TẠP CHÍ KHOA HỌC
TRƯỜNG ĐẠI HỌC SƯ PHẠM TP HỒ CHÍ MINH
Tập 19, Số 12 (2022): 2053-2063
Website: https://journal.hcmue.edu.vn
ISSN: 2734-9918

HO CHI MINH CITY UNIVERSITY OF EDUCATION
JOURNAL OF SCIENCE
Vol. 19, No. 12 (2022): 2053-2063
https://doi.org/10.54607/hcmue.js.19.12.3630(2022)

Bài báo nghiên cứu

ÁP DỤNG MÔ HÌNH HỌC SÂU NHẬN DẠNG MỨC ĐỘ HÀI LÒNG CỦA NGƯỜI HỌC

***Lê Hồng Thúy Vũ*[1], *Nguyễn Viết Hưng*[2], *Trịnh Huy Hoàng*[2*]**

[1]*Trường TH, THCS, THPT Trương Vĩnh Ký, Việt Nam*

[2]*Trường Đại học Sư phạm Thành phố Hồ Chí Minh, Việt Nam*

[*]*Tác giả liên hệ: Trịnh Huy Hoàng – Email: hoangth@hcmue.edu.vn*

Ngày nhận bài: 11-10-2022; ngày nhận bài sửa: 07-11-2022; ngày duyệt đăng: 22-11-2022

TÓM TẮT

Giáo viên dựa vào biểu hiện của người học sẽ biết được các hoạt động trong tiết dạy là thu hút hay nhàm chán, qua đó có những điều chỉnh phù hợp để chất lượng giảng dạy ngày càng tốt hơn. Trong giảng dạy trực tuyến, giáo viên và người học tương tác qua màn hình máy tính. Do đó, để đánh giá mức độ hài lòng của người học thì chủ yếu dựa vào cảm xúc trên khuôn mặt. Ngày nay, nhờ vào học sâu (deep learning), việc nhận dạng cảm xúc trên khuôn mặt người đã có những kết quả khả quan và giữ một vị trí quan trọng trong lĩnh vực thị giác máy tính và trí tuệ nhân tạo. Nghiên cứu đề xuất một mô hình học sâu phát hiện các cảm xúc khuôn mặt để từ đó hỗ trợ nhận dạng mức độ hứng thú của người học. Việc huấn luyện dựa trên bộ dữ liệu thu thập riêng là "HSTVK-EMO".

Từ khóa: phương pháp học sâu; phát hiện cảm xúc; mức độ hứng thú; giảng dạy trực tuyến

1 Giới thiệu

Từ 2020, tại Việt Nam, việc dạy và học đang diễn ra bình thường thì đại dịch Covid-19 xảy ra, làm thay đổi quy trình giảng dạy trực tiếp truyền thống quen thuộc, hình thức dạy và chuyển sang dạy học trực tuyến (online), giáo viên và người học chỉ giao tiếp qua màn hình. Mặc dù hiện nay, việc học trực tiếp đã thực hiện trở lại nhưng kịch bản dạy học trực tuyến vẫn được tính đến cho những thời điểm dịch bệnh có khả năng bùng phát trở lại. Cho nên nhiều trường học đã kết hợp hoạt động dạy học trực tiếp và dạy học trực tuyến để phát huy được những ưu điểm và hạn chế những nhược điểm của hai phương pháp này.

Cảm xúc đóng một vai trò quan trọng trong quá trình tiếp thu kiến thức, ra quyết định của một cá nhân. Do đó, chúng ảnh hưởng trực tiếp đến nhận thức, quá trình học tập và cách mọi người giao tiếp (Park et al., 2012). Nét mặt, cử chỉ, lời nói đều là các tín hiệu truyền đạt thông tin thể hiện thái độ tích cực và tiêu cực của một người. Trong đó thì lời nói chỉ hiệu quả 7%, cử chỉ hiệu quả nhiều hơn là 38%, riêng biểu hiện trên khuôn mặt hiệu quả đến 55%

Cite this article as: Le Hong Thuy Vu, Nguyen Viet Hung, & Trinh Huy Hoang (2022). Applying deep learning models to identify the satisfaction level of learners. *Ho Chi Minh City University of Education Journal of Science*, *19*(12), 2053-2063.

(Rinn, 1984). Khi đánh giá một tiết dạy có hay và hấp dẫn người học hay không, ngoài việc dựa vào những đánh giá của chuyên gia giáo dục hay những người trực tiếp giảng dạy thì có thể thấy thước đo chính xác nhất là ở biểu hiện qua mức độ hứng thú của người học, mà biểu hiện hứng thú xác định bởi cảm xúc của người học. Có thể nói từ những thay đổi cảm xúc của người học ta đánh giá được độ hứng thú. Dựa vào độ hứng thú của người học, người dạy có thể thay đổi phương pháp truyền tải cho phù hợp và nhà giáo dục có thể cải tiến để hoàn thiện chương trình học.

Dựa trên những trao đổi trên và xuất phát từ mục đích nâng cao hiệu quả giảng dạy học tập, nghiên cứu "Áp dụng mô hình học sâu nhận dạng mức độ hài lòng của người học" được thực hiện. Bộ dữ liệu hình ảnh của người học "HSTVK-EMO" được thu thập để làm dữ liệu huấn luyện mô hình.

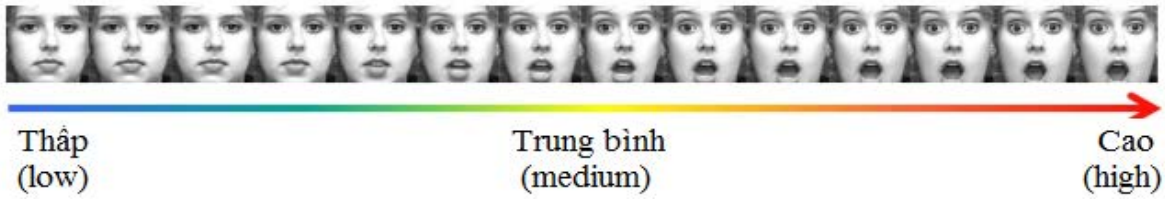
2 Cơ sở lí thuyết và một số nghiên cứu liên quan

Gương mặt người thường thể hiện sáu cảm xúc cơ bản là: *giận dữ (anger), ghê tởm (disgust), hạnh phúc (happiness), sợ hãi (fear), ngạc nhiên (surprise) và buồn bã (sadness)* (Nicolae Sfetcu, 2020; Ekman, 1971).

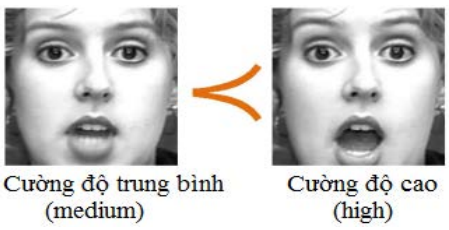
Hình 1. *Hình minh họa sáu loại biểu cảm cơ bản trên khuôn mặt*

Hầu hết các nghiên cứu nhận dạng cảm xúc trên khuôn mặt người tập trung vào việc nhận ra sáu biểu cảm phổ biến. Nhưng đôi khi điều này là không đủ cho các ứng dụng trong thế giới thực khi chúng ta cần nhận dạng theo mức độ của từng cảm xúc. Nhận dạng cảm xúc theo mức độ cũng đóng vai trò quan trọng đối với việc chọn một chiến lược phản ứng thích hợp cho sự tương tác giữa con người và máy tính (Chang et al., 2013).

Ví dụ, hình sau minh họa một chuỗi hình ảnh của cùng cảm xúc *ngạc nhiên* nhưng mức độ từ thấp đến vừa đến cao.



Hình 2. *Chuỗi hình ảnh của biểu cảm ngạc nhiên theo mức độ từ thấp đến vừa đến cao*



Hình 3. *Hai cường độ biểu cảm* "ngạc nhiên" *trên khuôn mặt*

Trong phạm vi nghiên cứu, bộ dữ liệu theo mức độ cảm xúc được thu thập, gán nhãn và kiểm thử. Đây là căn cứ để phân loại mức độ hứng thú của người học, từ đó đánh giá hiệu quả tiết học.

3 Kết quả và thảo luận

3.1. Phát biểu bài toán

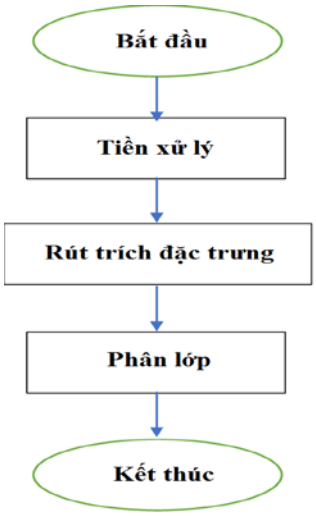
Bài toán tìm một ánh xạ từ không gian biểu diễn ảnh vào một trong L lớp cho trước chính là bài toán phân loại ảnh (Tran, Le, & Nguyen, 2018). Với giới hạn của nghiên cứu này, bài toán đánh giá độ hứng thú của người học có thể quy về bài toán phân loại ảnh mức độ cảm xúc trên khuôn mặt người. Do đó, nhóm tác giả đã tiến hành thí nghiệm, thu thập, phân tích, gán nhãn cho bộ dữ liệu ảnh biểu cảm trên gương mặt người học, chia thành 4 lớp mức độ hứng thú, đặt tên là "HSTVK-EMO". Và huấn luyện mô hình phân loại mức độ hứng thú người học được thực hiện trên bộ dữ liệu "HSTVK-EMO" này.

INPUT (đầu vào)	Bộ dữ liệu ảnh đã thực hiện tiền xử lí với: - Kích thước 48x48 pixel - Ảnh xám (1 kênh màu) - Chuẩn hóa [0,1]
OUTPUT (đầu ra)	Vectơ một chiều gồm 4 nút mạng đại diện cho 4 mức độ hứng thú Giá trị mỗi nút mạng trong phạm vi từ 0 đến 1

3.2 Giải bài toán bằng phương pháp học sâu

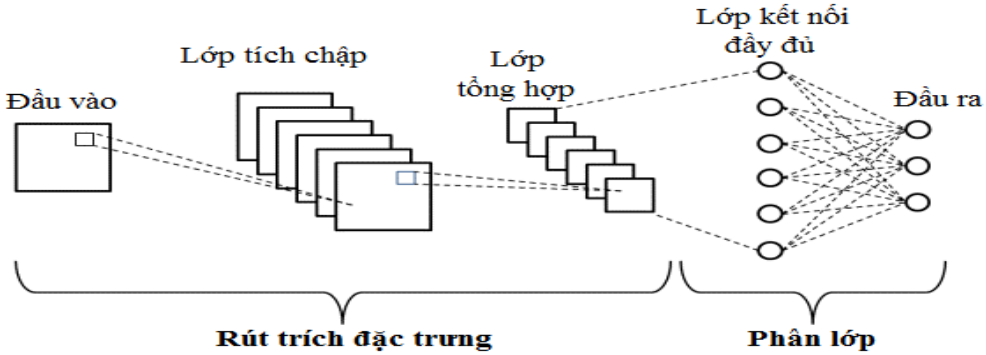
Sơ đồ sau đây làm rõ ba bước quan trọng khi tiến hành phân loại ảnh bằng mô hình học sâu mạng nơ–ron tích chập (Majeed & Srayyih, 2018).

.



Hình 4. *Các bước phân loại ảnh*

Trong đó, hai bước trích xuất đặc trưng (feature extraction) và phân loại (classification) khi thực hiện với mô hình CNN có sơ đồ như sau:



Hình 5. *Trích xuất đặc trưng và phân loại ảnh với CNN*

3.3 Mô hình CNN nhận dạng độ hứng thú của người học

3.3.1 Cơ sở dữ liệu ảnh biểu cảm khuôn mặt người học có nhãn mức độ hứng thú (HSTVK-EMO)

Bộ dữ liệu ảnh biểu cảm trên gương mặt người học được thu thập, phân tích, gán nhãn và kiểm thử, đặt tên là "HSTVK-EMO". Theo như kinh nghiệm giảng dạy nhiều năm của tác giả tại trường trung học phổ thông và kinh nghiệm từ đồng nghiệp thì các cảm xúc *giận dữ (angry), sợ hãi (fear) và ghê tởm (disgust)* rất hiếm khi xảy ra trong tiết học thực tế. Do đó, trong phạm vi của đề tài, không thu thập ba cảm xúc này.

Các mức độ cảm xúc trên ảnh thu được được phân loại như sau:

1. *Bình thường (neutral)*
2. *Hạnh phúc nhiều (happy-high)*
3. *Hạnh phúc vừa phải (happy-lowmed)*
4. *Buồn bã nhiều (sad-high)*
5. *Buồn bã vừa phải (sad-lowmed)*
6. *Ngạc nhiên nhiều (surprise-high)*
7. *Ngạc nhiên vừa phải (surprise-lowmed).*

Sau đó gom nhóm cảm xúc để tạo ra bộ dữ liệu "HSTVK-EMO" theo mức độ hứng thú của người học gồm bốn lớp như sau:

1. *Bình thường: bao gồm cảm xúc neutral.*
2. *Hứng thú vừa: bao gồm các cảm xúc tích cực là happy-lowmed và surprise-lowmed.*
3. *Rất hứng thú: bao gồm các cảm xúc rất tích cực là happy-high và surprise-high.*
4. *Không hứng thú: bao gồm các cảm xúc tiêu cực là sad-lowmed, sad-high.*

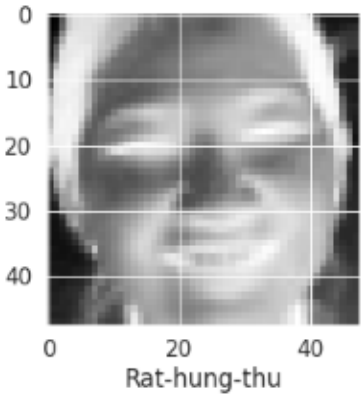
Cuối cùng, dữ liệu gán nhãn được đưa ra để khảo sát nhằm đánh giá mức độ chính xác và thêm phần khách quan. Kết quả độ chính xác trung bình của việc gán nhãn đạt 95%, trong đó lớp *không hứng thú* có độ chính xác thấp nhất (sad-lowmed đạt 90%, sad-high đạt 92%).

3.3.2 Tiền xử lí ảnh

Chuẩn hóa dữ liệu khuôn mặt gồm:

- Giảm kích thước ảnh (resized) xuống 48x48 (pixel).

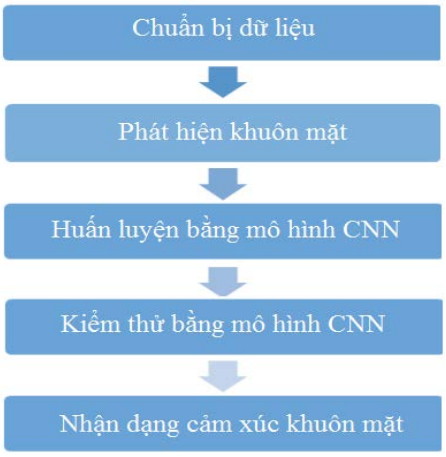
- Chuyển ảnh màu sang ảnh xám (gray scaled) để tăng hiệu suất cho mô hình CNN. Suy ra ma trận đầu vào cho mô hình CNN đề nghị là ảnh một kênh màu (48x48x1).
- Ảnh cũng được chuẩn hóa ảnh về các giá trị nằm trong khoảng [0,1] bằng tham số rescale=1/255.



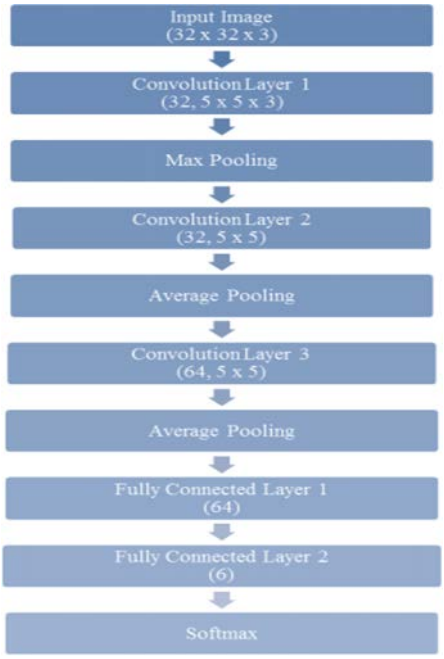
Hình 6. *Ảnh sau chuẩn hóa*

3.3.3 Mô hình nhận dạng cảm xúc trên khuôn mặt người

Có nhiều mô hình học sâu giúp phân loại cảm xúc trên khuôn mặt người. Trong đó có mô hình "*5-layer CNN*" (Dandıl & Özdemir, 2019). Để huấn luyện thì họ thu thập bộ dữ liệu cảm xúc trên khuôn mặt người từ trang tìm kiếm Google. Sau khi giả tăng cường dữ liệu thì mô hình "*5-layer CNN*" khi huấn luyện có tỉ lệ kiểm tra đúng là 62%. Sau đó các tác giả kế thừa kết quả rút trích đặc trưng trên khuôn mặt người dựa trên bộ dữ liệu lớn ImageNet của mô hình AlexNet, độ chính xác của mô hình học chuyển giao tăng lên 74%.



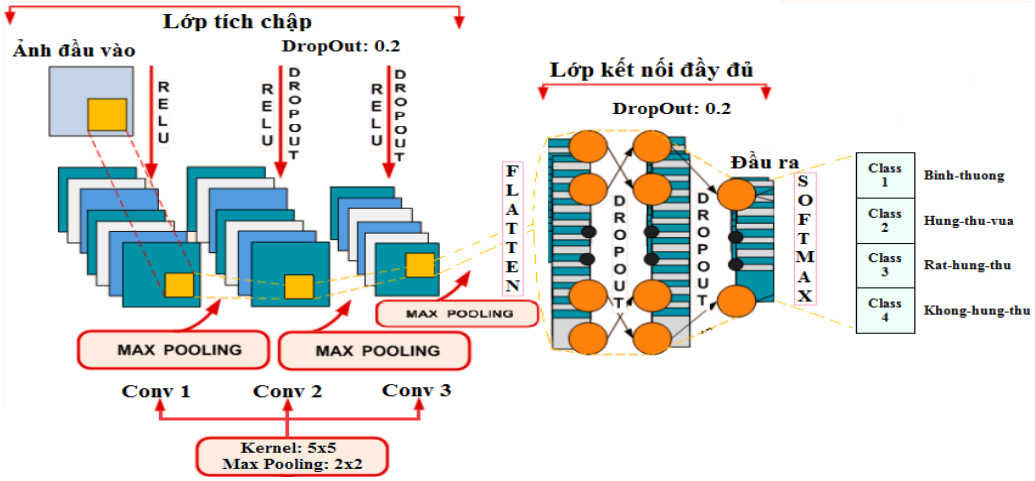
Hình 7. *Lưu đồ của hệ thống nhận dạng cảm xúc thời gian thực*



Hình 8. *Kiến trúc mô hình "5-layer CNN"*

3.3.4 Mô hình v-CNN nhận dạng độ hứng thú của người học

Mô hình v-CNN được đề xuất dựa trên kiến trúc của mô hình "5-layer CNN" được đặt tên là "5-layer v-CNN".



Hình 9. *Kiến trúc mạng* "5-layer v-CNN" *được đề xuất*

So với mô hình "5-layer CNN" thì kiến trúc mô hình "5-layer v-CNN" có một số cải tiến như sau:

- Số bộ lọc (kernel) trong mỗi lớp tích chập lần lượt là 32, 64, 128.
- Giảm kích thước ảnh với cả 3 lớp max pooling (nhằm tập trung rút trích các đặc trưng nổi trội của ảnh).
- Lớp kết nối đầy đủ thứ nhất giảm còn 128 nút mạng, lớp kết nối đầy đủ thứ 2 giảm còn 4 nút mạng tương ứng với 4 lớp mức độ hứng thú trong bộ dữ liệu "HSTVK-EMO".

3.4. Thực nghiệm

3.4.1 Dữ liệu đầu vào

```
print(augXtrain.shape)
print(augYtrain.shape)
```

```
(20060, 48, 48)
(20060,)
```

Hình 10. *Một số ảnh ngẫu nhiên trong tập huấn luyện sau khi tăng cường (phóng to, thu nhỏ, dịch chuyển, xoay...)*

3.4.2 Huấn luyện trên mô hình CNN với bộ dữ liệu "HSTVK-EMO"

Nhằm chọn ra được mô hình tốt để thí nghiệm thời gian thực. Nhóm tác giả tiến hành huấn luyện trên cả 2 mô hình "5-layer CNN" và "5-layer v-CNN".

Bảng 1. *Kết quả huấn luyện mô hình* "5-layer CNN"

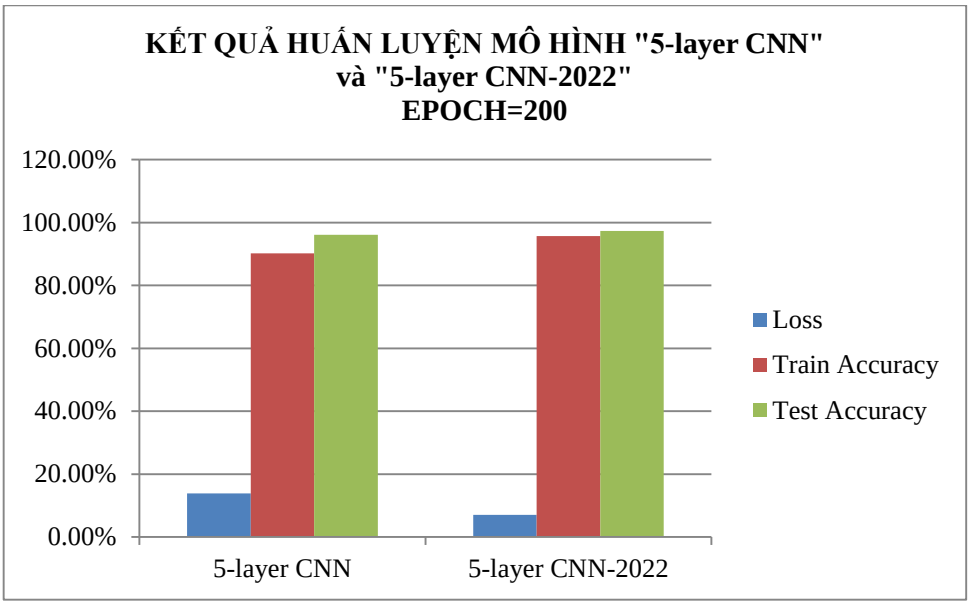
Epoch	Loss	Train Accuracy	Test Accuracy	Time
100	26.44%	83.11%	90.83%	502
125	21.94%	85.34%	92.82%	562
150	17.09%	87.51%	94.50%	743
200	13.82%	90.26%	96.17%	983

Bảng 2. *Kết quả huấn luyện mô hình "5-layer v-CNN"*

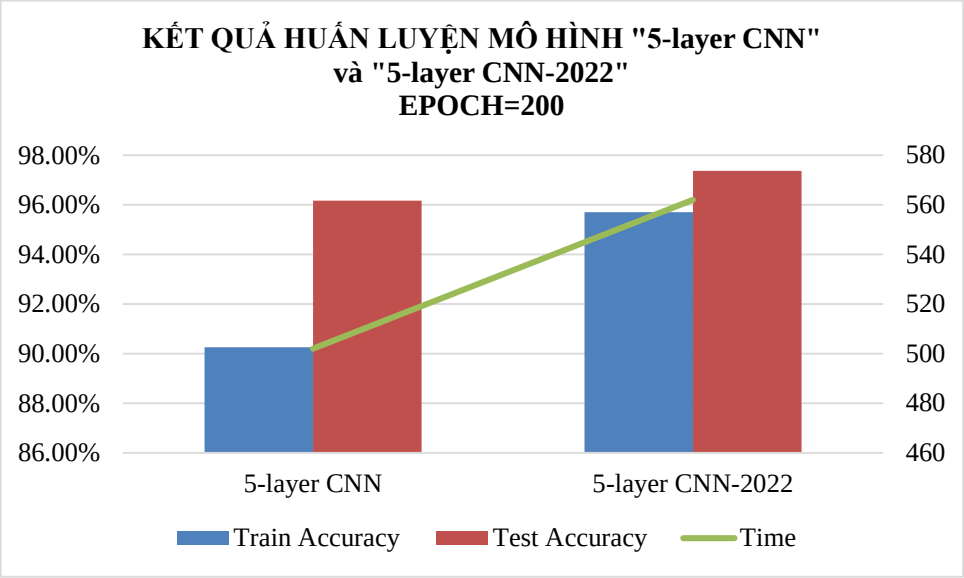
Epoch	Loss	Train Accuracy	Test Accuracy	Time
100	13.68%	90.70%	95.53%	482
125	10.76%	92.55%	96.97%	609
150	9.12%	94.20%	96.81%	745
200	7.00%	95.70%	97.37%	983

Dựa vào Bảng 1 và Bảng 2, nhận thấy khi tăng dần số lần huấn luyện (epoch) từ 100 đến 200, thì cả hai mô hình đều có độ lỗi giảm dần, độ chính xác tăng dần. Điều này là phù hợp. Không có hiện tượng học quá khớp (overfitting). Về mặt thời gian huấn luyện thì cả hai mô hình gần tương đương nhau.

Biểu đồ 1. *Mối tương quan giữa độ lỗi (%), tỉ lệ huấn luyện đúng (%) và tỉ lệ kiểm tra đúng (%)*

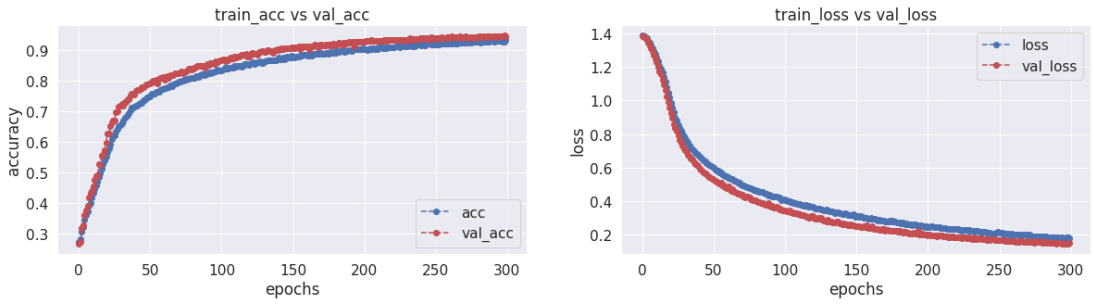


Biểu đồ 2. *Mối tương quan giữa tỉ lệ huấn luyện đúng (%), tỉ lệ kiểm tra đúng (%) và thời gian huấn luyện (giây)*

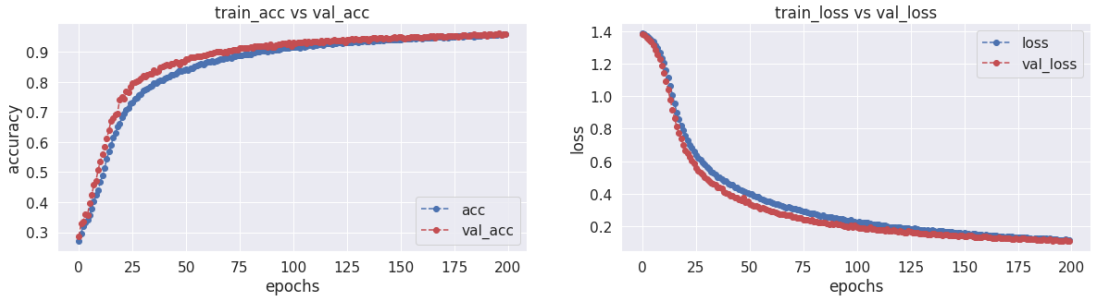


Dựa vào Biểu đồ 1 và Biểu đồ 2, có thể nhận thấy so với mô hình "5-layer CNN" thì mô hình "5-layer v-CNN" cho kết quả khả quan hơn với độ lỗi thấp hơn, độ chính xác cao hơn.

Biểu đồ 3. *Sự hội tụ của độ lỗi, độ chính xác ở mô hình* "5-layer CNN" *(a) và* "5-layer v-CNN" *(b)*



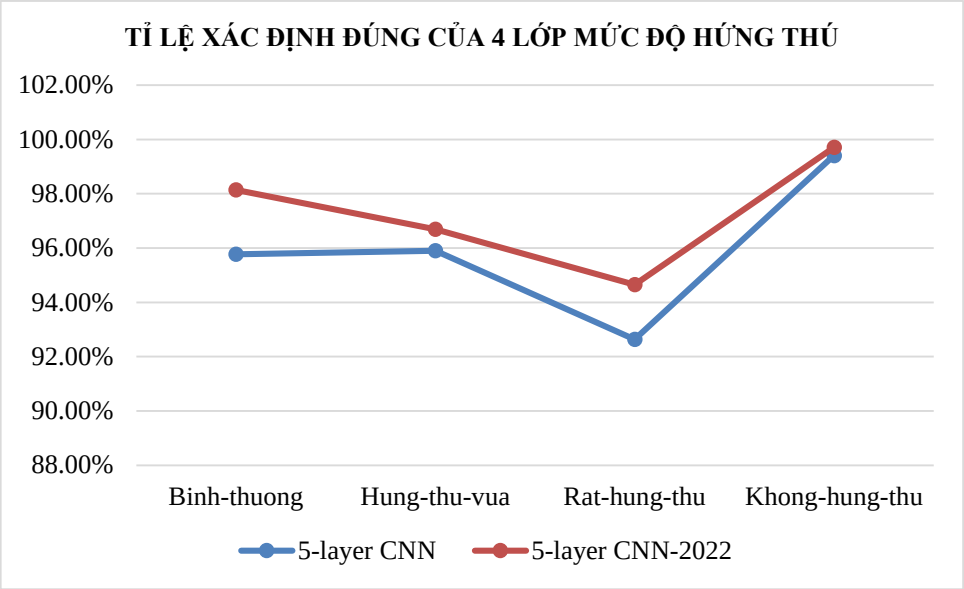
(a)



(b)

Dựa vào Biểu đồ 3, có thể nhận thấy mô hình "5-layer v-CNN" học nhanh hơn, hội tụ sớm hơn. Trong khi mô hình "5-layer CNN" sau 200 epoch cũng chưa hội tụ mà hội tụ quanh epoch thứ 300.

Biểu đồ 4. *Tỉ lệ xác định đúng 4 lớp*



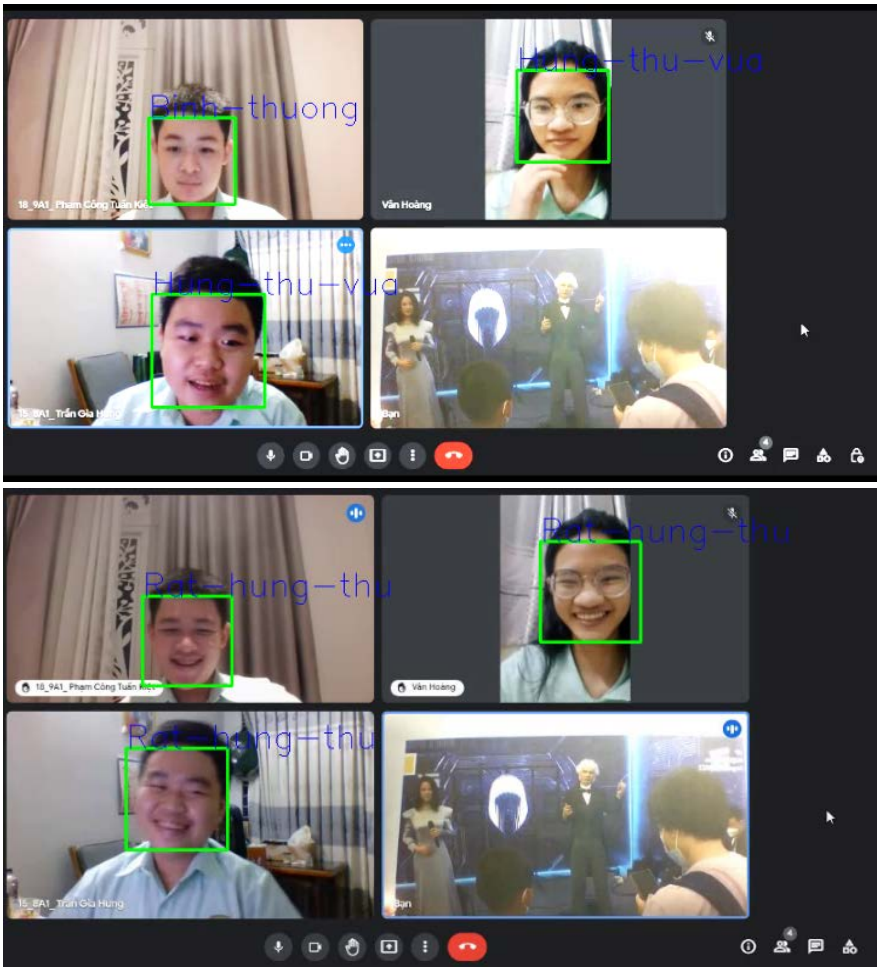
Dựa vào Biểu đồ 4, tác giả nhận thấy mô hình 32-64-128 có mức độ xác định đúng trên từng nhãn cao hơn hẳn hoặc xấp xỉ mức cao nhất của từng loại.

Tóm lại, so với mô hình "5-layer CNN" thì mô hình đề xuất "5-layer v-CNN" có những ưu điểm như thời gian thấp, độ lỗi thấp, độ chính xác cao. Cho nên tác giả chọn kết quả huấn luyện phân lớp của mô hình "5-layer v-CNN" với epoch=200 để thử nghiệm nhận dạng mức độ hứng thú thời gian thực (real-time).

4.3 Thực nghiệm thời gian thực (real-time) nhận dạng mức độ hứng thú dựa trên biểu cảm khuôn mặt người từ video lớp học online

Bước 1. Ghi lại một đoạn video của một hoạt động dạy online và đưa vào thí nghiệm thời gian thực.

Bước 2. Kiểm thử trên mô hình "5-layer v-CNN" với kiến trúc 32-64-128-2fc" đã huấn luyện sau 200 epoch và in ra nhãn dự đoán.



Hình 11. *Kết quả thí nghiệm nhận dạng mức độ hứng thú thời gian thực*

4 Kết luận và hướng phát triển

Mô hình CNN đề nghị huấn luyện trên bộ dữ liệu cảm xúc được xây dựng sẵn. Đến khi nhận dạng cảm xúc trên khuôn mặt người trong thời gian thực thì độ chính xác thường bị giảm. Đó là vì con người ngoài sáu cảm xúc cơ bản, còn có nhiều cảm xúc khác nữa xen lẫn nhau.

Ngoài nguyên nhân trên, có một số nghiên cứu chỉ ra ảnh trong thời gian thực là ảnh nhìn thấy được sẽ bị tác động bởi nhiều yếu tố đến từ môi trường như ánh sáng, khoảng cách… Họ đề xuất nhận dạng cảm xúc kết hợp ảnh nhiệt với ảnh nhìn thấy được để tăng độ chính xác, vì ảnh nhiệt không bị chi phối bởi tác nhân môi trường. Nhưng một thách thức không nhỏ đến từ môi trường giáo dục chung hiện nay, đó là nhà trường không đủ điều kiện để trang bị máy quay thu thập ảnh nhiệt.

Trong tương lai, tiếp tục thu thập ý kiến đóng góp chuyên gia để gán nhãn cảm xúc chính xác hơn nữa. Điều này giúp tăng độ chính xác cho kết quả nhận dạng của mô hình. Tiếp tục cải tiến mô hình “5-layer v-CNN” sâu hơn để tăng hiệu suất nhận dạng.

❖ ***Tuyên bố về quyền lợi:*** *Các tác giả xác nhận hoàn toàn không có xung đột về quyền lợi.*

TÀI LIỆU THAM KHẢO

Chang, K. Y., Chen, C. S., & Hung, Y. P. (2013, October). Intensity rank estimation of facial expressions based on a single image. In *2013 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 3157-3162). IEEE.

Dandıl, E., & Özdemir, R. (2019). Real-time Facial Emotion Classification Using Deep Learning. *Data Science and Applications*, *2*(1).

Ekman, P. (1971). Universals and cultural differences in facial expressions of emotion. In *Nebraska symposium on motivation*. University of Nebraska Press.

Majeed, M. A., & Srayyih, M. N. (2018). Using neural network for recognition handwritten indian numbers. *Misan Journal of Academic Studies*, *17*(33-2).

Park, B. J., Jang, E. H., Kim, S. H., Huh, C., & Sohn, J. H. (2012, April). Seven emotion recognition by means of particle swarm optimization on physiological signals: Seven emotion recognition. In *Proceedings of 2012 9th IEEE International Conference on Networking, Sensing and Control* (pp. 277-282). IEEE.

Sfetcu, N. (2020), Models of Emotion, A partial translation of: Sfetcu, Nicolae, "*Emoțiile și inteligența emoțională în organizații*". MultiMedia Publishing (ISBN 978-606-033-328-9).

Tran, S. H., & Le, H. T., & Nguyen, T. T. (2018). Phan lop anh dua tren to hop da dac trung [Image Classification Based On Multiple Feature Combination]. *Ho Chi Minh City University Of Education Journal Of Science*, *15*(12), 67-81.

Rinn, W. E. (1984). The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expressions. *Psychological bulletin*, *95*(1), 52.

APPLYING DEEP LEARNING MODELS TO IDENTIFY THE SATISFACTION LEVEL OF LEARNERS

Le Hong Thuy Vu[1], Nguyen Viet Hung[2], Trinh Huy Hoang[2*]

[1]Truong Vinh Ky High School, Middle School, High School, Vietnam

[2]Ho Chi Minh City University of Education, Vietnam

[*]Corresponding author: Trinh Huy Hoang – Email: hoangth@hcmue.edu.vn

Received: October 11, 2022; Revised: November 07, 2022; Accepted: November 21, 2022

ABSTRACT

Based on the students' expressions, teachers will know whether the lesson activities are attractive or boring, appropriately adjust their teaching. In online teaching, teachers and learners interact through computer screens. Therefore, assessing learners' satisfaction is mainly based on facial emotions. Today, thanks to deep learning, the facial recognition of emotions has had positive results and holds an important position in computer vision and artificial intelligence. The study proposes a deep learning model that detects facial emotions to help identify learners' interest levels. The training is based on the separately collected data set "HSTVK-EMO".

Keywords: deep learning methods; emotion detection; interest level; online teaching