MÔ HÌNH HỌC MÁY TRONG XÂY DỰNG CÁC DỊCH VỤ DỰA TRÊN ĐỊNH DANH

A MACHINE LEARNING MODEL FOR IDENTITY-BASED SERVICES

Phạm Ngọc Huân[1], Nguyễn Lương Bằng[2], Phí Trung Hiếu[3], Nguyễn Mạnh Cường[4,*]

TÓM TẮT

Các dịch vụ dựa trên định danh ngày càng phổ biến và mang lại nhiều tiện ích cho người dùng. Định danh tự động giúp mang lại các trải nghiệm dịch vụ cao cấp cho người thụ hưởng trong rất nhiều lĩnh vực như giáo dục, nghỉ dưỡng, chăm sóc sức khỏe, chăm sóc khách hàng. Trong bài báo này, chúng tôi giới thiệu một hệ thống hỗ trợ các dịch vụ dựa trên định danh tự động với mô hình được đề xuất là mô hình học chuyển tiếp kết hợp giữa mô hình mạng nơ ron nhân tạo CNN và mô hình máy véc tơ hỗ trợ SVM. Một kiến trúc CNN được đề xuất và được sử dụng như là bộ trích rút thuộc tính cho mô hình SVM làm nhiệm vụ phân lớp các đối tượng. Các kết quả thu được cho thấy sự cải thiện đáng kể về độ chính xác trong định danh cũng như thời gian huấn luyện so với các mô hình riêng lẻ. Hệ thống được xây dựng thành công hứa hẹn sẽ giúp tăng chất lượng, giá trị cho các dịch vụ và gia tăng sự hài lòng của người dùng.

Từ khóa: *Hệ thống định danh khuôn mặt, học chuyển tiếp, mạng nơ ron, trích chọn thuộc tính, SVM.*

ABSTRACT

Identity-based services are becoming more and more popular and bring many benefits to users. Particularly, automatic identification helps bring high-class service experiences to beneficiaries in many fields such as education, resort travel, health care, customer care. In this paper, we introduce a system that supports automatic identity-based services with the proposed model as a forward learning model combining the CNN artificial neural network model and the machine learning model support vector SVM. A CNN architecture is proposed and used as an attribute extractor for the SVM model to classify objects. The obtained results show a significant improvement in identification accuracy as well as training time compared to the individual models. The successfully built system promises to help increase the quality and value of services and increase user satisfaction.

Keywords**:** *Face recognition system, transfer learning, neural network, feature extraction, SVM.*

[1]Lớp CNT T03 - K14, hoa CNTT, Trường Đại học Công nghiệp Hà Nội
[2]Lớp CNTT 01- K12, hoa CNTT, Trường Đại học Công nghiệp Hà Nội
[3]Lớp CNTT04 - K14, hoa CNTT, Trường Đại học Công nghiệp Hà Nội
[4]Khoa CNTT, Trường Đại học Công nghiệp Hà Nội
[*]Email: manhcuong.nguyen@haui.edu.vn

1. GIỚI THIỆU

Con người khi sinh ra đã có những đặc điểm sinh học tự nhiên riêng biệt để phân biệt người này với người kia, rất khó để trùng lặp. Các đặc trưng trên khuôn mặt là những đặc điểm riêng trên khuôn mặt mỗi người gần như không thay đổi theo thời gian. Chính vì thế việc xác định, nhận dạng mặt người thông quan các đặc trưng sinh trắc học sẽ đảm bảo được độ chính xác, tin cậy cao.

Trong bài báo này, chúng tôi giới thiệu một hệ thống hỗ trợ các dịch vụ dựa trên định danh hỗ trợ các dịch vụ tự động sử dụng mô hình học chuyển tiếp (Transfer Learning) [2, 3], là sự kết hợp của hai mô hình CNN và SVM, gọi là DSVM [1]. Khác với các mô hình thay thế CNN-SVM được giới thiệu trong [4] khi mà CNN và SVM được huấn luyện đồng thời trong cùng một mô hình, DSVM tách rời quá trình huấn luyện của CNN và SVM. Trong đó, một kiến trúc CNN được thiết kế làm nhiệm vụ lọc nhiễu, trích rút đặc trưng và giảm số chiều trên ảnh. Véc tơ đặc trưng thu được từ mô hình này được xem như đầu vào của một mô hình SVM đa lớp. Sự kết hợp này được kỳ vọng sẽ tận dụng được ưu điểm của cả hai mô hình khi mà CNN thực hiện việc trích chọn đặc trưng rất hiệu quả trên ảnh, trong khi SVM lại có độ chính xác phân lớp tốt nếu dữ liệu đầu vào được tiền xử lý hiệu quả, kết quả thực nghiệm mô hình cải thiện đáng kể độ chính xác đáng kể so với các mô hình riêng lẻ.

Phần tiếp theo của bài báo có cấu trúc như sau: Phần 2 sẽ giới thiệu về bài toán định danh khuôn mặt. Trong phần 3, chúng tôi trình bày các mô hình được sử dụng trong hệ thống. Phần 4 được dành để trình bày một số kết quả thực nghiệm và cuối cùng là một số kết luận.

2. BÀI TOÁN ĐỊNH DANH DỰA TRÊN ẢNH KHUÔN MẶT

2.1. Tổng quan bài toán định danh dựa trên ảnh khuôn mặt

Định danh khuôn mặt bao gồm quá trình thu thập dữ liệu khuôn mặt từ hình ảnh hoặc video, sau đó sử dụng các mô hình trí tuệ nhân tạo để tiến hành phân loại. Mô hình thông thường của các hệ thống định danh gồm các quá trình sau:

- Trích chọn đặc trưng (Feature Engineering): là quá trình thực hiện trích xuất, trích chọn các đặc trưng (thuộc tính) quan trọng từ dữ liệu thô để sử dụng làm đại diện cho các mẫu dữ liệu huấn luyện.

- Phân lớp dữ liệu: Sau khi trích xuất các đặc trưng, các kỹ thuật học máy sẽ được áp dụng để xây dựng một bộ

phân lớp các khuôn mặt. Các bộ phân lớp dựa vào các thuật toán như KNN, Decision tree, Naive Bayes, SVM,…

2.2. Các bài toán con cần giải quyết

Phát hiện khuôn mặt(Face detection)

Face detection là một trong những bài toán điển hình trong lĩnh vực Computer Vision. Tương tự như bài toán Object detection nhưng thay vì phát hiện các vật thể một cách chung chung trong ảnh, bài toán face detection tập trung vào việc phát hiện các khuôn mặt. Đầu vào của các mô hình face detection là một hoặc một tập ảnh. Đầu ra của các mô hình này là một tập hợp các bounding box với mỗi bounding box bao gồm 4 giá trị để miêu tả một hình chữ nhật.

Nhận dạng khuôn mặt(Face *recognition)*

Nhận điện khuôn mặt là một phương pháp xác định danh tính của một người dựa trên dữ liệu khuôn mặt của họ. Các đặc trưng trên khuôn mặt mỗi người là riêng biệt và rất khó trùng lặp. Bài toán sử dụng dữ liệu đầu vào là hình ảnh khuôn mặt. Sau khi phân tích dữ liệu đầu vào hệ thống phân tích các đặc điểm sinh trắc học trên khuôn mặt như mắt, môi gò má, trán và đường viền của môi, tai và cằm... nhiệm vụ tiếp theo là đưa những dữ liệu về khuôn mặt đi so sánh với những khuôn mặt đã có trong database.

3. MÔ HÌNH HỌC CHUYỂN TIẾP CHO BÀI TOÁN ĐỊNH DANH

3.1. Mô hình phát hiện khuôn mặt TinyFace Detector

Tiny detector là một công cụ dò tìm khuôn mặt thời gian thực rất hiệu quả, nhỏ và tiêu tốn ít tài nguyên. Bù lại, nó hoạt động kém khi phát hiện các khuôn mặt nhỏ. Mô hình này cực kỳ thân thiện với thiết bị di động và web với kích cỡ chỉ 190KB. Nó cơ bản là phiên bản nhỏ hơn của Tiny Yolo V2 tuy nhiên nó thay thế các khối chập thông thường của Yolo bằng các khối chập có thể phân tách theo chiều sâu. Yolo hoàn toàn phức hợp, do đó có thể dễ dàng thích ứng với các kích thước hình ảnh đầu vào khác nhau để đánh đổi độ chính xác cho hiệu suất.

3.2. Định danh dựa trên mô hình học chuyển tiếp

Mô hình máy véc tơ hỗ trợ SVM

SVM là một thuật toán học có giám sát. Nó có thể sử dụng cho cả việc phân loại hoặc hồi quy. Tuy nhiên nó được sử dụng chủ yếu cho việc phân loại. Mục tiêu của mô hình SVM là tìm ra một siêu phẳng (hyperplane) phân tách tốt nhất tập dữ liệu huấn luyện thành hai phần riêng biệt.

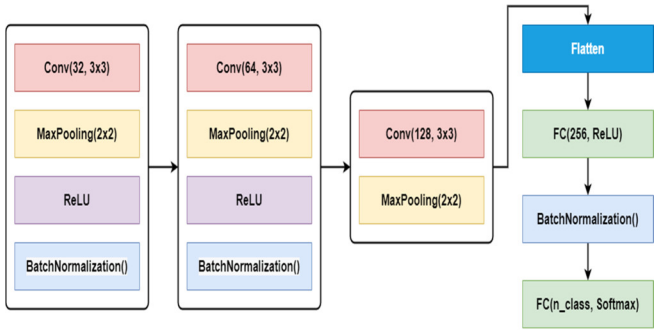
SVM có thể hoạt động hiệu quả khi sử dụng kết hợp với các hàm nhân (Kernel functions) giúp cho nó không còn bị giới hạn bởi việc chỉ phân lớp tốt trên dữ liệu có khả năng phân tách tuyến tính. Bản chất của phương pháp Kernel-SVM là chuyển không gian dữ liệu ban đầu thành một không gian mới nhiều chiều hơn mà ở đó cho khả năng phân lớp dễ dàng hơn.

Mô hình SVM được cho là rất hiệu quả trong bài toán phân lớp dữ liệu vector số. Tuy nhiên, nó có thể đòi hỏi các thao tác tiền xử lý dữ liệu hay trích chọn đặc trưng phức tạp.

Mô hình mạng nơ ron tích chập

Hiệu quả của mô hình mạng nơ ron tích chập (CNN) phụ thuộc rất nhiều vào kiến trúc của mạng. Trong phần này, chúng tôi đề xuất một kiến trúc mạng đơn giản và có thể phù hợp với mô hình học chuyển tiếp. Mục tiêu là thiết kế một mạng không quá phức tạp nhưng vừa đủ cho các mục đích trích chọn đặc trưng và cắt giảm số chiều của ảnh.

Mô hình được đề xuất (hình 1) bao gồm 3 khối chính với 3 lớp tích chập (*Convolutional layer*). Lớp tích chập có số lượng các bộ lọc (*Filter*) được đề xuất lần lượt là 32, 64 và 128. Sau mỗi lớp tích chập là một lớp gộp (Max pooling layer) với kích thước 2×2.

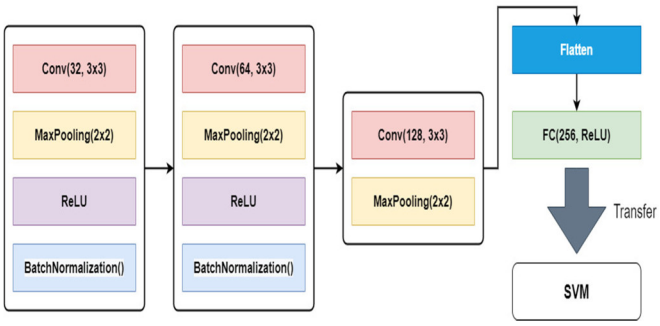


Hình 1. Mô hình mạng nơ ron tích chập được đề xuất

Mô hình học chuyển tiếp DSVM

Trong phần này chúng tôi đề xuất mô hình chuyển tiếp, gọi là DSVM, cho bài toán định danh dựa trên khuôn mặt. Mô hình này là sự kết hợp của các mô hình SVM và mạng CNN đã được đề xuất ở trên. Mạng CNN đóng vai trò là bộ lọc nhiễu, trích chọn đặc trưng trong khi mô hình SVM làm nhiệm vụ phân lớp.

Trước tiên mạng được huấn luyện với đầy đủ các bước. Các quá trình tiếp theo được thực hiện như sau: Mỗi ảnh đầu vào được đưa qua mạng CNN đã được huấn luyện ở trên nhưng bỏ đi hai bước cuối cùng. Đầu ra của được lưu lại dưới dạng một véc tơ đặc trưng 256 chiều. Trong pha tiếp theo, véc tơ đặc trưng này được chuyển tiếp tới một mô hình SVM sử dụng chiến lược phân đa lớp *one vs one*. Hình 2 dưới đây minh họa một cách trực quan hơn mô hình DSVM.



Hình 2. Mô hình học chuyển tiếp DSVM

4. MỘT SỐ KẾT QUẢ THỰC NGHIỆM

Chúng tôi tiến hành thực nghiệm các mô hình trên môi trường *Google Colab* với ngôn ngữ lập trình *Python*. Trước

tiên, mạng nơ ron tích chập với các bước như mô tả ở trên được huấn luyện và phân lớp trên các bộ dữ liệu. Các kết quả nhận dạng của mạng này được coi là kết quả của mô hình CNN riêng lẻ và được dùng để so sánh với mô hình học chuyển tiếp DSVM. Tương tự như vậy, một mô hình SVM riêng lẻ cũng được thực thi.

Các tham số cho mô hình SVM riêng lẻ và trong DSVM được xác định như sau. Tham số điều chỉnh *C* trong được chọn cố định là 1. Hàm nhân được sử dụng là *Gaussian Kernel* với tham số σ được chọn lần lượt trong tập {10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}}. Với mỗi bộ dữ liệu và mỗi bộ tham số, chúng tôi sử dụng thủ tục 5-*fold cross validation* để tính ra kết quả trung bình. Cuối cùng, kết quả thực nghiệm của các mô hình (bao gồm độ chính xác phân lớp - *Accuracy*, thời gian huấn luyện, thời gian phân lớp) trên mỗi bộ dữ liệu là kết quả trung bình tốt nhất theo *Accuracy* của các bộ tham số trên tập dữ liệu đó.

4.1. Dữ liệu thực nghiệm

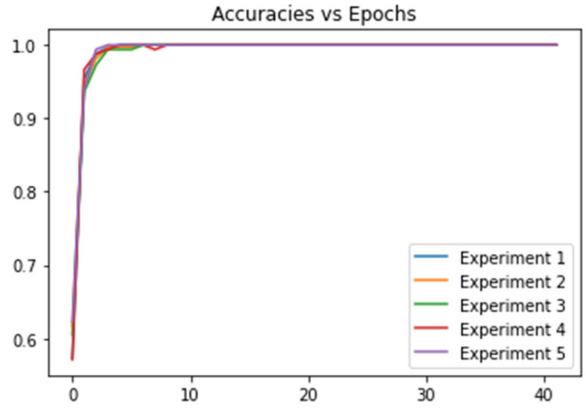
Quá trình thực nghiệm được tiến hành trên 6 bộ dữ liệu ảnh khuôn mặt người. Trong đó, bộ TLFace do chúng tôi tự thu thập, các bộ còn lại là các bộ dữ liệu phổ biến có thể dễ dàng thu thập trên Internet. Chi tiết của các bộ dữ liệu xin xem trong bảng 1.

Bảng 1. Dữ liệu thực nghiệm

Dữ Liệu	Số mẫu	Số lớp	Kích thước (pixel)	Số kênh màu
TLFace	411	20	120×80	3
AT&T	400	40	112×92	1
Georgia Tech Face	750	50	211×151	3
AR Face	2392	92	165×120	3
The extended Yale Face	2432	38	192×168	1
FEI Face	2800	200	240×320	3

4.2. Kết quả thực nghiệm

Để kiểm tra tính ổn định, tốc độ hội tụ của mô hình CNN đề xuất ở trên. Chúng tôi tiến hành thử nghiệm mô hình này trên bộ dữ liệu TLFace. Hình 3 biểu thị độ chính xác phân lớp và giá trị của hàm tổn thất qua các lần lặp (*Epochs*) trong quá trình huấn luyện.



Hình 3. Độ chính xác phân lớp của CNN qua các lần lặp (Epochs)

Các kết quả cho thấy tốc độ hội tụ khá nhanh của mô hình này trên TLFace (sau khoảng 5 bước lặp). Quá trình huấn luyện cũng diễn ra ổn định, không có hiện tượng giảm đột ngột (*drop*) của *Accuracy*. Các kết quả thử nghiệm tiếp theo trên ba mô hình CNN, SVM và DSVM được trình bày trong Bảng 3. Trong đó, chỉ số về độ chính xác phân lớp trên các bộ dữ liệu huấn luyện (*Training set*) và dữ liệu kiểm tra (*Test set*) được báo cáo.

Bảng 2. Độ chính xác phân lớp trên bộ dữ liệu TLFace

Dữ liệu	CNN	SVM	DSVM
Training set	**100%**	98,97%	99,57%
Test set	95,13%	89,29%	**99,51**%

Qua các kết quả này ta có thể thấy: Mô hình CNN được đề xuất đảm bảo độ chính xác phân lớp tốt hơn mô hình SVM trên dữ liệu TLFace và với các tập tham số được chọn. Trong trường hợp này, CNN cũng cho kết quả tốt nhất trong ba mô hình trên tập dữ liệu huấn luyện. Tuy nhiên, với dữ liệu kiểm tra (test set), DSVM có độ chính xác đạt hơn 99%, tốt hơn so với hơn 95% của CNN và hơn 89% của SVM.

Tiếp theo, chúng tôi tiến hành thực nghiệm chi tiết hơn với các bộ dữ liệu còn lại.

Bảng 3. Kết quả thực nghiệm trên một số bộ dữ liệu khác

Dữ liệu	Accuracy (%)			Training time (s)			Testing time (s)		
	CNN	SVM	DSVM	CNN	SVM	DSVM	CNN	SVM	DSVM
AT&T	98,25	97,75	**99,25**	6,18	3,69	0,10	0,11	0,47	**0,01**
Georgia Tech Face	95,47	85,33	**99,60**	35,96	126,97	0,28	0,15	15,78	**0,05**
AR Face	99,96	65,34	**100**	70,95	820,16	3,39	**0,21**	102,84	0,57
The extended Yale Face	99,18	83,76	**99,84**	95,79	338,68	0,70	**0,3**	53,54	0,34
FEI Face	93,68	-	**99,21**	540,23	-	3,98	1,26	-	**1,06**

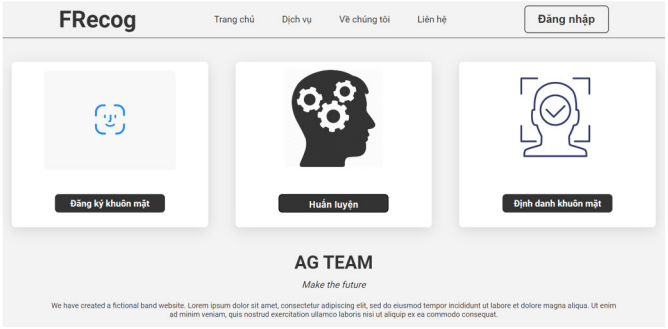
Bảng 3 trình bày các kết quả trên các độ đo *Accuracy*, thời gian huấn luyện và thời gian phân lớp của các mô hình trên các bộ dữ liệu còn lại. Các kết quả thử nghiệm cho thấy DSVM cho kết quả tốt hơn các mô hình CNN và SVM riêng lẻ về độ chính xác phân lớp trên tất cả các bộ dữ liệu thử nghiệm (giao động từ 99,21 tới 100%, trung bình đạt 99,58%). Các kết quả này cũng cho thấy một sự ổn định của DSVM trên các bộ dữ liệu khác nhau.

4.3. Hệ thống Demo

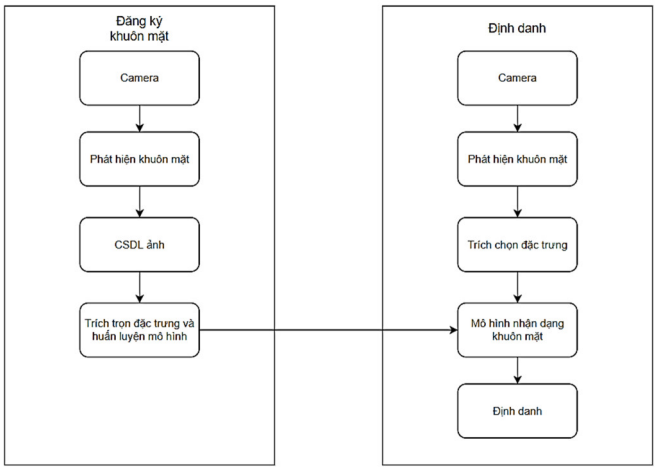
Sau quá trình thực nghiệm các mô hình, chúng tôi tiến hành xây dựng ứng dụng thử nghiệm. Ứng dụng này được cài đặt trên Django framework với ngôn ngữ lập trình Python và gồm 3 chức năng chính sau:

- Đăng ký khuôn mặt (Face Register)
- Huấn luyện mô hình (Train Model)
- Nhận diện khuôn mặt (Face Recognition)

Mô hình hoạt động của hệ thống được mô tả như trong hình 5.

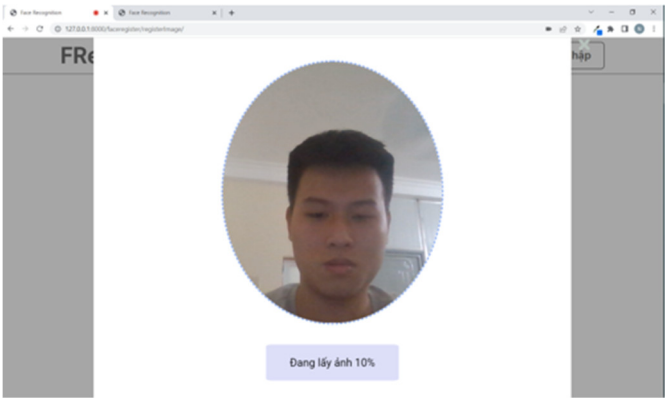


Hình 4. Giao diện trang chủ của hệ thống



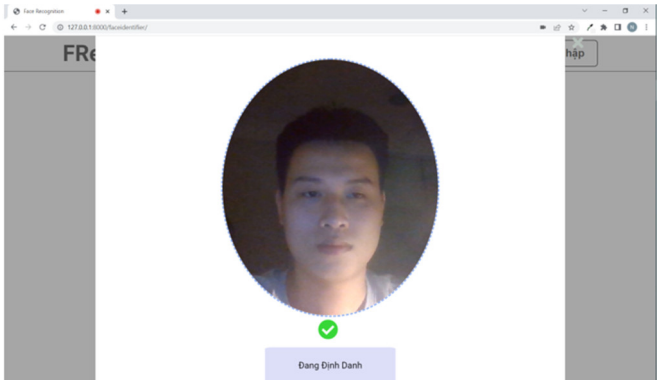
Hình 5. Mô hình hoạt động của hệ thống

Chức năng đăng ký khuôn mặt được sử dụng để lấy 20 ảnh của mỗi người dùng để phục vụ cho chức năng huấn luyện mô hình. Người dùng được yêu cầu di chuyển mặt vào vùng chỉ định để tiến hành đăng ký (hình 6).



Hình 6. Giao diện đăng ký khuôn mặt

Vì ứng dụng thử nghiệm hiện đang thực thi với số lớp nhỏ nên ta có thể huấn luyện trực tiếp trên CPU của máy chủ hoặc máy tính cá nhân trong khoảng 5 - 10 phút *với batch size 16*. Sau khi mô hình được huấn luyện, hệ thống sẽ sử dụng nó để tiến hành phân lớp (hình 7).



Hình 7. Giao diện chức năng định danh

5. KẾT LUẬN

Qua quá trình học hỏi, nghiên cứu, thử nghiệm và đánh giá, bằng việc tận dụng các nghiên cứu có trước về xử lý ảnh, phát hiện khuôn mặt, nhận diện khuôn mặt, xây dựng website, nhóm tác giả đã tìm hiểu, nghiên cứu, cài đặt thử nghiệm các thuật toán học sâu tiên tiến liên quan đến các bài toán phát hiện khuôn mặt, nhận diện khuôn mặt, thiết kế mô hình học sâu, học chuyển tiếp dành cho bài toán định danh dựa trên khuôn mặt và đã kết hợp các mô hình trên để xây dựng hệ thống hỗ trợ các dịch vụ dựa trên định danh tự động. Hệ thống định danh khuôn mặt cho độ chính xác tốt, trên 99% với bộ dữ liệu huấn luyện và độ chính xác tăng đáng kể so với sử dụng các mô hình đơn lẻ trong khi giảm thời gian huấn luyện với các mô hình truyền thống.

TÀI LIỆU THAM KHẢO

[1]. Nguyễn Mạnh Cường, Nguyễn Lương Bằng, Phạm Ngọc Huân, Phí Trung Hiếu, 2022. *Mô hình học chuyển tiếp cho các dịch vụ dựa trên định danh*. Tạp chí Khoa học và Công nghệ, trường Đại học Công nghiệp Hà Nội, tập 58, số 2, tháng 4.

[2]. Y. Jason, et al., 2014. *How transferable are features in deep neural networks*. Advances in Neural Information Processing Systems (NIPS).

[3]. S. J. Pan, Q. Yang, 2010. *A Survey on Transfer Learning*. IEEE Transaction on Knowledge and Data Engineering.

[4]. A. F. Agarap, 2019. *An architecture combining convolutional neu-ral network (cnn) and support vector machine (svm) for imageclassification*. arXiv:1712.03541v2.