**TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

**KHOA CÔNG NGHỆ THÔNG TIN**

**NGÀNH KHOA HỌC MÁY TÍNH**

**Nguyễn Trọng Triết – 1753113**

**Nguyễn Vương Đạt Vũ – 1753131**

**KHÓA LUẬN TỐT NGHIỆP**

**NHẬN DIỆN ĐƠN THUỐC SỬ DỤNG PHƯƠNG PHÁP HỌC SÂU**

KHÓA LUẬN TỐT NGHIỆP CỬ NHÂN

CHƯƠNG TRÌNH CHẤT LƯỢNG CAO

**GIÁO VIÊN HƯỚNG DẪN**

ThS. Lê Ngọc Thành

TP. Hồ Chí Minh, Ngày Tháng Năm 2021

MỤC LỤC

[**LỜI CẢM ƠN** 3](#_Toc53755474)

[**DANH MỤC HÌNH ẢNH** 4](#_Toc53755475)

[**DANH MỤC BẢNG BIỂU** 5](#_Toc53755476)

[**DANH MỤC THUẬT NGỮ** 6](#_Toc53755477)

[**TÓM TẮT KHÓA LUẬN** 7](#_Toc53755478)

[**CHƯƠNG 1: MỞ ĐẦU** 8](#_Toc53755479)

[**1.1** **Giới thiệu đề tài** 8](#_Toc53755480)

[**1.2** **Lý do chọn đề tài** 8](#_Toc53755481)

[**1.3** **Nghiên cứu liên quan** 9](#_Toc53755482)

[**1.4** **Hướng tiếp cận và giải quyết vấn đề** 9](#_Toc53755483)

[**1.5** **Mục tiêu luận văn** 9](#_Toc53755484)

[**1.6** **Phạm vi đề tài** 10](#_Toc53755485)

[**1.7** **Cấu trúc của luận văn** 10](#_Toc53755486)

[**CHƯƠNG 2: CƠ SỞ LÝ THUYẾT** 11](#_Toc53755487)

# **LỜI CẢM ƠN**

Lời đầu tiên, chúng em xin trân trọng gửi lời cảm ơn sâu sắc và chân thành nhất tới ThS. Lê Ngọc Thành đã chỉ dạy và hướng dẫn nhiệt tình trong quá trình học cũng như hoàn thành luận văn.

Chúng em xin chân thành cảm ơn quý thầy cô và ban giám hiệu trường đại học Khoa Học Tự Nhiên, đặc biệt là thầy cô khoa Công nghệ thông tin đã tận tình giảng dạy chúng em trong suốt 4 năm đại học. Đó không chỉ là kiến thức mà còn là hành trang để chúng em bước vào đời.

Chúng em xin cảm ơn quý nhà trường đã tạo điều kiện thuận lợi cho chúng em học tập và thực hiện đề tài tốt nghiệp này.

Chúng em xin cảm ơn sự quan tâm giúp đỡ và ủng hộ của các anh chị bạn bè trong quá trình thực hiện khóa đề tài. Mặc dù đã cố gắng hoàn thành khóa luận trong phạm vi và khả năng cho phép nhưng chắc chắn sẽ không tránh khỏi những thiếu sót.

Chúng em rất mong nhận được sự thông cảm, góp ý và tận tình chỉ bảo của quý thầy cô và các bạn.

TP. Hồ Chí Minh, Ngày Tháng Năm 2021

# **DANH MỤC HÌNH ẢNH**

# **DANH MỤC BẢNG BIỂU**

# **DANH MỤC THUẬT NGỮ**

# **TÓM TẮT KHÓA LUẬN**

Trong 10 năm gần đây, với sự phát triển của phương pháp học sâu nói riêng cũng như ngành trí tuệ nhân tạo nói chung đã tạo ra sự thu hút sự quan tâm đặc biệt đối với các sinh viên ngành công nghệ thông tin, các nghiên cứu sinh, các nhà khoa học… Trong đó, bài toán nhận diện ký tự quang học, đang được nhóm chúng em quan tâm. Luận văn này trình bày về nghiên cứu và xây dựng 1 hệ thống nhận diện và trính xuất thông tin của đơn thuốc, trong đó gồm thông tin của thuốc(được so sánh với ngân hàng thuốc Việt Nam). Trên cơ sở các nghiên cứu với điều kiện thực tế của các công cụ xử lý ngôn ngữ tiếng Việt, nhóm đưa ra đề xuất sử dụng các mô hình học máy để nhận diện, phân tách, trích xuất và so sánh ký tự nhằm đưa ra tỷ lệ chính xác nhất. Thực nghiệm và đánh giá các mô hình cho thấy kết quả khả quan tạo tiền đề cho các phát triển tiếp theo trong tương lai.

# **CHƯƠNG 1: MỞ ĐẦU**

## **Giới thiệu đề tài**

Đơn thuốc là tài liệu chỉ định dùng thuốc của bác sĩ cho người bệnh; Là cơ sở pháp lý cho việc chỉ định sử dụng thuốc, bán thuốc và cấp thuốc theo đơn. Các công việc liên quan tới đơn thuốc rất quan trọng, đặc biệt là việc lấy thông tin của thuốc. Việc thường thấy là bệnh nhân mang đơn thuốc bác sĩ đã in ra các cửa hàng thuốc để tìm mua thuốc hoặc là người bán thuốc tìm thông tin của thuốc qua đơn thuốc.

Trên thực tế, các công việc kiểm tra tên thuốc bằng máy vẫn chưa phổ biến và việc này vẫn đang thực hiện bằng cách thủ công bằng cách các dược sĩ sẽ nhập tên thuốc lên hệ thống để kiểm tra. Nếu số lượng đơn thuốc lớn thì cần rất nhiều thời gian để nhập tay từng loại thuốc hoặc cần thêm nhiều nguồn nhân lực để đẩy nhanh tiến độ.

Ngoài ra, khi kiểm tra thuốc bằng cách thủ công có thể dẫn đến sai sót cao đồng thời kéo theo rủi ro sai tên thuốc có thể gây nguy hiểm cho bệnh nhân.

Dựa vào sự phát triển của công nghệ 4.0, với sự giúp đỡ của các thiết bị công nghệ sẽ giúp con người giải quyết các công việc cần độ chính xác cao và cần sự nhanh chóng.

## **Lý do chọn đề tài**

Nhằm giúp ngành y tế phát triển, nhóm chúng em quyết định nghiên cứu một giải pháp giúp cho việc lấy thông tin trên đơn thuốc qua nhưng tấm hình đơn thuốc bình thường, chất lượng cao. Sau đó, có thể trích xuất tên loại thuốc rồi so sánh với ngân hàng thuốc của bộ y tế để đưa ra tên thuốc chính xác nhất. Công việc này có thể giúp cho người bệnh tìm hiểu các loại thuốc mình sử dụng và giúp cho các dược sĩ lấy được thông tin thuốc rồi kiểm tra với hệ thống thuốc trong cửa hàng.

## **Nghiên cứu liên quan**

Đọc và trích xuất văn bản thông qua hình ảnh là giải pháp không mới và đã có nhiều công trình nghiên cứu đã được sử dụng vào đời sống. Tuy nhiên, áp dụng công nghệ này vào ngành y tế nhằm trích xuất thông tin đơn thuốc vẫn chưa được phổ biến.

Hiện nay có rất nhiều thư viện nhận diện ký tự quang học phổ biến như OCR engines ABBYY Finereader, OCRopus, Tesseract, Calamari, EasyOCR... Nhưng đối với các ký tự có dấu của Tiếng Việt thì các thư viện này hỗ trợ nhận diện ký tự với độ chính xác không cao. Để đạt được độ chính xác cao thì người dung cần huấn luyện mô hình bộ dữ liệu Tiếng Việt riêng. Và để hỗ trợ cho khả năng đọc được tốt hơn thì hình ảnh đầu vào cần có độ tương phản cao, nên là chữ đen nền trắng và giảm thiểu các điểm nhiễu trên hình.

Luận văn này sẽ tiếp tục theo những hướng nghiên cứu trên bằng cách huấn luyện mô hình máy học để đọc Tiếng Việt và kết hợp mô hình đọc Tiếng Anh có sẵn để phù hợp với đọc tên thuốc. Ngoài ra vẫn sẽ có các công việc tiền xử lý hình ảnh và xác định vùng thông tin cần đọc. Thông tin sau khi được đọc sẽ so sánh với ngân hàng dữ liệu nhằm đưa ra được kết quả chính xác nhất.

## **Hướng tiếp cận và giải quyết vấn đề**

Đọc thông tin đơn thuốc qua hình chụp được chia thành ba phần:

1. Tiền xử lý hình ảnh đưa vào. Hình đưa vào sẽ đc xử lý và cắt giữ đơn thuốc.
2. Xác định vùng chứa văn bản và tìm vùng thông tin chứa tên tên thuốc.
3. Nhận diện ký tự trong vùng chứa văn bản và nâng cao tỉ lệ chính xác của tên thuốc.

## **Mục tiêu luận văn**

Các mục tiêu sinh viên tự đề ra để hoàn thành tốt đề tài luận văn:

* Xây dựng model nhận dạng đơn thuốc thông qua hình chụp. Hình ảnh đầu vào phải được tiền xử lý và bao quát được đươc đơn thuốc.
* Khoanh vùng các ký tự cần đọc trên hình ảnh.
* Ngân hàng dữ liệu phải luôn được cập nhật để thuận tiện cho việc so sánh.
* Đạt dược độ chính xác cao nhất với tên thuốc.
* Giảm thiểu thời gian đọc dữ liệu.
* Hỗ trợ vừa đọc được tên thuốc Tiếng Việt và Tiếng Anh.

## **Phạm vi đề tài**

Luận văn của nhóm được giới hạn trong phạm vi đọc các nội dung của đơn thuốc thông qua ảnh được chụp và đưa ra được tên thuốc chính xác nhất trên đơn thuốc. Nhóm sẽ tạo ra một ứng dụng có thể đáp ứng được yêu cầu trên gồm chụp ảnh hoặc tải ảnh lên để xử lý và trả về kết quả là thông tin của tên loại thuốc. Nội dung chính được sử dụng gồm các chữ số, ngôn ngữ Tiếng Việt và Tiếng Anh. Phạm vi trên không bao gồm các đơn thuốc viết tay mà là sử dụng các đơn thuốc được in thành văn bản.

## **Cấu trúc của luận văn**

# **CHƯƠNG 2: CƠ SỞ LÝ THUYẾT**