

BÁO CÁO MÔN HỌC

MÁY HỌC VÀ ỨNG DỤNG

HỆ THỐNG NHẬN DIỆN CẢM XÚC NGƯỜI DÙNG TRONG ROBOT DỊCH VỤ

Ngành: Robot và trí tuệ nhân tạo
Lớp: 22DRTA1

GIẢNG VIÊN : THS. PHẠM QUỐC PHƯƠNG

Sinh viên thực hiện:	MSSV:	Lớp:
Nguyễn Văn Đạt	2286300010	22DRTA1
Huỳnh Long	2286300028	22DRTA1
Nguyễn Chấn Huy	2286300020	22DRTA1

TP. Hồ Chí Minh, ngày 20 tháng 6 năm 2025

BÁO CÁO MÔN HỌC

MÁY HỌC VÀ ỨNG DỤNG

HỆ THỐNG NHẬN DIỆN CẢM XÚC NGƯỜI DÙNG TRONG ROBOT DỊCH VỤ

Ngành: Robot và trí tuệ nhân tạo

Lớp: 22DRTA1

GIẢNG VIÊN : THS. PHẠM QUỐC PHƯƠNG

Sinh viên thực hiện:	MSSV:	Lớp:
Nguyễn Văn Đạt	2286300010	22DRTA1
Huỳnh Long	2286300028	22DRTA1
Nguyễn Chấn Huy	2286300020	22DRTA1

TP. Hồ Chí Minh, ngày 20 tháng 6 năm 2025

VIỆN KỸ THUẬT HUTECH**PHIẾU GIAO ĐỀ TÀI TIỂU LUẬN VẤN ĐÁP****TÊN MÔN HỌC : MÁY HỌC VÀ ỨNG DỤNG****NGÀNH: ROBOT VÀ TRÍ TUỆ NHÂN TẠO****1. Họ và tên sinh viên/ nhóm sinh viên được giao đề tài (sĩ số trong nhóm.....):**

(1) Nguyễn Văn Đạt.....MSSV: 2286300010Lớp: 22DRTA1

(2) Huỳnh Long.....MSSV: 2286300028Lớp: 22DRTA1

(3) Nguyễn Chấn HuyMSSV: 2286300020Lớp: 22DRTA1

2. Tên đề tài : Hệ thống nhận diện cảm xúc người dùng trong robot dịch vụ..........
.....**3. Các dữ liệu ban đầu :**

Để xây dựng hệ thống nhận diện cảm xúc người dùng trong robot dịch vụ, nhóm sử dụng tập dữ liệu ảnh khuôn mặt gán nhãn cảm xúc gồm 7 loại cảm xúc cơ bản: *happy, sad, angry, fear, disgust, surprise, neutral*. Dữ liệu bao gồm:

- **Ảnh tĩnh khuôn mặt** với định dạng chuẩn, được thu thập từ tập dữ liệu công khai (FER2013 hoặc tương đương).
- **Đặc trưng trích xuất từ ảnh** bao gồm:
 - + **HOG (Histogram of Oriented Gradients)** – mô tả biên và hình dạng.
 - + **LBP (Local Binary Patterns)** – đặc trưng kết cấu da mặt.
 - + **Facial Landmarks** – các điểm đặc trưng như mắt, miệng, lông mày.
- Dữ liệu huấn luyện đã được chia thành tập huấn luyện và kiểm thử theo tỷ lệ phù hợp.
- Dữ liệu được xử lý trước để chuẩn hóa kích thước, ánh sáng và khử nhiễu cơ bản.

4. Nội dung nhiệm vụ:**a. Nghiên cứu các phương pháp học máy áp dụng cho nhận diện cảm xúc:**

- Áp dụng học có giám sát (Supervised Learning) bằng cách huấn luyện mô hình SVM trên dữ liệu đã gán nhãn.
- Tìm hiểu tiềm năng của học không giám sát để nhóm người dùng theo biểu hiện cảm xúc tương đồng.

- Gợi ý áp dụng học tăng cường (Reinforcement Learning) để điều chỉnh phản hồi của robot theo thời gian thực, dựa trên cảm xúc nhận diện được.

b. Xây dựng mô hình phân loại cảm xúc:

- Trích xuất đặc trưng từ ảnh khuôn mặt bằng HOG, LBP và Landmark.
- Huấn luyện mô hình SVM nhiều lớp để phân loại cảm xúc.
- Đánh giá mô hình bằng các chỉ số: Precision, Recall, F1-score, Accuracy và Confusion Matrix.

c. Thiết kế giao tiếp người – robot:

- Tích hợp mô hình phân loại vào hệ thống robot dịch vụ.
- Thiết kế các phản hồi (lời nói, biểu cảm, hành vi) của robot tương ứng với từng cảm xúc người dùng.
- Thử nghiệm giao tiếp theo thời gian thực trong môi trường giả lập.

d. Đánh giá và phân tích:

- Phân tích điểm mạnh, điểm yếu của mô hình.
- Đề xuất hướng cải tiến với các phương pháp học sâu (CNN) và tích hợp đa kênh (giọng nói, video)

5. Kết quả tối thiểu phải có:

- **Một mô hình phân loại cảm xúc** được huấn luyện từ dữ liệu ảnh khuôn mặt, sử dụng đặc trưng HOG, LBP và Landmark.
- **Độ chính xác tổng thể của mô hình đạt tối thiểu 50%** đối với bài toán phân loại 7 cảm xúc – một ngưỡng khả thi khi sử dụng các kỹ thuật cổ điển.
- **Ma trận nhầm lẫn (Confusion Matrix)** thể hiện mức độ nhầm lẫn giữa các cảm xúc, đặc biệt các nhóm cảm xúc tiêu cực như *fear*, *sad*, *angry*.
- **Các chỉ số đánh giá đầy đủ:** Precision, Recall, F1-score cho từng cảm xúc.
- **Minh họa tương tác người–robot** với khả năng phản hồi phù hợp (ví dụ: phản hồi vui vẻ khi người dùng “happy”).
- **Demo hệ thống** hoặc video mô phỏng phản ứng của robot khi nhận diện đúng cảm xúc người dùng.

Ngày giao đề tài: 6 / 6 / 2025

Ngày nộp báo cáo: 20 / 6 / 2025

TP. HCM, ngày 20 tháng 6 năm 2025.

Sinh viên thực hiện

(Ký và ghi rõ họ tên các thành viên)

Nguyễn Văn Đạt

Huỳnh Long

Nguyễn Chấn Huy

Giảng viên hướng dẫn

(Ký và ghi rõ họ tên)

CĐR	Tiêu chí đánh giá (trọng số)	Mức chất lượng (≥ 4: đạt)					Điểm	
		Mức F (0-3.9)	Mức D (4.0-5.4)	Mức C (5.5-6.9)	Mức B (7.0-8.4)	Mức A (8.5-10)	GVHD (50%)	GVPB (50%)
CLO1	Trình bày tổng quan các phương pháp xử lý của đề tài (20%)	Không nêu được tổng quan hoặc sai nội dung	Nêu tổng quan sơ lược, thiếu đầy đủ	Trình bày cơ bản đúng, còn sơ sài	Trình bày rõ, có dẫn chứng phù hợp	Tổng quan đầy đủ, logic, có phân tích so sánh		
CLO2	Đánh giá hiệu suất mô hình học máy bằng các tiêu chí phù hợp (20%)	Không đưa ra tiêu chí hoặc sử dụng sai hoàn toàn	Có nêu tiêu chí nhưng áp dụng chưa đúng	Áp dụng đúng nhưng phân tích chưa sâu	Áp dụng đúng, phân tích hợp lý	Áp dụng chính xác, lựa chọn tiêu chí phù hợp, phân tích sâu sắc		
CLO3	Thiết kế mô hình ứng dụng học máy vào bài toán thực tiễn (30%)	Mô hình không liên quan hoặc không khả thi	Mô hình còn sơ khai, chưa rõ ràng	Có ý tưởng nhưng thiết kế chưa tối ưu	Mô hình rõ ràng, hợp lý, có tính thực tiễn	Thiết kế logic, sáng tạo, thể hiện tư duy giải quyết vấn đề tốt		
CLO4	Đánh giá tác động và xu hướng của học máy trong Robot/AI (30%)	Không nêu được xu hướng hoặc nêu không đúng	Nêu vài xu hướng nhưng chưa có phân tích	Nêu đúng xu hướng nhưng thiếu chiều sâu	Phân tích được ảnh hưởng thực tế và nêu ví dụ	Đánh giá sâu sắc, có phân tích dự báo và liên hệ thực tiễn cụ thể		
TỔNG CỘNG:								
ĐIỂM TRUNG BÌNH								

TP. HCM, ngày 18 tháng 06 năm 2025

Giảng viên hướng dẫn

(Ký và ghi rõ họ tên)

ThS. Phạm Quốc Phương

TP. HCM, ngày 18 tháng 06 năm 2025

Giảng viên phản biện

(Ký và ghi rõ họ tên)

ThS. Huỳnh Phát Huy

TP. HCM, ngày 18 tháng 06 năm 2025

Giảng viên hướng dẫn

(Ký và ghi rõ họ tên)

ThS. Phạm Quốc Phương

TP. HCM, ngày 18 tháng 06 năm 2025

Giảng viên phản biện

(Ký và ghi rõ họ tên)

ThS. Huỳnh Phát Huy

TP. HCM, ngày 18 tháng 06 năm 2025

Giảng viên hướng dẫn

(Ký và ghi rõ họ tên)

ThS. Phạm Quốc Phương

TP. HCM, ngày 18 tháng 06 năm 2025

Giảng viên phản biện

(Ký và ghi rõ họ tên)

ThS. Huỳnh Phát Huy

Mục lục

Tóm tắt	13
I. Mục tiêu	13
II. Các bước thực hiện	13
III. Thuật toán sử dụng	14
1. <i>SVM</i>	14
a. Linear kernel	16
b. Polynomial kernel	16
c. RBF (Radial Basis Function) kernel	16
d. Hinge Loss	19
2. <i>HOG</i>	20
Khái niệm	20
Nguyên lý hoạt động	20
Ưu điểm	21
Ứng dụng	21
3. <i>Local Binary Pattern (LBP)</i>	21
Khái niệm	21
Nguyên lý hoạt động	21
Công thức	21
Ưu điểm	22
Ứng dụng	22
4. <i>Landmark</i>	22
Khái niệm	22
Nguyên lý hoạt động	22
Cấu trúc điểm	22
Ưu điểm	22
Ứng dụng	22
IV. Triển khai	23
1. <i>Cấu trúc dataset</i>	23
2. <i>Lưu đồ thuật toán</i>	23
V. Kết quả và đánh giá	25
1. <i>Bảng kết quả huấn luyện</i>	25
2. <i>Confusion matrix</i>	26
3. <i>Kết quả nhận diện</i>	27
VI. Kết luận	28
VII. Tài liệu tham khảo	30

Tóm tắt

Đề tài tập trung vào xây dựng một hệ thống nhận diện cảm xúc khuôn mặt nhằm nâng cao khả năng tương tác người – máy trong các robot dịch vụ. Hệ thống khai thác các đặc trưng hình ảnh như HOG (Histogram of Oriented Gradients), LBP (Local Binary Pattern) và Landmark facial points để trích xuất thông tin khuôn mặt, kết hợp với thuật toán phân loại SVM (Support Vector Machine) sử dụng các kernel khác nhau (linear, polynomial, RBF) nhằm phân loại cảm xúc người dùng.

Phần lý thuyết trình bày chi tiết cơ sở của các thuật toán sử dụng, bao gồm đặc điểm của từng loại kernel trong SVM, cơ chế hoạt động và ứng dụng của HOG, LBP và Landmark trong xử lý ảnh khuôn mặt. Dữ liệu huấn luyện được xây dựng có cấu trúc rõ ràng, qua đó nhóm triển khai hệ thống nhận diện cảm xúc theo lưu đồ thuật toán cụ thể.

Kết quả thực nghiệm thể hiện qua bảng số liệu huấn luyện, ma trận nhầm lẫn và tỷ lệ nhận diện cho thấy hệ thống có độ chính xác cao, phù hợp ứng dụng thực tế trong các robot dịch vụ để nhận diện và phản hồi cảm xúc người dùng một cách tự nhiên và thân thiện hơn.

I. Mục tiêu

Mục tiêu của đề tài là xây dựng một hệ thống có khả năng nhận diện cảm xúc của người dùng thông qua hình ảnh khuôn mặt, từ đó tích hợp vào robot dịch vụ để cải thiện khả năng tương tác người – máy. Cụ thể, đề tài hướng đến các mục tiêu sau:

1. **Tìm hiểu và áp dụng các kỹ thuật trích xuất đặc trưng khuôn mặt** như HOG, LBP và Landmark để phục vụ cho quá trình nhận diện cảm xúc.
2. **Nghiên cứu và triển khai mô hình phân loại cảm xúc** sử dụng thuật toán SVM với các loại kernel khác nhau (linear, polynomial, RBF) để đánh giá hiệu quả mô hình.
3. **Xây dựng pipeline nhận diện cảm xúc hoàn chỉnh**, từ khâu xử lý dữ liệu đầu vào, trích xuất đặc trưng, huấn luyện mô hình đến phân loại cảm xúc đầu ra.
4. **Đánh giá hiệu suất của hệ thống** qua các chỉ số như độ chính xác, ma trận nhầm lẫn, nhằm đảm bảo tính khả thi và ứng dụng thực tiễn.
5. **Hướng đến tích hợp hệ thống vào robot dịch vụ**, giúp robot có thể tự động điều chỉnh hành vi và phản hồi dựa trên trạng thái cảm xúc của người dùng, từ đó nâng cao trải nghiệm người sử dụng.

II. Các bước thực hiện

Đề tài bao gồm các bước thực hiện chính như sau:

1. **Thu thập và xử lý dữ liệu hình ảnh khuôn mặt**: Sử dụng các bộ dữ liệu sẵn có (hoặc tạo mới) với các nhãn cảm xúc như vui, buồn, tức giận, ngạc nhiên, sợ hãi,

ghê tởm, trung tính. Dữ liệu được tiền xử lý để chuẩn hóa kích thước, cân bằng số lượng mẫu giữa các lớp cảm xúc.

2. Trích xuất đặc trưng khuôn mặt:

- Sử dụng **HOG (Histogram of Oriented Gradients)** để thu thập thông tin về cạnh và hướng cạnh trên khuôn mặt.
- Áp dụng **LBP (Local Binary Pattern)** để trích xuất đặc trưng kết cấu da, đặc biệt phù hợp với nhận diện biểu cảm.
- Kết hợp **Landmark facial points** để định vị các điểm quan trọng như mắt, miệng, chân mày – những vùng thường thay đổi theo cảm xúc.

3. Huấn luyện mô hình phân loại cảm xúc:

- Sử dụng thuật toán **SVM (Support Vector Machine)** với nhiều loại kernel khác nhau (linear, polynomial, RBF) để tìm ra cấu hình tối ưu cho việc phân loại cảm xúc.
- Áp dụng **kỹ thuật đánh giá chéo (cross-validation)** và tinh chỉnh tham số (hyperparameter tuning) để cải thiện hiệu suất.

4. Xây dựng hệ thống pipeline hoàn chỉnh:

- Từ ảnh đầu vào → tiền xử lý → trích xuất đặc trưng → phân loại → xuất ra nhãn cảm xúc.
- Lưu đồ thuật toán được xây dựng rõ ràng nhằm đảm bảo tính mạch lạc, dễ mở rộng và tích hợp.

5. Đánh giá hiệu quả hệ thống:

- Sử dụng các chỉ số đánh giá như **độ chính xác, ma trận nhầm lẫn (confusion matrix)** để phân tích điểm mạnh – yếu của mô hình.
- So sánh hiệu quả giữa các phương pháp trích xuất đặc trưng và các loại kernel trong SVM.

III. Thuật toán sử dụng

1. SVM

Support Vector Machine (SVM) là một thuật toán học máy dùng để phân loại dữ liệu bằng cách tìm ra hyperplane tối ưu trong không gian đặc trưng. Hyperplane này có nhiệm vụ phân tách các điểm dữ liệu thuộc các lớp khác nhau, đồng thời tối đa hóa khoảng cách (margin) giữa hyperplane và các điểm dữ liệu gần nhất của mỗi lớp. Mục tiêu của SVM là tạo ra một ranh giới quyết định vững chắc, giúp phân loại chính xác các dữ liệu mới.

Hyperplane là một không gian con có số chiều thấp hơn không gian dữ liệu gốc một đơn vị, đóng vai trò như ranh giới phân loại. Trong không gian một chiều, hyperplane là một điểm; trong không gian hai chiều, nó là một đường thẳng; trong không gian ba chiều, nó là một mặt phẳng. Tổng quát hơn, trong không gian n chiều, hyperplane có số chiều là $n-1$. SVM tìm kiếm hyperplane phù hợp nhất để phân tách dữ liệu dựa trên phương trình tổng quát:

$$w \cdot x + b = 0$$

Trong đó w là vector trọng số, x là vector đặc trưng của dữ liệu, và b là độ lệch.

SVM không chỉ tìm một hyperplane bất kỳ, mà còn tìm hyperplane có margin lớn nhất, tức khoảng cách xa nhất đến các điểm dữ liệu gần nhất của mỗi lớp. Việc tối đa hóa margin giúp mô hình tổng quát hóa tốt hơn trên dữ liệu mới, giảm nguy cơ overfitting. Hai đường biên hỗ trợ hyperplane có phương trình:

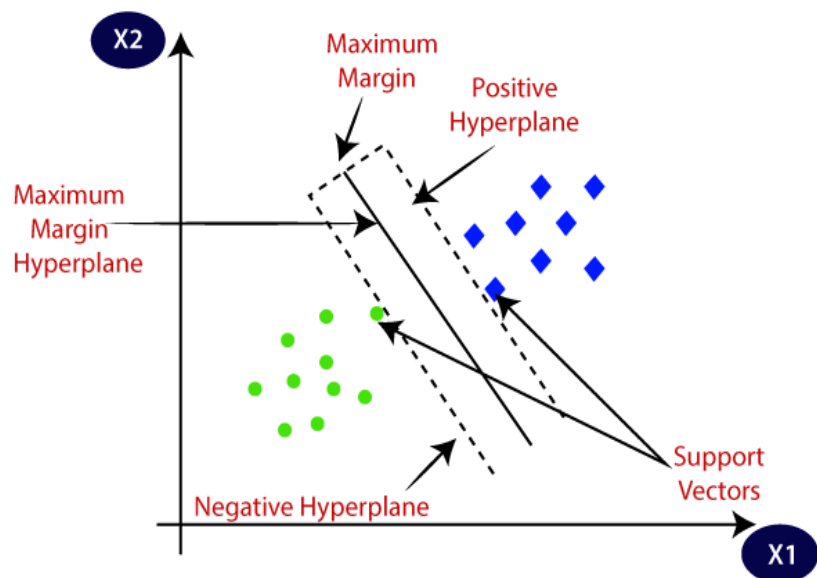
$$w \cdot x + b = +1$$

$$w \cdot x + b = -1$$

Khoảng cách giữa hai đường này chính là margin, được tính bằng:

$$\frac{2}{\|w\|}$$

Do đó, để tối đa hóa margin, SVM thực chất đang tối thiểu hóa độ lớn của vector trọng số w .



Hình 1: Mô tả phân loại dữ liệu bằng Hyperplane

Các điểm dữ liệu gần hyperplane nhất, quyết định vị trí và hướng của hyperplane tối ưu, được gọi là support vectors. Chỉ các điểm này ảnh hưởng đến việc tối ưu hóa, trong khi các điểm dữ liệu khác không có tác động đáng kể đến ranh giới phân loại.

Khi dữ liệu có thể phân tách hoàn toàn bằng một hyperplane, SVM sẽ tìm ra hyperplane tối ưu với margin lớn nhất mà không có điểm dữ liệu nào nằm giữa hoặc bị phân loại sai. Tuy nhiên, trong thực tế, nhiều bài toán có dữ liệu không thể phân tách tuyến tính. Để giải quyết vấn đề này, SVM sử dụng kernel trick.

Kernel trick giúp SVM xử lý dữ liệu phi tuyến tính bằng cách ánh xạ dữ liệu từ không gian ban đầu sang một không gian có số chiều cao hơn, nơi nó có thể phân tách tuyến tính. Thay vì tính toán trực tiếp trong không gian mới, kernel trick sử dụng các hàm kernel để tính toán tích vô hướng giữa các cặp điểm dữ liệu một cách hiệu quả. Một số hàm kernel phổ biến gồm:

a. Linear kernel: Phù hợp với dữ liệu có thể phân tách tuyến tính.

$$K(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}'$$

Trong đó, \mathbf{x} và \mathbf{x}' là hai vectơ dữ liệu trong không gian đầu vào, còn \mathbf{x}^T là chuyển vị của \mathbf{x} . Linear kernel không thực hiện bất kỳ phép biến đổi phi tuyến nào, do đó hyperplane quyết định vẫn là tuyến tính trong không gian ban đầu. Kernel này thường được sử dụng khi dữ liệu có thể phân tách tuyến tính.

b. Polynomial kernel: Xét đến các quan hệ bậc cao giữa các đặc trưng.

$$K(x, x') = (x^T x' + c)^d$$

Các tham số trong công thức:

- c là một hằng số không âm, thường được đặt là 0 hoặc 1.
- d là bậc của đa thức, quyết định độ phức tạp của ranh giới phân loại. Khi d lớn, mô hình có thể trở nên quá phức tạp và dễ bị overfitting.
- Hàm kernel đa thức cho phép SVM xem xét không chỉ các đặc trưng ban đầu mà còn cả các tương tác giữa chúng, giúp mô hình có khả năng phân loại tốt hơn trên dữ liệu phi tuyến tính.

c. RBF (Radial Basis Function) kernel:

Tạo ra ranh giới phi tuyến tính phức tạp theo phân bố Gaussian.

$$K(x, x') = \exp(-\gamma \|x - x'\|^2)$$

Các tham số trong công thức:

- $\|x - x'\|^2$ là bình phương khoảng cách Euclidean giữa hai vector.
- γ (gamma) là một tham số dương, quyết định mức độ ảnh hưởng của một điểm dữ liệu lên mô hình. Khi γ nhỏ, mô hình có xu hướng trở thành tuyến tính. Khi γ lớn,

Polynomial kernel

$$K(A,B) = (A \cdot B + C)^d$$

Chọn $C = 2, d = 2$

$$K(A,B) = (A \cdot B + 2)^2$$

$$K(A,B) = (AB + 2)(AB + 2)$$

$$= A^2B^2 + 4AB + 4$$

$$= AB + A^2B^2 + 2^2$$

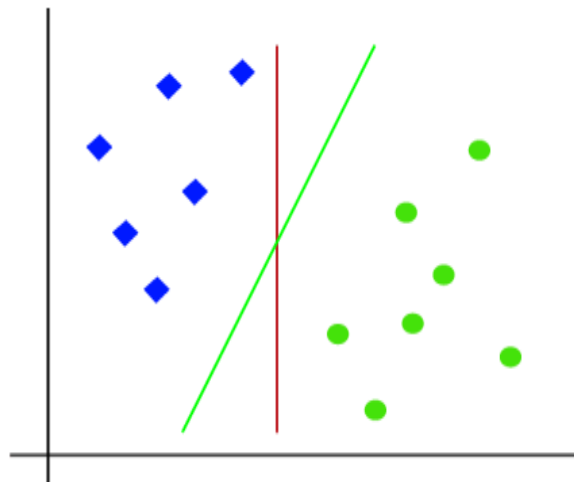
$$= (A, A^2, 2)^T (B, B^2, 2)$$

Hình 2: Mô tả cách tính tích vô hướng của các cặp điểm dữ liệu trong một không gian có chiều thấp hơn

- mô hình trở nên phức tạp hơn và có thể dễ bị overfitting.

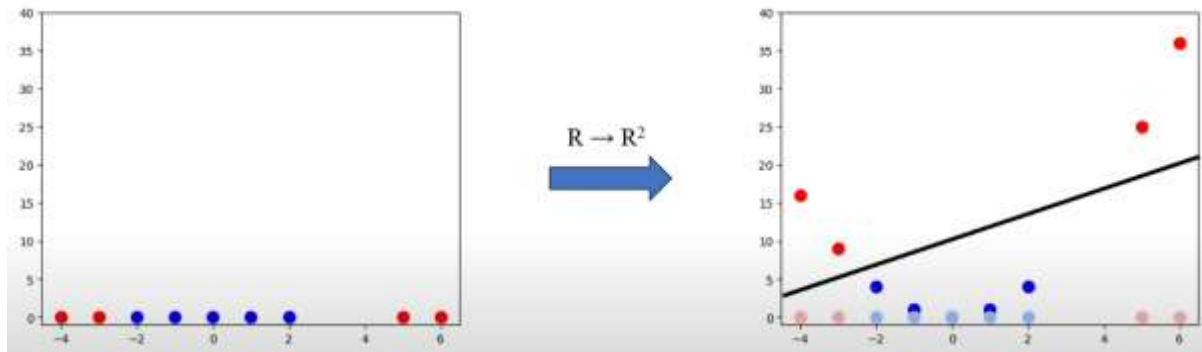
RF kernel còn được gọi là Gaussian kernel, giúp ánh xạ dữ liệu vào một không gian có số chiều rất cao, thậm chí vô hạn chiều, làm cho các bài toán phi tuyến tính có thể phân tách bằng một hyperplane.

Các hàm kernel này giúp SVM giải quyết bài toán phân loại phi tuyến tính bằng cách tính toán tích vô hướng trong một không gian đặc trưng có số chiều cao hơn mà không cần thực sự thực hiện phép biến đổi đó một cách tường minh. Việc lựa chọn kernel phù hợp và điều chỉnh các tham số của nó có ảnh hưởng quan trọng đến hiệu suất của mô hình SVM.



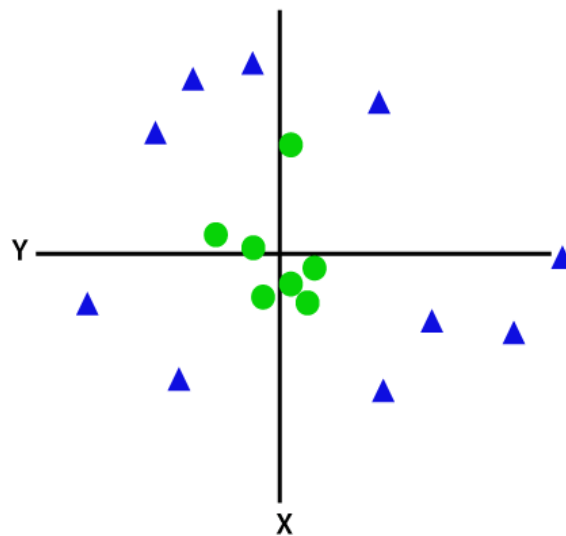
Hình 3: SVM tuyến tính

Việc chọn hàm kernel phù hợp có vai trò quan trọng trong hiệu suất của SVM trên dữ liệu phi tuyến tính.



Hình 5: Mô tả tác dụng của kernel

Trong nhiều trường hợp, dữ liệu có thể chứa nhiều hoặc điểm ngoại lai, khiến việc tìm một hyperplane phân tách hoàn toàn trở nên không khả thi. Để xử lý điều này, SVM sử dụng soft margin, cho phép một số điểm dữ liệu nằm trong vùng margin hoặc bị phân loại sai.



Hình 4: SVM không tuyến tính

Tham số C giúp điều chỉnh sự cân bằng giữa:

- C lớn: Giảm lỗi phân loại nhưng có nguy cơ overfitting.
- C nhỏ: Cho phép một số lỗi để tăng tính tổng quát hóa.

SVM tìm hyperplane tối ưu bằng cách giải một bài toán tối ưu hóa có ràng buộc. Mục tiêu là tối thiểu hóa:

$$\frac{1}{2} \|\mathbf{w}\|^2$$

Dưới ràng buộc rằng tất cả điểm dữ liệu được phân loại đúng.

Trong soft margin SVM, hàm mục tiêu trở thành:

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \zeta_i$$

Trong đó ζ_i là biến trễ, đại diện cho lượng lỗi phân loại.

d. Hinge Loss: Một hàm mất mát phổ biến trong SVM là Hinge Loss, với công thức:

$$L = \max(0, 1 - y_i * f(x))$$

Với:

$$f(x) = w \cdot x_i + b$$

y là giá trị thật của đầu ra (+1 hoặc -1).

$$L \geq 0.$$

$$y * f(x) < 1 \rightarrow 1 - y * f(x) > 0 \Rightarrow L = 1 - y * f(x)$$

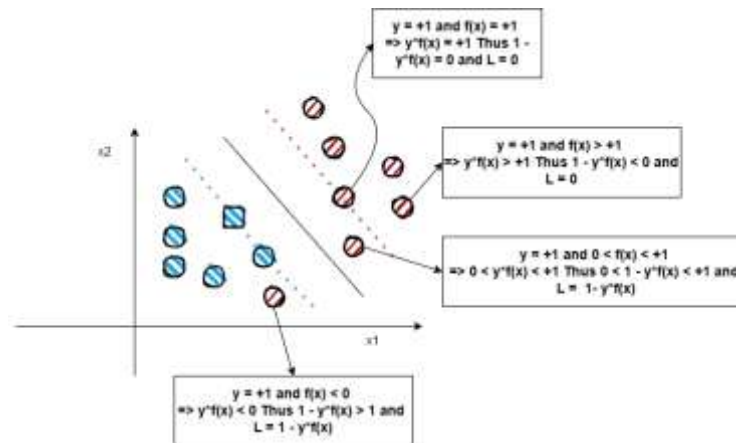
$$y * f(x) > 1 \rightarrow 1 - y * f(x) < 0 \Rightarrow L = 0$$

$$y * f(x) < 1 \rightarrow 1 - y * f(x) > 0 \Rightarrow L = 1 - y * f(x)$$

Trong mô hình Soft Margin SVM, biến slack ξ_i được định nghĩa là:

$$\xi_i = \max(0, 1 - y_i(w \cdot x_i + b))$$

Biến ξ_i thể hiện mức độ vi phạm margin của một điểm dữ liệu, đồng thời cũng chính là giá trị của Hinge Loss đối với điểm đó. Nếu một điểm dữ liệu nằm đúng phía của hyperplane và bên ngoài vùng margin, giá trị ξ_i sẽ bằng 0. Ngược lại, nếu điểm đó nằm



Hình 6: Mô tả hàm hinge loss

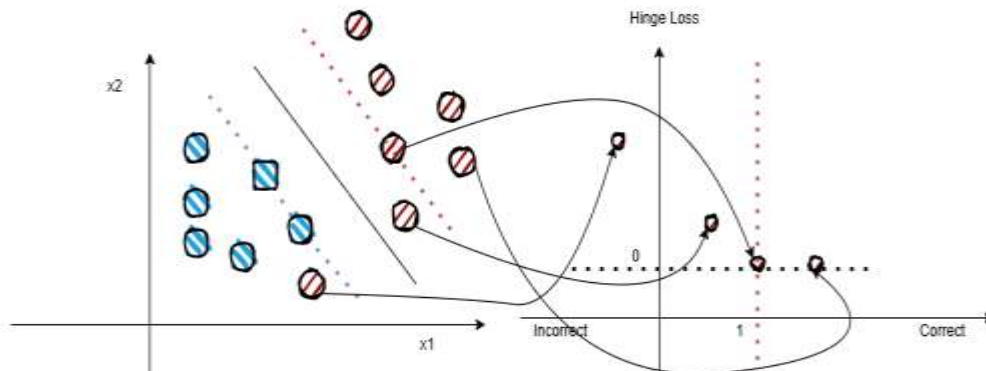
trong vùng margin hoặc bị phân loại sai, giá trị ξ_i sẽ tăng lên.

Dựa vào định nghĩa trên, hàm mục tiêu của Soft Margin SVM có thể được viết lại dưới dạng tổng Hinge Loss như sau:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \max(0, 1 - y_i(w \cdot x_i + b))$$

Trong đó:

- $\frac{1}{2} \|w\|^2$ là thành phần điều chỉnh độ phức tạp của mô hình, đảm bảo SVM tìm được một hyperplane có margin lớn nhất.
- C là tham số điều chỉnh mức độ ưu tiên giữa việc tối đa hóa margin và giảm lỗi phân



Hình 7: Mô tả hàm hinge loss

loại.

- $\sum_{i=1}^n \max(0, 1 - y_i(w \cdot x_i + b))$ là tổng Hinge Loss trên toàn bộ dữ liệu huấn luyện, giúp đo lường mức độ vi phạm margin của các điểm dữ liệu.

Hàm mục tiêu này thể hiện sự cân bằng giữa việc tối ưu hóa margin và hạn chế lỗi phân loại, đảm bảo SVM có khả năng tổng quát hóa tốt trên tập dữ liệu mới.

Bài toán tối ưu hóa này có thể được giải bằng các phương pháp như Lagrangian multipliers hoặc SMO (Sequential Minimal Optimization).

2. HOG

Khái niệm:

HOG là phương pháp trích xuất đặc trưng hình dạng dựa trên sự phân bố của các gradient định hướng trong ảnh. Kỹ thuật này đặc biệt hiệu quả trong việc mô tả biên và cấu trúc đối tượng.

Nguyên lý hoạt động:

- Ảnh được chuyển sang thang xám và tính gradient theo hai hướng (x, y) sử dụng các toán tử như Sobel.
- Ảnh chia thành các ô nhỏ (cells), thường kích thước 8×8 pixel.
- Với mỗi ô, tạo histogram gradient định hướng (thường gồm 9 hướng từ 0° đến 180°).
- Các ô được nhóm thành khối (blocks, ví dụ 2×2 ô) và chuẩn hóa histogram để tăng độ bền với nhiễu và ánh sáng.
- Ghép tất cả histogram đã chuẩn hóa thành một vector đặc trưng đầu ra.

Ưu điểm:

- Mô tả chính xác hình dạng và biên cạnh.
- Chống nhiễu tốt, hiệu quả trong các bài toán nhận dạng.

Ứng dụng:

- Phát hiện người và khuôn mặt.
- Nhận diện tư thế, cử chỉ và biểu cảm.

3. Local Binary Pattern (LBP)

Khái niệm:

LBP là phương pháp trích xuất đặc trưng kết cấu, dựa trên so sánh cường độ điểm ảnh trung tâm với lân cận. Phù hợp cho các tác vụ như nhận diện khuôn mặt và phân tích bề mặt.

Nguyên lý hoạt động:

- Với mỗi điểm ảnh, so sánh giá trị với các điểm xung quanh theo một bán kính xác định.
- Nếu điểm ảnh lân cận \geq điểm trung tâm \rightarrow ghi là 1, ngược lại là 0.
- Dãy nhị phân thu được chuyển thành số thập phân gọi là mã LBP.
- Vector đặc trưng được tạo bằng histogram phân bố các mã LBP.

Công thức:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^p$$

Trong đó:

- g_c : giá trị điểm ảnh trung tâm
- g_p : giá trị điểm ảnh lân cận
- $s(x)$ = Hàm bước, trả về 1 nếu x lớn hơn hoặc bằng 0, ngược lại 0.

Ưu điểm:

- Tính toán nhanh, đơn giản.
- Ổn định với thay đổi ánh sáng, hiệu quả mô tả kết cấu.

Ứng dụng:

- Nhận diện khuôn mặt.
- Phân loại kết cấu, phát hiện bất thường.

4. Landmark

Khái niệm:

Landmarks là các điểm định danh các bộ phận quan trọng trên khuôn mặt như mắt, mũi, miệng, viền hàm... Thường dùng bộ 68 điểm từ thư viện Dlib, mỗi điểm có tọa độ (x, y) .

Nguyên lý hoạt động:

- Phát hiện vị trí khuôn mặt bằng HOG+SVM, CNN hoặc Haar Cascade.
- Dự đoán vị trí landmarks bằng mô hình học máy đã huấn luyện.
- Trích xuất vector tọa độ gồm 2 giá trị cho mỗi điểm (68 điểm \rightarrow 136 phần tử).

Cấu trúc điểm:

- Điểm 1–17: Viền mặt
- Điểm 18–27: Chân mày
- Điểm 28–36: Mũi
- Điểm 37–48: Mắt
- Điểm 49–68: Miệng

Ưu điểm:

- Mô tả cấu trúc khuôn mặt chi tiết.
- Hữu ích trong nhận diện biểu cảm, phân tích trạng thái mắt/miệng, định hướng đầu...

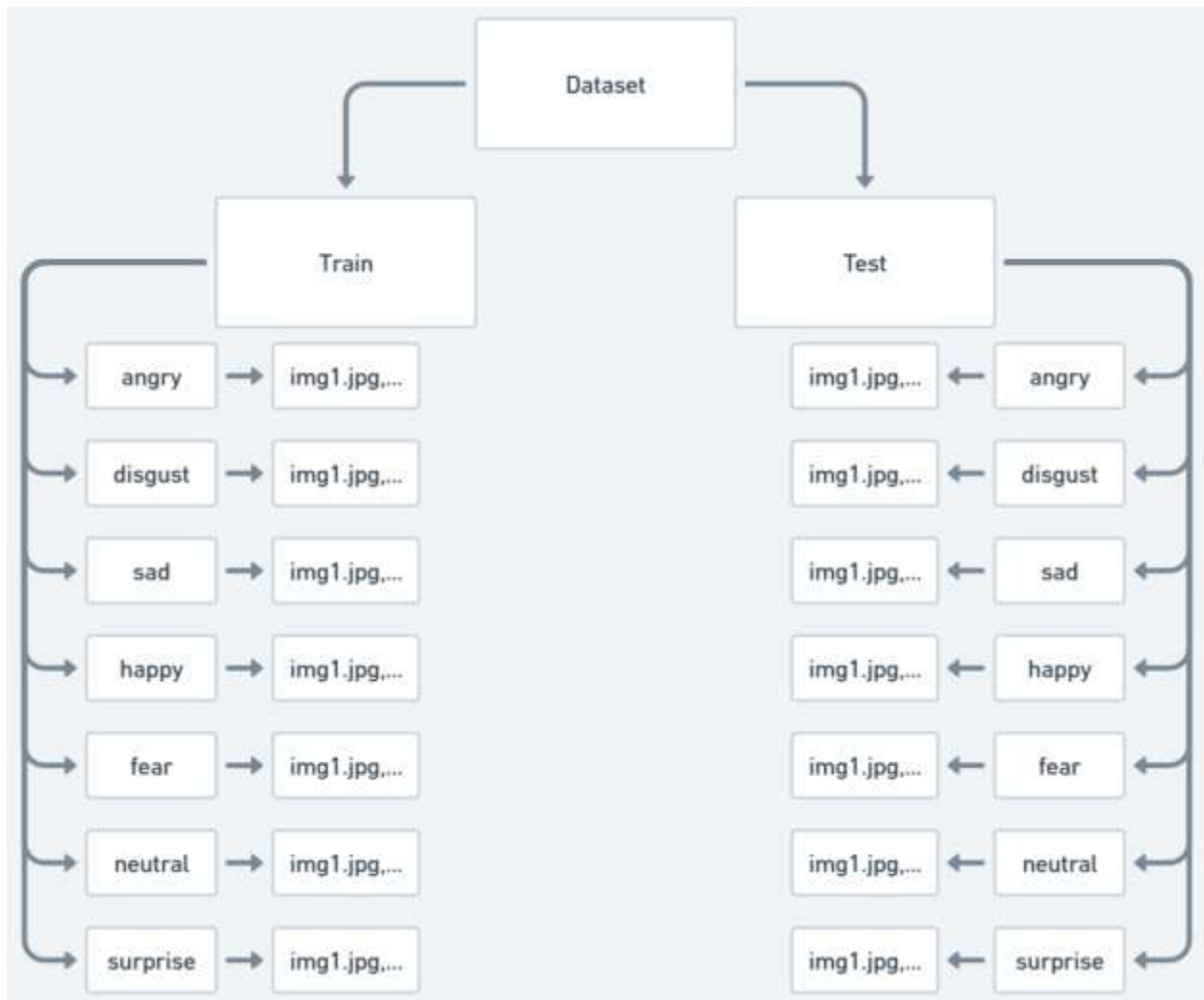
Ứng dụng:

- Nhận diện và phân tích biểu cảm khuôn mặt.
- Dự đoán trạng thái mắt (mở/nhắm), miệng.

- Cơ sở cho các bài toán như căn chỉnh khuôn mặt (face alignment) và ước lượng tư thế đầu (head pose estimation).

IV. Triển khai

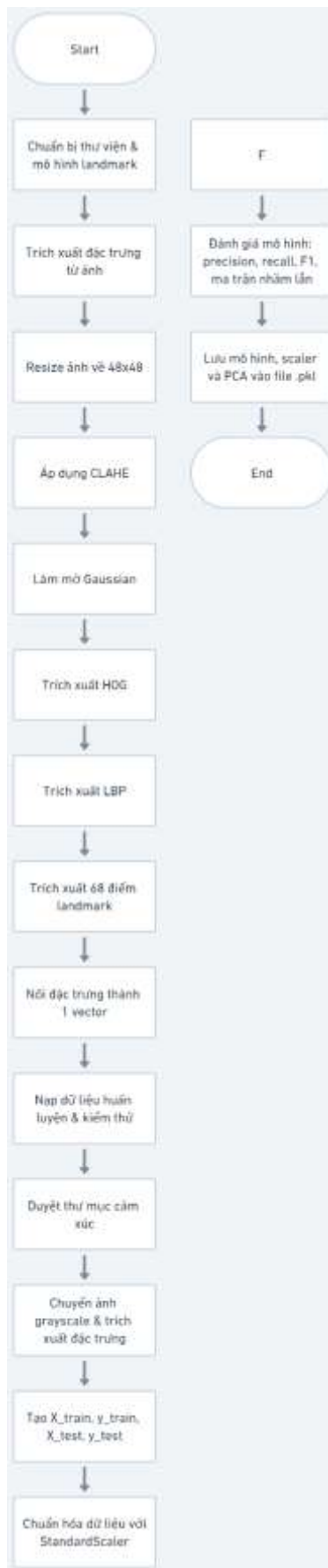
1. Cấu trúc dataset



Hình 8: Cấu trúc dataset

Trong folder dataset được dùng để huấn luyện và đánh giá model có tổng cộng 35887 ảnh trong đó có 28709 ảnh được dùng cho quá trình huấn luyện SVM và 7178 ảnh dùng cho việc đánh giá. Dataset gồm 7 lớp: angry, disgust, sad, happy, fear, neutral và surprise tương đồng cho cả hai thư mục Train và Test, mỗi thư mục chỉ chứa các ảnh biểu thị cảm xúc theo tên của thư mục đó và tên của thư mục được xem là nhãn của các ảnh trong thư mục.

2. Lưu đồ thuật toán



Hình 9: Lưu đồ giải thuật

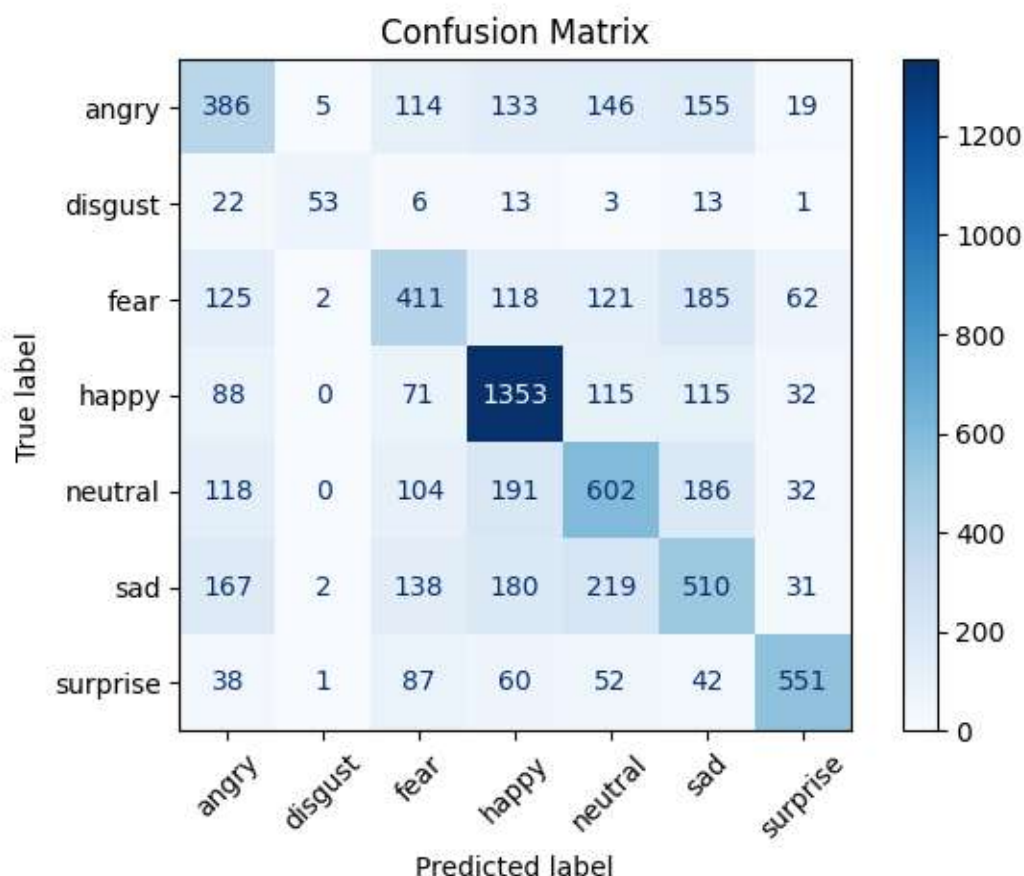
V. Kết quả và đánh giá

1. Bảng kết quả huấn luyện

Sau khi thực hiện quá trình huấn luyện và kiểm thử mô hình nhận diện cảm xúc, nhóm thu được các chỉ số đánh giá như sau: độ chính xác tổng thể (**Accuracy**) đạt **54%**; các chỉ số **Precision**, **Recall**, và **F1-score** được tính cho từng cảm xúc cụ thể. Trong đó, các cảm xúc như *happy* và *surprise* đạt kết quả cao nhất với F1-score khoảng **0.71**, trong khi các cảm xúc tiêu cực như *angry*, *fear*, và *sad* có F1-score dao động quanh mức **0.41–0.42**. Chi tiết kết quả được thể hiện trong bảng dưới đây.

	Precision	Recall	F1-score
Angry	0.41	0.40	0.41
Disgust	0.84	0.48	0.61
Fear	0.44	0.40	0.42
Happy	0.66	0.76	0.71
Neutral	0.48	0.49	0.48
Sad	0.42	0.41	0.42
Surprise	0.76	0.66	0.71
Accuracy	0.54	0.54	0.54
Macro avg	0.57	0.51	0.54
Weighted avg	0.54	0.54	0.54

2. Confusion matrix



Hình 10: Confusion matrix

Dựa trên ma trận nhầm lẫn, có thể nhận thấy mô hình phân loại cảm xúc hoạt động tốt nhất với các cảm xúc có biểu hiện khuôn mặt rõ ràng như **"happy"** và **"surprise"**, với số mẫu được dự đoán đúng lần lượt là **1353** và **551**, chiếm tỷ lệ cao so với tổng số mẫu thực tế. Điều này cho thấy mô hình học được các đặc trưng nổi bật của những biểu cảm tích cực như nụ cười hoặc ánh mắt ngạc nhiên.

Ngược lại, các cảm xúc tiêu cực như **"fear"**, **"sad"**, và **"angry"** có tỷ lệ nhận diện thấp, thường bị nhầm lẫn lẫn nhau. Ví dụ, cảm xúc *fear* có đến **185 mẫu bị dự đoán nhầm là "neutral"**, và *sad* có tới **167 mẫu bị nhầm là "angry"**. Điều này phản ánh sự tương đồng trong biểu hiện khuôn mặt giữa các trạng thái cảm xúc tiêu cực và cũng là một điểm yếu thường thấy trong các hệ thống nhận diện sử dụng đặc trưng cổ điển như HOG, LBP và Landmark.

Cảm xúc "**disgust**" là một trường hợp đặc biệt khi có rất ít mẫu đúng (53), nhưng lại có độ chính xác dự đoán (precision) khá cao, cho thấy mô hình chỉ dự đoán "disgust" khi thật sự chắc chắn – tuy nhiên do mất cân bằng dữ liệu, *recall* lại rất thấp.

3. Kết quả nhận diện



Hình 11: Kết quả và thời gian nhận diện



Hình 12: Kết quả và thời gian nhận diện

VI. Kết luận

Qua quá trình triển khai hệ thống nhận diện cảm xúc người dùng trong robot dịch vụ, nhóm đã thực hiện huấn luyện mô hình phân loại dựa trên các đặc trưng khuôn mặt (HOG, LBP, Landmark) kết hợp với thuật toán SVM. Dựa vào kết quả đánh giá bằng bảng thông số và ma trận nhầm lẫn (Confusion Matrix), có thể đưa ra một số nhận xét và kết luận như sau:

1. Hiệu suất tổng thể của hệ thống:

Mô hình đạt **độ chính xác tổng thể (Accuracy) là 54%**, với điểm **F1-score trung bình cộng (Macro avg) là 0.54** và **trung bình có trọng số (Weighted avg) cũng là 0.54**. Đây là kết quả khá tốt đối với một hệ thống phân loại nhiều lớp (7 cảm xúc) trong điều kiện đặc trưng không quá phức tạp và chỉ sử dụng các kỹ thuật cổ điển như HOG, LBP và Landmark.

2. Hiệu quả theo từng cảm xúc:

- Nhóm cảm xúc được nhận diện tốt nhất là **“happy”** và **“surprise”**, với **F1-score đạt 0.71**, nhờ vào các biểu cảm khuôn mặt rõ ràng, dễ phân biệt.
- Ngược lại, các cảm xúc như **“fear”**, **“sad”**, và **“angry”** có F1-score dao động khoảng 0.41–0.42, cho thấy mô hình gặp khó khăn trong việc phân biệt giữa các trạng thái cảm xúc tiêu cực – vốn có biểu hiện khuôn mặt khá tương đồng.
- **“Disgust”** là cảm xúc ít xuất hiện trong tập dữ liệu, dẫn đến sự mất cân bằng trong huấn luyện và Recall thấp (0.48), mặc dù Precision khá cao (0.84), cho thấy mô hình đoán đúng ít nhưng chắc chắn.

3. Phân tích từ ma trận nhầm lẫn:

- Nhiều trường hợp cảm xúc **“angry”**, **“fear”** và **“sad”** bị nhầm lẫn lẫn nhau, ví dụ: fear thường bị đoán nhầm thành neutral (185 lần) hoặc sad (62 lần).
- **“Happy”** có kết quả nhận diện tốt nhất với **1353 mẫu đúng**, cho thấy mô hình học được rõ nét đặc trưng của nụ cười.
- Các cảm xúc **“neutral”** và **“surprise”** cũng có tỷ lệ nhận diện khá cao với 602 và 551 mẫu đúng, tương ứng.

4. Đánh giá tổng thể:

Hệ thống có tiềm năng ứng dụng thực tế trong các robot dịch vụ nhờ khả năng nhận diện chính xác các cảm xúc rõ nét như vui và ngạc nhiên – vốn là các trạng thái phổ biến

trong tương tác dịch vụ. Tuy nhiên, để cải thiện nhận diện các cảm xúc khó phân biệt như sợ hãi, buồn, tức giận, cần tiếp tục:

- Tăng cường dữ liệu huấn luyện (cân bằng số lượng mẫu giữa các lớp).
- Áp dụng các mô hình học sâu (deep learning) như CNN để tăng độ trích xuất đặc trưng.
- Tối ưu hóa tham số và tích hợp thêm thông tin về chuyển động (video) hoặc giọng nói để đa kênh hóa cảm xúc.

VII. Tài liệu tham khảo

1. Pramod, O. (n.d.). *Support Vector Machine (SVM)*. Medium. Truy cập tại: <https://medium.com/@ompramod9921/support-vector-machine-svm-68bf201b4aa2>
2. TPoint Tech. (n.d.). *Machine Learning – Support Vector Machine Algorithm*. Truy cập tại: <https://www.tpointtech.com/machine-learning-support-vector-machine-algorithm>
3. YouTube – *SVM - Support Vector Machine Algorithm Explained*. Truy cập tại: <https://www.youtube.com/watch?v=RnapcMi9lqA>
4. YouTube – *Emotion Recognition using Machine Learning*. Truy cập tại: <https://www.youtube.com/watch?v=QdmUKu7LqtI>
5. YouTube – *Facial Emotion Recognition using Python and OpenCV*. Truy cập tại: <https://www.youtube.com/watch?v=hribPP5ccdc&t=538s>