BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC SƯ PHẠM TP. HỒ CHÍ MINH
KHOA CÔNG NGHỆ THÔNG TIN

BÁO CÁO TỔNG KẾT
ĐỀ TÀI NGHIÊN CỨU KHOA HỌC CỦA SINH VIÊN

XÂY DỰNG THUẬT TOÁN HỖ TRỢ GIÁO VIÊN ĐÁNH GIÁ PHẢN HỒI CỦA NGƯỜI HỌC

<Mã số đề tài>

Thuộc nhóm ngành khoa học: Công nghệ Thông tin

TP Hồ Chí Minh – 3/2023

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC SƯ PHẠM TP. HỒ CHÍ MINH
KHOA CÔNG NGHỆ THÔNG TIN

BÁO CÁO TỔNG KẾT
ĐỀ TÀI NGHIÊN CỨU KHOA HỌC CỦA SINH VIÊN

XÂY DỰNG THUẬT TOÁN HỖ TRỢ GIÁO VIÊN ĐÁNH GIÁ PHẢN HỒI CỦA NGƯỜI HỌC

<Mã số đề tài>

Thuộc nhóm ngành khoa học: Công nghệ thông tin

Nhóm sinh viên thực hiện:

Cao Đức Trung	47.01.104.222
Phan Lương Thùy Dương	47.01.104.074
Đào Xuân Tân	46.01.104.160
Trần Lê Chí Hải	47.01.104.084

Giảng viên hướng dẫn: TS. Nguyễn Viết Hưng

TP. Hồ Chí Minh, 03/2023

LỜI CAM ĐOAN

Nhóm nghiên cứu bao gồm Cao Đức Trung, Phan Lương Thùy Dương, Đào Xuân Tân, Trần Lê Chí Hải, sinh viên khoa Công nghệ Thông tin, trường Đại học Sư Phạm thành phố Hồ Chí Minh.

Nhóm xin cam đoan công trình nghiên cứu "Xây dựng thuật toán hỗ trợ giáo viên đánh giá phản hồi của người học" là do nhóm tìm hiểu, nghiên cứu và thực hiện dưới sự hướng dẫn của dưới sự hướng dẫn của TS. Nguyễn Viết Hưng. Công trình nghiên cứu không có sự sao chép từ các tài liệu, công trình nghiên cứu khác mà không ghi rõ nguồn trong tài liệu tham khảo.

Nhóm cam đoan báo cáo nghiên cứu này là kết quả của quá trình làm việc nghiêm túc, trung thực và tôn trọng đạo đức trong nghiên cứu khoa học. Kết quả thực nghiệm trong báo cáo nghiên cứu này là khách quan và chưa được công bố trong bất kì công trình nghiên cứu nào khác.

Nhóm xin chịu hoàn toàn trách nhiệm về kết quả thực hiện và lời cam đoan này.

Thành phố Hồ Chí Minh, ngày 25 tháng 03 năm 2023

LỜI CẢM ƠN

Trước hết nhóm nghiên cứu chúng em xin chân thành gửi lời cảm ơn sâu sắc đến Thầy của nhóm, Tiến sĩ Nguyễn Viết Hưng, người đã định hướng, chỉ bảo, giúp đỡ tận tình trong cả quá trình học tập, nghiên cứu và hoàn thiện nghiên cứu này.

Nhóm cũng xin bày tỏ lòng biết ơn đến quý thầy, cô giáo đã trực tiếp tham gia giảng dạy và truyền đạt kiến thức quý báu cho nhóm trong suốt quá trình học tại trường Đại học Sư phạm Thành phố Hồ Chí Minh.

Cuối cùng, Nhóm em muốn gửi lời cảm ơn đến gia đình và bạn bè của nhóm. Những người luôn động viên và ủng hộ để có đủ niềm tin, động lực để hoàn thành nghiên cứu khoa học.

Thành phố Hồ Chí Minh, ngày 25 tháng 03 năm 2023

MỤC LỤC

DANH MỤC CÁC CHỮ VIẾT TẮT

Từ viết tắt	Từ đầy đủ
AU	Action Units
CK	Cohn-Kanade
CK+	Extended Cohn-Kanade Dataset
CNN	Convolutional Neural Network
DL	Deep Learning
FACS	Facial Action Coding System
FC	Fully Connected
FER-2013	Facial Expression Recognition 2013 Dataset
JAFFE	The Japanese Female Facial Expression Dataset
ML	Machine Learning
MUG	Multimedia Understanding Group
RaFD	Radboud Faces Database
RAVNESS	Ryerson Audio-Visual Database of Emotional Speech and Song
ReLU	Rectified Linear Unit
ResNet	Residual Network
VGG	Visual Geometry Group

DANH MỤC CÁC BẢNG

DANH MỤC CÁC HÌNH VẼ, ĐỒ THỊ

MỞ ĐẦU

1. Lý do chọn đề tài.

Khi đại dịch Covid bùng phát đã gây ra nhiều tác động sâu sắc trong đời sống xã hội Việt Nam, một trong những lĩnh vực chịu nhiều ảnh hưởng nhất là giáo dục. Bộ Giáo dục và Đào Tạo (2021) buộc phải thay đổi hình thức giảng dạy trực tiếp thành dạy học trực tuyến để đảm bảo an toàn cho học sinh. Tuy nhiên, việc chuyển đổi hình thức dạy học từ trực tiếp sang trực tuyến trong thời gian dài sẽ gây những ảnh hưởng và thay đổi đáng kể đến hoạt động dạy học của tất cả các đối tượng liên quan. Hoạt động đánh giá phản hồi và tương tác từ người học là những trở ngại trong việc dạy học trực tuyến. Trong một khảo sát tại Trung Quốc, có đến 51,4% giáo viên trả lời là dạy học trực tuyến thiếu đi sự tương tác với học sinh (Song et al., 2020). Để dạy học trực tuyến một cách phù hợp, thầy cô đã thay đổi chương trình dạy học, phương tiện tương tác với người học cũng như thi cử trực tuyến. Nhưng để giáo viên biết được người học liệu có hào hứng hay cách giảng dạy hiệu quả như thế nào thông qua màn hình máy tính là một thách thức lớn. Giáo viên cần phải tập trung trong việc truyền đạt đủ kiến thức cho người học nên đôi khi không thể quan sát hết tất cả biểu hiện người học.

Theo nghiên cứu của Byoung-Jun Park và cộng sự (2012) của ông cho thấy sự phản hồi của người học được biểu hiện thông qua cảm xúc, điều này chứng minh rằng cảm xúc đóng một vai trò quan trọng trong quá trình tiếp thu kiến thức của học sinh, đồng thời, người dạy cũng dựa vào cảm xúc của học sinh trong giờ học để có thể đánh giá được mức độ hứng thú của học sinh đối với bài giảng. Từ đó, người dạy có thể thay đổi phương pháp giảng dạy để phù hợp và nâng cao chất lượng giảng dạy của mình.

Phản hồi của người học được coi là một yếu tố quan trọng và ảnh hưởng mạnh mẽ đến quá trình học và kết quả học tập. Tuy nhiên, tác động của nó có

thể tích cực hoặc tiêu cực tùy thuộc vào cách thức áp dụng. Phản hồi của người học thường được đề cập đến như một công cụ quan trọng trong giáo dục và thường xuất hiện trong các bài viết về giảng dạy và học tập (Hattie et al., 2017). Sự thành công của một tiết dạy phụ thuộc vào sự phản hồi của người học và kết quả mà người học đạt được sau khóa học. Lý do nhóm nghiên cứu chọn đề tài này là do việc đánh giá phản hồi của người học bởi giáo viên vẫn còn nhiều hạn chế và khó khăn. Thông thường, giáo viên thường đánh giá phản hồi của học sinh dựa trên kinh nghiệm và quan sát cá nhân, từ đó dễ bị ảnh hưởng bởi những yếu tố khác như tâm lý, quan điểm cá nhân, v.v. Với việc xây dựng thuật toán hỗ trợ giáo viên đánh giá phản hồi người học, nhóm nghiên cứu hy vọng rằng đề tài của mình sẽ hỗ trợ giáo viên trong quá trình đánh giá phản hồi từ người học thông qua mỗi tiết giảng dạy.

Từ những cơ sở trên, đề tài "**Xây dựng thuật toán hỗ trợ giáo viên đánh giá phản hồi của người học"** được thực hiện dựa trên nhận dạng cảm xúc trên khuôn mặt người học nhờ ứng dụng của thị giác máy tính. Cụ thể phân tích cảm xúc bằng ảnh có thể nhìn thấy và nét mặt của người học.

2. Mục tiêu và nhiệm vụ nghiên cứu

Mục tiêu: Xây dựng một thuật toán hỗ trợ giáo viên đánh giá phản hồi của người học thông qua việc sử dụng công nghệ mô hình học sâu. Giáo viên có thể đánh giá phản hồi của học sinh chính xác và khách quan hơn khi sử dụng mô hình. Từ đó giáo viên đưa ra phương thức giảng dạy hiệu quả cho người học.

Nhiệm vụ: Thực hiện các nhiệm vụ sau:

- Tìm hiểu tổng quan các công trình nghiên cứu về phân tích, nhận diện, dự đoán cảm xúc của người bằng ảnh thường.
- Tìm hiểu về ảnh thường và rút trích đặc trưng dựa vào mô hình học sâu

- Tìm hiểu phương pháp phân loại cảm xúc của khuôn mặt người bằng mô hình Convolutional Neural Network (CNN).
- Tìm hiểu về mô hình ResNet.
- Xây dựng mô hình hỗ trợ đánh giá sự phản hồi của người học.
- Tiến hành thực nghiệm và đánh giá kết quả đạt được.
- Định hướng phát triển trong tương lai của đề tài.

Commented [GU1]: chỗ này bổ sung Resnet nữa nha

Đóng góp của nghiên cứu

Đóng góp của nghiên cứu là xây dựng thuật toán hỗ trợ giáo viên đánh giá phản hồi người học dựa trên mô hình học sâu (Deep Learning) và xây dựng bộ dữ liệu đánh giá phản hồi của người học dựa trên bộ dữ liệu KTFE.

3. Đối tượng và phạm vi nghiên cứu

Đối tượng: Cảm xúc của người học thông qua khuôn mặt.

Phạm vi nghiên cứu:

- Nghiên cứu này quan tâm đến sự phản hồi của người học thông qua các cảm xúc khuôn mặt.
- Phân loại bốn loại cảm xúc cơ bản: ngạc nhiên, hạnh phúc, buồn bã và bình thường. Từ đó, xây dựng bộ dữ liệu riêng.
- Nhận biết, dự đoán sự thay đổi cảm xúc khuôn mặt người học dựa vào thông tin trên khuôn mặt người học (ảnh thường).
- Nghiên cứu thực hiện trên bộ dữ liệu công cộng là KTFE.

4. Phương pháp nghiên cứu:

Phương pháp nghiên cứu lý thuyết

- Tìm hiểu các công trình nghiên cứu liên quan.
- Tìm hiểu về bài toán nhận diện cảm xúc.
- Tìm hiểu các phương pháp rút trích đặc trưng và phân loại.

- Tìm hiểu về ảnh thường, ảnh nhiệt và rút trích đặc trưng quan trọng trên hai nguồn thông tin này.
- Tìm hiểu các mô hình cho bài toán đánh giá cảm xúc của người học.
- Tìm hiểu các phương pháp phân loại hình ảnh dựa trên học sâu (Deep Learning).
- Tìm hiểu phương pháp phân loại cảm xúc của khuôn mặt người bằng mô hình Convolutional Neural Network (CNN).

Phương pháp nghiên cứu thực nghiệm

- Tiến hành phân tích và cài đặt.
- So sánh và đánh giá kết quả đạt được

5. Ý nghĩa khoa học và thực tiễn

Về mặt lý thuyết

Xây dựng mô hình không chỉ phục vụ cho việc dạy học mà còn phục vụ cho thị giác máy tính, tâm lý học và nhiều hướng liên quan khác.

Về mặt thực tiễn

Việc sử dụng mô hình học sâu nói riêng và trí tuệ nhân tạo nói chung để giải quyết bài toán thực tế là cấp thiết trong quá trình "Chuyển đổi số" theo Chỉ thị số 05/CT-TTg ngày 23/2/2023 của Thủ tướng Chính phủ.

6. Nội dung văn bản

Đề tài này gồm 7 chương:

Chương mở đầu

Chương này giới thiệu tổng quan về đề tài gồm các nội dung như: lý do chọn đề tài, mục tiêu và nhiệm vụ nghiên cứu, đối tượng và phạm vi nghiên cứu, phương pháp nghiên cứu, ý nghĩa khoa học và thực tiễn cũng như cấu trúc chung của đề tài.

Chương 1. Tổng quan tình hình nghiên cứu

Chương này giới thiệu tổng quan về tình hình nghiên cứu, các thách thức về nhận diện cảm xúc bằng hình ảnh, sơ lược về dữ liệu cảm xúc khuôn mặt và một số mô hình học sâu cho phân loại cảm xúc trên ảnh gương mặt người.

Chương 2. Cơ sở lý thuyết

Chương này giới thiệu lý thuyết về cảm xúc, biểu cảm khuôn mặt, độ hứng thú, bài toán nhận diện khuôn mặt và mô hình CNN. Những kiến thức cơ bản này là tiền đề để áp dụng vào việc xây dựng thuật toán hỗ trợ đánh giá sự phản hồi của người học.

Chương 3. Xây dựng bộ dữ liệu

Chương này trình bày tình trạng của bộ dữ liệu, quá trình thu thập và gán nhãn cho bộ dữ liệu và xử lý bộ dữ liệu.

Chương 4. Thuật toán hỗ trợ giáo viên đánh giá người học

Chương này trình bày phương pháp đề xuất của đề tài.

Chương 5. Thực nghiệm và đánh giá

Chương này trình bày quá trình thực nghiệm và phân tích về những ưu điểm, nhược điểm, so sánh và đánh giá kết quả đạt được khi thực hiện chương trình.

Chương 6. Kết luận và hướng phát triển

Chương này tổng kết lại những gì đã đạt được và chưa đạt được sau khi nghiên cứu và tiến hành thực nghiệm. Từ đó nêu lên những hướng nghiên cứu và phát triển tiếp theo trong tương lai.

CHƯƠNG 1. TỔNG QUAN TÌNH HÌNH NGHIÊN CỨU

1.1. Một số công trình nghiên cứu liên quan

Trong những năm gần đây, nhận dạng cảm xúc bằng kỹ thuật học máy đã và đang thu hút được sự chú ý đáng kể trong nhiều lĩnh vực khác nhau, bao gồm cả giáo dục. Cảm xúc đóng một vai trò quan trọng trong giáo dục, nó có tác động đáng kể đến việc học tập và thành tích của người học. Khả năng nhận biết được và có động thái phù hợp với cảm xúc của người học là điều kiện cần thiết để tạo ra một môi trường học tập tích cực và hiệu quả (Roorda et al., 2011).

Do đó, nhận diện cảm xúc khuôn mặt luôn được phát triển, nhiều công trình nghiên cứu tập trung phân loại cảm xúc cơ bản. Nguyen. H và cộng sự (2022) đã nghiên cứu phương pháp kết hợp giữa ảnh thường và ảnh nhiệt để ước lượng bảy cảm xúc cơ bản. Liu và Wang đã dựa vào chuỗi ảnh nhiệt và bộ dữ liệu NVIE. Sau đó thống kê, tính toán biểu đồ về sự khác biệt nhiệt độ trên khuôn mặt, xây dựng mô hình Hiddeb Markov (HMM) phân biệt hạnh phú ghê tởm, sợ hãi với tỷ lệ được công nhận là 68,11%, 57,14% và 52,30%.

Hình 1.1: Bảy cảm xúc cơ bản của bộ dữ liệu KTFE

Tuy nhiên, cảm xúc con người không dừng lại ở những cảm xúc cơ bản đó, mà là sự kết hợp giữa những loại cảm xúc cơ bản đó. Và những nghiên cứu trên tập trung phân biệt các cảm xúc cơ bản, không đủ cho các ứng dụng xã hội khi chúng ta cần nhận dạng rõ hơn về 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.

Từ mong muốn trên, những nhà nghiên cứu bắt đầu sử dụng mô hình học sâu (Deep Learning) vào nghiên cứu đo lường cảm xúc khuôn mặt. Việc phát

hiện cảm xúc khuôn mặt bằng cách này đã giúp cho không chỉ giảng viên quan tâm đến độ hứng thú người học trong tiết học của mình mà còn giúp cho các ngành khác biết được mức độ hứng thú của khách hàng một cách nhanh chóng, từ đó đưa ra kế hoạch phù hợp cho sau này.

Cùng với sự phát triển của mô hình học sâu cụ thể là CNN, nhà nghiên cứu đã xây dựng mô hình CNN để nhận diện cảm xúc. Z. Rzayeva và cộng sự (2019) đã đề xuất một mô hình CNN để nhận diện cảm xúc với kích thước ảnh đầu vào khác nhau, các tác giả thực nghiệm với tỷ lệ chính xác là 88% và 92% trên cơ sở dữ liệu Cohn_Kanade và RAVNESS. A. Fathallah và cộng sự (2017) đã đề xuất mô hình CNN dựa trên kiến trúc mạng VGG để nhận diện sáu cảm xúc cơ bản. Các tác giả thực nghiệm trên ba cơ sở dữ liệu CK+, RaFD và MUG với tỷ lệ chính xác tương ứng là 99.33%, 93.33% và 87.65%.

1.2. Thách thức trong lĩnh vực nhận diện cảm xúc bằng hình ảnh.

Thiếu hụt dữ liệu được gắn nhãn: Một mô hình học sâu điển hình thường yêu cầu một số lượng lớn các phiên bản huấn luyện để đạt được hiệu suất tốt nhất trên các ví dụ thử nghiệm. Tuy nhiên, số lượng ảnh được gán nhãn thường bị hạn chế trong thực tế vì để có đủ ảnh được gán nhãn đòi hỏi nhiều nỗ lực thu thập của nhóm nghiên cứu. Làm thế nào để học một mô hình tốt từ một nguồn lực hạn chế như vậy trở thành một trở ngại rất lớn.

Bộ phân loại tổng quát: Bộ phân loại học sâu và học máy được đào tạo tốt trên một tập dữ liệu cụ thể thường không hoạt động tốt trên tập dữ liệu khác, đặc biệt là khi phân phối dữ liệu cực kỳ khác nhau. Việc phát triển một bộ phân loại chung có thể thực hiện trên các bộ dữ liệu khác nhau là một thách thức đặc biệt đối với các kỹ thuật nhận dạng hình ảnh.

Nhận diện cảm xúc tổng hợp: Hầu hết các nghiên cứu đang diễn ra tập trung vào việc phát hiện các cảm xúc đơn giản như niềm vui, bình thường, buồn,

hạnh phúc, tức giận, sợ hãi và ngạc nhiên. Nhưng những cảm xúc hỗn hợp chi tiết như đau đớn, vui mừng ngạc nhiên, tức giận ghê tởm rất khó phát hiện.

1.3. Sơ lược về dữ liệu cảm xúc từ gương mặt.

Hiện nay, việc nghiên cứu phát hiện và phân tích cảm xúc con người ngày càng được quan tâm do có tính ứng dụng cao trong nhiều lĩnh vực. Ngày càng có nhiều bộ dữ liệu phân tích cảm xúc được thu thập và tạo ra nhằm phục vụ cho mục đích nghiên cứu, phổ biến có thể đề cập đến bộ dữ liệu hình ảnh thường như FER2013 (Goodfellow, I. J. et al., 2013), CK+ (Ekman, P., 1993), JAFFE (Lucey, P. et al., 2010), và còn nhiều bộ dữ liệu cảm xúc khác.

FER2013 là cơ sở dữ liệu cảm xúc khuôn mặt do Kaggle cung cấp, được giới thiệu trong hội thảo ICML 2013 bởi Pierre Luc Carrier và Aaron Courville. Dữ liệu là tập các ảnh xám với kích thước 48x48 điểm ảnh. Tập dữ liệu có 35,887 ảnh với 7 cảm xúc.

Hình 1.2: Ảnh mẫu trong cơ sở dữ liệu JAFFE

CK+ : Cơ sở dữ liệu The Extended Cohn-Kanade Dataset là cơ sở dữ liệu cảm xúc khuôn mặt được xây dựng dành riêng cho những hệ thống FACS. CK+ gồm các dãy ảnh tương ứng với các thay đổi của các AU của 210 đối tượng có độ tuổi từ 18 đến 50 tuổi, nữ 69%, 81% người Mỹ gốc Âu, 13% người Mỹ gốc Phi và 6% các nhóm khác.

Hình 1.3: Ảnh mẫu trong cơ sở dữ liệu CK+

JAFFE : Japanese Female Facial Expressions là cơ sở dữ liệu cảm xúc khuôn mặt của phụ nữ Nhật Bản. JAFFE gồm 213 hình ảnh của 7 cảm xúc của 10 đối tượng. Dữ liệu là tập các ảnh xám với độ phân giải 256 × 256.

Hình 1.4: Ảnh mẫu trong cơ sở dữ liệu JAFFE

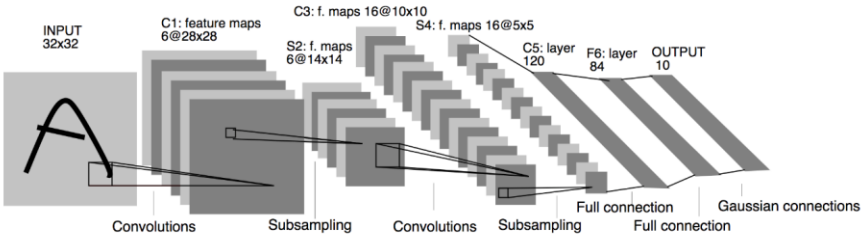
1.4. Một số mô hình học sâu cho phân loại cảm xúc trên ảnh gương mặt người

1.4.1. Lenet (1990s)

Năm 1989, LeCun đã thực hiện một nghiên cứu tại phòng thí nghiệm Bells, trong đó ông đã áp dụng phương pháp lan truyền ngược (backpropagation) trên các chữ số viết tay. Nghiên cứu này đã dẫn đến sự phát triển của mạng LeNet, một trong những mạng CNN đầu tiên được sử dụng để nhận dạng ký tự, như

đọc mã zip, chữ số, và nhiều tác vụ khác. Kiến trúc của LeNet là một trong những kiến trúc đơn giản nhất, với các lớp tích chập luôn đặt ngay sau các lớp tổng hợp (Gupta, R., 2017).

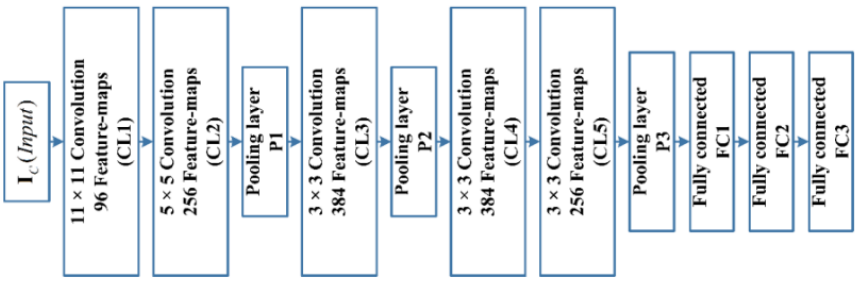
LeNet là một trong những kiến trúc CNN tiêu biểu nhất, đặc trưng là LeNet-5 gồm năm lớp với hai lớp tích chập và ba lớp kết nối đầy đủ (số 5 được đặt theo số lượng lớp tích chập và kết nối đầy đủ). Lớp tổng hợp trung bình, hiện nay đã được thay thế bằng lớp lấy mẫu phụ, có thể được huấn luyện trọng số (điều này khác với thiết kế của các CNN hiện đại). LeNet-5 có khoảng 60,000 tham số (*parameters*).



Hình 1.5: Kiến trúc mô hình Lenet-5 [18].

1.4.2. Alexnet (2012)

Công trình đầu tiên phổ biến mạng nơ-ron sử dụng CNN trong thị giác máy tính là AlexNet, được phát triển bởi Alex Krizhevsky, Ilya Sutskever và Geoff Hinton vào năm 2012. Mạng này đã cho thấy hiệu quả vượt trội hơn đáng kể so với các thuật toán học máy truyền thống khác trong cuộc thi ImageNet ILSVRC 2012. AlexNet có kiến trúc sâu hơn LeNet với tám lớp, gồm năm lớp tích chập (có áp dụng ReLU) và ba lớp kết nối đầy đủ (Gupta, R., 2017).



Hình 1.6: Kiến trúc mô hình AlexNet [19]

CHƯƠNG 2. CƠ SỞ LÝ THUYẾT

2.1. Cảm xúc, biểu cảm khuôn mặt và độ hứng thú

Khuôn mặt là một hệ thống truyền tải các tín hiệu, thông điệp phức tạp. Khuôn mặt truyền đạt ba loại tín hiệu: tín hiệu tĩnh, tín hiệu chậm và tín hiệu nhanh (Rzayeva, Z., & Alasgarov, E., 2019). *Các tín hiệu tĩnh* bao gồm các khía cạnh cố định của khuôn mặt - sự hình thành sắc tố da, hình dạng khuôn mặt, kích thước, hình dạng và vị trí của các đặc điểm khuôn mặt (lông mày, mũi, miệng, …). *Các tín hiệu chậm* bao gồm sự thay đổi bề ngoài của khuôn mặt theo thời gian như nếp nhăn, kết cấu da. *Các tín hiệu nhanh* do sự di chuyển của cơ mặt tạo thành, dẫn đến các thay đổi tạm thời trên khuôn mặt, thay đổi vị trí và hình dạng các đặc điểm. Những thay đổi này xuất hiện nhanh trong vài giây hoặc chưa đến một giây.

Ngoài ra, khuôn mặt còn có thể truyền tải các thông điệp về cảm xúc, tâm trạng, thái độ, tính cách, … và nhiều thông tin khác. Kết hợp với ngôn ngữ cơ thể, tiếng nói, khuôn mặt là một phương tiện quan trọng để truyền tải thông tin và giao tiếp giữa con người với con người.

Cảm xúc là một phản ứng tâm lý của con người trước tác động của ngoại cảnh. Cảm xúc liên quan đến những cảm giác tạm thời như sợ hãi, bất ngờ, vui vẻ, ... và được thể hiện thông qua những thay đổi về diện mạo khuôn mặt. Khi những cảm giác này xảy ra, các cơ mặt co lại dẫn đến sự thay đổi tạm thời về vị trí và hình dạng của các đặc điểm khuôn mặt. Chúng ta có thể nhìn thấy được những thay đổi này. Nếp nhăn xuất hiện và biến mất, vị trí và hình dạng của lông mày, mắt, mũi, môi, má và cằm tạm thời thay đổi. Các nghiên cứu đã chỉ ra rằng việc đánh giá cảm xúc chính xác có thể được thực hiện từ các tín hiệu nhanh trên khuôn mặt (Rzayeva, Z., & Alasgarov, E., 2019).

Các tín hiệu nhanh trên khuôn mặt thể hiện một cảm xúc là biểu hiện của cảm xúc đó trên khuôn mặt. Biểu hiện của cảm xúc trên khuôn mặt hay biểu cảm khuôn mặt là một hoặc nhiều chuyển động hoặc vị trí của các cơ (Rinn, W. E., 1984) bên dưới da của khuôn mặt. Biểu cảm có liên quan đến cảm xúc vì đó là sự biểu bộ, bộc lộ của cảm xúc. Ở con người, biểu cảm xuất phát từ cảm xúc và biểu hiện ra bên ngoài đa dạng tùy thuộc mức độ cảm nhận mang lại.

Biểu cảm trên khuôn mặt là một phương tiện giao tiếp rất quan trọng trong xã hội loài người. Đó là một hình thức phi ngôn ngữ để truyền tải thông tin và ý nghĩa. Các cơ kết nối với da và màng cơ sẽ tạo ra chuyển động và nếp gấp trên khuôn mặt, gây ra sự di chuyển của mắt, mũi, miệng, lông mày để tạo nên các biểu cảm khác nhau (Rinn, W. E., 1984)

Có thể nói dựa vào các biểu cảm khuôn mặt ta có thể dự đoán được cảm xúc của con người và có cách ứng xử phù hợp. Do đó, biểu cảm khuôn mặt đóng vai trò quan trọng trong các tình huống giao tiếp giữa con người, bao gồm cả trong những buổi học.

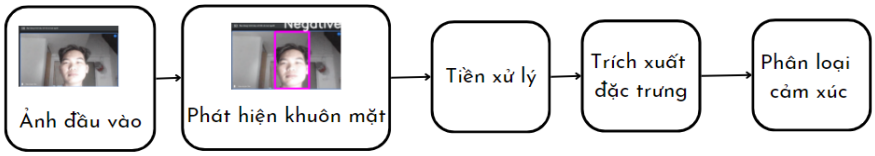
Hứng thú là thái độ con người đối với sự vật, hiện tượng nào đó, là biểu hiện của xu hướng về mặt nhận thức của cá nhân với hiện thực khách quan, biểu hiện sự ham thích của con người về sự vật, hiện tượng nào đó. Hứng thú của cá nhân được hình thành trong quá trình nhận thức và hoạt động thực tiễn.

Độ hứng thú của người học trong buổi học có liên quan đến biểu cảm khuôn mặt của họ. Biểu cảm khuôn mặt có thể phản ánh cảm xúc của người học khi tham gia vào hoạt động học tập. Nếu người học cảm thấy hứng thú và tham gia tích cực vào hoạt động hoạt động học tập thì biểu cảm khuôn mặt của họ có thể phản ánh sự hào hứng, niềm vui, sự tập trung và cảm giác hài lòng. Ngược lại, nếu người học cảm thấy không hứng thú hoặc buồn chán, biểu cảm khuôn mặt của họ có thể phản ánh sự mệt mỏi, hoặc không quan tâm.

Việc giáo viên quan sát và đánh giá biểu cảm khuôn mặt của người học có thể giúp họ một phần đánh giá phản hồi của người học và đưa ra các biện pháp giáo dục phù hợp để tạo động lực cho người học tham gia tích cực vào hoạt động học tập.

2.2. Bài toán nhận diện cảm xúc khuôn mặt

Các nhà nghiên cứu đã nghiên cứu rất nhiều về bài toán nhận diện cảm xúc khuôn mặt vì bài toán được sử dụng rộng rãi trong nhiều lĩnh vực khác nhau của đời sống như giáo dục, y tế, ... Nhìn chung, hệ thống nhận diện cảm xúc khuôn mặt sẽ được xử lý qua các bước khác nhau bao gồm: phát hiện khuôn mặt, trích xuất đặc trưng và phân loại cảm xúc.



*Hình 2.1: M*ô hình phát hiện cảm xúc sử dụng phương pháp học máy truyền thống.

Phát hiện khuôn mặt và tiền xử lý: ảnh khuôn mặt được lấy từ nguồn dữ liệu thô, để phát hiện cảm xúc cần phải thực hiện một số bước tiền xử lý để cải thiện chất lượng của ảnh, bao gồm căn chỉnh độ phân giải, chia lại tỷ lệ hình ảnh, tăng độ tương phản và áp dụng các phương pháp xử lý khác để cải thiện chất lượng của ảnh. Tất cả những bước tiền xử lý này sẽ giúp cho quá trình phát hiện cảm xúc trên ảnh khuôn mặt trở nên hiệu quả hơn.

Trích xuất đặc trưng: là một giai đoạn quan trọng trong quá trình phát hiện cảm xúc. Trong giai đoạn này, các thuật toán tính toán đặc trưng của khuôn mặt sẽ được sử dụng. Kết quả đầu ra của bước tính toán này là một vector đặc trưng được sử dụng làm đầu vào cho bước tiếp theo của quá trình phân loại cảm xúc.

Phân loại và nhận diện cảm xúc: đây là giai đoạn cuối cùng của hệ thống nhận diện cảm xúc khuôn mặt, để phân loại và nhận diện các loại cảm xúc trên khuôn mặt (hạnh phúc, sợ hãi, ngạc nhiên, bình thường).

Trong phương pháp Học máy truyền thống, một đặc điểm quan trọng là độ chính xác của mô hình phụ thuộc vào chất lượng của các đặc trưng được lựa chọn. Nếu các đặc trưng được lựa chọn phù hợp với bài toán, thì kết quả dự đoán của mô hình sẽ càng chính xác hơn. Do đó, việc lựa chọn đặc trưng là một yếu tố quan trọng trong việc xây dựng mô hình Học máy để phân loại cảm xúc trên khuôn mặt.

2.3. Mô hình mạng Nơ-ron Tích chập (Convolutional Neural Network)

Để giải quyết bài toán của bài nghiên cứu này, chúng tôi lựa chọn một phương pháp được sử dụng trong Deep Learning là Convolutional Neural Network (CNN).

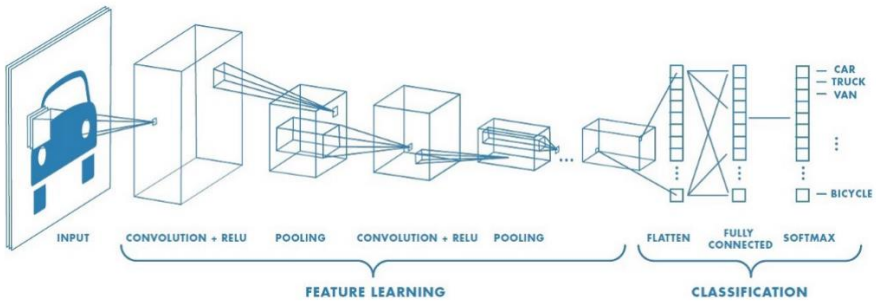
Sự ra đời của mạng CNN là dựa trên ý tưởng cải tiến cách thức các mạng nơ-ron nhân tạo truyền thống học thông tin trong ảnh. Ưu điểm của CNN là tận dụng được tính năng rút trích đặc trưng của lớp tích chập và bộ phân lớp được huấn luyện đồng thời.

Khác với phương pháp học máy truyền thống, mạng Nơ-ron tích chập CNN là một trong những mô hình mạng học sâu phổ biến nhất hiện nay, có khả năng nhận dạng và phân loại hình ảnh với độ chính xác rất cao, thậm chí còn tốt hơn con người trong nhiều trường hợp. Mô hình này đã và đang được phát triển, ứng dụng vào các hệ thống xử lý ảnh lớn của Facebook, Google hay Amazon… cho các mục đích khác nhau như tìm kiếm ảnh hoặc gợi ý sản phẩm cho người tiêu dùng.

Mạng Nơ-ron tích chập (*Convolutional Neural Network*) hay còn gọi là CNNs hoặc ConvNet là mô hình mạng được sử dụng rộng rãi, áp dụng rất nhiều trong trích xuất đặc trưng của ảnh, người ta sử dụng CNN nhiều trong các bài

toán nhận biết cũng như phân loại hình ảnh. Trong bài toán phân loại hình ảnh sử dụng CNN thì đầu vào là ảnh số, máy tính sẽ dựa vào các giá trị điểm ảnh sau đó đưa ra kết luận loại mà bức ảnh thuộc về cho bài toán phân loại, máy tính chỉ nhìn thấy bức ảnh như một mảng của các giá trị điểm ảnh. Mỗi bức ảnh thể hiện bởi ba thông số W, H, D, trong đó W là chiều rộng của ảnh, số lượng điểm ảnh trên một hàng của ma trận ảnh, H là chiều cao của ảnh, là số lượng điểm ảnh trên một cột của ma trận ảnh, D là độ sâu của ảnh.

Kiến trúc mô hình mạng CNN bao gồm lớp tích chập (*convolution*), lớp gộp (*pooling*) và lớp kết nối đầy đủ (*fully connected*), và ở lớp cuối cùng sẽ áp dụng hàm Softmax để đưa ra xác suất mà đối tượng thuộc về lớp trong bài toán phân loại. Khi các lớp này được xếp chồng lên nhau, một kiến trúc CNN sẽ được hình thành.



Hình 2.2: Mô hình CNN sử dụng cho bài toán phân loại

2.3.1. Conlovution Layer – Tầng tích chập

Đây là tầng đầu tiên của mạng CNN, tầng này giúp trích xuất đặc trưng của ảnh, lấy dữ liệu đầu vào và trải qua các phép biến đổi để tạo ra dữ liệu đầu vào cho tầng kế tiếp. Tầng tích chập trình bày mối quan hệ giữa các giá trị điểm ảnh bằng cách học các đặc trưng của ảnh thông qua việc sử dụng các hộp ô vuông (*bounding box*) đầu vào.

Đặc trưng ảnh là những chi tiết xuất hiện trong ảnh, từ đơn giản như cạnh, hình khối, chữ viết tới phức tạp như mắt, mặt, chó, mèo, bàn, ghế, xe, đèn giao thông, … Bộ lọc phát hiện đặc trưng là bộ lọc giúp phát hiện và trích xuất các đặc trưng của ảnh, có thể là bộ lọc góc, cạnh, đường chéo, hình tròn, hình vuông, …

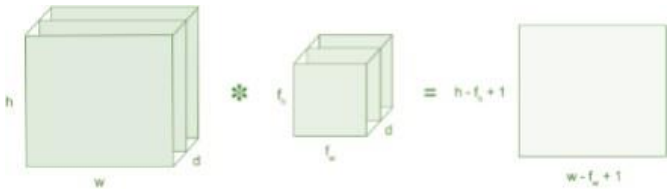
Để trích xuất đặc trưng ảnh, tầng tích chập sử dụng các bộ lọc (*filter*) để thực hiện phép tích chập khi đưa chúng đi qua đầu vào I theo các chiều của nó. Các siêu tham số của các bộ lọc này bao gồm kích thước bộ lọc F và độ trượt (*stride*) S. Kết quả đầu ra O được gọi là *feature map* hay *activation map*.

Ví dụ: Về trích xuất đặc trưng của ảnh sử dụng tích chập

Một ma trận ảnh có chiều: $h * w * d$

Một bộ lọc có: $f_h * f_w * d$

Đầu ra một ma trận ảnh có chiều $(h - f_h + 1) * (w - f_w + 1) * 1$

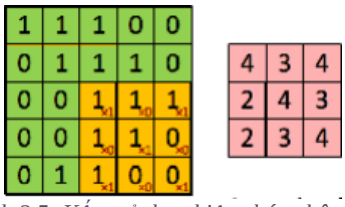


Hình 2.3:Ví dụ nhân ma trận ảnh với bộ lọc [10]

Ví dụ: Xem một ma trận có kích thước 5 x 5 và có giá trị các điểm ảnh là 0 hoặc 1, xét một bộ lọc có kích thước 3 x 3.

Hình 2.4: Ma trận và Bộ lọc [10]

Sau đó thực hiện tính tích chập của ma trận 5 x 5 với ma trận bộ lọc 3 x 3 sẽ thu được một ma trận đầu ra gọi là "*Feature Map*".



Hình 2.5: Kết quả thực hiện phép nhân [10]

Tích chập của một ảnh với các bộ lọc khác nhau sẽ đưa ra các kết quả khác nhau,như phát hiện cạnh, làm mờ, làm sắc nét ảnh.

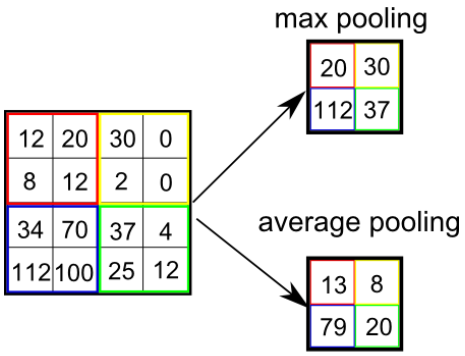
Như vậy, sau khi đưa một ảnh đầu vào cho lớp tích chập sẽ nhận được kết quả đầu ra là một loạt ảnh tương ứng với các bộ lọc đã được sử dụng để thực hiện phép tích chập. Các trọng số của các bộ lọc này được khởi tạo ngẫu nhiên trong lần đầu tiên và sẽ được cập nhật trong quá trình huấn luyện.

2.3.2. Pooling Layer – Tầng trích xuất

Tầng trích xuất (*Pooling Layer*) được sử dụng sau tầng tích chập, có chức năng giảm số lượng của tham số khi mà bức ảnh đầu vào lớn, giúp bỏ đi các thông tin dư thừa, giúp giảm chiều dữ liệu, tăng tốc độ tính toán và hiệu năng trong việc phát hiện các đặc trưng nhưng vẫn giữ được các thông tin quan trọng trong ảnh đầu vào.

Có nhiều phương thức trích xuất khác nhau để phù hợp với từng bài toán, phổ biến gồm 2 loại:

- *Max Pooling*: lấy giá trị điểm ảnh lớn nhất.
- *Average Pooling*: lấy giá trị trung bình của các điểm ảnh trong vùng ảnh cục bộ.

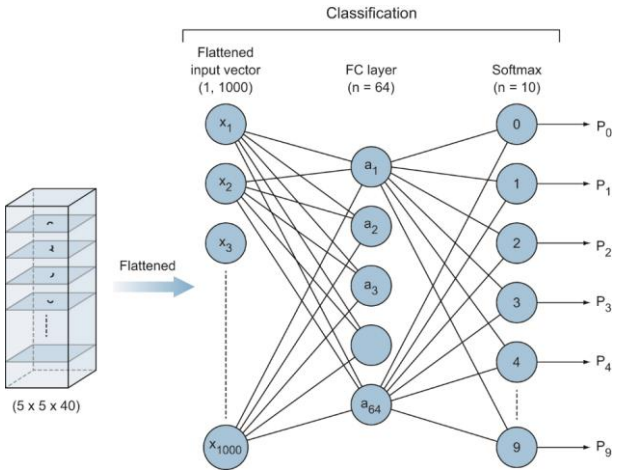


Hình 2.6: Ví dụ về Max Pooling và Average Pooling với pool size 2x2 *[10]*

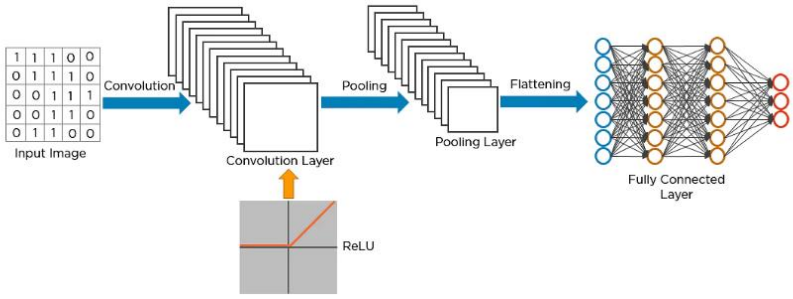
2.3.3. Fully Connected Layer – Tầng kết nối đầy đủ

Tầng kết nối đầy đủ (FC) nhận đầu vào là các dữ liệu đã được làm phẳng, mà mỗi đầu vào đó được kết nối đến tất cả các nơ-ron. Trong mô hình mạng CNNs, các tầng kếtnối đầy đủ thường được tìm thấy ở cuối mạng và được dùng để tối ưu hoá mục tiêu của mạng ví dụ như độ chính xác của tầng.

Mục đích của tầng kết nối đầy đủ là sử dụng các đặc trưng được trích xuất bởi phần *Convolution Layer* và *Pooling Layer* để phân loại hình ảnh đầu vào thành các lớp khác nhau dựa trên bộ dữ liệu huấn luyện. Tầng kết nối đầy đủ tiến hành phân lớp dữ liệu bằng cách kích hoạt hàm softmax để tính xác suất ở lớp đầu ra.



Hình 2.7: Minh họa kết nối đầy đủ và phân lớp [10]



Hình 2.8: Sơ đồ các bước từ đầu vào đến đầu ra trong mô hình CNN [10]

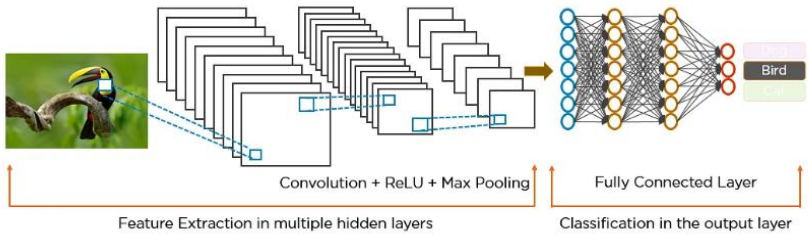
Ví dụ về các bước nhận dạng hình ảnh trong mô hình CNN:

Các *pixel* từ hình ảnh được đưa vào lớp tích chập thực hiện phép toán tích chập tạo ra các ma trận đặc trưng.

Mỗi ma trận đặc trưng được áp dụng hàm như ReLU để tạo ra một ma trận đặc trưng được chỉnh sửa. Cần nhiều lớp tích chập kết hợp ReLU để trích lọc đặc trưng rõ ràng hơn.

Các lớp tổng hợp khác nhau với các bộ lọc khác nhau được sử dụng để xác định các phần cụ thể của hình ảnh.

Các ma trận đặc trưng sau tổng hợp được làm phẳng và đưa vào một lớp được kết nối đầy đủ để có được kết quả phân loại cuối cùng.



Hình 2.9: Kết quả mô hình CNN nhận dạng hình ảnh một con chim [11]

CHƯƠNG 3. XÂY DỰNG BỘ DỮ LIỆU

3.1. Tình trạng cơ sở dữ liệu

Bộ cơ sở dữ liệu rất quan trọng trong việc xây dựng thuật toán hỗ trợ giáo viên đánh giá hứng thú người học dựa trên mô hình CNN *(Convolutional Neural Network)*. Có thể nói, độ chính xác của việc phát hiện và phân loại phụ thuộc rất nhiều vào bộ dữ liệu. nhóm lựa chọn bộ cơ sở dữ liệu KTFE bởi vì đây là cơ sở dữ liệu có thể nhìn thấy và nhiệt tự nhiên đầu tiên. Những cơ sở dữ liệu này sẽ cho phép các nhóm về biểu hiện trên khuôn mặt và cảm xúc có nhiều cách tiếp cận thực tế hơn. Hơn hết KTFE đã khắc phục lỗi trễ thời gian mà cơ sở dữ liệu cũ gặp khi thực hiện các thử nghiệm. Bên cạnh những ưu diểm của KTFE thì nhóm cũng lưu ý những nhược điểm của bộ dữ liệu như số lượng của mỗi cảm xúc không giống nhau và dữ liệu ảnh thường với cảm xúc chưa thể hiện rõ qua biểu cảm trên khuôn mặt.

3.2. Quá trình trình thu thập dữ liệu

Từ những ưu điểm trên phần 3.1 trên cũng như trong bộ dữ liệu KTFE có đối tượng phù hợp với phạm vi của nghiên cứu. Nhóm đã quyết định sử dụng bộ dữ liệu KTFE.

Từ bảy cảm xúc của bộ dữ liệu, nhóm tiến hành chọn lọc, chỉ lấy bốn cảm xúc là hạnh phúc, buồn bã, kinh ngạc và bình thường. Lý do nhóm chỉ lấy bốn loại cảm xúc trên vì trong quá trình học tập cũng như khảo sát bạn bè xung quanh, nhận thấy ba cảm xúc còn lại là lo sợ, kinh tởm, giận dữ hiếm khi có trong một tiết học.

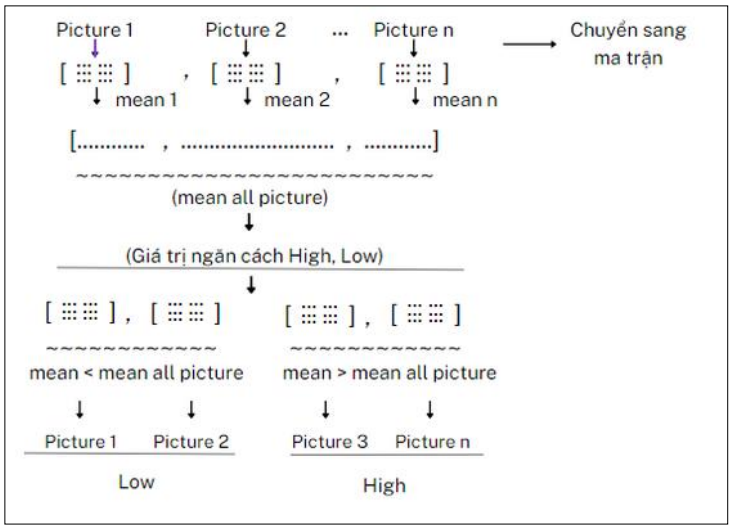
Sau khi đã có 4 cảm xúc cơ bản, chúng tôi sử dụng thuật toán *Haar cascade* nhận diện khuôn mặt và cắt mặt trong ảnh. Đưa những ảnh đó thành những tập tin riêng biệt phù hợp với cảm xúc của ảnh đó để tiện cho quá trình chạy thử. Như vậy nhóm đã có bộ dữ liệu đầu vào bao gồm: bình thường, buồn bã, kinh ngạc, hạnh phúc với tỉ lệ chia dữ liệu giữa tập train, test lần lượt là 90, 10.

	Huấn luyện (train)	**Kiểm tra (test)**	**Tổng**
Hạnh phúc	1476	95	1571
Buồn bã	2340	109	2449
Kinh ngạc	768	71	839
Bình thường	587	11	598

Bảng 3.1: Số lượng hình ảnh từng cảm xúc trong bộ dữ liệu "KTFE"

Sau đó chúng tôi tiến hành chuyển bộ dữ liệu đầu vào từ ảnh màu thành ảnh xám vì các lý do sau: Khi chuyển đổi ảnh màu thành ảnh xám, mỗi điểm ảnh chỉ có giá trị độ xám duy nhất, không còn thông tin về màu sắc. Việc tách các lớp theo giá trị điểm ảnh sẽ trở nên dễ dàng hơn và hiệu quả hơn so với việc tách các lớp theo màu sắc. Do lượng dữ liệu của bộ dữ liệu thấp nên việc sử dụng ảnh xám có thể làm giảm độ phức tạp của dữ liệu, tăng độ chính xác cho model và tốc độ tính toán trong quá trình huấn luyện.

Nhóm xử lý theo từng cảm xúc một trong 4 cảm xúc: Hạnh phúc, bình thường, buồn bã, ngạc nhiên. Khi đã chuyển từ ảnh thường về ảnh xám, ảnh xám sẽ được đưa về ma trận, nhóm tiến hành tính giá trị trung bình các chỉ số trong một bức ảnh và đưa vào mảng chứa các giá trị trung bình của mỗi ảnh. Sau đó chúng tôi tính giá trị trung bình của mảng trên để làm giá trị ngăn cách giữa mức độ high và lowmed của cảm xúc. Khi xử lý xong, nhóm đưa các ảnh vào các thư mục thuộc mức độ hứng thú của cảm xúc của ảnh đó và hoàn thành tập train, test, val cho dữ liệu. Sử dụng Ngôn ngữ lập trình Python để tạo ra bộ dữ liệu lớp thứ nhất bao gồm 7 mức độ hứng thú: Bình thường (neutral), hạnh phúc nhiều (happy-high), hạnh phúc vừa phải (happy-lowmed), buồn bã nhiều (sad-high), buồn bã vừa phải (sad-lowmed), ngạc nhiên nhiều (surprise-high), ngạc nhiên vừa phải (surprise-lowmed). Số lượng tập train và val được tách ra từ tập train ban đầu với tỉ lệ 80:20, tập test thì giữ nguyên số lượng ảnh.



Bảng 3.2: Sơ đồ thuật toán phân loại ảnh thành high và lowmed

	Train	**Test**
Happy	65.36	66.78
Sad	64.41	63.28
Surprise	57.37	59.48

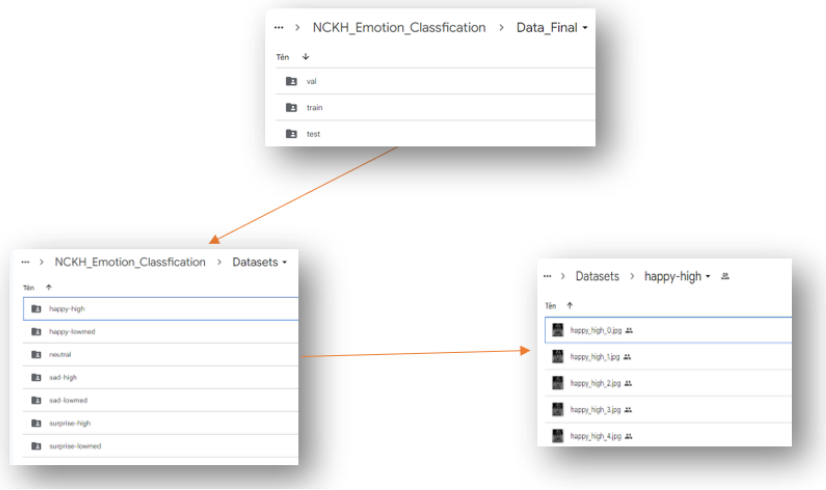
Bảng 3.3: Các giá trị "mean all picture" cho từng class

	Huấn luyện (train)	**Kiểm chứng (val)**	**Kiểm tra (test)**	**Tổng**
Happy-high	710	118	55	883
Happy-lowmed	552	96	40	688
Neutral	549	38	11	598
Sad-high	1082	196	57	1335
sad-lowmed	886	176	52	1114
surprise-high	325	45	34	404

surprise-lowmed	351	47	37	435

Bảng 3.4: Số lượng ảnh của từng loại cảm xúc trong bộ dữ liệu lớp thứ nhất "KTFE-2023-v1"

Sau khi có được bộ dữ liệu nhất, chúng tôi đã tạo ra bộ dữ liệu thứ 2 kế thừa từ các mức độ hứng thú ở trên và gom nhóm để đưa ra thang mức độ hứng thú là "bình thường", "hứng thú vừa", "rất hứng thú" và "không hứng thú". Chúng tôi sử dụng google drive để lưu trữ dữ liệu, thuận tiện cho việc sử dụng của các thành viên trong nhóm.



Hình 3.1: Nơi lưu trữ dữ liệu theo các thư mục riêng biệt trong drive

	Huấn luyện (train)	**Kiểm chứng (val)**	**Kiểm tra (test)**	**Tổng**
medium interest	903	143	77	1123
very-interest	1035	163	89	1287

	Huấn luyện (train)	Kiểm chứng (val)	Kiểm tra (test)	Tổng
neutral	549	38	11	598
uninterested	1968	372	109	2449

Bảng 3.5:Số lượng ảnh từng mức độ hứng thú trong bộ dữ liệu "KTFE-2023-v2"

Nhóm gom các cảm xúc ở bộ dữ liệu thứ nhất vào thành bộ "KTFE-2023-v2" bởi vì đánh giá hứng thú người không chỉ qua một loại cảm xúc cơ bản mà là sự kết hợp giữa những mức độ hứng thú khác nhau. Khi người học rất hứng thú với một đề tài mới họ sẽ có cảm xúc là vừa kinh ngạc nhiều vừa hạnh phúc nhiều. Kinh ngạc vì đây là đề tài mới, hạnh phúc khi tiếp thu được kiến thức bổ ích cho bản thân người học. Tương tự với như vậy, nhóm tìm hiểu và gom thành bộ dữ liệu thứ hai cho nghiên cứu.

Sau khi đã tạo ra được bộ dữ liệu thứ hai gồm bốn mức độ hứng thú, chúng tôi tiến hành gom nhóm để tạo ra bộ dữ liệu thứ ba gồm hai mức đánh giá phản hồi của người học là "tích cực" và "tiêu cực". Quy trình gom nhóm dữ liệu dựa vào thực tiễn của người học. Trong một buổi học, độ hứng thú của người học đóng vai trò quan trọng để đánh giá mức độ hiệu quả của buổi học. Khi gom nhóm dữ liệu "hứng thú vừa" và "rất hứng thú" vào nhóm "tích cực" sẽ chỉ ra những người học đang có xu hướng tham gia tích cực hơn trong buổi học. Nhóm "tiêu cực" bao gồm "bình thường" và "không hứng thú" để chỉ ra nhóm người học đang có mức độ hứng thú thấp, có xu hướng đứng ngoài quá trình học. Gom nhóm từ bốn mức độ hứng thú thành hai nhóm "tích cực" và "tiêu cực" hỗ trợ giáo viên đánh giá phản hồi của người học một cách tổng quan hơn từ đó có những thay đổi phù hợp trong quá trình dạy học.

Tên lớp	Huấn luyện (train)	Kiểm chứng (val)	Kiểm tra (test)	Tổng

Tiêu cực	2517	410	120	3047
Tích cực	1956	306	166	2428

Bảng 3.6: Số lượng của hai mức đánh giá trong bộ dữ liệu "KTFE-2023-v3"

3.3. Quá trình gán nhãn cho dữ liệu

Sau khi được xử lý size ảnh, phân loại thì tới bước gán nhãn cho ảnh. Vì đây là bài toán phân loại cho nên chúng tôi đã gán nhãn các ảnh thuộc tích cực và tiêu cực theo số như sau: 'Negative': 0, 'Positive': 1

Khi huấn luyện, các ảnh sẽ được lưu vào 1 mảng, song song với mảng chứa các ma trận ảnh sẽ có 1 mảng lưu trữ các nhãn gồm các số đã định nghĩa theo cảm xúc ở trên. Từ đó, máy sẽ tự nhận biết được ảnh nào là cảm xúc nào và được huấn luyện.

3.4. Quá trình khai phá và xử lý tạo thành bộ dữ liệu đã thống kê

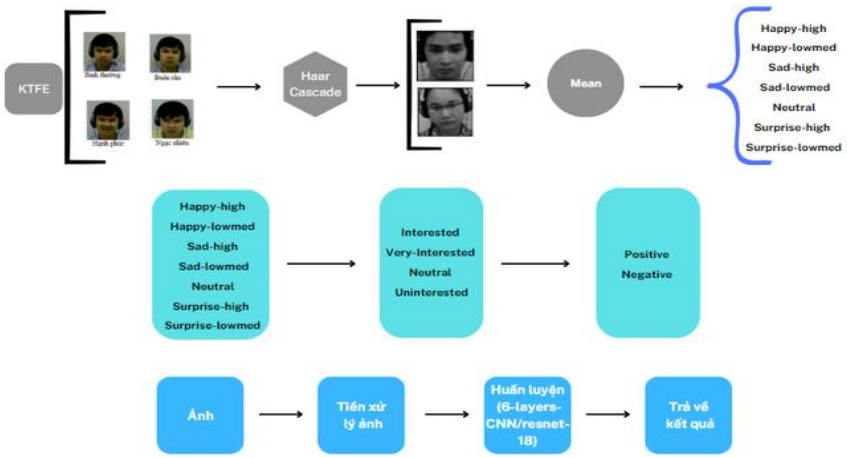
Chúng tôi đã sử dụng Jupyter Notebook trong quá trình xử lý và khai phá dữ liệu bằng máy ảo trên Google Colab. Trong quá trình khám phá, chúng tôi đã thống kê được số lượng ảnh và label cho từng tập train, val và test (Bảng 3.4.1).

	Images	**Label**
Train	4455	4455
Validation	716	716
Test	286	286

Bảng 3.7: Số lượng ảnh là label trong tập train, val, test

Sau khi đọc qua nhiều bài báo và nghiên cứu về nhận diện cảm xúc trên gương mặt, chúng tôi đã quyết định phát triển và đi theo cách huấn luyện mô hình nhận diện cảm xúc trên ảnh 48x48 chỉ chứa khuôn mặt. Từ bộ dữ liệu KTFE chúng tôi chọn, chúng tôi đã sử dụng python để chuyển đổi qua ảnh xám cùng với thuật toán nổi tiếng *Haar Cascade* để lấy được khuôn mặt trong ảnh. Vì thuật toán đã lâu đời và chưa được cập nhật gần đây cho nên vẫn còn nhiều

ảnh đã bị nhận diện khuôn mặt sai và không lấy ra được khuôn mặt cho nên chúng tôi đã kiểm tra và chắt lọc xóa những ảnh bị lỗi và chừa lại những ảnh đạt chỉ tiêu của bài. Sau khi chắt lọc và cắt khung chứa khuôn mặt với kích thước 48x48, chúng tôi tiến hành đưa các hình ảnh về ảnh xám và đọc thành ma trận với **OpenCV3** trong python.



Hình 3.2: Quy trình xử lý dữ liệu

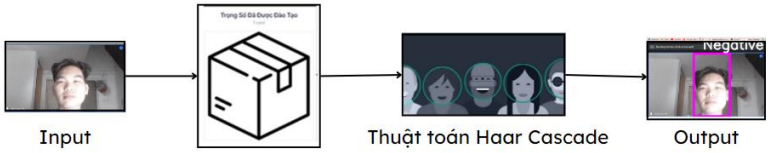
CHƯƠNG 4. MÔ HÌNH HỖ TRỢ GIÁO VIÊN ĐÁNH GIÁ NGƯỜI HỌC

4.1. Bài toán hỗ trợ giáo viên đánh giá cảm xúc người học

Nhằm đánh giá cảm xúc của người học trong hoạt động giảng dạy trực tuyến, chúng tôi đã nghĩ đến ứng dụng lý thuyết học sâu (Deep Learning) để phân tích cảm xúc trên khuôn mặt của người học trong các hình ảnh cắt từ video theo thời gian thực của lớp học thông thường, nhằm giúp người dạy nhận diện được phản hồi của người học nhanh chóng và chính xác. Điều này giúp người dạy có thể xác định mức độ hứng thú, yêu thích môn học của người học. Nghiên cứu của J. Hernik chỉ ra rằng, hứng thú ảnh hưởng tích cực đến quá trình dạy học, làm tăng sự hài lòng của người học và có thể tác động lớn đến việc ghi nhớ thông

tin [21]. Và từ những phản hồi đó, người dạy có thể đánh giá chất lượng bài giảng để có thể điều chỉnh giáo trình và phương pháp dạy học sao cho phù hợp. Kết quả là giúp nâng cao chất lượng giảng dạy và hiệu quả học tập trực tuyến.

Qúa trình dự đoán của chúng tôi sẽ sử dụng thuật toán *Haar Cascade* để tìm ra khuôn mặt trong bức ảnh và video, webcam. Sau khi có quy trình đánh giá độ hứng thú trên gương mặt của người học bằng mô hình học sâu, chúng tôi bắt đầu công cuộc xây dựng mô hình.



Hình 4.1: Quy trình đánh giá độ hứng thú qua mô hình học sâu

4.2. Tiến trình huấn luyện mô hình

Phần dữ liệu được đưa vào hệ thống phân loại thực nghiệm được phân chia ngẫu nhiên thành 3 phần là: tập dữ liệu huấn luyện *(training set*), tập dữ liệu thẩm định (*validation set*) và tập dữ liệu kiểm tra (*testing set*).

Quá trình huấn luyện: Dữ liệu hình ảnh làm đầu vào cho hệ thống học sâu. Với số lượng hình ảnh tổ hợp từ 2 lớp tiêu cực và tích cực trên tập *training set* và *validation set*. Mô hình sẽ thực hiện việc huấn luyện trên toàn bộ mạng CNN và *Fully Connected Layers*, với mục đích trích xuất đặc trưng từ các bức ảnh và phân loại chúng vào 1 trong 2 lớp đối tượng. Quá trình huấn luyện sẽ được thực hiện thông qua việc truyền dữ liệu qua các lớp và cập nhật trọng số của mô hình để giảm thiểu độ lỗi.

Sau khi huấn luyện mô hình, chúng tôi sử dụng các phương thức đánh giá đối với bài toán Classification để đánh giá độ chính xác của mô hình.

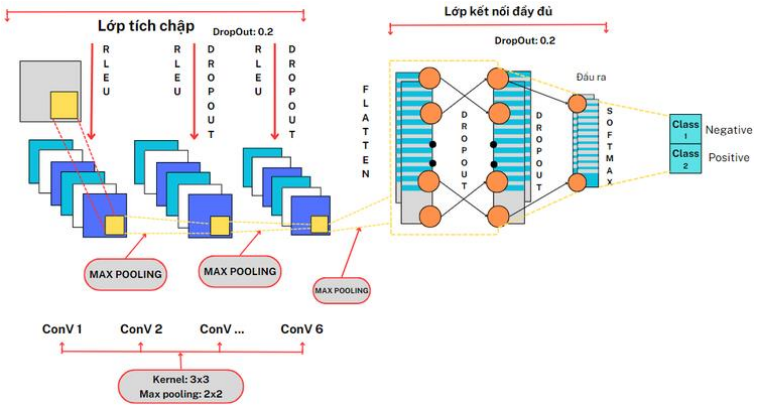
Phương thức	**Công dụng**	**Công thức**
Confusion Matrix	Thể hiện được có bao nhiêu điểm dữ liệu thực sự thuộc vào một class, và được dự đoán là rơi vào một class.	Là một matrix thể hiện phần trăm hoặc số lượng dự đoán đúng và sai của 2 class.
Accuracy	Độ đo của bài toán phân loại mà đơn giản nhất, tính toán bằng cách lấy số dự đoán đúng chia cho toàn bộ các dự đoán. Nhược điểm: chỉ cho biết bao nhiêu phần trăm lượng dữ liệu được phân loại đúng mà không chỉ ra được cụ thể mỗi loại được phân loại như thế nào, lớp nào được phân loại đúng nhiều nhất hay dữ liệu của lớp nào thường bị phân loại nhầm nhất vào các lớp khác.	Accuracy = (number of correct predictions) / (total number of predictions)
Precision	Cho biết thực sự có bao nhiêu dự đoán Positive hay Negative là thật sự True.	$Precision = \frac{True\ Positive}{\text{True Positive+False Positive}}$
Recall	Đo lường tỷ lệ dự báo chính xác các trường hợp positive trên toàn bộ các mẫu thuộc nhóm positive.	$Recall = \frac{True\ Positive}{\text{True Positive+False Negative}}$

	Recall cao đồng nghĩa với việc True Positive Rate cao, tức là tỷ lệ bỏ sót các điểm thực sự là positive là thấp.	
F1-Score	Kết hợp cả Recall và Precision.	$F1 - score = \frac{2*Precision*Recall}{\text{Precision+Recall}}$

Bảng 4.1: Ý nghĩa và công thức phương thức đánh giá mô hình trên dữ liệu

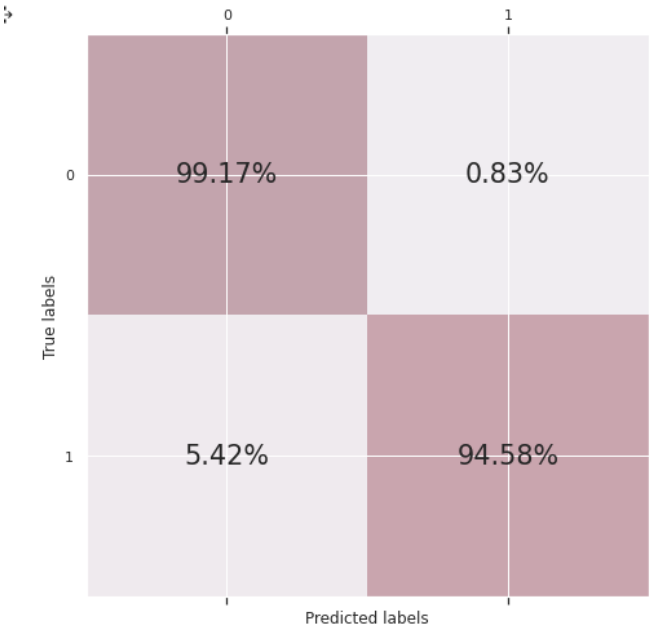
4.1.1. Mô hình mạng 6-Layers-CNN

Chúng tôi đã xây dựng một mô hình và đặt tên cho mô hình này là "6-Layers-CNN". Mô hình sử dụng phương pháp mạng nơ-ron học sâu với kiến trúc CNN để trích xuất đặc trưng từ hình ảnh và *Fully Connected Dense Layers* để phân loại ảnh vào các lớp của bộ dữ liệu KTFE. Mô hình này sử dụng các lớp Conv2D để trích xuất đặc trưng và các lớp MaxPooling2D để giảm kích thước đầu vào. Sau đó, các lớp Dense được sử dụng để phân loại dữ liệu. Mạng CNN bao gồm 6 lớp tích chập và 3 lớp pooling để giảm độ phức tạp của thuật toán. Các lớp tích chập sử dụng hàm kích hoạt ReLU để tăng tính phi tuyến và ngăn chặn tình trạng vanishing gradient. Chúng tôi cũng sử dụng lớp 3 lớp Dropout "0.2" để giảm *overfitting* và lớp padding là "*same*" để giữ nguyên kích thước ảnh. Cuối cùng, sử dụng lớp pooling để giảm kích thước của đầu vào trước khi truyền vào lớp kết nối đầy đủ Mạng nơ-ron đầy đủ sử dụng hai lớp kết nối đầy đủ với hàm kích hoạt ReLU và một lớp Softmax để tính xác suất phân loại của ảnh.

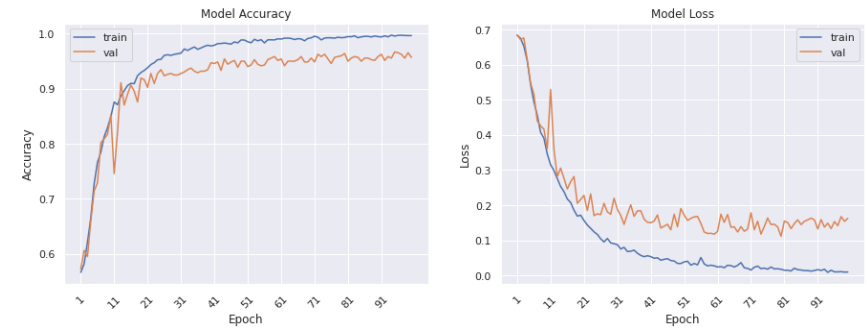


Hình 4.2: Sơ đồ mô hình 6-Layers-CNN

Quá trình dự đoán trên tập kiểm tra: Sau khi huấn luyện mạng trên tập *training set*, chúng tôi sử dụng mô hình đã được huấn luyện và phương thức *predict* của Keras để dự đoán lớp của từng ảnh trong tập kiểm tra. Kết quả dự đoán của mô hình sẽ được so sánh với nhãn thực tế của từng bức ảnh trong tập kiểm tra để tính toán độ chính xác và ma trận nhầm lẫn.



Hình 4.3: Hình thể hiện độ chính xác của mỗi lớp qua tỉ lệ phần trăm



Hình 4.4: Sự hội tụ của giá trị acc và loss của mô hình "6-Layers-CNN"

Sau khi chúng tôi huấn luyện, chúng tôi đã kiểm tra độ chính xác trên tập test và ra được tỉ lệ đúng là 96.5%. Sau đó, chúng tôi tiếp tục tìm ra các chỉ số của *precision*, *recall*, *f1-score* như bảng dưới đây (Bảng 4.2.1)

Mức độ hứng thú	Precision	Recall	F1-score
Tiêu cực	93%	99%	96%
Tích cực	99%	95%	97%

Bảng 4.2: Bảng thể hiện chỉ số Precision, Recall, F1-score của "6-layers-CNN"

Với số liệu trên, chúng tôi tự đánh giá được mô hình của mình đạt được kết quả khá cao và đạt được như mong đợi. Với chỉ số từ (Hình 4.2.2), mô hình của chúng tôi chưa có độ chính xác cao nên khi dự đoán vẫn còn xảy ra nhiều sai lệch về cảm xúc trong ảnh. Chỉ số ***Precision*** của Tiêu cực nhỏ hơn Tích cực nhưng ***Recall*** của Tích cực lại cao hơn, do đó độ chênh lệch chính xác của Tiêu cực và tích cực khá cân bằng. Để minh chứng cho việc mô hình của chúng tôi có tốt hay không, chúng tôi quyết định thử sử dụng thêm một mạng CNN nổi tiếng khác tên là ResNet-18 để đự đoán và so sánh xem 2 mô hình cái nào tối ưu trên tập train, val và test hơn. Rồi sau đó đưa ra quyết định sử dụng mô hình nào để đưa vào bài toán thực tế.

4.2.2. Mô hình mạng ResNet-18 (Residual Network 18)

ResNet (Mạng phần dư) là một trong những mạng huấn luyện CNN nổi tiếng được giới thiệu bởi Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun (2016). ResNet là một kiến trúc mạng nơ-tron sâu giúp giải quyết các vấn đề phổ biến của học sâu truyền thống, bao gồm *Vanishing gradients* và *degradation* (suy thoái) (Hochreiter, S., 1991). Khi mạng càng sâu, *gradient* có thể bị *vanishing* (biến mất) hoặc *exploding* (bùng nổ), và độ chính xác của mô hình có thể bị giảm do sự suy thoái. ResNet giải quyết vấn đề này bằng cách sử dụng các kết nối tránh, cho phép thông tin dễ dàng chảy qua mạng và huấn luyện các mô hình rất sâu. *Batch Normalization* (Ioffe, S., & Szegedy, C., 2015) là một trong các kỹ thuật được sử dụng để giúp cân bằng hệ số, giúp mô hình dễ hội tụ hơn. Các kết quả cho thấy, ResNet là một trong những kiến trúc mạng nơ-tron sâu hiệu quả nhất liên quan đến xử lý ảnh và phân loại.

ResNet hoạt động bằng cách sử dụng các khối *Convolutional Layer* để học các đặc trưng từ dữ liệu hình ảnh. Tuy nhiên, ResNet có một khối đặc biệt gọi là "*Identity shortcut*", cho phép thông tin từ đầu vào được truyền trực tiếp đến các lớp đầu ra, không qua các lớp trung gian khác. Điều này giúp tránh tình trạng thông tin bị mất dần đi trong quá trình lan truyền ngược và giúp mô hình huấn luyện tốt hơn.

Hiện tại thì có rất nhiều biến thể của kiến trúc ResNet với số lớp khác nhau như ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-152, … Với tên là ResNet theo sau là một số chỉ kiến trúc ResNet với số lớp nhất định.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		[3×3, 64; 3×3, 64] ×2	[3×3, 64; 3×3, 64] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3	[1×1, 64; 3×3, 64; 1×1, 256] ×3
conv3_x	28×28	[3×3, 128; 3×3, 128] ×2	[3×3, 128; 3×3, 128] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×4	[1×1, 128; 3×3, 128; 1×1, 512] ×8
conv4_x	14×14	[3×3, 256; 3×3, 256] ×2	[3×3, 256; 3×3, 256] ×6	[1×1, 256; 3×3, 256; 1×1, 1024] ×6	[1×1, 256; 3×3, 256; 1×1, 1024] ×23	[1×1, 256; 3×3, 256; 1×1, 1024] ×36
conv5_x	7×7	[3×3, 512; 3×3, 512] ×2	[3×3, 512; 3×3, 512] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3	[1×1, 512; 3×3, 512; 1×1, 2048] ×3
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

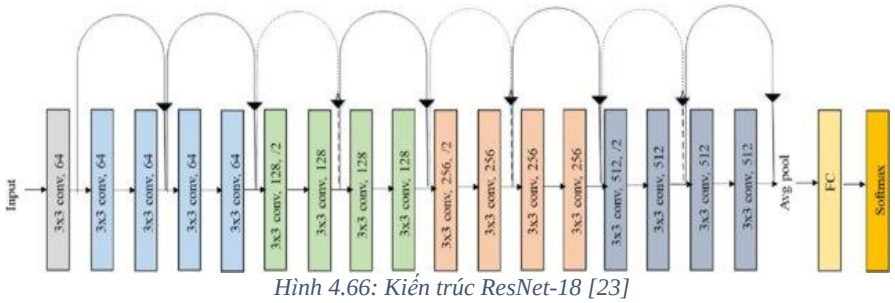
Hình 4.5: Bảng chi tiết kiến trúc mạng ResNet 18, 34, 50, 101 và 152 lớp [22]

ResNet-18 là một trong các kiến trúc ResNet đầu tiên được phát triển và có 18 lớp mạng. Trong đó có đường đi nối trực tiếp (*shortcut connections*) được thêm vào giữa các lớp để giảm thiểu *Vanishing gradient* và cho phép mô hình học được các đặc trưng phức tạp hơn.

Các đặc trưng cơ bản của ResNet-18 bao gồm: Lớp convolutional đầu tiên với 64 kernel kích thước 7x7 và stride bằng 2 để giảm kích thước ảnh đầu vào. Bốn khối tích chập (*convolutional block*) được lặp lại, mỗi khối gồm 2 lớp tích chập với *kernel* kích thước 3x3, stride bằng 1 và số lượng *kernel* tăng dần từ 64 đến 512. Mỗi khối tích chập được kết nối với một đường nối trực tiếp (*shortcut*

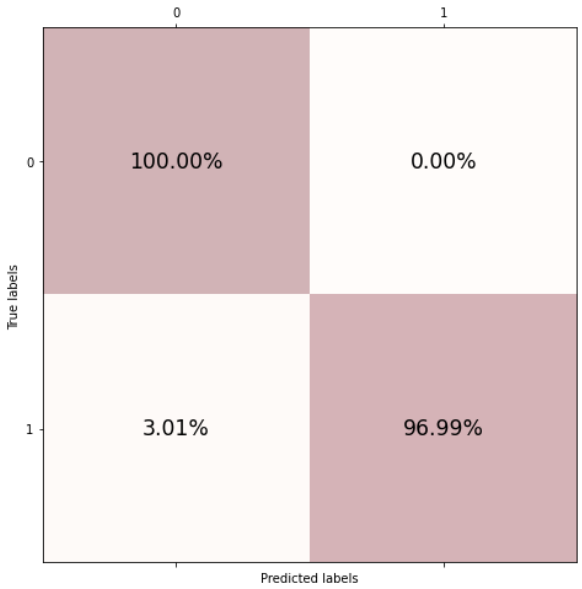
connection) để tạo thành một khối *residual block*. Đường nối trực tiếp này giúp mô hình học được các đặc trưng phức tạp hơn bằng cách giảm thiểu hiện tượng đóng băng và cho phép các đặc trưng được truyền từ lớp này sang lớp khác một cách hiệu quả hơn. Sau 4 khối tích chập là một lớp kết nối đầy đủ với 1000 nơ-ron để phân loại các đối tượng khác nhau. Cuối cùng kiến trúc có 18 tầng.

Chúng tôi lựa chọn kiến trúc ResNet-18 tiến hành huấn luyện để so sánh với mô hình "6-Layers-CNN" mà chúng tôi phát triển vì kiến trúc ResNet-18 cho phép sử dụng ảnh đầu vào là ảnh xám, phù hợp với bộ dữ liệu mà chúng tôi đã thu thập, còn các kiến trúc ResNet khác yêu cầu ảnh đầu vào là ảnh màu (RGB) và kiến trúc ResNet-18 thường được sử dụng để huấn luyện những trên các bộ dữ liệu nhỏ, đạt được độ chính xác cao trong thời gian ngắn.

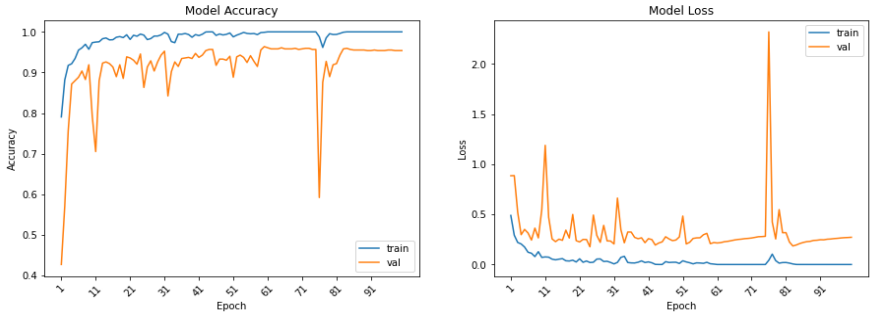


Hình 4.66: Kiến trúc ResNet-18 [23]

Sau khi huấn luyện mạng 6-Layers-CNN trên tập training set, chúng tôi tiếp tục sử dụng mô hình ResNet-18 đã được huấn luyện và phương thức ***Predict*** của Keras để dự đoán lớp của từng ảnh trong tập kiểm tra. Kết quả dự đoán của mô hình sẽ được so sánh với nhãn thực tế của từng bức ảnh trong tập kiểm tra để tính toán độ chính xác và ma trận nhầm lẫn.



Hình 4.77:Hình thể hiện độ chính xác của mỗi lớp qua tỉ lệ phần trăm



Hình 4.88: Sự hội tụ của giá trị acc và loss của mô hình "6-Layers-CNN"

Sau khi chúng tôi huấn luyện, chúng tôi đã kiểm tra độ chính xác trên tập test và ra được tỉ lệ đúng là 96.5%. Sau đó, chúng tôi tiếp tục tìm ra được các chỉ số của precision, recall, F1-score như bảng dưới đây (Bảng 4.3.1)

Mức độ hứng thú	Precision	Recall	F1-score
Tiêu cực	96%	100%	98%
Tích cực	100%	97%	98%

Bảng 4.33: Bảng thể hiện chỉ số Precision, Recall, F1-score của mô hình ResNet-18

Từ số liệu được dưa ra ở Bảng 4.3.1, chỉ số F1-score của 2 mức độ hứng thứ Tiêu cực và Tích cực cân bằng nhau. Chỉ số F1-score là sự kết hợp của chỉ số Precision và Recall, với việc F1-score bằng nhau từ 2 class mô hình ResNet-18 cho ta thấy được độ chính xác tốt và cân đối từ 2 mức độ hứng thú, giúp cho việc nhận diện được cân bằng ở cả 2 lớp.

4.2.3. So sánh "ResNet-18" với "6-Layers-CNN"

Từ số liệu bảng 4.2.1 và Bảng 4.3.1, chúng tôi tiến hành kết hợp để cho ra bảng thống kê tổng hợp các chỉ số ở cả 2 mô hình.

Mô hình	Precision	Recall	F1-score
6-Layers-CNN	96%	97%	97%
ResNet-18	98%	98%	98%

Bảng 4.2.2.1: Các chỉ số độ chính xác ở 2 mô hình ResNet-18, 6-Layers-CNN

Nhìn vào bảng trên, ta có thể thấy rõ được các chỉ số ở mô hình ResNet-18 cao hơn so với mô hình 6-Layers-CNN. Các chỉ số của ResNet-18 đồng đều và cho ra kết quả trên tập *test* tốt hơn với ResNet-18 là 98.25% còn 6-Layers-CNN là 96.5%. Tuy nhiên với 2 hình 4.2.3 và 4.3.4, ta có thể thấy giá trị *loss* và *acc* của ResNet-18 trên tập *validation* không được ổn định so với 6-Layers-CNN vì mô hình ResNet-18 có độ phức tạp cao hơn 6-Layers-CNN dẫn đến không thể tổng quát hóa tốt cho dữ liệu mới, có thể dễ gặp *overfitting*.

Như vậy, với độ chính xác tổng thể trên tập train và val cùng với dự đoán thử trên tập test của ResNet-18 cao hơn 6-Layers-CNN, cho nên chúng tôi quyết định dùng mạng ResNet-18 và lấy ra *weights* tốt nhất để thực nghiệm trên thực tế.

CHƯƠNG 5. THỰC NGHIỆM VÀ ĐÁNH GIÁ

5.1. Môi trường thực nghiệm

Về thông tin máy tính chạy thực nghiệm:

- Hệ điều hành: Window 10 – 64 bit.
- Bộ vi xử lý: Intel(R) Core (TM) i5-1035G4 CPU @ 1.10GHz.
- Bộ nhớ RAM: 8.0 GB.

Về ngôn ngữ lập trình:

- Sử dụng ngôn ngữ lập trình Python 3.10 cùng với các gói thư viện OpenCV3, Keras và Tensorflow.

5.2. Dữ liệu đầu vào

Đối với cơ sở dữ liệu, nhóm sử dụng bộ cơ sở dữ liệu đã xây dựng được trình bày ở trên từ cơ sở dữ liệu Kotani Thermal Facial Emotions (KTFE) [8] chứa 7 cảm xúc. Nhóm sử dụng 80% dữ liệu cho huấn luyện và thử nghiệm là 20%.

Bảng tiếp theo trình bày kết quả 7 loại cảm xúc thu được thông qua thực nghiệm trên tập dữ liệu kiểm thử.

Tên lớp	**Số ảnh từng mức độ hứng thú**			
	Huấn luyện (train)	**Kiểm chứng (val)**	**Kiểm tra(test)**	**Tổng**
happy-high	710	118	55	883
happy-lowmed	552	96	40	688
neutral	549	38	11	598
sad-high	1082	196	57	1335
sad-lowmed	886	176	52	1114
surprise-high	325	45	34	404

Tên lớp	Số ảnh từng mức độ hứng thú			
	Huấn luyện (train)	Kiểm chứng (val)	Kiểm tra(test)	Tổng
surprise-lowmed	351	47	37	435

Bảng 5.1:Bảng thể hiện bảy mức độ hứng thú của bộ dữ liệu

Từ bộ dữ liệu trên, tiếp tục gom nhóm và cho ra bộ dữ liệu sau.

Tên lớp	Số ảnh từng mức độ hứng thú			
	Huấn luyện (train)	Kiểm chứng (val)	Kiểm tra (test)	Tổng
medium interest	903	143	77	1123
very-interest	1035	163	89	1287
neutral	549	38	11	598
uninterested	1968	372	109	2449

Bảng 5.2: Bảng thể hiện bốn mức độ hứng thú thu được sau khi gom nhóm từ bộ dữ liệu "KTFE-2023-v1"

Từ bộ dữ liệu trên, tiếp tục gom nhóm và cho ra bộ dữ liệu thứ 3:

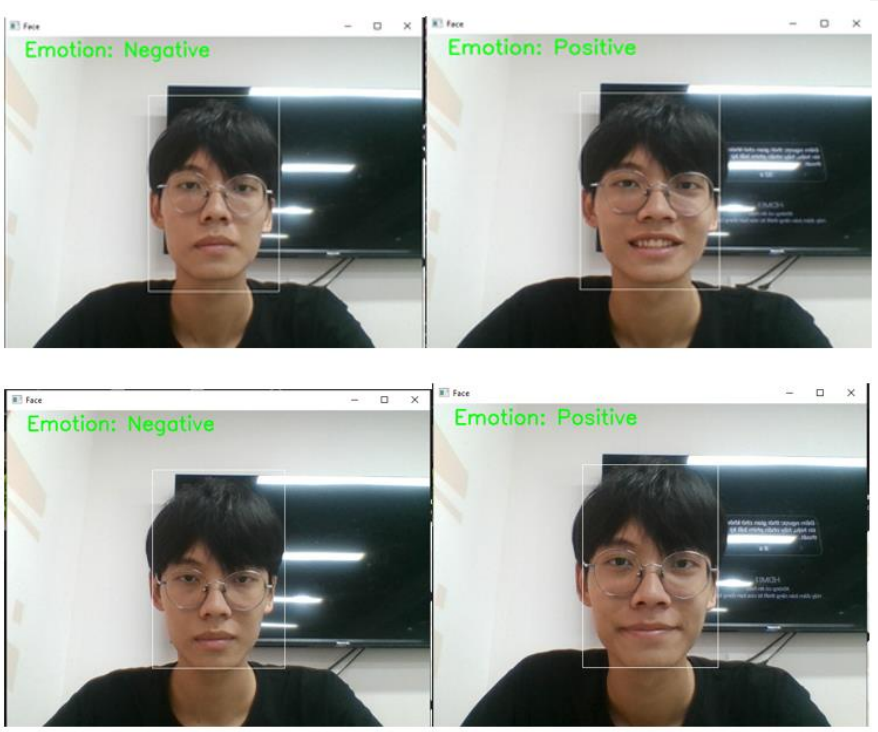
Tên lớp	Số ảnh từng mức độ hứng thú			
	Huấn luyện (train)	Kiểm chứng (val)	Kiểm tra (test)	Tổng
Tiêu cực	2517	410	120	3047
Tích cực	1956	306	166	2428

Bảng 5.3: Bảng thể hiện hai mức độ hứng thú để đánh giá cảm xúc khuôn mặt người học

5.3. Kết quả thực nghiệm

Sau khi huấn luyện phân loại hai mức độ hứng thú bằng mô hình ResNet-18 là đến bước thực nghiệm thời gian thực (real-time) trên đoạn video ghi lại hoạt động dạy học trực tuyến và ứng dụng với webcam trên máy tính chạy thực

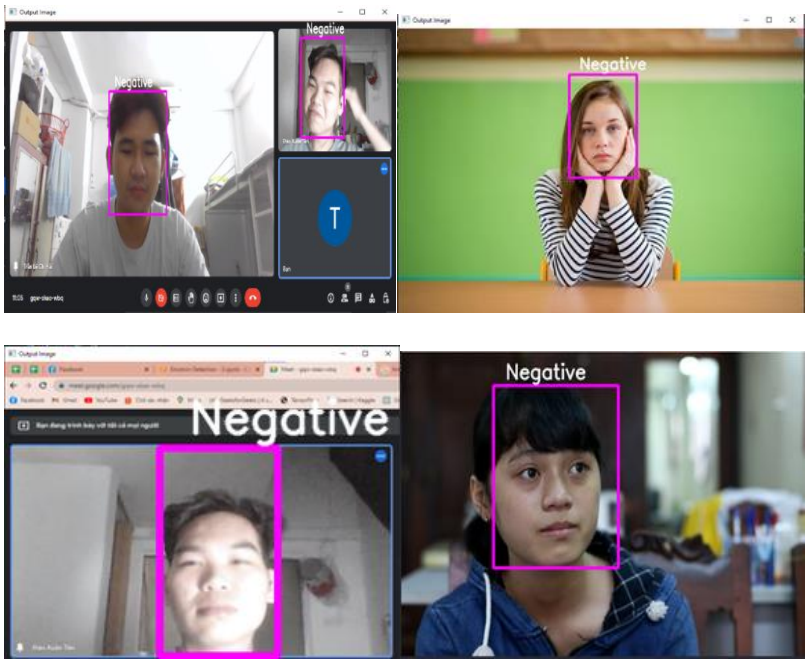
nghiệm. Qua đó đánh giá độ hứng thú của người học ở hai cấp độ: tích cực và tiêu cực.



Hình 5.3.11: Kết quả dự đoán với webcam (ảnh trái: Negative, ảnh phải: Positive)

Khi sử dụng video để dự đoán độ hứng thú trực tiếp trên gương mặt, các label được thay đổi liên tục vì thuật toán cũng liên tục dự đoán cảm xúc của người ở trước camera.

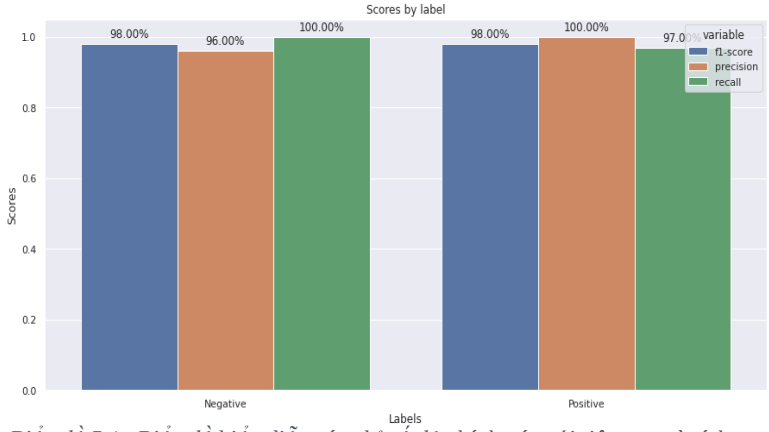
Bên cạnh đó, chúng tôi còn thực nghiệm mô hình để dự đoán các tấm ảnh có mặt học sinh trong lúc học và một vài ảnh lấy trên mạng để thử độ chính xác (Hình 5.3.2)



Hình 5.3.22: Kết quả dự đoán với ảnh

5.4. Đánh giá

Để đánh giá khách quan hiệu suất của phương pháp nghiên cứu được đề xuất, trong phần này chúng tôi tiến hành phân tích kết quả thực nghiệm đạt được.



Biểu đồ 5.1: Biểu đồ biểu diễn các chỉ số độ chính xác với tiêu cực và tích cực

Do bộ dữ liệu còn sự chênh lệch khá nhiều về số lượng các nhãn nên độ chính xác của mỗi độ hứng thú có sự chênh lệch và không quá đồng đều. Từ các chỉ số trên, ta có thể thấy mô hình được huấn luyện khá tốt nhưng khi ứng dụng real-time thì gặp khó khăn vì số lượng dữ liệu chưa đủ lớn để cố định được cảm xúc và nhiều ảnh cảm xúc chưa có sự khác nhau rõ rệt cho nên bị thay đổi nhãn liên tục và không dự đoán được chính xác nhiều. Khi ứng dụng để dự đoán qua ảnh, mô hình dự đoán được khá nhiều ảnh có sự thể hiện rõ ràng về cảm xúc, với những ảnh không thể hiện rõ cảm xúc vẫn còn bị sai.

CHƯƠNG 6. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Trong bài nghiên cứu, với mô hình CNN "6-Layers-CNN" mà chúng tôi xây dựng đã đạt được mục tiêu tự đề ra "Xây dựng mô hình hỗ trợ giáo viên đánh giá phản hồi của người học" với hai đánh giá "Tích cực" và "Tiêu cực" qua thang mức độ hứng thú là: "Hứng thú vừa", "Rất hứng thú", "Bình thường", "Không hứng thú". Với mô hình ResNet-18, chúng tôi đã thực hiện thành công việc ứng dụng một mạng CNN vào bài nghiên cứu của mình và sử dụng để thực nghiệm với ảnh chứa mặt người học trong lúc học online và real-time trên webcam.

Dữ liệu cảm xúc nhóm sử dụng là bộ dữ liệu KTFE thô, thực hiện gán nhãn để phù hợp với kích thước đầu vào của mô hình. Với dữ liệu KTFE, kết quả sau khi chạy mô hình 6-Layers-CNN tự xây dựng đạt được là 96.5%.

Bên cạnh những kết quả đã đạt được thì vẫn còn những vấn đề nhóm chưa thực hiện được. Mô hình CNN của nhóm và cả mô hình ResNet-18 được huấn luyện trên bộ dữ liệu còn nhiều ảnh khuôn mặt chưa có sự thay đổi rõ rệt trên gương mặt, nên khi áp dụng vào thời gian thực độ chính xác sẽ bị giảm và bị sai sót. Cảm xúc của người xen lẫn nhiều cảm xúc cơ bản với nhau. Bên cạnh đó, ảnh đầu vào trong thời gian thực đôi khi độ phân giải không được cao do ảnh hưởng bởi môi trường như ánh sáng, gió, đường dây mạng, khoảng cách, … Có rất nhiều đề xuất rằng nên kết hợp giữa ảnh thường với ảnh nhiệt hay kết hợp ảnh thường với hành động của người học. Nhưng nhóm vẫn chưa đủ điều kiện để thực hiện đề xuất trên.

Trong tương lai, nhóm sẽ tiếp tục nghiên cứu, xây dựng mô hình "6-Layers-CNN" cũng như tiếp thu những đề xuất trên để phát triển mô hình ngày càng tốt hơn, giúp đỡ giáo viên đánh giá người học có tỉ lệ chính xác cao hơn.

TÀI LIỆU THAM KHẢO

[1] Bộ Giáo dục và Đào tạo. (2021). *Công văn 4040/BGDĐTGDTrH ngày 16 tháng 9 năm021 về Hướng dẫn thực hiện Chương trình Giáo dục phổ thông cấp Trung học cơ sở, Trung học phổ thông ứng phó với dịch COVID-19 năm học 2021-2022*, Hà Nội, Việt Nam.

[2] Song, H., Wu, J., & Zhi, T. (2020), Online teaching for elementary and secondary schools during COVID-19. *ECNU Review of Education*, 3(4), 745-754. Retrieved January 18, 2023, from https://www.researchgate.net/publication/342227991_Results_of_Survey_on_Online_Teaching_for_Elementary_and_Secondary_Schools_During_COVID-19_Prevention_and_Control

[3] Toala, R., Durães, D., & Novais, P. (2021). Emotions and Intelligent Tutors. *Trends and Applications in Information Systems and Technologies: Volume 1* (pp. 488-496). Springer International Publishing.

[4] Park, B. J., Jang, E. H., Kim, S. H., Huh, C., & Sohn, J. H. (2012, 11-14 April). *Seven emotion recognition by means of particle swarm optimization on physiological signals: Seven em otion recognition* [Paper presentation]. Proceedings of 2012 9th IEEE International Conference on Networking, Sensing and Control, Stará Lesná, Slovakia.

[5] Keltner, D., & Haidt, J. (2001). Social functions of emotions at four levels of analysis. In W. G. Parrott (Ed.), *Emotions in social psychology: Essential readings* (pp. 175–184). Psychology Press.

[6] Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112. https://doi.org/10.3102/003465430298487

[7] Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). *The influence of affective teacher–student relationships on students' school engagement and achievement: A meta-analytic approach*. Review of Educational Research, 81(4), 493–529. https://doi.org/10.3102/0034654311421793

[8] Hưng, N. V., & Hoàng, T. H. (2022). Áp dụng mô hình học sâu nhận dạng mức độ hài lòng của người học. *Tạp chí Khoa học*, 19(12*)*, 2053.

[9] Rzayeva, Z., & Alasgarov, E. (2019, 23-25 October). *Facial emotion recognition using convolutional neural networks* [Paper presentation]. 2019 IEEE 13th international conference on application of information and communication technologies (AICT), Tashkent, Uzbekistan.

[10] Fathallah, A., Abdi, L., & Douik, A. (2017, 30 October-03 November). *Facial expression recognition via deep learning* [Paper presentation]. 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), Hammamet, Tunisia.

[11] Goodfellow, I. J., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., ... & Bengio, Y. (2013). *Challenges in representation learning: A report on three machine learning contests*. Springer berlin heidelberg.

[12] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010, 13–16 June). *The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression* [Paper presentation]. 2010 ieee computer society conference on computer vision and pattern recognition-workshops, San Francisco, California, USA.

[13] Lyons, M., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998, 14-16 April*).* *Coding facial expressions with gabor wavelets* [Paper presentation]. Proceedings Third IEEE international conference on automatic face and gesture recognition, Nara, Japan.

[14] Ekman, P. (1993). *Facial expression and emotion*. American Psychologist, 48(4), 384–392. https://doi.org/10.1037/0003-066X.48.4.384.

[15] Elgendy, M. (2020). *Deep learning for vision systems*. Simon and Schuster.

[16] Wu, J. (2017). *Introduction to convolutional neural networks*. National Key Lab for Novel Software Technology. Nanjing University. China, *5*(23), 495.

[17] 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. https://doi.org/10.1037/0033-2909.95.1.52

[18] Gupta, R. (2017). *Applying Deep Learning for Classifying Images of Hand Postures in the Rock-Paper-Scissors Game*. Otto-Friedrich University Bamberg, Germany.

[19] Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. Oxford University Research Archive https://doi.org/10.48550/arXiv.1409.1556

[20] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

[21] Hernik, J., & Jaworska, E. (2018). *The effect of enjoyment on learning*. IATED.

[22] He, K., Zhang, X., Ren, S., & Sun, J. (2016, 26-30 June). *Deep residual learning for image recognition* [Paper presentation]. Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, Nevada, USA.

[23] Ramzan, F., Khan, M. U. G., Rehmat, A., Iqbal, S., Saba, T., Rehman, A., & Mehmood, Z. (2020). A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using

resting-state fMRI and residual neural networks. *Journal of medical systems*, 44, 1-16. doi:10.1007/s10916-019-1475-2

[24] Hochreiter, S. (1991). *Untersuchungen zu dynamischen neuronalen Netzen*. [Diploma thesis, Institut für Informatik, Lehrstuhl Prof. Brauer, Technische Universität München] https://www.researchgate.net/publication/243781690_Untersuchungen_zu_dynamischen_neuronalen_Netzen

[25] Ioffe, S., & Szegedy, C. (2015, 6-11 July). *Batch normalization: Accelerating deep network training by reducing internal covariate shift*. International conference on machine learning 2015, Lille, France.