**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

**Cơ sở TP. HỒ CHÍ MINH**

**A logo of a company

Description automatically generated**

**BÁO CÁO BÀI TẬP LỚN**

**Môn học: Nhập môn trí tuệ nhân tạo**

**Giáo viên hướng dẫn: Nguyễn Thị Tuyết Hải**

**Đề tài: Ứng dụng học máy, học sâu để xây dựng chương trình nhận diện các đặc trưng dựa trên khuôn mặt con người**

**Tên thành viên MSSV**

**Phan Văn Thọ N22DCCN083**

**Lê Võ N22DCCN**

**Ngô Lê Hoài Trung N22DCCN**

**Trần Nhật Nguyên N22DCCN**

**I. Bài toán đặt ra**

**1. Giới thiệu đề tài**

Trong thời đại công nghệ số hiện nay, việc tích hợp trí tuệ nhân tạo (AI) vào các hệ thống giám sát và hỗ trợ con người ngày càng trở nên phổ biến. Một trong những ứng dụng thiết thực và tiềm năng là xây dựng hệ thống nhận diện trạng thái khuôn mặt nhằm phát hiện trạng thái mệt mỏi, cảm xúc, độ tuổi và giới tính của người sử dụng. Đề tài này tập trung vào việc phát triển một hệ thống thông minh có khả năng nhận diện trạng thái mắt (mở/nhắm), cảm xúc khuôn mặt, tuổi và giới tính của người dùng trong thời gian thực thông qua webcam.

Hệ thống sử dụng thư viện Mediapipe để trích xuất các điểm đặc trưng trên khuôn mặt, kết hợp với các mô hình học sâu đã được huấn luyện trước để dự đoán:

* Trạng thái mắt: Phát hiện người dùng có đang nhắm hay mở mắt, hỗ trợ các ứng dụng như giám sát sự tỉnh táo của người lái xe.
* Cảm xúc: Phân tích biểu cảm khuôn mặt để suy luận trạng thái cảm xúc (vui, buồn, giận, v.v.).
* Tuổi và giới tính: Ước lượng độ tuổi và xác định giới tính người dùng, hỗ trợ phân tích nhân khẩu học.

Dữ liệu đầu vào được thu nhận qua camera và xử lý trong thời gian thực với các luồng xử lý song song (multithreading) nhằm đảm bảo hiệu suất. Ngoài ra, hệ thống còn tích hợp chức năng phát âm bằng tiếng Anh để thông báo trạng thái mắt, giúp tăng tính tương tác và hỗ trợ người dùng không cần nhìn vào màn hình.

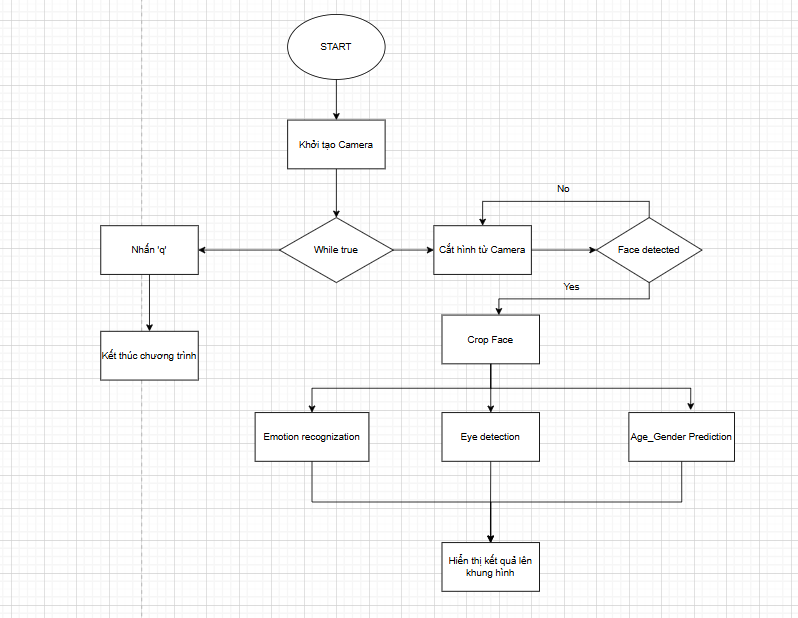
Đề tài mang lại giá trị thực tiễn cao trong các lĩnh vực như: an toàn giao thông, giám sát hành vi, phân tích dữ liệu người dùng, và có thể mở rộng thêm các chức năng nâng cao trong tương lai như nhận diện danh tính hoặc phân tích hành vi phức tạp hơn.

**2. Mục tiêu đề tài**

Mục tiêu chính của đề tài là xây dựng một hệ thống giám sát khuôn mặt trong thời gian thực, có khả năng nhận diện đồng thời nhiều đặc điểm quan trọng từ khuôn mặt người dùng, bao gồm:

1. Nhận diện trạng thái mắt (mở hoặc nhắm)
   * Phát hiện trạng thái nhắm mắt liên tục nhằm cảnh báo nguy cơ buồn ngủ hoặc mất tập trung, đặc biệt hữu ích trong các ứng dụng giám sát người lái xe hoặc nhân viên vận hành máy móc.
2. Nhận diện cảm xúc khuôn mặt
   * Phân tích biểu cảm để xác định cảm xúc hiện tại như vui, buồn, giận, sợ, ngạc nhiên… từ đó có thể ứng dụng trong các hệ thống tương tác người – máy hoặc chăm sóc sức khỏe tinh thần.
3. Dự đoán tuổi và giới tính
   * Ước lượng khoảng tuổi và giới tính của người dùng nhằm phục vụ cho mục đích phân tích nhân khẩu học hoặc cá nhân hóa dịch vụ.
4. Mở rộng thêm
   * Đảm bảo hiệu suất hệ thống trong quá trình xử lý đồng thời nhiều tác vụ (như dự đoán trạng thái mắt, cảm xúc, tuổi, giới tính và phát âm cảnh báo).
   * Tăng khả năng phản hồi và cảnh báo trực quan bằng âm thanh khi phát hiện người dùng đang nhắm mắt hoặc có biểu hiện mệt mỏi.

**II. Lưu đồ giải thuật**



**III. Các bước thực hiện**

1. **Khởi tạo camera, phát hiện khuôn mặt, cắt khuôn mặt, mắt(Thọ)**

**…**

1. **Nhận diện cảm xúc (Trung)**
   1. **Huấn luyện mô hình**

Mô hình nhận diện cảm xúc được huấn luyện trên dữ liệu ảnh khuôn mặt với nhãn cảm xúc tương ứng. Bộ dữ liệu điển hình là **FER2013**, bao gồm:

* Khoảng 1.**000 ảnh khuôn mặt** ở độ phân giải thấp.
* Mỗi ảnh có kích thước **48x48 pixels**, ảnh **màu xám (grayscale)**.
* 7 lớp cảm xúc bao gồm:
* Angry
* Disgust
* Fear
* Happy
* Neutral
* Sad
* Surprise

Trước khi đưa vào mô hình, ảnh được **chuẩn hóa về giá trị pixel trong khoảng [0, 1]**, và reshape về kích thước (48, 48, 1) để phù hợp với đầu vào của mạng nơ-ron tích chập (CNN).

* 1. **Dự đoán**

Trong quá trình chạy chương trình thời gian thực:

* Khuôn mặt được cắt từ khung hình video sử dụng thư viện **Mediapipe**.
* Ảnh khuôn mặt được chuyển thành **màu xám**, resize về **48x48**, chuẩn hóa và reshape đúng định dạng đầu vào.
* Mô hình sẽ dự đoán xác suất thuộc về mỗi lớp cảm xúc, và kết quả là nhãn có xác suất cao nhất.

1. **Nhận diện giới tính, đoán tuổi (Nguyên, Võ)**
   1. **Huấn luyện mô hình**

**…**

* 1. **Nhận diện**

**…**

1. **Nhận diện trạng thái mắt (Thọ)**
   1. **Huấn luyện mô hình**
   2. **Nhận diện**
2. **Nhận kết quả từ các mô hình và hiển thị lên màn hình (Thọ)**
3. **Kết thúc chương trình (Thọ)**