# GIẢI PHÁP PHÂN LOẠI CHỮ HÁN VIẾT TAY VỚI SỰ HỖ TRỢ CỦA TỐI ƯU SIÊU THAM SỐ

## HANDWRITTEN CHINESE CHARACTER CLASSIFICATION SOLUTION WITH HYPERPARAMETER OPTIMIZATION SUPPPORT

**Vũ Thị Duyên[1,*]**

DOI: http://doi.org/10.57001/huih5804.2024.203

**TÓM TẮT**

Gần đây, giải pháp phân loại chữ Hán viết tay dựa trên mạng nơ-ron tích chập đã trở nên phổ biến và đạt được một số thành công nổi bật. Tuy nhiên, để đạt được mạng nơ-ron tích chập với khả năng phân loại chính xác cao, các siêu tham số cho các mạng này cần được được tối ưu. Dựa trên kiến trúc mạng nơ-ron tích chập LeNet-5, bài báo này trình bày giải pháp phân loại chữ Hán viết tay dựa trên mạng nơ-ron tích chập với sự hỗ trợ của phương pháp tối ưu siêu tham số Hyperband. Kết quả thử nghiệm đã cho thấy mô hình mạng với sự hỗ trợ của tối ưu siêu tham số đã đạt được độ chính xác trên tập dữ liệu kiểm thử lên tới 96%, cao hơn độ chính xác dựa trên mô hình LeNet-5 và mô hình với siêu tham số ngẫu nhiên.

***Từ khóa:*** *Phân loại chữ Hán viết tay, CNN, tối ưu siêu tham số.*

**ABSTRACT**

Recently, handwritten Chinese character classification solutions based on convolutional neural networks have become popular and achieved some outstanding successes. However, to achieve convolutional neural networks with high classification accuracy, the hyperparameters for these networks need to be optimized. Based on LeNet-5 convolutional neural network architecture, this paper presents a solution for handwritten Chinese character classification based on convolutional neural network with the support of Hyperband hyperparameter optimization method. Experimental results have shown that the network model with the support of hyperparameter optimization has achieved accuracy on the test data set up to 96%, higher than the model based on the LeNet-5 model and the model with random hyperparameters.

***Keywords:*** *Chinese character classification, CNN, hyperparameter optimization.*

[1]Khoa Ngoại ngữ, Học viện Cảnh sát nhân dân
[*]Email: vtduyen80@gmail.com
Ngày nhận bài: 03/5/2024
Ngày nhận bài sửa sau phản biện: 15/6/2024
Ngày chấp nhận đăng: 25/6/2024

### 1. GIỚI THIỆU

Một trong những mạng quan trọng nhất trong lĩnh vực học sâu là mạng nơ ron tích chập (CNN: Convolutional Neural Network). Việc CNN đạt được những tiến bộ đáng chú ý trong nhiều lĩnh vực, bao gồm nhưng không giới hạn ở thị giác máy tính và xử lý ngôn ngữ tự nhiên, đã thu hút được rất nhiều sự quan tâm từ cả doanh nghiệp và giới học thuật trong những năm gần đây. Mọi người đã có thể đạt được những điều mà trước đây được cho là không thể, bao gồm nhận dạng khuôn mặt, xe không người lái, siêu thị tự phục vụ, phương pháp điều trị y tế thông minh và phân loại mẫu nhờ thị giác máy tính dựa trên CNN [1].

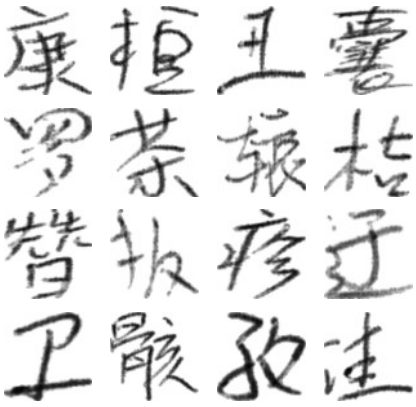
Lớp tích chập, lớp tổng hợp và lớp kết nối đầy đủ (FC: Fully Connected) là các thành phần cơ bản của mô hình CNN. Nó có thể hoàn thành hiệu quả nhiều tác vụ khác nhau bằng cách xếp chồng các lớp này một cách hợp lý trong một mạng sâu. Tuy nhiên, bất kỳ hiệu năng nào của CNN đều bị ảnh hưởng lớn bởi các siêu tham số bao gồm số lượng lớp tích chập, số lượng bộ lọc, kích thước của bộ lọc, bước trượt,… [2]. Kiểm thử thủ công các siêu tham số là một cách thông thường để xác định các siêu tham số thích hợp nhằm thu được các mô hình CNN hiệu suất cao. Tuy nhiên, do một số lý do, bao gồm nhiều siêu tham số, mô hình phức tạp, đánh giá mô hình tốn thời gian và sự tương tác siêu tham số phi tuyến tính, việc điều chỉnh thủ công không còn hữu dụng trong nhiều bài toán thực tế. Những yếu tố này đã thúc đẩy nhiều nghiên cứu hơn về các kỹ thuật tự động tối ưu siêu tham số, thường được gọi là tối ưu siêu tham số (HPO: Hyperparameter Optimization). Mục tiêu chính của HPO là tự động hóa quá trình điều chỉnh siêu tham số và cho phép người dùng triển khai thành công các mô hình CNN với siêu tham số tốt nhất cho các bài toán trong thế giới thực [2-4].

Phân loại chữ Hán viết tay được sử dụng trong nhiều ứng dụng, chẳng hạn như phân loại thư, đọc séc ngân hàng, ghi chú sách và ghi chú viết tay. Độ chính xác phân loại ký tự cao là điều cần thiết cho sự thành công của phân loại văn bản viết tay [5]. Vì vậy, bài báo tận dụng giải pháp tối ưu siêu tham số Hyperband [6, 7] để tìm kiếm siêu tham số tối ưu cho kiến trúc của mô hình Lenet-5 sao cho độ chính xác phân loại chữ Hán đạt được độ chính xác cao nhất.

### 2. GIẢI PHÁP PHÂN LOẠI CHỮ HÁN VIẾT TAY VỚI SỰ HỖ TRỢ CỦA TỐI ƯU SIÊU THAM SỐ

Tập dữ liệu chữ Hán viết tay, được sử dụng trong nghiên cứu này, được xây dựng bởi Phòng nghiên cứu Quốc gia về phân loại mẫu thuộc Viện Tự động hóa của Viện Hàn lâm Khoa học Trung Quốc (CASIA). Các mẫu dữ liệu viết tay được tạo ra bởi 1.020 người sử dụng bút Anoto trên giấy. Trong số

đó, tập dữ liệu HWDB1.1 được viết bởi 300 người và chứa khoảng 3.755 chữ Hán cấp 1 GB2312-80 và 171 chữ số và ký hiệu [8]. Để đơn giản, chúng tôi chọn sử dụng 100 lớp chữ Hán từ tập dữ liệu HWDB1.1 để kiểm chứng giải pháp. Hình 1 minh họa một số chữ Hán viết tay trong tập dữ liệu huấn luyện và thử nghiệm.



Hình 1. Chữ Hán viết tay trong tập dữ liệu

Để phân loại được chữ Hán viết tay dựa trên mô hình mạng nơ-ron tích chập có hiệu suất cao nhất, giải pháp tối ưu siêu tham số Hyperband sẽ được áp dụng vào mô hình LeNet-5. Hyperband là một biến thể của phương pháp tìm kiếm ngẫu nhiên, nhưng với một số lý thuyết khám phá-khai thác (explore-exploit) để tìm ra cách phân bổ thời gian tốt nhất cho từng tập cấu hình siêu tham số. Đây là một phương pháp giảm một nửa liên tiếp và nguyên tắc hoạt động của phương pháp được mô tả qua ví dụ sau đây:

- Lấy mẫu ngẫu nhiên 64 tập siêu tham số trong không gian tìm kiếm.

- Đánh giá sau 100 lần lặp lại sự mất xác nhận của tất cả những điều này.

- Loại bỏ những tập siêu tham số có hiệu suất thấp nhất và chỉ giữ lại một nửa.

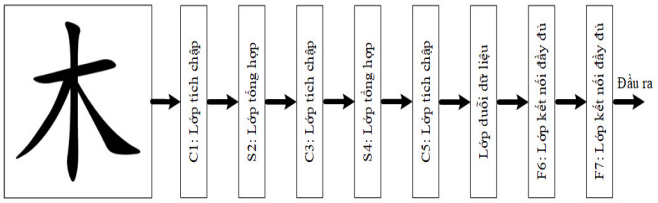
- Chạy những tập siêu tham số được giữ lại trong 100 lần lặp nữa và đánh giá.

- Tiếp tục lược bỏ đi một nửa tập siêu tham số.

- Chạy những cái tập siêu tham số tốt trong 200 lần lặp nữa và đánh giá.

- Lặp lại cho đến khi chỉ còn một mô hình duy nhất.

Mô hình mạng nơ-ron tích chập nổi tiếng LeNet-5 được đề xuất bởi LeCun và cộng sự [9] đã được ứng dụng trong hàng loạt các bài toán phân loại [10] như chữ số viết tay trong tập dữ liệu MNIST và đối tượng trong tập dữ liệu CIFAR. Mô hình được thiết kế dựa trên kiến trúc của LeNet-5 để thực hiện phân loại chữ Hán viết tay và được minh họa trong hình 2. Trong cấu trúc này, đầu vào là ảnh chữ viết tay có kích thước 64x64, và đầu ra là xác suất dự đoán về ký tự tương ứng trong tập dữ liệu. Mô hình bao gồm 3 lớp tích chập, 2 lớp tổng hợp, lớp chuỗi dữ liệu để chuyển đổi ma trận nhiều chiều thành một chiều và 2 lớp kết nối đầy đủ. Tương đương với kiến trúc LeNet-5, mô hình ngẫu nhiên và mô hình cần tối ưu hóa siêu tham số cũng có cùng số lượng lớp, dữ liệu đầu vào và kích thước lớp đầu ra. Các siêu tham số chung cho mỗi lớp, như kích thước bộ lọc, lớp tổng hợp, hàm kích hoạt và thuật toán tối ưu, được chọn trong không gian tìm kiếm, như mô tả chi tiết trong bài báo này. Điều này bao gồm việc xác định các thông số như kích thước bộ lọc, cách thức tổng hợp dữ liệu, loại hàm kích hoạt và thuật toán tối ưu để đảm bảo tính hiệu quả của mô hình. Siêu tham số tại các lớp này đối với mô hình LeNet-5, không gian tìm kiếm đối với mô hình ngẫu nhiên và mô hình dựa trên Hyperband được tổng hợp trong bảng 1. Mô hình được hỗ trợ bởi quá trình tối ưu siêu tham số (được gọi là mô hình dựa trên HPO) được so sánh với mô hình sử dụng các siêu tham số ngẫu nhiên, được gọi là mô hình ngẫu nhiên. Điều này nhằm chứng minh rằng quá trình lựa chọn siêu tham số đối với một mô hình mạng nơ-ron có tầm quan trọng lớn.




Hình 2. Mô hình mạng nơ-ron tích chập dựa trên kiến trúc LeNet-5

Bảng 1. Siêu tham số cho mạng nơ-ron tích chập

| Siêu tham số | Mô hình LeNet-5 | Mô hình ngẫu nhiên và Mô hình dựa trên Hyperband |
|---|---|---|
| Số bộ lọc tại lớp C1 | 6 | Số nguyên ngẫu nhiên từ 5 đến 150 với bước nhảy 1 |
| Số bộ lọc tại lớp C3 | 16 | |
| Số bộ lọc tại lớp C5 | 120 | |
| Số nơ-ron tại lớp F6 | 84 | |
| Kích thước bộ lọc tại C1, C3, C5 | 5 | Số nguyên ngẫu nhiên từ 3 đến 5 với bước nhảy 1 |
| Lớp tổng hợp tại S2 và S4 | Lớp tổng hợp trung bình | Lớp tổng hợp trung bình hoặc lớn nhất |
| Số nơ-ron tại lớp F7 | 100 | |
| Hàm kích hoạt tại C1, C3, C5, F6 | Hyperbolic Tangent (Tanh) | ReLu hoặc Tanh hoặc Sigmoid hoặc Elu hoặc Linear hoặc Softplus hoặc Swish |
| Hàm kích hoạt tại F7 | Softmax | |
| Hệ số học | 0,01 | Số thực ngẫu nhiên từ 1e-5 đến 1e-2 với bước nhảy 1e-5 |
| Hàm mất mát | Categorical cross-entropy | |
| Thuật toán tối ưu | SGD | Adamax hoặc Adam hoặc Nadam hoặc SGD |

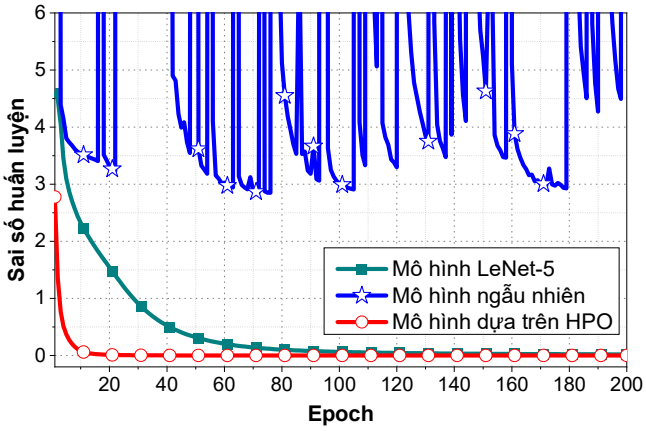
## 3. KẾT QUẢ THỬ NGHIỆM

Mô hình mạng nơ-ron tích chập được xây dựng dựa trên giao diện lập trình ứng dụng Keras [6, 7] và được thực thi trên Google Colaboratory với GPU NVIDIA Tesla T4 16GB GDDR6 PCIe 3.0. Bài báo này sử dụng phương pháp tối ưu siêu tham số Hyperband được tích hợp trong Keras Tuner để tối ưu siêu tham số trong mô hình LeNet-5. Kết quả được biểu diễn dưới dạng giá trị trung bình của 100 lần chạy độc lập. Để đảm bảo

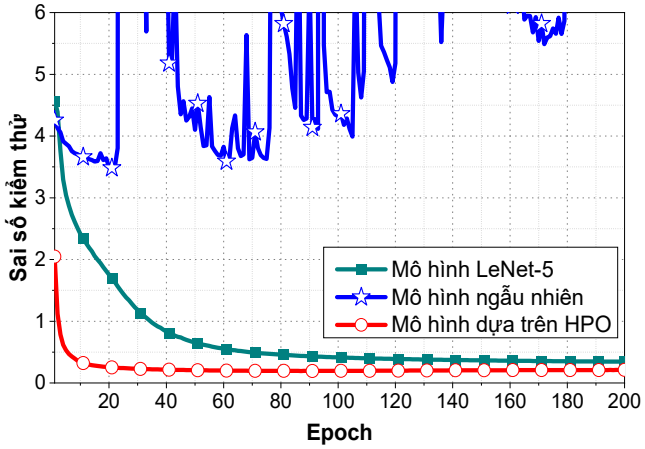
khả năng dự đoán của mô hình, quá trình huấn luyện được thực hiện với khoảng 25.000 hình ảnh với 200 epochs. Sau đó, mô hình được kiểm thử trên 5.000 hình ảnh không thuộc tập dữ liệu huấn luyện để đánh giá hiệu suất của nó.

Bảng 2. Siêu tham số tối ưu

| Siêu tham số | Mô hình LeNet-5 | Mô hình dựa trên Hyperband |
|---|---|---|
| Số bộ lọc tại lớp C1 | 6 | 23 |
| Số bộ lọc tại lớp C3 | 16 | 123 |
| Số bộ lọc tại lớp C5 | 120 | 81 |
| Số nơ-ron tại lớp F6 | 84 | 142 |
| Kích thước bộ lọc tại C1, C3, C5 | 5x5 | 5x5 |
| Lớp tổng hợp tại S2 và S4 | Lớp tổng hợp trung bình | Lớp tổng hợp lớn nhất |
| Số nơ-ron tại lớp F7 | 100 | |
| Hàm kích hoạt tại C1, C3, C5, F6 | Tanh | Tanh |
| Hàm kích hoạt tại F7 | Softmax | |
| Hệ số học | 0,01 | 0,00027 |
| Hàm mất mát | Categorical cross-entropy | |
| Thuật toán tối ưu | SGD | Nadam |

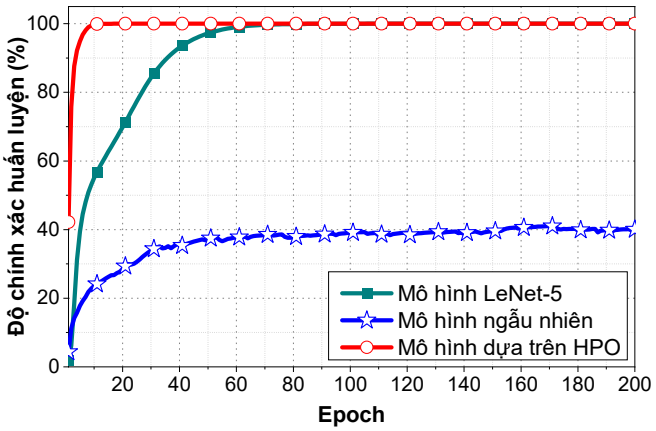



Hình 3. Sai số trên tập dữ liệu huấn luyện qua từng epoch

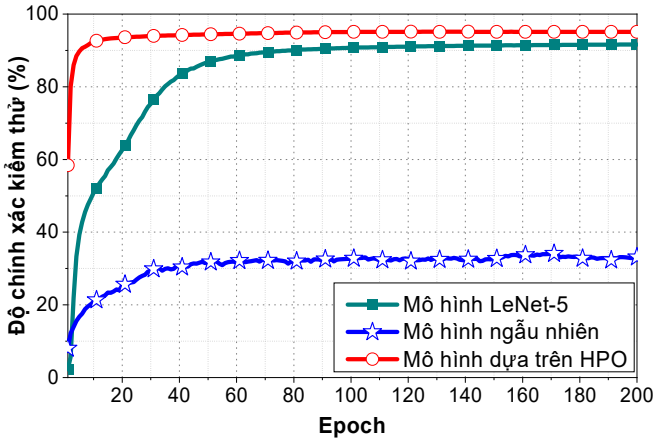



Hình 4. Sai số trên tập dữ liệu kiểm thử qua từng epoch

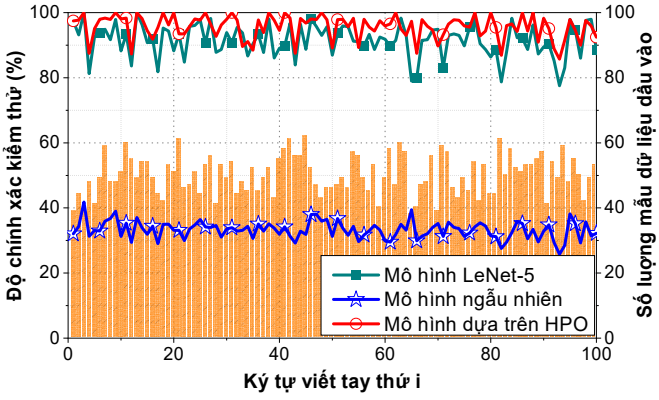
Sau khi áp dụng phương pháp tối ưu siêu tham số Hyperband, siêu tham tối ưu được tổng hợp như trong bảng 2. Qua từng epoch, sai số trên tập dữ liệu huấn luyện và kiểm thử lần lượt được minh họa trong hình 3 và 4. Có thể thấy rằng, chỉ với 20 epoch, mô hình với sự hỗ trợ của phương pháp Hyperband (được gọi là mô hình dựa trên HPO trong bài báo này) đã gần như hội tụ, trong khi mô hình LeNet-5 cần khoảng 80 epoch để hội tụ. Bên cạnh đó, giá trị sai số của mô hình dựa trên HPO đã đạt giá trị nhỏ hơn mô hình LeNet-5. Trong khi đó, mô hình sử dụng siêu tham số ngẫu nhiên đã không thể hội tụ với giá trị hàm sai số liên tục dao động qua từng epcoh; điều này dẫn đến việc phân loại các chữ Hán viết tay không được chính xác.




Hình 5. Độ chính xác phân loại trên tập dữ liệu huấn luyện




Hình 6. Độ chính xác phân loại trên tập dữ liệu thử nghiệm




Hình 7. Độ chính xác phân loại trên từng ký tự

Hình 5 và 6 minh họa độ chính xác phân loại trên tập dữ liệu huấn luyện và thử nghiệm qua từng epoch. Kết quả chỉ

ra rằng, mặc dù mô hình dựa trên HPO và mô hình LeNet-5 đầu đạt độ chính xác 100% trên tập dữ liệu huấn luyện lần lượt sau 20 và 80 epoch, nhưng độ chính xác trên tập dữ liệu kiểm thử của hai mô hình này là không giống nhau. Cụ thể, chỉ với khoảng 20 epoch trên tập dữ liệu kiểm thử, mô hình dựa trên HPO đã đạt độ chính xác lên tới 94%, trong khi đó, mô hình LeNet-5 chỉ đạt khoảng 62%. Sau khi hội tụ, mô hình dựa trên HPO đạt khoảng 96%, trong khi đó mô hình LeNet-5 đạt khoảng 91%. Bên cạnh đó, mô hình ngẫu nhiên chỉ đạt độ chính xác phân loại khoảng 30% sau 200 epoch. Điều này nói lên rằng, lựa chọn siêu tham số phù hợp hay với sự hỗ trợ của tối ưu siêu tham số, mô hình mạng nơ-ron tích sẽ có thể đạt được hiệu suất cao nhất có thể. Đối với từng chữ Hán viết tay, số lượng ảnh thử nghiệm và độ chính xác phân loại đối với từng mô hình được biểu diễn trong hình 7. Khi sử dụng mô hình dựa trên HPO, hầu hết các ký tự đều đạt độ chính xác phân loại trên 90%, trong khi đó, mô hình LeNet-5 và mô hình ngẫu nhiên lần lượt chỉ đạt trên 80% và đạt trong khoảng 25% đến 40%.

## 4. KẾT LUẬN

Bài báo này đã trình bày giải pháp phân loại chữ Hán viết tay dựa trên mô hình mạng nơ-ron tích chập với sự hỗ trợ của phương pháp tối ưu siêu tham số Hyperband. Mô hình dựa trên tối ưu siêu tham số đã được so sánh với mô hình LeNet-5 và mô hình với siêu tham số ngẫu nhiên. Kết quả kiểm thử đã cho thấy rằng, mô hình dựa trên HPO đã đạt được độ chính xác trên tập dữ liệu kiểm thử lên tới 96%, cao hơn độ chính xác dựa trên mô hình LeNet-5 và mô hình với siêu tham số ngẫu nhiên. Bên cạnh đó, chỉ với khoảng 20 epoch, mô hình dựa trên HPO cũng cho thấy chúng có tốc độ hội tụ nhanh hơn hai mô hình còn lại. Trong tương lai, nghiên cứu về giải pháp tối ưu hóa siêu tham số cũng như phương pháp phân loại chữ Hán viết tay trực tuyến hoặc sử dụng mô hình phức tạp trên các tập dữ liệu lớn hơn đồng thời triển khai mô hình sau khi được huấn luyện trên các thiết bị di động như Android, iOS, hoặc trên các thiết bị nhúng như Raspberry Pi và vi điều khiển, hứa hẹn sẽ là một đề tài nghiên cứu đầy tiềm năng.

**TÀI LIỆU THAM KHẢO**

[1]. Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," *IEEE Transactions on Neural Networks and Learning Systems*, 1-21, 2021.

[2]. B. Akay, D. Karaboga, R. Akay, "A comprehensive survey on optimizing deep learning models by metaheuristics," *Artificial Intelligence Review*, 55, 2, 829-894, 2021.

[3]. N. Bacanin, T. Bezdan, E. Tuba, I. Strumberger, M. Tuba, "Optimizing Convolutional Neural Network Hyperparameters by Enhanced Swarm Intelligence Metaheuristics," Algorithms, 13, 3, MDPI AG, 67, 2020.

[4]. L. Yang, A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, 415, Elsevier BV, 295-316, 2020.

[5]. X. Y. Zhang, Y. Bengio, C. L. Liu, "Online and offline handwritten Chinese character recognition: A comprehensive study and new benchmark," *Pattern Recognition*, 61, Elsevier BV, 348-360, 2017.

[6]. François Chollet, et al., *Keras*. 2015 [Online]. Available: https://keras.io.

[7]. Lisha Li, et al., "Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization," *Journal of Machine Learning Research*, 18, 1-52, 2018.

[8]. C. L. Liu, F. Yin, D. H. Wang, Q. F. Wang, "CASIA online and offline Chinese handwriting databases," in *Proceeding of the 11th International Conference on Document Analysis and Recognition (ICDAR)*, Beijing, China, 37-41, 2011.

[9]. Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, 86, 11, 2278-2324, 1998.

[10]. Z. H. Zhang, Z. Yang, Y. Sun, Y. F. Wu, Y. D. Xing, "Lenet-5 convolution neural network with mish activation function and fixed memory step gradient descent method," in *2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing*, 2019.

**AUTHOR INFORMATION**

**Vu Thi Duyen**

Faculty of Foreign Languages, People's Police Academy, Vietnam