

PHƯƠNG PHÁP KẾT HỢP ĐẶC TRƯNG THÍCH NGHI CHO BÀI TOÁN PHÂN LỚP BỆNH VIÊM PHỔI TỪ ẢNH X-QUANG

Bạch Thị Thiện Thanh, Nguyễn Khánh Linh,
Nguyễn Thanh Phong, Nguyễn Ngọc Thủy*

Khoa Công nghệ thông tin, Trường Đại học Khoa học, Đại học Huế

*Email: nnthuy.cs@hueuni.edu.vn

Ngày nhận bài: 7/5/2026; ngày hoàn thành phản biện: 29/5/2026; ngày duyệt đăng: 4/6/2026

TÓM TẮT

Ứng dụng học máy vào chẩn đoán viêm phổi từ ảnh X-quang là một bài toán quan trọng và thách thức. Trong nghiên cứu này, chúng tôi đề xuất một phương pháp kết hợp đặc trưng nhằm khai thác hiệu quả cả đặc trưng học sâu và đặc trưng thủ công. Tất cả các đặc trưng sau đó được giảm chiều bằng PCA và kết hợp thông qua cơ chế Gated Fusion để học biểu diễn đặc trưng. Các đặc trưng hợp nhất sau đó được sử dụng làm đầu vào cho các bộ phân lớp như KNN, hồi quy Logistic (LR) và MLP để học dự đoán. Kết quả thực nghiệm cho thấy hiệu quả dự đoán dựa trên đặc trưng được kết hợp bằng phương pháp đề xuất vượt trội đáng kể so với chỉ dựa trên đặc trưng học sâu hoặc đặc trưng thủ công. Cụ thể trên tập đặc trưng hợp nhất, bộ phân lớp KNN cho kết quả tốt nhất với độ chính xác 94.39% và F1-score 95.59%, LR cho kết quả gần tương đương 94.23% và 95.47%, trong khi MLP cũng đạt kết quả 93.59% và 95.02% tương ứng.

Từ khóa: Gated fusion, Kết hợp đa đặc trưng, Học sâu, Chest X-ray, Pneumonia.

1. GIỚI THIỆU

Học máy là một nhánh quan trọng của trí tuệ nhân tạo (AI), tập trung vào việc xây dựng các mô hình có khả năng học từ dữ liệu để giải quyết các bài toán dự đoán và nhận dạng. Các phương pháp học máy tiêu biểu bao gồm học có giám sát, học không giám sát và học tăng cường, được ứng dụng rộng rãi trong nhiều lĩnh vực như thị giác máy tính, xử lý ngôn ngữ tự nhiên và phân tích dữ liệu. Về mặt lịch sử, khái niệm học máy được hình thành từ những năm 1950 với công trình của Arthur Samuel. Trong giai đoạn 1960–1980, các mô hình truyền thống như hồi quy tuyến tính và cây quyết định được phát triển và ứng dụng rộng rãi. Tuy nhiên, phải đến sau năm 2000, cùng với sự

phát triển của dữ liệu lớn và năng lực tính toán, học máy mới thực sự bùng nổ và trở thành nền tảng của nhiều hệ thống thông minh hiện đại.

Trên nền tảng đó, học sâu (LeCun, Y., Bengio, Y. & Hinton, G., 2015) nổi lên như một hướng tiếp cận tiên tiến trong học máy, sử dụng các mạng nơ-ron nhiều tầng để học biểu diễn dữ liệu ở nhiều mức trừu tượng. Nhờ khả năng tự động trích xuất đặc trưng từ dữ liệu thô, học sâu vượt qua hạn chế của các phương pháp phụ thuộc vào đặc trưng thủ công và đạt hiệu suất cao trong nhiều bài toán phức tạp. Bắt nguồn từ các nghiên cứu về mạng nơ-ron từ những năm 1940 và được củng cố bởi thuật toán lan truyền ngược (1986), học sâu thực sự bùng nổ từ năm 2012 với thành công của AlexNet trên ImageNet, mở ra sự phát triển mạnh mẽ của các kiến trúc như CNN, RNN và Vision Transformer (ViT) (Dosovitskiy, A. et al., 2020).

Song song với sự phát triển của học sâu, phương pháp kết hợp đặc trưng (Feature Fusion) đã trở thành một hướng nghiên cứu quan trọng nhằm nâng cao hiệu quả của các mô hình học máy. Phương pháp này tích hợp nhiều nguồn đặc trưng để tạo biểu diễn dữ liệu toàn diện hơn. Xuất phát từ lĩnh vực nhận dạng mẫu và xử lý tín hiệu từ những năm 1990, các kỹ thuật kết hợp đã phát triển từ các cách đơn giản như nối đặc trưng, trung bình có trọng số đến các cơ chế hiện đại như attention và gated fusion, cho phép mô hình tự động điều chỉnh mức đóng góp của từng đặc trưng.

Trong những năm gần đây, AI đã đạt được nhiều thành tựu quan trọng trong lĩnh vực y sinh, đặc biệt là trong chẩn đoán hình ảnh y khoa (Kermany et al., 2018). Các hệ thống dựa trên học máy và học sâu có khả năng hỗ trợ bác sĩ trong việc phân tích ảnh nhanh chóng và chính xác hơn. Trong đó, bài toán chẩn đoán viêm phổi từ ảnh X-quang lồng ngực là một hướng nghiên cứu có ý nghĩa thực tiễn cao, góp phần phát hiện sớm bệnh và nâng cao hiệu quả điều trị. Trước đây, các phương pháp xử lý ảnh y khoa chủ yếu dựa trên các đặc trưng thủ công và các đặc trưng thống kê, vốn có ưu điểm về tính giải thích nhưng hạn chế về khả năng biểu diễn. Vì vậy, sự ra đời của các mô hình học sâu hiện đại, đặc biệt là ViT, đã mở ra khả năng học biểu diễn đặc trưng mạnh mẽ từ dữ liệu ảnh. Tuy nhiên, việc chỉ sử dụng một nguồn đặc trưng riêng lẻ vẫn chưa khai thác hết thông tin tiềm ẩn trong dữ liệu. Do đó, xu hướng kết hợp đa đặc trưng giữa đặc trưng học sâu và đặc trưng thủ công đang được quan tâm nhằm nâng cao hiệu suất mô hình. Tuy nhiên, trong bối cảnh các nguồn đặc trưng không đồng nhất mang những thông tin bổ sung nhưng có mức độ quan trọng khác nhau, các phương pháp kết hợp truyền thống như nối đặc trưng hoặc trung bình đơn giản thường áp dụng cơ chế cào bằng, không phản ánh đúng đóng góp thực sự của từng thành phần. Điều này đặc biệt bất lợi trong các bài toán y sinh, nơi dữ liệu thường hạn chế và dễ bị nhiễu, dẫn đến nguy cơ suy giảm hiệu năng và khả năng tổng quát hóa.

Nghiên cứu này hướng đến mục tiêu nâng cao hiệu quả chẩn đoán viêm phổi từ ảnh X-quang lồng ngực thông qua việc xây dựng một hệ thống kết hợp đa đặc trưng

hoàn chỉnh. Cụ thể, các nguồn đặc trưng bao gồm ViT, Radiomics, Statistical và Wavelet Scattering Transform (WST) được chuẩn hóa và giảm chiều thông qua PCA nhằm loại bỏ nhiễu và giảm độ phức tạp của không gian biểu diễn. Sau đó, cơ chế Gated fusion được áp dụng để điều chỉnh động mức độ đóng góp của từng thành phần thông qua các công học được, giúp mô hình tự động nhấn mạnh các thông tin quan trọng và hạn chế ảnh hưởng của các yếu tố kém liên quan. Biểu diễn hợp nhất thu được được sử dụng làm đầu vào cho các bộ phân lớp học máy truyền thống như KNN và LR cùng với bộ phân lớp học sâu MLP để thực hiện nhiệm vụ phân lớp.

Đóng góp chính của nghiên cứu nằm ở việc đề xuất một cách tiếp cận kết hợp hiệu quả giữa đặc trưng học sâu và đặc trưng thủ công, tận dụng đồng thời khả năng học biểu diễn mạnh mẽ và tính ổn định, dễ diễn giải. Đồng thời, cơ chế Gated Fusion được áp dụng để tự động điều chỉnh mức đóng góp của từng nguồn đặc trưng, giúp cải thiện chất lượng biểu diễn. Kết quả cũng cho thấy các bộ phân lớp truyền thống vẫn đạt hiệu năng cao khi sử dụng đặc trưng tốt, đồng thời mở ra tiềm năng ứng dụng trong hỗ trợ chẩn đoán y khoa.

Phần còn lại của bài báo được tổ chức như sau. Phần 2 trình bày các nghiên cứu liên quan. Phần 3 giới thiệu phương pháp đề xuất. Phần 4 mô tả thiết lập thực nghiệm và các kết quả đạt được. Phần 5 thảo luận và phân tích sâu về kết quả. Cuối cùng, Phần 6 đưa ra kết luận và định hướng nghiên cứu trong tương lai.

2. CÁC NGHIÊN CỨU LIÊN QUAN

Trong bối cảnh nhu cầu tự động hóa các hệ thống hỗ trợ chẩn đoán ngày càng gia tăng, việc khai thác thông tin từ ảnh X-quang ngực đã trở thành một hướng nghiên cứu quan trọng trong lĩnh vực y sinh. Các phương pháp hiện đại tập trung vào việc xây dựng những biểu hiện đặc trưng có khả năng phản ánh đầy đủ cả cấu trúc cục bộ lẫn ngữ cảnh toàn cục của ảnh, từ đó nâng cao độ chính xác của bài toán phân lớp bệnh. Đặc biệt, mô hình ViT đã mở ra hướng tiếp cận mới trong thị giác máy tính bằng cách sử dụng kiến trúc Transformer để học biểu diễn đặc trưng toàn cục. Không giống như các mạng tích chập truyền thống, ViT khai thác cơ chế self-attention để mô hình hóa mối quan hệ giữa các vùng ảnh, giúp cải thiện khả năng nhận diện cấu trúc giải phẫu. Bên cạnh đó, nhiều nghiên cứu đã áp dụng các mô hình học sâu tiên tiến cho bài toán phân lớp bệnh phổi từ ảnh X-quang và đạt được kết quả khả quan (Khánh, N., 2023) (Pham, T. A., Hoang, V. D., 2024).

Song song với đó, các phương pháp trích xuất đặc trưng thủ công vẫn đóng vai trò quan trọng trong phân tích ảnh y học. Phương pháp Radiomics (Lambin et al., 2012) cho phép trích xuất số lượng lớn các đặc trưng định lượng từ ảnh như kết cấu (texture), hình dạng (shape) và cường độ (intensity), hỗ trợ cải thiện khả năng giải thích của mô

hình. Bên cạnh đó, các đặc trưng kết cấu thống kê (statistical texture features) dựa trên ma trận đồng xuất hiện mức xám (GLCM) (R. M. Haralick, K. Shanmugam and I. Dinstein, 1973) vẫn được sử dụng rộng rãi nhờ khả năng mô tả chi tiết cấu trúc ảnh. Ngoài ra, cách tiếp cận WST (Mallat, S., 2012) cũng cung cấp các đặc trưng ổn định trước biến dạng và không yêu cầu quá trình huấn luyện. Đặc biệt, các nghiên cứu gần đây cho thấy việc kết hợp đặc trưng wavelet với radiomics có thể cải thiện đáng kể hiệu quả phân lớp trong các bài toán y sinh (Tang et al., 2023).

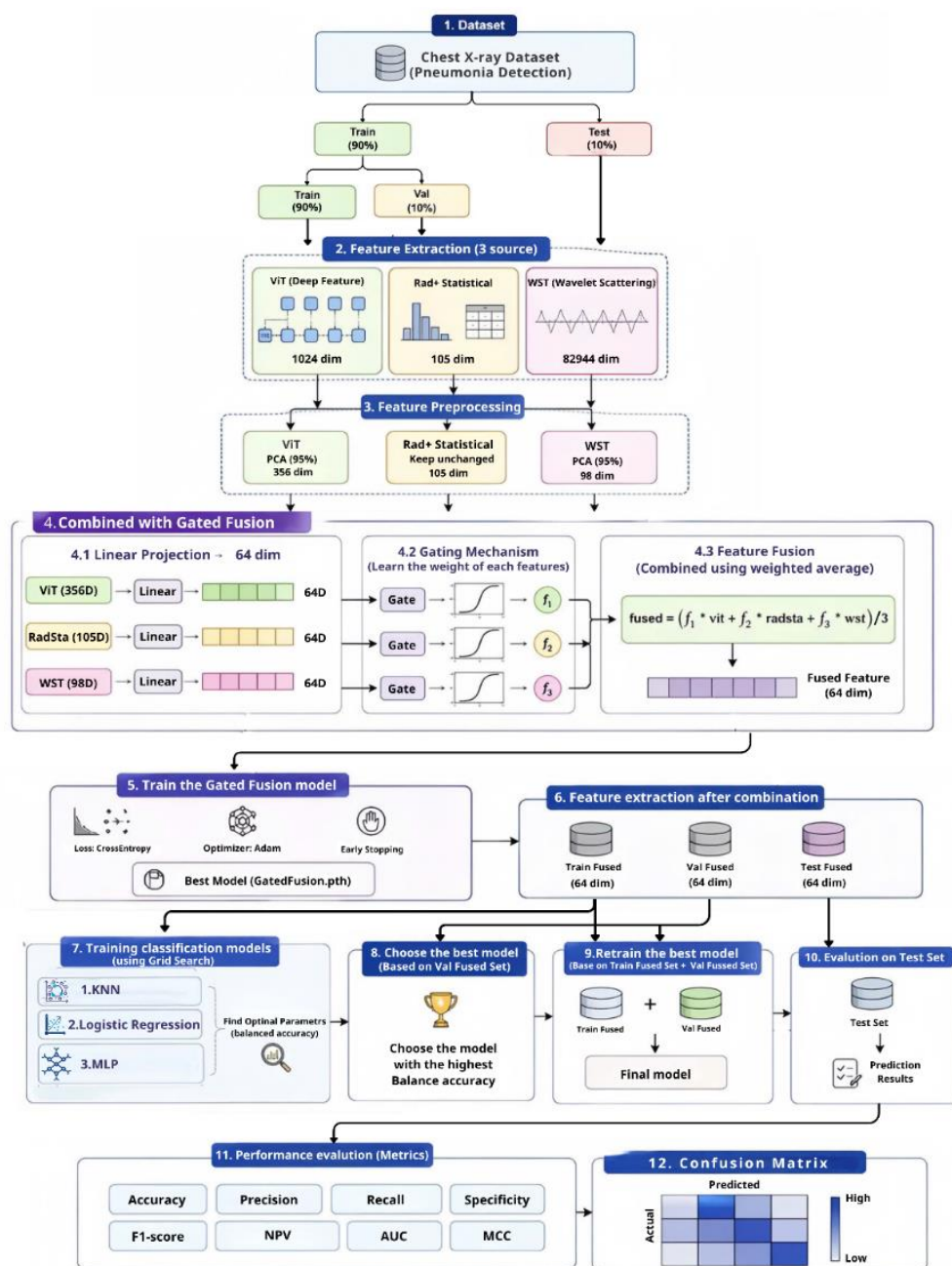
Nhằm tận dụng ưu điểm của các loại đặc trưng khác nhau, nhiều nghiên cứu đã đề xuất các phương pháp kết hợp đặc trưng. Công trình Multimodal Deep Learning (Ngiam et al., 2011) đặt nền móng cho việc học biểu diễn chung từ nhiều nguồn dữ liệu, mở ra hướng nghiên cứu về kết hợp đặc trưng trong học sâu. Các phương pháp kết hợp truyền thống như nối đặc trưng hoặc tổng hợp tuyến tính tuy đơn giản nhưng chưa phản ánh được mức độ đóng góp khác nhau của từng nguồn đặc trưng. Để giải quyết vấn đề này, các phương pháp kết hợp đặc trưng thích nghi đã được đề xuất. Tiêu biểu là cơ chế Gated Fusion (Arevalo et al., 2017), cho phép mô hình tự động điều chỉnh trọng số của từng nguồn đặc trưng dựa trên dữ liệu đầu vào. Ngoài ra, một số nghiên cứu khác cũng tập trung vào việc học biểu diễn kết hợp thông qua các chiến lược kết hợp nâng cao (Baltrusaitis et al., 2017), nhằm khai thác tốt hơn mối quan hệ giữa các đặc trưng khác nhau. Các nghiên cứu gần đây còn cho thấy hiệu quả của việc kết hợp đặc trưng trong phân tích ảnh y học, từ đó thúc đẩy phát triển các phương pháp kết hợp thích nghi (ADehbozorgi P, Ryabchykov O, Bocklitz TW., 2025).

Tuy nhiên, phần lớn các phương pháp hiện tại vẫn chỉ tập trung vào việc kết hợp một số lượng hạn chế các loại đặc trưng hoặc chưa khai thác đầy đủ tính bổ sung giữa các miền đặc trưng khác nhau. Đặc biệt, việc tích hợp đồng thời các đặc trưng đa dạng như đặc trưng học sâu ViT, Radiomics, Statistical và Wavelet Scattering Transform vẫn còn là một thách thức chưa được nghiên cứu đầy đủ trong bài toán phân lớp viêm phổi từ ảnh X-quang lồng ngực.

3. PHƯƠNG PHÁP ĐỀ XUẤT

3.1. Tổng quan phương pháp

Trong nghiên cứu này, chúng tôi đề xuất một phương pháp lai kết hợp giữa học sâu và học máy cho bài toán chẩn đoán viêm phổi từ ảnh X-quang lồng ngực, như minh họa trong Hình 1.



Hình 1: Sơ đồ phương pháp đề xuất

Phương pháp được xây dựng theo hướng kết hợp đa đặc trưng nhằm khai thác đồng thời thông tin từ các nguồn đặc trưng khác nhau. Nghiên cứu được tiến hành trên tập dữ liệu ảnh X-quang lồng ngực gồm hai lớp: NORMAL (không mắc bệnh) và PNEUMONIA (mắc viêm phổi). Tập dữ liệu được chia thành tập huấn luyện, tập kiểm thử và tập thẩm định trong đó tập huấn luyện được sử dụng để xây dựng mô hình và tập kiểm thử được dùng để đánh giá hiệu suất.

Cụ thể, đặc trưng sâu được trích xuất từ mô hình ViT, trong khi các đặc trưng thủ công bao gồm Radiomics, Statistical và WST nhằm bổ sung thông tin về kết cấu, hình dạng và đặc điểm tần số của ảnh. Các đặc trưng này sau đó được giảm chiều bằng PCA để loại bỏ nhiễu và tối ưu hóa không gian đặc trưng.

Sau đó, các đặc trưng sau khi được xử lý được tích hợp thông qua cơ chế Gated Fusion, cho phép học biểu diễn đặc trưng hợp nhất bằng cách điều chỉnh trọng số giữa các nguồn đặc trưng một cách thích nghi. Các đặc trưng sau khi điều chỉnh được kết hợp bằng phép trung bình có trọng số, tạo thành biểu diễn đặc trưng hợp nhất có kích thước cố định.

Biểu diễn đặc trưng thu được được sử dụng làm đầu vào cho các bộ phân lớp học máy truyền thống, bao gồm KNN, LR và bộ phân lớp học sâu MLP để thực hiện phân lớp. Quá trình tối ưu siêu tham số được thực hiện bằng Grid Search kết hợp xác thực chéo (k -fold Cross-Validation), sau đó mô hình tối ưu được sử dụng để suy luận trên tập kiểm thử.

Cuối cùng, hiệu suất của phương pháp được đánh giá trên tập kiểm thử thông qua các độ đo chuẩn như Accuracy, Precision, Recall, F1-score và AUC, nhằm đảm bảo đánh giá toàn diện và khách quan.

3.2. Thuật toán

Sau đây chúng tôi trình bày chi tiết quy trình đề xuất nhằm trích xuất, chuẩn hóa và kết hợp các đặc trưng đa nguồn thông qua cơ chế Gated Fusion. Mục tiêu của phương pháp là tận dụng đồng thời thông tin toàn cục từ ViT và thông tin thủ công từ Radiomics, Statistical và Wavelet, từ đó xây dựng một biểu diễn đặc trưng thống nhất và giàu tính phân biệt.

Quy trình được chia thành hai giai đoạn chính: (1) trích xuất và hợp nhất đặc trưng thông qua cơ chế học có trọng số động (Gated Fusion), và (2) tối ưu mô hình phân lớp và thực hiện suy luận trên tập dữ liệu kiểm thử. Chi tiết từng bước được mô tả trong các thuật toán dưới đây.

Thuật toán 1: Trích xuất và kết hợp đặc trưng bằng Gated Fusion

Input:

$D_{train} = \{X, y\}$: Tập dữ liệu huấn luyện gồm N mẫu

$F_{vit}, F_{rad}, F_{sta}, F_{wst} \in \mathbb{R}^{N \times d}$: Các đặc trưng đầu vào

d : Số chiều không gian đặc trưng ẩn

Output: $F_{fused}^{train} \in \mathbb{R}^{N \times d}$: Đặc trưng hợp nhất

1. Kết hợp đặc trưng thủ công:

$$F_{radsta} \leftarrow [F_{rad}, F_{sta}]$$

2. Giảm chiều đặc trưng:

$$F_{vit} \leftarrow PCA(F_{vit})$$

$$F_{wst} \leftarrow PCA(F_{wst})$$

3. Chiếu về không gian đặc trưng chung:

$$H_i \leftarrow W_i F_i, \quad i \in \{vit, radsta, wst\}$$

4. Chuẩn hóa đặc trưng:

$$H_i \leftarrow LayerNorm(H_i)$$

5. Tính hệ số gate:

$$G_i \leftarrow \sigma(W_g H_i)$$

6. Điều chỉnh đặc trưng:

$$\widetilde{H}_i \leftarrow H_i \odot (1 + G_i)$$

7. Kết hợp đặc trưng:

$$F_{fused} \leftarrow \frac{1}{3}(\widetilde{H}_{vit} + \widetilde{H}_{radsta} + \widetilde{H}_{wst})$$

8. Huấn luyện mô hình bằng Cross-Entropy

9. Trích xuất đặc trưng hợp nhất:

$$F_{fused}^{train} \leftarrow F_{fused}$$

Thuật toán 2: Tối ưu phân lớp và suy luận trên tập kiểm thử

Input:

$F_{fused}^{train}, y_{train}$: Đặc trưng và nhãn huấn luyện

F_{fused}^{val}, y_{val} : Đặc trưng và nhãn thẩm định

F_{fused}^{test} : Đặc trưng kiểm thử

Θ : Không gian siêu tham số

k : Số fold trong Cross-Validation

Output:

$\hat{y}$: Nhãn dự đoán

Các độ đo đánh giá: Accuracy, Precision, Recall, F1-score, AUC, MCC, Specificity, NPV.

Giai đoạn 1: Tối ưu siêu tham số (Grid Search)

1. Khởi tạo:

$$BestScore \leftarrow 0$$

2. Duyệt lưới:

$$for \theta \in \Theta do$$

3. Huấn luyện mô hình với θ bằng k -fold Cross-Validation

4. Tính độ đo trung bình: S_θ

5. If ($S_\theta > BestScore$) then

6. $BestScore \leftarrow S_\theta$

7. $\theta^* \leftarrow \theta$

8. $M_{best} \leftarrow M$

9. End if

10. End for

11. Huấn luyện lại mô hình tối ưu:

$$M^* \leftarrow M(\theta^*) \text{ trên } F_{fused}^{train} \cup F_{fused}^{val}$$

Giai đoạn 2: Suy luận trên tập kiểm thử

12. Dự đoán nhãn: $\hat{y} \leftarrow M^*(F_{fused}^{test})$

13. Đánh giá mô hình: Accuracy, Precision, Recall, F1-score, AUC, MCC, Specificity, NPV

Phân tích độ phức tạp

Độ phức tạp tính toán của phương pháp đề xuất được phân tích theo từng thành phần chính.

Trong giai đoạn Gated Fusion, các phép biến đổi tuyến tính, chuẩn hóa và tính toán hệ số gate đều được thực hiện trên từng mẫu với chi phí tuyến tính theo số chiều đặc trưng. Do đó, độ phức tạp của giai đoạn này là $O(N \cdot d)$.

Đối với quá trình tối ưu siêu tham số bằng Grid Search, mô hình cần được huấn luyện lặp lại với K tổ hợp tham số khác nhau. Mỗi lần huấn luyện có chi phí xấp xỉ $O(N \cdot d)$, dẫn đến tổng chi phí là $O(K \cdot N \cdot d)$.

Trong giai đoạn suy luận, việc dự đoán nhãn trên tập kiểm thử cũng có độ phức tạp tuyến tính theo số mẫu và số chiều đặc trưng, tương ứng $O(N \cdot d)$.

Tổng thể, độ phức tạp của toàn bộ phương pháp là $O(K \cdot N \cdot d)$. Trong trường hợp số tổ hợp siêu tham số K và số chiều đặc trưng d được xem là hằng số, độ phức tạp có thể được rút gọn về $O(N)$, cho thấy phương pháp có khả năng mở rộng tốt đối với các tập dữ liệu có kích thước lớn.

4. THỰC NGHIỆM

4.1. Tập dữ liệu

Trong nghiên cứu này, dữ liệu được sử dụng cho cả quá trình huấn luyện và kiểm thử lấy từ bộ dữ liệu Chest X-Ray Images (Pneumonia) được phát hành bởi National Institutes of Health (NIH) và được tổng hợp bởi nhóm nghiên cứu từ trung tâm Guangzhou Women And Children's Medical Center, Trung Quốc. Bộ dữ liệu bao gồm các ảnh X-quang lồng ngực của trẻ em từ 1 đến 5 tuổi và đã được sử dụng rộng rãi trong các nghiên cứu về học máy và thị giác máy tính cho bài toán chẩn đoán viêm phổi. Bảng 1 mô tả tập dữ liệu bao gồm 5.856 hình ảnh được chia thành hai lớp riêng biệt là Pneumonia và Normal.

Bảng 1: Tập dữ liệu chẩn đoán viêm phổi

Lớp	Tổng số	Tỷ lệ (%)
Pneumonia	4,273	73
Normal	1,583	27
Tổng	5,856	100

Bộ dữ liệu ban đầu được chia thành ba phần: tập huấn luyện (train set) 5.216 ảnh, tập thẩm định (validation set) 16 ảnh và tập kiểm thử (test set) 624 ảnh, tương ứng với 88,7%, 0,3% và 10% tổng số mẫu. Tuy nhiên, do tập thẩm định có kích thước quá nhỏ và không đảm bảo tính đại diện, phần dữ liệu này được gộp vào tập huấn luyện. Sau khi gộp, tập huấn luyện mới gồm 5.232 ảnh, bao gồm 1.349 mẫu NORMAL và 3.883 mẫu PNEUMONIA. Tiếp đó, tập này được chia lại thành hai phần train và validation theo tỉ lệ 90%–10%, tương ứng với 4.708 ảnh cho huấn luyện và 524 ảnh cho thẩm định. Chi tiết về bộ dữ liệu này được mô tả ở Bảng 2.

Bảng 2: Bảng phân chia tập dữ liệu chẩn đoán viêm phổi.

Lớp (Class)	Số mẫu Train	Số mẫu Test	Số mẫu Val	Tổng số	Tỷ lệ (%)
Pneumonia	3,494	390	389	4,273	73

Normal	1,214	234	135	1,583	27
Tổng	4708	624	524	5,856	100



NORMAL



PNEUMONIA

Hình 2: Hình ảnh minh họa từ bộ dữ liệu chẩn đoán viêm phổi

Trong nghiên cứu này, dữ liệu ảnh X-quang ngực được tiền xử lý nhằm phục vụ mô hình học đa đặc trưng. Tất cả ảnh được chuẩn hóa về kích thước cố định 128×128 , đảm bảo tính đồng nhất giữa các tập train và test.

Trên cơ sở đó, dữ liệu được xử lý theo ba nhánh đặc trưng khác nhau. Đối với nhánh RadSta, ảnh được chuyển sang dạng grayscale để trích xuất các đặc trưng thủ công. Với nhánh WST, ảnh grayscale sau khi được chuẩn hóa kích thước sẽ được đưa vào quá trình trích xuất đặc trưng Wavelet Scattering, sau đó áp dụng PCA để giảm chiều không gian đặc trưng. Đối với nhánh ViT, ảnh grayscale được chuyển sang dạng RGB bằng cách lặp lại ba kênh nhằm phù hợp với đầu vào của mô hình Vision Transformer, đồng thời các đặc trưng sâu sau khi trích xuất cũng tiếp tục được giảm chiều bằng PCA trước khi thực hiện hợp nhất đặc trưng. Trong cả hai nhánh WST và ViT, PCA được xây dựng trên tập huấn luyện và sau đó được sử dụng để biến đổi tập validation và tập kiểm tra nhằm đảm bảo tính khách quan của quá trình đánh giá mô hình.

Như vậy, cùng một dữ liệu đầu vào nhưng được xử lý theo ba hướng biểu diễn khác nhau, tương ứng với đặc trưng thống kê (RadSta), đặc trưng tần số-kết cấu (WST) và đặc trưng ngữ nghĩa (ViT), làm cơ sở cho mô hình học hợp nhất trong bài toán phân loại viêm phổi.

4.2. Tham số huấn luyện mô hình

Trong nghiên cứu này, các đặc trưng đầu vào gồm đặc trưng học sâu ViT và ba nhóm đặc trưng thủ công: Radiomics, Statistical và WST. Sự kết hợp này nhằm khai thác

đồng thời thông tin ngữ nghĩa mức cao và các đặc trưng cường độ-kết cấu mức thấp từ ảnh X-quang ngực.

Đối với đặc trưng ViT, ảnh X-quang RGB được thay đổi kích thước về 128×128 pixel bằng hàm *tf.image.resize()*. Sau đó, đặc trưng được trích xuất bằng mô hình pretrained *google/vit-large-patch16-224-in21k* thông qua thư viện *Hugging Face Transformers*. Bộ tiền xử lý *ViTImageProcessor.from_pretrained()* được sử dụng với *do_rescale=False*, và vector đặc trưng được lấy từ đầu ra trước tầng phân loại cuối cùng.

Đối với đặc trưng *Statistical*, ảnh được làm phẳng thành vector một chiều và trích xuất 12 đặc trưng mô tả phân bố mức xám như trung bình, phương sai, độ lệch chuẩn, *skewness*, *kurtosis*, năng lượng và entropy, sử dụng *NumPy* và *SciPy*. Các đặc trưng sau đó được chuẩn hóa bằng *StandardScaler()*.

Đối với đặc trưng *Radiomics*, sử dụng thư viện *PyRadiomics* kết hợp *SimpleITK* thông qua *RadiomicsFeatureExtractor()* với *binWidth = 25* và *normalize = True*, kích hoạt các nhóm đặc trưng gồm *first-order*, *GLCM* và *GLSZM*. Sau khi loại bỏ các trường không liên quan, thu được 93 đặc trưng mỗi ảnh, tiếp tục được chuẩn hóa bằng *StandardScaler()*.

Đối với đặc trưng WST, sử dụng thư viện *Kymatio* qua *Scattering2D()* với $J = 2$, $L = 8$. Sau khi tính các hệ số scattering, áp dụng *Global Average Pooling* để giảm chiều, thu được vector đặc trưng 81 chiều cho mỗi ảnh.

Cuối cùng, hai nhóm đặc trưng *Statistical* và *Radiomics* được nối lại, tạo thành vector đặc trưng 105 chiều cho bước xử lý tiếp theo. Sau đó, mô hình Gated Fusion được huấn luyện từ đầu để học cách chuẩn hóa và kết hợp các nguồn đặc trưng.

Mạng fusion được huấn luyện nhằm học biểu diễn đặc trưng kết hợp trước khi đưa vào các bộ phân lớp. Do đặc trưng đầu vào đã được trích xuất từ các mô hình tiền huấn luyện, quá trình huấn luyện được thiết kế gọn nhẹ để tránh overfitting và đảm bảo ổn định. Như trình bày ở Bảng 3, mô hình được huấn luyện trong 15 epoch và sử dụng toàn bộ tập huấn luyện để học tham số. Sau khi huấn luyện, mạng fusion không được dùng trực tiếp làm bộ phân lớp mà chỉ được sử dụng để trích xuất đặc trưng hợp nhất phục vụ cho các bộ phân lớp phía sau.

Bảng 3: Cấu hình huấn luyện mô hình

Tham số	Giá trị
Optimizer	Adam
Learning rate	1×10^{-4}
Batch size	32
Number of epochs	15
Loss function	Cross-Entropy

Các đặc trưng hợp nhất sau đó được sử dụng làm đầu vào cho các bộ phân lớp. Việc lựa chọn siêu tham số được thực hiện bằng Grid Search kết hợp 5-fold cross-validation, với thước đo đánh giá là Balanced Accuracy, đảm bảo đánh giá công bằng đối với dữ liệu mất cân bằng như được sử dụng trong thực nghiệm này.

4.3. Các độ đo đánh giá hiệu năng

Để đánh giá một cách toàn diện hiệu quả của mô hình đề xuất, chúng tôi sử dụng nhiều độ đo khác nhau nhằm phản ánh đầy đủ các khía cạnh của bài toán phân lớp. Lưu ý rằng, do dữ liệu có sự mất cân bằng giữa các lớp, do đó việc chỉ sử dụng độ chính xác tổng thể (accuracy) có thể không phản ánh đúng năng lực thực sự của mô hình. Vì vậy, các độ đo như precision, recall, F1-score, Specificity, FDR, FNR, NPV, BAcc, AUC và MCC được lựa chọn để cung cấp cái nhìn chi tiết và đáng tin cậy hơn. Các chỉ số này giúp đánh giá khả năng của mô hình trong việc nhận diện chính xác các trường hợp bệnh và không bệnh, từ đó đảm bảo tính ứng dụng trong thực tế chẩn đoán viêm phổi từ ảnh X-quang lồng ngực.

Các độ đo được tính toán dựa trên ma trận nhầm lẫn (confusion matrix), bao gồm bốn thành phần: số mẫu dương tính dự đoán đúng (TP), số mẫu âm tính dự đoán đúng (TN), số mẫu dương tính dự đoán sai (FN) và số mẫu âm tính dự đoán sai (FP). Cụ thể, các độ đo được xác định như sau:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$FDR = \frac{FP}{TP + FP} = 1 - Precision \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$FNR = \frac{FN}{TP + FN} = 1 - Recall \quad (5)$$

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

$$NPV = \frac{TN}{TN + FN} \quad (7)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (8)$$

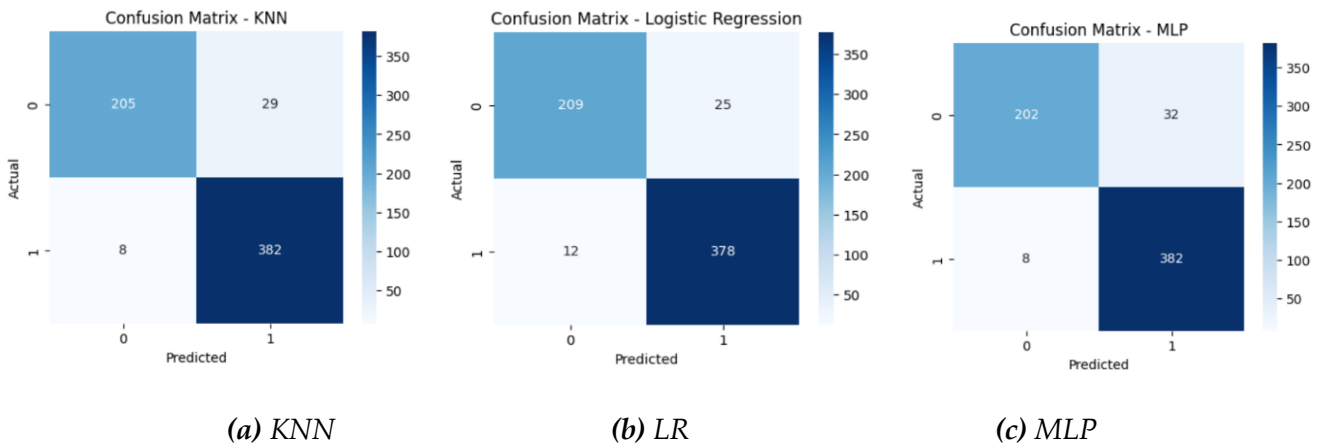
$$BAcc = \frac{Recall + Specificity}{2} \quad (9)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (10)$$

Trong đó, các độ đo như Recall và Specificity phản ánh khả năng phát hiện đúng các trường hợp bệnh và không bệnh, trong khi MCC và AUC cung cấp đánh giá tổng thể về chất lượng mô hình trong điều kiện dữ liệu mất cân bằng.

4.4. Kết quả

Kết quả thực nghiệm được thể hiện ở Hình 3 cho thấy tất cả các bộ phân lớp đều đạt hiệu suất cao trên không gian đặc trưng kết hợp, khẳng định được hiệu quả của phương pháp được đề xuất. Trong số các mô hình được đánh giá, KNN đạt độ chính xác tổng thể cao nhất, cho thấy sự phù hợp của nó với không gian đặc trưng được phân tách rõ ràng. Tuy nhiên, trong chẩn đoán y khoa, việc giảm thiểu sai sót âm tính là vô cùng quan trọng. Về mặt này, bộ phân lớp MLP cho thấy hiệu suất vượt trội, đạt số lượng sai sót âm tính thấp nhất (FN=8) và độ nhạy cao nhất đối với lớp bệnh viêm phổi (97,95%). Điều này cho thấy MLP đáng tin cậy hơn trong việc phát hiện các trường hợp dương tính, giảm nguy cơ bỏ sót chẩn đoán.



Hình 3: Ma trận nhầm lẫn của các mô hình

Mặt khác, KNN tạo ra ít kết quả dương tính giả hơn, do đó nó thận trọng hơn và có khả năng phù hợp hơn trong các trường hợp cần giảm thiểu chẩn đoán quá mức. Hồi quy Logistic cho thấy hiệu suất cạnh tranh nhưng hơi kém hơn so với KNN và MLP.

Nhìn chung, kết quả cho thấy sự đánh đổi giữa độ nhạy và độ đặc hiệu, trong đó MLP ưu tiên độ nhạy trong khi KNN mang lại sự cân bằng tốt hơn về độ chính xác tổng thể. Để quan sát hiệu quả của đặc trưng được kết hợp bởi phương pháp đề xuất so với các đặc trưng của mỗi phương pháp riêng lẻ, chúng tôi cũng tiến hành so sánh chúng trên nhiều độ đo khác nhau. Kết quả từ Bảng 4 cho thấy phương pháp đề xuất đạt hiệu suất vượt trội và nhất quán so với các phương pháp sử dụng đặc trưng đơn lẻ trên cả ba bộ phân lớp KNN, Logistic Regression và MLP. Cụ thể, với KNN, độ chính xác tăng đáng kể từ 87,50% (ViT_large) lên 94,39%, đồng thời hệ số MCC cũng được cải thiện mạnh từ 73,04% lên 87,99%, phản ánh khả năng phân lớp cân bằng và đáng tin cậy hơn. Xu hướng tương tự cũng được ghi nhận với Logistic Regression (Acc đạt 94,23%, MCC

đạt 87,65%) và MLP (Acc đạt 93,59%, MCC đạt 86,34%), cho thấy tính ổn định của phương pháp khi áp dụng trên nhiều mô hình khác nhau. Đáng chú ý, các chỉ số Recall đều đạt mức rất cao (trên 97%) trong khi FNR giảm xuống dưới 3%, cho thấy mô hình có khả năng phát hiện bệnh hiệu quả và hạn chế tối đa trường hợp bỏ sót, điều đặc biệt quan trọng trong các bài toán chẩn đoán y khoa. Ngoài ra, các chỉ số như Precision, F1-score và Specificity cũng được cải thiện đồng đều, cho thấy phương pháp đề xuất không chỉ nâng cao độ chính xác tổng thể mà còn duy trì được sự cân bằng giữa các lớp. Những kết quả này khẳng định rằng việc kết hợp đa đặc trưng thông qua cơ chế Gated Fusion giúp khai thác hiệu quả thông tin bổ sung giữa các nguồn đặc trưng khác nhau, từ đó nâng cao đáng kể hiệu suất phân lớp.

Bảng 4: So sánh hiệu suất giữa các bộ phân lớp và đặc trưng

Classifier	Đặc trưng	Acc	Pre	FDR	Rec	FNR	Spe	NPV	F1	BAcc	AUC	MCC
KNN	ViT_large	0.8750	0.8786	0.1214	0.9282	0.0718	0.7863	0.8679	0.9027	0.8573	0.8573	0.7304
	Radiomics + Statistical	0.8317	0.8174	0.1826	0.9410	0.0590	0.6496	0.8686	0.8749	0.7953	0.7953	0.6365
	WST	0.8606	0.8499	0.1501	0.9436	0.0564	0.7222	0.8848	0.8943	0.8329	0.8329	0.6999
	Phương pháp đề xuất	0.9439	0.9404	0.0596	0.9718	0.0282	0.8974	0.9502	0.9559	0.9346	0.9346	0.8799
LR	ViT_large	0.9263	0.9257	0.0743	0.9590	0.0410	0.8718	0.9273	0.9421	0.9154	0.9154	0.8418
	Radiomics + Statistical	0.8718	0.8523	0.1477	0.9615	0.0385	0.7222	0.9185	0.9036	0.8419	0.8419	0.7260
	WST	0.8654	0.8660	0.1340	0.9282	0.0718	0.7607	0.8641	0.8960	0.8444	0.8444	0.7092
	Phương pháp đề xuất	0.9423	0.9381	0.0619	0.9718	0.0282	0.8932	0.9500	0.9547	0.9325	0.9325	0.8765
MLP	ViT_large	0.9119	0.9017	0.0983	0.9641	0.0359	0.8248	0.9324	0.9318	0.8944	0.8944	0.8112
	Radiomics + Statistical	0.7756	0.7441	0.2559	0.9769	0.0231	0.4402	0.9196	0.8448	0.7085	0.7085	0.5262
	WST	0.8590	0.8528	0.1472	0.9359	0.0641	0.7308	0.8724	0.8924	0.8333	0.8333	0.7093
	Phương pháp đề xuất	0.9359	0.9227	0.0773	0.9795	0.0205	0.8632	0.9619	0.9502	0.9214	0.9214	0.8634

5. THẢO LUẬN

Kết quả thực nghiệm cho thấy phương pháp kết hợp đa đặc trưng đề xuất giúp cải thiện đáng kể hiệu năng phân lớp so với việc sử dụng từng nhóm đặc trưng riêng lẻ, với các chỉ số Accuracy, F1-score, AUC và MCC đều tăng rõ rệt. Điều này cho thấy việc tích hợp nhiều nguồn thông tin giúp nâng cao khả năng phân biệt trong bài toán chẩn đoán ảnh X-quang. Sự cải thiện này xuất phát từ tính bổ sung giữa các loại đặc trưng: đặc trưng học sâu (ViT_large) nắm bắt ngữ cảnh toàn cục, trong khi Radiomics, Statistical và WST cung cấp thông tin về kết cấu, hình dạng và tần số. Nhờ đó, không gian biểu diễn trở nên toàn diện hơn.

Cơ chế Gated Fusion đóng vai trò quan trọng khi cho phép mô hình học cách điều chỉnh đóng góp của từng nhóm đặc trưng một cách linh hoạt, giúp giảm nhiễu và tăng cường các đặc trưng có tính phân biệt cao—đặc biệt phù hợp với dữ liệu y sinh phức tạp. Ngoài ra, khi sử dụng đặc trưng hợp nhất, các bộ phân lớp như KNN, Logistic Regression và MLP cho kết quả tương đương nhau, cho thấy chất lượng đặc trưng quan trọng hơn độ phức tạp mô hình. Điều này mang lại lợi thế thực tiễn khi có thể sử dụng các mô hình đơn giản nhưng vẫn đạt hiệu quả cao.

Khi đối chiếu với một số công trình đã công bố trên cùng bộ dữ liệu Chest X-ray Pneumonia của Kaggle, phương pháp đề xuất cho thấy hiệu năng cạnh tranh và có khả năng cải thiện so với nhiều mô hình học sâu truyền thống. Chẳng hạn, nghiên cứu của Harsh Bhatt và Manan Shah (2023) đã đề xuất một mô hình CNN kết hợp đạt Accuracy 84,12% và F1-score 88,56% trên tập kiểm thử. Trong khi đó, phương pháp đề xuất không chỉ tận dụng đặc trưng học sâu mà còn tích hợp thêm Radiomics, Statistical và WST nhằm khai thác thông tin đa mức từ ảnh X-quang. Kết quả cho thấy phương pháp của chúng tôi vượt trội với Accuracy đạt 94,39% và F1-score đạt 95,59% trên bộ phân lớp KNN. Điều này chứng minh rằng hướng tiếp cận đa đặc trưng kết hợp dựa trên cơ chế Gated Fusion có tiềm năng cải thiện khả năng tổng quát hoá và tăng độ ổn định của mô hình chẩn đoán.

6. KẾT LUẬN

Nghiên cứu này đã đề xuất một phương pháp kết hợp đặc trưng thích nghi cho bài toán phân lớp viêm phổi từ ảnh X-quang lồng ngực. Bằng cách tích hợp nhiều nguồn đặc trưng khác nhau và sử dụng cơ chế Gated Fusion để điều chỉnh trọng số một cách linh hoạt, phương pháp đã xây dựng được biểu diễn dữ liệu có khả năng phân tách tốt giữa các lớp. Kết quả thực nghiệm cho thấy các đặc trưng sau khi kết hợp giúp cải thiện hiệu năng phân lớp và hỗ trợ hiệu quả cho các mô hình học máy truyền thống. Điều này nhấn mạnh vai trò quan trọng của bước biểu diễn đặc trưng trong toàn bộ hệ thống, đồng thời cho thấy rằng việc nâng cao chất lượng đầu vào có thể mang lại hiệu quả đáng kể mà không cần tăng độ phức tạp của mô hình phân lớp.

Trong tương lai, phương pháp có thể được mở rộng theo hướng đánh giá trên các tập dữ liệu đa dạng hơn nhằm kiểm chứng khả năng tổng quát hóa. Bên cạnh đó, việc tích hợp cơ chế Gated Fusion vào các mô hình học sâu đầu-cuối hoặc kết hợp thêm các nguồn dữ liệu khác cũng là những hướng nghiên cứu tiềm năng.

LỜI CẢM ƠN

Các tác giả cảm ơn sự hỗ trợ của Trường Đại học Khoa học, Đại học Huế.

TÀI LIỆU THAM KHẢO

- [1]. Adehbozorgi P, Ryabchykov O, Bocklitz TW. (2025). *A comparative study of statistical, radiomics, and deep learning feature extraction techniques for medical image classification in optical and radiological modalities*. Comput Biol Med. 2025 Mar;187:109768. doi: 10.1016/j.compbiomed.2025.109768. Epub 2025 Jan 31. PMID: 39891957.
- [2]. Arevalo, John & Solorio, Tamar & Montes, Manuel & González, Fabio. (2017). *Gated Multimodal Units for Information Fusion*. 10.48550/arXiv.1702.01992.
- [3]. Baltrusaitis, Tadas & Ahuja, Chaitanya & Morency, Louis-Philippe. (2017). *Multimodal Machine Learning: A Survey and Taxonomy*. 10.48550/arXiv.1705.09406.
- [4]. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. (2020). *An image is worth 16x16 words: Transformers for image recognition at scale*. arXiv preprint arXiv:2010.11929. 2020.
- [5]. Harsh Bhatt, Manan Shah, (2023). A Convolutional Neural Network ensemble model for Pneumonia Detection using chest X-ray images, Healthcare Analytics, Volume 3.
- [6]. Kermany, Daniel & Goldbaum, Michael & Cai, Wenjia, et al. (2018). Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell. 172. 1122-1131.e9. 10.1016/j.cell.2018.02.010.
- [7]. Khánh, N. (2023). Nghiên cứu hiệu quả của các mô hình học sâu tiên tiến trong phân lớp bệnh phổi bằng hình ảnh X-quang ngực. TNU Journal of Science and Technology, 228(15), 45 - 53. doi:https://doi.org/10.34238/tnu-jst.8408
- [8]. Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., van Stiphout, R. G. P. M., Granton, P., Zegers, C. M. L., Gillies, R., Boellard, R., Dekker, A., & Aerts, H. J. W. L. (2012). Radiomics: Extracting more information from medical images using advanced feature analysis. European Journal of Cancer, 48(4), 441-446. https://doi.org/10.1016/j.ejca.2011.11.036
- [9]. LeCun, Y., Bengio, Y. & Hinton, G. (2015). *Deep learning*. Nature 521, 436–444. https://doi.org/10.1038/nature14539
- [10]. Mallat, S. (2012), *Group Invariant Scattering*. Comm. Pure Appl. Math., 65: 1331-1398. https://doi.org/10.1002/cpa.21413
- [11]. Ngiam, Jiquan & Khosla, Aditya & Kim, Mingyu & Nam, Juhan & Lee, Honglak & Ng, Andrew. (2011). *Multimodal Deep Learning*. Proceedings of the 28th International Conference on Machine Learning, ICML 2011. 689-696.
- [12]. Pham, T. A., & Hoang, V. D. (2024). Chest X-ray image classification using transfer learning and hyperparameter customization for lung disease diagnosis. Journal of Information and Telecommunication, 8(4), 587–601. https://doi.org/10.1080/24751839.2024.2317509
- [13]. R. M. Haralick, K. Shanmugam and I. Dinstein, (1973). *Textural Features for Image Classification*, IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-3, no. 6, pp. 610-621, Nov. 1973, doi: 10.1109/TSMC.1973.4309314.
- [14]. Tang, V.H., Duong, S.T.M., Nguyen, C.D.T. et al. (2023). *Wavelet radiomics features from multiphase CT images for screening hepatocellular carcinoma: analysis and comparison*. Sci Rep 13, 19559. https://doi.org/10.1038/s41598-023-46695-8

AN ADAPTIVE FEATURE FUSION FRAMEWORK FOR PNEUMONIA CLASSIFICATION IN CHEST X-RAY IMAGES

Bach Thi Thien Thanh, Nguyen Khanh Linh,
Nguyen Thanh Phong, Nguyen Ngoc Thuy*

Faculty of Information Technology, University of Sciences, Hue University

*Email: nnthuy.cs@hueuni.edu.vn

ABSTRACT

Applying machine learning to pneumonia diagnosis from chest X-ray images is an important and challenging problem. In this study, we propose a feature fusion method that effectively exploits both deep features and handcrafted features. All extracted features are subsequently reduced using PCA and integrated through a Gated Fusion mechanism to learn a unified feature representation. The fused features are then used as input for classifiers such as KNN, Logistic Regression (LR), and MLP for prediction. Experimental results demonstrate that the proposed fusion-based approach significantly outperforms models relying solely on either deep features or handcrafted features. Specifically, on the fused feature set, the KNN classifier achieves the best performance with an accuracy of 94.39% and an F1-score of 95.59%. Logistic Regression yields comparable results with 94.23% accuracy and a 95.47% F1-score, while MLP also achieves strong performance with 93.59% accuracy and a 95.02% F1-score.

Keywords: Gated Fusion, Multi-feature fusion, Deep learning, Chest X-ray, Pneumonia.