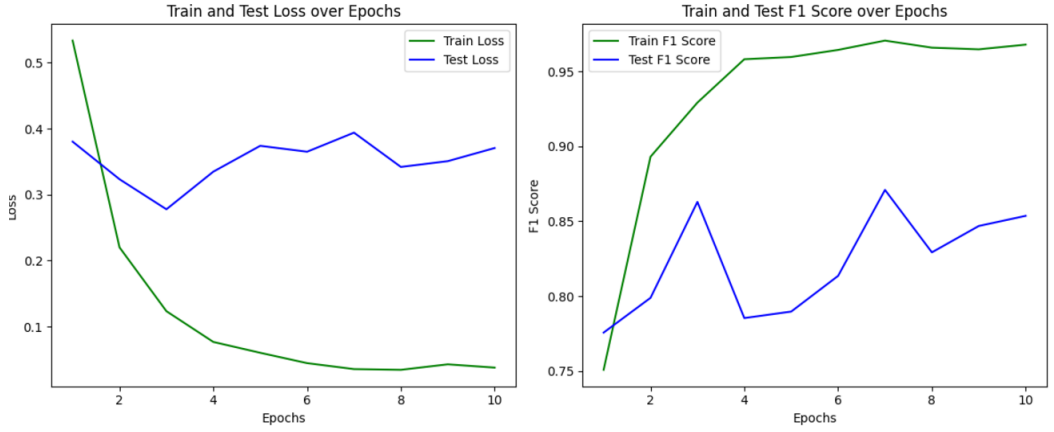
Biểu đồ Train và Test F1-score (bên phải): F1-score trên tập Train tăng đều và đạt đỉnh ở gần mức 0,96 vào khoảng 8-10 epoch. F1-score trên tập Test tăng ổn định trong các epoch đầu và đạt giá trị khoảng 0,87 ở epoch 7, nhưng giảm nhẹ về sau.

Như vậy, mô hình AlexNet hội tụ từ khoảng 8-10 epoch, với độ lỗi thấp và độ chính xác cao. Tuy nhiên, sự chênh lệch giữa Train và Test F1-score (Train khoảng 0,96, Test khoảng 0,85) cho thấy mô hình có dấu hiệu overfitting

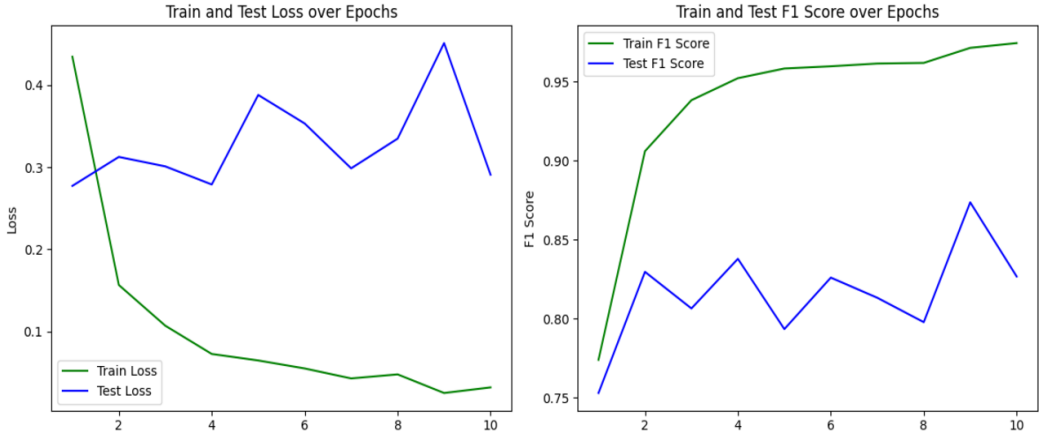
Mô hình VGG16: Biểu đồ trực quan hóa hiệu suất của mô hình VGG16 trong quá trình huấn luyện được thể hiện ở Hình 5.



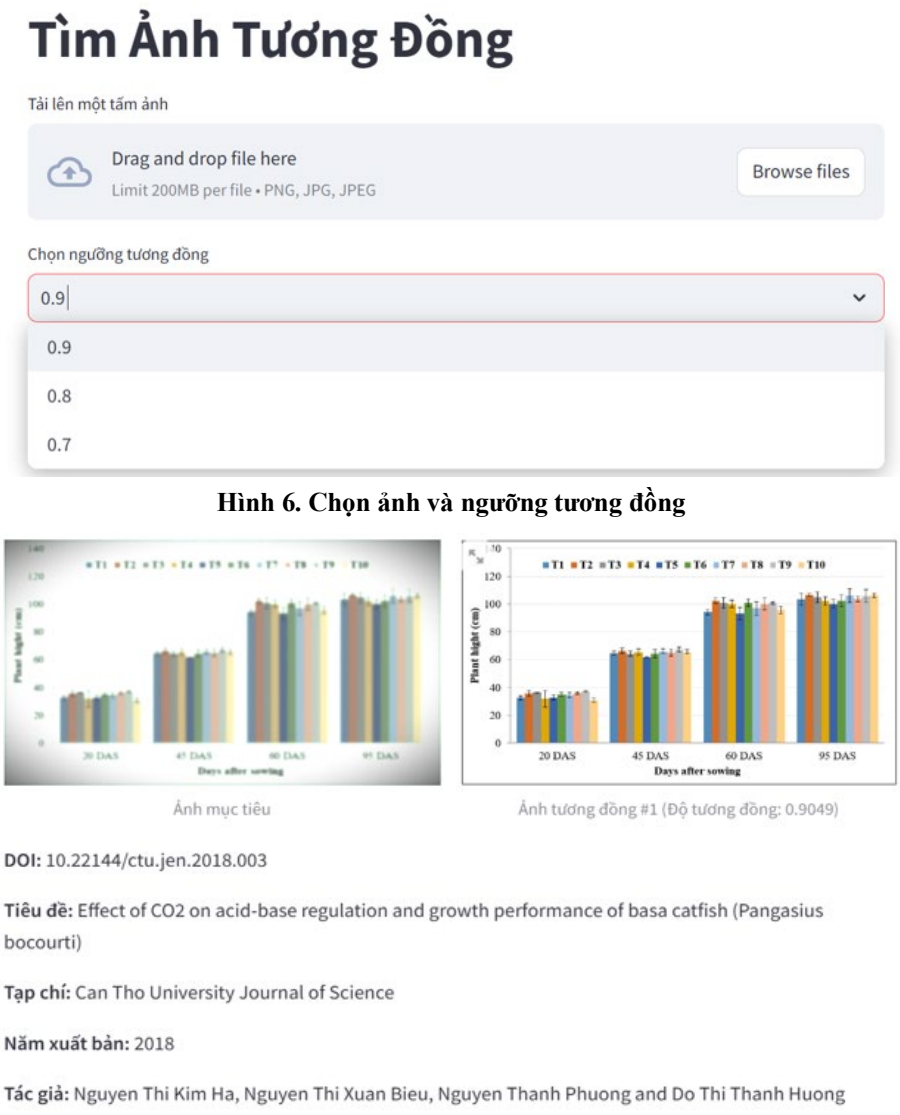
Hình 3. Biểu đồ trực quan hóa hiệu suất của mô hình ResNet50 trong huấn luyện



Hình 4. Biểu đồ trực quan hóa mô hình AlexNet trong quá trình huấn luyện



Hình 5. Biểu đồ trực quan hóa mô hình VGG16 trong quá trình huấn luyện



Hình 6. Chọn ảnh và ngưỡng tương đồng

Hình 7. Kết quả tìm kiếm ảnh tương đồng (ảnh bên phải) so với ảnh mục tiêu (ảnh bên trái)

Biểu đồ Train và Test Loss (bên trái): Loss trên tập Train giảm trong 2 epoch đầu từ khoảng 0,4 xuống khoảng 0,1, sau đó hội tụ ở mức khoảng 0,05 sau 10 epoch. Loss trên tập Test dao động nhiều và không giảm đều, cho thấy mô hình có thể gặp phải vấn đề overfitting.

Biểu đồ Train và Test F1-score (bên phải): F1-score trên tập Train tăng trong 2 epoch đầu từ khoảng 0,75 lên 0,9, sau đó hội tụ ở mức khoảng 0,97 sau 10 epoch. F1-score trên tập Test tăng trong 3 epoch đầu từ khoảng 0,75 lên 0,85 và dao động từ epoch 4 đến epoch 10 trong khoảng 0,8 - 0,85.

Như vậy, mô hình VGG16 học hiệu quả trên tập huấn luyện nhưng biểu hiện trên tập kiểm tra chưa hoàn hảo, cho thấy mô hình chưa hội tụ tốt trên tập kiểm tra.

Các độ đo của 3 mô hình ResNet50, AlexNet và VGG16 được thể hiện như Bảng 3.

Bảng 3. So sánh độ đo các mô hình ResNet50, AlexNet và VGG16

Models	Precision	Recall	F1-score	Thời gian xử lý (s)
ResNet50	**0,8795**	**0,9275**	**0,9004**	196
AlexNet	0,8503	0,8980	0,8709	157
VGG16	0,8818	0,8785	0,8735	257

Như vậy, thực nghiệm cho thấy, ResNet50 chứng minh được khả năng tổng quát hóa tốt nhất, với Test Loss thấp và F1-score cao nhất. AlexNet và

VGG16 tuy vẫn là những mô hình mạnh, nhưng hiệu suất chưa đạt được như ResNet50 trong bài toán phân loại ảnh này, mặc dù thời gian xử lý của mô hình ResNet50 chưa phải tốt nhất. Do đó, ResNet50 là lựa chọn tối ưu cho các bài toán nhận diện và phân loại ảnh có độ phức tạp cao.

Từ kết quả thực nghiệm trên, hệ thống thực hiện tính toán tương đồng ảnh thông qua ảnh đầu vào so với tập cơ sở dữ liệu ảnh được trích xuất trước đó từ các bài báo khoa học được xây dựng. Khi nhận ảnh đầu vào, hệ thống thực hiện các bước tiền xử lý dữ liệu và phân tích, sau đó trả về danh sách các hình ảnh có mức độ tương đồng ngưỡng tương đồng cho trước.

Đầu tiên, người dùng chọn ảnh đầu vào (ảnh mục tiêu) cần tìm ảnh tương đồng, sau đó chọn ngưỡng tương đồng là 0,9 (xem Hình 6). Sau đó, hệ thống tìm được một ảnh trong cơ sở dữ liệu có độ tương đồng là 0,9049, kèm theo những thông tin về bài báo khoa học mà ảnh được sử dụng như: DOI, tiêu đề, tạp chí, năm xuất bản và các tác giả (xem Hình 7).

5. KẾT LUẬN

Hệ thống cơ sở dữ liệu chứa hình ảnh từ các bài báo khoa học, kết hợp với phương pháp phân loại và trích xuất đặc trưng phục vụ cho tìm kiếm hình ảnh tương đồng với độ chính xác cao được xây dựng trong nghiên cứu. Bằng cách áp dụng mô hình ResNet50, hệ thống có khả năng nhận diện và so sánh hình ảnh một cách hiệu quả, giúp tối ưu hóa quá trình truy xuất thông tin trong lĩnh vực khoa học. Hơn nữa, nghiên cứu đã chứng minh tính hiệu quả của việc sử dụng các thuật toán tối ưu hóa nhằm cải thiện độ chính xác trong phân tích hình ảnh. Hiệu suất của mô hình ResNet50 cũng được đánh giá thông qua việc so sánh với hai mô hình học sâu phổ biến khác là AlexNet và VGG16 trong bài toán phân loại hình ảnh khoa học. Kết quả cho thấy ResNet50 là mô hình phù hợp nhất để áp dụng vào các bài toán phân loại ảnh có độ phức tạp cao.

Tuy nhiên, để cải thiện hơn nữa hiệu suất, nghiên cứu trong tương lai có thể tập trung vào tối ưu hóa quá trình huấn luyện, áp dụng các phương pháp giảm overfitting và thử nghiệm với nhiều mô hình học sâu tiên tiến hơn.

LỜI CẢM TẠ

Đề tài này được tài trợ bởi Trường Đại học Cần Thơ, Mã số: T2024-60.

TÀI LIỆU THAM KHẢO (REFERENCES)

Chechik, G., Sharma, V., Shalit, U., & Bengio, S. (2010). Large scale online learning of image similarity through ranking. *Journal of Machine Learning Research*, 11, 1109–1135. https://doi.org/10.1007/978-3-642-02172-5_2

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. *International Conference on Machine Learning*.

Guyon, I., & Elisseeff, A. (2006). An introduction to feature extraction. *In Feature Extraction: Foundations and Applications* (pp. 1–25). Springer.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

Hirematch, P. S., & Puijari, J. (2007). Content-based image retrieval based on color, texture, and shape feature using image and its complement. *International Journal of Computer Science and Security*, *1*(4), 25–35.

Khan, A. S., & Shafique, M. (2018). Content-based image retrieval using histogram of oriented gradients and SIFT features. *International Journal of Computer Applications*, *179*(26), 1–8. https://doi.org/10.5120/ijca2018917508

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (pp. 1097–1105).

Li, X., & Qian, Y. (2019). A novel image similarity measurement method based on deep learning. *Sensors*, *19*(14), 3060. https://doi.org/10.3390/s19143060

Lu, J., Ma, C. X., Zhou, Y. R., Luo, M. X., & Zhang, K. B.. (2019). Multi-feature fusion for enhancing image similarity learning. *IEEE Access*, 7, 167547–167556. https://doi.org/10.1109/ACCESS.2019.2953078

Luu, T. H., Phuc, P. N. K., Yu, Z., Pham, D. D., & Cao, H. T. (2022). Face Mask Recognition for Covid-19 Prevention. *Computers, Materials & Continua*, 73(2). https://doi.org/10.32604/cmc.2022.029663

Luu, T., Ky Phuc, P., Lam, T., Yu, Z., & Lam, V. (2023). Ensembling techniques in solar panel quality classification. *International Journal of Electrical and Computer Engineering (IJECE),*

13(5), 5674-5680. doi:http://doi.org/10.11591/ijece.v13i5.pp5674-5680

McCorduck, P. (2004). *Machines who think* (2nd ed.). AK Peters. https://doi.org/10.1201/9780429258985

Russakoff, D. B., Tomasi, C., Rohlfing, T., & Maurer, C. R. (2004). Image similarity using mutual information of regions. In *European Conference on Computer Vision* (pp. 596–607). Springer.

Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach (4th ed.). Pearson*. https://doi.org/10.1007/978-3-540-24672-5_47

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D.,& Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-9). https://doi.org/10.1109/CVPR.2015.7298594

Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning* (pp. 6105-6114).

Zagoruyko, S. (2016). *Wide residual networks. arXiv preprint arXiv:1605.07146*.

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J. B., Yuan, J., & Yan, L. Li. (2021). A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity*, *2021*(1), 8812542.