

Phát triển mô hình dự báo tử vong ở bệnh nhân viêm phổi cộng đồng nhập viện dựa trên học máy

Development of a machine learning-based model for predicting mortality in hospitalized patients with community-acquired pneumonia

Đỗ Thanh Hòa¹, Nguyễn Hải Ghi² và Lê Đức Duẩn^{2,*}

Bệnh viện Trung ương Quân đội 108

Tóm tắt

Mục tiêu: Phát triển mô hình dự báo tử vong 30 ngày ở bệnh nhân viêm phổi cộng đồng nhập viện bằng thuật toán rừng ngẫu nhiên (RF) và so sánh khả năng dự báo với thang điểm CURB-65. **Đối tượng và phương pháp:** Nghiên cứu mô tả, hồi cứu tại Khoa Cấp cứu - Bệnh viện TWQĐ 108 từ năm 2021 - 2022 ở các bệnh nhân người lớn nhập viện vì viêm phổi cộng đồng. Dữ liệu ban đầu gồm 21 biến lâm sàng và cận lâm sàng thu thập trong 24 giờ đầu nhập viện. Các bệnh nhân chia ngẫu nhiên thành tập huấn luyện (70%) và tập kiểm tra (30%). Hồi quy LASSO được áp dụng trên tập huấn luyện để chọn biến dự báo độc lập, sau đó dùng để xây dựng mô hình RF. Hiệu suất mô hình đánh giá trên tập kiểm tra qua các chỉ số: AUC, Se, Sp, PPV, NPV và độ chính xác. **Kết quả:** 350 bệnh nhân viêm phổi cộng đồng, tuổi trung bình $71,7 \pm 16,0$ với tỷ lệ tử vong 30 ngày là 6,3%. Hồi quy LASSO đã chọn ra 5 biến dự báo độc lập (rối loạn ý thức, thâm nhiễm đa thùy, nồng độ urê, albumin và lactate máu). Mô hình RF được xây dựng từ 5 biến này có hiệu suất vượt trội hơn với thang điểm CURB-65 trên tập kiểm tra, với $AUC_{RF} = 0,969$ và $AUC_{CURB-65} = 0,703$ ($p=0,004$). Mức độ quan trọng của các biến trong mô hình RF theo thứ tự giảm dần là nồng độ lactate, albumin, urê, rối loạn ý thức và thâm nhiễm đa thùy phổi. **Kết luận:** Mô hình rừng ngẫu nhiên sử dụng 5 biến lâm sàng và cận lâm sàng cho thấy hiệu suất dự báo tử vong 30 ngày ở bệnh nhân CAP nhập viện tốt hơn đáng kể so với thang điểm CURB-65.

Keywords: Viêm phổi cộng đồng, dự báo tử vong, rừng ngẫu nhiên.

Summary

Objective: To develop a 30-day mortality prediction model for hospitalized patients with community-acquired pneumonia (CAP) using the Random Forest (RF) algorithm and compare its predictive ability with the CURB-65 score. **Subject and method:** A descriptive, retrospective study conducted at the Emergency Department of 108 Military Central Hospital from 2021 to 2022, involving adult patients hospitalized due to CAP. Initial data included 21 clinical and paraclinical variables collected within the first 24 hours of admission. Patients were randomly divided into a training set (70%) and a testing set (30%). LASSO regression was applied to the training set to select independent predictors, which were then used to build the RF model. Model performance was evaluated on the testing set using metrics: AUC, Se, Sp, PPV, NPV, and Accuracy. **Result:** A total of 350 CAP patients were included, with a mean age of 71.7 ± 16.0 years and a 30-day mortality rate of 6.3%. LASSO regression

Ngày nhận bài: 20/10/2025, ngày chấp nhận đăng: 26/11/2025

* Tác giả liên hệ: Duanhocvienquany@gmail.com - Bệnh viện Trung ương Quân đội 108

selected 5 independent predictors: altered mental status, multilobar infiltrates, and blood urea nitrogen, albumin and lactate levels. The RF model constructed from these 5 variables demonstrated superior performance compared to the CURB-65 score on the testing set, with $AUC_{RF}=0.969$ and $AUC_{CURB-65}=0.703$ ($p=0.004$). The feature importance in the RF model in descending order was: lactate level, albumin level, blood urea nitrogen, altered mental status, and multilobar infiltrates. *Conclusion:* The Random Forest model using 5 clinical and paraclinical variables showed significantly better performance in predicting 30-day mortality in hospitalized CAP patients compared to the CURB-65 score.

Keywords: Community-Acquired Pneumonia, Mortality prediction, Random Forest.

I. ĐẶT VẤN ĐỀ

Viêm phổi cộng đồng (Community-Acquired Pneumonia - CAP) là nguyên nhân tử vong hàng đầu do nhiễm khuẩn trên toàn cầu với tỷ lệ tử vong 30 ngày ở bệnh nhân CAP khoảng 7,5 - 9,8%^{1,2}. Việc phân tầng nguy cơ tử vong sớm là rất quan trọng để quyết định nơi điều trị và chiến lược quản lý bệnh nhân.

Hiện nay, thang điểm CURB-65 (lú lẫn, urê máu, tần số hô hấp, huyết áp và tuổi ≥ 65) được sử dụng phổ biến nhờ tính đơn giản và dễ áp dụng tại phòng cấp cứu. Tuy nhiên, các phân tích gần đây cho thấy hiệu suất dự báo tử vong của thang điểm này 0,67 - 0,75^{3,4} và còn tồn tại nhiều hạn chế như bỏ sót nhiều yếu tố tiên lượng quan trọng khác như tình trạng viêm, suy đa tạng hay các chỉ số sinh học mới nổi (ví dụ: lactate, albumin, D-dimer...). Điều này dẫn đến dẫn đến độ nhạy và độ đặc hiệu còn chưa tối ưu.

Những năm gần đây, Machine Learning đã cho thấy tiềm năng vượt trội trong phân tích dữ liệu lớn và phức tạp, từ đó xây dựng các mô hình dự báo chính xác hơn. Ví dụ mô hình rừng ngẫu nhiên (Random Forest) dự báo tử vong nội viện ở bệnh nhân viêm phổi nặng và cho thấy hiệu suất tốt hơn với các thang điểm truyền thống như APACHE II, SOFA hay CURB-65⁵. Đồng thời mô hình này có thể tận dụng nhiều biến số khác nhau cùng lúc để xác định mối quan hệ phi tuyến tính phức tạp giúp tiên lượng chính xác hơn⁶. Xuất phát từ thực tế đó, nghiên cứu này được thực hiện với mục tiêu: Phát triển mô hình dự báo tử vong 30 ngày ở bệnh nhân viêm phổi cộng đồng nhập viện bằng thuật toán Rừng ngẫu nhiên (RF) nhằm đưa ra một công cụ hỗ trợ quyết định lâm sàng hiệu quả hơn trong việc phân tầng nguy cơ bệnh nhân CAP.

II. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP

2.1. Đối tượng

Nghiên cứu người lớn ≥ 18 tuổi

Tiêu chuẩn lựa chọn: Bệnh nhân được chẩn đoán xác định là viêm phổi cộng đồng (CAP) dựa theo hướng dẫn chẩn đoán và điều trị viêm phổi cộng đồng của BYT năm 2020⁷. Bệnh nhân được nhập viện điều trị tại khoa lâm sàng (hô hấp, truyền nhiễm, khoa hồi sức tích cực). Có đầy đủ các dữ liệu lâm sàng và cận lâm sàng cần thiết trong vòng 24 giờ đầu nhập viện.

Tiêu chuẩn loại trừ: Được chẩn đoán ban đầu là viêm phổi bệnh viện hoặc viêm phổi thở máy. Bệnh nhân tử vong trong vòng 24 giờ đầu. Có bệnh lý gây tử vong không liên quan trực tiếp đến viêm phổi cộng đồng (ví dụ như: chấn thương sọ não nặng, nhồi máu cơ tim cấp...). Không đầy đủ dữ liệu để xây dựng mô hình.

2.2. Phương pháp

Thiết kế nghiên cứu: Nghiên cứu mô tả, hồi cứu tại khoa cấp cứu - bệnh viện Trung ương Quân đội 108 từ năm 2021 đến năm 2022.

Các dữ liệu lâm sàng và cận lâm sàng: Thu thập các dữ liệu tại thời điểm chẩn đoán tại khoa cấp cứu gồm thông số lâm sàng (tuổi, giới, mạch, huyết áp trung bình, tần số thở, rối loạn ý thức, thân nhiệt), xét nghiệm máu (nồng độ glucose, urê, creatinin, natri, albumin, protein, lactate, số lượng bạch cầu, neutrophil, lymphocyte, tiểu cầu, RDW) và X-quang phổi (thâm nhiễm đa thùy, tràn dịch màng phổi).

Định nghĩa kết cục tử vong 30 ngày: Là các bệnh nhân tử vong tại bệnh viện hoặc bệnh nhân đã ra viện

được xác nhận là tử vong bằng giấy báo tử/gọi điện thoại. Thời gian tử vong được ghi nhận từ khi bệnh nhân nhập khoa cấp cứu đến ngày tử vong.

Phát triển mô hình rừng ngẫu nhiên (RF)

Chúng tôi phân chia ngẫu nhiên dữ liệu thành 2 phần theo tỷ lệ 70:30. Với 70% để huấn luyện mô hình (training set) và 30% để kiểm tra mô hình (testing set).

Để xác định các yếu tố dự báo liên quan đến tử vong 30 ngày, chúng tôi áp dụng hồi quy LASSO (Least Absolute Shrinkage and Selection Operator) bằng gói *glmnet* trong R. Mô hình gồm 21 biến lâm sàng và cận lâm sàng. Tham số phạt tối ưu λ được xác định thông qua kiểm chứng chéo 10 lần (10 fold cross-validation) với hiệu suất đánh giá bằng AUC. Phương pháp LASSO giúp giảm hiện tượng quá khớp (overfitting) và tạo ra một tập hợp các biến dự báo cô đọng hơn bằng cách đưa hệ số của các biến không quan trọng về 0. Các biến có hệ số khác 0 được chọn làm tập hợp biến dự báo đầu vào để xây dựng mô hình rừng ngẫu nhiên.

Mô hình rừng ngẫu nhiên (RF) được xây dựng nhằm dự báo tử vong 30 ngày. Thuật toán RF sử dụng kỹ thuật bagging (Bootstrap Aggregating) để tạo ra một tập hợp cây quyết định được huấn luyện từ các mẫu bootstrap và lựa chọn ngẫu nhiên các biến đầu vào để giảm hiện tượng quá khớp và tăng tính khái quát của mô hình. Quá trình huấn luyện và

tối ưu hóa siêu tham số (hyperparameter) được thực hiện bằng kiểm chứng chéo 5 lần (5 fold cross-validation) sử dụng gói *caret* trong R. Hiệu suất của mô hình được tối ưu dựa trên AUC. Mức độ quan trọng của các biến dự báo được đánh giá dựa trên chỉ số Gini.

Phân tích thống kê

Các biến liên tục được trình bày dưới dạng trung bình (độ lệch chuẩn) hoặc trung vị (khoảng tứ phân vị) tùy theo phân bố của số liệu. Các biến phân loại trình bày dưới dạng số lượng (tỷ lệ phần trăm). So sánh giữa các nhóm thực hiện bằng Student t test hoặc Mann-Whitney U test đối với biến liên tục và kiểm định χ^2 hoặc Fisher exact test đối với biến phân loại tùy theo trường hợp cụ thể.

Đánh giá hiệu suất của mô hình RF thông qua các metric: AUC, độ chính xác (accuracy), độ nhạy (Se), độ đặc hiệu (Sp), giá trị tiên đoán dương (PPV), giá trị tiên đoán âm (NPV). So sánh hiệu suất (AUC) với thang điểm CURB-65 trên cùng testing set bằng Delong test.

Phân tích số liệu bằng phần mềm R 4.4.1.

III. KẾT QUẢ

Nghiên cứu tuyển chọn được 350 bệnh nhân với độ tuổi trung bình $71,7 \pm 16,0$, nam giới chiếm 228 người (65,1%), tử vong trong 30 ngày là 22 bệnh nhân (6,3%).

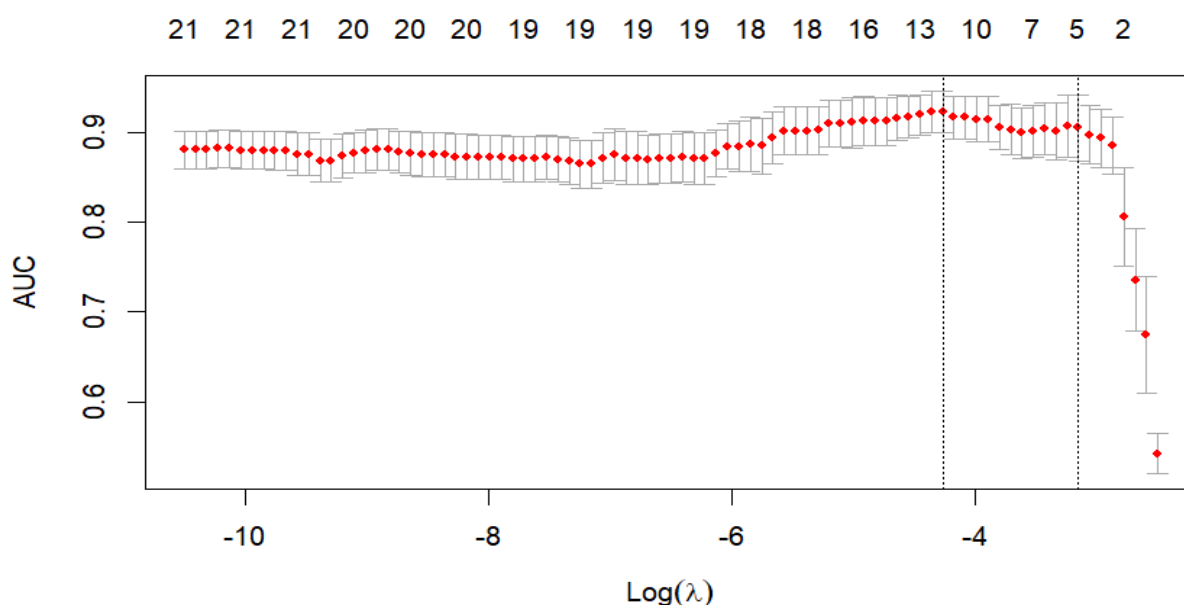
Bảng 1. Một số đặc điểm lâm sàng và cận lâm sàng theo kết cục

Đặc điểm	Nhóm tử vong (n = 22)	Nhóm sống sót (n = 328)	P
Tuổi (năm)	79,8 $\pm$ 9,5	72,2 $\pm$ 15,8	0,020
Giới nam (%)	18 (81,8%)	210 (64,0)	0,108
Rối loạn ý thức, n (%)	16 (72,7%)	61 (18,6%)	< 0,001
Tần số thở (lần/phút)	21,0 (10,8)	18,0 (5,0)	0,077
Tần số tim (lần/phút)	102,0 (46,0)	92,0 (25,0)	0,427
Huyết áp trung bình (mmHg)	84,0 (16,6)	95,0 (17,4)	0,013
Thân nhiệt (độ C)	37,8 (3,0)	37,0 (1,1)	0,015
Số lượng BC (G/L)	14,7 (5,1)	11,1 (6,8)	0,020
Số lượng N (G/L)	16,0 (8,5)	9,3 (8,3)	< 0,001
Số lượng L (G/L)	0,7 (0,8)	1,1 (1,0)	0,018

Đặc điểm	Nhóm tử vong (n = 22)	Nhóm sống sót (n = 328)	P
Số lượng TC (G/L)	211 (162)	257 (152)	0,052
RDW	16,9 (3,6)	13,9 (2,1)	0,001
Glucose (mmol/l)	7,6 (4,9)	7,1 (3,9)	0,751
Urê (mmol/l)	11,5 (10,8)	6,5 (4,7)	< 0,001
Creatinin (μ mol/l)	96,5 (58,3)	74,0 (42,5)	0,020
Natri (mmol/l)	133 (10,7)	134 (6,0)	0,979
Protein (mg/dl)	55,6 (14,0)	60,0 (8,0)	0,152
Albumin (mg/dl)	26,4 (5,5)	29,0 (6,3)	0,051
Lactate (mmol/l)	2,8 (2,1)	1,4 (1,0)	<0,001
Thâm nhiễm đa thùy, n (%)	18 (81,8%)	134 (40,9%)	<0,001
Tràn dịch màng phổi, n (%)	8 (13,2%)	92 (28,0%)	0,403

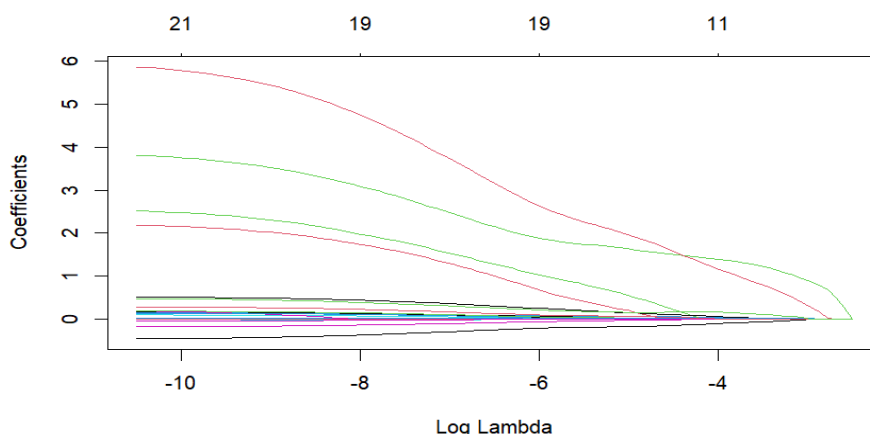
Nhận xét: Có sự khác biệt có ý nghĩa thống kê về độ tuổi ($p=0,020$), rối loạn ý thức ($p<0,001$), huyết áp trung bình ($p=0,013$), thân nhiệt ($p=0,015$), số lượng bạch cầu ($p=0,020$), số lượng neutrophil ($p<0,001$), số lượng lymphocyte ($P=0,018$), RDW ($p=0,001$), nồng độ urê ($p<0,001$), creatinin ($p=0,020$), lactate ($p<0,001$) và thâm nhiễm đa thùy phổi ($p<0,001$) giữa nhóm tử vong và nhóm sống sót.

Các bệnh nhân được chia ngẫu nhiên thành 2 phần: tập huấn luyện mô hình (training set) có 245 bệnh nhân (70%) và tập kiểm tra mô hình (testing set) có 105 bệnh nhân (30%). Tập huấn luyện mô hình dùng để lựa chọn biến dự báo độc lập bằng phương pháp LASSO và xây dựng mô hình rừng ngẫu nhiên dưới đây.



Hình 1. Biểu đồ kiểm chứng chéo minh họa sự thay đổi của AUC theo log λ

Nhận xét: Hình 1 cho thấy giá trị log λ cho AUC tối đa là $\approx -4,2$ tương ứng với khoảng 13 biến có hệ số khác 0. Chúng tôi chọn giá trị log λ lớn nhất mà hiệu suất AUC vẫn nằm trong một độ lệch chuẩn (1 SE) của giá trị AUC tối đa (log $\lambda \approx -3$). Tại log λ này, mô hình LASSO chọn ra 5 biến dự báo độc lập có hệ số khác 0.



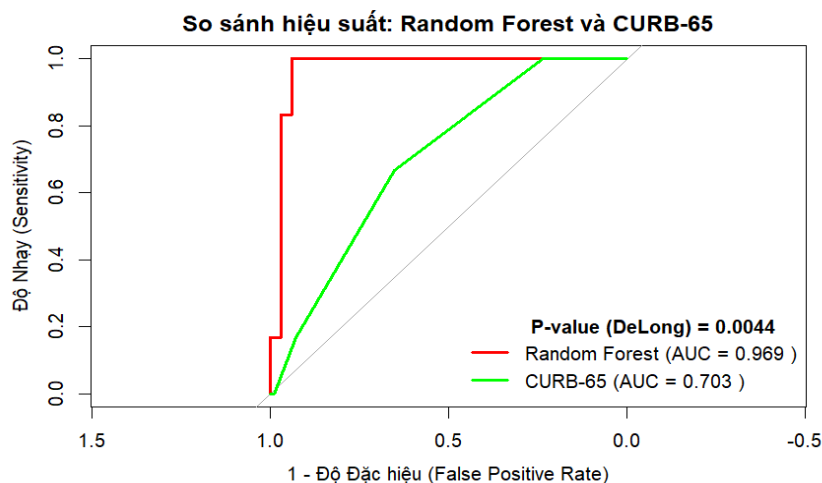
Hình 2. Biểu đồ đường đi hệ số

Nhận xét: Hình 2 cho thấy một số biến (biểu thị bằng đường màu đỏ và xanh lá cây) duy trì hệ số lớn và khác 0 ngay cả khi mức phạt cao (log λ lớn). Cho thấy các biến này là yếu tố dự báo mạnh. Một số đường có hệ số hồi quy LASSO > 0 thể hiện tăng giá trị của biến này làm tăng tử vong, ngược lại một số đường có hệ số hồi quy LASSO < 0 thể hiện giảm giá trị các biến này làm tăng tử vong.

Bảng 2. Các biến dự báo độc lập được chọn lọc bằng Hồi quy LASSO

STT	Tên biến	Hệ số LASSO
1	Rối loạn ý thức	+ 1,013
2	Thâm nhiễm đa thùy phổi	+ 0,479
3	Nồng độ urê máu	+ 0,017
4	Nồng độ Albumin máu	- 0,028
5	Nồng độ lactate máu	+ 0,053

Nhận xét: Có 5 biến được lựa chọn dựa trên mô hình LASSO tại vị trí $\log \lambda \approx -3$. Trong đó các biến rối loạn ý thức, thâm nhiễm đa thùy, nồng độ urê cao, nồng độ lactate cao liên quan đến tăng nguy cơ tử vong; nồng độ albumin máu giảm liên quan đến tăng nguy cơ tử vong. Năm biến này được đưa vào xây dựng mô hình RF.

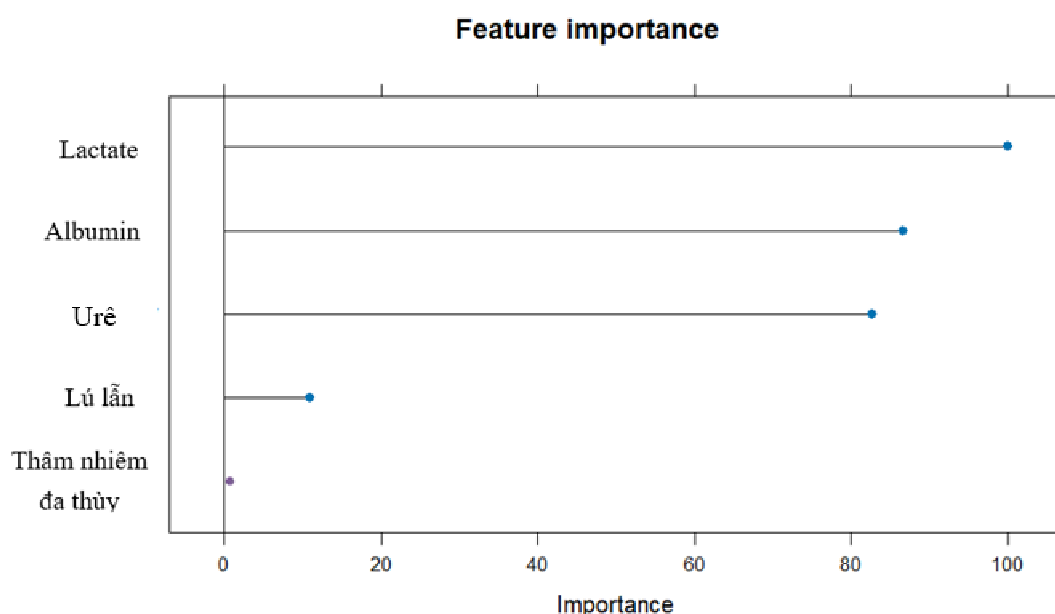


Hình 3. Diện tích dưới đường cong so sánh hiệu suất mô hình RF với thang điểm CURB-65

Bảng 3. So sánh hiệu suất mô hình RF và thang điểm CURB-65

Mô hình	Accuracy	AUC	Se	Sp	PPV	NPV	p
RF	0,971	0,969	0,979	0,80	0,25	0,99	0.004
CURB-65	0,654	0,703	0,667	0,653	0,105	0,969	

Nhận xét: Sau khi xây dựng mô hình RF từ 5 biến độc lập, kiểm tra mô hình trên testing set (cutoff tại xác suất phân định 0,5) và so sánh các metric với thang điểm CURB-65 (tại cutoff là 2) cho thấy mô hình RF có hiệu suất tốt hơn (AUC = 0,979) so với thang điểm CURB-65 với (AUC = 0,703) có ý nghĩa thống kê ($P = 0,004$).

**Hình 4.** Phân bố tầm quan trọng của các biến dự báo tử vong

Nhận xét: Trong 5 biến đưa vào mô hình RF, mức độ quan trọng của các biến trong dự báo tử vong theo thứ tự giảm dần là nồng độ lactate -> nồng độ albumin -> nồng độ urê -> rối loạn ý thức -> thâm nhiễm đa thùy phổi.

IV. BÀN LUẬN

Nghiên cứu của chúng tôi đã thành công trong việc xây dựng mô hình dự báo tử vong 30 ngày ở bệnh CAP nhập viện bằng thuật toán RF. Mô hình RF cho thấy khả năng phân biệt vượt trội, đạt AUC 0,969 và độ chính xác 0,971, cao hơn đáng kể so với thang điểm CURB-65 đạt mức khá với AUC 0,703 ($P=0,004$). Các biến số tiên lượng quan trọng nhất được chọn lọc bằng hồi quy LASSO là lactate máu, albumin máu, urê máu, rối loạn ý thức và thâm nhiễm đa thùy phổi.

Hiệu suất dự báo tốt của mô hình RF trong nghiên cứu này phù hợp với xu hướng y văn gần đây, vốn ủng hộ việc áp dụng các thuật toán học máy để cải thiện khả năng tiên lượng CAP so với các thang điểm kinh điển. Nghiên cứu của tác giả Zhao và cộng sự (2025) về mô hình học máy dự báo tử vong nội viện do viêm phổi nặng cho thấy mô hình học máy đạt AUC 0,8779 (95% CI, 0,738 - 0,974), tốt hơn so với các thang điểm truyền thống như APACHE II, SOFA và CURB-65⁵.

Một nghiên cứu sử dụng các mô hình học máy lâm sàng dự báo tử vong ở bệnh nhân viêm phổi cộng đồng ở người cao tuổi cho thấy khả năng dự báo tử vong 180 ngày, 90 ngày và 30 ngày với AUC lần lượt là 0,829; 0,832 và 0,904. Mô hình này tốt hơn có ý nghĩa thống kê so với thang điểm CURB-65 với AUC chỉ đạt 0,573⁸.

Nghiên cứu của tác giả Yuan và cộng sự (2022) sử dụng rừng ngẫu nhiên và XGBoost tích hợp hệ thống học mở rộng để tiên lượng tử vong 28 ngày bệnh nhân viêm phổi cộng đồng cho thấy mô hình rừng ngẫu nhiên trên tập huấn luyện đạt AUC 0,979 (95% CI: 0,963 - 0,996) và tập kiểm tra đạt 0,962 (95% CI: 0,936 - 0,988)⁹. Các kết quả này hoàn toàn phù hợp với báo cáo của chúng tôi, củng cố nhận định về tiềm năng của học máy. Điều này cho thấy thuật toán RF đã tối ưu hóa được mối quan hệ phi tuyến tính và tương tác phức tạp giữa các yếu tố liên lượng và thang điểm CURB-65 không thể nắm bắt được.

Năm biến số được xác nhận là quan trọng trong nghiên cứu này đều liên quan rõ ràng đến tiên lượng xấu ở bệnh nhân CAP. Lactate máu cao là biến quan trọng hàng đầu. Lactate cao là chỉ điểm của thiếu oxy mô và là một trong chỉ điểm tiên lượng xấu trong viêm phổi¹⁰. Albumin đóng vai trò quan trọng giúp điều hòa cân bằng dịch, duy trì áp lực keo và hỗ trợ chức năng miễn dịch và giảm albumin là chỉ điểm của mức độ nặng viêm¹¹. Nồng độ albumin thấp có liên quan độc lập đến suy tạng và tăng tỷ lệ tử vong¹². Urê máu cao và lú lẫn chính là 2 thành tố quan trọng trong thang điểm CURB-65, phản ánh tình trạng suy giảm chức năng thận hoặc mất nước và tổn thương hệ thống thần kinh do nhiễm khuẩn hệ thống hoặc suy hô hấp. Tổn thương phổi đa thùy liên quan trực tiếp đến mức độ viêm và phá hủy cấu trúc phổi, liên quan trực tiếp đến nguy cơ suy hô hấp cấp và tử vong cao hơn. Phân tích tổng hợp của tác giả Xie và cộng sự (2024) cũng chỉ ra nhiều yếu tố tăng nguy cơ tử vong ở bệnh nhân viêm phổi nặng như sốc nhiễm khuẩn, tổn thương thận cấp, nhiễm toan chuyển hóa, tăng urê máu lúc nhập viện, tăng CRP, giảm albumin, tổn thương 2 bên hoặc đa thùy, nhiễm khuẩn gram âm¹³.

Mô hình dự báo rừng ngẫu nhiên này có tiềm năng ứng dụng cao trong lâm sàng. Tất cả 5 biến số dự báo trên đều là các thông số lâm sàng, cận lâm sàng cơ bản, dễ dàng thu thập trong 24 giờ đầu nhập viện. Tuy nhiên, nghiên cứu này cũng có một số hạn chế quan trọng như dữ liệu trích xuất hồi cứu và đơn trung tâm. Biến cố tử vong thấp do vậy cần cỡ mẫu lớn hơn để đảm bảo khả năng tổng quát

hóa. Mô hình này còn thiếu ngoại kiểm trên quần thể độc lập khác do đó cần thêm những nghiên cứu đa trung tâm để có thể áp dụng mô hình này vào thực tế lâm sàng.

V. KẾT LUẬN

Mô hình Rừng ngẫu nhiên dự báo tử vong 30 ngày ở bệnh nhân viêm phổi cộng đồng nhập viện đã cho thấy hiệu suất vượt trội (AUC = 0,969) so với thang điểm CURB-65 (AUC = 0,703, P=0,004). Các biến dự báo quan trọng nhất gồm: tăng lactate máu, giảm albumin máu, tăng urê máu, lú lẫn và thâm nhiễm đa thùy phổi. Mô hình RF là một công cụ hỗ trợ lâm sàng chính xác hơn đáng kể, giúp tiên lượng tốt bệnh nhân CAP nguy cơ cao.

TÀI LIỆU THAM KHẢO

1. Reichel F, Tesch F, Berger S, et al. Burden and risk factors for 30-day readmission and mortality after CAP hospitalization: A population-based cohort study. *Eur J Intern Med.* Aug 2025;138:60-68. doi:10.1016/j.ejim.2025.04.025
2. Barlas RS, Clark AB, Loke YK, et al. Comparison of the prognostic performance of the CURB-65 and a modified version of the pneumonia severity index designed to identify high-risk patients using the International Community-Acquired Pneumonia Collaboration Cohort. *Respir Med.* Aug-Sep 2022;200:106884. doi:10.1016/j.rmed.2022.106884
3. Hincapié C, Ascuntar J, León A, Jaimes F. Community-acquired pneumonia: comparison of three mortality prediction scores in the emergency department. *Colomb Med (Cali).* Oct-Dec 2021;52(4):e2044287. doi:10.25100/cm.v52i4.4287
4. Tokioka F, Okamoto H, Yamazaki A, Itou A, Ishida T. The prognostic performance of qSOFA for community-acquired pneumonia. *J Intensive Care.* 2018;6:46. doi:10.1186/s40560-018-0307-7
5. Zhao W, Li X, Gao L, et al. Machine learning-based model for predicting all-cause mortality in severe pneumonia. *BMJ Open Respir Res.* Mar 22 2025;12(1)doi:10.1136/bmjresp-2023-001983
6. Langsetmo L, Schousboe JT, Taylor BC, et al. Advantages and Disadvantages of Random Forest

- Models for Prediction of Hip Fracture Risk Versus Mortality Risk in the Oldest Old. *JBMR Plus*. Aug 2023;7(8):e10757. doi:10.1002/jbm4.10757
7. Lương Ngọc Khuê, Ngô Quý Châu, Nguyễn Việt Nhung và cộng sự (2020) *Hướng dẫn chẩn đoán và điều trị viêm phổi mắc phải tại cộng đồng ở người lớn*. Bộ Y tế.
 8. Weng B, Huang L, Jiao W et al (2025) *Development and validation of a clinical-friendly model for predicting 180-day mortality in the older with community-acquired pneumonia*. *BMC Geriatr* 25(1): 271. doi:10.1186/s12877-025-05834-8
 9. Yuan J, Liu X, Wang WF, Zhang JJ (2022) *A Broad Learning System to Predict the 28-Day Mortality of Patients Hospitalized with Community-Acquired Pneumonia: A Case-Control Study*. *Comput Math Methods Med* :7003272. doi:10.1155/2022/7003272
 10. Demirel B (2018) *Lactate levels and pneumonia severity index are good predictors of in-hospital mortality in pneumonia*. *Clin Respir J* 12(3): 991-995. doi:10.1111/crj.12616
 11. Fanali G, di Masi A, Trezza V, Marino M, Fasano M, Ascenzi P (2012) *Human serum albumin: From bench to bedside*. *Mol Aspects Med* 33(3): 209-290. doi:10.1016/j.mam.2011.12.002
 12. Unal Cetin E, Kurtkulagi O, Kamis F et al (2025) *Advancing ICU mortality prediction in community-acquired pneumonia: Combining fibrinogen-to-albumin ratio, CT severity score, PSI, and CURB-65*. *Biomol Biomed* 25(9): 2083-2091. doi:10.17305/bb.2025.12127
 13. Xie K, Guan S, Kong X et al (2024) *Predictors of mortality in severe pneumonia patients: A systematic review and meta-analysis*. *Syst Rev* 13(1): 210. doi:10.1186/s13643-024-02621-1