

ỨNG DỤNG MÔ HÌNH SWIN TRANSFORMER TRONG HỖ TRỢ CHẨN ĐOÁN BỆNH BỤI PHỔI SILIC NGHỀ NGHIỆP

Nguyễn Thị Tân Tiến^{*,*}, Nguyễn Văn Tảo^{*}, Ngô Hữu Huy^{*}

^{*} Trường Đại học Công nghệ Thông tin và Truyền thông, Đại học Thái Nguyên

^{*} Trường Đại học Y-Dược, Đại học Thái Nguyên

Tóm tắt: Bệnh bụi phổi silic nghề nghiệp là một trong những bệnh lý nghề nghiệp nghiêm trọng, đòi hỏi được phát hiện sớm để nâng cao hiệu quả điều trị. Nghiên cứu này đề xuất sử dụng kiến trúc Swin Transformer nhằm tự động phân loại ảnh X-quang ngực thành hai lớp: bình thường và mắc bệnh bụi phổi silic. Tập dữ liệu gồm 2089 ảnh được thu thập thực tế ở các bệnh viện tại tỉnh Thái Nguyên, sau đó ảnh X-quang gốc được xử lý và sử dụng một số kỹ thuật tăng cường (rotation, horizontal flipping, zooming and brightness adjustment) trước khi đưa vào mô hình. Swin Transformer tận dụng cơ chế attention phân cấp (Hierarchical attention) để trích xuất đặc trưng hình ảnh hiệu quả, phục vụ cho nhiệm vụ phân loại. Kết quả thử nghiệm cho thấy mô hình Swin Transformer đạt hiệu suất vượt trội so với một số kiến trúc học sâu phổ biến (DenseNet201, MobileNetV3-Small). Cụ thể, Swin Transformer đạt độ chính xác 87,77% trên tập dữ liệu kiểm thử, trong khi DenseNet201 và MobileNetV3-Small chỉ đạt lần lượt 74,34% và 76,98%, cho thấy tiềm năng ứng dụng cao của mô hình trong hỗ trợ chẩn đoán sớm bệnh bụi phổi silic từ ảnh X-quang ngực.

Từ khóa: Bụi phổi silic, CNN, học sâu, swin transformer, X-quang ngực.

I. GIỚI THIỆU

Bệnh bụi phổi silic nghề nghiệp (Occupational Silicosis viết tắt là BPSi) tiếp tục là một mối đe dọa sức khỏe cộng đồng dai dẳng và nghiêm trọng trên phạm vi toàn cầu [1]. Căn bệnh tàn phá này phát sinh chủ yếu từ việc phơi nhiễm lâu dài và lặp đi lặp lại với bụi silic tinh thể tự do, phổ biến trong các ngành công nghiệp trọng yếu như khai thác mỏ, xây dựng, sản xuất vật liệu chịu lửa và công nghiệp thủy tinh [2, 3]. Cơ chế bệnh sinh đặc trưng bởi sự lắng đọng của các hạt silic siêu mịn trong nhu mô phổi, kích hoạt phản ứng viêm mãn tính và dần tiến triển thành xơ hóa (fibrosis) lan tỏa, dẫn đến tổn thương nhu mô phổi không hồi phục. Hậu quả của bệnh bụi phổi silic là rất nặng nề, từ suy giảm chức năng hô hấp nghiêm trọng đến các biến chứng phức tạp. Với tiên lượng xấu và khả năng gây tử vong nếu không được can thiệp kịp thời, đặc biệt ở những người lao động tiếp xúc nghề nghiệp kéo dài [4], việc phát hiện sớm trở nên quan trọng. Chẩn đoán và

điều trị kịp thời không chỉ cải thiện đáng kể chất lượng cuộc sống của người bệnh mà còn là chìa khóa để kiểm soát sự tiến triển của bệnh, ngăn ngừa biến chứng nguy hiểm và giảm thiểu tỷ lệ tử vong liên quan [5, 6]. Theo truyền thống, chẩn đoán bệnh bụi phổi silic phụ thuộc chủ yếu vào việc diễn giải hình ảnh X-quang ngực kết hợp với đánh giá lâm sàng và tiền sử nghề nghiệp. Tuy nhiên, phương pháp này đối mặt với những thách thức. Sự đa dạng, đôi khi không điển hình của các biểu hiện tổn thương trên ảnh X-quang đòi hỏi chuyên môn sâu và kinh nghiệm dày dặn từ các bác sĩ chẩn đoán hình ảnh. Điều này làm chậm quá trình chẩn đoán chính xác, đặc biệt nghiêm trọng ở các khu vực thiếu hụt bác sĩ chuyên khoa phổi hoặc chẩn đoán hình ảnh [7]. Hơn nữa, tính chủ quan vốn có trong quá trình diễn giải hình ảnh dẫn đến nguy cơ biến thiên kết quả giữa các chuyên gia khác nhau (inter-rater variability), ảnh hưởng đến tính đồng nhất và tin cậy của chẩn đoán [8].

Trong thập kỷ qua, công nghệ học sâu (deep learning) đã cách mạng hóa lĩnh vực phân tích ảnh y tế. Các kiến trúc mạng nơ-ron tích chập (CNN) đã chứng tỏ khả năng vượt trội trong việc tự động trích xuất đặc trưng từ ảnh X-quang và hỗ trợ chẩn đoán nhiều bệnh lý phổi khác nhau [9]. Tuy nhiên, đối với những thách thức đặc thù của ảnh X-quang BPSi – nơi đặc trưng bệnh lý có thể mờ nhạt, rải rác và đòi hỏi hiểu biết ngữ cảnh không gian rộng lớn của phổi – các mô hình CNN truyền thống với cơ chế tích chập cục bộ vẫn bộc lộ những hạn chế nhất định trong việc nắm bắt mối quan hệ xa giữa các vùng tổn thương và tổng quát hóa hiệu quả các đặc trưng hình thái phức tạp.

Sự xuất hiện của kiến trúc Transformer trong lĩnh vực thị giác máy tính đã mở ra một kỷ nguyên mới cho việc mô hình hóa các mối quan hệ phụ thuộc dài hạn trên dữ liệu hình ảnh [10, 11]. Mặc dù Vision Transformer (ViT) áp dụng cơ chế self-attention toàn cục, nó thường yêu cầu lượng dữ liệu huấn luyện khổng lồ và gặp vấn đề về hiệu quả tính toán trên ảnh độ phân giải cao [12]. Để giải quyết những thách thức này, Swin Transformer (Shifted Windows Transformer) đã được phát triển, kết hợp một cách khéo léo cấu trúc phân cấp của CNN với cơ chế self-attention theo cửa sổ dịch chuyển [13]. Sự lai ghép này giúp Swin Transformer vừa học được đặc trưng cục bộ chi tiết, vừa nắm bắt được ngữ cảnh toàn cục một cách hiệu quả về mặt tính toán, trở thành một kiến trúc đầy tiềm năng cho phân tích ảnh y tế, đặc biệt là các bài toán đòi hỏi hiểu biết đa tỷ lệ và quan hệ không gian phức tạp như chẩn đoán BPSi [14, 15].

Nghiên cứu này tập trung khai thác sức mạnh của Swin Transformer để phát triển một hệ thống hỗ trợ chẩn đoán

Tác giả liên hệ: Ngô Hữu Huy

Email: nhuy@ictu.edu.vn

Đến tòa soạn: 04/2025, chỉnh sửa: 05/2025, chấp nhận đăng: 06/2025

BPSi từ ảnh X-quang ngực. Bằng cách tận dụng khả năng học biểu diễn phân cấp và mô hình hóa quan hệ không gian hiệu quả của Swin Transformer, chúng tôi hướng tới việc xây dựng một mô hình có khả năng phân loại chính xác ảnh X-quang ngực thành nhóm bình thường và nhóm mắc BPSi, góp phần cải thiện quy trình sàng lọc, phát hiện sớm và quản lý bệnh cho người lao động.

II. CÁC NGHIÊN CỨU LIÊN QUAN

A. Phương pháp chẩn đoán bụi phổi Silic bởi chuyên gia

Chẩn đoán hình ảnh dựa trên X-quang ngực (CXR) từ lâu đã là phương tiện chẩn đoán hình ảnh tiêu chuẩn vàng để phát hiện và đánh giá các tổn thương nhu mô phổi, bao gồm cả các biểu hiện đặc trưng của bệnh bụi phổi silic như các đám mờ nhỏ và lớn. Phương pháp này không chỉ đóng vai trò nền tảng trong chẩn đoán ban đầu mà còn hỗ trợ phân loại giai đoạn bệnh theo các hệ thống tiêu chuẩn [16]. Phương pháp này thường dựa vào kinh nghiệm của bác sĩ. Tuy nhiên, phương pháp này phụ thuộc rất lớn vào kinh nghiệm chuyên môn và kỹ năng diễn giải hình ảnh của bác sĩ chẩn đoán. Sự phức tạp và biến thiên trong biểu hiện X-quang của bệnh bụi phổi silic đặt ra thách thức đáng kể, dẫn đến quá trình đánh giá tốn nhiều thời gian và có thể gây ra sự biến thiên kết quả đáng kể giữa các bác sĩ hoặc chuyên gia khác nhau (Inter-observer variability) [17, 18]. Điều này đặc biệt khó khăn trong bối cảnh cần chẩn đoán nhanh chóng và chính xác, cũng như tại các khu vực thiếu hụt nguồn nhân lực y tế chuyên sâu. Trước những hạn chế cố hữu của diễn giải hình ảnh chủ quan, các hệ thống hỗ trợ chẩn đoán bằng máy tính (Computer-Aided Diagnosis - CAD) đã được phát triển như những nỗ lực ban đầu nhằm tự động hóa một phần quy trình phân tích ảnh X-quang [19]. Các hệ thống CAD sử dụng các thuật toán xử lý ảnh và học máy truyền thống để trích xuất các đặc trưng được thiết kế thủ công từ hình ảnh, cung cấp thông tin bổ trợ giúp bác sĩ giảm bớt thời gian và tăng cường tính nhất quán trong đánh giá [20, 21]. Bên cạnh CAD, các hệ thống hỗ trợ quyết định lâm sàng (Clinical Decision Support Systems - CDSS) tích hợp thông tin hình ảnh với dữ liệu lâm sàng của bệnh nhân để đưa ra khuyến nghị toàn diện hơn [22], cùng với các hệ thống báo cáo chuẩn hóa nhằm giảm thiểu sự khác biệt trong diễn giải [23].

B. Phương pháp hỗ trợ chẩn đoán bụi phổi Silic dựa trên các mô hình học sâu

Sự ra đời của học sâu đã mang lại một bước tiến lớn trong phân tích ảnh y tế. Các kiến trúc CNN, với khả năng tự động học các bộ lọc đặc trưng từ dữ liệu, đã được áp dụng rộng rãi để phân loại và phát hiện các bệnh lý trên ảnh X-quang ngực [9]. Đối với BPSi, nhiều nghiên cứu đã sử dụng các mô hình CNN như ResNet và DenseNet để tự động nhận diện các nốt mờ silic [12, 24]. Những mô hình này đã đạt được những kết quả ban đầu đầy hứa hẹn trong việc phân biệt giữa ảnh bình thường và ảnh có dấu hiệu BPSi.

Tuy nhiên, các mô hình CNN truyền thống thường tập trung vào việc trích xuất đặc trưng cục bộ thông qua các lớp tích chập, điều này có thể không đủ hiệu quả để nắm bắt toàn bộ bối cảnh không gian và mối quan hệ giữa các tổn thương rải rác trên ảnh X-quang BPSi, đặc biệt là trong các trường hợp bệnh phức tạp hoặc ở giai đoạn sớm khi các nốt mờ chưa rõ ràng và có thể bị ảnh hưởng bởi cấu trúc giải phẫu lân cận. Hơn nữa, việc huấn luyện các mô hình CNN từ đầu thường đòi hỏi một lượng lớn dữ liệu

được gán nhãn chất lượng cao, điều này có thể là một thách thức đối với các bộ dữ liệu y tế chuyên biệt như ảnh X-quang BPSi [25].

C. Ứng dụng mô hình Swin Transformer trong y tế

Kiến trúc Transformer, ban đầu được phát triển cho các bài toán xử lý ngôn ngữ tự nhiên dựa trên cơ chế self-attention, đã tạo ra một làn sóng mới trong lĩnh vực thị giác máy tính. Vision Transformer (ViT) là mô hình tiên phong chứng minh rằng một kiến trúc chỉ dựa vào self-attention, không cần tích chập, vẫn có thể đạt hiệu suất cạnh tranh với CNN trong các tác vụ phân loại ảnh [12]. ViT chia ảnh đầu vào thành các patch cố định và áp dụng self-attention toàn cục để mô hình hóa mối quan hệ giữa tất cả các patch. Khả năng nắm bắt phụ thuộc dài hạn này là một ưu điểm lớn so với CNN. Tuy nhiên, ViT có nhược điểm là yêu cầu tập dữ liệu huấn luyện rất lớn để có thể khái quát hóa tốt và chi phí tính toán cho self-attention toàn cục tăng theo bình phương số lượng patch, làm cho nó kém hiệu quả khi xử lý ảnh độ phân giải cao [12, 26].

Để giải quyết những hạn chế của ViT, Swin Transformer đã được giới thiệu. Kiến trúc này tích hợp một cách thông minh cơ chế self-attention theo cửa sổ (window-based self-attention) và chiến lược dịch chuyển cửa sổ (shifted window) để giảm chi phí tính toán xuống còn tuyến tính với số lượng patch, đồng thời vẫn cho phép thông tin được truyền qua các ranh giới cửa sổ, mở rộng trường tiếp nhận theo chiều sâu [13, 26]. Đặc biệt quan trọng đối với ảnh y tế, Swin Transformer còn xây dựng biểu diễn đặc trưng theo cấu trúc phân cấp, tương tự như các kiến trúc CNN hiệu quả [12, 15]. Sự kết hợp giữa self-attention hiệu quả, cấu trúc phân cấp và khả năng mô hình hóa cả đặc trưng cục bộ lẫn ngữ cảnh toàn cục làm cho Swin Transformer trở nên rất phù hợp cho các bài toán phân tích ảnh y tế, bao gồm cả ảnh X-quang ngực. Các nghiên cứu gần đây đã chứng minh hiệu quả của Swin Transformer trong phân loại bệnh phổi, bao gồm cả lao phổi và các nốt phổi [13, 14, 15]. Điều này cho thấy tiềm năng lớn của Swin Transformer trong việc giải quyết những thách thức đặc thù của chẩn đoán BPSi.

Bên cạnh việc ứng dụng các kiến trúc học sâu đơn lẻ, xu hướng hiện nay còn hướng tới việc kết hợp các mô hình khác nhau hoặc tích hợp các cơ chế bổ sung để nâng cao hiệu suất hỗ trợ chẩn đoán. Việc kết hợp CNN và Transformer đã được đề xuất để tận dụng ưu điểm của cả hai. Hơn nữa, các kỹ thuật Few-Shot Learning đang được nghiên cứu để giải quyết bài toán thiếu dữ liệu trong các ứng dụng y tế chuyên biệt như BPSi [25]. Một hướng phát triển đầy hứa hẹn khác là việc kết hợp các mô hình dựa trên ảnh với các mô hình dựa trên đồ thị (Graph-based models), như Graph Transformer Post-hoc Modeling, để mô hình hóa mối quan hệ ngữ nghĩa giữa các ảnh hoặc các vùng tổn thương, từ đó cung cấp cái nhìn tổng thể và tăng cường khả năng hỗ trợ chẩn đoán [27]. Những hướng nghiên cứu này mở ra tiềm năng lớn để xây dựng các hệ thống hỗ trợ chẩn đoán BPSi toàn diện và hiệu quả hơn trong tương lai.

III. BỘ DỮ LIỆU

A. Thu thập dữ liệu

Quá trình thu thập, xây dựng bộ dữ liệu BSIPTIT được thực hiện một cách khoa học, tập trung chủ yếu vào các hình ảnh X-quang ngực - phương pháp tiêu chuẩn vàng

trong chẩn đoán tổn thương nhu mô phổi liên quan đến bệnh bụi phổi silic. Bộ dữ liệu này được chia thành hai nhóm chính dựa trên đặc điểm bệnh lý rõ ràng

Nhóm mắc bệnh bụi phổi silic: Bao gồm những hình ảnh X-quang ngực thể hiện các dấu hiệu đặc trưng của bệnh bụi phổi silic như các nốt mờ nhỏ (small opacities), đám mờ lớn (large opacities), dày thành phế quản (bronchial wall thickening), và phù tổ chức kẽ (interstitial pattern).

Nhóm bình thường: Bao gồm những hình ảnh X-quang ngực không thể hiện bất kỳ dấu hiệu tổn thương phổi nào liên quan đến bệnh bụi phổi silic

Để đảm bảo tính chính xác và độ tin cậy cao nhất, toàn bộ dữ liệu hình ảnh đã trải qua quá trình gắn nhãn (Annotation) tỉ mỉ bởi các bác sĩ chuyên gia trong lĩnh vực y học lao động và chẩn đoán hình ảnh.

Quy trình chẩn đoán và gắn nhãn tuân thủ nghiêm ngặt các hướng dẫn quốc tế và quốc gia, cụ thể là Hướng dẫn của Tổ chức Lao động Quốc tế (ILO) năm 2011 về phân loại X-quang bệnh bụi phổi và Thông tư 15/2016/TT-BYT của Bộ Y tế Việt Nam. Chỉ những trường hợp có kết quả chẩn đoán rõ ràng và hình ảnh chất lượng cao mới được đưa vào bộ dữ liệu; các trường hợp không điển hình, không rõ ràng hoặc hình ảnh bị nhiễu nặng đã được loại bỏ để duy trì tính toàn vẹn của tập dữ liệu.

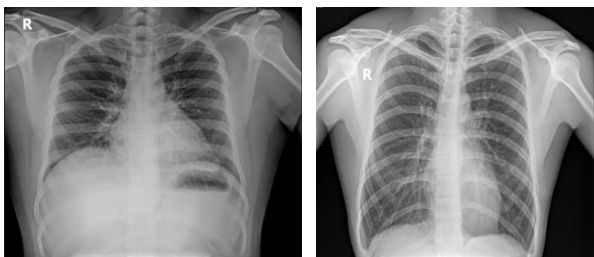
- **Đối tượng:** Bộ dữ liệu thu thập trực tiếp từ 2089 công nhân có nguy cơ cao mắc bệnh bụi phổi silic do tiếp xúc thường xuyên với bụi silic trong các ngành công nghiệp như khai thác mỏ, xây dựng và sản xuất.

- **Địa điểm:** Các hình ảnh được thu thập từ các bệnh viện chuyên khoa ở các tỉnh như Thái Nguyên, Hải Dương, Bình Định, Phú Yên, và Đồng Nai.

- **Tiêu chuẩn chẩn đoán:** Tất cả các trường hợp bệnh bụi phổi silic đều được xác nhận theo tiêu chuẩn của ILO 2000 để đảm bảo tính nhất quán và độ tin cậy. Các hình ảnh X-quang được chụp với kích thước tiêu chuẩn là 35 x 43 cm.

- **Số lượng mẫu:** Từ 2089 người lao động này, có 535 trường hợp được chẩn đoán mắc bệnh bụi phổi silic, với hình ảnh X-quang ngực của họ được đưa vào lớp Silicosis trong bộ dữ liệu.

Hình 1 và 2 minh họa hai ví dụ điển hình về hình ảnh X-quang ngực trong bộ dữ liệu, đại diện cho lớp ‘Bình thường’ và ‘lớp Bụi phổi silic’, làm nổi bật sự khác biệt về mặt hình ảnh giữa hai nhóm.



Hình 1. Hình ảnh X-quang ngực của người không mắc bệnh



Hình 2. Hình ảnh X-quang ngực của người mắc bệnh bụi phổi silic nghề nghiệp

B. Tiền xử lý dữ liệu và tăng cường dữ liệu

Tiền xử lý dữ liệu: Đây là một bước không thể thiếu nhằm chuyển đổi dữ liệu thô sang định dạng phù hợp và tối ưu cho việc huấn luyện mô hình học sâu.

Định dạng dữ liệu: Dữ liệu hình ảnh ban đầu được lưu trữ ở định dạng DICOM (.dcm), là định dạng chuẩn trong y tế, chứa đầy đủ thông tin về hình ảnh và metadata liên quan. Mặc dù DICOM lý tưởng cho lưu trữ và xem bằng phần mềm chuyên dụng (như MicroDicom, RadiAnt), việc xử lý trực tiếp trong các framework học sâu thường yêu cầu định dạng hình ảnh phổ biến hơn như JPG hoặc PNG.

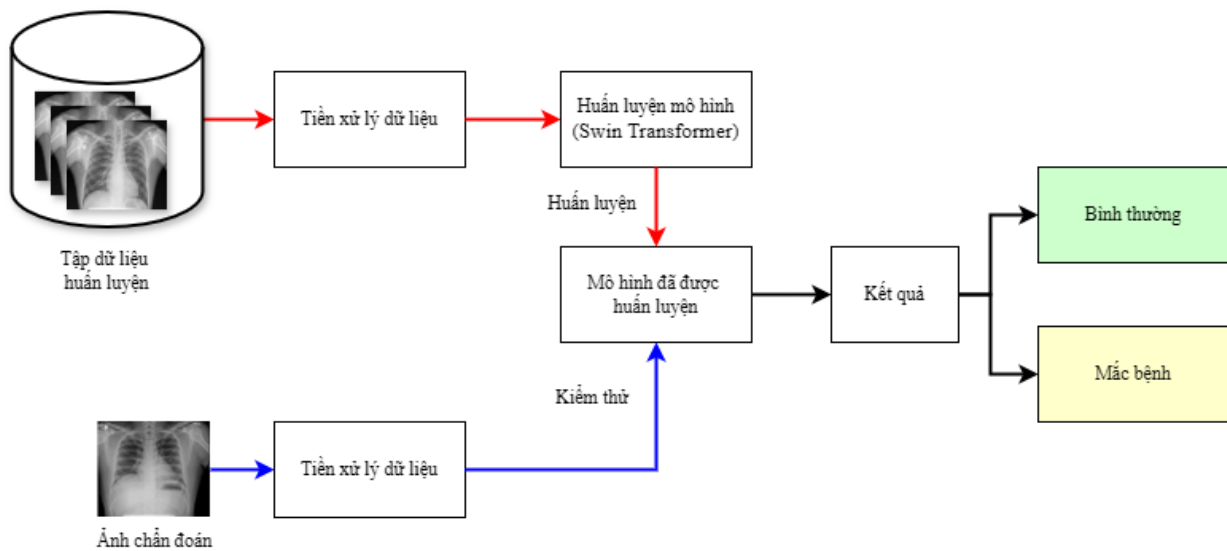
Quy trình xử lý: Để chuẩn hóa và tự động hóa quy trình này, chúng tôi đã phát triển một script Python để chuyển đổi hàng loạt các tệp DICOM sang định dạng JPG. Sau khi chuyển đổi, tất cả các hình ảnh được chuẩn hóa về kích thước đồng nhất là $224 \times 224 \times 3$ (chiều rộng $\times$ chiều cao $\times$ số kênh màu) để phù hợp với yêu cầu đầu vào của kiến trúc DenseNet được sử dụng trong nghiên cứu này. Để giảm độ phức tạp tính toán và nâng cao hiệu suất mô hình, các giá trị điểm ảnh được chuẩn hóa bằng cách chia cho 255. Việc này giúp đưa các giá trị về cùng một thang chuẩn trong khoảng $[0, 1]$, từ đó cải thiện tính ổn định và hiệu quả trong quá trình huấn luyện và kiểm thử mô hình.

Tổng quan bộ dữ liệu: Sau quá trình tiền xử lý, bộ dữ liệu cuối cùng bao gồm 2089 hình ảnh X-quang, với phân bố cụ thể là:

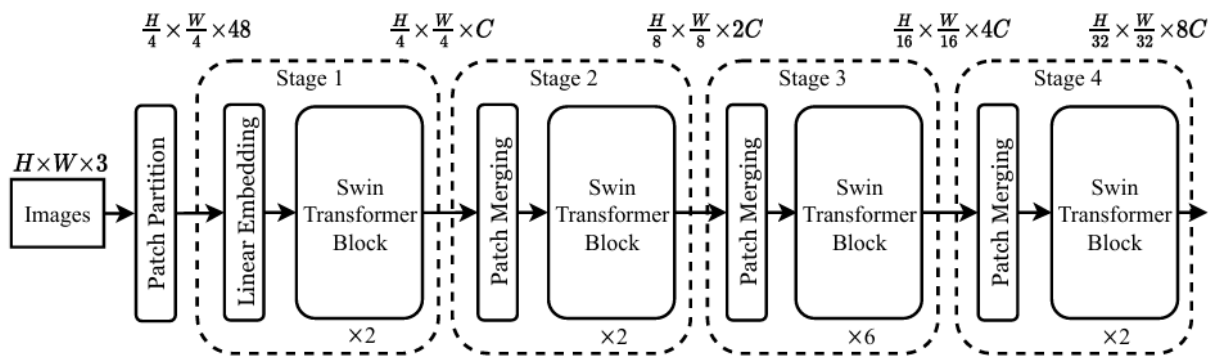
- 535 hình ảnh của bệnh nhân mắc bệnh silicosis.
- 1554 hình ảnh của người không mắc bệnh.

Để nâng cao khả năng tổng quát hóa (Generalization) và giảm thiểu hiện tượng quá khớp (Overfitting), đặc biệt khi làm việc với tập dữ liệu có kích thước hạn chế, kỹ thuật tăng cường dữ liệu (Data augmentation) đã được áp dụng. Các phép biến đổi ngẫu nhiên được thực hiện trên các hình ảnh huấn luyện bao gồm: xoay ảnh (Rotation), lật ngang (Horizontal flipping), phóng to (Zooming), và thay đổi độ sáng (Brightness adjustment). Bằng cách tạo ra nhiều biến thể của cùng một hình ảnh gốc, tăng cường dữ liệu giúp mở rộng không gian mẫu, buộc mô hình phải học các đặc trưng mạnh mẽ và bất biến hơn đối với các biến đổi nhỏ, từ đó cải thiện đáng kể độ chính xác và khả năng áp dụng của mô hình trên dữ liệu mới. Cấu hình tham số cho các phép biến đổi dữ liệu cụ thể như sau:

- **Xoay ảnh (Rotation):** Ảnh X-quang được xoay ngẫu nhiên trong khoảng từ -10° đến $+10^\circ$, nhằm mô phỏng sự thay đổi nhẹ trong tư thế chụp mà không làm biến dạng cấu trúc giải phẫu.



Hình 3. Mô hình tổng quan



Hình 4. Kiến trúc mô hình Swin Transformer

- Lật ngang (Horizontal Flipping): Ảnh có thể được lật đối xứng từ trái sang phải hoặc ngược lại, giúp mô hình học được đặc trưng bệnh lý ở cả hai bên phổi và giảm sự phụ thuộc vào hướng ảnh gốc.

- Phóng to (Zooming): Phép biến đổi phóng to được áp dụng ngẫu nhiên trong phạm vi lên đến 20%, giúp mô phỏng các trường hợp ảnh chụp ở khoảng cách khác nhau, đồng thời tăng tính đa dạng cho dữ liệu huấn luyện.

- Điều chỉnh độ sáng (Brightness Adjustment): Độ sáng của ảnh được thay đổi ngẫu nhiên trong khoảng $\pm 20\%$ nhằm phản ánh các điều kiện chiếu sáng khác nhau trong quá trình thu nhận ảnh, đồng thời giúp mô hình trở nên linh hoạt hơn trước sự thay đổi ánh sáng.

IV. ÁP DỤNG MÔ HÌNH SWIN TRANSFORMER PHÁT HIỆN BỆNH BỤI PHỔI SILIC NGHỀ NGHIỆP

A. Mô hình tổng quan

Hình 3 minh họa mô hình tổng quan sử dụng kiến trúc Swin Transformer để hỗ trợ phát hiện bệnh bụi phổi silic nghề nghiệp từ hình ảnh X-quang ngực.

Input: Lô ảnh X-quang ngực trong tập dữ liệu được chia thành hai nhóm: Nhóm bình thường, nhóm mắc bệnh bụi phổi silic nghề nghiệp.

Output: Mô hình dự đoán nhãn phân loại tương ứng cho từng hình ảnh (0 - bình thường, 1 - mắc bệnh).

Trong nghiên cứu này, Swin Transformer được áp dụng vào bài toán chẩn đoán bệnh bụi phổi silic nghề nghiệp theo một quy trình bốn giai đoạn: tiền xử lý dữ liệu, xây dựng kiến trúc mô hình, cấu hình huấn luyện và đánh giá hiệu suất. Ảnh X-quang ngực được chuyển đổi từ định dạng DICOM sang PNG/JPEG, resize về kích thước 224×224 pixel, và áp dụng kỹ thuật tăng cường dữ liệu để giảm thiểu mất cân bằng lớp. Dữ liệu sau xử lý được đưa vào kiến trúc Swin Transformer phân cấp, với các tầng Window-based và Shifted Window-based Multi-head Self-Attention nhằm trích xuất các đặc trưng cục bộ và toàn cục hiệu quả. Lớp phân loại Dense ở đầu ra thực hiện nhiệm vụ phân biệt hai lớp: bình thường và mắc bệnh bụi phổi silic.

B. Kiến trúc mô hình Swin Transformer

Kiến trúc mô hình được áp dụng trong nghiên cứu này là Swin Transformer, một mạng nơ-ron tiên tiến kết hợp hiệu quả cơ chế Self-Attention theo cửa sổ cục bộ với cấu trúc phân cấp (Hierarchical). Việc lựa chọn kiến trúc này nhằm tận dụng khả năng học biểu diễn đa tỷ lệ và mô hình hóa quan hệ không gian hiệu quả của Transformer, vốn rất quan trọng trong phân tích ảnh X-quang ngực để phát hiện bệnh bụi phổi silic nghề nghiệp.

Mô hình xử lý ảnh X-quang đầu vào qua nhiều giai đoạn. Ban đầu, dữ liệu được đưa qua lớp tích chập để trích xuất đặc trưng cấp thấp và chia thành các patch. Tiếp theo, các giai đoạn của Swin Transformer Blocks và Patch Merging layers lần lượt thực hiện học biểu diễn ngữ cảnh cục bộ (sử dụng Window-based Self-Attention - W-MSA và Shifted Window-based Self-Attention - SW-MSA) và giảm dần kích thước không gian trong khi tăng số chiều đặc trưng, cho phép mô hình nắm bắt các đặc trưng ở các mức độ trừu tượng khác nhau. Sau cùng, các đặc trưng được tổng hợp và đưa qua lớp kết nối đầy đủ để thực hiện phân loại nhị phân (Bình thường/Mắc bệnh).

Chi tiết về cấu trúc mạng, bao gồm trình tự các lớp, vai trò chức năng và kích thước tensor đầu ra tại mỗi bước, được trình bày cụ thể trong Hình 4. Sự kết hợp của các cơ chế này giúp mô hình Swin Transformer đạt được hiệu suất cao trong việc phân tích các đặc trưng tinh tế và phân bố không đồng nhất trên ảnh X-quang BPSi, đồng thời tối ưu hóa hiệu quả tính toán so với các kiến trúc Transformer toàn cục.

V. KẾT QUẢ THỰC NGHIỆM

A. Thiết lập thực nghiệm

Các thí nghiệm được tiến hành trên nền tảng phần cứng mạnh mẽ bao gồm GPU NVIDIA Tesla T4 và 15GB RAM. Việc xây dựng và huấn luyện mô hình được thực hiện hiệu quả bằng thư viện học máy PyTorch.

Quá trình huấn luyện sử dụng Adam Optimizer với learning rate ban đầu 0.0001, hàm mất mát Cross-Entropy và kỹ thuật early stopping nhằm tối ưu hóa tốc độ hội tụ và hạn chế overfitting. Sau huấn luyện, mô hình được đánh giá trên tập kiểm tra dựa trên các chỉ số Accuracy, Precision, Recall, F1-score và F1-marco, đồng thời phân tích thêm thông qua ma trận nhầm lẫn (Confusion Matrix) và trực quan hóa Grad-CAM. Quy trình này đảm bảo mô hình không chỉ đạt hiệu suất cao về mặt định lượng mà còn có khả năng giải thích tốt đối với các quyết định chẩn đoán dựa trên hình ảnh.

Mô hình được huấn luyện trong 50 epoch đầy đủ, với kích thước batch cố định là 32 mẫu trên mỗi lần lặp, đảm bảo quá trình học diễn ra ổn định.

Tập dữ liệu nghiên cứu bao gồm hai lớp: Bình thường và Bụi phổi silic. Tập dữ liệu được chia thành tập huấn luyện (80%) và tập kiểm tra (20%) để đánh giá độc lập hiệu năng mô hình. Chi tiết về số lượng mẫu trong từng tập trước khi tăng trưởng và sau khi tăng cường được trình bày cụ thể trong Bảng I.

Bảng I. Chi tiết bộ dữ liệu

Dữ liệu	Bụi phổi Silic	Bình thường	Tổng số
Tập huấn luyện	428	1244	1672
Tập kiểm tra	107	310	417
Tổng số	535	1554	2089

B. Kết quả

1) Thang đo đánh giá

Tiến hành đánh giá hiệu suất của Swin Transformer, mô hình được so sánh với hai kiến trúc học sâu phổ biến

là DenseNet201 và MobileNetV3-Small được huấn luyện trên bộ dữ liệu X-quang ngực thu thập từ bệnh nhân mắc bệnh bụi phổi silic và người bình thường. Hai nhóm chính được phân loại gồm: Bình thường (Normal) được gán nhãn là 0 và Bụi phổi silic (Silicosis) được gán nhãn là 1. Các thước đo hiệu suất chính được tính toán dựa trên các công thức dưới đây:

Accuracy (Độ chính xác tổng quát):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision (Độ chính xác):

$$Độ\ chính\ xác = \frac{TP}{TP + FP}$$

Recall (Độ nhạy):

$$Độ\ nhạy = \frac{TP}{TP + FN}$$

F1-score:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

F1-marco:

$$F1 - macro = \frac{1}{N} \sum_{i=1}^N F1_i$$

Trong đó:

TP (True Positive): Số lượng dự đoán đúng cho lớp bệnh bụi phổi silic.

TN (True Negative): Số lượng dự đoán đúng cho lớp bình thường.

FP (False Positive): Số lượng dự đoán sai cho lớp bình thường thành bụi phổi silic.

FN (False Negative): Số lượng dự đoán sai cho lớp bụi phổi silic thành bình thường.

N: Số lượng lớp.

Những thước đo này được áp dụng nhằm đánh giá toàn diện hiệu suất phân loại của từng mô hình trên tập dữ liệu kiểm tra (test set). Kết quả đạt được giúp làm rõ mức độ hiệu quả và độ chính xác của các kiến trúc học sâu khi áp dụng vào bài toán phát hiện bệnh bụi phổi silic từ ảnh X-quang ngực.

2) So sánh độ chính xác giữa các mô hình

Để kiểm chứng hiệu quả của Swin Transformer, mô hình được so sánh với hai kiến trúc học sâu phổ biến là DenseNet201 và MobileNetV3-Small. Các mô hình này đều được huấn luyện và đánh giá trên cùng một tập dữ liệu ảnh X-quang ngực đã tiền xử lý, với tập huấn luyện chiếm 80% và tập kiểm tra 20%.

Kết quả từ Bảng II cho thấy Swin Transformer vượt trội hơn các mô hình như DenseNet201 và MobileNetV3-Small với độ chính xác đạt 87,77% trên tập dữ liệu kiểm thử. Đặc biệt, mô hình Swin Transformer đạt giá trị F1-macro cao nhất – một chỉ số quan trọng trong bài toán mất cân bằng dữ liệu, phản ánh khả năng cân bằng giữa độ nhạy

và độ chính xác. Điều này cho thấy Swin Transformer không chỉ học tốt mà còn tổng quát hóa hiệu quả trên dữ liệu chưa từng thấy.

Bảng II. Độ chính xác giữa Swin Transformer và các mô hình

Mô hình	Accuracy		F1-macro
	Huấn luyện	Kiểm thử	
MobileNetV3-Small	73,29	74,34	42,64
DenseNet201	78,88	76,98	57,32
Swin Transformer	95,36	87,77	84,64

3) Phân tích ma trận nhầm lẫn

Bảng III thể hiện ma trận nhầm lẫn (Confusion Matrix) của mô hình Swin Transformer khi dự đoán các hình ảnh X-quang ngực thuộc hai lớp: "Bụi phổi silic" và "Bình thường".

Swin Transformer cho thấy khả năng nhận diện tốt các Mô hình nhận diện đúng 89 ca mắc bệnh và 277 ca bình thường. Số trường hợp bị bỏ sót bệnh (False Negative) là 18, chiếm tỷ lệ đáng kể, ảnh hưởng đến độ nhạy. Số trường hợp bình thường bị chẩn đoán sai thành bệnh (False Positive) là 33, tuy ít hơn nhưng vẫn có thể ảnh hưởng đến độ tin cậy lâm sàng.

Mặc dù tỷ lệ FN vẫn tồn tại, nhưng độ chính xác tổng thể và độ chính xác trong nhóm bình thường (TN) là rất cao, khẳng định mô hình học được đặc trưng bệnh một cách đáng tin cậy.

4) Trực quan hóa vùng đặc trưng trong chẩn đoán bụi phổi silic

Grad-CAM (Gradient-weighted Class Activation Mapping) là một phương pháp hữu hiệu trong việc trực quan hóa các vùng mà mô hình học sâu tập trung khi đưa ra dự đoán [28]. Công cụ này không chỉ giúp đánh giá độ tin cậy của mô hình mà còn cung cấp thông tin chi tiết để cải thiện dữ liệu hoặc tối ưu hóa cấu trúc mô hình.

Hình ảnh Grad-CAM trong Hình 5 minh họa các vùng tập trung của từng mô hình khi phân loại. Hình 5 cho thấy Swin Transformer tập trung mạnh vào vùng phổi, đặc biệt

Bảng III. Ma trận nhầm lẫn của mô hình Swin Transformer

		Dự đoán	
		Bụi phổi silic	Bình thường
Thực tế	Bụi phổi silic	89	18
	Bình thường	33	277

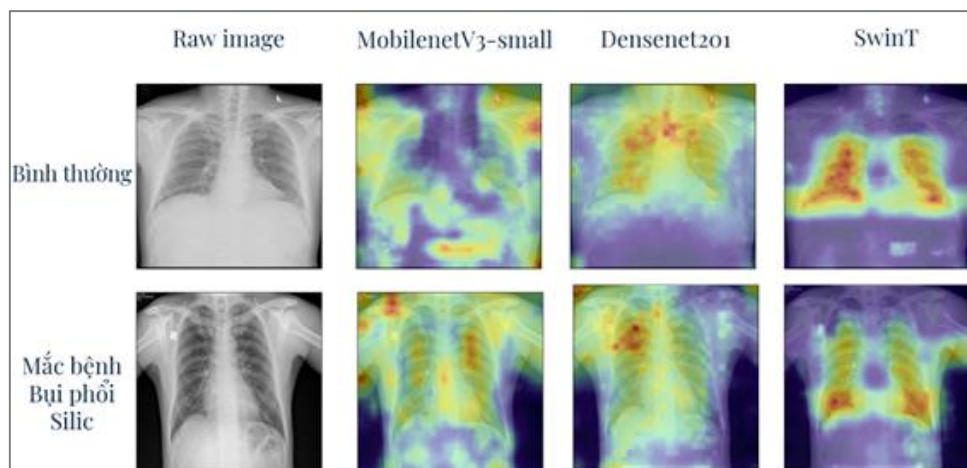
tại các vị trí tổn thương dạng nốt hoặc xơ hóa, giúp mô hình đưa ra quyết định chính xác. DenseNet201 và MobileNetV3-Small có xu hướng phân tán sự chú ý ra khỏi vùng tổn thương, khiến chất lượng phân loại giảm.

Trên các ảnh bình thường, Swin Transformer vẫn duy trì sự tập trung chủ yếu vào cấu trúc phổi, tránh nhận diện sai vào các vùng không liên quan (như xương sườn, mô mềm ngoài phổi...). Những kết quả này chứng minh rằng cơ chế attention trong Swin Transformer không chỉ mang lại hiệu suất tốt về định lượng, mà còn đảm bảo tính hợp lý về mặt ngữ nghĩa và cấu trúc ảnh y học.

Tóm lại, Grad-CAM đã chỉ ra rằng Swin Transformer không chỉ có khả năng học đặc trưng phức tạp hơn mà còn có khả năng tập trung chính xác vào các vùng tổn thương đặc thù của bệnh bụi phổi silic, điều mà DenseNet201 và MobileNetV3-Small không đạt được. Sự khác biệt rõ ràng trong việc xác định các vùng đặc trưng quan trọng khẳng định ưu thế nổi trội của Swin Transformer, đặc biệt trong các ứng dụng y khoa cần độ chính xác cao. Kết quả này củng cố tiềm năng của Swin Transformer trong việc hỗ trợ các bác sĩ đưa ra chẩn đoán chính xác và đáng tin cậy hơn.

VI. KẾT LUẬN

Trong nghiên cứu này, Swin Transformer đã được chứng minh là một kiến trúc học sâu hiệu quả trong nhiệm vụ phân loại ảnh X-quang ngực nhằm phát hiện bệnh bụi phổi silic nghề nghiệp. Nhờ thiết kế Self-Attention cục bộ với cơ chế cửa sổ trượt và cấu trúc phân cấp đặc trưng, mô hình đạt được F1-macro 84,64%, vượt trội so với các mạng CNN truyền thống như DenseNet201 và MobileNetV3-Small. Bên cạnh khả năng nhận diện tổn thương nhỏ và



Hình 5: So sánh vùng đặc trưng được mô hình tập trung các mô hình bằng Grad-CAM

phức tạp, Swin Transformer cũng cho thấy mức độ tổng quát hóa cao trong bối cảnh dữ liệu y khoa nhiễu và phân bố không đồng đều. Tuy nhiên, các hạn chế liên quan đến mất cân bằng dữ liệu, tỷ lệ bỏ sót bệnh và thiếu khả năng khoanh vùng tổn thương cụ thể vẫn đặt ra những yêu cầu cấp thiết cho các hướng cải tiến tiếp theo.

Để nâng cao hơn nữa hiệu quả mô hình, nghiên cứu tương lai sẽ tập trung tích hợp Swin Transformer với Graph Transformer Post-hoc nhằm khai thác các quan hệ ngữ nghĩa giữa các vùng ảnh hoặc giữa các mẫu trong tập dữ liệu thông qua biểu diễn đồ thị. Cách tiếp cận này được kỳ vọng không chỉ cải thiện độ nhạy bằng cách phát hiện các tín hiệu bệnh lý tiềm ẩn mà còn tăng cường khả năng suy luận ngữ cảnh, từ đó nâng cao độ tin cậy khi triển khai trong các hệ thống hỗ trợ quyết định lâm sàng. Ngoài ra, việc kết hợp với các kiến trúc phân đoạn như Swin-Unet cũng sẽ được nghiên cứu nhằm bổ sung khả năng xác định vị trí tổn thương, góp phần hoàn thiện hơn quy trình hỗ trợ chẩn đoán hình ảnh y khoa.

LỜI CẢM ƠN

Nghiên cứu này là sản phẩm của đề tài KH&CN cấp Trường có mã số YD2025-T-03, được tài trợ bởi kinh phí của Trường Đại học Y-Dược, Đại học Thái Nguyên.

TÀI LIỆU THAM KHẢO

- [1]. P. D. Blanc and A. Seaton, "Pneumoconiosis Redux. Coal Workers' Pneumoconiosis and Silicosis are Still a Problem," *American Journal of Respiratory and Critical Care Medicine*, vol. 193, no. 6, pp. 603–605, 2016.
- [2]. T. Li, X. Yang, H. Xu and H. Liu, "Early Identification, Accurate Diagnosis, and Treatment of Silicosis," *Canadian Respiratory Journal*, vol. 2022, pp. 1–6, 2022.
- [3]. F. Brims, D. Chambers, D. Deller, G. Edwards, R. Hoy, C. Jones, R. Knight, D. Yates, C. Kelaher, J. Shepherd, K. Stone and N. Wojcik, "National Guidance for Doctors Assessing Workers Exposed to Respirable Crystalline Silica Dust," *Australain Government Department of Health*, pp. 1–64, 2022.
- [4]. E. K. Austin, C. James and J. Tessier, "Early Detection Methods for Silicosis in Australia and Internationally: A Review of the Literature," *International Journal of Environmental Research and Public Health*, vol. 18, no. 15, pp. 1–24, 2021.
- [5]. X. Liu, Q. Jiang, P. Wu, L. Han and P. Zhou, "Global Incidence, Prevalence and Disease Burden of Silicosis: 30 Years' Overview and Forecasted Trends," *BMC public health*, vol. 23, no. 1, pp. 1–8, 2023.
- [6]. C. P. Bondarchuk, B. Grobman, A. Mansur and C. Y. Lu, "National Trends in Pneumonia-Related Mortality in the United States, 1999-2019," *Infectious Diseases (London, England)*, vol. 57, no. 1, pp. 56–65, 2025.
- [7]. K.-I. Kim, C. W. Kim, M. K. Lee, K. S. Lee, C.-K. Park, S. J. Choi and J. G. Kim, "Imaging of Occupational Lung Disease," *RadioGraphics*, 2001.
- [8]. W. Xuan, Z. Han and L. Zheng, "Non-Invasive Detection of Silicosis Based on Array Sensing and Pattern Recognition," *E3S Web of Conferences*, vol. 271, pp. 1–4, 2021.
- [9]. N. Shivaanivarsha, P. Kavipriya and S. Vigita, "A Deep Transfer Learning Approach for the Detection of Silicosis Using Chest Radiography," in *Proc. National Conference on Control Instrumentation System Conference*, Manipal, India, pp. 575–589, 2023.
- [10]. K. Xia and J. Wang, "Recent Advances of Transformers on Medical Image Analysis: A Comprehensive Review," *MedComm – Future Medicine*, vol. 2, no. 1, pp. 1–46, 2023.
- [11]. K. He, C. Gan, Z. Li, I. Rekik, Z. Yin, W. Ji, Y. Gao, Q. Wang, J. Zhang and D. Shen, "Transformers in Medical Image Analysis: A Review," pp. 1–23, 2022.
- [12]. M. Ahmad, S. Usman, F. Ullah, I. Batyrshin, M. Usman, M. Muzammil, K. Sajid, M. Hasnain, M. Jalal and G. Sidorov, "Automated Diagnosis of Lung Diseases Using Vision Transformer: A Comparative Study on Chest X-Ray Classification," *Journal of Population Therapeutics and Clinical Pharmacology*, vol. 31, no. 10, pp. 162–173, 2024.
- [13]. N. S. Mahajaya, P. D. W. Ayu and R. R. Huizen, "Classification of Lung Diseases in X-Ray Images Using Transformer-Based Deep Learning Models," *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI*, vol. 13, no. 3, pp. 494–505, 2024.
- [14]. S. A. H. and A. Jesser, "Advanced Hierarchical Imaging Techniques in TB Diagnosis: Leveraging Swin Transformer for Enhanced Lung Tuberculosis Detection," *Signal & Image Processing: An International Journal (SIPIJ)*, vol. 14, no. 14, pp. 1–10, 2023.
- [15]. H. Jin, C. Yu, J. Zhang, R. Zheng, Y. Fu and Y. Zhao, "Multitask Swin Transformer for Classification and Characterization of Pulmonary Nodules in CT Images," *Quantitative Imaging in Medicine and Surgery*, vol. 15, no. 3, pp. 1845–1861, 2025.
- [16]. C. M. Jones, S. S. Pasricha, S. B. Heinze and S. MacDonald, "Silicosis in Artificial Stone Workers: Spectrum of Radiological High-Resolution CT Chest Findings," *Journal of Medical Imaging and Radiation Oncology*, vol. 64, no. 2, pp. 241–249, 2020.
- [17]. L. Kong and J. Cheng, "Classification and Detection of COVID-19 X-Ray Images Based on Densenet and VGG16 Feature Fusion," *Biomedical Signal Processing and Control*, vol. 77, p. 103772, 2022.
- [18]. F. R. Song, W. Qiu, B. Ruan, W. M. Zhu, G. J. Yang, L. Li, W. H. Geng and H. Y. Zhao, "A Comparative Study on Diagnosis of Silicosis By Digital and High kV Film-Screen Chest Radiography," *Chinese journal of industrial hygiene and occupational diseases*, vol. 38, no. 12, pp. 919–921, 2020.
- [19]. C. Young, S. Barker, R. Ehrlich, B. Kistnasamy and A. Yassi, "Computer-Aided Detection for Tuberculosis and Silicosis in Chest Radiographs of Gold Miners of South Africa," *The International Journal of Tuberculosis and Lung Disease*, vol. 24, no. 4, pp. 444–451, 2020.
- [20]. R. Ehrlich, S. Barker, J. te Water Naude, D. Rees, B. Kistnasamy, J. Naidoo and A. Yassi, "Accuracy of Computer-Aided Detection of Occupational Lung Disease: Silicosis and Pulmonary Tuberculosis in Ex-Miners from the South African Gold Mines,"

- International Journal of Environmental Research and Public Health*, vol. 19, no. 19, pp. 1–14, 2022.
- [21]. F. Mushtaq, S. Bhattacharjee, S. Mandia, K. Singh, S. S. Chouhan, R. Kumar and P. Harjule, “Artificial Intelligence for Computer Aided Detection of Pneumoconiosis: A Succinct Review Since 1974,” *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108516, 2024.
- [22]. Y. Lee, Y. Chae and S. Jeon, “Integration and Evaluation of Clinical Decision Support Systems for Diagnosis of Idiopathic Pulmonary Fibrosis (IPF),” *Healthcare Informatics Research*, vol. 16, no. 4, pp. 260–272, 2010.
- [23]. F. Pesapane, M. Codari and F. Sardanelli, “Artificial Intelligence in Medical Imaging: Threat or Opportunity? Radiologists Again at The Forefront of Innovation in Medicine,” *European Radiology Experimental*, vol. 2, no. 1, pp. 1–10, 2018.
- [24]. G. K. Sharma, P. Harjule, B. Agarwal and R. Kumar, “Silicosis Detection Using Extended Transfer Learning Model,” in *Proc. International Conference on Recent Trends in Image Processing and Pattern Recognition*, pp. 111–126, 2024.
- [25]. T. T. T. Nguyen, Q. B. Bui and C. Pham, “A Data-Constrained Approach for Occupational Silicosis Detection on Chest X-Rays with Few-Shot Learning,” *Journal of Computer Science and Cybernetics*, vol. 41, no. 1, pp. 1–16, 2025.
- [26]. R. Yulvina, S. A. Putra, M. Rizkinia, A. Pujitresnani, E. D. Tenda, R. E. Yunus, D. H. Djumaryo, P. A. Yusuf and V. Valindria, “Hybrid Vision Transformer and Convolutional Neural Network for Multi-Class and Multi-Label Classification of Tuberculosis Anomalies on Chest X-Ray,” *Computers*, vol. 13, no. 12, pp. 1–29, 2024.
- [27]. B. Q. Bui, T. T. T. Nguyen, D. M. Le, C. Tran and C. Pham, “Innovative Silicosis and Pneumonia Classification: Leveraging Graph Transformer Post-hoc Modeling and Ensemble Techniques,” *arXiv*, 2024.
- [28]. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” in *Proc. IEEE International Conference on Computer Vision*, pp. 618–626, 2017.

extract image features for the classification task. Experimental results demonstrate that the Swin Transformer significantly outperforms several widely used deep learning architectures, such as DenseNet201 and MobileNetV3-Small. Specifically, the Swin Transformer achieved an accuracy of 87.77% on the test set, while DenseNet201 and MobileNetV3-Small attained 74.34% and 76.98%, respectively. These findings highlight the strong potential of the Swin Transformer model in supporting early diagnosis of silicosis from chest X-ray images.

Keywords: Silicosis, CNN, Deep Learning, Swin Transformer, Chest X-ray.



Nguyễn Thị Tân Tiến, nhận bằng Đại học năm 2007 tại trường Đại học Sư phạm, Đại học Thái Nguyên. Nhận bằng Thạc sĩ năm 2010 tại trường Đại học Sư phạm Hà Nội. Hiện nay là giảng viên Trường Đại học Y-Dược, Đại học Thái Nguyên. Lĩnh vực nghiên cứu chính là học máy, học sâu, xử lý ảnh và thị giác máy tính.



Nguyễn Văn Tảo, nhận bằng nhận bằng Thạc sĩ tại Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội năm 1999, nhận bằng Tiến sĩ tại Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam năm 2009. Hiện nay là giảng viên Trường Đại học Công nghệ Thông tin và Truyền thông, Đại học Thái Nguyên. Hướng nghiên cứu chính là xử lý hình ảnh, an toàn và bảo mật thông tin.



Ngô Hữu Huy, nhận bằng Đại học và Thạc sĩ tại Trường Đại học Công nghệ Thông tin và Truyền thông, Đại học Thái Nguyên năm 2010 và 2012, tương ứng, nhận bằng Tiến sĩ chuyên ngành Kỹ thuật thông tin và Khoa học máy tính tại Trường Đại học Feng Chia, Đài Loan năm 2021. Hiện nay là giảng viên Trường Đại học Công nghệ Thông tin và Truyền thông, Đại học Thái Nguyên. Lĩnh vực nghiên cứu chính là thị giác máy tính, học sâu, hệ thống nhúng, mạng nơ ron, và phát hiện đối tượng.

APPLICATION OF THE SWIN TRANSFORMER MODEL IN SUPPORTING THE DIAGNOSIS OF OCCUPATIONAL SILICOSIS

Abstract: Occupational silicosis is a serious lung disease that necessitates early detection to improve treatment outcomes. This study proposes the use of the Swin Transformer architecture to automatically classify chest X-ray images into two categories: normal and silicosis-affected. A dataset comprising 2,089 chest X-ray images was collected from hospitals in Thai Nguyen province. The original images underwent preprocessing and data augmentation techniques, including rotation, horizontal flipping, zooming and brightness adjustment, prior to being input into the model. The Swin Transformer leverages a hierarchical attention mechanism to effectively