

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ GIAO THÔNG VẬN TẢI KHOA CÔNG NGHỆ THÔNG TIN**

-----🙞🙜🕮🙞🙜-----

**BÁO CÁO ĐỀ TÀI TRÍ TUỆ NHÂN TẠO**

# TÊN ĐỀ TÀI :

**ỨNG DỤNG KỸ THUẬT HỌC MÁY TRONG PHÂN LOẠI CẢM XÚC**

**GIÁO VIÊN HƯỚNG DẪN : ĐOÀN THỊ THANH HẰNG SINH VIÊN THỰC HIỆN : HOÀNG VĂN VŨ**

**NGUYỄN HUY DŨNG VŨ TRÍ MINH**

# BÙI VIỆT ĐỨC NGUYỄN VĂN TRẬN

**LỚP : 73DCTT22**

**HÀ NỘI**

# MỤC LỤC

[TÓM TẮT 3](#_bookmark0)

[GIỚI THIỆU 5](#_bookmark1)

[CHƯƠNG 1 - TỔNG QUAN VỀ ĐỀ TÀI 7](#_bookmark2)

* 1. [Giới thiệu chung 7](#_bookmark3)
     1. [Định nghĩa phân loại cảm xúc 7](#_bookmark4)
     2. [Tầm quan trọng của phân loại cảm xúc 7](#_bookmark5)
  2. [Vai trò của kỹ thuật học máy trong phân loại cảm xúc 7](#_bookmark6)
     1. [Ưu điểm của học máy 7](#_bookmark7)
     2. [Các kỹ thuật học máy phổ biến 8](#_bookmark8)
  3. [Các phương pháp phân loại cảm xúc phổ biến 9](#_bookmark9)
     1. [Phân tích văn bản 9](#_bookmark10)
     2. [Nhận diện cảm xúc từ giọng nói 9](#_bookmark11)
     3. [Phân tích khuôn mặt 10](#_bookmark12)

[CHƯƠNG 2 - CƠ SỞ LÝ THUYẾT 11](#_bookmark13)

* 1. [Tổng quan về deep learning 11](#_bookmark14)
     1. [Định nghĩa 11](#_bookmark15)
     2. [Các bài toán phát hiện cảm xúc khuôn mặt 11](#_bookmark16)
  2. [Phân tích cảm xúc 12](#_bookmark17)
  3. [Ứng dụng học sâu vào bài toán phát triển cảm xúc khuôn mặt 14](#_bookmark18)

[CHƯƠNG 3 - KẾT QUẢ THỰC NGHIỆM 15](#_bookmark19)

* 1. [Dataset Fer – 2013 15](#_bookmark20)
  2. [Quá trình huấn luyện 15](#_bookmark21)
     1. [Tải dữ liệu từ Kaggle 15](#_bookmark22)
     2. [Lưu trữ và Tải lại mô hình huấn luyện 16](#_bookmark23)
     3. [Tiền xử lý dữ liệu ảnh và Dự đoán cảm xúc từ ảnh 17](#_bookmark24)
  3. [Kết quả thí nghiệm 19](#_bookmark25)

[TÀI LIỆU THAM KHẢO 22](#_bookmark26)

2

# TÓM TẮT

Phát hiện cảm xúc khuôn mặt sử dụng các phương pháp học máy là chủ đề quan trọng trong lĩnh vực thị giác máy tính. Trong những năm gần đây, học sâu (Deep learning ) đã thể hiện được ưu thế trong bài toán xử lý dữ liệu ảnh, âm thanh cả trong nghiên cứu và công nghiệp. Trong bài báo này, một mô hình học sâu với kiến trúc mạng tích chập được giới thiệu với thiết kế gồm 8 khối chính, trong đó 7 khối mạng tích chập và khối cuối là đầu ra softmax. Kiến trúc này hướng đến việc nhận dạng các thành phần trên mặt và cảm xúc của khuôn mặt. Tập dữ liệu phổ biến về nhận dạng mặt người FER-2013 được dùng trong quá trình thực nghiệm, kết quả cho thấy việc phát hiện cảm xúc khuôn mặt của mô hình đề xuất đạt độ chính xác tương đương với những mô hình tốt nhất đã được công bố.

# GIỚI THIỆU

Bài toán phát hiện cảm xúc khuôn mặt đã có lịch sử nghiên cứu lâu dài. Từ năm 1964, Bledsoe là người đầu tiên xây dựng chương trình nhận dạng khuôn mặt tự động kết hợp với hệ thống máy tính, bằng cách phân loại khuôn mặt trên cơ sở mốc chuẩn được nhập vào bằng tay. Các thông số để phân loại là khoảng cách chuẩn, tỉ lệ giữa các điểm như góc, mắt, miệng, chóp mũi và chóp cằm. Sau này, tại Bell Labs đã phát triển một kĩ thuật dựa trên vector với 21 thuộc tính khuôn mặt được phát hiện bằng cách sử dụng kỹ thuật phân loại tiêu chuẩn mẫu. Các thuộc tính được lựa chọn đánh giá chủ yếu là: màu tóc, chiều dài của đôi tai, độ dày môi... Năm 1986, hệ thống WISARD dựa trên mạng nơron đã có thể nhận biết được tình trạng và biểu cảm khuôn mặt một cách hạn chế.

Phát hiện cảm xúc khuôn mặt là bước phát triển tiếp sau của việc phát hiện khuôn mặt, tuy nhiên có nhiều quan điểm trong việc định nghĩa khái niệm cảm xúc, vốn rất không rõ ràng. Matsumoto phân chia cảm xúc khuôn mặt thành 7 nhóm thể hiện chính: Vui vẻ, Ngạc nhiên, Hài lòng, Buồn bực, Cáu giận, Phẫn nộ và Sợ hãi. Tuy nhiên, nhóm của Mase và Pentland cho rằng chỉ 4 loại cảm xúc được thể hiện một cách rõ ràng là Hạnh phúc, Ngạc nhiên, Giận giữ và Căm phẫn; các loại cảm xúc khác thường không rõ ràng và tùy thuộc nhiều vào kinh nghiệm của người quan sát (tức là không thể định lượng một cách chính xác). Cơ sở dữ liệu Radboud Faces Database thì phân chia cảm xúc khuôn mặt thành 8 loại: Tức giận, Căm phẫn, Sợ hãi, Hạnh phúc, Buồn rầu, Bất ngờ, Khinh miệt và Trung lập. Dataset Kaggle FER-F2013 thì lại chỉ có 7 loại cảm xúc: Giận dữ, Căm phẫn, Sợ hãi, Hạnh phúc, Buồn rầu, Bất ngờ và Trung lập.

Do việc định nghĩa khái niệm cảm xúc tương đối mờ, nên việc đánh giá chất lượng các phương pháp phát hiện cảm xúc rất tùy thuộc vào tập dữ liệu huấn luyện và kiểm tra. Ví

4

dụ như báo cáo của Mase đề xuất phương pháp nhận diện cảm xúc dựa trên các đặc trưng chuyển động cơ mặt và chỉ sử dụng phương pháp K-láng giềng gần nhất nhưng đạt mức độ chính xác lên đến 88%. Trong khi đó, với tập dữ liệu FER-2013, phương pháp tốt nhất hiện nay sử dụng RBM (máy boltzmann hạn chế) đạt độ chính xác 71%, mô hình này sử dụng khoảng 5 triệu tham số , các phương pháp còn lại đều cho kết quả dưới 70%.

Trong bài báo này, chúng em thử nghiệm một kiến trúc học sâu dựa trên nhiều lớp tích chập (ConvNet) để phát hiện cảm xúc khuôn mặt. Dữ liệu thu được từ webcam sẽ được định vị khuôn mặt bằng phương pháp haar cascade từ thư viện OpenCV, sau đó dữ liệu được chuyển vào mạng học sâu với đầu ra xác suất (softmax), trả về xác suất của 7 loại cảm xúc do hệ thống tính toán được. Kết quả thử nghiệm trên bộ dữ liệu FER-2013 đạt 66.3%, nằm trong TOP5 mô hình học máy tốt nhất của dataset này.

# CHƯƠNG 1 - TỔNG QUAN VỀ ĐỀ TÀI

### Giới thiệu chung

* + 1. Định nghĩa phân loại cảm xúc

Phân loại cảm xúc là một lĩnh vực quan trọng trong trí tuệ nhân tạo, tập trung vào việc xác định và phân loại các trạng thái cảm xúc từ dữ liệu đầu vào như văn bản, hình ảnh, âm thanh, hoặc video. Kỹ thuật học máy đóng vai trò cốt lõi trong việc xây dựng các mô hình phân loại tự động, hỗ trợ các ứng dụng như chăm sóc khách hàng, giám sát sức khỏe tinh thần, và hệ thống gợi ý.

* + 1. Tầm quan trọng của phân loại cảm xúc

Ứng dụng này được sử dụng rộng rãi trong các lĩnh vực như kinh doanh, y tế, và giáo dục. Trong kinh doanh, phân loại cảm xúc giúp phân tích phản hồi của khách hàng, từ đó cải thiện sản phẩm và dịch vụ. Trong y tế, nó hỗ trợ phát hiện sớm các vấn đề tâm lý. Trong giáo dục, việc nhận diện cảm xúc của học sinh có thể giúp giáo viên điều chỉnh phương pháp giảng dạy phù hợp.

### Vai trò của kỹ thuật học máy trong phân loại cảm xúc

* + 1. Ưu điểm của học máy
* Tự động hóa trong xử lý dữ liệu: Kỹ thuật học máy giúp tự động hóa quá trình phân tích và nhận diện cảm xúc mà không cần xây dựng các quy tắc thủ công. Điều này đặc biệt hữu ích khi xử lý dữ liệu phức tạp và có khối lượng lớn, như hàng triệu bài đăng mạng xã hội hoặc hàng ngàn video mỗi ngày.
* Khả năng học từ dữ liệu lớn: Học máy phát huy hiệu quả cao nhất khi có lượng dữ liệu lớn. Các thuật toán có thể phân tích các mẫu phức tạp trong dữ liệu, từ đó đưa ra dự đoán chính xác ngay cả trong các trường hợp không rõ ràng. Ví dụ, khi một văn bản

6

có các từ vừa mang tính tích cực vừa mang tính tiêu cực, mô hình học máy có thể dựa vào ngữ cảnh tổng thể để đưa ra dự đoán cuối cùng.

* Nâng cao hiệu suất và khả năng mở rộng: Học máy không chỉ giúp phân loại cảm xúc nhanh chóng mà còn cho phép xử lý song song khối lượng dữ liệu khổng lồ. Khi các hệ thống cần mở rộng quy mô (như phân tích hàng trăm ngàn phản hồi khách hàng), các mô hình học máy có thể dễ dàng điều chỉnh để đáp ứng nhu cầu này.
  + 1. Các kỹ thuật học máy phổ biến
* *Học không giám sát (Unsupervised Learning*): Khác với học có giám sát, học phi giám sát sử dụng những dữ liệu chưa được gán nhãn từ trước để suy luận. Phương pháp này thường được sử dụng để tìm cấu trúc của tập dữ liệu. Tuy nhiên lại không có phương pháp đánh giá được cấu trúc tìm ra được là đúng hay sai. Ví dụ như phân cụm dữ liệu, chiết xuất thành phần chính của một chất nào đó.

## *Học có giám sát (Supervised Learning):* Là phương pháp sử dụng những dữ liệu đã được gán nhãn từ trước để suy luận ra quan hệ giữa đầu vào và đầu ra. Các dữ liệu này được gọi là dữ liệu huấn luyện và chúng là cặp các đầu vào-đầu ra. Học có giám sát sẽ xem xét các tập huấn luyện này để từ đó có thể đưa ra dự đoán đầu ra cho 1 đầu vào mới chưa gặp bao giờ. Ví dụ: dự đoán giá nhà, phân loại email.

* *Học bán giám sát (Semi-supervised Learning):* Là một phương pháp học máy nằm giữa học có giám sát và học không giám sát. Trong học bán giám sát, hệ thống sử dụng một lượng nhỏ dữ liệu có gán nhãn (labeled data) cùng với một lượng lớn dữ liệu không có gán nhãn (unlabeled data) để huấn luyện mô hình.

### Các phương pháp phân loại cảm xúc phổ biến

* + 1. Phân tích văn bản

**Khái niệm:** Phân tích văn bản là một kỹ thuật trích xuất cảm xúc từ các đoạn văn bản như bài đăng mạng xã hội, bình luận sản phẩm, email, hoặc blog. Đây là phương pháp phổ biến nhất do lượng dữ liệu văn bản trên internet ngày càng tăng.

### Ứng dụng:

* Phân tích phản hồi khách hàng từ các bài đánh giá sản phẩm hoặc dịch vụ.
* Phát hiện các cuộc khủng hoảng truyền thông thông qua bài đăng mạng xã hội tiêu cực.
  + 1. Nhận diện cảm xúc từ giọng nói

**Khái niệm:** Nhận diện cảm xúc từ giọng nói dựa trên việc phân tích các đặc trưng âm thanh như tần số, cao độ, cường độ, tốc độ nói và các thay đổi âm lượng. Mỗi loại cảm xúc (vui, buồn, giận dữ, lo lắng) thường có các dấu hiệu đặc trưng trong giọng nói.



### Ứng dụng:

8

* Phân tích cảm xúc trong cuộc gọi chăm sóc khách hàng.
* Hỗ trợ điều trị tâm lý bằng cách theo dõi cảm xúc của bệnh nhân qua giọng nói.
  + 1. Phân tích khuôn mặt

**Khái niệm:** Phân tích khuôn mặt là quá trình nhận diện cảm xúc dựa trên các biểu cảm khuôn mặt như cười, cau mày, nhíu mày, hoặc mở rộng mắt. Phương pháp này thường sử dụng dữ liệu hình ảnh hoặc video.



### Ứng dụng

* Xây dựng các hệ thống kiểm tra sự hài lòng của khách hàng qua biểu cảm khuôn mặt.
* Theo dõi trạng thái cảm xúc của học sinh trong lớp học trực tuyến.

# CHƯƠNG 2 - CƠ SỞ LÝ THUYẾT

### Tổng quan về deep learning

* + 1. Định nghĩa

Học sâu (deep learning) là tập các thuật toán học máy với ý tưởng xây dựng mô hình dữ liệu có mức độ trừu tượng cao dựa trên các dữ liệu có mức độ trừu tượng hóa thấp hơn, bằng cách phân lớp dữ liệu và các biến đổi phi tuyến .

Nghiên cứu từ rất lâu cho thấy mạng nơron được chứng minh khả năng xấp xỉ vạn năng chỉ với không quá 4 lớp, nhưng chưa có phương pháp nào cụ thể ước lượng số nơron cần thiết trên mỗi lớp . Việc nghiên cứu về các mạng có số lớp lớn chỉ trở nên phổ biến sau thành công của mạng AlexNet khi mô hình này thắng giải ImageNet 2012 với khoảng cách rất xa so với các mô hình cạnh tranh , mặc dù kiến trúc CNN đã được LeCun giới thiệu từ trước đó rất lâu.

Ngoài kiến trúc CNN, các mô hình mạng học sâu còn nhiều dạng kiến trúc khác như các lớp truyền thẳng kết nối đầy đủ (fully connected layer), RNN, LSTM, GRU, DBN,…

* + 1. Các bài toán phát hiện cảm xúc khuôn mặt

Đây là một bài toán phân lớp tương đối tiêu chuẩn, đã được nghiên cứu trong một thời gian khá dài. Một hệ thống nhận diện cảm xúc khuôn mặt thường được triển khai gồm 3 bước .

* Nhận ảnh và tiền xử lý: Ảnh khuôn mặt được lấy từ nguồn dữ liệu tĩnh (chẳng hạn như từ file, database), hoặc động (từ livestream, webcam, camera,…), nguồn dữ liệu này

10

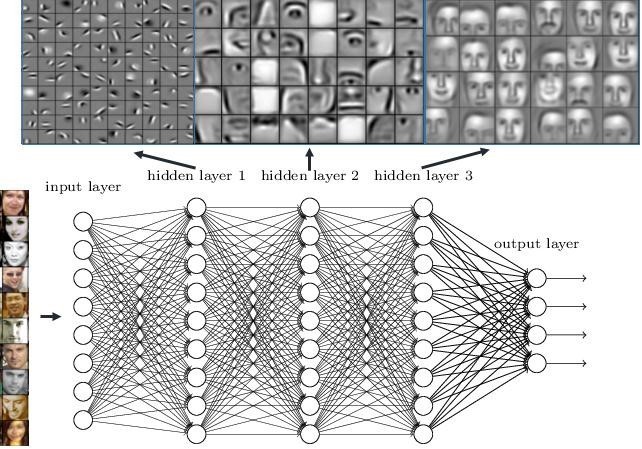
có thể trải qua một số bước tiền xử lý nhằm tăng chất lượng hình ảnh để giúp việc phát hiện cảm xúc trở nên hiệu quả hơn.

* Trích xuất các đặc trưng: Bước rất quan trọng, đặc biệt với các phương pháp truyền thống, các đặc trưng khuôn mặt được tính toán dựa trên các thuật toán có sẵn, kết quả thường là một vector đặc trưng làm đầu vào cho bước sau.
* Phân lớp và nhận diện cảm xúc: Đây là một bài toán phân lớp điển hình, rất nhiều các thuật toán có thể áp dụng trong bước này như KNN, SVM, LDA, HMM,…

Một vấn đề lớn đối với bài toán phát hiện cảm xúc khuôn mặt là thiếu sót các dataset tiêu chuẩn đủ lớn và sự chuẩn hóa các loại cảm xúc. Một trong những dataset đầu tiên cho bài toán này (năm 2009) là CK+ chỉ có 593 loạt ảnh, bộ dataset MMI cũng chỉ có 740 ảnh và 2900 video. Một số dataset xuất hiện gần đây có số lượng mẫu lớn hơn như EmotionNet có 1 triệu mẫu hoặc AffectNet có 450 nghìn mẫu. Các dataset cũng có nhiều khác biệt nhau về số lượng và cách phân loại cảm xúc, cũng như cách tính hiệu suất của các phương pháp phân loại.

### Phân tích cảm xúc

Hình dưới biểu diễn một mô hình học sâu tiêu biểu sử dụng trong nhận dạng mặt người, trong đó dữ liệu đầu vào của mạng có thể là dữ liệu ở dạng thô nhất là các điểm ảnh RGB (thậm chí không cần qua tiền xử lý). Các đặc trưng được tổ hợp và tạo thành các chi tiết nhỏ ở lớp ẩn đầu tiên, sau đó tiếp tục được tái tạo và tổ hợp mức chi tiết lớn ở lớp ẩn thứ hai, và cuối cùng các hình ảnh đặc trưng của toàn bộ khuôn mặt ở lớp ẩn thứ 3. Lớp output cho ra đánh giá xác suất khuôn mặt thuộc phân lớp nào (người nào ) .



Một mô hình học sâu thường có 3 nhiệm vụ được kết hợp trong một kiến trúc mạng duy nhất :

**Các lớp đặc trưng (features):** có nhiệm vụ chuyển đổi các đặc trưng thành dạng dữ liệu phù hợp để xử lý, chẳng hạn như các tầng tích chập (convolution), mẫu (subsampling), pooling,…

**Các lớp mô hình (modeling):** sử dụng các thuật toán học để khái quát hóa dữ liệu, chẳng hạn nơron network, restricted BM, DBN, autoencoder,…

**Các lớp giải mã (decoding):** dựa trên dữ liệu khái quát biến đổi thành đầu ra (markov random field hoặc những công cụ tương tự).

Các mạng học sâu đều có cấu trúc xác định trước, như vậy bài toán tập huấn vẫn là việc xác định giá trị các tham số trên mạng. Hiện chưa có phương pháp tập huấn nào cho phép điều chỉnh cấu trúc mạng hiệu quả.

12

### Ứng dụng học sâu vào bài toán phát triển cảm xúc khuôn mặt

Các mạng học sâu được ứng dụng rộng rãi vào bài toán phát hiện cảm xúc khuôn mặt, đặc biệt các loại mạng phù hợp với việc xử lý dữ liệu hình ảnh như CNN, DBN (deep belief network), DAE (deep autoencoder). Ngoài ra, một số tác giả sử dụng kết quả của các pre-trained model như AlexNet, VGG-face, GoogleNet,… và sử dụng các đặc trưng được trích xuất từ các mô hình này làm đầu vào cho hệ thống phân loại của họ .

Tuy được ứng dụng rộng rãi, nhưng bài toán phát hiện cảm xúc khuôn mặt vẫn là một thách thức lớn vì độ chính xác của những hệ thống hiện nay vẫn còn khá thấp; chẳng hạn như mô hình CNN của Liu et al. Cho dataset MMI mới đạt khoảng 78,5 % (tốt nhất cho dataset này); mô hình kết hợp VGG16-LSTM của Vielzeuf et al. Cho dataset AffectNet mới đạt được 48,6 % (tốt nhất cho dataset này).

# CHƯƠNG 3 - KẾT QUẢ THỰC NGHIỆM

### Dataset Fer – 2013

Dữ liệu FER-2013 được công bố bởi trang Kaggle trong khuôn khổ workshop của hội thảo ICML 2013. Dữ liệu gồm các ảnh đa cấp xám cỡ 48x48 chỉ gồm khuôn mặt hầu như được căn giữa ảnh và tỉ lệ khuôn mặt được điều chỉnh chiếm phần lớn diện tích của ảnh. Một ảnh sẽ được gán nhãn nằm một trong bảy loại cảm xúc giá trị từ 0 đến 6 (0: giận dữ, 1: căm phẫn, 2: sợ hãi, 3: hạnh phúc, 4: buồn rầu, 5: bất ngờ, 6: trung lập).

Bộ dữ liệu này gồm 28.709 mẫu huấn luyện, mẫu kiểm tra công khai có 3.589 ảnh. Khi thực hiện đánh giá mô hình, Kaggle sẽ sử dụng một bộ kiểm tra khác cũng có

3.589 ảnh, vì vậy kết quả đánh giá của ban giám khảo có thể có sai lệch so với sử dụng bộ test công khai, một số trường hợp đặc biệt sai lệch có thể lên đến 5% .

Chúng em sử dụng bộ dữ liệu này cho mô hình thử nghiệm vì bộ dữ liệu có số mẫu khá lớn, phù hợp với việc huấn luyện với mạng học sâu, vốn đòi hỏi nhiều mẫu hơn các phương pháp học máy thông thường. Ngoài ra, bộ dữ liệu được cấu trúc dễ dàng xử lý bởi thư viện Keras/TensorFlow và có nhiều kết quả đối chứng khi thực hiện so sánh mô hình của chúng em với các kết quả của những nhóm nghiên cứu khác.

### Quá trình huấn luyện

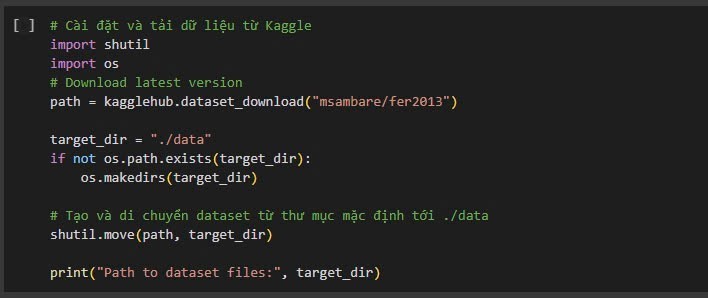
* + 1. Tải dữ liệu từ Kaggle

### Quy trình thực hiện:

* Dữ liệu từ Kaggle (tên dataset: msambare/fer2013) được tải về sử dụng lệnh kagglehub.dataset\_download().

14

* Dữ liệu từ Kaggle (tên dataset: msambare/fer2013) được tải về sử dụng lệnh kagglehub.dataset\_download().
* Thư mục mặc định để lưu trữ dữ liệu là ./data.
* Nếu thư mục ./data chưa tồn tại, chương trình sẽ tự động tạo bằng cách sử dụng hàm os.makedirs().
* Dữ liệu tải xuống được di chuyển vào thư mục này bằng lệnh shutil.move().

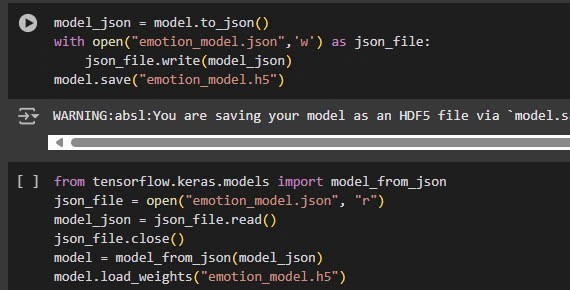


* + 1. Lưu trữ và Tải lại mô hình huấn luyện

**Lưu Trữ Mô hình:** Lưu kiến trúc mô hình: Sử dụng phương pháp model.to\_json(), kiến trúc mô hình được lưu dưới dạng tệp JSON (emotion\_model.json). Lưu trọng số mô hình: Sử dụng phương pháp model.save(), trọng số của mô hình được lưu trong tệp HDF5 (emotion\_model.h5).

**Tải lại Mô hình:** Tải kiến trúc mô hình: Tệp JSON (emotion\_model.json) được đọc và sử dụng để tái tạo lại mô hình bằng hàm model\_from\_json(). Tải trọng số mô

hình:Tệp trọng số (emotion\_model.h5) được tải lên và áp dụng lại cho mô hình sử dụng phương pháp model.load\_weights().



* + 1. Tiền xử lý dữ liệu ảnh và Dự đoán cảm xúc từ ảnh

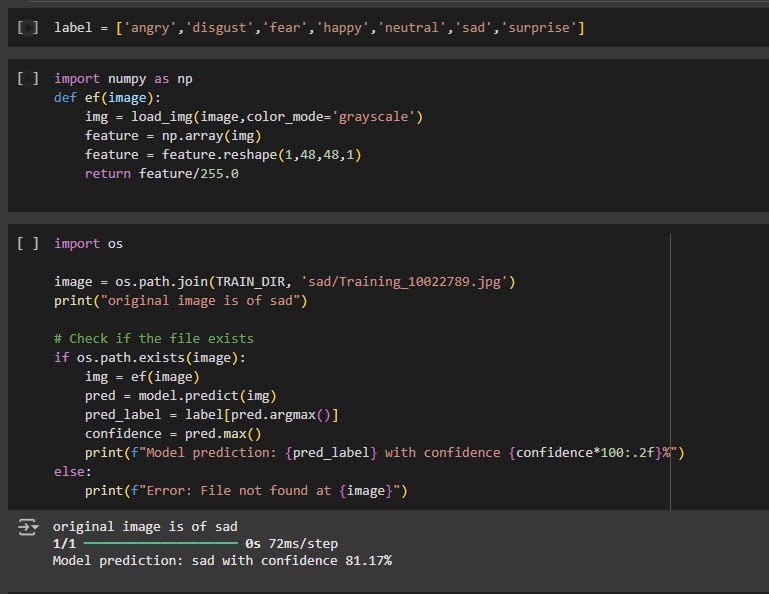
**Danh mục nhãn cảm xúc:** Các nhãn cảm xúc được sử dụng trong mô hình bao gồm: angry, disgust, fear, happy, neutral, sad, surprise.

**Hàm tiền xử lý ảnh (hàm ef):** Ảnh được tải lên và chuyển sang chế độ grayscale bằng load\_img(image, color\_mode='grayscale'). Ảnh được chuyển thành mảng numpy và reshape về kích thước (48, 48, 1) bằng np.array() và reshape(). Giá trị pixel được chuẩn hóa về khoảng [0, 1] bằng cách chia cho 255.

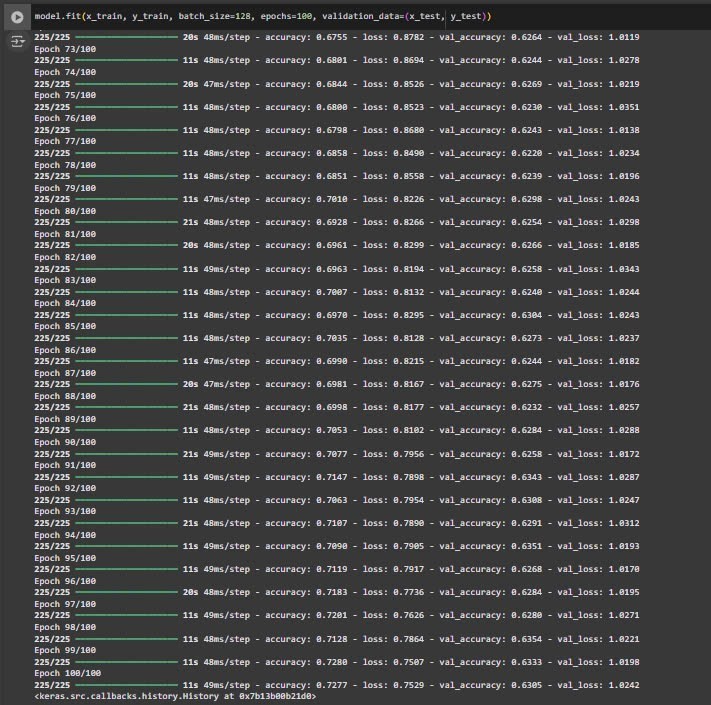
**Kiểm tra tệp ảnh:** Chương trình kiểm tra xem tệp ảnh có tồn tại trong thư mục huấn luyện hay không.

**Dự đoán nhãn cảm xúc:** Ảnh sau khi tiền xử lý được đưa vào mô hình để dự đoán nhãn. Mô hình trả về chỉ số nhãn (pred.argmax()) và giá trị độ tin cậy (confidence).

16



Ảnh minh họa dưới đây thể hiện chi tiết sự thay đổi của các chỉ số trong từng epoch:



### Kết quả thí nghiệm

Để triển khai huấn luyện và thử nghiệm mô hình đề xuất, ngôn ngữ Python và thư viện tensorflow, numpy, opencv-python, screeninfo, matplotlib, kagglehub, pandas, scikit-learn được sử dụng cho việc xây dựng mô hình mạng CNN. Dữ liệu FER-2013 được tiền xử lý không đáng kể, ngoại trừ việc chuyển đổi đa cấp xám từ dạng số nguyên 0 đến 255 về miền số thực [0, 1] nhằm hỗ trợ tốt hơn cho dữ liệu đầu vào của mạng tích chập.

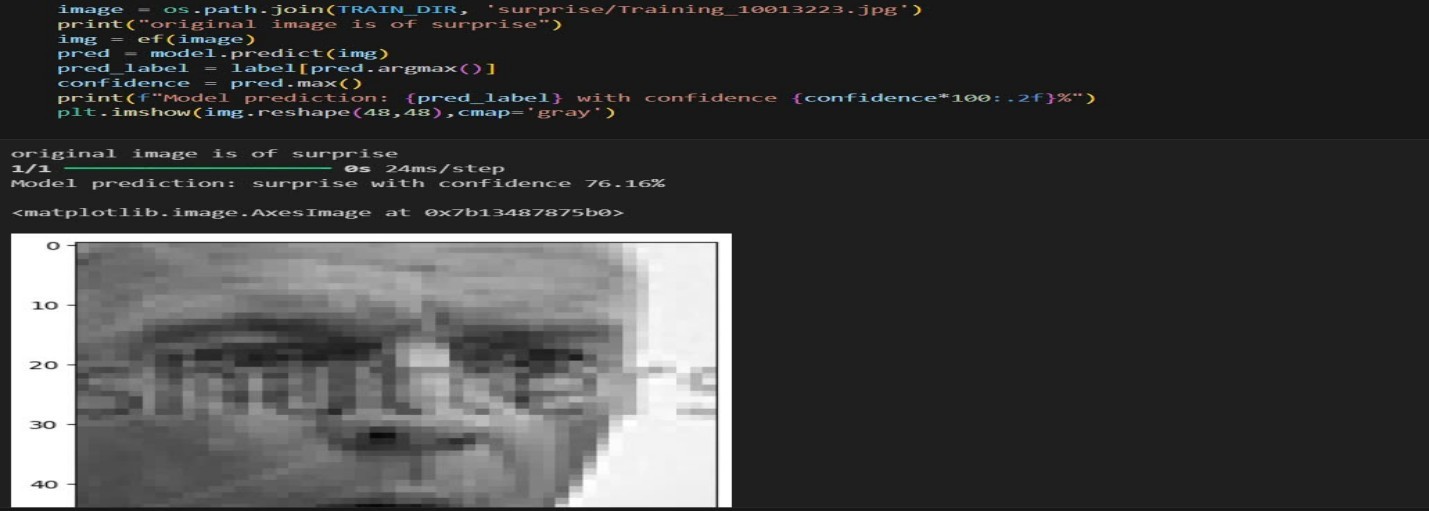
Ngôn ngữ Python kết hợp thêm OpenCV cũng được sử dụng để viết chương trình minh họa hỗ trợ cho việc xử lý dữ liệu đầu vào từ webcam/camera. Quá trình xử lý qua 5 bước như sau:

18

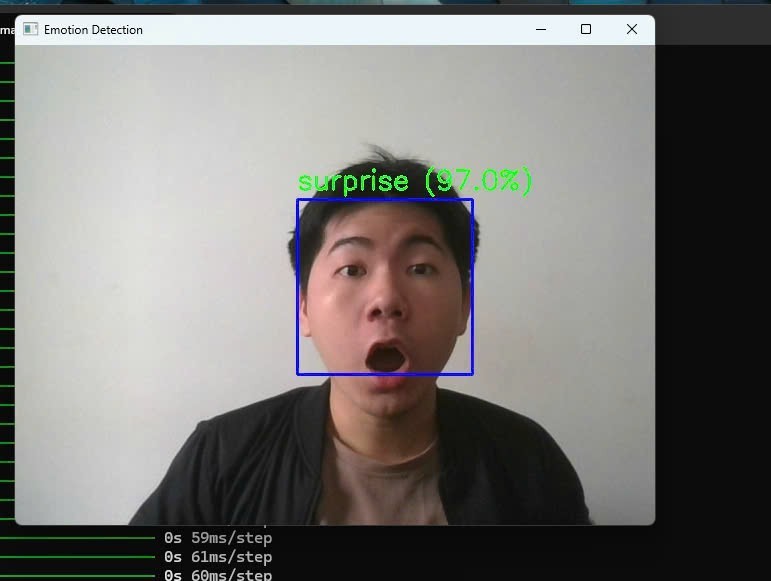
* Ảnh đầu vào được chuyển thành đa cấp xám;
* Dùng haar cascade (OpenCV) tìm kiếm vùng mặt người trên ảnh đầu vào;
* Vùng ảnh mặt người được chuyển đổi về kích thước 48x48;
* Ảnh 48x48 đa cấp xám chuyển đổi về miền [0, 1] sau đó đưa vào mô hình CNN;
* Đầu ra của CNN là xác suất của các cảm xúc, chọn cảm xúc có xác xuất cao nhất làm kết quả cuối cùng.

Tất cả các thí nghiệm được chạy trên máy trạm sử dụng bộ xử lý Intel i9-7920X, RAM 64 GB và GPU GTX 1080 Ti, hệ điều hành Ubuntu 18.04; các thư viện hỗ trợ Keras 2.2.4, TensorFlow 1.12, CUDA 10.0.130, cuDNN 7.4.1.

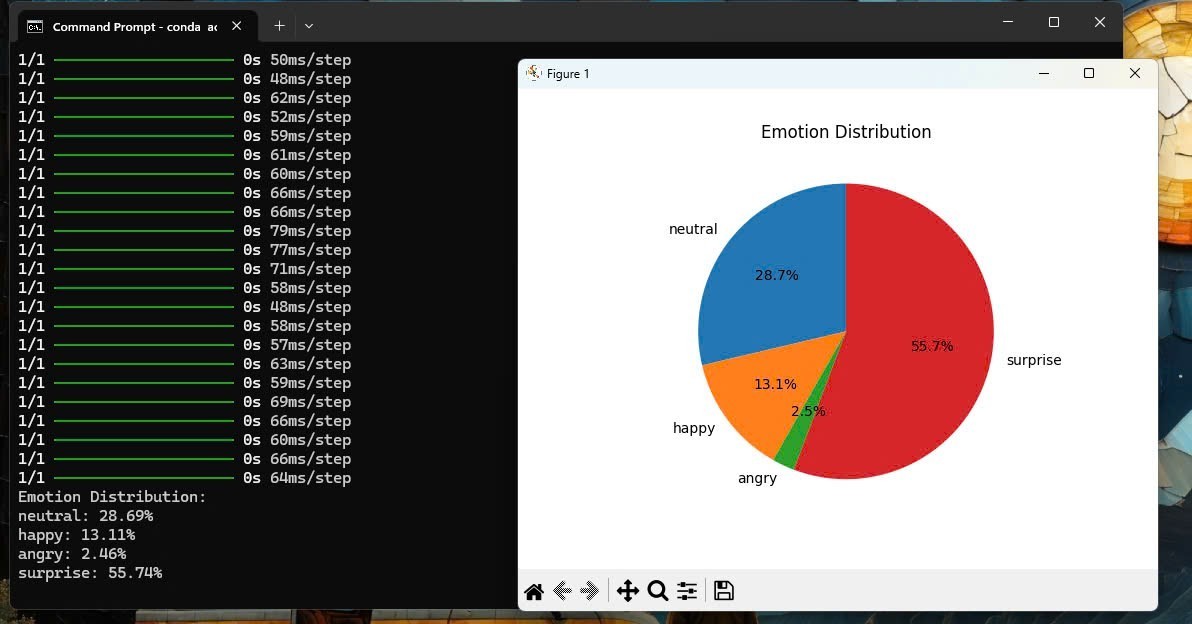
Mô hình được huấn luyện với epochs = 100, tuy nhiên kết quả về độ chính xác trên tập huấn luyện và tập kiểm tra gần như ổn định sau bước 70 khi kiểm nghiệm thực tế. Độ chính xác trên tập dữ liệu kiểm tra không bị giảm sau khi mạng đã ổn định, như vậy có thể thấy mô hình không bị hiện tượng quá khớp. Muốn tăng độ chính xác của mô hình, chúng tôi điều chỉnh phù hợp về số filter trên mỗi lớp và có thể tăng thêm một số lớp ẩn trong mạng CNN nhằm tăng khả năng nhận biết các cấu trúc phức tạp trên khuôn mặt.



Kết quả thử nghiệm thực tế cho thấy mô hình khá nhạy khi nhận biết cảm xúc hạnh phúc (happy), khá kém với cảm xúc căm phẫn (disgust). Việc hầu hết các mô hình được công bố với tập dữ liệu FER-2013 đều chỉ đạt độ chính xác thấp (dưới 70%), điều này có thể cho thấy bộ dữ liệu này có những yếu tố mất cân bằng hoặc nhiễu khi gán nhãn dữ liệu.



Nhận diện khuôn mặt bằng cảm xúc



20

Biểu đồ thể hiện cảm xúc

Ứng dụng này không chỉ cung cấp giải pháp phát hiện cảm xúc thời gian thực, mà còn mở ra các cơ hội ứng dụng trong nhiều lĩnh vực khác nhau. Việc kết hợp các công nghệ như OpenCV, Keras và matplotlib cho thấy tiềm năng mạnh mẽ của hệ thống trong việc phân tích và hiểu rõ hơn về cảm xúc của người dùng.

# TÀI LIỆU THAM KHẢO

1. Bledsoe, W. W (1964). “The Model Method in Facial Recognition”, *Technical Report PRI 15*. Panoramic Research, Inc., Palo Alto, California.
2. Matsumoto, David, and Hyi Sung Hwang (2011). "Reading facial expressions of emotion", *Psychological Science Agenda*, Vol 25, No5, pp. 10-18.
3. K. Mase, A. Pentland (1991), “Recognition of facial expression from optical flow”, *IEEE TRANSACTIONS on Information and Systems*, Vol E74-D, No10, pp. 3474- 3483.
4. I Goodfellow, D Erhan, PL Carrier, A Courville, M Mirza, B Hamner, W Cukierski, Y Tang, DH Lee, Y Zhou, C Ramaiah, F Feng, R Li, X Wang, D Athanasakis, J Shawe- Taylor, M Milakov, J Park, R Ionescu, M Popescu, C Grozea, J Bergstra, J Xie, L Romaszko, B Xu, Z Chuang, and Y. Bengio (2013). "Challenges in Representation Learning: A report on three machine learning contests." arXiv 2013.
5. Paul Viola and Michael Jones (2001). "Rapid Object Detection using a Boosted Cascade of Simple Features", Conference on Computer vision and Pattern recognition 2001.
6. Docs, OpenCV. “Face Detection Using Haar Cascades.”, OpenCV: Face Detection Using Haar Cascades, 4 Aug. 2017.
7. François Chollet. “Xception: Deep Learning with Depthwise Separable Convolutions”. arXiv 2017.
8. P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression", *Proceedings of IEEE on CVPR for Human Communicative Behavior Analysis*, San Francisco, USA, 2010.
9. Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., & van Knippenberg, A. (2010). *Presentation and validation of the Radboud Faces Database*. Cognition & Emotion, 24(8), 1377-1388. DOI:

10.1080/02699930903485076.

22