

**NATIONAL UNIVERSITY HO CHI MINH CITY
UNIVERSITY OF INFORMATION TECHNOLOGY
ADVANCED PROGRAM IN INFORMATION SYSTEMS**

NGUYEN NHAT THUONG - 20522000

**THE COMBINATION OF ACNE ANALYSIS
MODELS WITH CUSTOMIZED LARGE
LANGUAGE MODELS TO IMPROVE SKINCARE**

BACHELOR OF ENGINEERING IN INFORMATION SYSTEMS

**THESIS ADVISOR
PhD. NGUYEN THANH BINH**

HO CHI MINH CITY, 2024

ASSESSMENT COMMITTEE

The Assessment Committee is established under the Decision., date
.....by Rector of the University of Information Technology.

- 1 - Chairman.
- 2 - Secretary.
- 3 - Member
- 4 - Member.

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to the Faculty of Information Systems, University of Information Technology - Vietnam National University, Ho Chi Minh City for providing me with favorable conditions to study and complete this research project. I would like to extend my deepest gratitude to Phd. Nguyen Thanh Binh, who dedicated his time and effort to impart valuable knowledge, provide guidance, and accompany me throughout the entire process.

I have strived to apply the knowledge and skills acquired during my studies to complete this project. However, due to my limited practical experience and expertise, there are undoubtedly shortcomings in my report. I sincerely hope to receive constructive feedback from the professors to further improve its quality.

Once again, I would like to express my heartfelt appreciation for the support and guidance provided by all professors during this endeavor.

Ho Chi Minh City, October 10, 2024

Nguyen Nhat Thuong

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

- 1.1 Problem Statement
- 1.2 Scope and methods of research
- 1.3 Contribute of Thesis

CHAPTER 2 THEORETICAL BASIS

- 2.1 Dataset overview
 - 2.1.1 Acne dataset
 - 2.1.2 RAG dataset
- 2.2 Advice Treatment
 - 2.2.1 Retrieval-augmented generation overview
 - 2.2.2 Data extraction
 - 2.2.3 Embedding
 - 2.2.4 Cosine Similarity
 - 2.2.5 ElasticSearchDB
 - 2.2.6 Chatbox
- 2.3 Acne Detection
 - 2.3.1 YOLOv8
 - 2.3.2 YOLOV8 and Slicing Aided Hyper Inference
- 2.4 Evaluate
 - 2.4.1

CHAPTER 3 RESULT AND ANALYSIS

- 3.1 Methodological approach
- 3.2 Result and analysis
 - 3.2.1 Acne detection using YOLOv8
 - 3.2.2 Acne detection using YOLOv8 and SAHI
 - 3.2.3 Chunk Evaluate
 - 3.2.4 RAG Evaluate Chunk Size

CHAPTER 4 SYSTEM IMPLEMENT APPLICATION

- 4.1 System architecture

CHAPTER 5 CONCLUSION AND FEATURE WORK

REFERENCES

CHAPTER 1 INTRODUCTION

1.1 Problem Statement

Mụn là một trong những vấn đề da liễu phổ biến nhất, ảnh hưởng nghiêm trọng đến chất lượng cuộc sống và tâm lý của người mắc phải. Theo thống kê của Hiệp hội Da liễu Việt Nam, có đến 85% thanh thiếu niên Việt Nam từng gặp vấn đề về mụn, trong đó nhiều trường hợp không được điều trị đúng cách, dẫn đến các biến chứng như sẹo rỗ, thâm mụn hoặc viêm da kéo dài [1]. Tuy nhiên, việc nhận diện chính xác loại mụn (như mụn đầu đen, mụn đầu trắng, mụn viêm, mụn bọc, v.v.) và lựa chọn phương pháp điều trị phù hợp lại là một thách thức lớn đối với nhiều người. Nguyên nhân chủ yếu là do thiếu kiến thức chăm sóc da khoa học hoặc khó tiếp cận với các dịch vụ da liễu chuyên nghiệp, đặc biệt tại các khu vực nông thôn và vùng sâu vùng xa. Ngoài ra, chi phí điều trị tại các phòng khám da liễu hoặc bệnh viện lớn thường cao, khiến nhiều người tự ý sử dụng các sản phẩm không rõ nguồn gốc hoặc điều trị sai cách, làm tình trạng da trở nên tồi tệ hơn [2].

Trong bối cảnh đó, việc phát triển một ứng dụng AI hỗ trợ nhận diện mụn và đưa ra lời khuyên điều trị cá nhân hóa là giải pháp tiềm năng, đáp ứng nhu cầu thực tế của người dùng tại Việt Nam. Công nghệ trí tuệ nhân tạo, đặc biệt là học sâu (deep learning), đã được chứng minh hiệu quả trong việc phân tích hình ảnh và chẩn đoán các vấn đề da liễu [3]. Ứng dụng này có thể giúp người dùng dễ dàng chụp ảnh da mặt, nhanh chóng xác định loại mụn, và nhận được các bước điều trị phù hợp dựa trên tình trạng da. Bên cạnh đó, ứng dụng còn cung cấp thông tin khoa học về chăm sóc da, giúp nâng cao nhận thức cộng đồng về việc điều trị mụn đúng cách, từ đó giảm thiểu các rủi ro do tự điều trị sai lầm. Với sự tiện lợi, chi phí thấp và khả năng tiếp cận rộng rãi, ứng dụng AI này không chỉ hỗ trợ cá nhân hóa việc chăm sóc da mà còn góp phần nâng cao chất lượng cuộc sống và sự tự tin của người dùng tại Việt Nam.

みなさん、こんにちは。私の名前はグエン・ニャット・トゥオンです。

みなさん、こんにちは。私の名前はグエン・ニャット・トゥオンです。

みなさん、こんにちは。私の名前はグエン・ニャット・トゥオンです。

みなさん、こんにちは。私の名前はグエン・ニャット・トゥオンです。

1.2 Scope and methods of research

Nghiên cứu tập trung vào việc phát triển một ứng dụng trí tuệ nhân tạo (AI) hỗ trợ nhận diện các loại mụn và cung cấp lời khuyên điều trị cá nhân hóa, với phạm vi cụ thể như sau: Đối tượng nghiên cứu bao gồm các loại mụn phổ biến như mụn đầu đen, mụn đầu trắng, mụn viêm, mụn bọc và các vấn đề liên quan đến da dầu, sẹo mụn.... Người dùng tiềm năng được xác định là thanh thiếu niên và người trưởng thành tại Việt Nam, đặc biệt ở độ tuổi từ 15-35, những người thường xuyên gặp phải các vấn đề về mụn. Không gian nghiên cứu tập trung vào thị trường Việt Nam, với trọng tâm là cả khu vực thành thị lẫn nông thôn, nơi khả năng tiếp cận dịch vụ da liễu có sự khác biệt rõ rệt. Đồng thời, nghiên cứu sử dụng các cơ sở dữ liệu hình ảnh da liễu, bao gồm ảnh y tế và ảnh thực tế từ người dùng, để huấn luyện và thử nghiệm thuật toán AI nhận diện mụn YOLOv8, YOLOv8 kết hợp với SAHI (Slicing Aided Hyper Inference) Mô hình ngôn ngữ lớn ChatGPT4o. Về mặt công nghệ, nghiên cứu sẽ ứng dụng các phương pháp học sâu (deep learning) và xử lý hình ảnh nhằm đảm bảo khả năng nhận diện chính xác và cung cấp các khuyến nghị điều trị phù hợp với từng cá nhân.

1.3 Contribute of Thesis

Luận văn mang lại hai đóng góp quan trọng như sau: Thứ nhất, nghiên cứu đã phát triển một ứng dụng trí tuệ nhân tạo (AI) với khả năng nhận diện các loại mụn phổ biến như mụn đầu đen, mụn đầu trắng, mụn viêm, mụn bọc,... thông qua phân tích hình ảnh, đồng thời đưa ra các khuyến nghị điều trị cá nhân hóa. Ứng dụng này không chỉ hỗ trợ người dùng tự đánh giá tình trạng da mà còn mở ra cơ hội tiếp cận giải pháp chăm sóc da hiệu quả cho những người gặp khó khăn trong việc tiếp cận các dịch vụ y tế và làm đẹp hiện đại. Đây là bước tiến quan trọng, tạo nền tảng cho việc mở rộng nghiên cứu và ứng dụng công nghệ AI trong lĩnh vực da liễu trong tương lai. Thứ hai, nghiên cứu góp phần nâng cao nhận thức cộng đồng về chăm sóc da khoa học thông qua việc cung cấp thông tin chính xác và hướng dẫn các phương pháp điều trị phù hợp. Điều này giúp ngăn ngừa các sai lầm trong điều trị, từ đó cải thiện sức khỏe làn da và giảm thiểu nguy cơ biến chứng lâu dài.

CHAPTER 2 THEORETICAL BASIS

2.1 Dataset overview

2.1.1 Acne dataset

Bộ dữ liệu sử dụng trong nghiên cứu này bao gồm hai tập hợp về các bệnh lý da liễu, tập trung vào 14 loại tổn thương da khác nhau. Với tổng cộng 11,533 hình ảnh, bộ dữ liệu

được phân loại theo các nhãn sau: sẹo mụn, mụn đầu đen/trắng, mụn conglobata, mụn kết tinh, mụn nang, mụn cóc phẳng, viêm nang lông, sẹo lồi, u milia, mụn sần, mụn mủ, mụn kết hợp, u mô hôi, và nốt ruồi. Và được chia thành 3 tập dữ liệu là train, test và validate

| | | |
|----------------|-------------------|---------------|
| sẹo mụn | mụn đầu đen/trắng | conglobata |
| mụn kết tinh | mụn nang | mụn cóc phẳng |
| viêm nang lông | sẹo lồi, u milia | mụn sần |
| mụn mủ | mụn kết hợp | u mô hôi |
| nốt ruồi | | |

2.1.2 RAG dataset

2.2 Retrieval-augmented generation

CHAPTER 3 RESULT AND ANALYSIS

3.1 Methodological approach

3.2 Result and analysis Acne Detection

3.2.1 Validation YOLOv8

| Class | Precision | Recall | mAP50 | mAP50-95 |
|--------------------|-----------|--------|-------|----------|
| acne scars | | | | |
| comedone | | | | |
| conglobata | | | | |
| crystalline | | | | |
| cystic | | | | |
| flat wart | | | | |
| folliculitis | | | | |
| keloid | | | | |
| milium | | | | |
| papule | | | | |
| pustule | | | | |
| sebo-crystn-conglo | | | | |
| syringoma | | | | |
| birthmark | | | | |
| all | 0.906 | 0.849 | 0.91 | 0.724 |

3.2 Result and analysis Advice Treatment

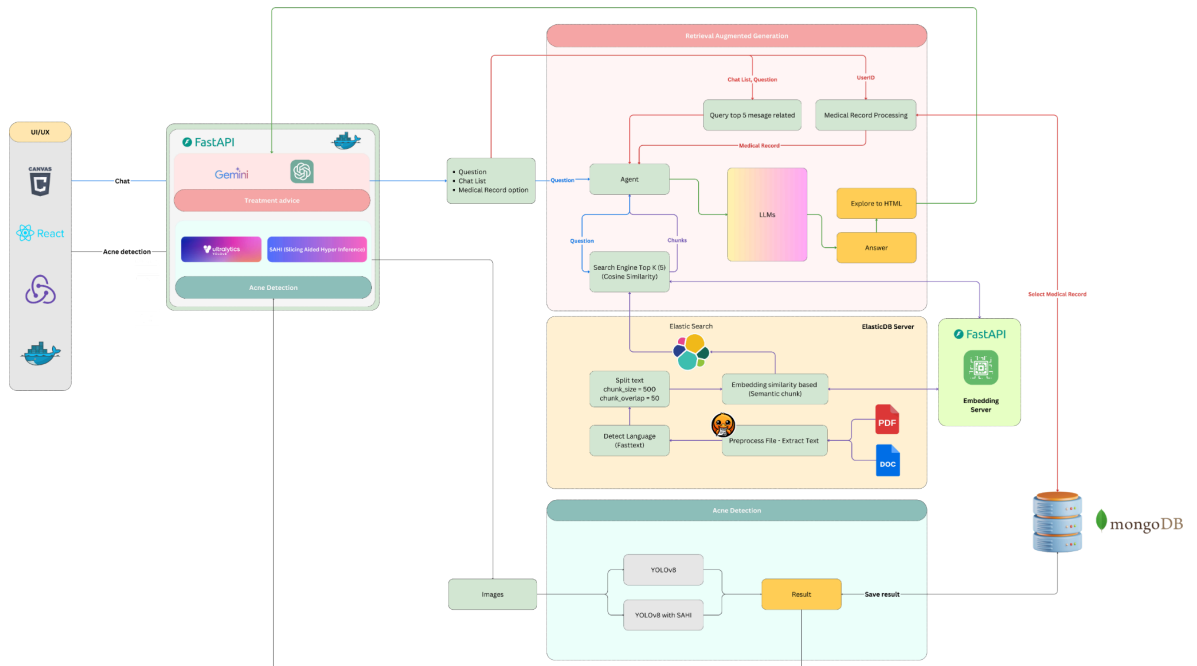
3.2.3 RAG Evaluate (Model as a Judge)

| Compare RAG with Model | Chunk Size | Average Factual Correctness | Average Faithfulness | Average Context Recall | Average Relevancy |
|------------------------|------------|-----------------------------|----------------------|------------------------|-------------------|
| GPT-4o | 512 | 0.949 | 0.888 | 0.548 | 0.765 |
| Gemini 1.5 Pro | | 0.921 | | | |
| Claude 3.5 Sonnet | | 0.946 | | | |
| GPT-4o | 256 | 0.9 | 0.919 | 0.359 | 0.772 |
| Gemini 1.5 Pro | | 0.928 | | | |
| Claude 3.5 Sonnet | | 0.919 | | | |

3.2.4 Chunk Evaluate

CHAPTER 4 SYSTEM IMPLEMENT APPLICATION

4.1 System architecture



REFERENCES

- [1] Hiệp hội Da liễu Việt Nam. (2022). *Thống kê về tình trạng mụn trứng cá ở thanh thiếu niên Việt Nam*. Retrieved from <http://dalieu.org.vn>
- [2] Bộ Y tế Việt Nam. (2021). *Báo cáo về tình trạng da liễu và các vấn đề sức khỏe cộng đồng*. Retrieved from <https://moh.gov.vn>
- [3] Nguyen, T., Le, M., & Tran, P. (2022). *Deep learning applications in dermatology: A systematic review*. Journal of Medical Imaging and Health Informatics. Retrieved from <https://www.jmihi.org>