



THỰC TẬP CNTT5: TRIỂN KHAI ỨNG DỤNG AI, IOT

ĐỀ TÀI: HỆ THỐNG TƯƠNG TÁC THÔNG MINH DỰA TRÊN
NHẬN DIỆN CẢM XÚC

Giảng viên hướng dẫn:

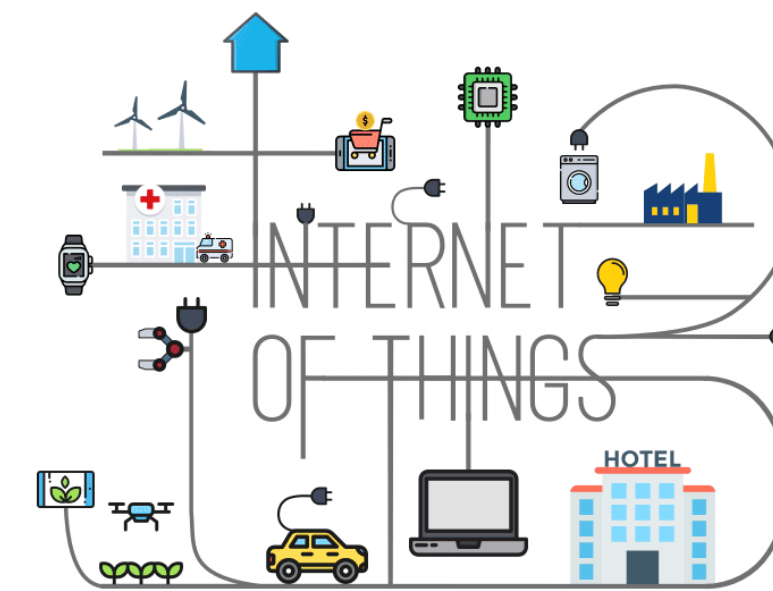
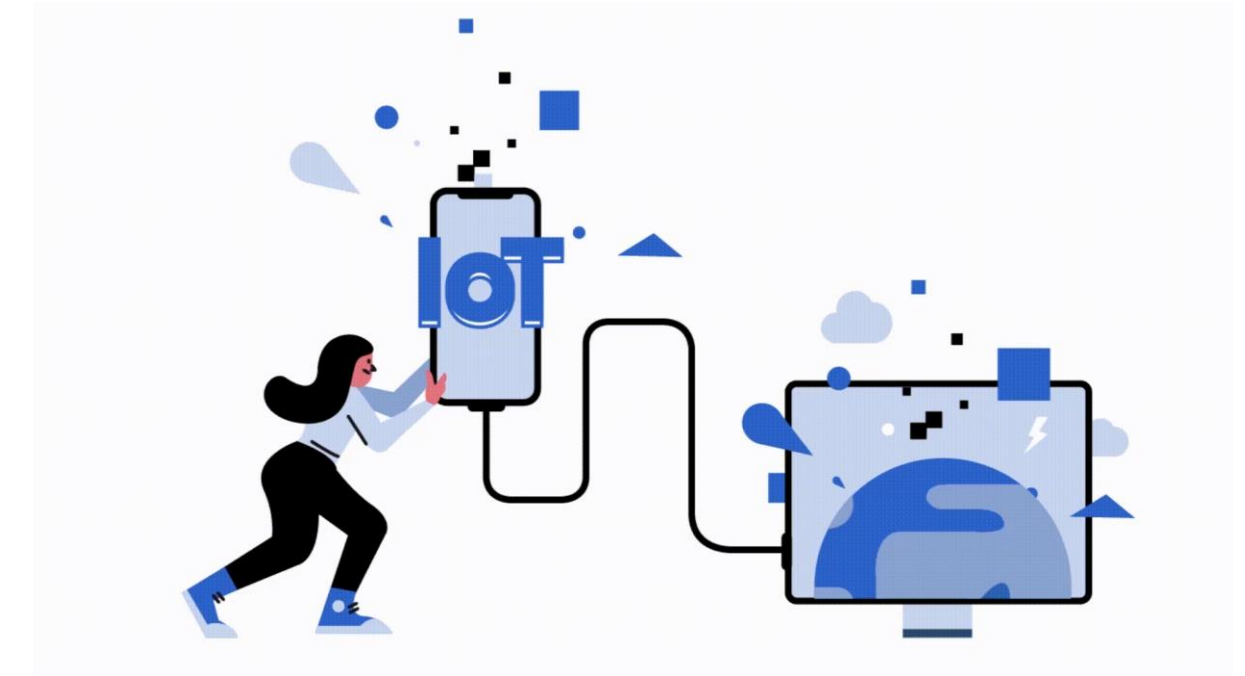
ThS. Lê Trung Hiếu
ThS. Nguyễn Thái Khánh

Trình bày:

Nhóm 11 – Lê Văn Việt

MỤC LỤC

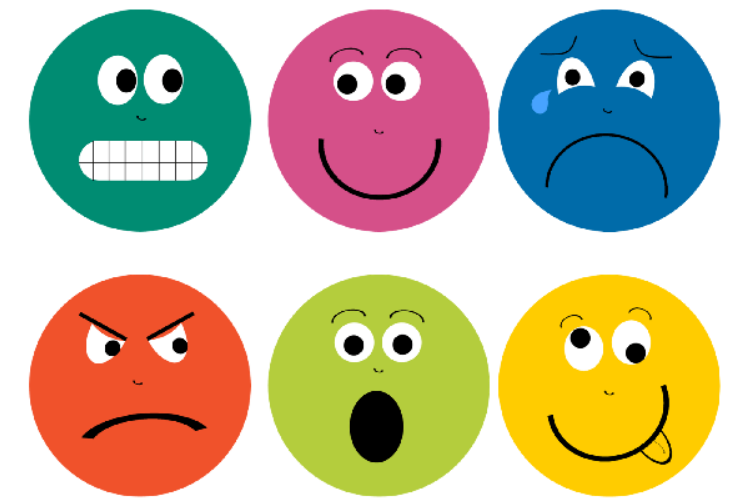
1. Giới thiệu/Đặt vấn đề
2. Mục tiêu và đề xuất
3. Nghiên cứu liên quan
4. Phương pháp triển khai
5. Sơ đồ hệ thống
6. Thiết kế và triển khai
7. Thực nghiệm và đánh giá
8. Ưu điểm và hạn chế
9. Kết luận và hướng phát triển



Smart IoT Solutions Shaping
the Business and Future

1. GIỚI THIỆU / ĐẶT VẤN ĐỀ

- Cảm xúc là một phần quan trọng trong giao tiếp giữa con người.
- Máy tính thông thường không có khả năng nhận biết cảm xúc.
- Ứng dụng AI-IOT xây dựng hệ thống có khả năng nhận diện cảm xúc qua khuôn mặt, tương tác, điều khiển thiết bị IoT theo ngữ cảnh cảm xúc.



2. MỤC TIÊU VÀ ĐỀ XUẤT

- Xây dựng hệ thống nhận diện cảm xúc khuôn mặt theo thời gian thực
- Kết hợp nhận diện cảm xúc với khả năng tương tác bằng giọng nói
- Điều khiển các thiết bị IoT dựa trên cảm xúc và lệnh thoại
- Tạo một môi trường thông minh thích ứng với trạng thái cảm xúc người dùng



3. NGHIÊN CỨU LIÊN QUAN

- FER (Facial Emotion Recognition): Phát triển từ phương pháp truyền thống (Haar Cascade) đến CNN

Github: <https://github.com/justinshenk/fer>

- DeepFace: Framework cho nhận diện khuôn mặt sâu, cung cấp các mô hình pre-trained

Github: <https://github.com/serengil/deepface>



3. NGHIÊN CỨU LIÊN QUAN

EmotiEffLib:

- Thư viện nhận diện cảm xúc khuôn mặt sử dụng Deep Learning, hỗ trợ Python và C++, tối ưu cho ứng dụng thời gian thực.
- Sử dụng các framework như PyTorch, ONNX Runtime để tối ưu hóa hiệu suất.



Github: <https://github.com/av-savchenko/EmotiEffLib>

4. PHƯƠNG PHÁP TRIỂN KHAI

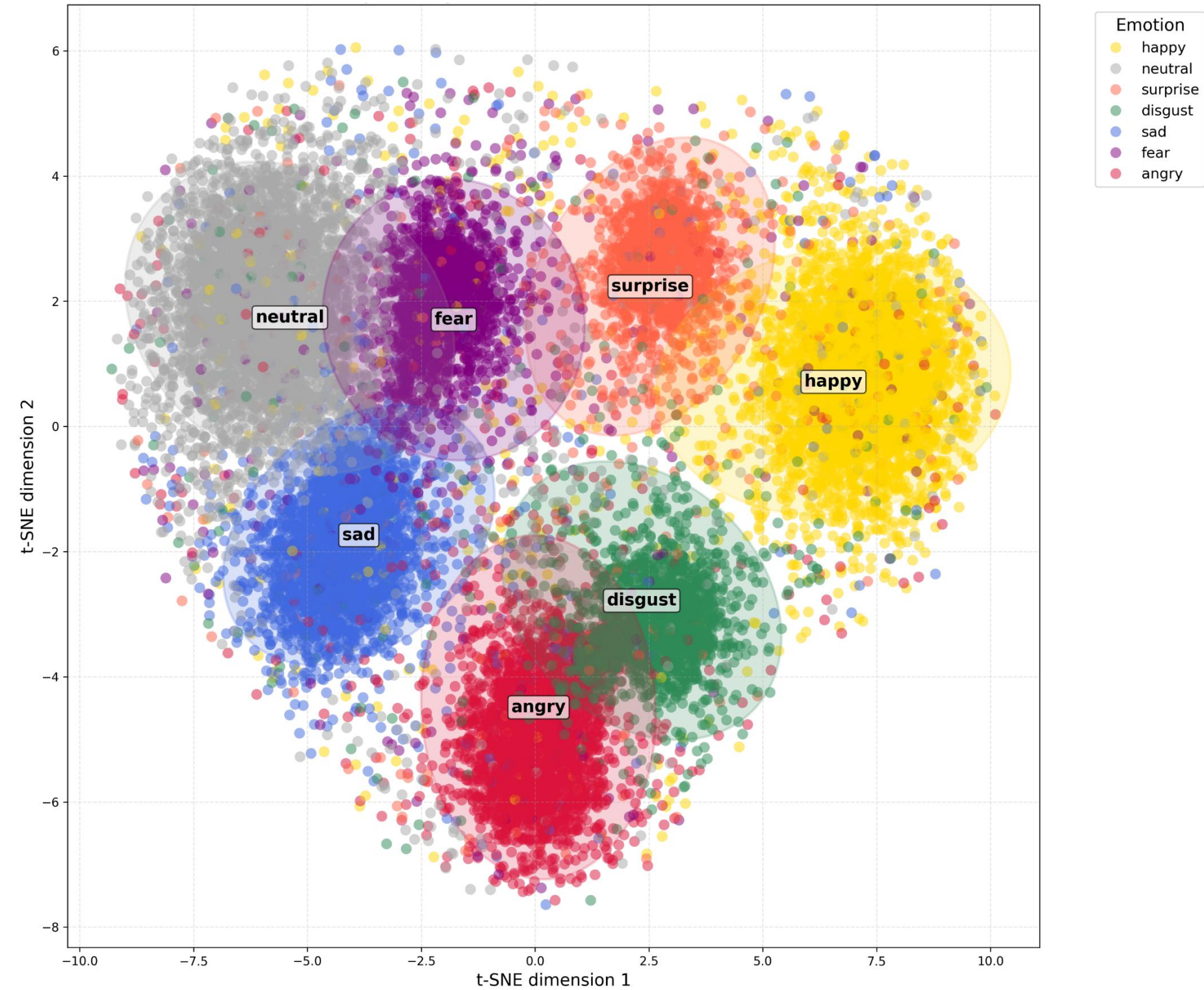
Dataset:

- FER2013: Bộ dữ liệu 35.887 ảnh – 7 cảm xúc, 48x48 pixel, ở dạng grayscale (thang độ xám)
- CK+: 593 chuỗi biểu hiện cảm xúc từ 123 đối tượng
- AFEW: Dữ liệu từ các đoạn phim.
- AffectNet: 400.000 ảnh được gán nhãn thủ công.
- KDEF: 4.900 ảnh cảm xúc từ 70 người.



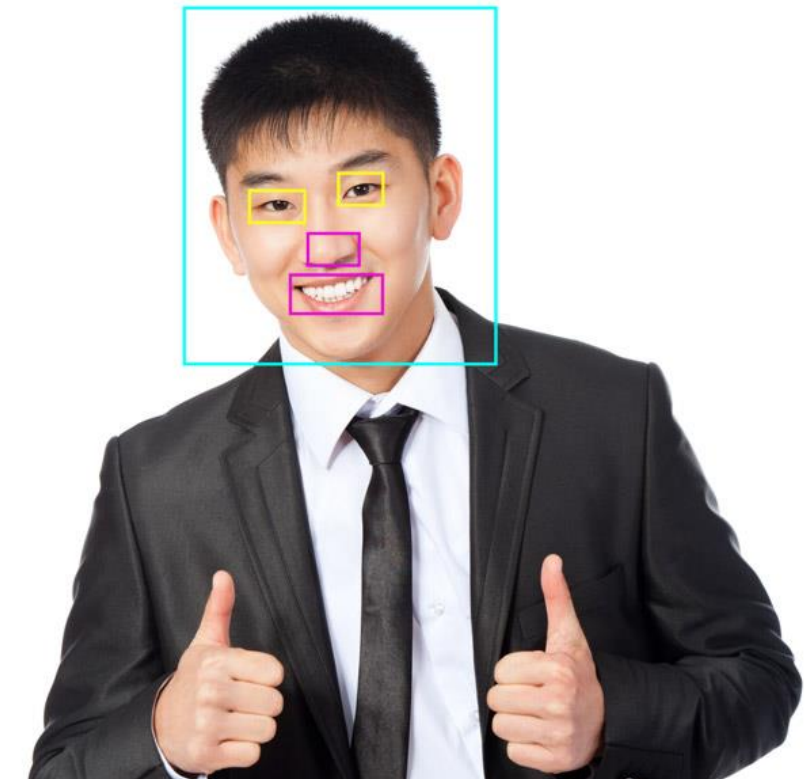
4. PHƯƠNG PHÁP TRIỂN KHAI

t-SNE:



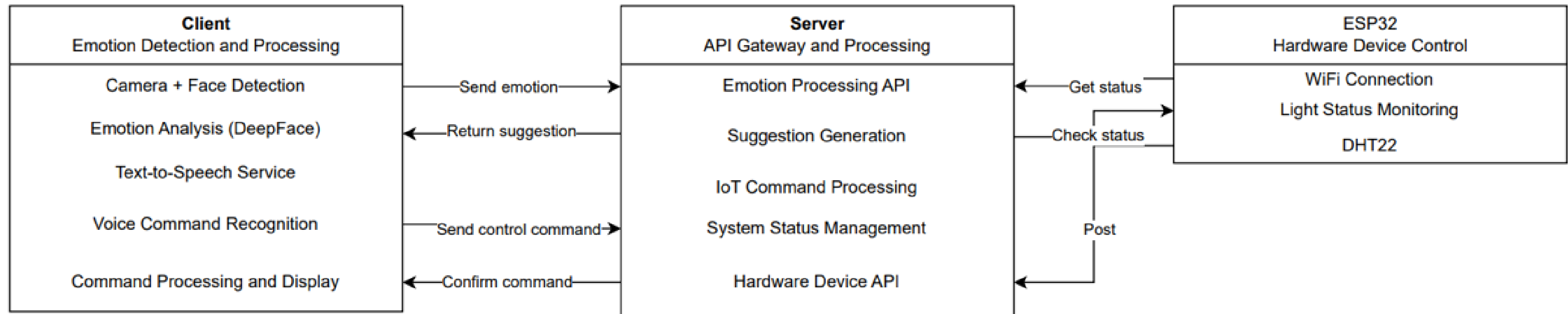
4. PHƯƠNG PHÁP TRIỂN KHAI

- Thu nhận hình ảnh từ webcam theo thời gian thực
- Phát hiện khuôn mặt và trích xuất vùng mặt
- Phân tích cảm xúc bằng DeepFace
- Đề xuất hành động dựa trên cảm xúc phát hiện
- Tương tác giọng nói nhận lệnh điều khiển
- Gửi lệnh điều khiển đến thiết bị IoT qua WiFi



5. SƠ ĐỒ HỆ THỐNG

Sơ đồ luồng hoạt động của hệ thống:



Emotion Detection → Generate Suggestion → Voice Command Recognition → IoT Device Control → ESP32 Status Update

6. THIẾT KẾ VÀ TRIỂN KHAI

Client System:

- Emotion Service: Sử dụng DeepFace và MTCNN để phân tích cảm xúc
- Voice Service: Xử lý lệnh giọng nói tiếng Việt

```
Đang phát hiện cảm xúc...  
Đã phát hiện cảm xúc: neutral (59.5%)  
Đang gửi cập nhật cảm xúc neutral (59.5%) đến server  
Phản hồi từ server: {'confidence': 59, 'emotion': 'neutral', 'message': 'Đã cập nhật  
cảm xúc: neutral (độ tin cậy: 59%)', 'status': 'success', 'suggestions': []}  
Đang phát âm (kết hợp): Phát hiện cảm xúc BÌNH THƯỜNG
```


6. THIẾT KẾ VÀ TRIỂN KHAI

Server System:

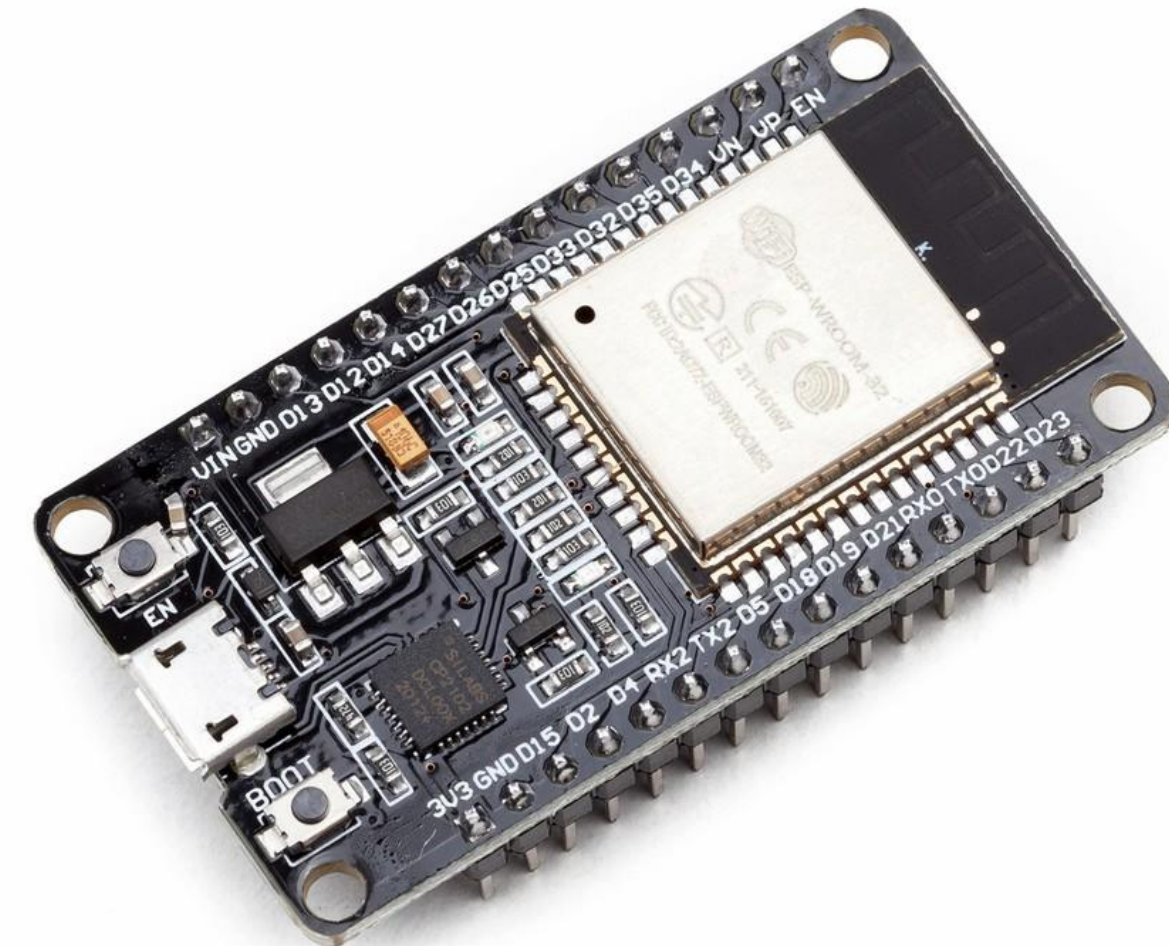
- Flask API Endpoints: Cung cấp các endpoint cho việc điều khiển thiết bị
- Emotion Handler: Xử lý và đề xuất dựa trên cảm xúc

```
Đã cập nhật cảm xúc: neutral (độ tin cậy: 59%)
172.20.10.3 - - [09/Mar/2025 16:30:52] "POST /emotion/update HTTP/1.1" 200 -
Đã cập nhật cảm xúc: happy (độ tin cậy: 96%)
172.20.10.3 - - [09/Mar/2025 16:31:56] "POST /emotion/update HTTP/1.1" 200 -
Đã cập nhật cảm xúc: fear (độ tin cậy: 95%)
172.20.10.3 - - [09/Mar/2025 16:32:01] "POST /emotion/update HTTP/1.1" 200 -
Đã BẬT đèn
172.20.10.3 - - [09/Mar/2025 16:32:11] "GET /light/on HTTP/1.1" 200 -
Đã cập nhật cảm xúc: fear (độ tin cậy: 53%)
172.20.10.3 - - [09/Mar/2025 16:32:13] "POST /emotion/update HTTP/1.1" 200 -
```

6. THIẾT KẾ VÀ TRIỂN KHAI

ESP32 IoT Device:

- Kết nối WiFi và HTTP Client
- Điều khiển đèn LED theo lệnh từ server
- DHT11/DHT22
- Đồng bộ trạng thái với server



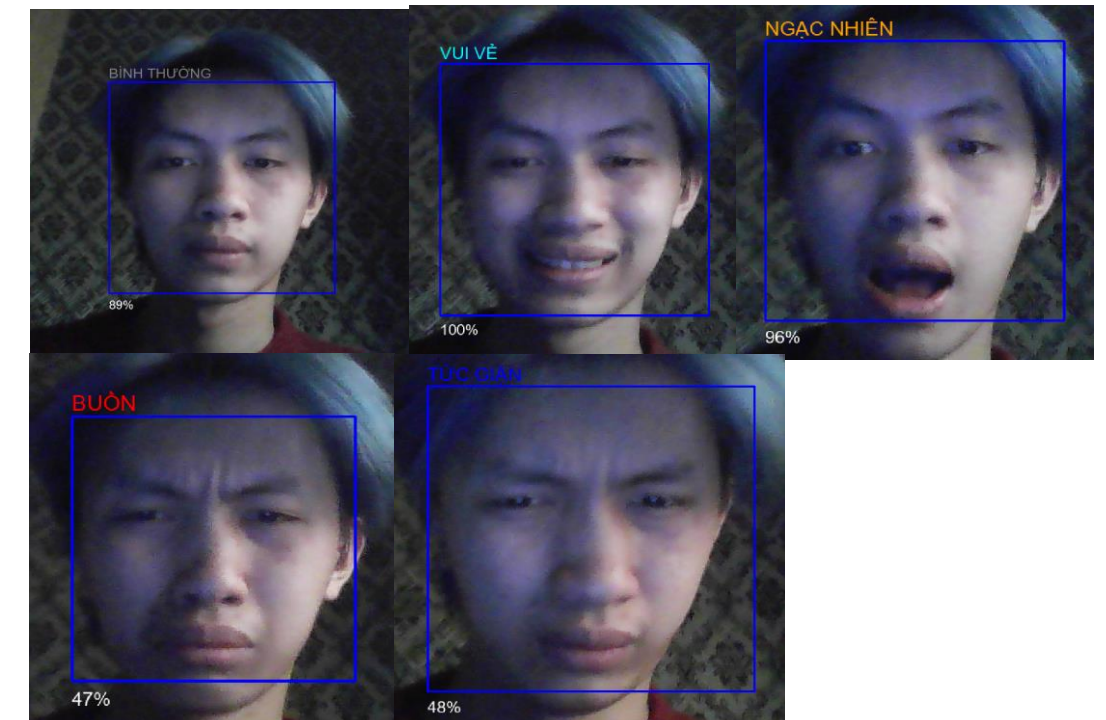
6. THIẾT KẾ VÀ TRIỂN KHAI

Dataset và Training:

- Sử dụng mô hình pre-trained DeepFace
- Mô hình nhận diện 7 cảm xúc: happy, sad, angry, neutral, surprise, fear, disgust

Giao diện:

- Hiển thị trực quan cảm xúc phát hiện
- Tương tác trực quan



7. THỰC NGHIỆM VÀ ĐÁNH GIÁ

Thực nghiệm hệ thống:

- Nhận diện cảm xúc: Test với các biểu cảm khác nhau
- Tương tác giọng nói: Test các lệnh điều khiển thiết bị
- Đề xuất dựa trên cảm xúc: Đánh giá độ phù hợp của đề xuất
- Điều khiển IoT: Kiểm tra độ trễ và độ tin cậy



7. THỰC NGHIỆM VÀ ĐÁNH GIÁ

Đánh giá:

- Độ chính xác nhận diện: ~85% trong điều kiện ánh sáng tốt (thử nghiệm trên tập test 100 ảnh).
- Độ nhận diện giọng nói: ~80% với các lệnh đơn giản
- Độ trễ hệ thống: ~500ms cho nhận diện cảm xúc, ~1-2s cho xử lý giọng nói.
- Tích hợp IoT: Hoạt động ổn định trong mạng LAN.

ĐANG LẮNG NGHE

Lệnh: bật đèn

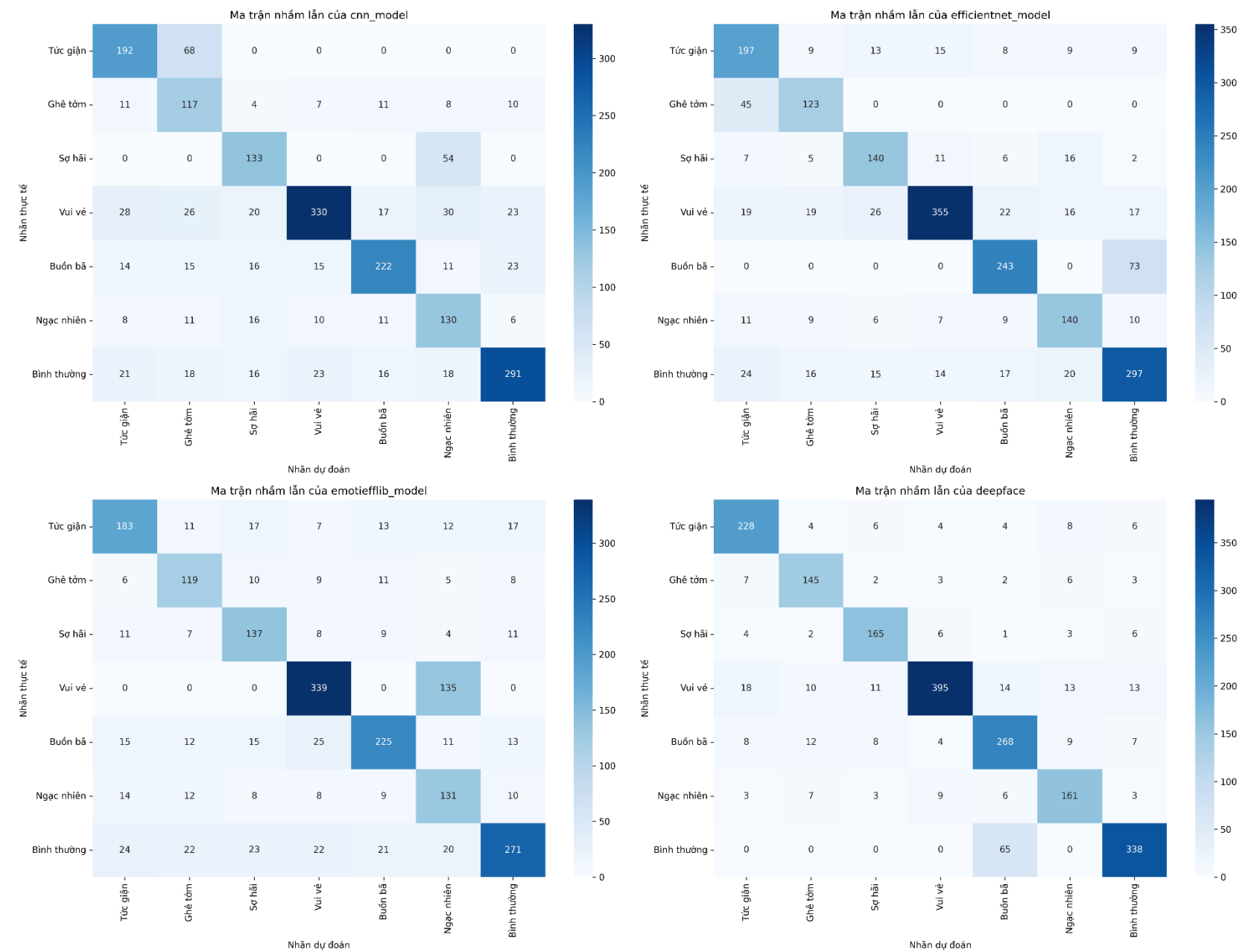
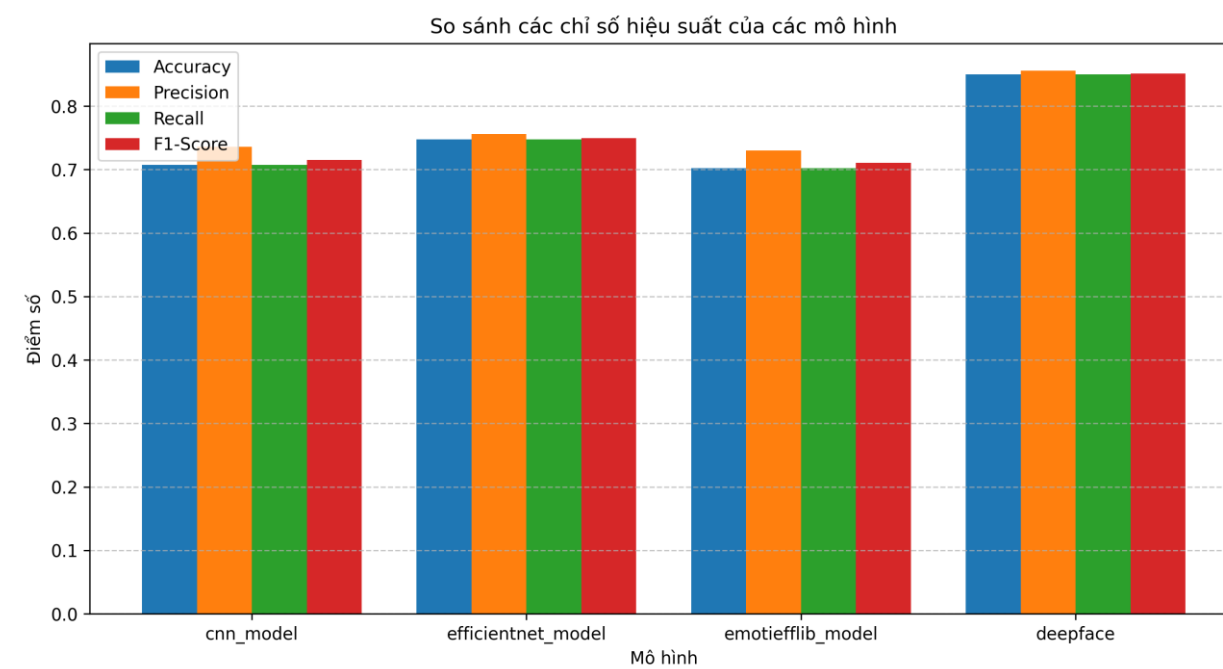
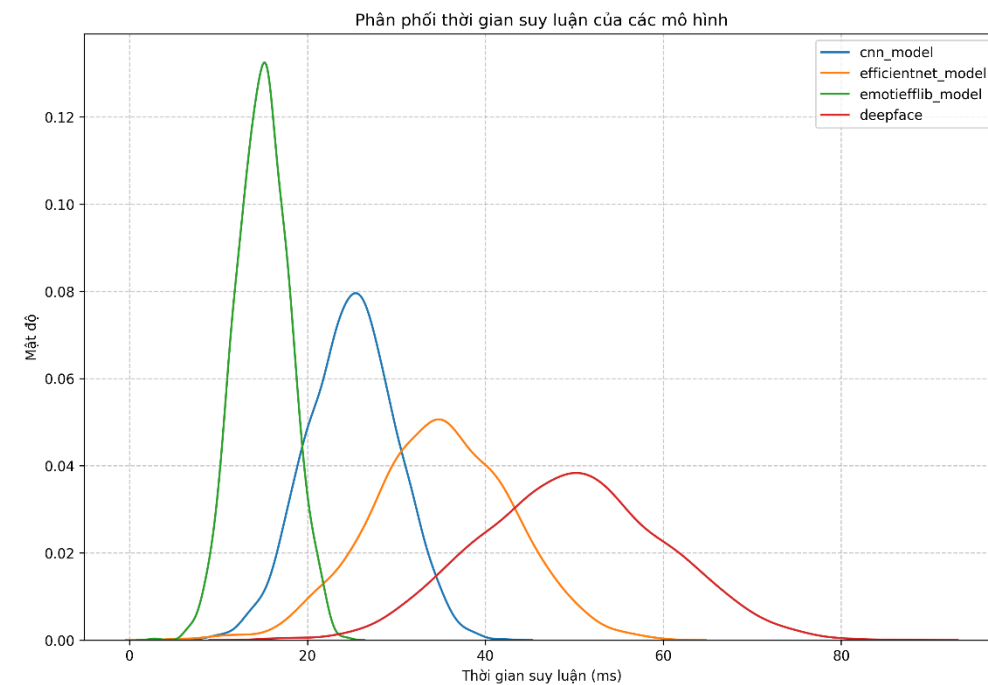
Kết quả: Đèn đã được bật

ĐỀ XUẤT:

1. Bạn có vẻ vui vẻ! Bạn có muốn bật đèn...
2. Bật nhạc vui để tiếp thêm năng lượng ...

7. THỰC NGHIỆM VÀ ĐÁNH GIÁ

Model:



7. THỰC NGHIỆM VÀ ĐÁNH GIÁ

So sánh mô hình:

<i>Model</i>	<i>Accuracy (%)</i>	<i>Precision (%)</i>	<i>Recall (%)</i>	<i>F1-score (%)</i>
<i>cnn_model</i>	70.75	74.59	70.75	71.53
<i>efficientnet_model</i>	74.75	75.58	74.75	74.96
<i>emotiefflib_model</i>	70.22	73.03	70.25	71.04
<i>deepface</i>	85.00	85.58	85.34	85.12

Nguồn: paperswithcode.com

8. ƯU ĐIỂM VÀ HẠN CHẾ

Ưu điểm:

- Nhận diện cảm xúc theo thời gian thực
- Tương tác tiếng Việt
- Mở rộng cao

Hạn chế:

- Phụ thuộc vào nhiều điều kiện (ánh sáng,..)
- Độ chính xác nhận diện giọng nói trong môi trường ồn kém
- Chức năng IoT cơ bản

9. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN



Kết luận:

- Đã xây dựng hệ thống nhận diện cảm xúc và tương tác người – máy
- Kết hợp hiệu quả computer vision, speech recognition và IoT

Hướng phát triển:

- Mở rộng với nhiều thiết bị IoT
- Mở rộng khả năng LLM cho hệ thống



Thank You