**TRƯỜNG ĐẠI HỌC PHENIKAA**

**KHOA CÔNG NGHỆ THÔNG TIN**

⸎⸎⸎⸎⸎



**TIỂU LUẬN GIỮA HỌC PHẦN**

**THỊ GIÁC MÁY TÍNH**

Đề bài: “Xây dựng ứng dụng nhận diện chữ số viết tay”

| **Nhóm 6:** | **Họ và tên** | **Mã sinh viên** |
| --- | --- | --- |
|  | Trương Hương Giang | 21010631 |
|  | Lưu Hiểu Khánh | 21012878 |
|  | Ngô Thị Linh | 21011613 |

**Giảng viên hướng dẫn:** TS. Phạm Tiến Lâm

Hà Nội, tháng 6 năm 2024

| **Họ và tên** | **Công việc** | **Điểm** |
| --- | --- | --- |
| Trương Hương Giang | - Tạo Model  - Tạo mã nguồn Flask App  - Huấn luyện mô hình | 9 |
| Lưu Hiểu Khánh | - Chỉnh sửa Model  - Viết báo cáo | 9 |
| Ngô Thị Linh | - Tạo website  - Viết báo cáo | 9 |

*Bảng 1. Kế hoạch phân công*

**1. MÔ TẢ ỨNG DỤNG**

**1.1. Tổng quan**

Ứng dụng nhận diện chữ số viết tay là một hệ thống web sử dụng Python Flask và mô hình học sâu CNN (Convolutional Neural Network) để nhận diện chữ số viết tay từ hình ảnh. Người dùng chỉ cần tải lên ảnh chứa các chữ số viết tay, và hệ thống sẽ tự động xử lý ảnh để nhận diện các chữ số và hiển thị chuỗi số tương ứng.

**1.2. Các thành phần của ứng dụng**

- Mô hình học sâu CNN được sử dụng để nhận diện các chữ số viết tay. Mô hình này được huấn luyện trên tập dữ liệu MNIST, bao gồm 60.000 ảnh chữ số viết tay cho tập huấn luyện và 10.000 ảnh cho tập kiểm tra. Sau khi huấn luyện, mô hình đạt được độ chính xác cao trên cả tập huấn luyện và kiểm tra, lần lượt là ~99% và ~98%.

- Ứng dụng web Flask được xây dựng để cung cấp giao diện người dùng cho phép tải lên ảnh, tương tác với mô hình CNN và hiển thị kết quả nhận diện. Ứng dụng web này được xây dựng đơn giản, dễ sử dụng và hỗ trợ nhiều tính năng web.

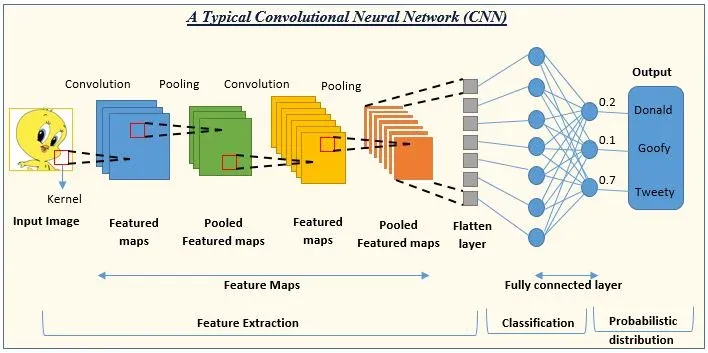
- Kỹ thuật xử lý ảnh được sử dụng để tiền xử lý ảnh, tách riêng từng chữ số, chuẩn hóa ảnh và xử lý kết quả nhận diện.

**2. MÔ TẢ THUẬT TOÁN HỌC MÁY CONVOLUTIONAL NEURAL NETWORK (CNN)**

**2.1. Giới thiệu**

CNN là từ viết tắt của cụm Convolutional Neural Network hay là mạng nơ ron tích chập. Mạng nơ ron này cho phép người dùng xây dựng những hệ thống phân loại và dự đoán với độ chính xác cực cao. Hiện nay, mạng CNN được ứng dụng nhiều hơn trong xử lý ảnh, cụ thể là nhận diện đối tượng trong ảnh.

**2.2. Các lớp cơ bản của CNN**

****

- Convolution Layer (Lớp tích chập): Là lớp đầu tiên trong mô hình CNN, có nhiệm vụ trích xuất các đặc trưng từ hình ảnh đầu vào.

- Stride (Bước nhảy): Xác định số pixel mà kernel di chuyển sau mỗi lần tính toán tích chập.

- Padding (Đường viền): Sử dụng để xử lý vấn đề kernel không phù hợp với kích thước ảnh đầu vào. Có hai phương pháp phổ biến là Zero padding và Valid padding.

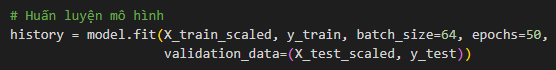
- ReLU (Hàm phi tuyến): Là hàm kích hoạt phổ biến trong mạng nơ-ron nhân tạo. Hàm ReLU có giá trị đầu ra là f(x) = max(0, x), nghĩa là nếu giá trị đầu vào nhỏ hơn 0 thì sẽ được gán bằng 0.

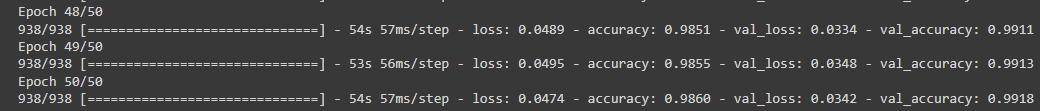
- Pooling Layer (Lớp gộp): Giúp giảm kích thước của ảnh đầu ra sau khi thực hiện tích chập, giảm số lượng tham số trong mô hình. Có 3 phương pháp phổ biến nhất là:

* Max Pooling: Lấy giá trị pixel lớn nhất trong một vùng pooling.
* Average Pooling: Lấy giá trị trung bình của các pixel trong một vùng pooling.
* Sum Pooling: Cộng các giá trị pixel trong một vùng pooling.

**3. HUẤN LUYỆN MÔ HÌNH**

**3.1. Kết quả huấn luyện**





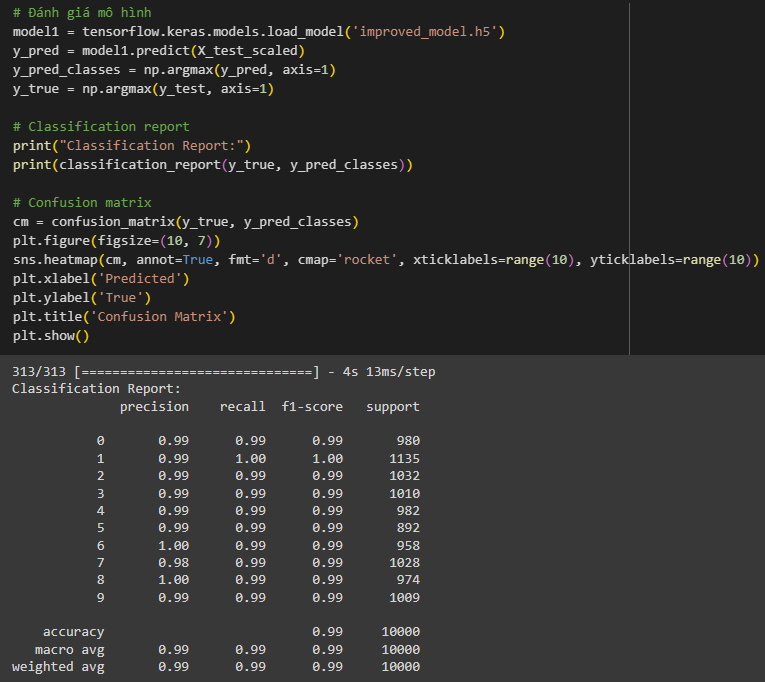
- Số epoch: 50

- Batch size: 64

- Độ chính xác trên tập huấn luyện: ~99%

- Độ chính xác trên tập kiểm tra: ~98%

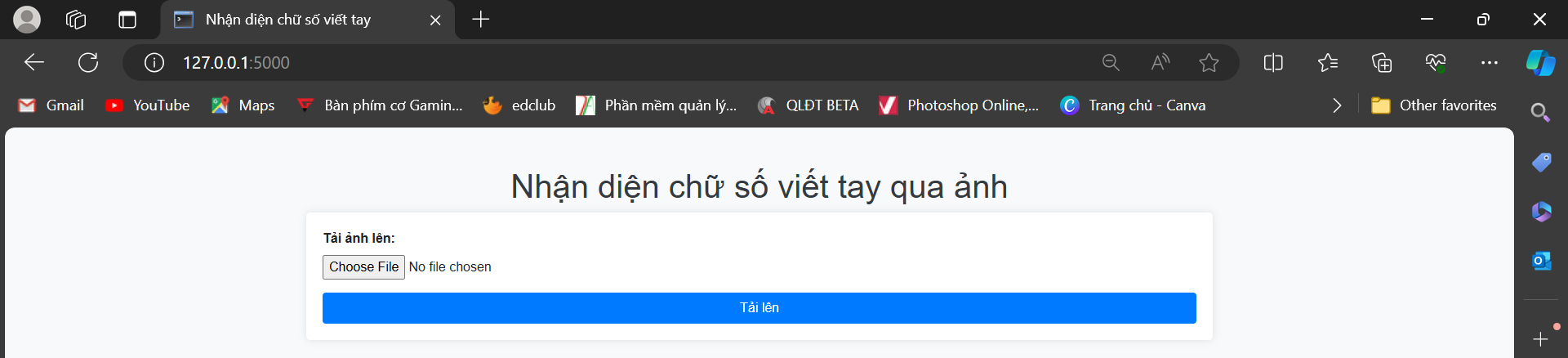
**3.2. Đánh giá Mô hình**

****

**4. MÔ TẢ ỨNG DỤNG WEB**

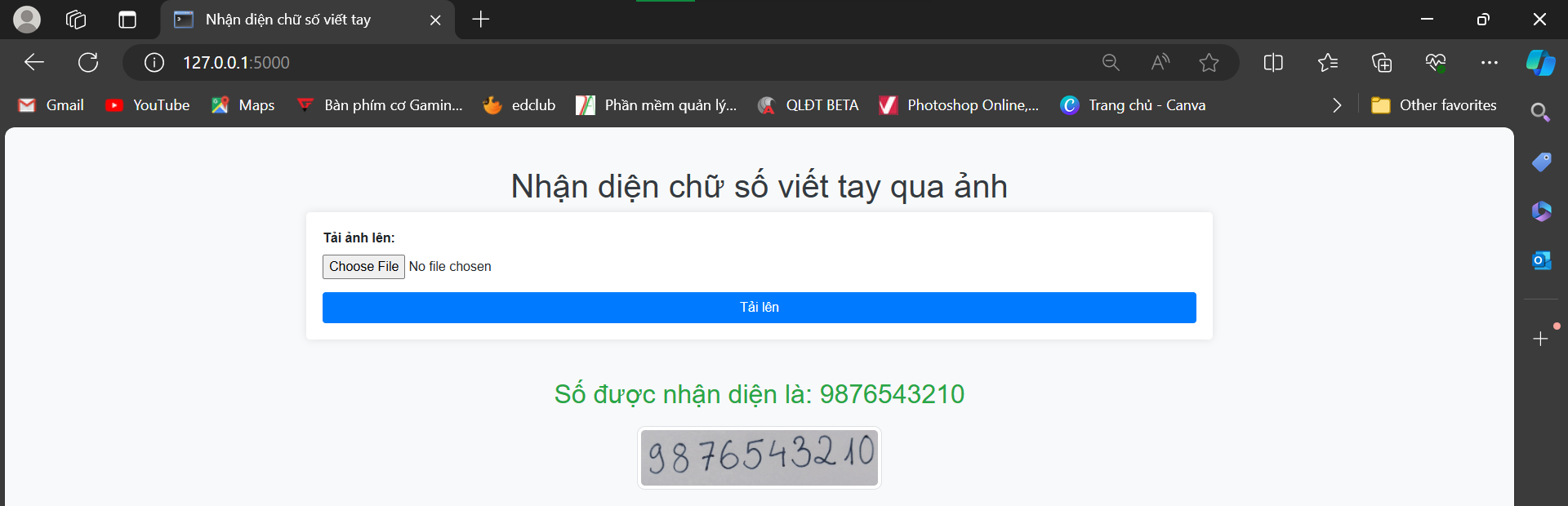
**4.1. Ứng dụng Web**

Ứng dụng web cho phép người dùng tải lên một ảnh chứa hàng chữ số viết tay (ví dụ: số điện thoại). Sau khi tải ảnh lên, hệ thống sẽ xử lý ảnh để tách riêng từng chữ số, dự đoán từng chữ số bằng mô hình CNN, và hiển thị chuỗi số dự đoán trên giao diện web.



*Hình 1. Hình ảnh minh họa ứng dụng Web*

Người dùng truy cập trang chủ và được yêu cầu tải lên một ảnh chứa các chữ số viết tay.



*Hình 2. Hình ảnh kết quả*

Sau khi người dùng tải ảnh lên và nhấn nút "Predict", hệ thống sẽ hiển thị kết quả nhận diện trong một ô textbox. Tuy nhiên độ chính xác của mô hình này cho ra khoảng 80-90%.

**2. Điều kiện của ảnh được tải lên**

- Ảnh cần có nền trắng và chữ số màu đen để hệ thống dễ dàng xử lý và nhận diện.

- Các chữ số trong ảnh nên được viết cách nhau một khoảng đủ để hệ thống tách biệt từng chữ số một cách chính xác.

- Định dạng ảnh: PNG, JPG.