

VIETNAM NATIONAL UNIVERSITY, HO CHI MINH CITY
UNIVERSITY OF TECHNOLOGY
FACULTY OF COMPUTER SCIENCE AND ENGINEERING



Đồ án tổng hợp - Hướng trí tuệ nhân tạo

Course Project

“EHR Transfer Learning”

Instructor(s): Nguyễn Tuấn Khôi
Ngô Hoàng Anh

Students: Bùi Đăng Khoa - 2252344
Trần Thành Trọng - 2353238

HO CHI MINH CITY, September 2025



Lời cảm ơn

Chúng em xin gửi lời cảm ơn chân thành đến các thầy Nguyễn Tuấn Khôi, thầy Ngô Hoàng Anh. Các thầy đã trực tiếp hướng dẫn, cung cấp tài liệu và ủng hộ chúng em trong suốt quá trình thực hiện đồ án này. Sự chỉ bảo, hướng dẫn và những kiến thức chuyên môn của các thầy đã giúp chúng em hoàn thiện bài báo cáo với hiểu biết sâu sắc hơn về môn học. Chúng em rất trân trọng những gì Thầy đã đóng góp và hy vọng sẽ tiếp tục nhận được sự hướng dẫn quý báu từ Thầy trong các dự án tương lai. Kính chúc các Thầy luôn mạnh khỏe, hạnh phúc và gặt hái nhiều thành công hơn nữa trong sự nghiệp giảng dạy và nghiên cứu.

Member list & Workload

No.	Fullname	Student ID	Problems	% done
1	Bùi Đăng Khoa	2252344	- Gaussian Noise	100%
2	Trần Thành Trọng	2353238	- SMOTE và XGBClassifier/XGBoost Regressor	100%



Mục lục

1 Tóm tắt	3
2 Giới thiệu	3
2.1 Bối cảnh và các nghiên cứu liên quan	3
2.2 Mục tiêu dự án	4
3 Tổng quan tài liệu	4
3.1 Nghiên cứu dựa trên X-quang ngực (CXR)	4
3.2 Nghiên cứu dựa trên EHR	4
3.3 Các phương pháp kết hợp CXR và EHR	5
3.4 Khoảng trống nghiên cứu trước đó	5
3.5 Dóng góp của dự án	5
3.6 Tổng kết	5
4 Phương pháp	6
4.1 Lựa chọn thiết kế	6
4.1.1 Nguồn dữ liệu	6
4.1.2 Hình thành dữ liệu	6
4.2 Thu thập và tổ chức dữ liệu	7
4.2.1 Xác định nhóm dự án	7
4.2.2 Làm sạch dữ liệu	8
4.3 Dánh giá	8
4.3.1 Xử lý dữ liệu và huấn luyện mô hình – Mô hình 1: SMOTE và XGBClassifier/XGBRegressor	8
4.3.2 Xử lý dữ liệu và huấn luyện mô hình – Mô hình 2: Gaussian Noise	10
5 Kết quả	10
5.1 Mô hình 1: SMOTE và XGBClassifier/XGBRegressor	10
5.2 Mô hình 2: Gaussian Noise	12
6 Thảo luận	16
6.1 Nhận xét kết quả	16
6.1.1 SMOTE và XGBClassifier/XGBRegressor	16
6.1.2 Gaussian Noise	17
6.2 Hạn chế và khuyến nghị	17
6.2.1 SMOTE và XGBClassifier/XGBRegressor	17
6.2.2 Gaussian Noise	18
7 Kết luận	18
8 Tài liệu tham khảo	18



1 Tóm tắt

Chụp X-quang ngực (CXR) là phương pháp phổ biến để phát hiện các bệnh hô hấp mạn tính (CRD). Mặc dù học sâu đã được ứng dụng rộng rãi trên CXR, độ tin cậy vẫn còn đáng ngờ do cơ chế hoạt động chưa rõ ràng. Hồ sơ sức khỏe điện tử (EHR) chứa nhiều thông tin hữu ích nhưng thường không thể sử dụng được do dữ liệu thưa thớt. Việc kết hợp CXR với EHR sử dụng học chuyển giao (Transfer Learning) nhằm cải thiện chẩn đoán 8 loại CRD cả về hiệu suất lẫn khả năng giải thích. Chúng em sử dụng kiến trúc DenseNet để trích xuất 18 đặc trưng tiềm ẩn từ CXR, áp dụng kỹ thuật tăng cường dữ liệu SMOTE, đồng thời sử dụng XGBClassifier và XGBRegressor để xử lý 12 đặc trưng EHR, sau đó kết hợp cả hai nguồn dữ liệu đưa vào mô hình XGBoost. Ngoài ra, nghiên cứu còn đề xuất một hướng tiếp cận khác dựa trên kỹ thuật Gaussian Noise Data Augmentation, trong đó nhiễu Gaussian được bổ sung trực tiếp lên các đặc trưng hình ảnh của lớp thiểu số nhằm tăng tính đa dạng dữ liệu huấn luyện, cải thiện khả năng khái quát hóa của mô hình khi kết hợp với dữ liệu EHR. Phương pháp được đánh giá qua các chỉ số độ chính xác (accuracy), độ chính xác dự đoán dương tính (precision) và độ nhạy (recall) để giải quyết vấn đề mất cân bằng dữ liệu. Kết quả cho thấy cả hai mô hình đều cải thiện dự đoán chẩn đoán dương tính, đồng thời quá trình ra quyết định của XGBoost có sự tương đồng cao với các chuyên gia y tế.

2 Giới thiệu

2.1 Bối cảnh và các nghiên cứu liên quan

Bệnh hô hấp mạn tính (CRD) là một nhóm các rối loạn ảnh hưởng nghiêm trọng đến các cơ quan đường hô hấp, chiếm ít nhất 5% tỷ lệ mắc bệnh và tử vong trên toàn cầu, với tỷ lệ cao hơn ở những khu vực có ô nhiễm không khí hoặc tiếp xúc lâu dài với khói thuốc lá. Việc phát hiện sớm CRD giúp giảm nguy cơ tiến triển nặng, cải thiện chất lượng cuộc sống và giảm chi phí điều trị y tế. Chụp X-quang ngực (CXR) là phương pháp phổ biến và không xâm lấn để chẩn đoán CRD thông qua việc đánh giá các bất thường như các đốm trăng trên hình ảnh. Tuy nhiên, phương pháp này đòi hỏi kiến thức chuyên môn cao và kinh nghiệm lâm sàng, vì các dấu hiệu bệnh lý thường tinh vi và dễ nhầm lẫn.

Gần đây, học máy (ML), đặc biệt là học sâu (DL), đã được áp dụng rộng rãi trên CXR để phát hiện CRD, với độ chính xác cao và khả năng xử lý khối lượng dữ liệu lớn. Các kỹ thuật như mạng nơ-ron tích chập (CNN) đã cải thiện đáng kể tốc độ chẩn đoán và giảm sai sót do con người. Tuy nhiên, việc dựa hoàn toàn vào DL vẫn gặp thách thức, bởi cơ chế ra quyết định của mô hình còn thiếu minh bạch, khiến khó nhận biết các sai lệch hoặc hiểu lầm tiềm ẩn. Các nỗ lực cải thiện tính giải thích, ví dụ như bản đồ nhiệt (heatmap), chỉ cung cấp thông tin trực quan nhưng chưa giải thích đầy đủ cơ sở ra quyết định.

Một vấn đề khác là thiếu dữ liệu dương tính, dẫn đến mất cân bằng dữ liệu nghiêm trọng, khiến các mô hình ML khó học được đặc trưng của lớp ít xuất hiện. Hồ sơ sức khỏe điện tử (EHR) chứa thông tin y tế quan trọng của bệnh nhân nhưng thường bị thiếu hoặc thưa thớt, nên không thể sử dụng trực tiếp trong DL. Các phương pháp imputation truyền thống cũng khó đảm bảo tính chính xác khi dữ liệu thưa thớt. Nhiều nghiên cứu trước đây đã kết hợp DL trên CXR với EHR để cải thiện dự đoán, nhưng hầu hết chưa áp dụng các kỹ thuật tăng cường dữ liệu (data augmentation) nhằm cân bằng lớp và nâng cao khả năng khái quát của mô hình.



2.2 Mục tiêu dự án

Dự án này xây dựng một hệ thống dự đoán đa phương diện, kết hợp học chuyển giao (Transfer Learning) để tích hợp dữ liệu EHR vào chẩn đoán CRD dựa trên CXR. Hệ thống hướng tới vừa duy trì độ chính xác cao của DL, vừa khai thác thông tin hữu ích từ EHR và các mô hình ML truyền thống. Để cải thiện hiệu suất dự đoán khi số lượng mẫu âm tính và dương tính chênh lệch, dự án áp dụng tăng cường dữ liệu cho các đặc trưng, cân bằng số lượng mẫu giữa các lớp, giảm ảnh hưởng của dữ liệu thiếu và nâng cao khả năng học của mô hình. Tập dữ liệu hình ảnh sau tăng cường được kết hợp với các đặc trưng EHR đầy đủ, sẵn sàng cho huấn luyện các mô hình XGBoost và các mô hình ML khác. Quá trình kiểm tra và đánh giá được thực hiện theo hai giai đoạn: đầu tiên đánh giá hiệu suất tổng thể của hệ thống, sau đó phân tích chi tiết quá trình ra quyết định của mô hình. Thông qua dự án, nhóm nghiên cứu kỳ vọng trả lời câu hỏi: việc kết hợp dữ liệu EHR và tăng cường dữ liệu có giúp cải thiện hiệu suất và độ minh bạch trong chẩn đoán CRD hay không.

3 Tổng quan tài liệu

3.1 Nghiên cứu dựa trên X-quang ngực (CXR)

Bệnh hô hấp mạn tính (CRD) là vấn đề y tế toàn cầu nghiêm trọng, gây tỷ lệ tử vong cao và ảnh hưởng đáng kể đến chất lượng cuộc sống. Chẩn đoán sớm đóng vai trò quan trọng trong việc giảm thiểu các biến chứng lâu dài. Trong thập kỷ qua, học sâu (DL) đã trở thành công cụ phổ biến để tự động chẩn đoán CRD từ hình ảnh X-quang ngực (CXR). Các mạng nơ-ron tích chập (CNN) như DenseNet, ResNet, CheXNet được sử dụng rộng rãi nhờ khả năng trích xuất đặc trưng chi tiết, nắm bắt các mẫu phức tạp và phát hiện các bất thường tinh vi mà mắt thường khó nhận ra.

DenseNet được đánh giá cao nhờ khả năng truyền thông tin trực tiếp giữa các lớp, giúp giảm mất mát thông tin và cải thiện học các đặc trưng tinh vi. ResNet sử dụng cơ chế residual giúp mô hình học sâu mà không gặp vấn đề vanishing gradient, còn CheXNet được phát triển đặc biệt cho CXR, đạt hiệu suất tương đương hoặc vượt trội so với chuyên gia y tế trên một số bệnh như viêm phổi.

Tuy nhiên, các nghiên cứu chỉ dựa trên hình ảnh gấp phải vẫn đề mất cân bằng lớp. Các bệnh hiếm như lao, viêm khớp hoặc vẹo cột sống thường có ít mẫu dương tính, dẫn đến recall thấp mặc dù accuracy tổng thể cao. Ngoài ra, các mô hình DL hoạt động như “hộp đen”, khó giải thích các quyết định, khiến việc áp dụng trong lâm sàng còn hạn chế.

3.2 Nghiên cứu dựa trên EHR

Hồ sơ sức khỏe điện tử (EHR) là nguồn dữ liệu lâm sàng phong phú, bao gồm các thông tin như tuổi, giới tính, chỉ số sinh tồn, bệnh lý đi kèm, kết quả xét nghiệm và tiền sử điều trị. Các mô hình ML như XGBoost, Random Forest và Gaussian Naive Bayes (GNB) thường được sử dụng để dự đoán CRD từ EHR nhờ khả năng giải thích và đánh giá tầm quan trọng của từng đặc trưng. Ví dụ, XGBoost có thể xác định ảnh hưởng của tiểu đường, ho, hút thuốc hoặc chỉ số sinh tồn đến nguy cơ CRD, giúp bác sĩ hiểu rõ các yếu tố đóng góp vào dự đoán.



Tuy nhiên, EHR thường thiếu dữ liệu do nhập liệu không đầy đủ hoặc bệnh nhân không được kiểm tra toàn diện. Các phương pháp imputation có thể bổ sung dữ liệu thiếu nhưng với tỷ lệ thưa thoát cao, điều này khó đảm bảo độ chính xác. Nếu không xử lý mất cân bằng lớp, các mô hình ML vẫn khó phát hiện các ca dương tính thuộc lớp thiểu số, dẫn đến sai lệch trong dự đoán.

3.3 Các phương pháp kết hợp CXR và EHR

Một số nghiên cứu gần đây đã kết hợp đặc trưng hình ảnh CXR với dữ liệu EHR nhằm nâng cao hiệu suất dự đoán. Việc kết hợp này không chỉ cải thiện accuracy mà còn giúp tăng interpretability, nhờ khả năng đánh giá tầm quan trọng của từng đặc trưng lâm sàng và hình ảnh. Các mô hình đa mô thức cho thấy khả năng dự đoán toàn diện hơn, phản ánh tốt hơn tình trạng sức khỏe của bệnh nhân.

Tuy nhiên, hầu hết các nghiên cứu chưa áp dụng kỹ thuật tăng cường dữ liệu (data augmentation) cho lớp thiểu số trong CXR, dẫn đến recall và precision thấp cho các bệnh ít gặp, mặc dù accuracy tổng thể cao. Ngoại lệ là một số nghiên cứu về viêm phổi hoặc nhiễm trùng phổi, do mất cân bằng lớp ít nghiêm trọng hơn, mô hình vẫn đạt hiệu quả chấp nhận được.

3.4 Khoảng trống nghiên cứu trước đó

Khoảng trống chính trong nghiên cứu hiện tại là việc xử lý mất cân bằng lớp trong học đa mô thức. Dù DL và ML đều có tiềm năng riêng, có ít bằng chứng về việc tạo dữ liệu tổng hợp hoặc tăng cường dữ liệu có thể nâng cao hiệu suất khi kết hợp ảnh và EHR. Các kỹ thuật như SMOTE và Gaussian Noise Data Augmentation đã hiệu quả trong các lĩnh vực khác nhưng chưa được khai thác nhiều trong chẩn đoán CRD. SMOTE tạo các mẫu nhân tạo cho lớp thiểu số dựa trên khoảng cách giữa các điểm dữ liệu, cải thiện khả năng học từ các ca dương tính hiếm. Gaussian Noise thêm nhiễu nhẹ vào đặc trưng, giúp mô hình học các biểu diễn ổn định và tổng quát hơn.

3.5 Đóng góp của dự án

Dự án này mở rộng nghiên cứu trước bằng cách kết hợp cả đặc trưng CXR và EHR, đồng thời áp dụng augmentation để giảm mất cân bằng lớp. SMOTE được áp dụng trên đặc trưng ảnh để cân bằng mẫu dương và âm, và Gaussian Noise được thêm vào các đặc trưng lớp thiểu số trích xuất từ DenseNet pre-trained. Nghiên cứu đánh giá xem các chiến lược này có cải thiện recall và precision mà không làm giảm accuracy không. So sánh các mô hình với và không có EHR cũng giúp đánh giá trị bổ sung của đặc trưng lâm sàng. Phương pháp này giải quyết các khoảng trống trước đây bằng cách: (i) giảm mất cân bằng lớp qua augmentation, (ii) xác thực vai trò của EHR kết hợp với ảnh, và (iii) đánh giá hiệu suất trên nhiều CRD tiêu biểu để đảm bảo đa dạng bệnh.

3.6 Tổng kết

Tóm lại, nghiên cứu trước đây cho thấy tiềm năng của DL với CXR, ML với EHR, và các phương pháp kết hợp. Tuy nhiên, vấn đề mất cân bằng lớp, hạn chế trong tăng cường dữ liệu và thiếu



đánh giá hệ thống vẫn còn tồn tại. Dự án này khắc phục bằng SMOTE và Gaussian Noise để cải thiện học từ lớp thiểu số, đồng thời kết hợp dữ liệu ảnh và lâm sàng, hướng tới dự đoán đáng tin cậy và có thể giải thích.

4 Phương pháp

4.1 Lựa chọn thiết kế

4.1.1 Nguồn dữ liệu

Dự án trong báo cáo này sử dụng hai nguồn dữ liệu: Medical Information Mart for Intensive Care (MIMIC) và CheXpert.

MIMIC-IV: chứa 299.712 hồ sơ chi tiết về các lần thăm khám của bệnh nhân tại Beth Israel Deaconess Medical Center (BIDMC). Cơ sở dữ liệu này được sử dụng để trích xuất EHR của bệnh nhân, cũng như làm ground truth cho nhiệm vụ này: mã International Classification of Diseases (ICD), ghi lại các bệnh được phát hiện khi bệnh nhân xuất viện.

MIMIC-CXR: bao gồm 227.835 nghiên cứu hình ảnh đã được ẩn danh của 64.588 bệnh nhân tại BIDMC và MIMIC-IV. Cơ sở dữ liệu này được sử dụng để truy xuất hình ảnh CXR của bệnh nhân.

CheXpert: là một bộ dữ liệu CXR khác được sử dụng trong dự án. Nó chứa 224.316 hình ảnh CXR của 65.240 bệnh nhân tại Stanford Hospital, hoàn toàn độc lập với BIDMC. Trong dự án này, bộ dữ liệu được dùng để pre-train mô hình DL đã chọn, nhằm thực hiện Transfer Learning trên dữ liệu CXR từ MIMIC. Việc này đảm bảo tính tổng quát của mô hình đối với dữ liệu thực tế, khi các trường hợp mới chưa được biết trước bởi mô hình đã huấn luyện.

Data access control: nhằm tuân thủ các khía cạnh đạo đức liên quan đến dữ liệu y tế, các tác giả đã hoàn thành tất cả các khóa đào tạo cần thiết để truy cập cơ sở dữ liệu. Để đảm bảo an toàn, dữ liệu sử dụng trong dự án này sẽ không được cung cấp công khai. Thay vào đó, các script lập trình được cung cấp, trong đó yêu cầu đăng nhập để xác nhận trạng thái đào tạo trước khi truy cập và tạo bộ dữ liệu.

4.1.2 Hình thành dữ liệu

Các CRD cần dự đoán: do có quá nhiều CRD, việc bao quát tất cả là không khả thi, vì vậy chúng ta sẽ chọn 8 bệnh tiêu biểu, được kỳ vọng có sự đa dạng về đặc điểm và phân bố dữ liệu. Bảng tóm tắt các bệnh được trình bày trong Figure 1. Bệnh nhân được xác định đã mắc CRD dựa trên mô tả mã ICD được cấp khi xuất viện, lấy trực tiếp từ MIMIC-IV.

Các đặc trưng EHR: dựa trên 8 CRD đã chọn, 12 đặc trưng EHR được lựa chọn để trích xuất từ MIMIC-IV. Đối với các đặc trưng này, chúng ta loại bỏ tất cả thông tin về thời gian nhằm giảm rò rỉ thông tin bệnh nhân.

Các đặc trưng hình ảnh DL: DenseNet là kiến trúc mạng nơ-ron sâu có khả năng duy trì thông tin ở nhiều mức chi tiết khác nhau. Mô hình này được chọn để trích xuất các chi tiết quan trọng

nhất từ hình ảnh CXR. Dự án này sẽ triển khai mô hình CheXpert đã được pre-train sẵn, có sẵn trong gói torchxrayvision. Lớp dense thứ hai tính từ cuối sẽ được chọn để trích xuất đặc trưng, tạo ra 18 đặc trưng tiềm ẩn từ mỗi hình ảnh CXR.

Bệnh	Keywords ICD	Số ca bệnh	% Tỷ lệ của từng bệnh
arthritis	arthritis	1191	4.10%
bronchitis	bronchitis	1485	5.1%
fracture	fracture, broken	2078	7.1%
lung_cancer	tumor, cancer	7787	26.8%
lung_infection	infection, infectious	952	3.2%
pneumonia	pneumonia	7980	27.4%
scoliosis	scoliosis, curvature	150	0.5%
tuberculosis	tuberculosis	771	2.7%

Figure 1: Thống kê số ca và tỷ lệ mắc các bệnh CRD trong tập dữ liệu

4.2 Thu thập và tổ chức dữ liệu

4.2.1 Xác định nhóm dự án

Các bệnh nhân trong MIMIC-IV sẽ bị loại trừ nếu họ không có CXR. Những bệnh nhân còn lại sẽ được kiểm tra mã ICD để xác định xem có khớp với một trong 8 CRD đã chọn hay không, từ đó chia bệnh nhân thành hai nhóm. Tất cả bệnh nhân có mã ICD khớp sẽ được đưa vào nhóm dự án, trong khi những bệnh nhân còn lại, có số lượng lớn hơn đáng kể, sẽ được chọn ngẫu nhiên với số lượng bằng nhau đầu tiên. Quá trình này được tóm tắt trong Figure 2C.

Tiếp theo, 11 đặc trưng EHR sẽ được trích xuất từ cơ sở dữ liệu MIMIC-IV, như minh họa trong Figure 2A. Mỗi bệnh nhân trong nhóm dự án được chọn sẽ có CXR gần nhất trước khi xuất viện được lựa chọn. Việc này giúp đơn giản hóa mô hình và tránh trùng lặp dữ liệu, vì một bệnh nhân có thể có nhiều kết quả CXR trong MIMIC. Trong tập dữ liệu cuối cùng của nhóm dự án, mỗi bệnh nhân dự kiến có 18 đặc trưng hình ảnh và 12 đặc trưng EHR, tổng cộng 30 đặc trưng.

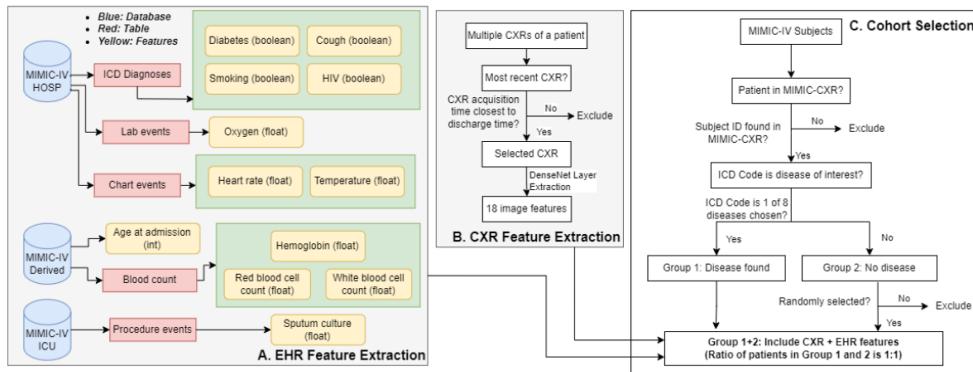


Figure 2: Sơ lược quy trình trích xuất nhóm đối tượng nghiên cứu

4.2.2 Làm sạch dữ liệu

Trong khi các đặc trưng hình ảnh không cần xử lý thêm do giao thức dữ liệu và việc làm sạch trong MIMIC-CXR đã được thực hiện đầy đủ, thì dữ liệu EHR lại chứa một số phép đo bất thường, chẳng hạn như nhiệt độ cơ thể là 150°C. Do đó, để đảm bảo tính toàn vẹn của dữ liệu đối với các đặc trưng liên tục, chúng ta đã loại bỏ bất kỳ phép đo nào là outlier trong phân bố đặc trưng, tức là những giá trị cách trung bình hơn 3 độ lệch chuẩn.

4.3 Dánh giá

Giai đoạn này nhằm mục đích đánh giá hiệu suất của các mô hình ML được chọn trên tập dữ liệu nhóm dự án, giúp làm rõ việc sử dụng Transfer Learning (TL) và dữ liệu EHR có cải thiện hiệu suất chẩn đoán CRD hay không. Trong mỗi quy trình, dữ liệu được chia thành dữ liệu huấn luyện và xác thực theo tỷ lệ 65-35, dự kiến sẽ đảm bảo đa dạng dữ liệu tốt cho việc đánh giá. Có 5 mô hình được chạy đồng thời: mô hình DenseNet gốc được triển khai để thực hiện phân loại nhị phân cho từng CRD, và mỗi hai mô hình ML được chọn sẽ thực hiện dự đoán trên dữ liệu có và không có các đặc trưng EHR. Với mỗi cặp mô hình-CRD, quy trình này được lặp lại 5 lần để thu thập đủ biến thiên.

Các chỉ số đánh giá: Để đánh giá mô hình, chúng tôi chọn độ chính xác (accuracy), độ chính xác dự đoán (precision) và khả năng phát hiện (recall) làm các chỉ số hiệu suất. Những chỉ số này hữu ích để đánh giá khả năng cân bằng giữa dự đoán chính xác và giảm bớt sót các mẫu thuộc lớp thiểu số, đặc biệt khi các trường hợp âm tính giả (False Negatives) được coi là ít mong muốn do hậu quả nghiêm trọng hơn.

4.3.1 Xử lý dữ liệu và huấn luyện mô hình – Mô hình 1: SMOTE và XGBClassifier/XGBRegressor

Trong tập dữ liệu, số lượng nhãn âm tính và dương tính chênh lệch lớn, ảnh hưởng đến khả năng dự đoán. SMOTE được áp dụng để tạo mẫu nhân tạo cho lớp thiểu số, giúp cân bằng dữ liệu và

cải thiện khả năng học của mô hình. Sau đó, XGBoost được huấn luyện trên dữ liệu tăng cường, đồng thời xử lý tốt mất cân bằng lớp và các giá trị thiếu.

SMOTE cho đặc trưng ảnh: SMOTE được áp dụng trên 18 đặc trưng ảnh từ các hình chụp X-quang. Mục đích là đảm bảo rằng số lượng mẫu dương tính và âm tính trong tập huấn luyện gần bằng nhau, giúp mô hình học được các đặc trưng của bệnh một cách chính xác hơn. Sau khi tăng cường, các mẫu hình ảnh mới được kết hợp với dữ liệu EHR còn thiếu để tạo ra tập dữ liệu đầy đủ, sẵn sàng cho bước huấn luyện mô hình.

XGBClassifier và XGBRegressor cho đặc trưng EHR: Dữ liệu EHR bao gồm cả các đặc trưng phân loại (categorical: diabetes, HIV, smoke, cough) và các đặc trưng số (numeric: age, oxygen, heart_rate, temperature, hemoglobin, rbc, wbc). Để dự đoán các đặc trưng này:

- XGBClassifier được dùng cho các đặc trưng phân loại. Nó giúp mô hình xác định giá trị nhãn (0 hoặc 1) dựa trên các đặc trưng của bệnh nhân (Figure 3.B).
- XGBRegressor được dùng cho các đặc trưng số (Figure 3.D). Trước đó, mô hình sẽ dự đoán mask (Figure 3.C) bằng XGBClassifier để biết giá trị đặc trưng có tồn tại hay không, sau đó XGBRegressor dự đoán giá trị thực tế của các đặc trưng số nếu mask = 1 (Figure 3.E).

Sự kết hợp giữa SMOTE, XGBClassifier và XGBRegressor giúp đảm bảo rằng tập dữ liệu đầu vào cho mô hình XGBoost cuối cùng là cân bằng, đầy đủ và phản ánh cả thông tin hình ảnh lẫn thông tin EHR, từ đó nâng cao hiệu suất dự đoán CRD.

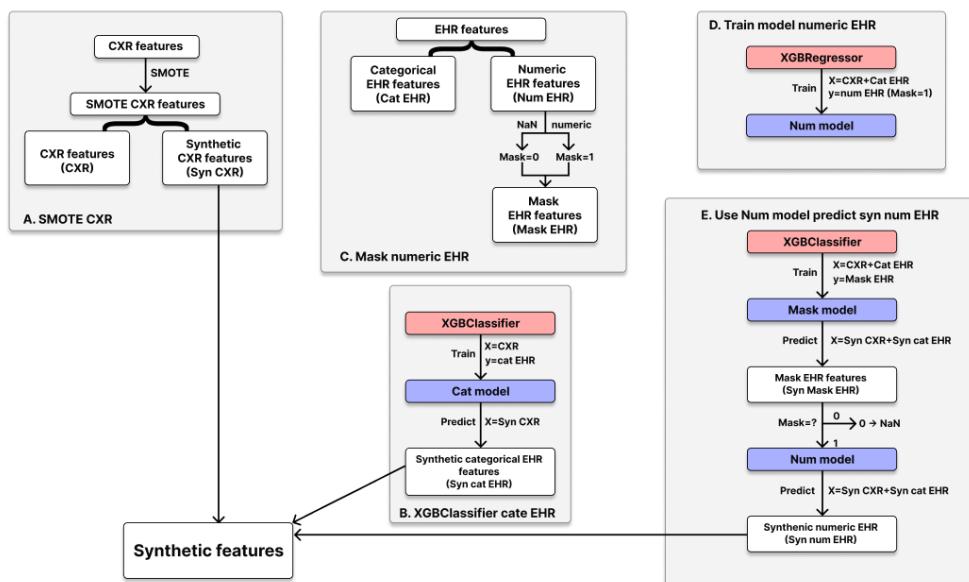


Figure 3: Quy trình tăng cường dữ liệu cho 2 đặc trưng CXR và EHR

4.3.2 Xử lý dữ liệu và huấn luyện mô hình – Mô hình 2: Gaussian Noise

Trong tập dữ liệu nghiên cứu, phân bố nhãn giữa bệnh nhân mắc bệnh và không mắc bệnh có sự mất cân bằng đáng kể, gây ảnh hưởng tiêu cực đến khả năng học của mô hình. Để khắc phục vấn đề này, mô hình thứ hai áp dụng phương pháp *Gaussian Noise Data Augmentation* nhằm tăng cường dữ liệu cho lớp thiểu số trong quá trình huấn luyện.

Gaussian Noise được áp dụng trực tiếp lên các đặc trưng hình ảnh được trích xuất từ mô hình DenseNet đã được huấn luyện trước trên bộ dữ liệu CheXpert. Cụ thể, các mẫu thuộc lớp dương tính trong tập huấn luyện được lấy mẫu lặp lại và bổ sung nhiễu Gaussian với kỳ vọng bằng 0 và độ lệch chuẩn tỷ lệ với độ lệch chuẩn của từng đặc trưng hình ảnh tương ứng. Hệ số nhiễu được lựa chọn ở mức nhỏ nhằm đảm bảo rằng các đặc trưng tổng hợp vẫn nằm trong không gian biểu diễn hợp lý của dữ liệu gốc. Các đặc trưng lâm sàng từ hồ sơ bệnh án điện tử (EHR) trong bộ dữ liệu MIMIC-IV không được áp dụng augmentation nhằm tránh làm sai lệch ý nghĩa y khoa của các biến lâm sàng.

Sau khi thực hiện data augmentation, tập dữ liệu huấn luyện trở nên cân bằng hơn giữa hai lớp, qua đó giúp mô hình học được các biến thể nhẹ của đặc trưng hình ảnh và cải thiện khả năng khái quát hóa. Trên tập dữ liệu đã được xử lý, mô hình Random Forest được huấn luyện với hai cấu hình: (i) sử dụng kết hợp đặc trưng hình ảnh và đặc trưng EHR, và (ii) chỉ sử dụng đặc trưng hình ảnh. Ngoài ra, mô hình Logistic Regression chỉ sử dụng đặc trưng hình ảnh được xây dựng như một mô hình baseline nhằm đánh giá hiệu quả của Transfer Learning và phương pháp Gaussian Noise Data Augmentation.

Cách tiếp cận này cho phép đánh giá riêng rẽ tác động của dữ liệu lâm sàng và Transfer Learning, đồng thời tạo cơ sở cho các phân tích thực nghiệm chi tiết ở các giai đoạn tiếp theo.

5 Kết quả

5.1 Mô hình 1: SMOTE và XGBClassifier/XGBRegressor

Kết quả từ Table 1 cho thấy khi chỉ sử dụng các đặc trưng hình ảnh từ ảnh X-quang ngực (CXR), mô hình XGBoost đạt độ chính xác tổng thể rất cao, dao động từ 71% đến gần 100% tùy thuộc vào từng bệnh. Tuy nhiên, các chỉ số Precision và Recall đều tương đối thấp, đặc biệt với những bệnh có sự mất cân bằng lớp nghiêm trọng như tuberculosis, arthritis và scoliosis. Điều này phản ánh xu hướng của mô hình khi học từ dữ liệu mất cân bằng: mô hình thường dự đoán nhãn âm tính phổ biến hơn, dẫn đến việc bỏ sót nhiều ca dương tính, trong khi các bệnh như pneumonia và lung infection lại có Precision và Recall cao hơn nhờ mức độ mất cân bằng lớp ít nghiêm trọng hơn.

Table 2 minh họa hiệu quả của phương pháp Data Augmentation thông qua SMOTE. Sau khi áp dụng SMOTE trên 18 đặc trưng hình ảnh, Recall của các bệnh như pneumonia tăng đáng kể từ 9,1% lên 42,5%, lung cancer từ 7,8% lên 43,1%. Một số chỉ số Precision giảm nhẹ, nhưng tổng thể, mô hình học được nhiều mẫu dương tính hơn, cho thấy SMOTE giúp giải quyết hiệu quả vấn đề mất cân bằng lớp trong dữ liệu ảnh. Khi kết hợp các đặc trưng EHR vào mô hình, cả Precision và Recall đều được cải thiện rõ rệt, ví dụ pneumonia đạt Precision 70,0% và Recall 37,8%, tuberculosis Precision 96,1% và Recall 19,5%, scoliosis Precision 100% và Recall 24,8%. Điều này chứng tỏ rằng thông tin từ EHR bổ sung giá trị dự đoán, giúp mô hình nhận diện

chính xác hơn các ca bệnh dương tính.

Dánh giá sự thay đổi giữa các thiết lập có và không có Data Augmentation (Table 3) cho thấy SMOTE giúp tăng Recall mạnh mẽ cho hầu hết các bệnh dựa trên đặc trưng ảnh, trong khi ảnh hưởng đến Accuracy và Precision là tương đối nhỏ. Ngược lại, khi kết hợp thêm các đặc trưng EHR, sự thay đổi các chỉ số gần như không đáng kể, hoặc chỉ dương nhẹ, cho thấy rằng EHR duy trì hiệu suất ổn định, nhưng không tác động mạnh mẽ đến cải thiện Recall như dữ liệu ảnh đã được tăng cường. Kết quả này nhấn mạnh vai trò quan trọng của việc cân bằng dữ liệu trong không gian đặc trưng hình ảnh đối với việc cải thiện khả năng nhận diện các ca bệnh dương tính, đồng thời cho thấy việc tích hợp EHR hỗ trợ ổn định kết quả mà không tạo ra ảnh hưởng quá lớn, phản ánh sự phối hợp cân bằng giữa hai nguồn dữ liệu trong quá trình học của mô hình XGBoost.

Model & Disease		Image features only			Image features with EHR features		
Model	Disease	Accuracy	Precision	Recall	Accuracy	Precision	Recall
XGBoost	pneumonia	70.4%	33.6%	9.1%	78.5%	70.1%	36.6%
	tuberculosis	97.3%	9.0%	0.1%	97.8%	95.7%	19.8%
	bronchitis	94.8%	19.7%	0.4%	95.2%	77.9%	10.9%
	arthritis	95.9%	2.9%	0.0%	96.5%	93.7%	16.1%
	fracture	92.7%	11.1%	0.3%	93.2%	71.2%	8.9%
	lung cancer	71.0%	33.9%	7.8%	77.2%	66.3%	31.6%
	scoliosis	99.4%	0.0%	0.0%	99.6%	97.5%	18.9%
	lung infection	96.7%	9.7%	0.2%	97.1%	92.2%	11.8%

Table 1: Kết quả đánh giá khi không có Data Augmentation

Model & Disease		Image features only			Image features with EHR features		
Model	Disease	Accuracy	Precision	Recall	Accuracy	Precision	Recall
XGBoost	pneumonia	58.4%	30.5%	42.5%	78.8%	70.0%	37.8%
	tuberculosis	88.5%	3.7%	12.9%	97.8%	96.1%	19.5%
	bronchitis	81.4%	6.7%	20.6%	95.3%	79.4%	10.0%
	arthritis	83.5%	4.9%	16.1%	96.4%	92.6%	14.7%
	fracture	76.5%	8.6%	24.6%	93.3%	69.3%	8.7%
	lung cancer	58.6%	30.6%	43.1%	77.1%	65.4%	30.8%
	scoliosis	97.8%	0.7%	2.3%	99.6%	100.0%	24.8%
	lung infection	86.3%	4.0%	13.3%	96.9%	94.8%	8.8%

Table 2: Kết quả đánh giá khi thêm Data Augmentation

Model & Disease		Image features only			Image with EHR features		
Model	Disease	Accuracy	Precision	Recall	Accuracy	Precision	Recall
XGBoost	pneumonia	-12.0	-3.1	33.4	0.3	-0.1	1.2
	tuberculosis	-8.8	-5.3	12.8	0.0	0.4	-0.3
	bronchitis	-13.4	-13.0	20.2	0.1	1.5	-0.9
	arthritis	-12.4	2.0	16.1	-0.1	-1.1	-1.4
	fracture	-16.2	-2.5	24.3	0.1	-1.9	-0.2
	lung cancer	-12.4	-3.3	35.3	-0.1	-0.9	-0.8
	scoliosis	-1.6	0.7	2.3	0.0	2.5	5.9
	lung infection	-10.4	-5.7	13.1	-0.2	2.6	-3.0

Table 3: Dánh giá độ lệch giữa không Data Augmentation và có Data Augmentation

5.2 Mô hình 2: Gaussian Noise

Trong giai đoạn đánh giá, nghiên cứu tập trung phân tích hiệu quả của phương pháp *data augmentation* bằng Gaussian Noise trong bối cảnh bài toán phân loại bệnh hô hấp mãn tính (CRD) từ ảnh X-quang ngực. Gaussian Noise được áp dụng chỉ trên các đặc trưng hình ảnh của lớp thiểu số (bệnh nhân mắc bệnh), trong khi các đặc trưng lâm sàng trích xuất từ MIMIC-IV được giữ nguyên nhằm bảo toàn ý nghĩa y khoa của dữ liệu EHR.

Mô hình Random Forest được lựa chọn làm mô hình chính và được đánh giá dưới ba cấu hình khác nhau:

- **Enhanced model (E):** sử dụng đặc trưng hình ảnh (Transfer Learning) kết hợp với dữ liệu lâm sàng từ MIMIC-IV.
- **Original model (O):** chỉ sử dụng đặc trưng hình ảnh.
- **Baseline model (L):** mô hình Logistic Regression chỉ sử dụng đặc trưng hình ảnh.

Kết quả được tổng hợp dưới dạng giá trị trung bình qua 5 lần chạy cho 8 bệnh CRD, bao gồm: pneumonia, tuberculosis, bronchitis, arthritis, fracture, lung cancer, scoliosis và lung infection.

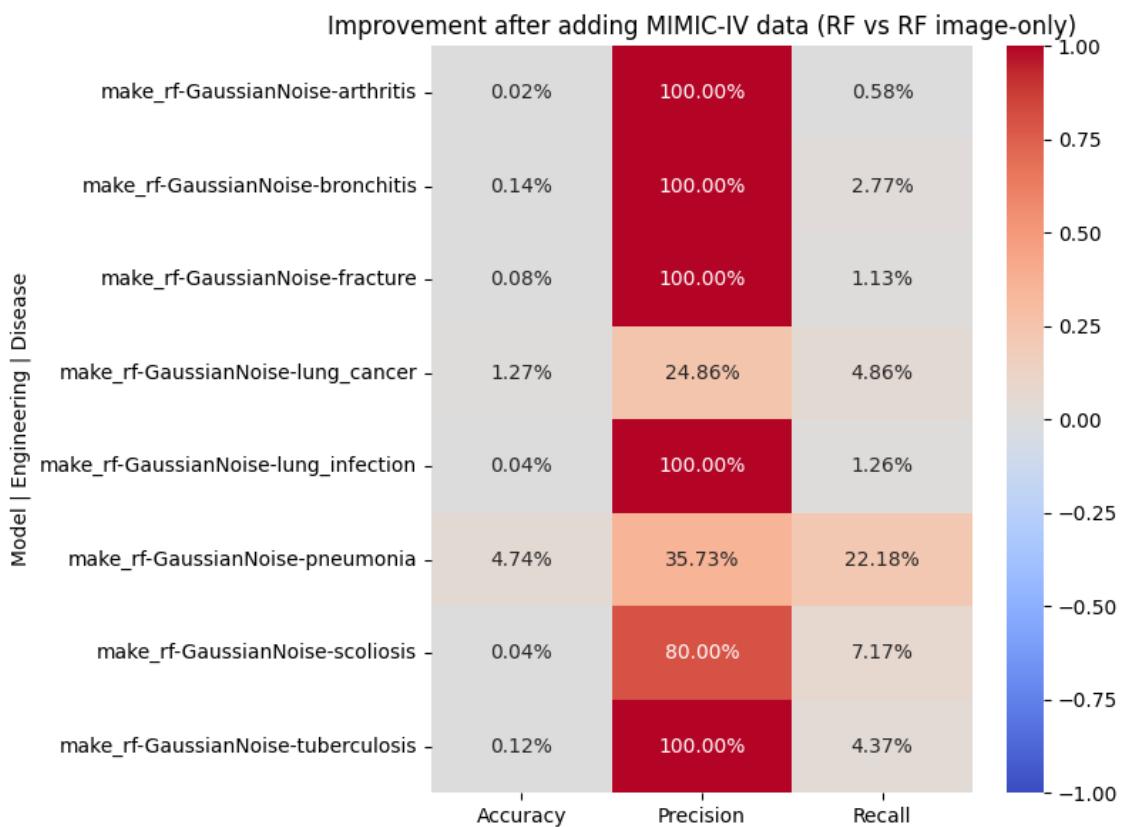


Figure 4: RF vs RF image-only

Kết quả của mô hình Gaussian Noise cho thấy một trade-off rõ ràng giữa Precision và Recall, đặc trưng cho các bài toán phân loại y sinh với dữ liệu mất cân bằng. Việc tăng mạnh Precision cho thấy mô hình ít đưa ra dự đoán bệnh sai, điều này đặc biệt quan trọng trong các hệ thống hỗ trợ chẩn đoán nhằm tránh can thiệp y tế không cần thiết.

Tuy nhiên, sự suy giảm Recall cho thấy mô hình có xu hướng bỏ sót một số ca bệnh. Trong bối cảnh y khoa, Recall cao giúp giảm nguy cơ bỏ sót bệnh nhân. Precision cao giúp giảm báo động giả. Do đó, mô hình Gaussian Noise phù hợp hơn cho các ứng dụng hỗ trợ quyết định hoặc sàng lọc bước đầu, nơi độ tin cậy của dự đoán được ưu tiên.

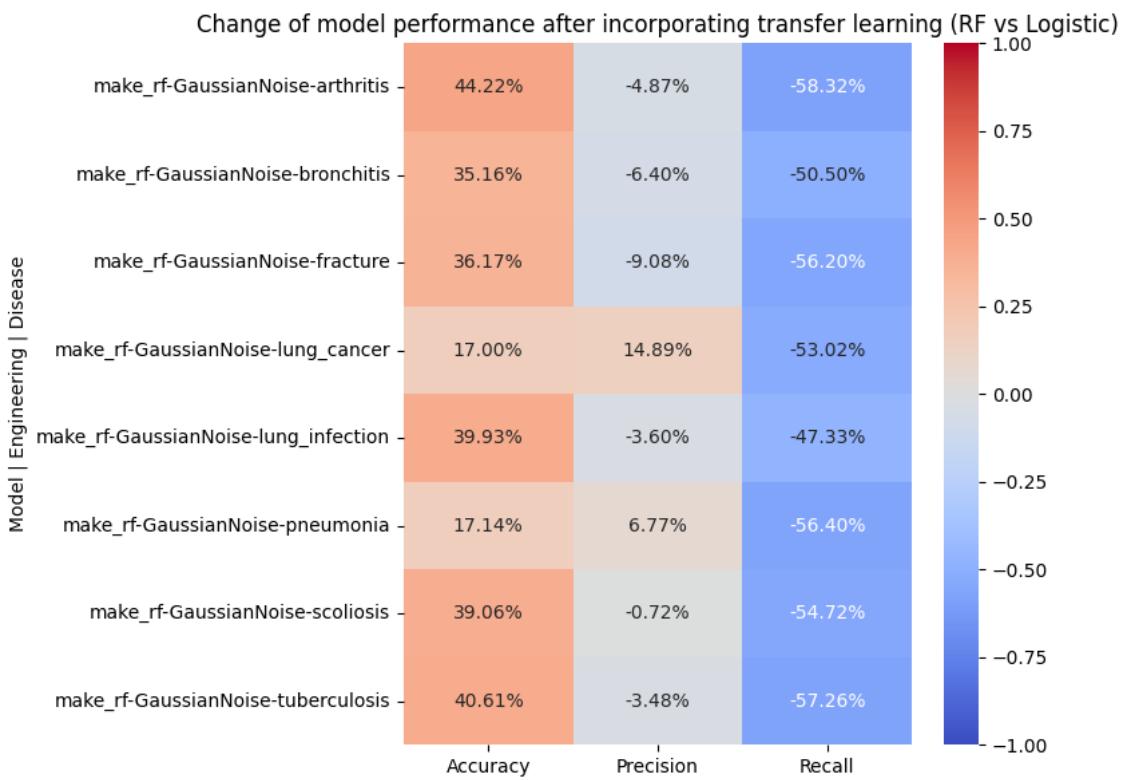


Figure 5: RF vs Logistic

Khi so sánh Random Forest sử dụng embedding DenseNet (image-only) với Logistic Regression image-only, kết quả cho thấy Accuracy tăng đáng kể trên cả 8 bệnh, với mức cải thiện dao động từ khoảng 17% đến hơn 40%. Điều này chứng tỏ các đặc trưng hình ảnh trích xuất từ mô hình deep learning pre-trained trên CheXpert mang thông tin ngữ nghĩa phong phú hơn so với các mô hình tuyến tính truyền thống.

Tuy nhiên, kết quả cũng cho thấy Recall của lớp dương tính giảm mạnh, thường từ 50% đến gần 60%. Hiện tượng này phản ánh xu hướng mô hình trả lời bảo thủ hơn: chỉ dự đoán bệnh khi mức độ chắc chắn cao, từ đó làm giảm số ca bệnh được phát hiện.

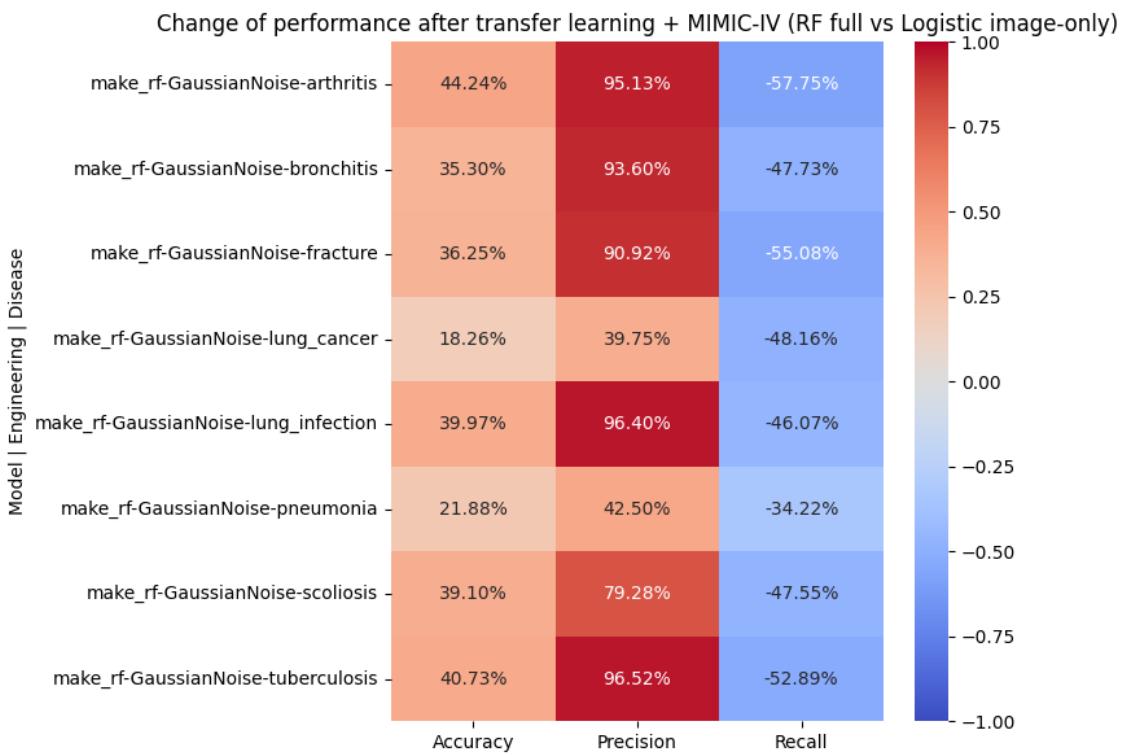


Figure 6: RF full vs Logistic image-only

So sánh giữa Enhanced Random Forest (Gaussian Noise + image + EHR) và Baseline Logistic Regression image-only cho thấy hiệu quả tổng hợp rõ rệt. Accuracy và Precision đều tăng ổn định trên toàn bộ 8 bệnh, trong đó Precision đạt mức cải thiện rất cao (có bệnh vượt quá 90%). Điều này khẳng định rằng việc kết hợp Transfer Learning, Gaussian Noise augmentation và dữ liệu lâm sàng giúp mô hình đạt được độ chính xác cao nhất.

Tuy nhiên, Recall của lớp dương tính tiếp tục giảm so với baseline, cho thấy mô hình tổng hợp là cấu hình thận trọng nhất trong việc dự đoán bệnh.

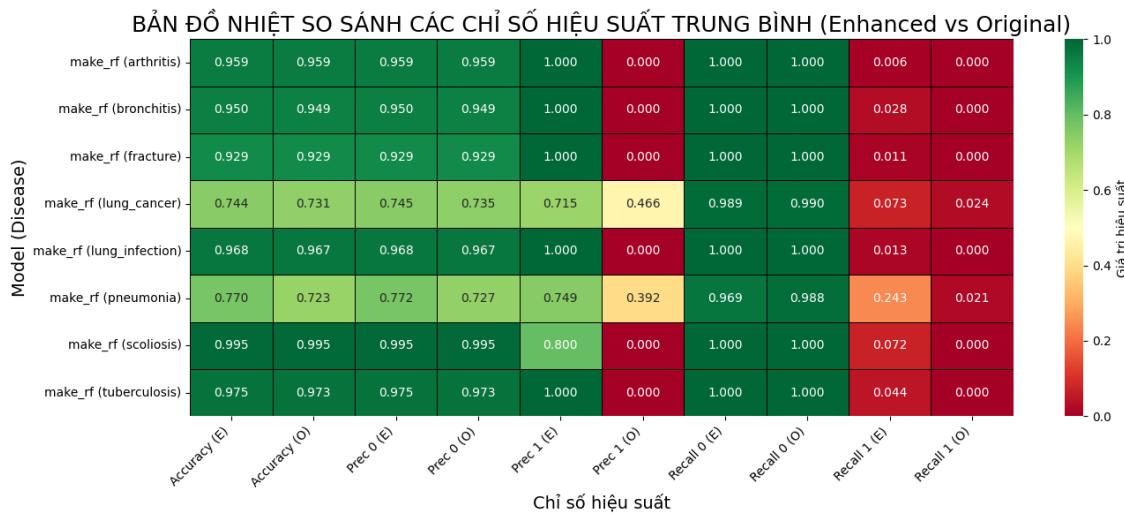


Figure 7: Performance heatmap (Gaussian Noise)

Bản đồ nhiệt so sánh chi tiết các chỉ số (Accuracy, Precision và Recall cho cả hai lớp) cho thấy:

- Precision lớp dương của mô hình Enhanced luôn cao hơn mô hình Original.
- Recall lớp âm gần như đạt 1.0 ở cả hai mô hình, cho thấy khả năng nhận diện ca không bệnh rất tốt.
- Recall lớp dương của mô hình Enhanced thấp hơn Original, xác nhận sự tồn tại của trade-off giữa Precision và Recall.

6 Thảo luận

6.1 Nhận xét kết quả

6.1.1 SMOTE và XGBClassifier/XGBRegressor

Kết quả thực nghiệm cho thấy khi mô hình XGBoost chỉ sử dụng các đặc trưng hình ảnh trích xuất từ CXR, độ chính xác tổng thể (Accuracy) đạt mức cao. Tuy nhiên, Precision và Recall lại thấp đối với các bệnh có mức độ mất cân bằng lớp nghiêm trọng, cho thấy mô hình có xu hướng dự đoán phần lớn các mẫu là âm tính. Hiện tượng này phản ánh hạn chế phổ biến của các mô hình học máy khi làm việc với dữ liệu y sinh không cân bằng, trong đó các ca bệnh dương tính hiếm bị mô hình bỏ sót, làm giảm khả năng phát hiện bệnh trong thực tế lâm sàng.

Việc áp dụng kỹ thuật SMOTE trên các đặc trưng hình ảnh đã giúp cải thiện đáng kể Recall ở hầu hết các bệnh, đặc biệt là các bệnh có tỷ lệ dương tính thấp. Nhờ việc tạo thêm các mẫu tổng hợp cho lớp thiểu số, mô hình học được các đặc trưng đại diện tốt hơn cho các ca bệnh hiếm, từ đó giảm hiện tượng thiên lệch về lớp âm tính. Mặc dù Precision có xu hướng giảm trong một



số trường hợp, sự gia tăng Recall cho thấy SMOTE giúp cải thiện khả năng phát hiện ca bệnh, vốn là yếu tố quan trọng trong các bài toán chẩn đoán y khoa.

Tổng thể, kết quả cho thấy việc kết hợp tăng cường dữ liệu ảnh và tích hợp dữ liệu EHR có thể giúp cải thiện hiệu suất dự đoán CRD ở một mức độ nhất định, đặc biệt đối với các lớp thiểu số, đồng thời nâng cao tính phù hợp của mô hình trong các ứng dụng hỗ trợ chẩn đoán lâm sàng.

6.1.2 Gaussian Noise

Kết quả của mô hình Gaussian Noise cho thấy một trade-off rõ ràng giữa Precision và Recall, đặc trưng cho các bài toán phân loại y sinh với dữ liệu mất cân bằng. Việc tăng mạnh Precision cho thấy mô hình ít đưa ra dự đoán bệnh sai, điều này đặc biệt quan trọng trong các hệ thống hỗ trợ chẩn đoán nhằm tránh can thiệp y tế không cần thiết.

Tuy nhiên, sự suy giảm Recall cho thấy mô hình có xu hướng bỏ sót một số ca bệnh. Trong bối cảnh y khoa, Precision và Recall đều có ý nghĩa quan trọng: Precision cao giúp giảm báo động giả, trong khi Recall cao giúp giảm nguy cơ bỏ sót bệnh nhân. Do đó, mô hình Gaussian Noise phù hợp hơn cho các ứng dụng hỗ trợ quyết định hoặc sàng lọc bước đầu, nơi độ tin cậy của dự đoán được ưu tiên.

6.2 Hạn chế và khuyến nghị

6.2.1 SMOTE và XGBClassifier/XGBRegressor

Mặc dù các đặc trưng hình ảnh (CXR) sau khi được tăng cường bằng SMOTE được sử dụng như một dạng dữ liệu bổ sung để huấn luyện mô hình XGBoost cùng với các đặc trưng EHR, kết quả cho thấy hiệu quả cải thiện là không đáng kể. Cụ thể, các chỉ số Precision, Recall và Accuracy của mô hình sử dụng EHR gần như không thay đổi so với trường hợp không áp dụng SMOTE, hoặc chỉ tăng ở mức rất nhỏ.

Điều này cho thấy việc tăng cường dữ liệu chỉ dựa trên các đặc trưng hình ảnh không đủ để bù đắp cho sự khan hiếm và thiếu thông tin trong EHR. Mặc dù số lượng mẫu huấn luyện tăng lên, các mẫu tổng hợp được tạo bởi SMOTE không cung cấp thêm thông tin lâm sàng mới cho EHR, khiến mô hình không học được các mối quan hệ có ý nghĩa hơn giữa đặc trưng hình ảnh và đặc trưng lâm sàng.

Ngoài ra, việc Recall tăng khi chỉ sử dụng đặc trưng hình ảnh sau SMOTE nhưng đi kèm với sự suy giảm Precision cho thấy mô hình có xu hướng dự đoán dương tính quá mức. Khi kết hợp với EHR, hiện tượng này không được cải thiện rõ rệt, dẫn đến việc chiến lược augmentation này chưa thực sự hữu dụng cho mô hình đa mô thức trong bối cảnh dữ liệu EHR thừa thớt.

Ngoài ra số lượng CRD và CXR mỗi bệnh nhân còn hạn chế, ground truth dựa vào mã ICD khi xuất viện có thể sai số, và thiếu sự tham gia của chuyên gia y tế để hướng dẫn mô hình học các đặc trưng quan trọng.



6.2.2 Gaussian Noise

Mặc dù đạt được kết quả khả quan, phương pháp Gaussian Noise vẫn tồn tại một số hạn chế. Thứ nhất, Gaussian Noise là một dạng augmentation tuyến tính, có thể làm sai lệch cấu trúc ngữ nghĩa tiềm ẩn của embedding hình ảnh. Thứ hai, dữ liệu EHR không được augment, dẫn đến khả năng mô hình vẫn bị giới hạn bởi sự khan hiếm thông tin lâm sàng ở một số bệnh.

Ngoài ra, mô hình chưa được tối ưu hóa ngưỡng phân loại theo mục tiêu y khoa cụ thể, điều này có thể cải thiện Recall nếu được điều chỉnh phù hợp. Trong tương lai, các hướng nghiên cứu có thể bao gồm việc kết hợp Gaussian Noise với các kỹ thuật augmentation khác, tối ưu ngưỡng dự đoán, hoặc fine-tune trực tiếp mô hình deep learning trên dữ liệu MIMIC-CXR.

7 Kết luận

Dự án này đã phát triển một hệ thống chẩn đoán bệnh hô hấp mạn tính (CRD) bằng cách tích hợp đặc trưng hình ảnh X-quang ngực (CXR) và dữ liệu hồ sơ bệnh án điện tử (EHR). Trên nền tảng đó, các phương pháp xử lý mắt cánh bằng dữ liệu đã được đánh giá nhằm cải thiện hiệu suất mô hình. Kết quả cho thấy SMOTE giúp cải thiện đáng kể Recall đối với các đặc trưng hình ảnh, trong khi các đặc trưng EHR gần như không thay đổi, phản ánh tác động hạn chế của SMOTE trong bối cảnh dữ liệu lâm sàng này. Bên cạnh đó, mô hình XGBoost thể hiện hiệu suất dự đoán cao và khả năng giải thích tốt thông qua phân tích tầm quan trọng đặc trưng, với các mẫu đặc trưng phù hợp với nhận định của chuyên gia y tế.

Song song với SMOTE, dự án cũng đề xuất và đánh giá phương pháp Gaussian Noise Data Augmentation như một hướng tiếp cận thay thế để tăng cường hiệu suất chẩn đoán CRD. Gaussian Noise được áp dụng trực tiếp lên các đặc trưng hình ảnh trích xuất từ mô hình DenseNet pre-trained, tập trung vào lớp thiểu số nhằm giảm ảnh hưởng của mắt cánh bằng dữ liệu trong không gian đặc trưng. Thực nghiệm trên 8 bệnh hô hấp sử dụng dữ liệu MIMIC-IV cho thấy việc kết hợp Gaussian Noise với Transfer Learning giúp cải thiện rõ rệt độ chính xác dự đoán dương tính (precision) và độ chính xác tổng thể (accuracy) so với các mô hình baseline chỉ sử dụng đặc trưng hình ảnh. Điều này cho thấy tăng cường dữ liệu trong không gian đặc trưng hình ảnh có thể giúp mô hình học được các biểu diễn ổn định và đáng tin cậy hơn.

Mặc dù vẫn tồn tại các hạn chế như số lượng mẫu CRD còn hạn chế, số lượng ảnh CXR trên mỗi bệnh nhân chưa nhiều, việc phụ thuộc vào mã ICD và thiếu sự tham gia trực tiếp của chuyên gia lâm sàng trong quá trình gán nhãn và đánh giá, hệ thống được đề xuất vẫn cho thấy tiềm năng ứng dụng thực tế. Các nghiên cứu trong tương lai nên mở rộng tập dữ liệu, kết hợp ý kiến chuyên gia, và thử nghiệm mô hình trong các môi trường y tế đa dạng nhằm nâng cao độ tin cậy, khả năng khái quát và tính ứng dụng lâm sàng của hệ thống.

8 Tài liệu tham khảo

- [1] A. L. Byrne, B. J. Marais, C. D. Mitnick, L. Lecca, and G. B. Marks, “Tuberculosis and chronic respiratory disease: A systematic review,” *International Journal of Infectious Diseases*, vol. 32, pp. 138–146, Mar. 2015.

- [2] C. J. L. Murray and A. D. Lopez, "Measuring the global burden of disease," *New England Journal of Medicine*, vol. 369, no. 5, pp. 448–457, Aug. 2013.
- [3] E. A. Simoes, T. Cherian, J. Chow, S. A. Shahid-Salles, R. Laxminarayan, and T. J. John, *Disease Control Priorities in Developing Countries*, World Bank, 2006.
- [4] D. Meedeniya, H. Kumarasinghe, S. Kolonne, C. Fernando, I. De la Torre Díez, and G. Marques, "Chest X-ray analysis empowered with deep learning: A systematic review," *Applied Soft Computing*, vol. 126, Art. no. 109319, Sep. 2022.
- [5] B. van Ginneken, B. M. ter Haar Romeny, and M. A. Viergever, "Computer-aided diagnosis in chest radiography: A survey," *IEEE Transactions on Medical Imaging*, vol. 20, no. 12, pp. 1228–1241, 2001.
- [6] A. Ait Nasser and M. A. Akhloufi, "A review of recent advances in deep learning models for chest disease detection using radiography," *Diagnostics*, vol. 13, no. 1, Art. no. 159, Jan. 2023.
- [7] N. P. Tembhare, P. U. Tembhare, and C. U. Chauhan, "Chest X-ray analysis using deep learning," *International Journal for Research in Applied Science and Engineering Technology*, vol. 11, no. 1, pp. 1441–1447, Jan. 2023.
- [8] C. M. Jones, Q. D. Buchlak, L. Oakden-Rayner, M. Milne, J. Seah, N. Esmaili, and B. Hachey, "Chest radiographs and machine learning – past, present and future," *Journal of Medical Imaging and Radiation Oncology*, vol. 65, no. 5, pp. 538–544, Jun. 2021.
- [9] A. J. Larrazabal, N. Nieto, V. Peterson, D. H. Milone, and E. Ferrante, "Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis," *Proceedings of the National Academy of Sciences*, vol. 117, no. 23, pp. 12592–12594, May 2020.
- [10] S. Singla, M. Eslami, B. Pollack, S. Wallace, and K. Batmanghelich, "Explaining the black-box smoothly—a counterfactual approach," *Medical Image Analysis*, vol. 84, Art. no. 102721, Feb. 2023.
- [11] 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 Int. Conf. Computer Vision (ICCV)*, pp. 618–626, 2017.
- [12] X. Wang et al., "Machine learning-enabled risk prediction of chronic obstructive pulmonary disease with unbalanced data," *Computer Methods and Programs in Biomedicine*, vol. 230, Art. no. 107340, Mar. 2023.
- [13] L. Álvarez-Rodríguez, J. de Moura, J. Novo, and M. Ortega, "Does imbalance in chest X-ray datasets produce biased deep learning approaches for COVID-19 screening?" *BMC Medical Research Methodology*, vol. 22, no. 1, Apr. 2022.
- [14] J. Irvin et al., "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proc. AAAI Conf. Artificial Intelligence*, vol. 33, no. 1, pp. 590–597, Jul. 2019.
- [15] D. Aletaha and J. S. Smolen, "Diagnosis and management of rheumatoid arthritis," *JAMA*, vol. 320, no. 13, pp. 1360–1372, Oct. 2018.



- [16] J. A. Janicki and B. Alman, "Scoliosis: Review of diagnosis and treatment," *Paediatrics & Child Health*, vol. 12, no. 9, pp. 771–776, Nov. 2007.
- [17] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, 2017.
- [18] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, pp. 785–794, 2016.
- [19] C. Mottez, "Enhancing Chest X-ray Diagnosis for Respiratory Diseases using Transfer Learning, https://github.com/clemence-mottez/mimic_iv/tree/main", accessed on Oct. 16, 2025.



PHỤ LỤC

Link github: <https://github.com/tronglinux123/AI-Project>