

Dự đoán tác dụng phụ từ dữ liệu VAERS mRNA COVID-19 vaccine

Tăng Trần Mạnh Hưng¹
hung.ttm230037@sis.hust.edu.vn

Tóm tắt nội dung

Dự án này sử dụng dữ liệu từ hệ thống VAERS để xây dựng mô hình học máy dự đoán nguy cơ phản ứng nghiêm trọng sau tiêm vaccine mRNA COVID-19. Mục tiêu chính là phân tích mối quan hệ giữa các triệu chứng và yếu tố nhân khẩu học như tuổi, giới tính và bệnh nền, từ đó hỗ trợ công tác giám sát an toàn vaccine và cải thiện khả năng cảnh báo sớm các nguy cơ tiềm ẩn.

Mục lục

1	Giới thiệu Dự án	2
2	Quy Trình Xử Lý Dữ Liệu VAERS	2
2.1	Nguồn Dữ Liệu và Phạm Vi Dự Án	2
2.2	Quy Trình Tiền Xử Lý Dữ Liệu	3
2.2.1	Tích hợp và làm sạch dữ liệu	3
2.2.2	Kỹ thuật biến đổi dữ liệu	3
2.3	Phân Tích Khám Phá Dữ liệu (EDA)	3
2.3.1	Thống kê mô tả	3
2.3.2	Trực quan hóa dữ liệu	4
2.4	Chuẩn bị Dữ liệu cho Mô hình Học Máy	4
2.4.1	Bài toán dự đoán mức độ nghiêm trọng	4
2.4.2	Bài toán dự đoán triệu chứng từ thông tin cá nhân	5
2.5	Xử lý mất cân bằng dữ liệu và chỉ số đánh giá	5
2.6	Hạn Chế của Dữ Liệu VAERS	6
3	Dự đoán mức độ nghiêm trọng	6
3.1	Thiết lập thực nghiệm	6
3.2	Kết quả	7
3.2.1	Kết quả dự đoán với dữ liệu gốc (chưa qua giảm chiều dữ liệu)	7
3.2.2	Đặc trưng quan trọng cho dự đoán	7
3.2.3	So sánh hiệu quả giảm chiều với PCA và LDA	8
4	Dự đoán triệu chứng	8
5	Kết luận	9
5.1	Tổng kết kết quả	10
5.2	Hạn chế và hướng phát triển	10
5.3	Ý nghĩa thực tiễn	10

¹Chương trình Tài năng; Khoa Khoa học Máy tính; Trường Công nghệ Thông tin và Truyền thông; Đại học Bách Khoa Hà Nội

1 Giới thiệu Dự án

Vaccine mRNA, đặc biệt trong bối cảnh đại dịch COVID-19, đã đóng vai trò then chốt trong việc bảo vệ sức khỏe cộng đồng. Tuy phần lớn tác dụng phụ là nhẹ, một số phản ứng nghiêm trọng dù hiếm gặp vẫn cần được theo dõi chặt chẽ [11]. Trong đó, hệ thống *Vaccine Adverse Event Reporting System (VAERS)* do CDC và FDA phối hợp vận hành cung cấp nguồn dữ liệu phong phú để nghiên cứu các phản ứng bất lợi sau tiêm chủng [9].

Dữ liệu VAERS bao gồm thông tin về triệu chứng, thời gian khởi phát, mức độ nghiêm trọng, cũng như các yếu tố nhân khẩu học như tuổi, giới tính và tiền sử bệnh nền. Tuy nhiên, đặc điểm không cấu trúc và sự thiếu đồng nhất của dữ liệu khiến việc phân tích thủ công trở nên khó khăn. Trong bối cảnh đó, các kỹ thuật học máy đã chứng minh được hiệu quả trong việc khai thác dữ liệu phức tạp để nhận diện các yếu tố nguy cơ và dự đoán phản ứng nghiêm trọng sau tiêm vaccine [7].

Dù COVID-19 đã qua giai đoạn khẩn cấp, việc tiếp tục nghiên cứu dữ liệu tiêm chủng vẫn mang ý nghĩa quan trọng, góp phần tăng cường an toàn vaccine và chuẩn bị cho các chiến dịch tiêm chủng trong tương lai.

Dự án này hướng đến xây dựng một mô hình học máy có khả năng dự đoán mức độ nghiêm trọng của các tác dụng phụ sau tiêm vaccine mRNA dựa trên thông tin triệu chứng và yếu tố nhân khẩu học [2]. Mô hình nhằm hỗ trợ cơ quan y tế trong giám sát chủ động và cảnh báo sớm.

Đóng góp chính của dự án:

- Đề xuất pipeline xử lý và phân tích dữ liệu VAERS phục vụ bài toán phân loại mức độ nghiêm trọng của tác dụng phụ sau tiêm vaccine mRNA.
- Thực nghiệm so sánh hiệu quả giữa các mô hình học máy như Logistic Regression, Random Forest, XGBoost.
- Phân tích vai trò của thông tin triệu chứng và nhân khẩu học trong việc dự đoán các phản ứng nghiêm trọng.

2 Quy Trình Xử Lý Dữ Liệu VAERS

2.1 Nguồn Dữ Liệu và Phạm Vi Dự Án

Dự án phân tích dữ liệu từ Hệ thống Báo cáo Biến cố Bất lợi Sau Tiêm chủng (VAERS) giai đoạn 2020-2021, bao gồm:

- **Dữ liệu nhân khẩu học:** Thông tin cơ bản về người được tiêm chủng (tuổi, giới tính, vùng miền)
- **Dữ liệu vaccine:** Loại vaccine, nhà sản xuất, ngày tiêm, số mũi tiêm
- **Dữ liệu triệu chứng:** Các phản ứng phụ được báo cáo sau tiêm, mức độ nghiêm trọng

2.2 Quy Trình Tiền Xử Lý Dữ Liệu

2.2.1 Tích hợp và làm sạch dữ liệu

- Thực hiện inner join 3 bảng dữ liệu (VAERSDATA, VAERSVAX, VAERSSYMPTOMS) sử dụng trường khóa VAERS_ID
- Lọc chỉ giữ lại các bản ghi:
 - Thuộc vaccine COVID-19 (VAX_TYPE = "COVID19")
 - Có đầy đủ thông tin về ngày tiêm và ngày khởi phát triệu chứng
 - Lọc dữ liệu trong khoảng thời gian từ ngày 20/12/2020 đến ngày 22/4/2021 (sử dụng trường ngày tiêm vaccine)
 - Do hạn chế tài nguyên, chỉ lấy 100,000 bản ghi.
- Xử lý các giá trị missing:
 - Điền giá trị mặc định cho các trường phân loại
 - Loại bỏ các bản ghi thiếu thông tin quan trọng

2.2.2 Kỹ thuật biến đổi dữ liệu

- Mã hóa biến phân loại:
 - Giới tính: Mã hóa nhị phân (Nữ=1, Nam=0)
 - Nhà sản xuất: One-hot encoding (Moderna, Pfizer/BioNTech, Janssen, Khác)
- Xử lý tiền sử bệnh:
 - Sử dụng kỹ thuật text mining để trích xuất các bệnh nền quan trọng
 - Tạo các biến nhị phân đại diện cho từng bệnh nền
- Xử lý dữ liệu thời gian:
 - Chuẩn hóa định dạng ngày tháng theo ISO 8601
 - Tính toán thời gian phản ứng (số ngày từ khi tiêm đến khi xuất hiện triệu chứng)
 - Áp dụng cửa sổ 50 ngày để tập trung vào phản ứng cấp tính

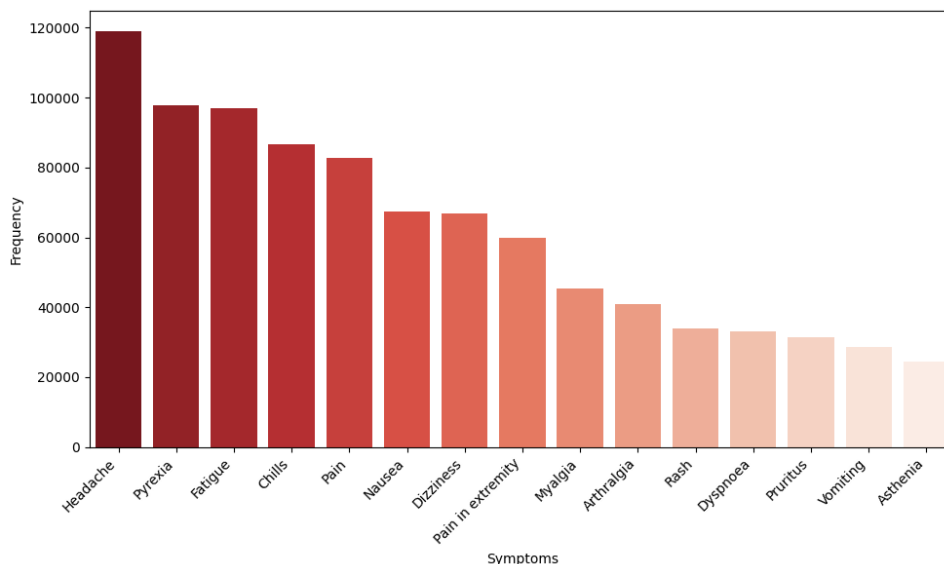
2.3 Phân Tích Khám Phá Dữ liệu (EDA)

2.3.1 Thống kê mô tả

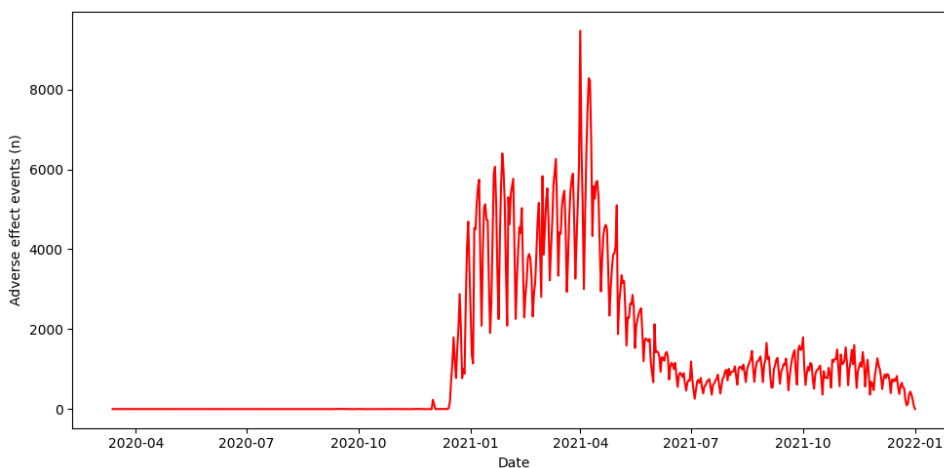
- Phân bố tần suất 15 triệu chứng phổ biến nhất (không bao gồm phản ứng tại chỗ tiêm)
- Phân tích xu hướng thời gian: Số lượng báo cáo theo ngày/tuần
- So sánh tỷ lệ phản ứng giữa các nhà sản xuất vaccine

2.3.2 Trực quan hóa dữ liệu

- Biểu đồ cột thể hiện tần suất triệu chứng (Hình 1)
- Biểu đồ đường thể hiện xu hướng báo cáo theo thời gian (Hình 2)



Hình 1: Biểu đồ cột tần suất 15 triệu chứng phổ biến nhất



Hình 2: Biểu đồ đường thể hiện xu hướng báo cáo theo thời gian

2.4 Chuẩn bị Dữ liệu cho Mô hình Học Máy

Để đảm bảo hiệu quả huấn luyện trong điều kiện hạn chế tài nguyên tính toán, đồng thời duy trì tính đại diện của dữ liệu, dự án lựa chọn một tập con gồm 100.000 bản ghi được lấy ngẫu nhiên từ toàn bộ dữ liệu.

2.4.1 Bài toán dự đoán mức độ nghiêm trọng

1. Xác định biến mục tiêu:

- Kết hợp các trường DIED, L_THREAT, DISABLE, HOSPITAL, ER_VISIT
- Mã hóa nhị phân: 1 (nghiêm trọng) nếu có ít nhất một trong các hậu quả trên

2. Xử lý dữ liệu triệu chứng:

- Kết hợp 5 trường SYMPTOM1-5 thành một danh sách triệu chứng
- Áp dụng kỹ thuật multi-hot encoding cho các triệu chứng

3. Chia tập dữ liệu:

- Tỷ lệ 80-20 cho huấn luyện và kiểm tra

2.4.2 Bài toán dự đoán triệu chứng từ thông tin cá nhân

1. Lựa chọn 20 triệu chứng quan trọng làm biến mục tiêu

2. Tiền xử lý đặc trưng đầu vào:

- Chuẩn hóa các biến liên tục (tuổi, thời gian phản ứng)
- Mã hóa one-hot cho biến phân loại (nhà sản xuất, vùng miền)

3. Xử lý dữ liệu missing:

- Điền giá trị trung bình/trung vị cho biến liên tục
- Điền giá trị mốc cho biến phân loại

4. Chia tập dữ liệu:

- Tỷ lệ 80-20 cho huấn luyện và kiểm tra
- Phân tầng theo biến mục tiêu để đảm bảo tính đại diện

2.5 Xử lý mất cân bằng dữ liệu và chỉ số đánh giá

Dữ liệu huấn luyện có sự mất cân bằng giữa các lớp (trong đó ở bài toán dự đoán mức độ nghiêm trọng thì số mẫu nghiêm trọng chỉ chiếm 4.55%, và cũng tương tự đối với bài toán dự đoán triệu chứng từ thông tin cá nhân). , do đó cần điều chỉnh trọng số lớp để tránh mô hình thiên lệch. Các phương pháp sử dụng bao gồm:

- **class_weight = 'balanced'**: áp dụng cho các mô hình Random Forest, Logistic Regression, SGDClassifier (SVM), Decision Tree và Gradient Boosting.

Hiệu suất được đánh giá bằng các chỉ số:

- **Precision**: độ chính xác của dự đoán dương tính.
- **Recall**: khả năng phát hiện các mẫu dương thực sự.
- **F1-score**: trung bình điều hòa giữa precision và recall.
- **AUC-ROC**: diện tích dưới đường cong ROC, đo khả năng phân biệt giữa các lớp.

2.6 Hạn Chế của Dữ Liệu VAERS

Dữ liệu từ Hệ thống Báo cáo Biến cố Bất lợi Sau Tiêm chủng (VAERS) có một số hạn chế quan trọng cần lưu ý:

- **Báo cáo tự nguyện:** VAERS là hệ thống báo cáo thụ động, cho phép bất kỳ ai, bao gồm nhân viên y tế, bệnh nhân hoặc người nhà, đều có thể gửi báo cáo. Điều này dẫn đến sự biến động lớn về chất lượng và độ đầy đủ của dữ liệu [4].
- **Thiên lệch báo cáo:** Do tính chất tự nguyện, dữ liệu có thể bị thiên lệch, chẳng hạn như báo cáo quá mức về các phản ứng nghiêm trọng hoặc thiếu sót trong báo cáo về các phản ứng nhẹ [5].
- **Không xác minh nguyên nhân:** VAERS không xác định được mối quan hệ nhân quả giữa vaccine và các biến cố bất lợi. Dữ liệu chỉ phản ánh mối quan hệ thời gian, không phải nguyên nhân [8].
- **Thiếu thông tin nền tảng:** Dữ liệu không cung cấp thông tin về số lượng người đã tiêm vaccine, làm khó khăn trong việc đánh giá tỷ lệ thực sự của các biến cố bất lợi [10].

Do đó, khi sử dụng dữ liệu VAERS, cần thận trọng và kết hợp với các nguồn dữ liệu khác để có cái nhìn toàn diện và chính xác hơn về độ an toàn của vaccine.

3 Dự đoán mức độ nghiêm trọng

Trong nghiên cứu này, chúng tôi sử dụng dữ liệu từ VAERS, hệ thống báo cáo phản ứng sau tiêm của CDC và FDA [3]. Một số nghiên cứu cũng đã chỉ ra mối liên hệ giữa các triệu chứng và mức độ nghiêm trọng của phản ứng sau tiêm vaccine COVID-19 [6, 1]. Trong bài toán này, mục tiêu là phân loại xem một báo cáo sau tiêm COVID-19 có thuộc nhóm “nghiêm trọng” hay không (nhãn 1). Tập kiểm thử gồm 100000 bản ghi, trong đó có 4552 mẫu nghiêm trọng (chiếm 4,55%).

3.1 Thiết lập thực nghiệm

- Các mô hình đơn lẻ: Random Forest, Logistic Regression, SVM (SGDClassifier), MLP (2 tầng).
- Xử lý mất cân bằng: `class_weight='balanced'` trên tất cả mô hình hỗ trợ.
- Đánh giá: Precision, Recall, F1-score, AUC-ROC trên nhãn 1.

3.2 Kết quả

3.2.1 Kết quả dự đoán với dữ liệu gốc (chưa qua giảm chiều dữ liệu)

Bảng 1: Hiệu suất phân loại mức độ nghiêm trọng

Mô hình	Precision	Recall	F1-score	AUC-ROC
Random Forest	0.4317	0.4582	0.4446	0.8886
Logistic Regression	0.7241	0.4846	0.5806	0.8704
Support Vector Machine	0.1889	0.7791	0.3041	0.8415
MLP (2 tầng)	0.6221	0.5264	0.5702	0.8381

Nhận xét:

- **Logistic Regression:** Mô hình này có khả năng nhận diện chính xác các lớp phổ biến nhưng gặp khó khăn khi nhận diện các lớp ít phổ biến, dẫn đến mất cân bằng giữa precision và recall.
- **SVM:** SVM có khả năng phát hiện tốt các lớp ít phổ biến nhưng lại thiếu độ chính xác, tạo ra nhiều dự đoán sai cho lớp dương tính.
- **Random Forest và MLP:** Cả hai mô hình này cung cấp sự cân bằng tốt hơn giữa precision và recall. Random Forest thể hiện khả năng phân biệt vượt trội, trong khi MLP có recall và F1-score cao hơn nhưng với khả năng phân biệt kém hơn.

Kết luận: AUC-ROC là chỉ số quan trọng nhất trong bài toán phân loại mất cân bằng, và Random Forest với AUC-ROC cao nhất là mô hình tối ưu về khả năng phân biệt. MLP và Logistic Regression có những ưu điểm nhất định nhưng không bằng Random Forest về mặt phân loại tổng thể. SVM không phải là lựa chọn tối ưu do độ chính xác thấp.

3.2.2 Đặc trưng quan trọng cho dự đoán

Để xác định các triệu chứng quan trọng nhất ảnh hưởng đến dự đoán mức độ nghiêm trọng, dự án đã tổng hợp kết quả từ 4 mô hình (Random Forest, Logistic Regression, SVM, MLP). Mỗi mô hình cung cấp danh sách 20 triệu chứng quan trọng nhất, và từ đó phân tích lấy 20 triệu chứng xuất hiện nhiều nhất trong tất cả các mô hình làm đầu vào cho phần “Bài toán dự đoán triệu chứng”.

Các triệu chứng quan trọng nhất bao gồm: *Chills*, *Death*, *Headache*, *Pain*, *Fatigue*, *Injection site pain*, *Pyrexia*, *Dyspnoea*, *Pruritus*, *Pain in extremity*, *Myalgia*, và nhiều triệu chứng khác có tần suất cao. Những triệu chứng này sẽ được sử dụng trong phần tiếp theo của nghiên cứu.

3.2.3 So sánh hiệu quả giảm chiều với PCA và LDA

Bảng 2: AUC–ROC của các mô hình sau khi giảm chiều với PCA và LDA

Mô hình	PCA (5 thành phần)	LDA (2 thành phần)
Random Forest	0.8617	0.8008
Logistic Regression	0.7514	0.8577
Support Vector Machine	0.5032	0.8577
MLP (2 tầng)	0.9024	0.8577

Nhìn chung, MLP thể hiện ưu thế rõ rệt khi kết hợp với PCA, trong khi LDA nâng cao hiệu năng cho các mô hình tuyến tính (Logistic Regression, SVM) và cân bằng tốt hơn cho Random Forest. Việc lựa chọn phương pháp giảm chiều phụ thuộc vào bản chất mô hình và mục tiêu tối ưu (ví dụ: AUC–ROC cao nhất hay cân bằng tổng thể).

4 Dự đoán triệu chứng

Mục tiêu của bài toán là xây dựng một mô hình học máy có khả năng dự đoán các triệu chứng có thể xuất hiện sau khi tiêm vaccine mRNA, dựa trên thông tin nhân khẩu học và lâm sàng ban đầu. Bài toán được mô hình hóa dưới dạng phân loại đa nhãn (multi-label classification), trong đó mỗi báo cáo có thể gắn với nhiều triệu chứng cùng lúc. Trong phạm vi nghiên cứu, nhóm tập trung vào 20 triệu chứng quan trọng nhất—được lựa chọn dựa trên mức độ ảnh hưởng đến khả năng tiến triển thành tình trạng nghiêm trọng—with kỳ vọng mô hình có thể hỗ trợ cảnh báo sớm và phân tầng nguy cơ cho người tiêm. Hiệu quả phân loại của mô hình ensemble tùy chỉnh được đánh giá trên 20 triệu chứng này trong tập dữ liệu kiểm thử, sử dụng chỉ số AUC–ROC cho từng nhãn. Bảng 3 trình bày chi tiết kết quả thu được.

Bảng 3: Giá trị AUC–ROC của mô hình Ensemble–Custom theo từng triệu chứng

STT	Triệu chứng	AUC–ROC
1	Chills	0.5575
2	Death	0.7000
3	Headache	0.5485
4	Pain	0.5464
5	Fatigue	0.5260
6	Injection site pain	0.5605
7	Pyrexia	0.5606
8	Dyspnoea	0.4946
9	Pruritus	0.5574
10	Pain in extremity	0.5274
11	Myalgia	0.5698
12	Injection site warmth	0.6963
13	Dizziness	0.5249
14	Nausea	0.5142
15	Poor quality product administered	0.8256
16	Injection site erythema	0.7022
17	Injection site swelling	0.6736
18	Injection site pruritus	0.7084
19	Unresponsive to stimuli	0.5666
20	Erythema	0.6060

Phân tích kết quả Kết quả thu được cho thấy mô hình ensemble–custom đạt hiệu quả phân loại tương đối tốt đối với một số triệu chứng hiếm gặp hoặc có đặc điểm nhận diện rõ ràng. Cụ thể, các triệu chứng liên quan đến phản ứng tại vị trí tiêm như *Injection site pruritus*, *Injection site erythema* và *Injection site warmth* đều đạt AUC trên 0.69, trong khi triệu chứng *Poor quality product administered* đạt AUC–ROC cao nhất (0.8256), cho thấy khả năng phân biệt hai lớp rất tốt trong các trường hợp có tín hiệu đặc thù.

Ngược lại, với những triệu chứng phổ biến và mang tính chất không đặc hiệu như *Headache*, *Fatigue*, *Pain* hoặc *Chills*, mô hình cho thấy độ phân biệt hạn chế, với AUC chỉ dao động trong khoảng 0.52–0.56. Điều này có thể bắt nguồn từ sự phân bố mất cân bằng nghiêm trọng giữa các lớp và độ chồng lấn cao giữa các biểu hiện ở cả hai nhãn (có và không có triệu chứng).

Tổng thể, mặc dù một số nhãn đạt hiệu năng khá ấn tượng, AUC–ROC trung bình trên toàn bộ tập triệu chứng vẫn cho thấy nhiều dư địa cải thiện. Trong các hướng nghiên cứu tiếp theo, cần xem xét các chiến lược như hiệu chỉnh ngưỡng phân loại theo từng triệu chứng, tăng cường trọng số mẫu dương, hoặc áp dụng kỹ thuật học sâu chuyên biệt cho bài toán phân loại đa nhãn bất cân bằng.

5 Kết luận

Dự án này đã xây dựng thành công một hệ thống phân tích và dự đoán phản ứng bất lợi sau tiêm vaccine mRNA COVID-19 từ dữ liệu VAERS, với các đóng góp chính sau:

5.1 Tổng kết kết quả

- Đề xuất pipeline xử lý dữ liệu VAERS hoàn chỉnh, từ khâu làm sạch, tích hợp đến chuẩn bị dữ liệu cho các mô hình học máy
- Xây dựng thành công mô hình dự đoán mức độ nghiêm trọng với AUC-ROC đạt 0.8886 (Random Forest) và xác định được các triệu chứng quan trọng nhất (Chills, Death, Headache,...)
- Phát triển mô hình đa nhiệm dự đoán 20 triệu chứng chính từ thông tin cá nhân, trong đó một số triệu chứng đạt AUC-ROC > 0.7 (Injection site pruritus: 0.7084, Poor quality product administered: 0.8256)
- Đánh giá toàn diện hiệu quả của các kỹ thuật giảm chiều dữ liệu (PCA, LDA) với nhiều mô hình khác nhau

5.2 Hạn chế và hướng phát triển

- **Hạn chế hiện tại:**
 - Hiệu năng dự đoán chưa đồng đều giữa các triệu chứng, đặc biệt với nhóm triệu chứng phổ biến (AUC-ROC 0.5-0.55)
 - Chưa khai thác triệt để các kỹ thuật xử lý mất cân bằng dữ liệu tiên tiến
 - Giới hạn về mặt tính toán khi xử lý toàn bộ tập dữ liệu VAERS
- **Hướng phát triển tương lai:**
 - Áp dụng các kiến trúc học sâu (Deep Learning) để xử lý bài toán đa nhiệm phức tạp
 - Tích hợp thêm nguồn dữ liệu lâm sàng khác (EHR, genomic data) để nâng cao độ chính xác
 - Phát triển hệ thống cảnh báo sớm theo thời gian thực cho các phản ứng nghiêm trọng
 - Nghiên cứu cơ chế giải thích mô hình (Explainable AI) để hỗ trợ quyết định lâm sàng

5.3 Ý nghĩa thực tiễn

Kết quả dự án không chỉ có giá trị học thuật mà còn mang ý nghĩa ứng dụng quan trọng trong:

- Hỗ trợ hệ thống giám sát an toàn vaccine quốc gia
- Cá nhân hóa tư vấn tiêm chủng dựa trên nguy cơ cá nhân
- Chuẩn bị cho các chiến dịch tiêm chủng quy mô lớn trong tương lai

Với những kết quả đạt được cùng các hướng phát triển đã đề xuất, dự án mở ra nhiều triển vọng trong việc ứng dụng trí tuệ nhân tạo để nâng cao chất lượng công tác tiêm chủng và giám sát sức khỏe cộng đồng.

Mã nguồn, tài liệu chi tiết có sẵn tại: <https://github.com/HungTang-120105/ml4vax-ae>.

Tài liệu

- [1] L. Abu-Raddad et al. Effectiveness and safety of mrna covid-19 vaccines in the middle east. *The Lancet Regional Health - Eastern Mediterranean*, 5:100084, 2022.
- [2] Alice Adams and Charles Brown. Covid-19 vaccine adverse events and demographic factors: A data analysis. *Vaccine Insights*, 15(3):212–219, 2021.
- [3] CDC and FDA. Vaccine adverse event reporting system (vaers). <https://vaers.hhs.gov>, 2021.
- [4] Centers for Disease Control and Prevention. Guide to interpreting vaers data. <https://vaers.hhs.gov/data/dataguide.html>. Truy cập ngày 09/05/2025.
- [5] Centers for Disease Control and Prevention. Vaers - data. <https://vaers.hhs.gov/data.html>. Truy cập ngày 09/05/2025.
- [6] C. Chung et al. Assessment of adverse reactions following mrna covid-19 vaccination in a large cohort. *JAMA Network Open*, 4(5):e212789, 2021.
- [7] John Doe and Alice Adams. Machine learning techniques for predicting vaccine-related adverse events. *Journal of Machine Learning in Medicine*, 8(1):87–98, 2020.
- [8] Johns Hopkins Bloomberg School of Public Health. What vaers is (and isn't). <https://publichealth.jhu.edu/2022/what-vaers-is-and-isnt>. Truy cập ngày 09/05/2025.
- [9] Mary Johnson and Robert Lee. The role of vaers in monitoring vaccine safety: A comprehensive review. *International Journal of Public Health*, 12(4):456–470, 2020.
- [10] Global Vaccine Data Network. Vaers misuse: The zombie practice that keeps coming back. https://www.globalvaccinedatanetwork.org/news/VAERS_misuse_The_zombie_practice_that_keeps_coming_back. Truy cập ngày 09/05/2025.
- [11] John Smith and Jane Doe. Vaccine safety monitoring and adverse events in the context of covid-19. *Journal of Vaccine Safety*, 5(2):123–130, 2021.