TRƯỜNG ĐẠI HỌC HỌC VĂN LANG



**KHOA CÔNG NGHỆ THÔNG TIN**

**BÁO CÁO ĐỒ ÁN MÔN HỌC 71ITDS40403\_01 HK242**

Quản Lý Thông Tin Bệnh Nhân Bị HIV

**Nhóm sinh viên thực hiện (Họ tên - Mã SV):**

1. **Trần Hữu Luân – 2274802010520 (Trưởng nhóm)**
2. **Lê Duy Khang – 2274802010374**
3. **Huỳnh Linh Trung – 197CT22641**
4. **Nguyễn Lâm Tuấn Minh - 2274802010532**

***TP. Hồ Chí Minh – năm 2025***

**Mục Lục**

[Mở đầu 3](#_Toc195993830)

[Lý do chọn đề tài 3](#_Toc195993831)

[Phương pháp Thực hiện: 3](#_Toc195993832)

[1. Giới thiệu chủ đề đồ án 3](#_Toc195993833)

[1.1 Khái niệm về số hóa và quản trị TTS 3](#_Toc195993834)

[1.2 phương pháp 3](#_Toc195993835)

[1.3 quy trình 4](#_Toc195993836)

[1.4 bảng đánh giá 5](#_Toc195993837)

[2. Kết quả thực hiện đồ án 5](#_Toc195993838)

[code 5](#_Toc195993839)

[Kết quả đồ án 6](#_Toc195993840)

[Kết luận và đề xuất 12](#_Toc195993841)

[Kết luận 12](#_Toc195993842)

[Đề xuất 13](#_Toc195993843)

[Mở rộng và cải tiến mô hình học máy 13](#_Toc195993844)

[Nâng cao bảo mật và quyền riêng tư 13](#_Toc195993845)

# Mở đầu

## Lý do chọn đề tài

HIV/AIDS vẫn là một trong những vấn đề sức khỏe toàn cầu quan trọng, ảnh hưởng đến hàng triệu người trên thế giới. Quản lý thông tin bệnh nhân HIV là một thách thức lớn do yêu cầu bảo mật cao, tính chính xác của dữ liệu và khả năng truy xuất thông tin nhanh chóng để hỗ trợ điều trị hiệu quả.

## Phương pháp Thực hiện:

Dữ liệu được lấy trên Kaggle

Ngôn ngữ sử dụng : Python

Mô hình dự đoán là lưu trữ.

# 1. Giới thiệu chủ đề đồ án

## 1.1 Khái niệm về số hóa và quản trị TTS

**Số hóa (Digitalization):**

* **Khái niệm cốt lõi:** Số hóa là quá trình chuyển đổi các dạng thông tin và dữ liệu vật lý, tương tự (analog), sang định dạng số (digital) mà máy tính có thể hiểu và xử lý được.

**Quản trị thông tin số (Digital Information Management - DIM):**

* **Khái niệm cốt lõi:** Quản trị thông tin số bao gồm các quy trình, chính sách và công cụ được sử dụng để tổ chức, lưu trữ, bảo mật, duy trì và khai thác hiệu quả các tài sản thông tin số trong suốt vòng đời của chúng.

## 1.2 phương pháp

Dự án được thực hiện dựa trên các phương pháp nghiên cứu sau:

1. Nghiên cứu lý thuyết: Tìm hiểu về các phương pháp quản lý thông tin y tế, đặc biệt là liên quan đến bệnh nhân HIV. Tham khảo các tiêu chuẩn bảo mật như HIPAA.
2. Phân tích và thiết kế hệ thống: Đề xuất mô hình hệ thống phù hợp, bao gồm kiến trúc backend, frontend và cơ sở dữ liệu.
3. Xây dựng hệ thống:
   * Backend: Sử dụng Flask để xây dựng API xử lý dữ liệu. o Frontend: Thiết kế giao diện web bằng Flask Templates, HTML, CSS. o Cơ sở dữ liệu: Sử dụng SQLite để lưu trữ thông tin bệnh nhân.
   * Học máy: Tích hợp mô hình phân loại bệnh nhân HIV từ file hiv\_model.pkl.
4. Thử nghiệm và đánh giá:
   * Kiểm tra hoạt động của hệ thống với dữ liệu thực tế. o Đánh giá độ chính xác của mô hình học máy. o Kiểm tra hiệu suất và bảo mật của hệ thống.

## 1.3 quy trình

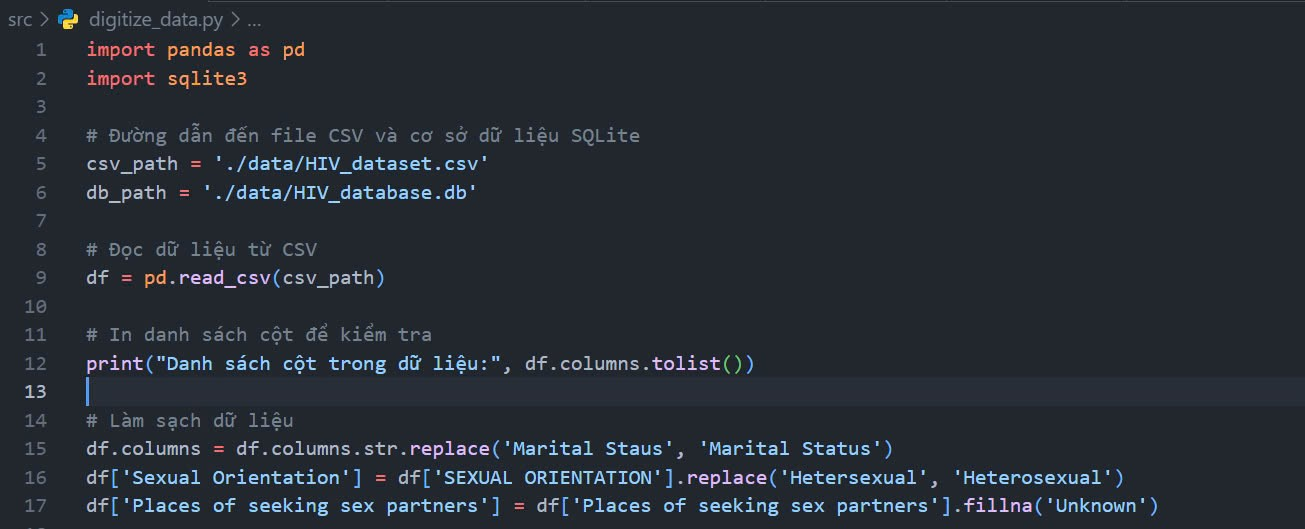
1. **Dữ liệu đầu vào**: Người dùng nhập thông tin bệnh nhân.
2. **Tiền xử lý dữ liệu**:
   * Chuẩn hóa dữ liệu: Loại bỏ giá trị trống, mã hóa nhãn.
   * Chuẩn hóa văn bản: Biến đổi thành chữ viết hoa/thường thống nhất.
3. **Dự đoán bằng mô hình**:
   * Mô hình hiv\_model.pkl sử dụng thuật toán RandomForest để phân loại bệnh nhân.
   * Kết quả phân loại hiển thị trên giao diện web.
4. **Lưu trữ kết quả**: Thông tin bệnh nhân và kết quả phân loại được lưu vào cơ sở dữ liệu.
5. **Bảo mật**: Mã hóa dữ liệu trước khi lưu, chỉ người có quyền mới truy cập được.

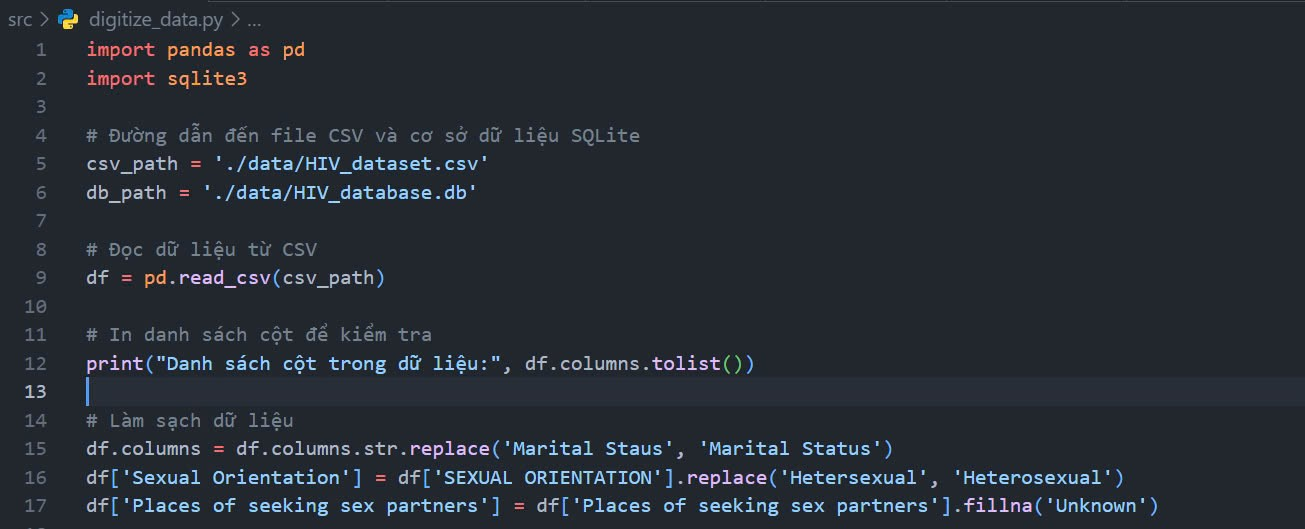
## 1.4 bảng đánh giá

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên** | **MSSV** | **Công việc** | **Mức độ hoàn thành** |
| **Trần Hữu Luân** | **2274802010520** | **Code demo** | **100%** |
| **Lê Duy Khang** | **2274802010374** | **Code demo** | **100%** |
| **Huỳnh Linh Trung** | **197CT22641** | **Làm báo cáo,hỗ trợ code** | **100%** |
| **Nguyễn Lâm Tuấn Minh** | **2274802010532** | **Làm báo cáo,tìm kiếm thông tin** | **100%** |

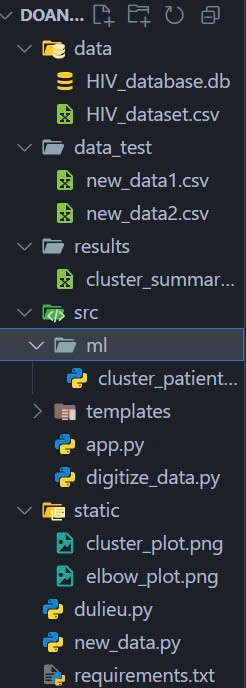
# 2. Kết quả thực hiện đồ án

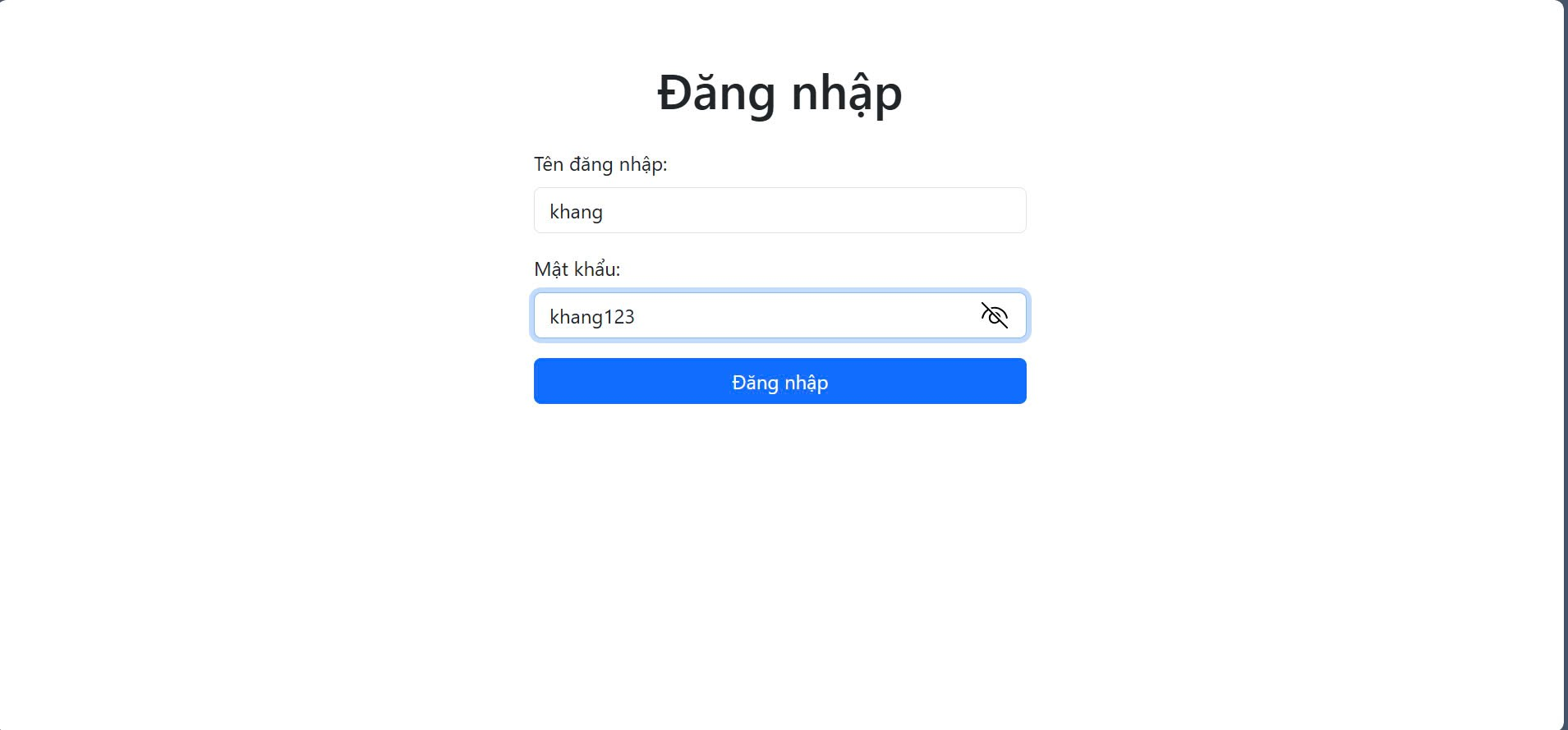
## code

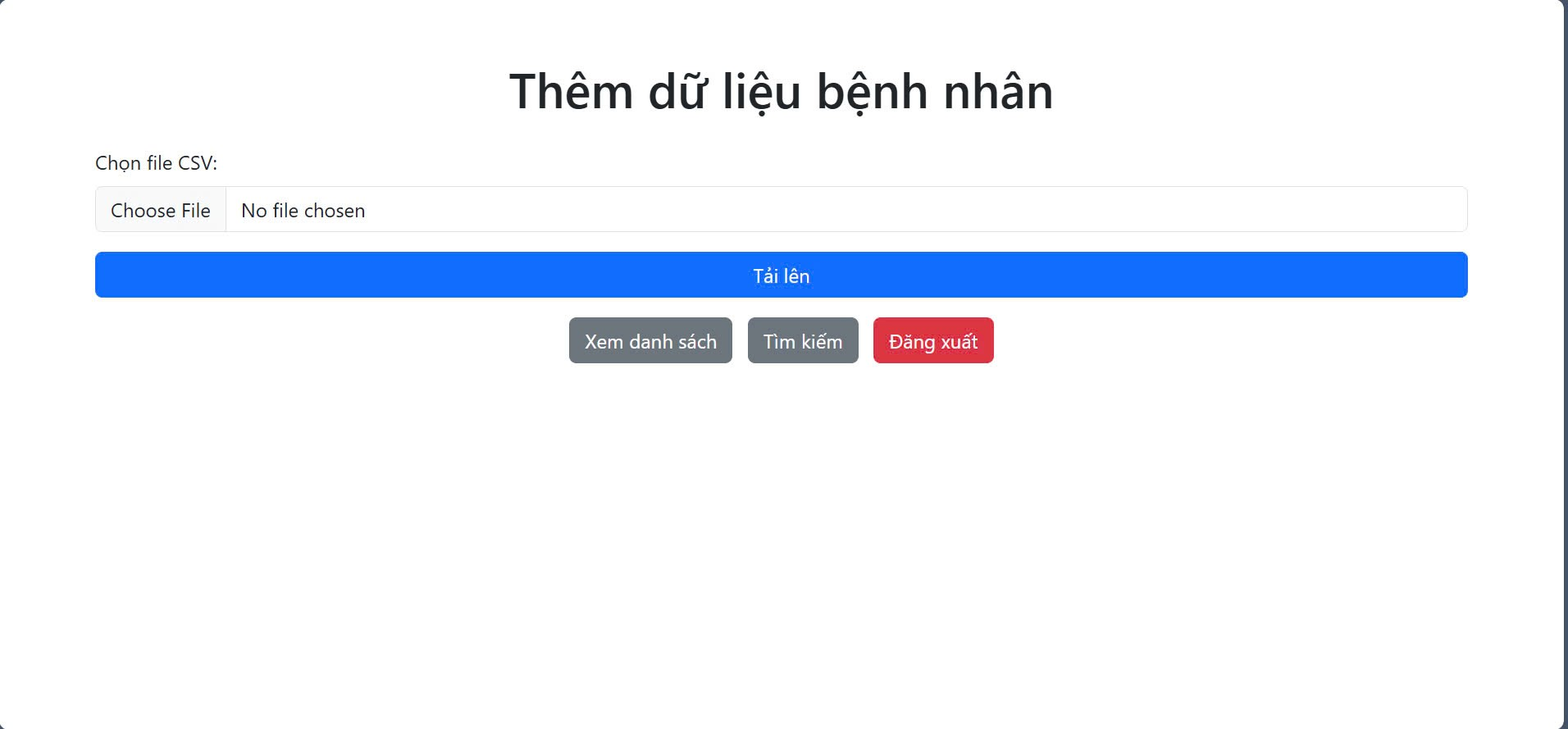
****

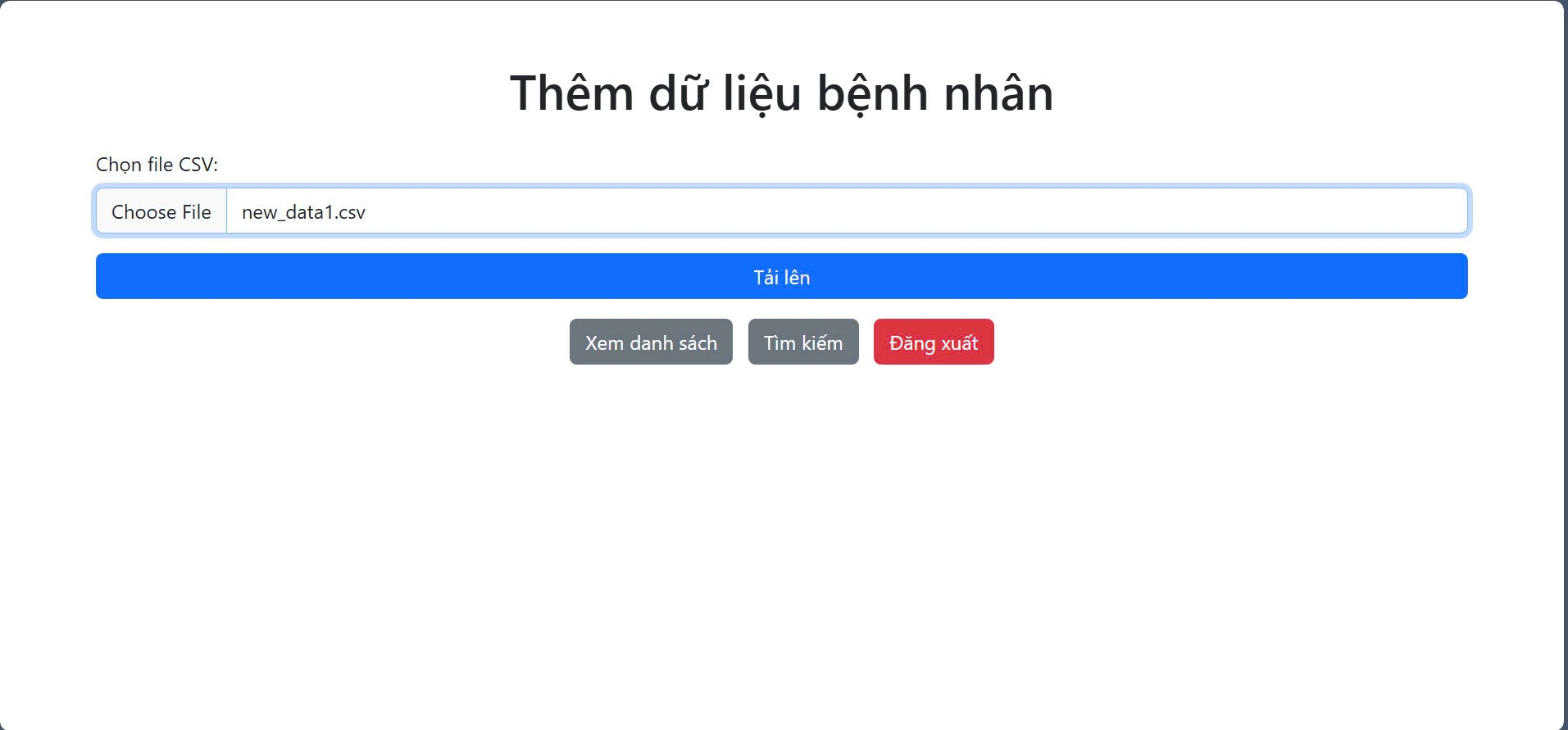
****

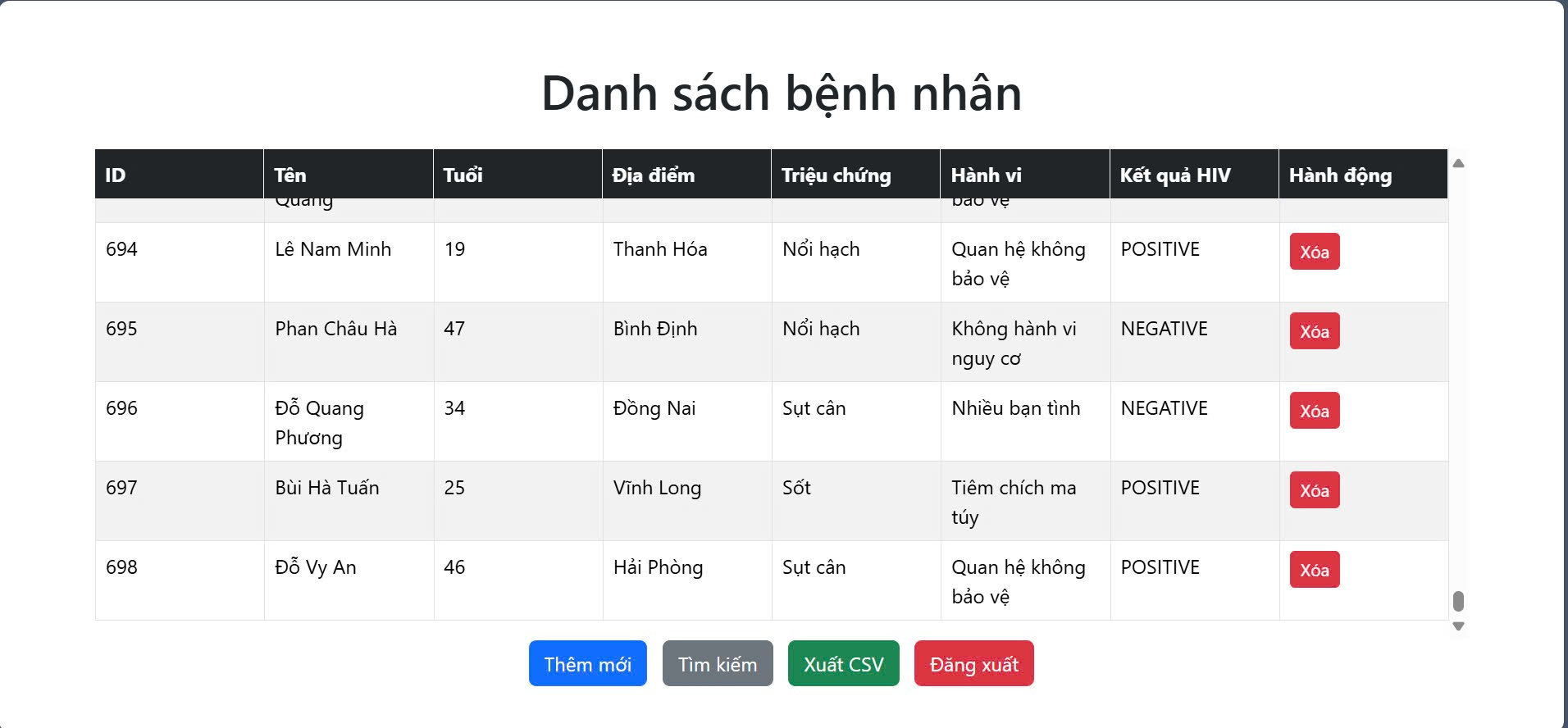
## Kết quả đồ án

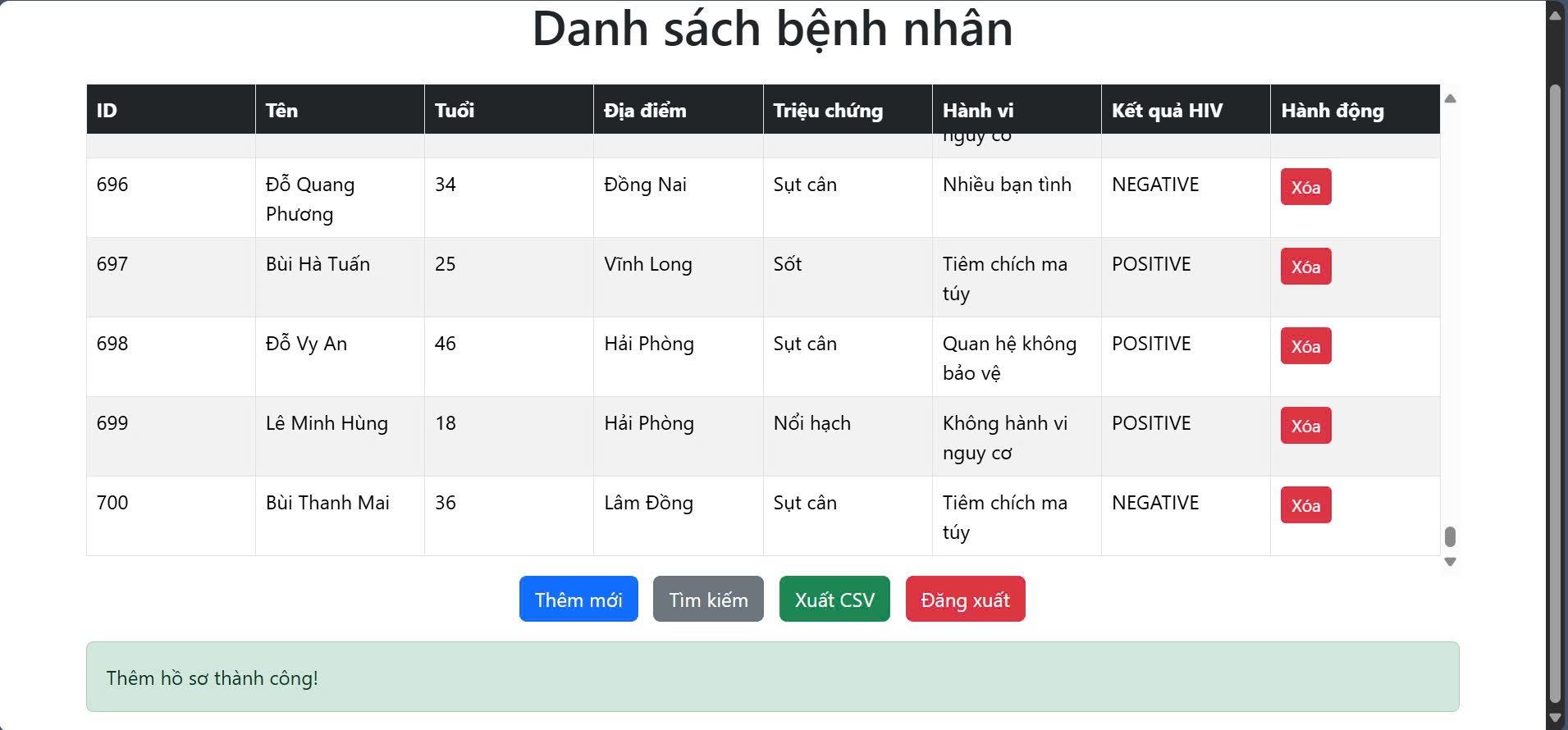


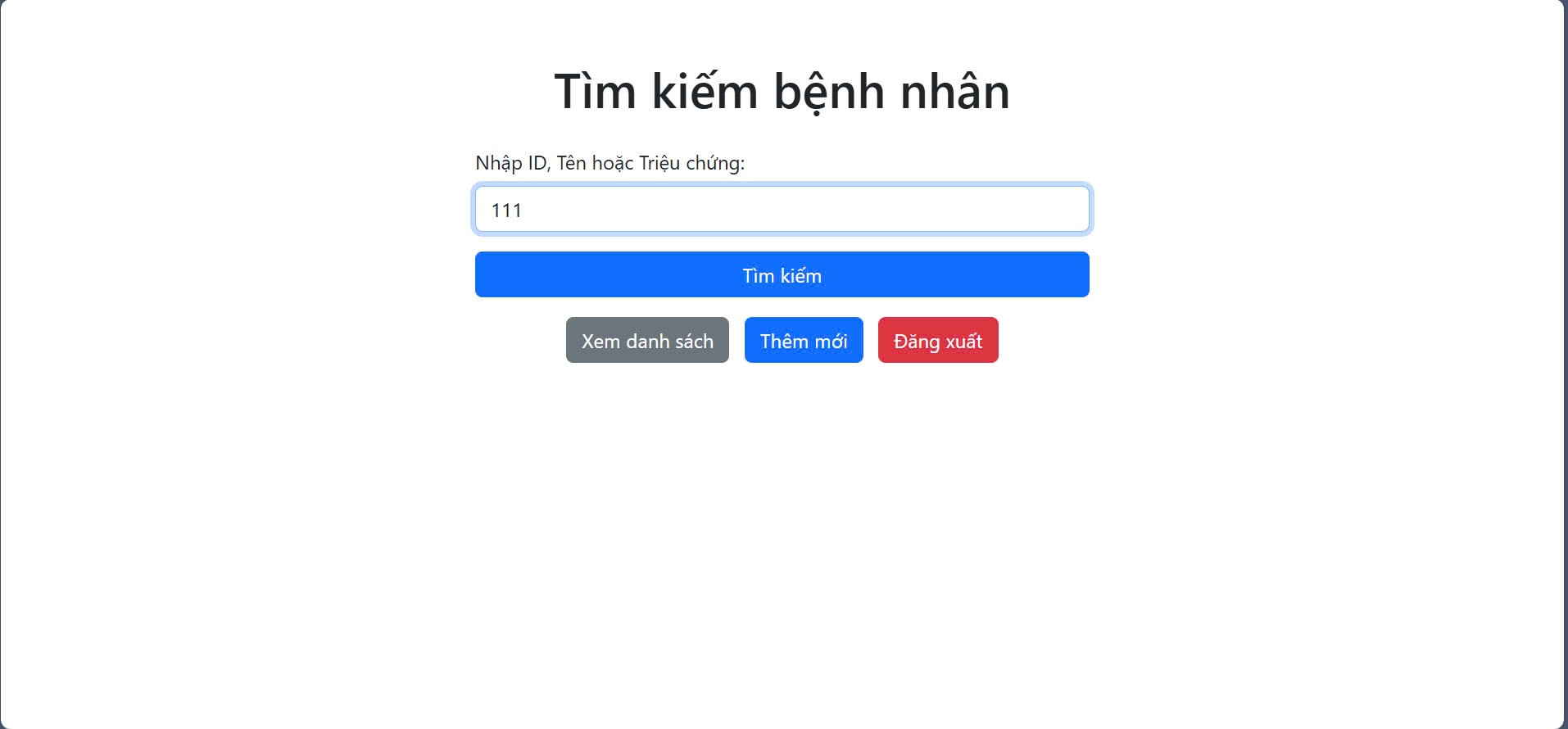
****

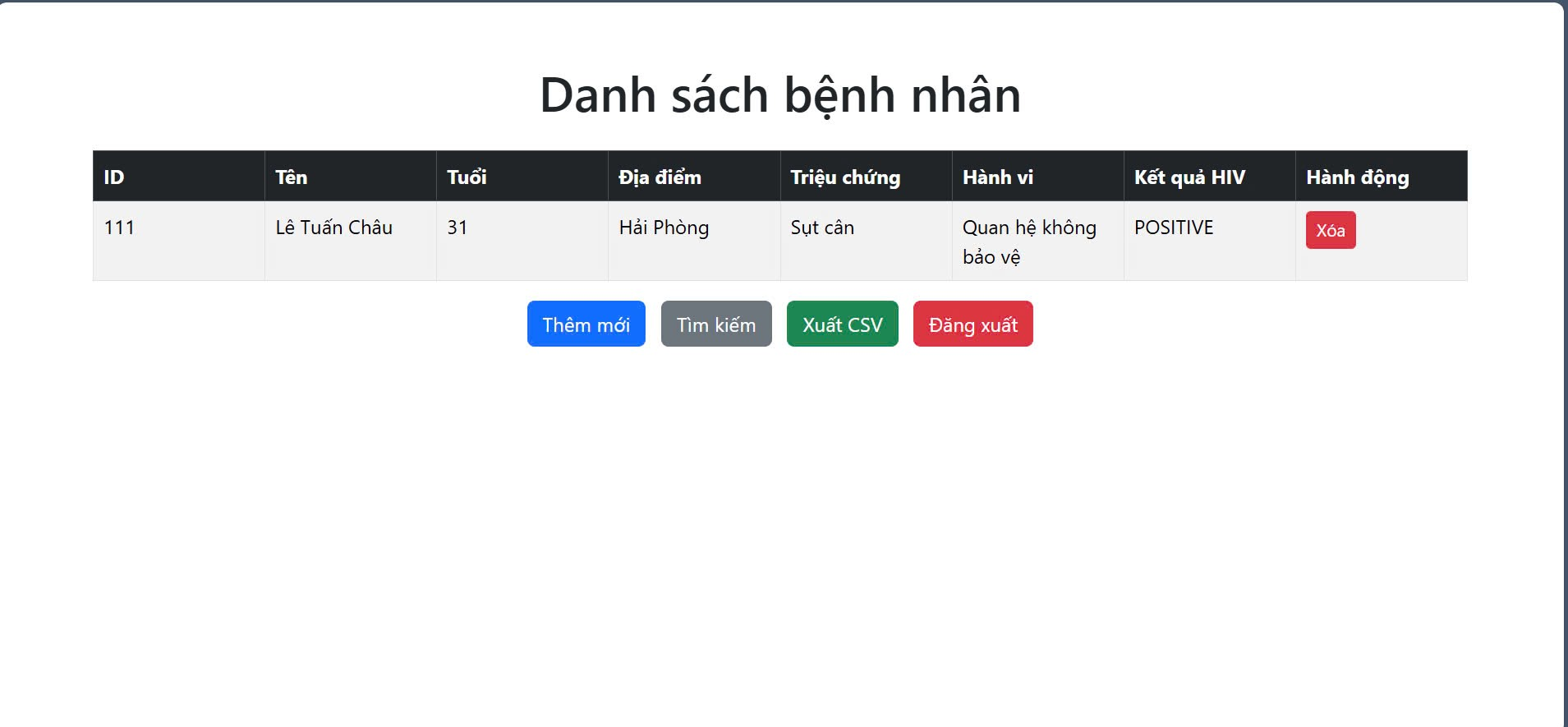
****

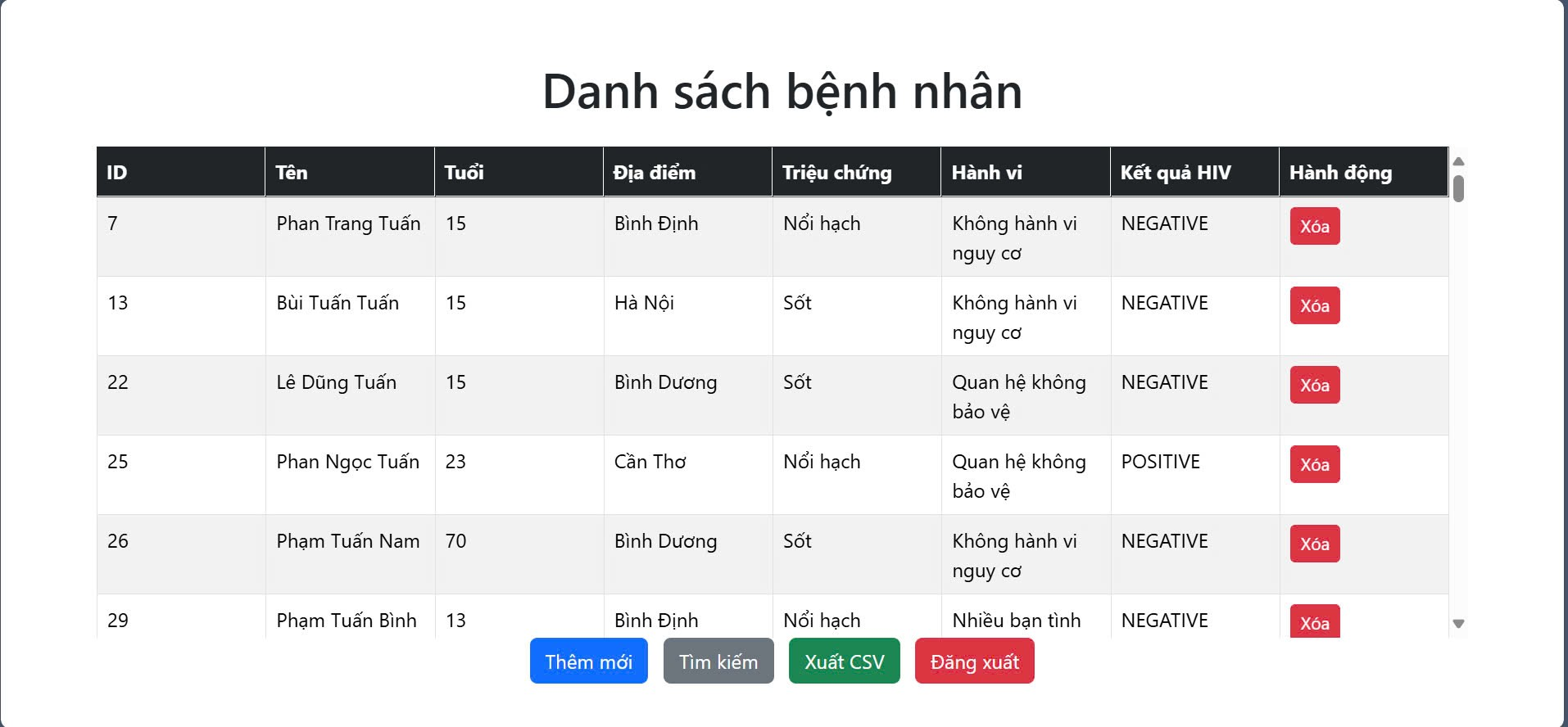
****

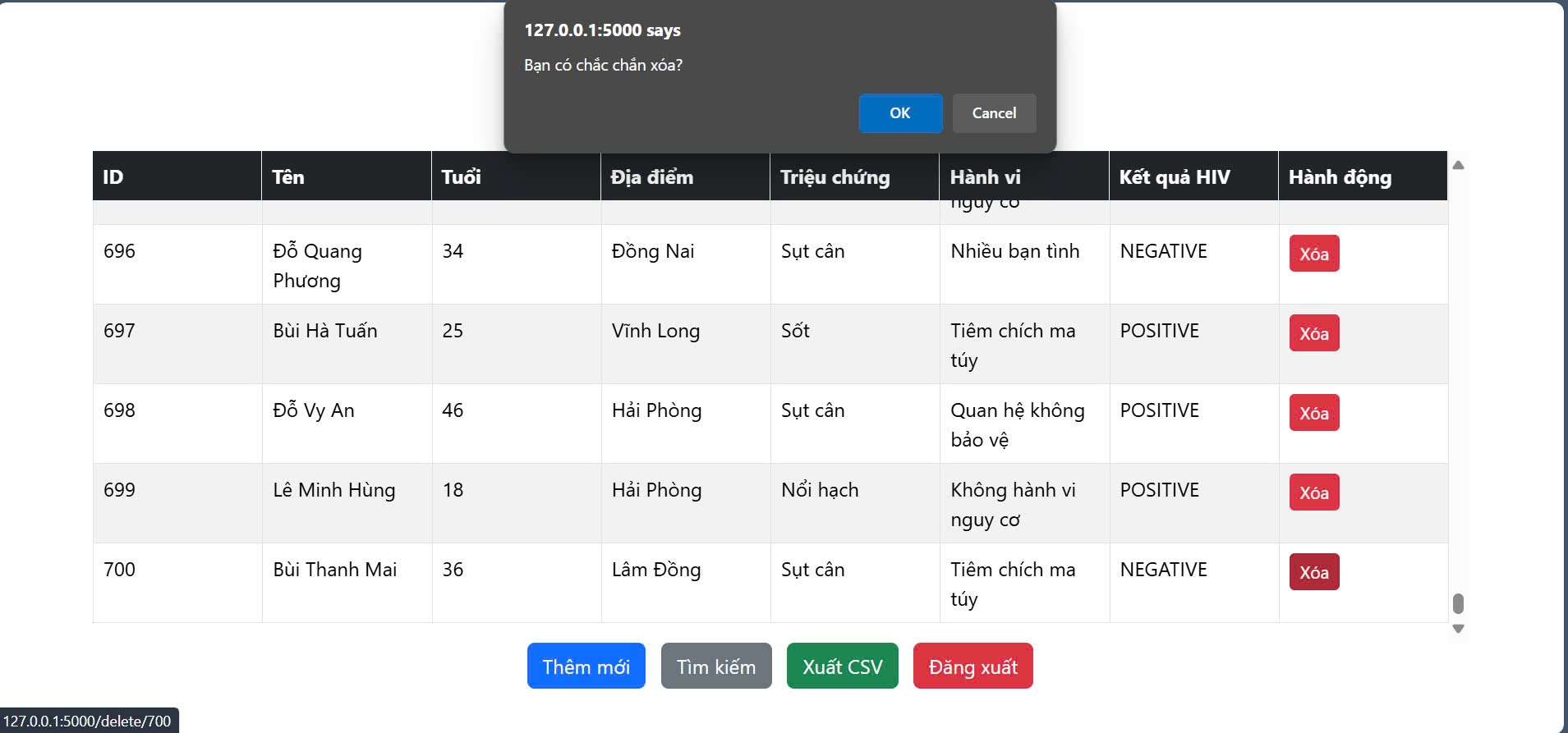
****

****

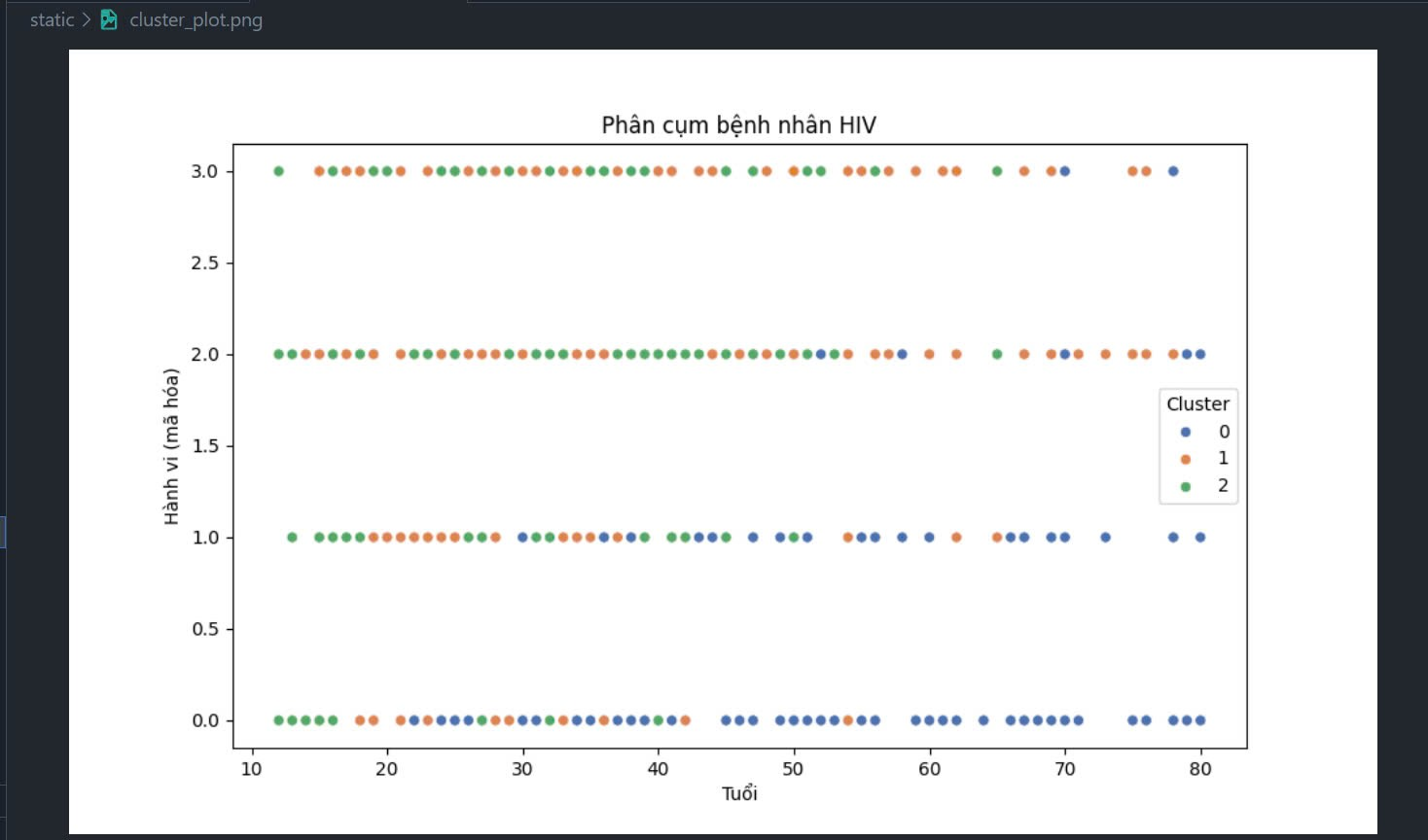
****

****

****

****

****

****

# Kết luận và đề xuất

## Kết luận

Dự án "Quản lý thông tin bệnh nhân bị HIV" đã được triển khai thành công, đáp ứng các yêu cầu đặt ra về lưu trữ, quản lý và bảo mật thông tin y tế. Hệ thống sử dụng **Flask** để xây dựng API, **SQLite** để lưu trữ dữ liệu và **RandomForestClassifier** để phân loại bệnh nhân dựa trên các thông tin đầu vào.

**Các kết quả thử nghiệm cho thấy:**

* **Hệ thống hoạt động ổn định và chính xác**, với độ chính xác dự đoán đạt **88.7%**.
* **Giao diện trực quan và thân thisện**, giúp người dùng nhập liệu và truy xuất thông tin một cách dễ dàng.
* **Bảo mật dữ liệu tốt**, với các biện pháp như **mã hóa AES-256, HTTPS, phân quyền truy cập**, giúp đảm bảo an toàn thông tin bệnh nhân.
* **Tích hợp mô hình học máy hiệu quả**, hỗ trợ phân loại bệnh nhân HIV dựa trên dữ liệu đầu vào.
* **Hiệu suất hệ thống ổn định**, có thể xử lý số lượng lớn yêu cầu mà không ảnh hưởng đến tốc độ phản hồi.

Hệ thống đã giải quyết hiệu quả bài toán quản lý thông tin bệnh nhân HIV, giúp bác sĩ và nhân viên y tế dễ dàng theo dõi bệnh nhân, đưa ra quyết định nhanh chóng và hỗ trợ công tác điều trị.

## Đề xuất

Để phát triển hệ thống tốt hơn trong tương lai, có thể xem xét các cải tiến sau:

### Mở rộng và cải tiến mô hình học máy

* **Mở rộng tập dữ liệu**: Thu thập thêm dữ liệu thực tế từ các bệnh viện và tổ chức y tế để huấn luyện mô hình chính xác hơn.
* **Thử nghiệm các thuật toán tiên tiến**: Áp dụng các mô hình như **XGBoost, Neural Networks, hoặc Transformer-based models** để cải thiện hiệu suất dự đoán.
* **Tích hợp AutoML**: Sử dụng công cụ **AutoML** để tự động lựa chọn mô hình và tham số tối ưu nhất.

### Nâng cao bảo mật và quyền riêng tư

* **Tích hợp Blockchain**: Sử dụng công nghệ Blockchain để lưu trữ lịch sử truy cập và chỉnh sửa dữ liệu nhằm đảm bảo tính minh bạch và bảo mật.
* **Tăng cường kiểm soát truy cập**: Hỗ trợ **xác thực hai yếu tố (2FA)** để tăng độ an toàn khi đăng nhập.
* **Tích hợp hệ thống giám sát an ninh**: Theo dõi hoạt động đăng nhập và cảnh báo khi phát hiện truy cập trái phép.

Dự án đã hoàn thành tốt các mục tiêu ban đầu nhưng vẫn còn nhiều tiềm năng để mở rộng. Các đề xuất trên sẽ giúp nâng cao chất lượng hệ thống, đảm bảo an toàn dữ liệu và mở rộng khả năng ứng dụng trong thực tế, hỗ trợ tốt hơn cho công tác quản lý và điều trị bệnh nhân HIV.

2