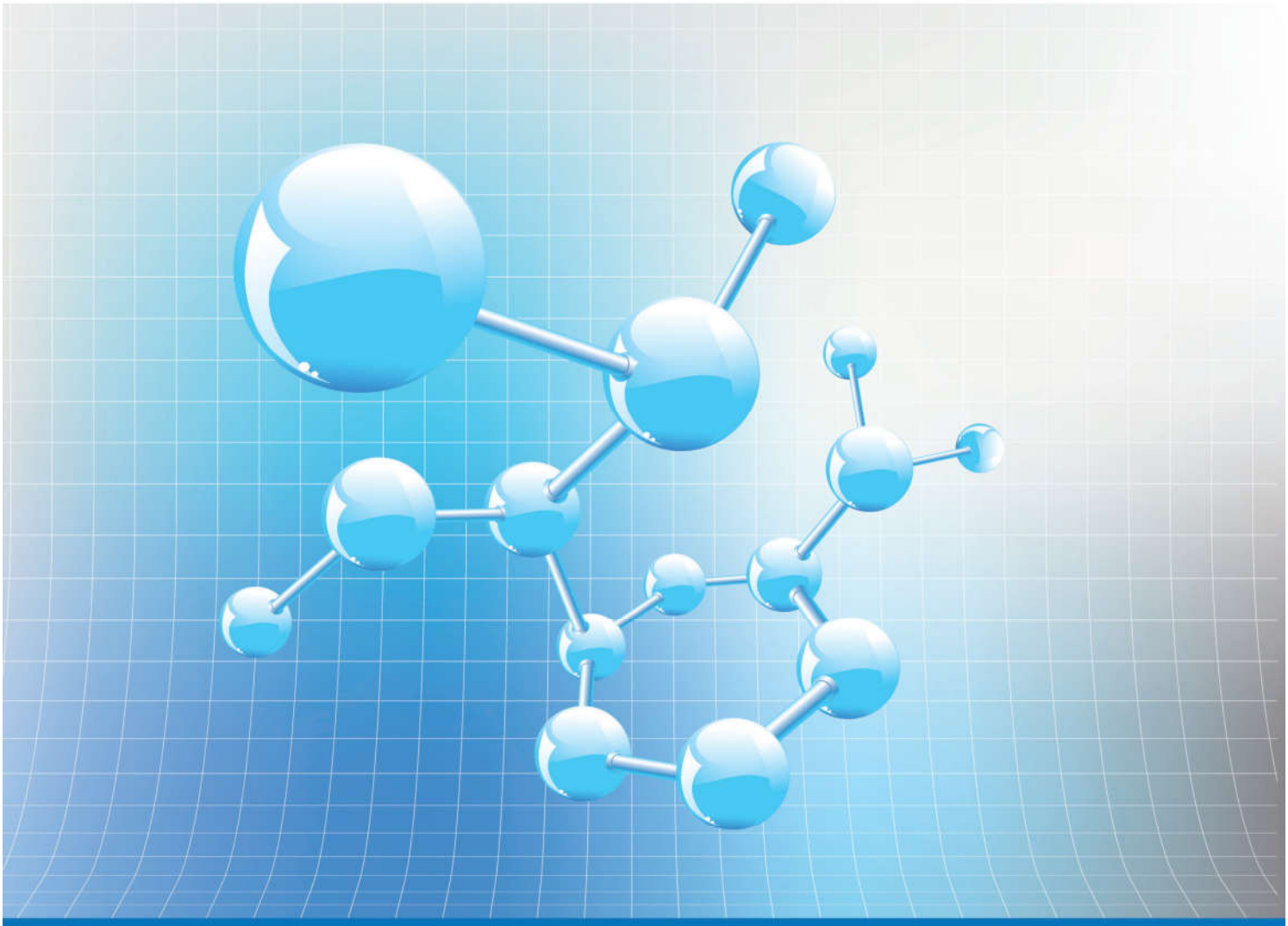
Tạp chí
NGHIÊN CỨU KHOA HỌC

ĐẠI HỌC SAO ĐỎ

SCIENTIFIC JOURNAL - SAO DO UNIVERSITY

P. ISSN 1859-4190
E. ISSN 2815-553X



Số 1 (84)
2024

P. ISSN 1859-4190
E. ISSN 2815-553X

■ **Tổng Biên tập**
TS. Đỗ Văn Đỉnh

■ **Phó Tổng biên tập**
TS. Nguyễn Thị Kim Nguyên

■ **Thư ký Tòa soạn**
PGS.TS. Ngô Hữu Mạnh

■ **Hội đồng Biên tập**

NGND.TS. Đinh Văn Nhượng - Chủ tịch Hội đồng
GS.TS. Phạm Thị Ngọc Yến
PGS.TSKH. Trần Hoài Linh
PGS.TS. Nguyễn Quốc Cường
PGS.TS. Nguyễn Văn Liễn
GS.TSKH. Thân Ngọc Hoàn
GS.TSKH. Bành Tiến Long
GS.TS. Trần Văn Địch
GS.TS. Phạm Minh Tuấn
PGS.TS. Nguyễn Doãn Ý
GS.TS. Đinh Văn Sơn
PGS.TS. Trương Thị Thủy
TS. Vũ Quang Thập
PGS.TS. Nguyễn Thị Bất
GS.TS. Đỗ Quang Kháng
TS. Bùi Văn Ngọc
PGS.TS. Ngô Sỹ Lương
PGS.TS. Khuất Văn Ninh
GS.TSKH. Phạm Hoàng Hải
PGS.TS. Đoàn Ngọc Hải
PGS.TS. Nguyễn Ngọc Hà
GS.TS. Yu Ming Zhang
TS. Nguyễn Văn Anh

■ **Ban Biên tập**

ThS. Đoàn Thị Thu Hằng - Trưởng ban
ThS. Đào Thị Vân

■ **Editor-in-Chief**
Dr. Do Van Dinh

■ **Vice Editor-in-Chief**
Dr. Nguyen Thi Kim Nguyen

■ **Office Secretary**
Assoc.Prof.Dr. Ngo Huu Manh

■ **Editorial Board**

Poeple's Teacher, Dr. Dinh Van Nhuong - Chairman
Prof.Dr. Pham Thi Ngoc Yen
Assoc.Prof.Dr.Sc. Tran Hoai Linh
Assoc.Prof.Dr. Nguyen Quoc Cuong
Assoc.Prof.Dr. Nguyen Van Lien
Prof.Dr.Sc. Than Ngoc Hoan
Prof.Dr.Sc. Banh Tien Long
Prof.Dr. Tran Van Dich
Prof.Dr. Pham Minh Tuan
Assoc.Prof.Dr. Nguyen Doan Y
Prof.Dr. Dinh Van Son
Assoc.Prof.Dr. Truong Thi Thuy
Dr. Vu Quang Thap
Assoc.Prof.Dr. Nguyen Thi Bat
Prof.Dr. Do Quang Khang
Dr. Bui Van Ngoc
Assoc.Prof.Dr. Ngo Sy Luong
Assoc.Prof.Dr. Khuat Van Ninh
Prof.Dr.Sc. Pham Hoang Hai
Assoc.Prof.Dr. Doan Ngoc Hai
Assoc.Prof.Dr. Nguyen Ngoc Ha
Prof.Dr. Yu Ming Zhang
Dr. Nguyen Van Anh

■ **Editorial**

MSc. Doan Thi Thu Hang - Head
MSc. Dao Thi Van

Địa chỉ Tòa soạn:
Trường Đại học Sao Đỏ.
Số 76, Nguyễn Thị Duệ, Thái Học 2, phường Sao Đỏ, thành phố Chí Linh, tỉnh Hải Dương.
Điện thoại: (0220) 3587213, Fax: (0220) 3882 921, Hotline: 0912 107858/0936 847980.
Website: http://tapchikhcn.saodo.edu.vn/Email: tapchikhcn@saodo.edu.vn.

Giấy phép xuất bản số: 620/GP-BTTTT ngày 17/9/2021 của Bộ Thông tin và Truyền thông.
In 2.000 bản, khổ 21 × 29,7cm, tại Công ty TNHH in Tre Xanh, cấp ngày 17/02/2011.

TẠP CHÍ
NGHIÊN CỨU KHOA HỌC
ĐẠI HỌC SAO ĐỎ

TRONG SỐ NÀY

Số 1(84) 2024

LIÊN NGÀNH ĐIỆN - ĐIỆN TỬ - TỰ ĐỘNG HÓA

Xây dựng một hệ thống nhận dạng tiếng nói và hình ảnh sử dụng các mạng học sâu trên vi điều khiển hạn chế tài nguyên và bộ nhớ	5	Vũ Văn Nghĩa Cung Thành Long
Nhận diện khuôn mặt với OPENCV và thuật toán LBPH	11	Lê Hải Thanh Đoàn Vân Chi Nguyễn Hữu Phát Nguyễn Trọng Các
Truyền thông không dây giữa phương tiện giao thông trên sóng Milimet	18	Vũ Bảo Tạo Nguyễn Thị Quyên Nguyễn Thị Phương Oanh
Xây dựng hệ thống thu thập và cảnh báo sự cố của động cơ trên tàu thủy dựa trên tín hiệu độ rung	24	Nguyễn Đức Thành Trần Hoài Linh Nguyễn Công Phương Đỗ Văn Đỉnh Phạm Văn Nam

LIÊN NGÀNH CƠ KHÍ - ĐỘNG LỰC

Nghiên cứu động lực học quay vòng của xe ô tô tải khi đi trên đường nhựa khô	31	Đào Đức Thụ Nguyễn Đình Cường Phùng Đức Hải Anh Lương Quý Hiệp
Phương pháp ghép nối các tấm thép silicon trong stator của động cơ điện: Tổng quan - Phần 1	37	Nguyễn Hoàng Minh Trí Ngô Hữu Mạnh Trịnh Văn Cường Mạc Thị Nguyên
So sánh chất lượng bề mặt của chi tiết máy khi hóa bền bằng các phương pháp biến dạng dẻo khác nhau	44	Nguyễn Văn Hinh Mạc Thị Nguyên
Nghiên cứu xây dựng hệ thống cỡ số cơ thể học sinh nam tiểu học tại thành phố Chí Linh	49	Bùi Thị Loan Nguyễn Thị Hồi

NGÀNH KINH TẾ

Giải pháp xây dựng thương hiệu cho nông sản Việt Nam	55	Nguyễn Thị Thủy

TẠP CHÍ
NGHIÊN CỨU KHOA HỌC
ĐẠI HỌC SAO ĐỎ

TRONG SỐ NÀY

Số 1(84) 2024

NGÀNH KINH TẾ

Thực trạng kế toán quản trị chi phí tại các doanh nghiệp sản xuất gạch Tuynel trên địa bàn tỉnh Hải Dương	61	Đinh Thị Kim Thiết
Phát triển kinh tế gắn với bảo vệ môi trường trên địa bàn tỉnh Hải Dương	68	Ngô Thị Luyện Nguyễn Thị Ngọc Mai
Kiểm soát thu bảo hiểm xã hội bắt buộc trên địa bàn thành phố Chí Linh, tỉnh hải Dương	75	Vũ Thị Thanh Thủy

LIÊN NGÀNH HÓA HỌC - THỰC PHẨM

Nghiên cứu khả năng hấp thụ Phenol của vật liệu chế tạo từ vỏ trấu	82	Vũ Hoàng Phương

NGÀNH GIÁO DỤC HỌC

Dạy học Hóa học ứng dụng theo định hướng phát triển năng lực vận dụng kiến thức vào thực tiễn cho sinh viên ngành Điện tại Trường Đại học Sao Đỏ	88	Phạm Thị Điệp Lê Ngọc Hòa

LIÊN NGÀNH KHOA HỌC TRÁI ĐẤT - MỎ

Đánh giá ảnh hưởng của dịch Covid-19 đến sự kiện du lịch ở Việt Nam	95	Nguyễn Thị Sao Tăng Thị Hồng Minh

LIÊN NGÀNH TRIẾT HỌC - XÃ HỘI HỌC - CHÍNH TRỊ HỌC

Nâng cao hiệu quả giáo dục lý tưởng cách mạng, đạo đức, lối sống và khát vọng cống hiến cho sinh viên Trường Đại học Sao Đỏ	101	Phạm Xuân Đức
Tư tưởng Hồ Chí Minh về công nghiệp hóa và sự vận dụng của Đảng Cộng sản Việt Nam hiện nay	107	Trần Thị Hồng Nhung
"Đề cương văn hóa Việt Nam" nội dung và ý nghĩa	113	Phạm Văn Dự
Quan hệ biện chứng giữa tính cách mạng và tính khoa học trong công tác tư tưởng của Đảng ta hiện nay	117	Trần Thị Hồng Nhung
Bảo vệ nền tảng tư tưởng về đại đoàn kết dân tộc, đấu tranh chống lại quan điểm sai trái, thù địch của kẻ thù theo quan điểm Đại hội XIII của Đảng	123	Nguyễn Thị Hiền

SCIENTIFIC JOURNAL
SAO DO UNIVERSITY

CONTENTS

No 1(84) 2024

TITLE FOR ELECTRICITY - ELECTRONICS - AUTOMATION

Deep learning on microcontroller limited resources and memory an application to a speech and image recognition system	5	Vu Van Nghia Cung Thanh Long
Face recognition with OPENCV and LBPH algorithm	11	Le Hai Thanh Doan Van Chi Nguyen Huu Phat Nguyen Trong Cac
Vehicle to vehicle wireless communications on millimeter wave	18	Vu Bao Tao Nguyen Thi Quyen Nguyen Thi Phuong Oanh
Developing a System for Collecting and Alerting Incidents of Ship Engine Failures Based on Vibration Signals	24	Nguyen Duc Thanh Tran Hoai Linh Nguyen Cong Phuong Do Van Dinh Pham Van Nam

TITLE FOR MECHANICAL AND DRIVING POWER ENGINEERING

Research on the turning dynamics of trucks when traveling on dry asphalt roads	31	Dao Duc Thu Nguyen Dinh Cuong Phung Duc Hai Anh Luong Quy Hiep
Joining of the silicon sheets steel in stator of the electric motors: Review - Part 1	37	Nguyen Hoang Minh Tri Ngo Huu Manh Trinh Van Cuong Mac Thi Nguyen
Comparison of the quality of the surface layer of parts reiforced by various methods plastic deformation	44	Nguyen Van Hinh Mac Thi Nguyen
Study on building body size system for of primary school boys in Chi Linh city	49	Bui Thi Loan Nguyen Thi Hoi

TITLE FOR ECONOMICS

Solutions for building brand for Vietnam agriculture products	55	Nguyen Thi Thuy

SCIENTIFIC JOURNAL
SAO DO UNIVERSITY

CONTENTS

No 1(84) 2024

TITLE FOR ECONOMICS

Current status of cost management accounting at Tuynel brick producing enterprises in Hai Duong province	61	Dinh Thi Kim Thiet
Economic development associated with environmental protection in Hai Duong province	68	Ngo Thi Luyen Nguyen Thi Ngoc Mai
Control of compulsory social insurance collection in Chi Linh city, Hai Duong province	75	Vu Thi Thanh Thuy

TITLE FOR CHEMISTRY AND FOOD TECHNOLOGY

Study on Phenol adsorption capacity of materials made from rice husks	82	Vu Hoang Phuong

TITLE FOR EDUCATION

Teaching applied chemistry with the orientation of developing the ability to apply knowledge into practice for electrical students at Sao Do University	88	Pham Thi Diep Le Ngoc Hoa

TITLE FOR EARTH SCIENCE - MINING

Assessment of the impact of Covid-19 and economic domnturn to tourism in Viet Nam	95	Nguyen Thi Sao Tang Thi Hong Minh

TITLE FOR PHILOSOPHY - SOCIOLOGY - POLITICAL SCIENCE

Improve the educational effectiveness of revolutionary ideals, ethics, lifestyle and arouse the desire to devote to students of Sao Do University	101	Pham Xuan Duc
Ho Chi Minh's ideology on industrialization and its application by the Communist Party of Viet Nam today	107	Tran Thi Hong Nhung
"Vietnamese cultural outline" content and meaning	113	Pham Van Du
Dialectical relationship between revolutionary and scientific nature in the ideological work of our Party today	117	Tran Thi Hong Nhung
Protecting the ideological foundation of great national unity, fighting against the wrong and hostile views of the enemy according to the viewpoint of the 13th Party Congress	123	Nguyen Thi Hien

Xây dựng một hệ thống nhận dạng tiếng nói và hình ảnh sử dụng các mạng học sâu trên vi điều khiển hạn chế tài nguyên và bộ nhớ

Deep learning on microcontroller limited resources and memory an application to a speech and image recognition system

Vũ Văn Nghĩa, Cung Thành Long*

**Tác giả liên hệ: long.cungthanh@hust.edu.vn*

Đại học Bách khoa Hà Nội

Ngày nhận bài: 16/02/2024
Ngày nhận bài sửa sau phản biện: 29/3/2024
Ngày chấp nhận đăng: 29/3/2024

Tóm tắt

Nhận dạng tiếng nói và nhận dạng hình ảnh là các chủ đề thu hút nghiên cứu rộng rãi trong lĩnh vực trí tuệ nhân tạo. Trong đó, việc sử dụng các mạng nơron nhân tạo mang lại kết quả nhận dạng tốt với tiềm năng ứng dụng cao. Tuy nhiên, các kỹ thuật này thường yêu cầu có dữ liệu và bộ nhớ lớn, khiến việc sử dụng chúng trên các thiết bị hạn chế về tài nguyên, công suất thấp trở nên khó khăn. Trong bài báo này chúng tôi trình bày việc thực hiện hai bài toán nhận dạng tiếng nói và nhận dạng hình ảnh, sử dụng các mạng học sâu (Convolution Neural Network - CNN) trên module vi điều khiển ESP32. Giải pháp đạt độ chính xác trên 97% với nhận dạng tiếng nói và 98% với nhận dạng ảnh, sử dụng khoảng 200KB bộ nhớ. Một ứng dụng điện thoại cũng được xây dựng để nhận thông tin trực tiếp từ bộ vi điều khiển, cho phép tương tác từ xa với hệ thống nhận dạng qua Wi-Fi hoặc Bluetooth.

Từ khóa: *Trí tuệ nhân tạo; bộ vi điều khiển; học sâu; thị giác máy tính; xử lý ngôn ngữ tự nhiên; học máy nhúng; TinyML.*

Abstract

Speech recognition and image recognition are topics that attract widespread research in the field of artificial intelligence. In particular, the use of artificial neural networks brings good recognition results with high application potential. However, these techniques often require large amounts of data and memory, making their use on low-power, resource-constrained devices difficult. In this article, we present the implementation of speech recognition and image recognition problems, using deep learning networks (Convolution Neural Network - CNN) on the ESP32 microcontroller module. The solution achieves over 97% accuracy with voice recognition and 98% with image recognition, using about 200KB of memory. A phone application was also built to receive information directly from the microcontroller, allowing remote interaction with the recognition system via Wi-Fi or Bluetooth.

Keywords: *AI; microcontrollers; deep learning; computer vision; natural language processing; embedded machine learning; TinyML.*

1. ĐẶT VẤN ĐỀ

Một vài năm trước, việc triển khai các thuật toán học máy hoặc mạng nơron trên các bộ vi điều khiển ít được tính đến, chủ yếu do các hạn chế về phần cứng của chúng. Để cải thiện một số vấn đề của điện toán đám mây [1-3] liên quan đến tính toán, lưu trữ, chi phí, băng thông mạng, độ trễ phản hồi, mức tiêu thụ điện năng, quyền riêng tư và bảo mật, điện toán biên bắt đầu xuất hiện, và cùng với đó, ý tưởng chạy các thuật toán chính trên vi điều khiển đã có thể thực hiện.

Các bộ vi điều khiển (Micro-Controller Unit - MCU) thường có khả năng tính toán thấp và công suất nhỏ. Song, chỉ riêng trong năm 2021, có tới 31,2 tỷ bộ vi điều khiển mới được sản xuất [4], vượt xa số lượng của các hệ thống khác có khả năng cao hơn. Đồng thời, khả năng tính toán của các bộ vi điều khiển cũng đang ngày một mạnh hơn, hiệu suất tính toán cao hơn với mức tiêu thụ năng lượng thấp. Các MCU mới thường có kiến trúc tập lệnh được tối ưu hóa đặc biệt cho các ứng dụng học sâu.

Người phản biện: 1. PGS.TS. Trần Hoài Linh
2. TS. Phạm Văn Nam

Đã xuất hiện một số công cụ nhằm triển khai học sâu trên MCU đang nhận được nhiều sự quan tâm. Chẳng hạn, TensorFlow Lite Micro là khung phần mềm [5, 6], cho phép thực thi các mô hình mạng nơron trên các bộ vi điều khiển. Hay TinyML cho phép triển khai các mô hình học máy vào thiết bị có dung lượng bộ nhớ nhỏ, tiêu thụ công suất thấp, cho phép phân tích dữ liệu cục bộ trên thiết bị và có khả năng đáp ứng trong thời gian thực [7].

Ứng dụng học sâu trên vi điều khiển đang dần trở thành xu hướng bởi những ưu điểm về năng lượng và chi phí. Có thể điểm qua một số công bố trong lĩnh vực này. Chẳng hạn, hãng chế tạo vi điều khiển STMicroelectronics giới thiệu giải pháp phát hiện người, sử dụng thị giác máy tính trên dòng vi điều khiển STM32 của họ [8]. Hay nhóm tác giả Alessio Canepa, Edoardo Ragusa, Rodolfo Zunino, Paolo Gastaldo công bố ứng dụng phát hiện đối tượng trên khung hình video giám sát, sử dụng mạng nơron sâu trên bộ vi điều khiển STM32 [9]. Giải pháp đạt độ chính xác 97%, trong khi tiêu thụ ít hơn 400 mW. Hoặc ứng dụng nhận dạng lệnh bằng giọng nói, sử dụng học sâu trên STM32 [10], dùng điều khiển robot từ xa, của nhóm nghiên cứu Ivana Guarneri, Giuseppe Messina, Arcangelo Bruna, Davide Giacalone. Ngoài ra, còn nhiều nghiên cứu khác trong lĩnh vực học máy trên vi điều khiển cũng đã được thực hiện và công bố [11-13], với nhiều ứng dụng đa dạng, như nhận dạng đối tượng, điều khiển rotor động cơ, chăm sóc sức khỏe,… Trong nước, các nghiên cứu theo hướng đưa học máy lên vi điều khiển được công bố chưa nhiều. Một công bố hiếm hoi theo hướng này có thể kể tới là, tác giả trong [14] triển khai mạng Lenet5 trên vi điều khiển STM32 nhằm nhận dạng hình ảnh chữ số viết tay. Kết quả đạt được tương đối khả quan, với độ chính xác tổng thể lên tới 96%.

Trên các công bố trong và ngoài nước gần đây, chưa thấy ứng dụng nào kết hợp hai bài toán nhận dạng tiếng nói và nhận dạng hình ảnh, sử dụng các mạng nơron học sâu, thực hiện trên vi điều khiển.

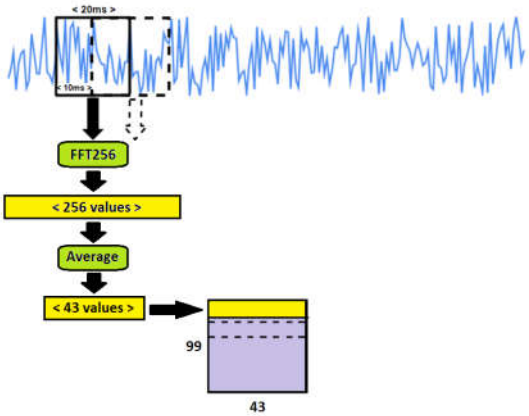
Trong nghiên cứu này, chúng tôi sẽ thực hiện bài toán nhận dạng tiếng nói kết hợp bài toán nhận dạng hình ảnh trên module ESP32, hạn chế về bộ nhớ và tiêu thụ năng lượng thấp. Trong đó, phần nhận dạng tiếng nói dùng kích hoạt thực hiện bài toán nhận dạng hình ảnh, với nhiệm vụ phân loại sản phẩm may mặc. Hai mô hình mạng nơron cho hai nhiệm vụ trên được xây dựng, huấn luyện và cài đặt trên bộ vi điều khiển ESP32.

Các phần tiếp theo của bài báo được tổ chức như sau: Phần 2 trình bày việc xây dựng các mô hình nhận dạng; phần 3 trình bày hệ thống kết hợp hai mô hình nhận dạng đã xây dựng; phần 4 trình bày một số thử nghiệm và đánh giá các mô hình nhận dạng; cuối cùng, một số kết luận và hướng phát triển được trình bày trong phần 5.

2. XÂY DỰNG CÁC MÔ HÌNH NHẬN DẠNG

2.1. Mô hình nhận dạng tiếng nói

Chúng tôi sử dụng tập dữ liệu Speech Commands [15] gồm 65000 mẫu có độ dài một giây của 30 từ ngắn, kết hợp với một số từ khóa được thêm vào (Marvin, Up, Down, On). Các từ khóa dùng làm lệnh đánh thức và điều khiển một số hoạt động của hệ thống nhận dạng hình ảnh. Chúng sẽ được nhận dạng trên nền nhiều loại tạp âm khác nhau từ tập dữ liệu đào tạo. Tất cả dữ liệu âm thanh dài 1 giây trong tập dữ liệu được chuyển về dạng ảnh phổ, là các mảng dữ liệu hai chiều với 99 hàng và 43 cột.



Hình 1. *Biến đổi âm thanh thô thành ảnh phổ*

Từ mỗi đoạn âm thanh thô dài 1 s trong miền thời gian, chúng tôi chia nhỏ thành 99 đoạn 20 ms liên tiếp, với tỉ lệ chồng chập 50% (chung 10 ms cuối cùng), (Hình 1). Sau đó, mỗi đoạn 20 ms trên sẽ được biến đổi, sử dụng thuật toán Fast Fourier Transform (FFT256) để thu được mảng 256 phần tử. Tiếp đó, các thông tin này được nhóm 6 phần tử liên tiếp và tính trung bình, để tạo ra 43 phần tử trong 1 hàng của ma trận ảnh phổ. Với 99 đoạn 20 ms, sẽ thu được 99 hàng trong một ảnh phổ đầu vào.

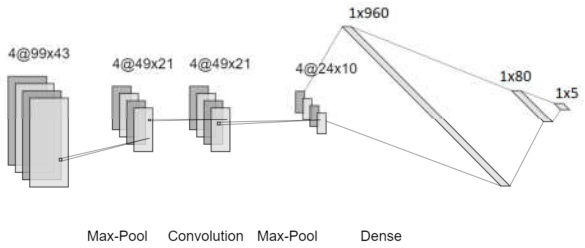
Bảng 1. *Huấn luyện các mạng CNN nhận dạng tiếng nói*

Cấu trúc mạng	Layer	Số filter	Epoch	Độ chính xác	Kích thước model
1	1 Conv + 2 FC	4	10	88%	Phù hợp
2	2 Conv + 2 FC	4	10	90%	Phù hợp
3	2 Conv + 2 FC	4	50	95%	Phù hợp
4	3 Conv + 2FC	4	50	97%	Tràn bộ nhớ

Mạng CNN được sử dụng để xây dựng mô hình nhận dạng tiếng nói. Số lượng các lớp tích chập (convolution - Conv) và kết nối đầy đủ (fully connected - FC) được điều chỉnh nhằm đạt được độ chính xác tối thiểu trên 90% theo yêu cầu. Do kích thước dữ liệu chỉ là 99×43,

đồng thời cần có cấu trúc mạng nhỏ nhất, đủ để tích hợp lên vi điều khiển, nên chúng tôi chỉ thử nghiệm các cấu trúc mạng với 1, 2 hoặc 3 lớp convolution. Các thử nghiệm cho thấy cấu trúc mạng với 2 lớp convolution và 2 lớp fully connected, là phù hợp cả về kích thước và kết quả nhận dạng. Bảng 1 trình bày một số kết quả huấn luyện và lựa chọn cấu trúc mạng.

Hình 2 và Bảng 2 trình bày cụ thể hơn về kiến trúc mạng và các tham số của mô hình được chọn.



Hình 2. *Kiến trúc mạng tích chập cho nhận dạng tiếng nói*

Bảng 2. *Các tham số mô hình nhận dạng tiếng nói*

Lớp	Số kernel	Kích thước kernel	Bước nhảy	Kích thước ảnh
Input	0	0	None	99×43
Conv2D	4	3×3	1	99×43×4
Max-pooling	4	2×2	2	49×21×4
Conv2D	4	3×3	1	49×21×4
Max-pooling	4	2×2	2	24×10×4
Flatten				1×1×960
Dropout		Dropout=0.1		1×1×960
Dense	80			1×1×80
Dropout		Dropout=0.1		1×1×80
Output	0	0	None	1×1×5

2.2. Mô hình nhận dạng hình ảnh

Trong bài báo này, chúng tôi sử dụng tập dữ liệu có sẵn Fashion_mnist [16] để triển khai hệ thống nhận dạng, phân loại 10 sản phẩm thời trang. Cách triển khai hệ thống này có thể thực hiện trên các tập dữ liệu khác, cho ứng dụng cụ thể, với các loại đối tượng bất kỳ.

Fashion_mnist gồm 70.000 sản phẩm thời trang thuộc 10 danh mục. Mỗi danh mục gồm 7.000 ảnh với kích thước 28×28 pixels, theo thang độ xám. Chúng tôi sử dụng 60.000 hình ảnh cho tập huấn luyện và 10.000 hình ảnh cho tập kiểm tra. Dữ liệu được sử dụng trực tiếp cho quá trình huấn luyện mô hình mà không cần xử lý gì thêm.

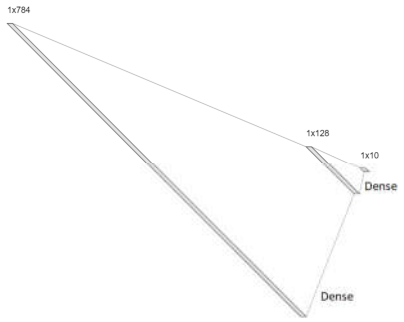
Mạng CNN được sử dụng làm mô hình nhận dạng sản phẩm. Do tín hiệu vào với kích thước chỉ là ma trận 28×28, đồng thời yêu cầu mạng có kích thước nhỏ, nên các cấu trúc mạng với 1 và 2 lớp convolution, kết hợp 2 lớp fully connected được chọn thử nghiệm đầu tiên. Tuy đảm bảo độ chính xác nhận dạng tối thiểu trên 95%, song hai cấu trúc mạng này có kích thước lớn, không phù hợp cho việc sử dụng trên module vi điều khiển ESP32.

Bảng 3. *Huấn luyện các mạng CNN nhận dạng hình ảnh*

Cấu trúc mạng	Layer	Số filter	Epoch	Độ chính xác	Kích thước model
1	1 Conv + 2 FC	20	50	98%	Tràn bộ nhớ Flash
2	1 Conv + 2 FC	10	50	97%	Tràn bộ nhớ động
3	2 FC	None	150	98%	Phù hợp

Do đó, cấu trúc mạng đơn giản hơn chỉ sử dụng 2 lớp fully connected được thử nghiệm. Cấu trúc này đủ nhỏ để chạy trên vi điều khiển và có độ chính xác đạt yêu cầu, khi tăng thời gian huấn luyện. Kết quả một số thử nghiệm được thể hiện tóm tắt trong Bảng 3.

Hình 3 và Bảng 4 thể hiện cụ thể hơn về kiến trúc mạng và các tham số của mô hình được chọn.



Hình 3. *Kiến trúc mạng tích chập cho nhận dạng hình ảnh*

Bảng 4. *Các tham số mô hình nhận dạng hình ảnh*

Lớp	Số kernel	Kích thước kernel	Kích thước ảnh
Flatten			1×1×784
Dense	128	Activation='relu'	1×1×128
Dense	10	Activation='softmax'	1×1×10

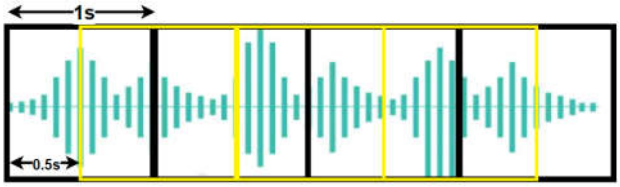
Hai mạng nơron cho nhiệm vụ nhận dạng tiếng nói và nhận dạng hình ảnh, sau quá trình huấn luyện và tối ưu hóa được chuyển sang định dạng phù hợp với vi điều khiển, sử dụng thư viện TinyMLGen. Sau đó, TensorFlow Lite Micro được dùng để nạp mô hình đã chuyển đổi lên vi điều khiển.

3. XÂY DỰNG HỆ THỐNG NHẬN DẠNG TRÊN VI ĐIỀU KHIỂN

3.1. Nhận dạng tiếng nói

Âm thanh thu được từ microphone được chuyển đổi liên tục sang dạng số và lưu trong bộ đệm. Khi đọc dữ

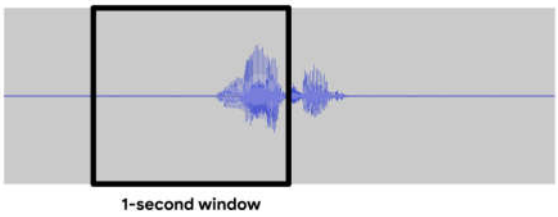
liệu, các đoạn âm thanh dài 1 giây với tỉ lệ chồng chập 50% được trích xuất như minh họa trong Hình 4.



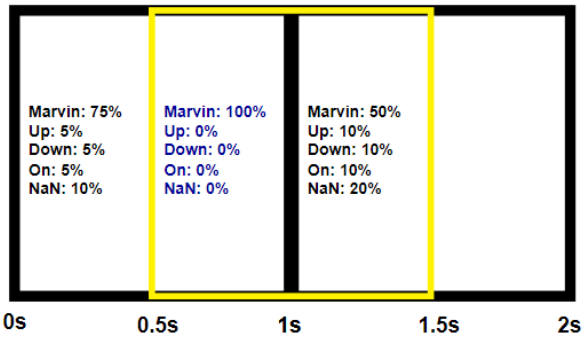
Hình 4. *Cách thức thu thập dữ liệu tiếng nói*

Các đoạn dữ liệu âm thanh được chuyển sang dạng ảnh phổ sử dụng thuật toán FFT, với kích thước 99×43 phù hợp với dữ liệu đầu vào của mạng nơron. Các ảnh phổ này được đưa tới mạng nơron đã được huấn luyện để dự đoán lệnh kích hoạt bằng giọng nói. Kết quả dự đoán là xác suất mỗi từ nhận dạng xuất hiện trong khung dữ liệu 1 giây.

Mỗi dự đoán thực hiện trên một cửa sổ dữ liệu dài một giây và thực hiện liên tiếp trên các cửa sổ trượt. Để đảm bảo có kết quả nhận dạng chính xác, tránh trường hợp cửa sổ chỉ bao trùm một phần dữ liệu của một từ nguyên vẹn (Hình 5), các cửa sổ được chồng chập 0.5 giây.



Hình 5. *Phổ của từ "byte" trong khung nhận dạng*



Hình 6. *Cách thực hiện nhận dạng một từ khóa*

Để tăng độ chính xác nhận dạng của hệ thống, chúng tôi lấy trung bình kết quả nhận dạng của mô hình CNN trên ba khung cửa sổ liên tiếp, như thể hiện trên Hình 6. Từ nào có kết quả nhận dạng trung bình lớn nhất thì kết luận nhận dạng được từ đó. Hình 6 minh họa trường hợp từ "Marvin" được nhận dạng.

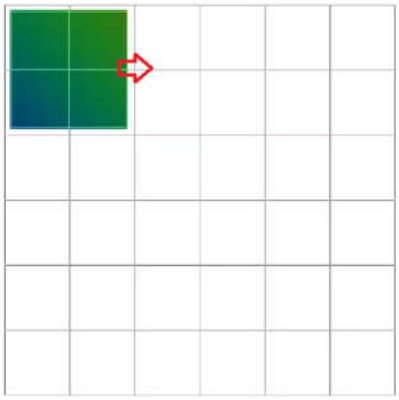
Có 4 từ khóa được sử dụng trong hệ thống nhận dạng tiếng nói và mỗi từ có một chức năng riêng. Trong đó, từ "Marvin" dùng để đánh thức hệ thống; các từ "Up", "Down" dùng để ra lệnh cài đặt thông số sau khi hệ thống được đánh thức; và từ "On" để ra lệnh cho hệ thống chuyển sang quá trình nhận dạng ảnh.

3.2. Nhận dạng hình ảnh

Chúng tôi sử dụng camera OV7670 với đầu ra là ảnh màu RGB565 và kích thước khung hình 160×120 pixels. Do đó, hình ảnh thu được cần chuẩn hóa về định dạng phù hợp với đầu vào mạng nơron đã được huấn luyện (ảnh đen trắng và kích thước ảnh 28×28 pixels). Đầu tiên, ảnh RGB565 được chuyển thành RGB888 [17] để đưa 3 giá trị đại diện cho red, green, blue về cùng một kích thước. Tiếp theo, chuyển từ ảnh màu về ảnh thang độ xám:

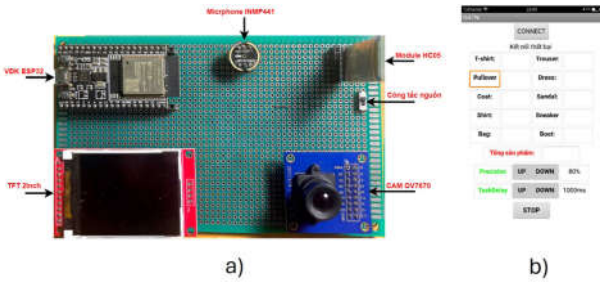
$$Grayscale = (red + blue + green)/3 \quad (1)$$

Sau đó, cắt viền ảnh từ 160×120 pixels thành 112×112 pixels và lấy trung bình giá trị mỗi điểm ảnh mới từ một khung 4×4 để giảm độ phân giải và kích thước ảnh (Hình 7). Theo đó, từ ảnh 112×112 pixels thu được ảnh 28×28 pixels theo yêu cầu.



Hình 7. *Chuẩn hóa kích thước ảnh*

Dữ liệu được đưa vào mạng nơron đã huấn luyện để nhận dạng sản phẩm. Kết quả dự đoán được hiển thị lên màn hình TFT (Thin Film Transistor) tại chỗ, đồng thời sử dụng bluetooth để truyền, hiển thị và lưu trữ kết quả lên ứng dụng điện thoại.



Hình 8. *Các linh kiện phần cứng (a) và giao diện trên ứng dụng điện thoại (b)*

Hệ thống gồm microphone INMP441 để thu dữ liệu âm thanh, camera OV7670 để thu dữ liệu hình ảnh, với các đầu ra số có thể kết nối trực tiếp với vi điều khiển và giao tiếp lần lượt qua các chuẩn I2S và I2C; màn hình TFT để hiển thị dữ liệu tại chỗ, module HC05 để kết nối không dây với app điện thoại (Hình 8a). Các thiết bị trên được ghép nối với mô đun vi điều khiển

ESP32, cho phép điều khiển toàn bộ hoạt động của hệ thống. Một ứng dụng điện thoại đơn giản, với giao diện như trong Hình 8b đã được xây dựng, để hiển thị dữ liệu nhận dạng và truyền lệnh điều khiển hệ thống qua giọng nói.

4. THỬ NGHIỆM VÀ ĐÁNH GIÁ

Để đánh giá hiệu suất và mức độ chính xác của mô hình nhận dạng, chúng tôi áp dụng phương pháp kiểm tra chéo (cross-validation). Các thông số Precision (Pre), Recall (Rec) và F1-score (F1) được sử dụng để lượng hóa kết quả.

$$Pre = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (2)$$

$$Rec = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (3)$$

$$F1 = 2 \times \frac{\mathrm{Pr}\,ecision \times \mathrm{Re}\,call}{\mathrm{Pr}\,ecision + \mathrm{Re}\,call} \quad (4)$$

Bảng 5. *Một số kết quả thử nghiệm*

Chức năng	Kịch bản		Pre	Rec	F1
Nhận dạng tiếng nói	Điều kiện tối ưu (không tạp âm)		1	1	1
	Môi trường tiếng nói	SNR = 12dB	1	0.97	0.98
		SNR = 6dB	1	0.94	0.97
		SNR = 2.5dB	0.97	0.94	0.95
	Chứa tạp âm (tiếng kêu, tiếng xe)	SNR = 12dB	1	1	1
		SNR = 6dB	1	1	1
		SNR = 2.5dB	1	0.97	0.98
Nhận dạng hình ảnh	Điều kiện tối ưu (đặt đúng vị trí)		1	1	1
	Đặt chéo ảnh	5°	1	1	1
		10°	1	0.97	0.98
		15°	0.97	0.97	0.97
	Mất 1 phần ảnh	5%	1	1	1
		10%	0.97	0.97	0.97
		15%	0.94	0.97	0.95

Mô hình nhận dạng được thử nghiệm với một số kịch bản khác nhau để đánh giá chất lượng (Bảng 5). Mỗi kịch bản được thử nghiệm 60 lần độc lập, với dữ liệu thực, thu từ microphone và camera của hệ thống, không thuộc tập dữ liệu sử dụng để huấn luyện và tối ưu hóa mô hình.

Với ứng dụng nhận dạng tiếng nói, chúng tôi thử nghiệm trong các môi trường có hai nguồn tạp âm khác nhau (tiếng nói và tiếng ồn không phải tiếng nói). Tỉ lệ tín hiệu trên nhiễu (Signal to Noise Ratio - SNR) được chọn với ba mức 2,5dB, 6dB và 12dB. Các kết quả thử nghiệm tổng hợp trong Bảng 5 cho thấy phần nhận dạng từ khóa hoạt động tốt. Khi nguồn nhiễu không phải tiếng nói, kết quả Precision đều là 1 ở cả 3 mức SNR khác nhau. Trong môi trường tạp âm là tiếng nói gần với từ nhận dạng và ở mức SNR thấp nhất, độ chính xác dù giảm nhưng precision vẫn đạt 0.97.

Với ứng dụng nhận dạng hình ảnh (phân loại sản phẩm) chúng tôi thử nghiệm hai kịch bản là ảnh bị đặt lệch và ảnh bị mất góc. Ba góc lệch 5°, 10° và 15° được sử dụng. Kết quả thử nghiệm cho thấy, với hai góc lệch nhỏ, đối tượng được nhận dạng chính xác 100%, với góc lệch lớn nhất (15°) độ chính xác giảm, song vẫn thỏa mãn yêu cầu đề ra (97%). Các thử nghiệm ảnh mất 5%, 10% và 15% diện tích góc, cho thấy khi mất 15% diện tích ảnh, độ chính xác chỉ còn 94%, dưới ngưỡng 95% đề ra (Bảng 5).

5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Trong nghiên cứu này, chúng tôi đã xây dựng và huấn luyện được hai mạng nơron học sâu, cài đặt thành công trên module vi điều khiển hạn chế tài nguyên. Đồng thời, một hệ thống ứng dụng đơn giản cũng được xây dựng, nhằm thử nghiệm khả năng nhận dạng hình ảnh thông qua ra lệnh bằng giọng nói. Các thử nghiệm bước đầu, trong điều kiện có nhiễu ở các mức độ khác nhau, cho kết quả tốt. Khả năng nhận dạng từ khóa ngắn đạt độ chính xác trên 97%. Khả năng nhận dạng hình ảnh với độ lệch và tỉ lệ mất diện tích dưới 15% đạt độ chính xác trung bình trên 98%. Một ứng dụng trên điện thoại thông minh cũng được xây dựng để thử nghiệm việc thu thập dữ liệu và điều khiển hệ thống từ xa.

Trong nghiên cứu tiếp theo, chúng tôi sẽ thử nghiệm tối ưu hóa tỉ lệ chồng chập dữ liệu để tăng độ chính xác nhận dạng từ khóa bằng tiếng nói. Đồng thời, đưa các từ tiếng Việt vào hệ thống ra lệnh bằng giọng nói. Nhóm cũng sẽ chuẩn bị cơ sở dữ liệu thực tế cho ứng dụng nhận dạng chất lượng sản phẩm trong dây chuyền chế biến nông sản, góp phần vào xu hướng phát triển nông nghiệp sạch, nông nghiệp thông minh.

TÀI LIỆU THAM KHẢO

[1]. Y. Perwej, M.A. AbuoGhaly, B. Kerim, H.A.M. Harb, *An extended Review on Internet of Things (IoT) and its Promising Applications, Communication on Applied Electronics*, Vol.7, No.26, pp. 8-22, 2019.

[2]. S. Shekhar and A. Gokhale, *Dynamic resource management across cloud-edge resources for performance-sensitive applications*, in proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, Madrid, Spain, May 2017, pp. 707-710.

[3]. Signoretti, G.; Silva, M.; Andrade, P.; Silva, I.; Sisinni, E.; Ferrari, P., *An Evolving Tinyml Compression Algorithm for IoT Environments Based on Data Eccentricity*, Sensors Vol.21, No.12, pp.41-53, 2021.

[4]. Microcontroller Unit (MCU) Shipments Worldwide from 2015 to 2023. https://www.statista.com/statistics/935382/worldwide-microcontroller-unit-shipments/

[5]. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M. et al, *Tensorflow: A system for large-scale machine learning*, in proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), Savannah, GA, USA, 2–4 November 2016, pp. 265–283.

[6]. David, R.; Duke, J.; Jain, A.; Janapa Reddi, V.; Jeffries, N.; Li, J.; Kreeger, N.; Nappier, I.; Natraj, M.; Wang, T. et al, *TensorFlow Lite Micro: Embedded Machine Learning for TinyML Systems*, in proceedings of the 4th MLSys Conference, San Jose, CA, USA, 4–7 April 2021, p. 08678.

[7]. Warden, P.; Situnayake, D., *Tinyml* ,- *Machine Learning with Tensorflow Lite on Arduino and Ultra-Low-Power Microcontrollers*, O'Reilly Media, Sebastopol, CA, USA, 2019.

[8]. STMicroelectronics, *Human presence detection using CNN-based computer vision application running on STM32*, [https://www.st.com/en/solutions-reference-designs/sl-snpsn011701v1.html#1].

[9]. Alessio Canepa, Edoardo Ragusa, Rodolfo Zunino, Paolo Gastaldo, *Detection-based video surveillance using Deep Neural Networks on STM32 microcontroller*, in proceeding of the 29th IEEE International Conference on Electronics, Circuits and Systems (ICECS), Glasgow, UK, Oct.24-26, 2022, pp. 1-4.

[10]. Ivana Guarneri, Giuseppe Messina, Arcangelo Bruna, Davide Giacalone, *A Deep Learning Short Commands Recognition for MCU in Robotics Applications*, in proceeding of the 7th International Conference on Automation, Robotics and Applications (ICARA), Prague, Czech, Feb. 04-06, 2021, pp. 43-47.

[11]. Pierre-Emmanuel Novac, *Micro AI: Illteligence Artificielle embarquee pour la Reconnaissance d'Activites Physiques sur Lunettes Intelligences*, Doctor Thesis, Cote D'Azur University, France, 2022.

[12]. Inteligent Control of quad-rotor aircrafts with a STM32 microcontroller using deep neural networks, Industrial Robot, pp.700-709, 2021.

[13]. Maha S. Diab, Esther Rodriguez-Villegas, *Embeeded machine learning using Microcontrollers in Wearable and Ambulatory Systems for health and Care Applications: A review*, IEEE access, pp.98450 – 98474, 2022.

[14]. Huỳnh Việt Thắng, *Nghiên cứu triển khai mạng học sâu Lenet5 trên vi điều khiển STM32 ứng dụng trong nhận dạng hình ảnh*, Tạp chí Khoa học và Công nghệ - Đại học Thái Nguyên, tập 226, số 11, 2021.

[15]. TensorFlow, *Speech commands,* [https://www.tensorflow.org/datasets/catalog/speech_commands].

[16]. TensorFlow, *Fashion_mnist,* [https://www.tensorflow.org/datasets/catalog/fashion_mnist].

[17]. Forlinx, *Difference between RGB565 and RGB888,* [https://www.forlinx.net/industrial-news/difference-between-rgb565-and-rgb888-423.html].

AUTHORS INFORMATION

Vu Van Nghia, Cung Thanh Long*

**Corresponding Author: long.cungthanh@hust.edu.vn*

Hanoi University of Science and Technology.

THỂ LỆ GỬI BÀI
TẠP CHÍ NGHIÊN CỨU KHOA HỌC, TRƯỜNG ĐẠI HỌC SAO ĐỎ

Tạp chí Nghiên cứu khoa học, Trường Đại học Sao Đỏ (P. ISSN 1859-4190, E. ISSN 2815-553X), thường xuyên công bố kết quả, công trình nghiên cứu khoa học và công nghệ của các nhà khoa học, cán bộ, giảng viên, nghiên cứu sinh, học viên cao học, sinh viên ở trong và ngoài nước.

1. Tạp chí xuất bản 01 số/quý bằng hai ngôn ngữ tiếng Việt và tiếng Anh. Tạp chí nhận đăng các bài báo khoa học thuộc các lĩnh vực: Điện - Điện tử - Tự động hóa; Cơ khí - Động lực; Kinh tế; Triết học - Xã hội học - Chính trị học; Các lĩnh vực khác gồm: Công nghệ thông tin; Hóa học - Công nghệ thực phẩm; Ngôn ngữ học; Toán học; Vật lý; Văn hóa - Nghệ thuật - Thể dục thể thao...
2. Bài nhận đăng là những công trình nghiên cứu khoa học chưa công bố trong bất kỳ ấn phẩm khoa học nào.
3. Tòa soạn chỉ nhận bài báo gửi online trên website http://tapchikhcn.saodo.edu.vn. Bài báo gửi về toà soạn dưới dạng file điện tử (*.doc *.docx và *.pdf); cuối bài báo, tác giả ghi rõ thông tin địa chỉ liên hệ, số điện thoại, email và cập nhật thông tin trên website. Bài báo phải được trình bày đúng định dạng, rõ ràng; Trường hợp bài báo phải chỉnh sửa theo thể lệ hoặc theo yêu cầu của Phản biện thì tác giả sẽ cập nhật trên website. Người phản biện sẽ do tòa soạn mời. Tòa soạn không gửi lại bài nếu không được đăng.
4. Các công trình thuộc đề tài nghiên cứu có Cơ quan quản lý cần kèm theo giấy phép cho công bố của cơ quan (Tên đề tài, mã số, tên chủ nhiệm đề tài, cấp quản lý,...).
5. Tên bài báo trình bày bằng hai ngôn ngữ (tiếng Việt và tiếng Anh), font Arial, cỡ chữ 14, in đậm, căn giữa.
6. Tên tác giả (không ghi học hàm, học vị), font Arial, cỡ chữ 10, in đậm, căn lề phải; cơ quan công tác của các tác giả, font Arial, cỡ chữ 9, in nghiêng, căn lề phải.
7. Chữ "Tóm tắt" in đậm, font Arial, cỡ chữ 10; Nội dung tóm tắt của bài báo không quá 10 dòng, trình bày bằng hai ngôn ngữ (tiếng Việt và tiếng Anh), font Arial, cỡ chữ 10, in thường.
8. Chữ "Từ khóa" in đậm, nghiêng, font Arial, cỡ chữ 10; Có từ 03÷05 từ khóa, font Arial, cỡ chữ 10, in nghiêng, ngăn cách nhau bởi dấu chấm phẩy, cuối cùng là dấu chấm.
9. Nội dung bài báo viết bằng tiếng Việt hoặc tiếng Anh; Nếu là bài báo viết bằng tiếng Việt: Tiêu đề tiếng Việt trước, tiếng Anh sau; Tóm tắt tiếng Việt trước, tiếng Anh sau; Từ khóa tiếng Việt trước, tiếng Anh sau; Nếu là bài báo viết bằng tiếng Anh: Tiêu đề tiếng Anh trước, tiếng Việt sau; Tóm tắt tiếng Anh trước, tiếng Việt sau; Từ khóa tiếng Anh trước, tiếng Việt sau.
10. Bài báo được đánh máy trên khổ giấy A4 (21 × 29,7cm) có độ dài không quá 8 trang, font Arial, cỡ chữ 10, giãn dòng At least 12pt, Before 3pt, After 3pt; căn lề trên 2.5cm, dưới 2.5cm, trái 3cm, phải 2cm; hình vẽ phải rõ ràng, đủ nét và được định dạng dưới dạng file ảnh (*.jpg); Phương trình, công thức phải soạn thảo bằng Mathtype hoặc Equation; Phần nội dung bài báo được chia thành 02 cột, khoảng cách cột là 1cm; Trong trường hợp hình vẽ, hình ảnh có kích thước lớn, bảng biểu có độ rộng lớn hoặc công thức, phương trình dài thì cho phép trình bày dưới dạng 01 cột.
11. Tài liệu tham khảo được sắp xếp theo thứ tự tài liệu được trích dẫn trong bài báo.

    - Nếu là sách/luận án: Tên tác giả (năm), Tên sách/luận án/luận văn, Nhà xuất bản/Trường/Viện, lần xuất bản/tái bản.

    - Nếu là bài báo/báo cáo khoa học: Tên tác giả (năm), Tên bài báo/báo cáo, Tạp chí/Hội nghị/Hội thảo, Tập/Kỷ yếu, số, trang.

    - Nếu là trang web: Phải trích dẫn đầy đủ tên website và đường link, ngày cập nhật.
12. Định dạng mẫu bài báo tham khảo tại địa chỉ http://tapchikhcn.saodo.edu.vn/news/detail/198/format_paper

    Bài báo sau khi xuất bản sẽ được công bố trên http://tapchikhcn.saodo.edu.vn.

THÔNG TIN LIÊN HỆ:

Ban Biên tập Tạp chí Nghiên cứu khoa học, Trường Đại học Sao Đỏ

Phòng 203, Tầng 2, Nhà B1, Trường Đại học Sao Đỏ.

Địa chỉ: Số 76, Nguyễn Thị Duệ, Thái Học 2, phường Sao Đỏ, thành phố Chí Linh, tỉnh Hải Dương.

Điện thoại: (0220) 3587213, Fax: (0220) 3882921, Hotline: 0912 107858/0936 847980.

Website: http://tapchikhcn.saodo.edu.vn

Email: tapchikhcn@saodo.edu.vn

BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC SAO ĐỎ

Địa chỉ:
- **Số 1:** Số 76, Nguyễn Thị Duệ, Thái Học 2, phường Sao Đỏ, thành phố Chí Linh, tỉnh Hải Dương.
- **Số 2:** Số 72, đường Nguyễn Thái Học, phường Thái Học, thành phố Chí Linh, tỉnh Hải Dương.
- **Điện thoại:** (0220) 3882 269 **Fax:** (0220) 3882 921 **Website:** http://saodo.edu.vn **Email:** info@saodo.edu.vn

P. ISSN 1859-4190
E. ISSN 2815-553X

Số 1 (84)
2024

Địa chỉ Tòa soạn:
Trường Đại học Sao Đỏ.
Số 76, Nguyễn Thị Duệ, Thái Học 2, phường Sao Đỏ, thành phố Chí Linh, tỉnh Hải Dương.
Điện thoại: (0220) 3587213, Fax: (0220) 3882 921, Hotline: 0912 107858/0936 847980.
Website: http://tapchikhcn.saodo.edu.vn/Email: tapchikhcn@saodo.edu.vn.

Giấy phép xuất bản số: 620/GP-BTTTT ngày 17/9/2021 của Bộ Thông tin và Truyền thông.
In 2.000 bản, khổ 21 × 29,7cm, tại Công ty TNHH in Tre Xanh, cấp ngày 17/02/2011.