**Ngày bắt đầu:** 20/5/2024

**Người thực hiện:** Võ Hoài Trọng (Github: HiTrong)

Xây dựng mô hình nhận dạng ung thư phổi thông qua ảnh chụp CT-Scan ở ngực

# Lý do và mục tiêu thực hiện đề tài

Hiện nay trên thế giới ung thư phổi là một trong những nguyên nhân gây tử vong hàng đầu trên thế giới. Ung thư phổi là một căn bệnh nguy hiểm, gây tử vong nhiều nhất trên toàn thế giới. Tại Việt Nam, mỗi năm ghi nhận khoảng 26.262 ca mắc mới ung thư phổi và 23.797 ca tử vong vì căn bệnh này.

Các nguyên nhân phổ biến cho ung thư phổi:

* Hút thuốc lá: là nguyên nhân chính gây ra ung thư phổi. Hiện nay, 90% bệnh nhân bị ung thư phổi do hút thuốc lá, và 4% bệnh nhân hít phải khói thuốc hàng ngày.
* Môi trường làm việc: Các tác nhân từ môi trường như khói bụi, làm việc trong môi trường luyện thép, niken, crom và khí than cũng có thể gây ra bệnh ung thư phổi.
* Tiếp xúc với tia phóng xạ: Đây là nguyên nhân gây ra nhiều bệnh ung thư, trong đó có ung thư phổi. Bệnh nhân làm việc trong các mỏ uranium, fluorspar và hacmatite có thể tiếp xúc với tia phóng xạ do hít thở không khí có chứa khi radon.

Chính vì vậy nếu ta có thể chẩn đoán sớm và chính xác khả năng ung thư phổi có thể giúp cải thiện tỷ lệ sống sót và chất lượng cuộc sống của bệnh nhân. Đó chính là một trong những lí do hàng đầu để đề tài này được thực hiện.

Ngoài ra đề tài còn giúp chúng ta tìm hiểu và ứng dụng thực tế đối với các mô hình học máy và deep learning để phân tích hình ảnh y tế có thể giúp các bác sĩ chẩn đoán bệnh nhanh hơn và chính xác hơn, giảm thiểu sai sót do con người gây ra. Giải quyết được vấn đề thiếu hụt nguồn lực y tế. Ở nhiều nơi, đặc biệt là các khu vực nông thôn và các quốc gia đang phát triển, việc tiếp cận các chuyên gia y tế có thể rất hạn chế. Một hệ thống tự động có thể giúp giảm tải cho các chuyên gia và cung cấp hỗ trợ chẩn đoán ở những nơi thiếu hụt nguồn lực y tế. Tất cả đã làm nên ý nghĩa của đề tài này.

# Tìm hiểu về các loại ung thư phổi

Để có thể xây dựng được một mô hình nhận dạng ung thư phổi thông qua ảnh chụp CT-Scan ở ngực thì với vị trí là đại diễn kĩ thuật ta rất cần hiểu sơ lược về kiến thức y khoa đối với loại bệnh này. Điều này sẽ giúp ích cho chúng ta trong quá trình xây dựng mô hình, kiểm nghiệm, hiểu rõ bản chất hơn. Ung thư phổi phổ biến bao gồm 3 loại: Adenocarcinoma, Large Cell Carcinoma, Squamous Cell Carcinoma.

## Adenocarcinoma – Ung thư biểu mô tuyến

Đây là dạng ung thư phổi phổ biến nhất, chiếm khoảng 30% trong tổng số các trường hợp ung thư phổi và khoảng 40% trong các trường hợp ung thư phổi không phải tế bào nhỏ (NSCLC). Thường được tìm thấy ở vùng ngoại biên của phổi, trong các tuyến tiết ra chất nhầy và giúp chúng ta thở.

## Large Cell Carcinoma – Ung thư tế bào lớn

Loại ung thư phổi này phát triển và lan rộng nhanh chóng, chiếm khoảng 10-15% trong tổng số các trường hợp ung thư phổi không phải tế bào nhỏ (NSCLC). Có thể xuất hiện ở bất kỳ vị trí nào trong phổi.

## Squamous Cell Carcinoma – Ung thư biểu mô tế bào vảy

Chiếm khoảng 30% trong tổng số các trường hợp ung thư phổi không phải tế bào nhỏ (NSCLC). Thường liên quan đến việc hút thuốc lá và được tìm thấy ở trung tâm của phổi, nơi các phế quản lớn nối với khí quản và phổi.

# Tìm hiểu về nền tảng hỗ trợ xây dựng mô hình nhận dạng

Hiện nay với sự phát triển của AI, các hệ thống nhận dạng hình ảnh đã đạt được những tiến bộ vượt bậc. Máy tính đã có thể xử lý và phân tích ảnh, học hỏi từ dữ liệu đó để giải quyết các bài toán bằng Deep Learning. Deep Learning, một nhánh của học máy, đặc biệt mạnh mẽ trong việc trích xuất và học các đặc trưng phức tạp từ dữ liệu hình ảnh.

Các Framework và công cụ phổ biến và mạnh mẽ nhất hiện nay là:

* TensorFlow: Được phát triển bởi Google, TensorFlow là một framework mã nguồn mở mạnh mẽ và linh hoạt, được sử dụng rộng rãi trong cộng đồng AI và học máy. TensorFlow hỗ trợ việc xây dựng và huấn luyện các mô hình deep learning phức tạp.
* Keras: Là một API cấp cao chạy trên nền TensorFlow, Keras giúp đơn giản hóa quá trình xây dựng và huấn luyện các mô hình học sâu. Keras cung cấp các công cụ dễ sử dụng và nhanh chóng, giúp nhà phát triển tập trung vào thiết kế mô hình mà không cần quá lo lắng về các chi tiết kỹ thuật phức tạp.
* PyTorch: Được phát triển bởi Facebook's AI Research lab (FAIR), PyTorch là một framework mã nguồn mở linh hoạt và dễ sử dụng cho các mô hình deep learning. PyTorch hỗ trợ tính toán động (dynamic computation graph), giúp việc debug và phát triển mô hình trở nên dễ dàng hơn.
* OpenCV: Là thư viện mã nguồn mở mạnh mẽ dành cho xử lý ảnh và thị giác máy tính. OpenCV cung cấp nhiều công cụ và thuật toán để tiền xử lý và phân tích ảnh trước khi đưa vào các mô hình deep learning.
* scikit-learn: Là một thư viện học máy mã nguồn mở cho Python, scikit-learn cung cấp các công cụ đơn giản và hiệu quả cho phân tích và khai thác dữ liệu. Mặc dù không chuyên về deep learning, scikit-learn có thể được tích hợp với các framework khác để xây dựng các pipeline học máy phức tạp.
* MXNet: Được phát triển bởi Apache Software Foundation, MXNet là một framework deep learning mã nguồn mở, được tối ưu hóa cho hiệu năng cao và hỗ trợ tính toán phân tán. MXNet cũng hỗ trợ tính toán động và được sử dụng rộng rãi trong nghiên cứu và phát triển AI.
* FastAI: Là một thư viện mã nguồn mở xây dựng trên nền PyTorch, FastAI cung cấp các công cụ đơn giản và mạnh mẽ để xây dựng và huấn luyện các mô hình deep learning. FastAI hướng đến việc làm cho deep learning trở nên dễ tiếp cận hơn cho mọi người, từ người mới bắt đầu đến các chuyên gia.

Sự kết hợp của các framework và công cụ này giúp các nhà nghiên cứu và phát triển có thể xây dựng, huấn luyện và triển khai các mô hình deep learning một cách hiệu quả và nhanh chóng. Chúng cung cấp các công cụ cần thiết để giải quyết các bài toán phức tạp trong nhận dạng hình ảnh và các lĩnh vực khác của AI.

# Các phương pháp và kỹ thuật trong việc xây dựng mô hình

# Xây dựng mô hình

# Đánh giá mô hình và kết luận