**TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI**

**KHOA CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**



**BÁO CÁO BÀI TẬP LỚN MÔN HỆ CƠ SỞ TRI THỨC**

**ĐỀ TÀI: Chuẩn đoán bệnh qua các triệu chứng lâm sàng**

|  |  |
| --- | --- |
|  | **Sinh viên thực hiện:**  **1. Nguyễn Bình Minh - 20152453**  **2. Nguyễn Tiến Thiện – 2015**  **3. Nguyễn Trọng Tuyền**  **4. Phạm Minh Hiếu**  **5.** |

**Gv hướng dẫn: Phạm Văn Hải**

Năm học 2017 - 2018

# Mục lục

**Chuẩn đoán bệnh qua các triệu chứng lâm sàng**

[Mục lục 2](#_Toc513312031)

[Lời nói đầu 2](#_Toc513312032)

[1 Phân tích yêu cầu 3](#_Toc513312033)

[1.1 Vấn đề 3](#_Toc513312034)

[1.2 Mục đích 3](#_Toc513312035)

[2 Phân tích bài toán 3](#_Toc513312036)

[2.1 Các chức năng cơ bản của hệ thống 3](#_Toc513312037)

[2.1.1 Đăng ký, đăng nhập tài khoản vào hệ thống. 3](#_Toc513312038)

[2.1.2 Chuẩn đoán bệnh qua các triệu chứng nhập vào. 3](#_Toc513312039)

[2.1.3 Tra cứu thông tin các bệnh. 4](#_Toc513312040)

[2.2 Phương pháp tiếp cận 4](#_Toc513312041)

[2.2.1 Cơ sở lý thuyết 4](#_Toc513312042)

[2.2.2 Áp dụng vào bài toán 6](#_Toc513312043)

[3 Thiết kế hệ thống 7](#_Toc513312044)

[3.1 Cơ sở tri thức 7](#_Toc513312045)

[3.2 Cơ sở dữ liệu 7](#_Toc513312046)

[3.3 Giao diện chương trình 7](#_Toc513312047)

[4 Đánh giá hệ thống 7](#_Toc513312048)

[5 Hướng phát triển tương lai 7](#_Toc513312049)

# Lời nói đầu

Trong thời đại hiện nay, mỗi người trên thế giới đều phụ thuộc vào điều trị y tế và thuốc. Mỗi ngày chúng ta có thể nghe một số bệnh hoặc triệu chứng của bệnh đã có mới được tìm ra. Nhưng với sự gia tăng số lượng bệnh và triệu chứng của chúng, mỗi người không thể biết được hết các bệnh và triệu chứng này. Vì vậy, để giải quyết vấn đề này, nhóm em đã phát triển một ứng dụng giúp chuẩn đoán bệnh qua các triệu chứng lâm sàng. Ứng dụng này bao gồm một số lượng lớn các bệnh, triệu chứng, phương pháp điều trị để chữa trị.

Một vấn đề lớn hiện nay là vấn đề đi thuê bác sĩ với chi phí khá cao. Nên những tầng lớp trung lưu và những người nghèo không đủ điều kiện thuê bác sĩ chữa trị. Với ứng dụng này, chúng ta có thể dễ dàng tìm kiếm bệnh với đầu vào là các triệu chứng lâm sàng. Ngoài ra còn một số tính năng khác như tìm kiếm, tham khảo thêm về các bệnh.

# Phân tích yêu cầu

## Vấn đề

* Số lượng bệnh và triệu chứng ngày càng nhiều
* Chi phí khám chữa bệnh ngày càng cao
* Tỷ lệ chính xác khám chữa bệnh bác sĩ chưa đạt đến 100%

## Mục đích

* Chuẩn đoán bệnh qua triệu chứng lâm sang
* Tra cứu tìm hiểu về các bệnh cho bệnh nhân
* Là kênh tư vấn cho bác sĩ trước khi chuẩn đoán

# Phân tích bài toán

## Các chức năng cơ bản của hệ thống

### 2.1.1 Đăng ký, đăng nhập tài khoản vào hệ thống.

Người dùng đăng ký tài khoản để lưu các thông tin liên quan đến hệ thống: tên, tuổi, chiều cao, cân nặng, các thông tin tiền sử bệnh án. Từ đó có thể đưa ra các chuẩn đoán cho người dùng tại thời điểm hiện tại thêm chính xác.

### 2.1.2 Chuẩn đoán bệnh qua các triệu chứng nhập vào.

Khi người dùng nhập vào các triệu chứng của bệnh xuất hiện và nhấn tìm kiếm thì hệ thống sẽ dựa vào tri thức đã có để chuẩn đoán bệnh của bệnh nhân.

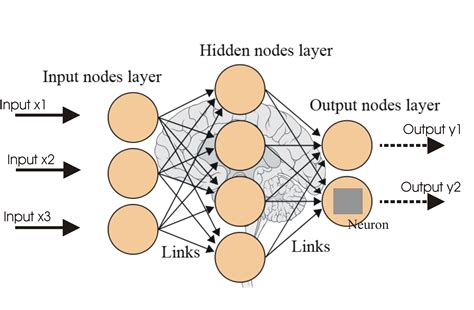
### 2.1.3 Tra cứu thông tin các bệnh.

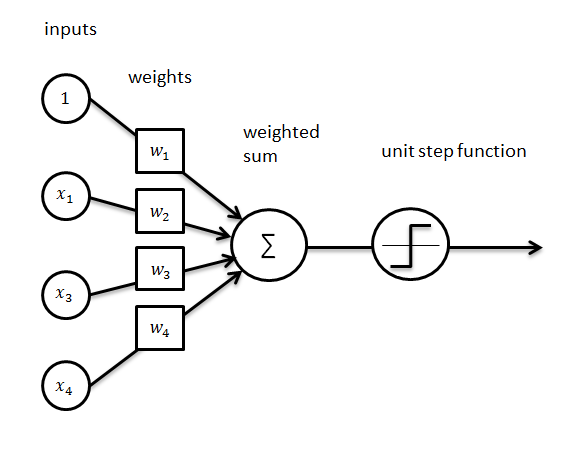
Hệ có các thông tin về khá nhiều bệnh, người dùng có thể tra cứu thông tin về tên, nguyên nhân , triệu chứng và cách điều trị đối với bệnh.

## Phương pháp tiếp cận

### 2.2.1 Cơ sở lý thuyết

Mạng nơron nhân tạo (Artificial Neural Network) bao gồm các nơron ( còn gọi là một đơn vị xử lý hay một node) được nối với nhau bởi các liên kết nơron. Mỗi liên kết kèm theo một trọng số nào đó, đặc trưng cho tính kích hoạt/ ức chế giữa các nơron. Có thể xem các trọng số là các phương tiện để lưu trữ thông tin dài hạn trong mạng nơron. Nhiệm vụ chung của quá trình huấn luyện mạng là cập nhật các trọng số khi có thêm các thông tin về mẫu học.



Cấu trúc của một nơ ron như hình bên dưới: Các thành phần cơ bản của một nơ ron nhân tạo bao gồm:

* Tập các đầu vào (inputs): là các tín hiệu vào (input signal) của nơ ron, các tín hiệu này thường được đưa dưới dạng vector N chiều.
* Tập các liên kết: Mỗi liên kết được thể hiện bởi một trọng số (gọi là trọng số liên kết – Synaptic weight). Thông thường các trọng số này được khởi tạo một cách ngẫu nhiên ở thời điểm khởi tạo mạng và được cập nhật liên tục trong quá trình học mạng.
* Bộ tổng: Thường dùng để tính tổng của các tích đầu vào với trọng số liên kết của nó:
* Trọng số điều chỉnh (bias): Thường được đưa vào như một thành phần của hàm truyền.

Hàm tác động / truyền (Activation / Transfer function): Hàm này được dùng để giới hạn phạm vi đầu ra của mỗi nơ ron. Nó nhận đầu vào là kết quả của hàm tổng và ngưỡng đã cho. Thông thường, phạm vi đầu ra của mỗi nơ ron được giới hạn trong đoạn [0, 1] hoặc [-1, 1]. Các hàm truyền rất đa dạng, có thể là các hàm tuyến tính hoặc phi tuyến. Việc lựa chọn hàm truyền nào là tùy thuộc vào từng bài toán và kính nghiệm của người thiết kế mạng.

### 2.2.2 Áp dụng vào bài toán

Áp dụng mạng nơ ron để xây dựng hệ cơ sở tri thức. Từ hệ cơ sở tri thức này hệ thống có thể đưa ra các chuẩn đoán với đầu vào là các triệu chứng, cho ra đầu ra là các bệnh có thể mắc phải của bệnh nhân.

Xây dựng một mạng nơ ron có nhiều tầng ẩn, với mỗi tầng ẩn có nhiều nơ ron. Sau khi xây dựng được mô hình mạng, mạng nơ ron sẽ được huấn luyện qua tập dữ liệu có sẵn để tạo ra tri thức. Tri thức trong mô hình mạng nơ ron là tri thức không tường minh, nó được biểu hiện qua cấu trúc mạng nơ ron và các trọng số liên kết sau khi huấn luyện.

Tập dữ liệu huấn luyện mạng nơ ron là một tập các bản ghi. Mỗi bản ghi là một trường hợp của bệnh nhân đã có các triệu chứng và bệnh mắc phải tương ứng.

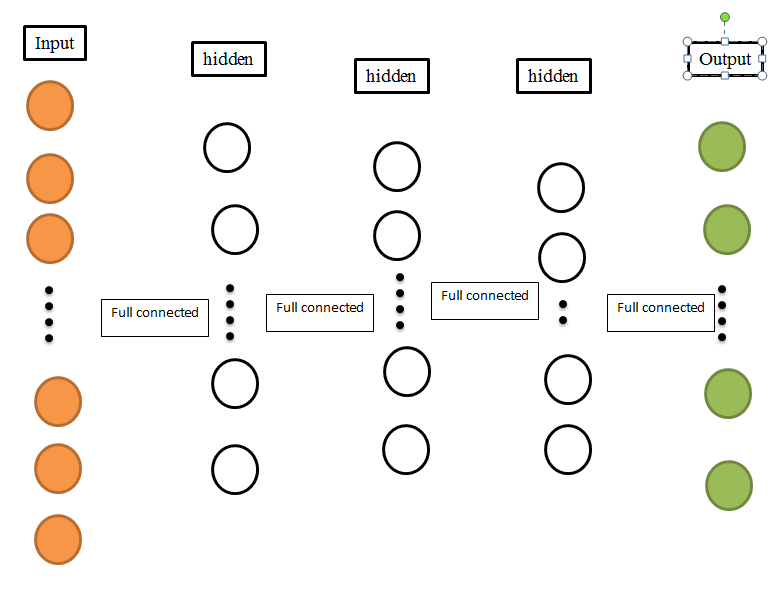
Sau khi huấn luyện mạng ta có được một cơ sở tri thức về chuẩn đoán bệnh. Từ đó, ta sẽ sử dụng nó vào việc chuẩn đoán bệnh khi có các triệu chứng mà người dùng nhập vào.

Ngoài ra, cơ sở tri thức còn chứa các thông tin khác về các bệnh. Các thông tin giúp người dùng có thể tra cứu, tìm hiểu thêm về bệnh.

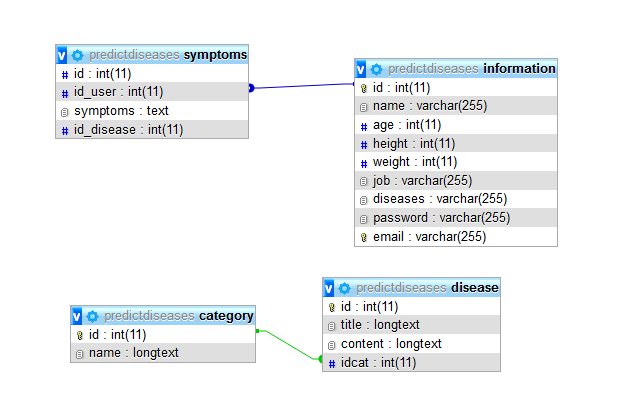
# Thiết kế hệ thống

## Cơ sở tri thức

* Tri thức được biểu diễn dưới dạng mạng nơ ron. Sau khi mạng nơ ron đã được huấn luyện chúng ta có được cơ sở tri thức tùy thuộc vào cấu trúc mạng và tập dữ liệu huấn luyện mạng.
* Biểu diễn dữ liệu:
  + Đầu vào của mạng nơ ron: là một vector n chiều (n là số loại triệu chứng trong tập dữ liệu). Mỗi phần tử của vector có một trong hai giá trị: 1 thể hiện có triệu chứng tại vị trí của phần tử đó xuất hiện trong đầu vào; ngược lại, phần tử có giá trị là 0 nếu triệu chứng tại vị trí đó của vector không xuất hiện trong đầu vào.  
    Ví dụ: Tập triệu chứng có 5 triệu chứng: đau đầu, đau nửa đầu, đau dạ dày, ho, mệt mỏi. Đầu vào được biểu diễn bởi vector 5 chiều với chỉ số các phần tử tương ứng với các triệu chứng trong tập triệu chứng. Với đầu vào có các triệu chứng: ho, đau đầu thì vector đầu vào sẽ là: [1, 0, 0, 1, 0].
  + Đầu ra của mạng nơ ron: là một vector m chiều (m là số loại bệnh có trong tập dữ liệu). Mỗi phần tử của vector là số thực có giá trị trong khoảng 0 đến 1 thể hiện xác suất mắc bệnh tương ứng với chỉ số của phần tử đó.   
    Ví dụ: Tập bệnh có 3 bệnh: cao huyết áp, dị ứng, tiểu đường. Đầu ra được biểu diễn bởi vector 3 chiều với chỉ số các phần tử tương ứng với các bệnh trong tập bệnh. Với đầu ra [0.99421, 0.211122, 0. 124432] thì xác suất mắc bệnh cao huyết áp là cao nhất : 99.42%.
* Mô tả dữ liệu:
  + Dữ liệu là một bảng có 133 cột và 4920 hàng.
  + Mỗi cột tương ứng với một triệu chứng, ngoại trừ cột cuối cùng (cột thứ 133) để chỉ tên bệnh tương ứng.
  + Mỗi hàng là một bản ghi tương ứng trường hợp mắc bệnh của bệnh nhân.
* Cơ chế suy diễn:
  + Từ giao diện người dùng, lấy các triệu chứng mà người dùng nhập vào.
  + Biểu diễn dữ liệu nhận được dưới dạng vector đầu vào của mạng nơ ron.
  + Dùng mạng nơ ron đã huấn luyện để chuẩn đoán bệnh với đầu vào là vector ở trên.
  + Sau khi truyền cho mạng nơ ron đầu vào ta được đầu ra là vector chỉ xác suất xảy ra các bệnh.
  + Dựa vào xác suất xảy ra bệnh để đưa ra chuẩn đoán cho người dùng bệnh có thể mắc phải nhất.
* Hướng tiếp cận cài đặt:
  + Thư viện: Dùng thư viện keras để xây dựng mô hình mạng nơ ron.
  + Mô hình mạng nơ ron: Mạng gồm có 5 tầng. Tầng thứ nhất (tầng đầu vào) có 132 nơ ron, tầng 2, 3, 4 là các tầng ẩn với số nơ ron lần lượt là 100, 80, 60; tầng thứ 5 (tầng đầu ra) có 42 nơ ron.



## Cơ sở dữ liệu



# Cài đặt

## Công nghệ sử dụng

Các ngôn ngữ sử dụng bao gồm: python, php, javascript

### Framework Codeigniter

Codeigniter Framework là một trong các Framework được sử dụng để thiết kế web. CodeIgniter mạnh mẽ với footprint rất nhỏ, dành cho các lập trình viên cần một bộ công cụ đơn giản và thông minh để tạo ra các ứng dụng web với đầy đủ các tính năng

Lý do dùng Codeigniter:

* Footprint nhỏ 2MB
* Hướng dẫn sử dụng rõ rang dễ hiểu
* Khả năng tương thích với lưu trữ chuẩn
* **Được thiết kế theo mô hình Model-View-Controller**
* **Hệ thống thư viện đa dạng**
* **Tấc độ nhanh**
* **Và quan trọng là miễn phí**

### Framework Bootstrap

Bootstrap là một framework cho phép thiết kế website reponsive nhanh hơn và dễ dàng hơn. Bootstrap là bao gồm các HTML templates, CSS templates và Javascript tao ra những cái cơ bản có sẵn như: typography, forms, buttons, tables, navigation, modals, image carousels và nhiều thứ khác. Trong bootstrap có thêm các plugin Javascript trong nó. Giúp cho việc thiết kế reponsive của bạn dễ dàng hơn và nhanh chóng hơn.

Lý do dùng bootstrap:

* Rất dễ để sử dụng
* Tính năng responsive

Bootstrap’s xây dựng sẵn reponsive css trên các thiết bị phones, tablets, và desktops

* Bootstrap mobile style là 1 phần core của framework
* Tương thích với trình duyệt

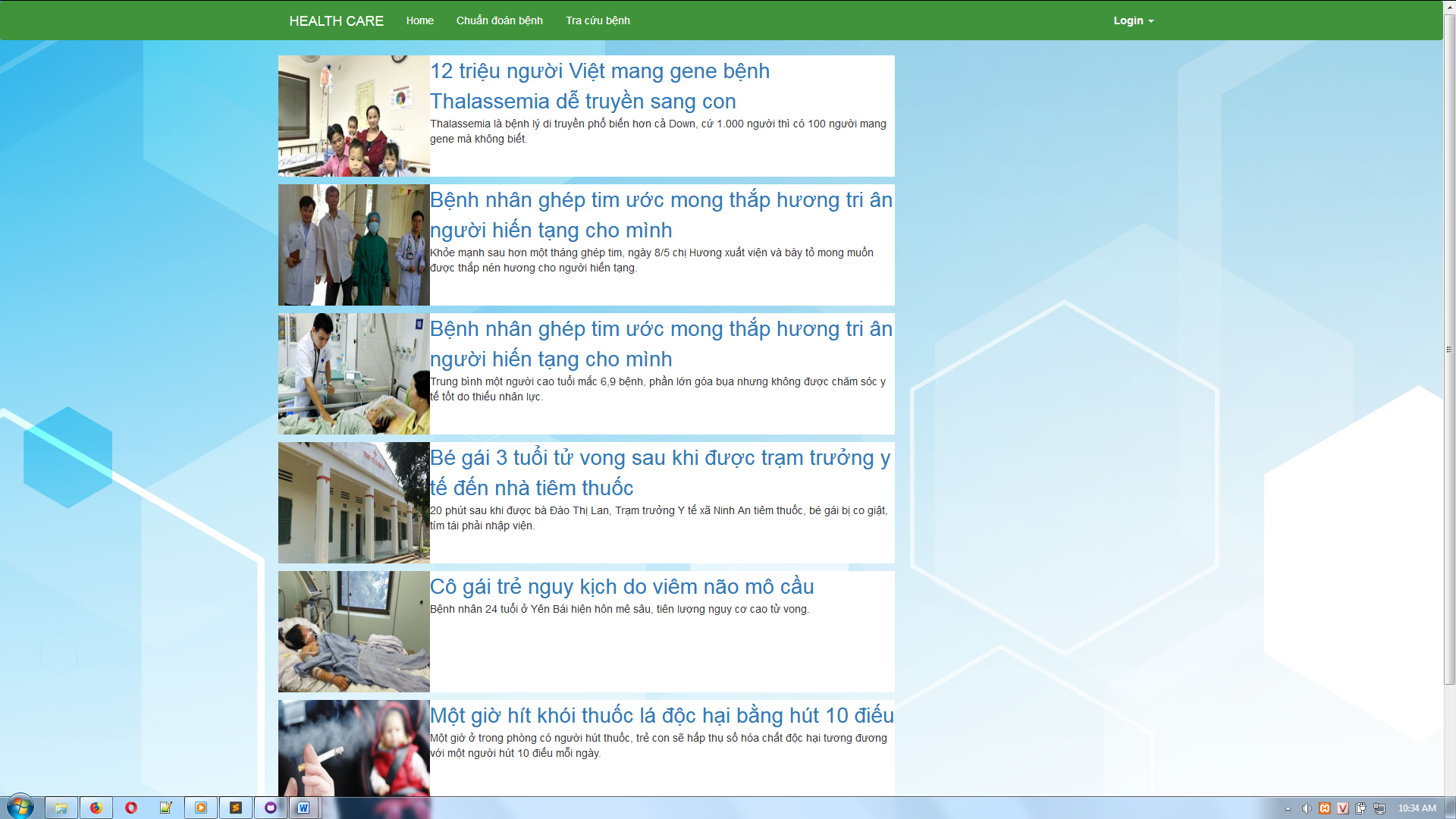
Áp dụng vào bài toán lấy bootstrap từ CDN

### Python

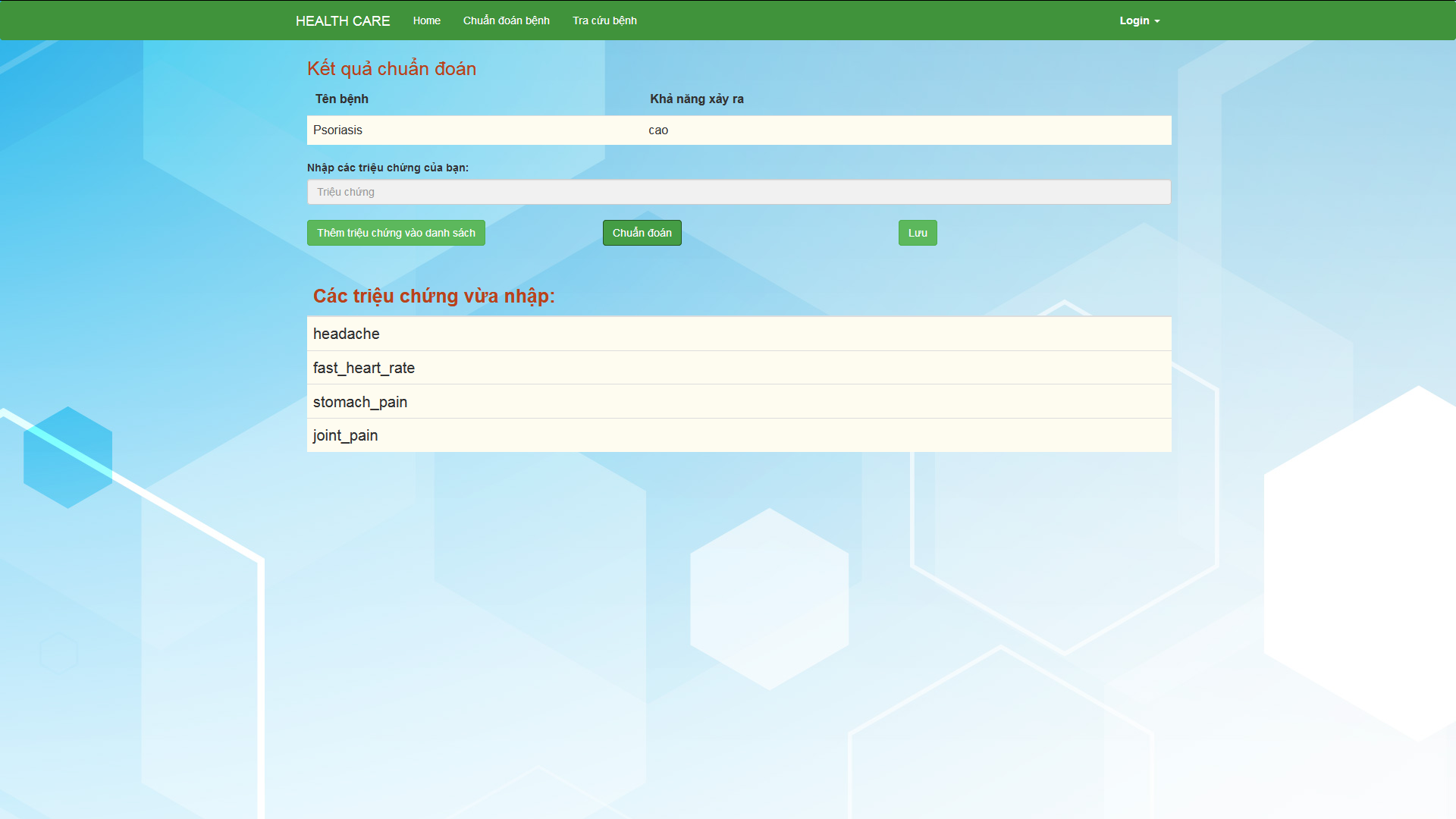
### Php

## Giao diện chương trình

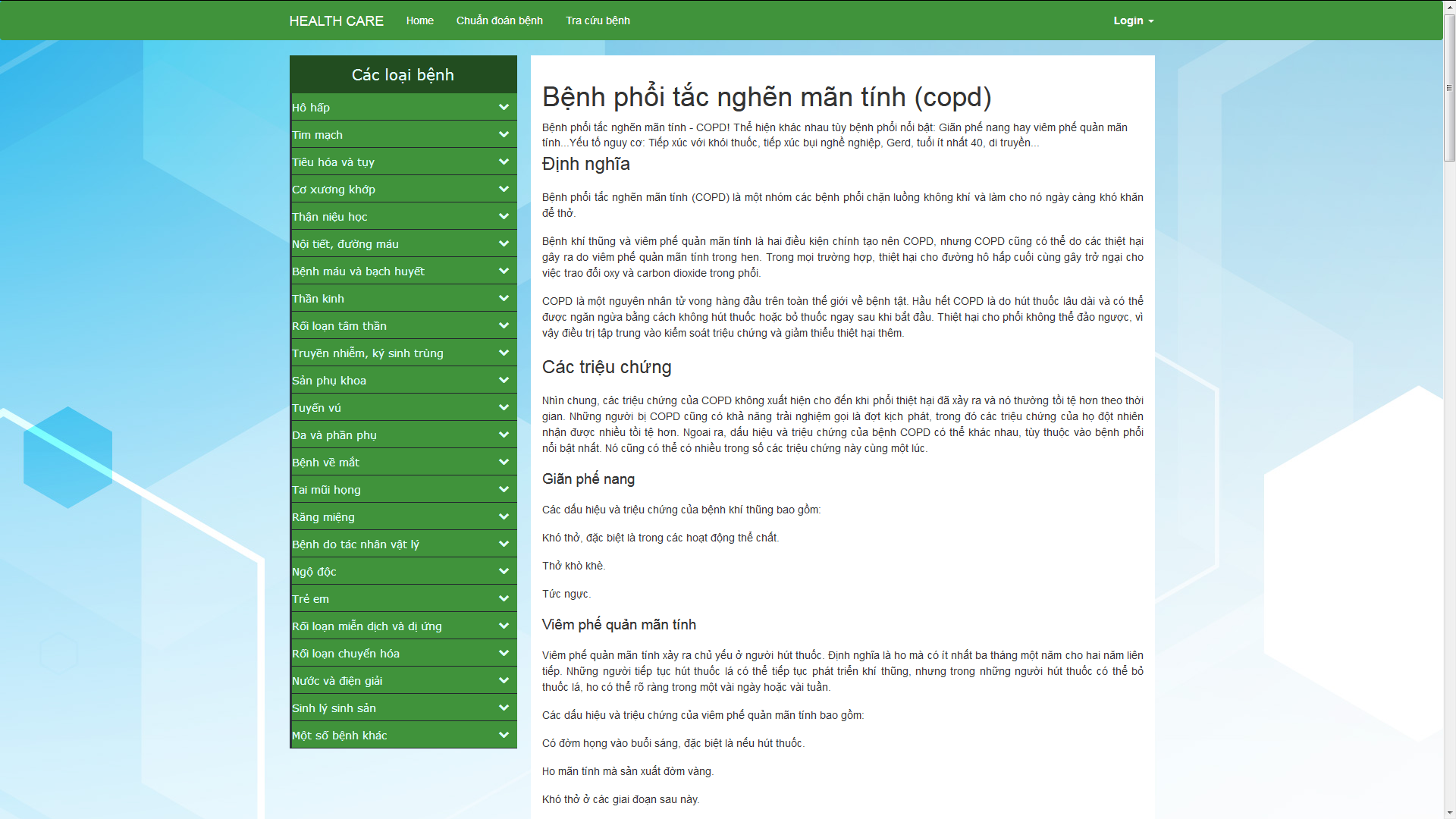
### Giao diện trang chủ hiển thị các tin tức



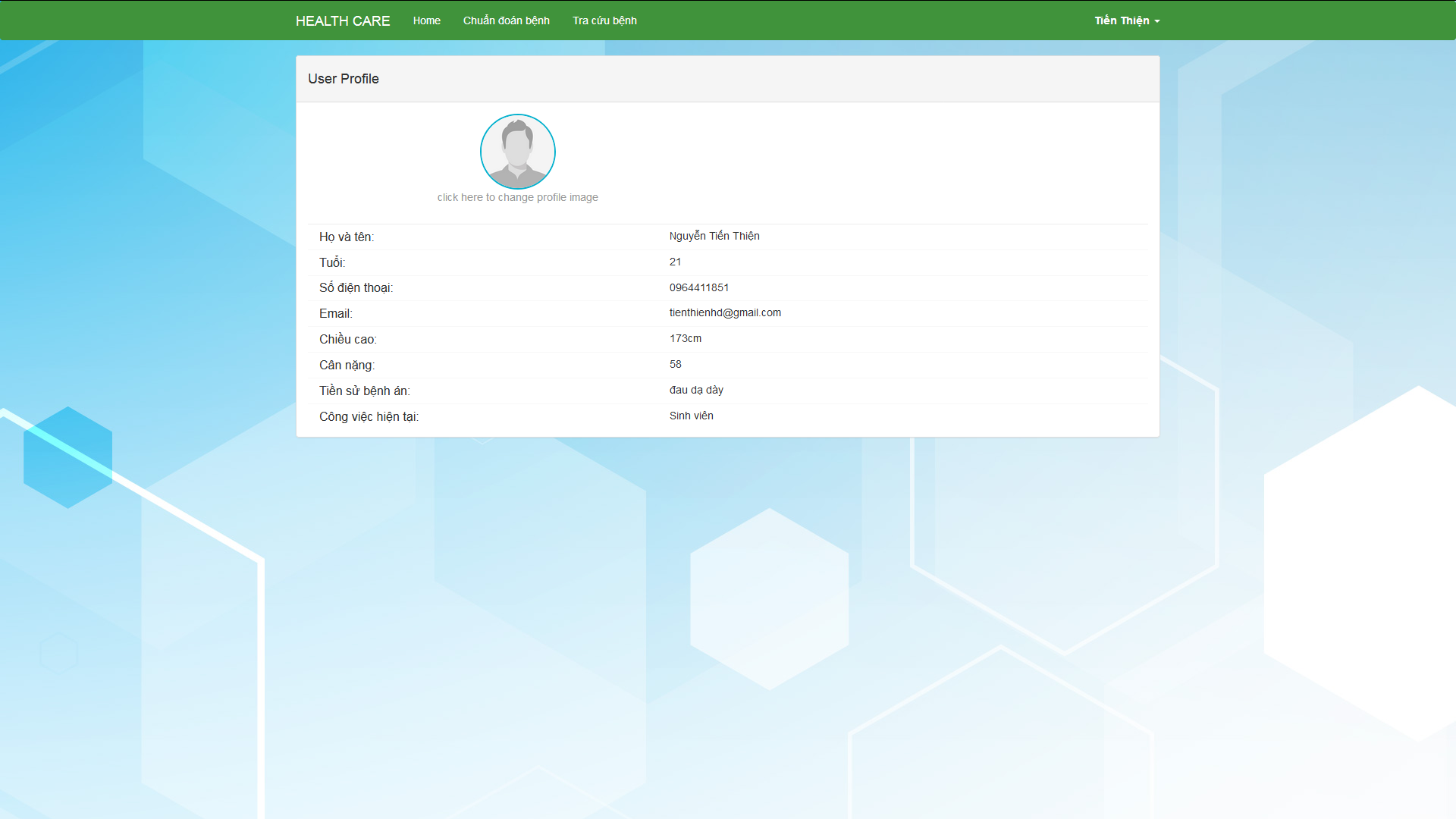
### Giao diện chuẩn đoán bệnh



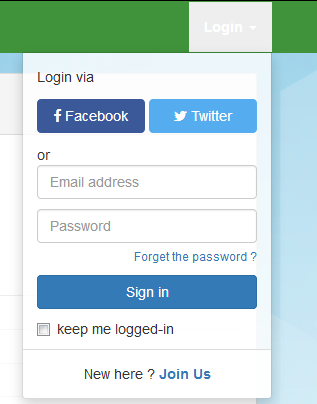
### Giao diện tra cứu bệnh



### Giao diện hiển thị thông tin cá nhân



### Giao diện đăng nhập



# Đánh giá hệ thống

# Hướng phát triển tương lai