

PHƯƠNG PHÁP DỰ BÁO CHUỖI THỜI GIAN TRÊN CHUỖI DỮ LIỆU HẠN CHẾ: TRƯỜNG HỢP ĐIỂN HÌNH TẠI MỘT PHÒNG KHÁM CHUYÊN KHOA VIỆT NAM

Đặng Hữu Phúc¹

Võ Nguyễn Kim Thảo^{1,2}

Nguyễn Văn Bắc³

¹Đại học Kinh tế Thành phố Hồ Chí Minh (UEH), Việt Nam

²Bệnh viện Đại học Y Dược Thành phố Hồ Chí Minh (UMC), Việt Nam

³Bệnh viện Quân Y 175, Việt Nam

*Tác giả liên hệ: Đặng Hữu Phúc – Email: phucdh@ueh.edu.vn

(Ngày nhận bài: 10/3/2026, ngày nhận bài chỉnh sửa: 18/4/2026, ngày duyệt đăng: 23/4/2026)

TÓM TẮT

Nghiên cứu nhằm xác định phương pháp dự báo chuỗi thời gian có khả năng ứng dụng thực tiễn trong điều kiện dữ liệu ngắn, hỗ trợ nhà quản trị phòng khám ra quyết định vận hành trong thời gian ngắn hạn. Ba phương pháp gồm hồi quy tuyến tính đơn giản, san bằng hàm mũ có điều chỉnh xu hướng và ARIMA được thử nghiệm trên chuỗi dữ liệu lưu lượng khám của một phòng khám chuyên khoa. Các chỉ số sai số dự báo (MAE, RMSE, MAPE) được sử dụng đánh giá hiệu quả mô hình. Kết quả nghiên cứu cho thấy mô hình TAF-c, một biến thể của phương pháp san bằng hàm mũ có điều chỉnh xu hướng, đạt chỉ số sai số thấp nhất (MAE = 20, RMSE = 26 và MAPE = 8,73), cho thấy khả năng ứng dụng hiệu quả trong dự báo ngắn hạn (MAPE < 10%). Nghiên cứu cung cấp bằng chứng thực nghiệm hỗ trợ nhà quản trị phòng khám trong hoạch định nguồn lực, tối ưu dòng chảy vận hành và phát triển chiến lược. Nghiên cứu nhấn mạnh giá trị vai trò của quản trị dựa trên dữ liệu trong bối cảnh chuyển đổi số y tế Việt Nam.

Từ khóa: ARIMA; dự báo chuỗi thời gian; lưu lượng ngoại trú; san bằng hàm mũ có điều chỉnh xu hướng (TAF)

1. Giới thiệu

Những dao động bất thường trong lưu lượng người bệnh là một trong những nguyên nhân dẫn đến tình trạng tắc nghẽn hệ thống vận hành tại các cơ sở y tế. Các gián đoạn trong dòng chảy công việc, xuất phát từ những tình huống ngoài dự đoán, gây sức ép lên nguồn lực vận hành làm suy giảm chất lượng công việc cũng như tăng tỉ lệ kiệt sức và mất gắn kết bên trong đội ngũ nhân viên (Kadri & cộng sự, 2014; Mohamed & Mohamad, 2020). Cùng với sự phát triển của khoa học dữ liệu, các mô hình ra quyết định quản trị dựa trên dữ liệu đã được ứng dụng rộng rãi và mang đến giá trị thực tiễn trong lĩnh vực y tế. Theo Feng và cộng sự (2021), các quyết định quản trị dựa trên dữ liệu được nhận định

là thành tố không thể thiếu trong phân bổ nguồn lực và hoạch định vận hành bệnh viện. Việc tích hợp khoa học dữ liệu vào y tế Việt Nam ngày càng trở nên cấp thiết. Năng lực dự báo của tổ chức trở thành yếu tố trọng yếu trong ước tính những dao động bất thường hằng ngày.

2. Tổng quan nghiên cứu

Nhiều nghiên cứu trước đây đã chứng minh vai trò quan trọng và sự cần thiết của phương pháp dự báo chuỗi thời gian trong bối cảnh y tế. Phần lớn các nghiên cứu chủ yếu được tiến hành tại phòng cấp cứu và các bệnh viện đa khoa lớn, nơi sở hữu hệ thống thông tin y tế với nguồn dữ liệu dài hạn (Arora & cộng sự, 2023; Gul & Celik, 2020; Hu & cộng sự, 2025; Khaldi & cộng sự, 2019). Năm 2016, Gopakumar và cộng sự đã tiến

hàng một nghiên cứu nhằm ước tính nhu cầu sử dụng giường bệnh tại một bệnh viện ở Úc. Với mục tiêu xác định dự đoán cho lưu lượng xuất viện vào ngày kế tiếp, chuỗi dữ liệu bao gồm tổng 12.141 người bệnh trong vòng 1.826 ngày đã được thu thập. Gopakumar đã sử dụng năm mô hình (ARIMA, ARIMAX, k-nearest neighbors, random forest và support vector regression) dựa trên cơ sở thuật toán hồi quy, thực hiện đánh giá mô hình thông qua bốn chỉ số sai số dự báo MFE, MAE, RMSE và sMAPE. Kết quả nghiên cứu đã cho thấy rằng mô hình random forest thể hiện tốt vai trò khi đã cải thiện 22,7% hiệu suất trên MAE khi so sánh với phương pháp naive (Gopakumar & cộng sự, 2016). Năm 2019, Ordu và cộng sự tiến hành một nghiên cứu khác nhằm mục đích xây dựng một khung cơ sở toàn diện cho toàn bộ các dịch vụ y tế cấp cứu tại Anh. Nghiên cứu đối sánh mô hình ARIMA, san bằng hàm mũ, hồi quy tuyến tính đa biến và Seasonal-Trend decomposition thông qua chỉ số MASE. Nghiên cứu đã nhận diện được 64 mô hình hiệu quả từ tổng thể 760 mô hình đề xuất, từ đó nhóm tác giả hình thành các hỗ trợ quyết định cho nhà quản trị bệnh viện trong các hoạt động vận hành liên quan đến nhu cầu sử dụng dịch vụ và phân bổ nguồn lực (Ordu & cộng sự, 2019). Một nghiên cứu khác của Vieira và cộng sự (2023) được thực hiện nhằm khám phá sự khác biệt của các mô hình dự báo chuỗi thời gian đơn giản và đa biến như hồi quy, ARIMA, SARIMA và san bằng hàm mũ. MAE, RMSE, MAPE được sử dụng để xác định mô hình thích hợp nhất trong dự đoán lưu lượng nhập phòng cấp cứu ở một bệnh viện tại Bồ Đào Nha (Vieira & cộng sự, 2023).

Tại Việt Nam, Nguyễn Tấn Phát và Dương Tuấn Anh (2022) đã ứng dụng mô hình ARIMA, Holt-Winters and Seasonal Artificial Neural Network

(SANN) trong ước tính lưu lượng nhập phòng cấp cứu tại một bệnh viện đa khoa khu vực. Kết quả từ nghiên cứu cho thấy các mô hình đề xuất đều đạt được những hiệu quả được chấp nhận sử dụng, trong đó SANN đạt được thành tích tốt nhất (Nguyễn Tấn Phát & Dương Tuấn Anh, 2022). Năm 2024, một nghiên cứu khác nhằm dự đoán nhu cầu cấp cứu tại Thành phố Hồ Chí Minh trong giai đoạn 2021-2025 đã được tiến hành từ bộ dữ liệu thu thập từ năm 2014 đến năm 2020. San bằng hàm mũ, ARIMA và một mô hình do tác giả đề xuất đã được giới thiệu và đưa vào đánh giá hiệu quả. Cuối cùng mô hình ARIMA có yếu tố xu hướng đã được lựa chọn sử dụng đáp ứng yêu cầu của nghiên cứu (Nguyễn Duy Long & cộng sự, 2024).

Đặc điểm chung của các nghiên cứu đã được thực hiện trước đây là đều được ghi nhận thực hiện trên chuỗi dữ liệu dài hạn, nơi các thông tin dữ liệu đủ dài để ứng dụng các phương pháp dự báo nâng cao mà không bị giới hạn trong xây dựng mô hình. Tuy nhiên, thực tế các tập dữ liệu y tế tại các phòng khám chuyên khoa hiện nay tại Việt Nam vẫn còn tồn tại nhiều ở dạng ngắn hạn hoặc phân mảnh do những hạn chế về trang thái thu thập và khả năng lưu trữ. Vì vậy, các nghiên cứu dự báo trên các tập dữ liệu ngắn hạn tại các phòng khám chuyên khoa Việt Nam còn ít được quan tâm triển khai. Trong khi, đối với các phòng khám, các dự báo ngắn hạn có thể hỗ trợ tốt nhà quản trị trong vận hành phòng khám hiệu quả hơn trong điều kiện nguồn lực giới hạn hiện có, giúp đơn vị phòng tránh nhiều rủi ro đến từ các tình huống bất ngờ.

Từ khoảng trống nghiên cứu này, bài báo “Phương pháp dự báo chuỗi thời gian trên chuỗi dữ liệu hạn chế: Trường hợp điển hình tại một phòng khám chuyên khoa Việt Nam” được tiến hành với mục đích bổ sung vào khoảng trống nghiên cứu dự báo tại các phòng khám

chuyên khoa với đặc điểm dữ liệu ngắn. Nghiên cứu này bao gồm hai mục tiêu chính: (1) Xác định phương pháp dự báo phù hợp với đặc điểm dữ liệu chuỗi thời gian ngắn hạn trong bối cảnh phòng khám chuyên khoa tại Việt Nam; (2) Cung cấp bằng chứng thực nghiệm nhằm hỗ trợ cải thiện các thực hành quản trị dựa trên dữ liệu trong vận hành phòng khám. Hai câu hỏi nghiên cứu chính nhằm giải quyết vấn đề nghiên cứu đó là: (1) Phương pháp dự báo chuỗi thời gian nào là phù hợp với đặc điểm dữ liệu hạn chế tại phòng khám chuyên khoa Việt Nam? (2) Các kết quả dự báo có thể hỗ trợ nhà quản trị như thế nào trong công tác ra quyết định quản trị vận hành một phòng khám ngoại trú?

3. Cơ sở lý thuyết và phương pháp nghiên cứu

3.1. Dự báo chuỗi thời gian

Theo Stevenson (2018), dự báo chuỗi thời gian là một kỹ thuật thống kê sử dụng dữ liệu thu thập trong quá khứ để tiến hành dự đoán tương lai. Các công cụ dự báo cung cấp nhiều thông tin có giá trị để tổ chức có thể ước tính nhu cầu trong tương lai từ những quan sát nhu cầu trong quá khứ (Basha & cộng sự, 2017). Chính vì vậy, dự báo đóng vai trò quan trọng trong hoạch định năng lực, thiết lập mục tiêu tổ chức và phát hiện sớm các thay đổi bất thường trong vận hành (Taylor & Letham, 2018). Kết quả thu được từ dự báo cũng được đánh giá là những nguồn đầu vào cơ sở cho quá trình ra quyết định vận hành (Stevenson, 2018).

Một đặc điểm của các mô hình dự báo chuỗi thời gian là khả năng nắm bắt sự phụ thuộc của các quan sát ở thời điểm hiện tại t (y_t) dựa trên các giá trị $y_1, y_2, \dots, y_n$ (Bowerman & cộng sự, 2005). Trong dự báo, nhiều cách tiếp cận được sử dụng đa dạng để tận dụng các quan sát trong quá khứ, nhằm nắm bắt toàn bộ các kiểu thông tin ẩn chứa trong toàn bộ chuỗi dữ liệu, tạo lập cơ sở cho các dự

đoán (Hyndman & Athanasopoulos, 2018). Sự đa dạng của mô hình dự báo được thể hiện ở khả năng quan sát và thu thập tại các tần suất khác nhau (hàng giờ, hàng ngày, hàng tháng hoặc hàng năm) hoặc những thành phần tác động điển hình trong chuỗi dữ liệu như thành phần xu hướng, yếu tố mùa vụ hay các biến ngẫu nhiên.

3.2. Các phương pháp dự báo đề xuất

3.2.1. Hồi quy tuyến tính đơn giản

Phương pháp hồi quy tuyến tính đơn sử dụng mô hình hóa mối quan hệ tuyến tính giữa biến phụ thuộc và biến độc lập trong phương trình hồi quy. Trong khi biến phụ thuộc đại diện biểu thị cho kết quả dự báo cần xác định, biến độc lập thể hiện vai trò giải thích, cung cấp thông tin đầu vào để xác định sự biến động của biến phụ thuộc. Mục tiêu cơ bản của phương pháp này là xây dựng một phương trình tuyến tính thể hiện mối quan hệ giữa hai biến, trong đó các tham số mô hình được ước tính bằng phương pháp bình phương tối thiểu nhằm tối thiểu hóa tổng bình phương sai số giữa giá trị quan sát và giá trị dự đoán, thể hiện trong công thức (1):

$$y_c = a + bx \quad (1)$$

(Nguồn: Stevenson, 2018)

Trong đó: y_c đại diện giá trị dự báo (biến phụ thuộc), x là biến giải thích (biến độc lập), b thể hiện hệ số góc và a là giao điểm khi $x = 0$.

3.2.2. San bằng hàm mũ có điều chỉnh xu hướng (TAF)

Phương pháp san bằng hàm mũ có điều chỉnh xu hướng là một mở rộng của phương pháp san bằng hàm mũ đơn giản, trong đó yếu tố xu hướng được tích hợp để phản ánh sự thay đổi có hệ thống của chuỗi dữ liệu theo thời gian. Không giống như san bằng đơn giản, phương pháp san bằng hàm mũ có điều chỉnh xu hướng liên tục cho phép cập nhật cả thành phần cấp độ và xu hướng tại mỗi giai đoạn Thành phần cấp độ phản ánh

giá trị san bằng tại thời điểm hiện tại của chuỗi, trong khi yếu tố xu hướng nắm bắt tốc độ thay đổi trung bình của chuỗi theo thời gian. Hai tham số α và β đại diện lần lượt hệ số làm mịn cấp độ và hệ số điều chỉnh xu hướng, với giá trị thường nằm trong khoảng $0 < \alpha, \beta < 1$ phổ biến. Giá trị dự báo cho giai đoạn tiếp theo được tính theo công thức (2):

$$TAF_{t+1} = S_t + T_t \quad (2)$$

(Nguồn: Stevenson, 2018)

Trong đó: S_t đại diện thành phần cấp độ tại thời điểm t , T_t đại diện ước tính xu hướng tại thời điểm t . Công thức cập nhật thành phần ước tính theo Công thức (3) và (4):

$$S_t = TAF_t + \alpha(A_t - TAF_t) \quad (3)$$

$$T_t = T_{t-1} + \beta(TAF_t - TAF_{t-1} - T_{t-1}) \quad (4)$$

(Nguồn: Stevenson, 2018)

Trong đó, A_t đại diện giá trị quan sát tại thời điểm t , TAF_t là giá trị dự báo điều chỉnh xu hướng tại thời điểm t .

3.2.3. Phương pháp ARIMA

Phương pháp ARIMA được phát triển bởi Box và Jenkins nhằm mô hình hóa và dự báo dựa trên cấu trúc tự tương quan nội tại của dữ liệu (Box & cộng sự 2015). Trong mô hình này, các quan sát được sắp xếp theo thứ tự thời gian và các giá trị trong quá khứ được sử dụng để giải thích và dự báo tương lai. Một điều kiện quan trọng khi áp dụng ARIMA là chuỗi dữ liệu phải ở trạng thái dừng hoặc có thể được chuyển đổi về trạng thái dừng thông qua các phép biến đổi phù hợp. Theo Wang (2008), một chuỗi thời gian được xem đạt tính dừng khi giá trị trung bình và phương sai không thay đổi theo thời gian, đồng thời không tồn tại xu hướng hệ thống.

Phương pháp luận Box-Jenkins mô tả quy trình xây dựng mô hình ARIMA qua bốn bước: Nhận diện mô hình - Ước lượng tham số - Kiểm tra chẩn đoán - Dự báo (Bowerman & cộng sự, 2005; Wilson & Keating, 2009). Mô hình ARIMA được biểu diễn qua bộ ba tham

số: tự hồi quy (AR), tích hợp (I) và trung bình động (MA). Mô hình có dạng ARIMA (p,d,q) trong đó p đại diện thành phần tự hồi quy, d số lần sai phân để đạt tính dừng và q là thành phần trung bình động (Cipra, 2020; Wang, 2008).

Việc lựa chọn mô hình ARIMA phù hợp thường dựa trên các tiêu chí thông tin nhằm cân bằng giữa độ phù hợp mô hình và mức độ phức tạp tham số. Hai tiêu chí phổ biến là Akaike Information Criterion (AIC) theo Công thức (5) và Bayesian Information Criterion (BIC) theo Công thức (6). Mô hình có giá trị AIC hoặc BIC thấp hơn thường được ưu tiên, do phản ánh mức độ phù hợp tốt hơn với số lượng tham số tối ưu.

$$AIC = 2k - 2\ln(L) \quad (5)$$

$$BIC = k\ln(n) - 2\ln(L) \quad (6)$$

(Nguồn: Zhang & Meng, 2023)

Trong đó, k là số lượng tham số của mô hình, L đại diện giá trị hàm hợp lí (likelihood) của mô hình và n là kích thước mẫu (Zhang & Meng, 2023).

3.3. Đánh giá độ chính xác dự báo

Đánh giá độ chính xác dự báo là một tiêu chí quan trọng trong quá trình lựa chọn mô hình dự báo phù hợp. Theo Stevenson (2018), việc đánh giá sai số dự báo cung cấp bằng chứng định lượng nhằm hỗ trợ các quyết định quản trị. Trong thực tiễn nghiên cứu và ứng dụng, các thước đo sai số phổ biến bao gồm Sai số tuyệt đối trung bình (MAE), Căn bậc hai sai số bình phương trung bình (RMSE) và Sai số phần trăm tuyệt đối trung bình (MAPE). Các chỉ số này phản ánh mức độ sai lệch giữa giá trị dự báo và giá trị quan sát thực tế của chuỗi dữ liệu. Thông thường, mô hình dự báo có giá trị sai số thấp hơn thường được ưu tiên có độ chính xác dự báo tốt hơn, được ưu tiên sử dụng.

3.4. Địa điểm và dữ liệu nghiên cứu

Dữ liệu nghiên cứu được tiến hành thu thập từ Phòng khám Suy tim, tầng 14, Bệnh viện Đại Học Y Dược Thành

phổ Hồ Chí Minh. Phòng khám Suy tim đã được thành lập và đi vào hoạt động từ tháng 8 năm 2022, thuộc hệ thống khám ngoại trú với cơ chế hoạt động định kỳ ba buổi mỗi tuần và cung cấp dịch vụ khám chuyên sâu, điều trị, tư vấn và quản lý bệnh mạn tính cho người bệnh suy tim. Với đặc điểm bệnh lý đòi hỏi theo dõi liên tục, điều chỉnh điều trị và tư vấn lối sống thường xuyên, mô hình phòng khám chuyên biệt cho phép tăng cường tương tác giữa người bệnh và bác sĩ chuyên khoa, qua đó hỗ trợ cải thiện kết quả điều trị và trải nghiệm người bệnh (Bệnh viện Đại học Y Dược Thành phố Hồ Chí Minh, 2022).

Chuỗi dữ liệu sử dụng trong nghiên cứu bao gồm lưu lượng người bệnh ngoại trú tại phòng khám suy tim, với 40 tháng quan sát (tháng 8/2022 đến tháng 11/2025). Bộ dữ liệu này phù hợp cho việc áp dụng các phương pháp dự báo chuỗi thời gian trong bối cảnh dữ liệu ngắn. Dữ liệu được trích xuất từ hệ thống thông tin phòng khám, đảm bảo tuân thủ các quy định về bảo mật và an toàn thông tin. Quá trình tiền xử lý cho thấy không tồn tại dữ liệu khuyết trong chuỗi. Phân tích định lượng và dự báo được thực hiện bằng Microsoft Excel và Stata. Dựa trên tổng quan tài liệu và điều kiện dữ liệu thực tiễn, các mô hình được lựa chọn để thử nghiệm bao gồm hồi quy tuyến tính đơn giản, các biến thể của phương pháp san bằng hàm mũ có điều chỉnh xu hướng và mô hình ARIMA.

3.5. Quy trình nghiên cứu

3.5.1. Xây dựng mô hình dự báo

A. Mô hình hồi quy tuyến tính đơn giản

Mô hình hồi quy tuyến tính đơn giản được ước tính và kiểm định bằng phần mềm Stata. Thời gian (tháng) đóng vai trò như biến giải thích cho biến động lưu lượng, trong khi lưu lượng người bệnh được xem là biến phụ thuộc cần đo lường. Khả năng dự báo của mô hình hồi quy tuyến tính đơn giản được đánh giá

khi sử dụng giá trị thống kê R^2 hiệu chỉnh và giá trị ý nghĩa thống kê p-value trong khoảng tin cậy 95%.

B. Mô hình san bằng hàm mũ có điều chỉnh xu hướng

Mô hình san bằng hàm mũ có điều chỉnh xu hướng được ước tính bằng công cụ Solver trong Microsoft Excel nhằm xác định các tham số làm mịn tối ưu. Các tham số làm mịn α và β được thiết lập là các biến quyết định trong khi hàm mục tiêu được xác định là tối thiểu hóa giá trị sai số tuyệt đối trung bình MAE. Các tham số α và β ràng buộc trong khoảng $(0,1)$, phù hợp với nguyên tắc san bằng hàm mũ. Quá trình tối ưu hóa được thực hiện lặp lại nhằm tìm kiếm cặp tham số làm mịn cho phép giảm thiểu sai số dự báo, từ đó xác định cặp tham số phù hợp cho mô hình. Nhằm cung cấp thêm bằng chứng so sánh và tăng độ tin cậy của kết quả nghiên cứu, ba biến thể bổ sung của mô hình được xây dựng để kiểm tra độ nhạy, dựa trên khác biệt về cách thức khởi tạo giá trị dự báo ban đầu trong từng chu trình tính toán.

Biến thể TAF-a tạo giá trị khởi đầu tại $t = 2$ dựa trên phương pháp naive. Cụ thể, thành phần cấp độ được khởi tạo bằng giá trị quan sát thực tế tại giai đoạn trước đó ($S_{t=2} = A_{t=1}$), trong khi thành phần xu hướng được ước tính = 0 với giả định chuỗi chưa có xu hướng rõ ràng tại thời điểm đầu.

Biến thể TAF-b bắt đầu với giá trị dự báo tại thời điểm $t = 5$ với giá trị khởi đầu của cấp độ và xu hướng ($S_{t=5}$ và $T_{t=5}$) được xác định dựa trên trung bình động thay đổi của ba giai đoạn liên tiếp gần nhất. Cách khởi tạo này cho phép mô hình tận dụng thông tin từ nhiều quan sát ban đầu nhằm ước lượng đồng thời cả mức độ và xu hướng của chuỗi dữ liệu.

Biến thể TAF-c cũng bắt đầu tại thời điểm $t = 5$. Tuy nhiên khác với TAF-b, biến thể TAF-c sử dụng phương pháp naive để khởi tạo thành phần cấp độ ($S_{t=5}$

$= A_{t=4}$), trong khi thành phần xu hướng khởi đầu ($T_{t=5}$) được ước tính bằng trung bình động thay đổi xu hướng trong ba giai đoạn trước đó. Cách tiếp cận này kết hợp giữa giả định đơn giản về cấp độ và ước lượng xu hướng dựa trên thông tin lịch sử gần nhất.

C. Mô hình ARIMA

Mô hình ARIMA(p,d,q) được xây dựng theo phương pháp luận Box-Jenkins và được triển khai trên phần mềm Stata. Kiểm định Augmented Dickey-Fuller (ADF) được thực hiện trước tiên nhằm đánh giá tính dừng của chuỗi dữ liệu thời gian. Trong trường hợp chuỗi chưa dừng, phép sai phân được tiến hành để chuyển đổi chuỗi về trạng thái dừng trước khi mô hình hóa. Các cấu trúc tự tương quan của chuỗi được phân tích thông qua hàm tự tương quan (ACF) và hàm tự tương quan một phần (PACF) nhằm hỗ trợ nhận diện các bậc thích hợp của thành phần (p, q).

Độ phù hợp và ổn định của mô hình ARIMA được đánh giá qua giá trị log-likelihood và các tiêu chí thông tin như Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), trong đó các mô hình có giá trị AIC hoặc BIC thấp hơn thường được ưu tiên lựa chọn. Kiểm định phần dư là giai đoạn quan trọng với mục tiêu xác định phần dư là nhiễu trắng nhằm đảm bảo mô hình đã nắm bắt toàn bộ cấu trúc phụ thuộc của chuỗi dữ liệu và không còn tồn tại các mẫu tự tương quan có ý nghĩa thống kê trong phần dư. Các kiểm định tự

tương quan phần dư, kiểm định Ljung-box Q, kiểm định phân phối chuẩn sẽ được tiến hành. Kiểm định nghiệm nghịch đảo cũng được thực hiện để kiểm tra điều kiện ổn định và khả nghịch của mô hình ARIMA.

3.5.2. Đánh giá mô hình đề xuất

Hiệu suất dự báo mô hình được đánh giá nhằm so sánh độ chính xác dự báo giữa các phương pháp. Dựa vào đặc điểm cấu trúc của chuỗi dữ liệu thời gian, ba chỉ số sai số dự báo MAE, RMSE, MAPE được lựa chọn sử dụng. Quyết định lựa chọn cuối cùng được thực hiện với dự báo sai số thấp nhất trong nhóm đề xuất sẽ làm cơ sở đề xuất nền tảng cho phân tích và ứng dụng dự báo.

3.5.3. Chuyển đổi kết quả dự báo thành hành động quản trị

Kết quả dự báo lưu lượng người bệnh tái khám dựa trên mô hình hiệu suất tốt hơn, được chuyển hóa thành các hàm ý quản trị mang tính thực tiễn. Các kết quả dự báo hỗ trợ nhà quản trị phòng khám trong việc đánh giá hiệu quả vận hành, cải thiện dòng chảy người bệnh và dòng chảy lâm sàng, phân bổ nguồn lực hợp lý và hoạch định các chiến lược mở rộng công suất phòng khám trong tương lai.

4. Kết quả nghiên cứu

4.1. Kết quả phân tích mô hình

Kết quả phân tích thống kê mô tả của mô hình hồi quy tuyến tính đơn giản với khoảng tin cậy 95%, thu được từ phần mềm Microsoft Excel được thể hiện trong bảng 1.

Bảng 1: Kiểm định thống kê mô hình hồi quy tuyến tính đơn

Thống kê hồi quy	
R^2	0,61
R^2 hiệu chỉnh	0,60

(Nguồn: Kết quả phân tích dữ liệu của nhóm tác giả, 2025)

Ba biến thể của mô hình san bằng hàm mũ có điều chỉnh xu hướng thu được từ công cụ Solver trên phần mềm

Microsoft Excel. Kết quả ba cặp tham số α và β thể hiện trong bảng 2.

Bảng 2: Tổng hợp các cặp tham số mô hình TAF

Mô hình TAF	Tham số α	Tham số β
Biến thể TAF-a	0,52	0,17
Biến thể TAF-b	0,40	0,10
Biến thể TAF-c	0,40	0,10

(Nguồn: Kết quả phân tích dữ liệu của nhóm tác giả, 2025)

Mô hình ARIMA được xây dựng bằng phần mềm Stata. Thông qua kiểm định ADF cho thấy chuỗi dữ liệu gốc chưa đạt tính dừng, vì vậy kiểm định tiên hành thực hiện sai phân ($d=1$) để chuyển đổi thành chuỗi dừng. Các mô hình ARIMA được nhận diện thông qua các

phân tích tự tương quan ACF và tự tương quan một phần PACF bao gồm ARIMA (1,1,1), ARIMA (5,1,1), ARIMA (1,1,5), ARIMA (2,1,1), ARIMA (5,1,5). Mức độ phù hợp của mô hình ARIMA được trình bày trong bảng 3.

Bảng 3: Tổng hợp tiêu chí lựa chọn mô hình ARIMA

Model	Log likelihood	AIC	BIC
ARIMA (1,1,1)	-185	379	386
ARIMA (5,1,1)	-182	381	394
ARIMA (1,1,5)	-183	381	394
ARIMA (2,1,1)	-184	379	387
ARIMA (5,1,5)	-178	380	400

(Nguồn: Kết quả phân tích dữ liệu của nhóm tác giả, 2025)

Mô hình ARIMA (1,1,1) đạt giá trị AIC, BIC thấp nhất và chỉ số hàm hợp lí tốt nhất. Kiểm định phần dư mô hình ARIMA (1,1,1) cho thấy phần dư là nhiễu trắng (Kiểm định Ljung-Box Q với p -value = 0,52; phần dư không còn tự tương quan; phần dư phân phối chuẩn; Bartlett's B = 0,32), và nghiệm nghịch

đạo thoả điều kiện mô hình ổn định. Vì vậy, ARIMA (1,1,1) trở thành mô hình ARIMA tối ưu nhất trong các mô hình ARIMA.

4.2. Kết quả đánh giá mô hình

Độ chính xác dự báo của các mô hình đề xuất được đánh giá thể hiện trong bảng 4.

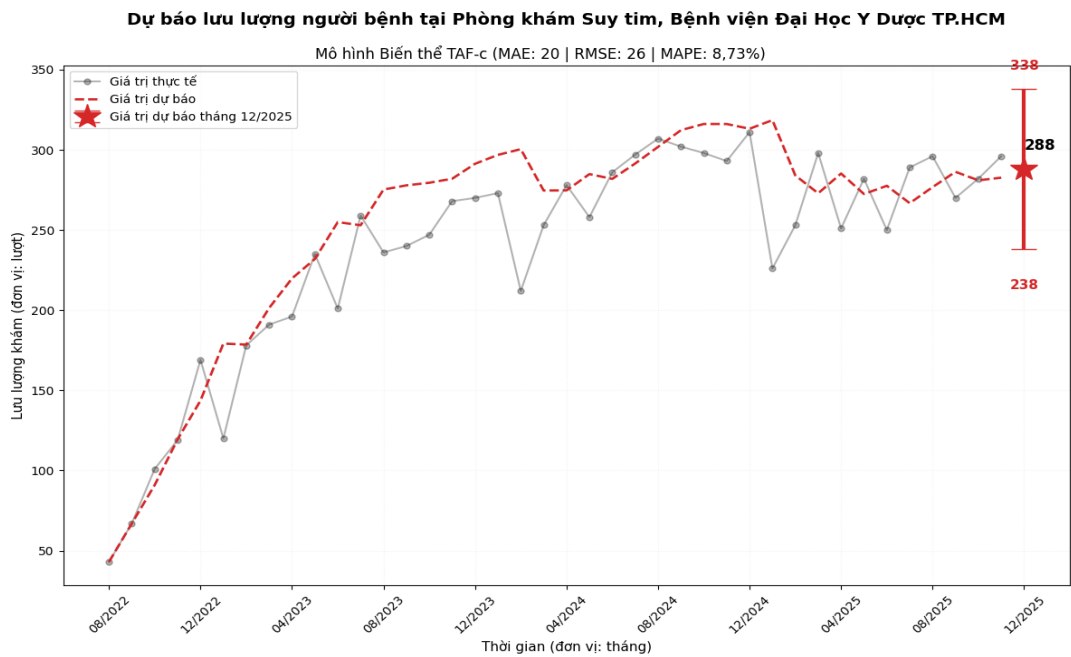
Bảng 4: Tổng hợp đánh giá độ chính xác dự báo các mô hình

Mô hình đề xuất	Chỉ số MAE	Chỉ số RMSE	Chỉ số MAPE
Hồi quy tuyến tính đơn	36	43	23,15%
ARIMA (1,1,1)	22	28	10,85%
Biến thể TAF-a	23	30	12,38%
Biến thể TAF-b	21	27	9,01%
Biến thể TAF-c	20	26	8,73%

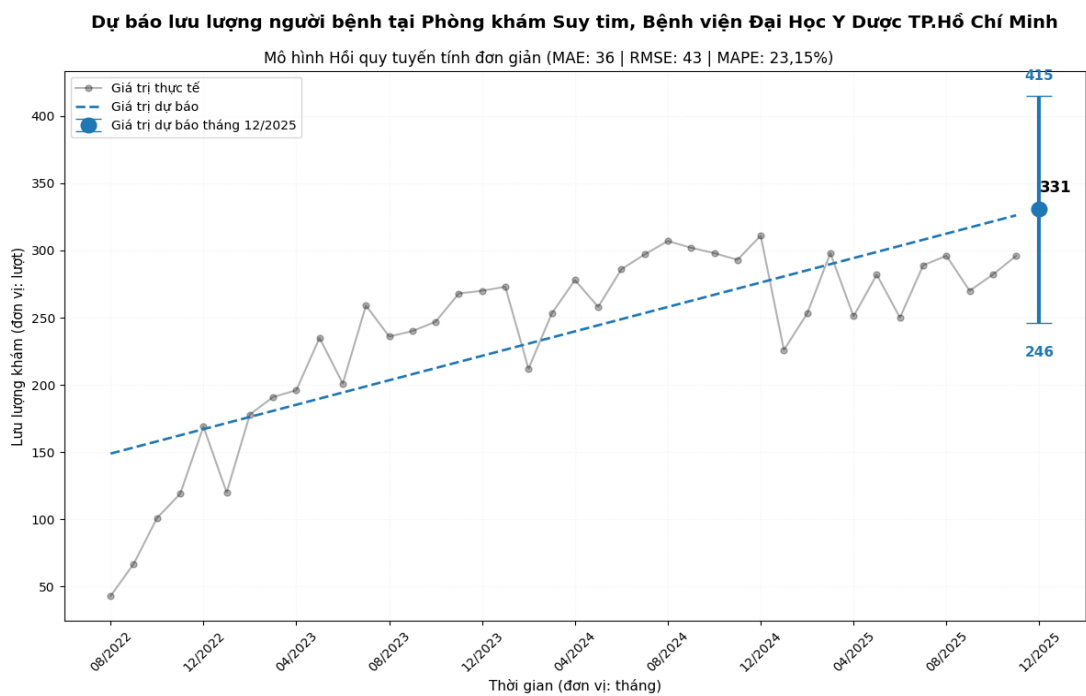
(Nguồn: Kết quả phân tích dữ liệu của nhóm tác giả, 2025)

Các biểu đồ hình 1 - hình 5 thể hiện so sánh giữa giá trị dự báo và giá trị thực tế về lưu lượng khám theo thời gian, thực hiện bằng các mô hình đề xuất. Dựa trên kết quả đánh giá mô hình dựa trên độ chính xác dự báo (bảng 4), biến thể TAF-c được lựa chọn là mô hình phù hợp nhất để tiến hành dự báo, nghiên cứu tại phòng khám, khi đạt được độ sai số thấp

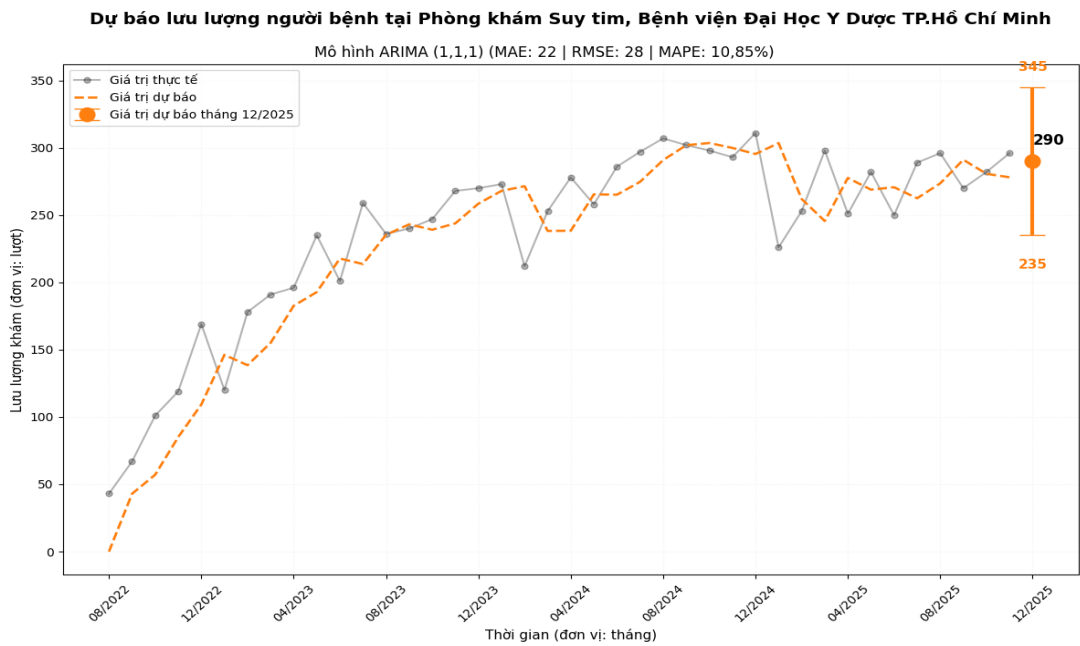
nhất ở cả ba chỉ số tham gia xem xét ($MAE = 20$, $RMSE = 26$, $MAPE = 8,73\%$). Tiến hành dự báo lưu lượng người bệnh khám bằng biến thể TAF-c ghi nhận kết quả 288 người bệnh sẽ đến khám trong tháng 12 năm 2025 (dao động 238 đến 338 người bệnh với khoảng tin cậy 95%) (hình 1).



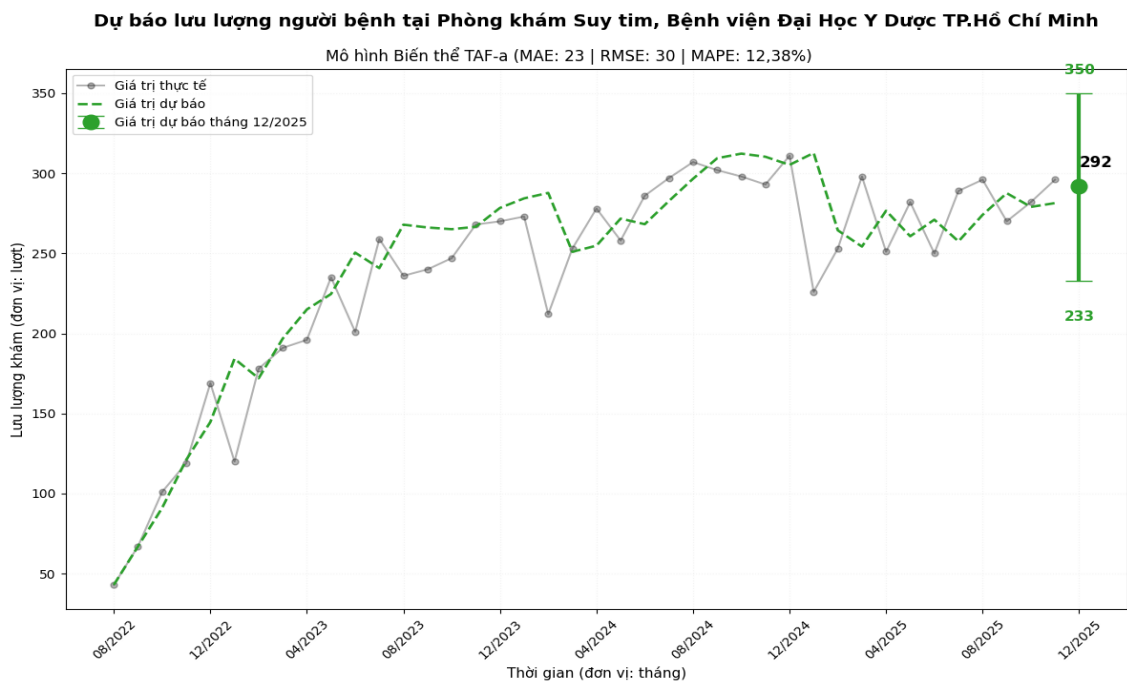
Hình 1: Biểu đồ dự báo lưu lượng người bệnh đến khám ngoại trú tại Phòng khám Suy tim tháng 12/2025 bằng mô hình TAF-c (Nguồn: Kết quả phân tích dữ liệu của nhóm tác giả, 2025)



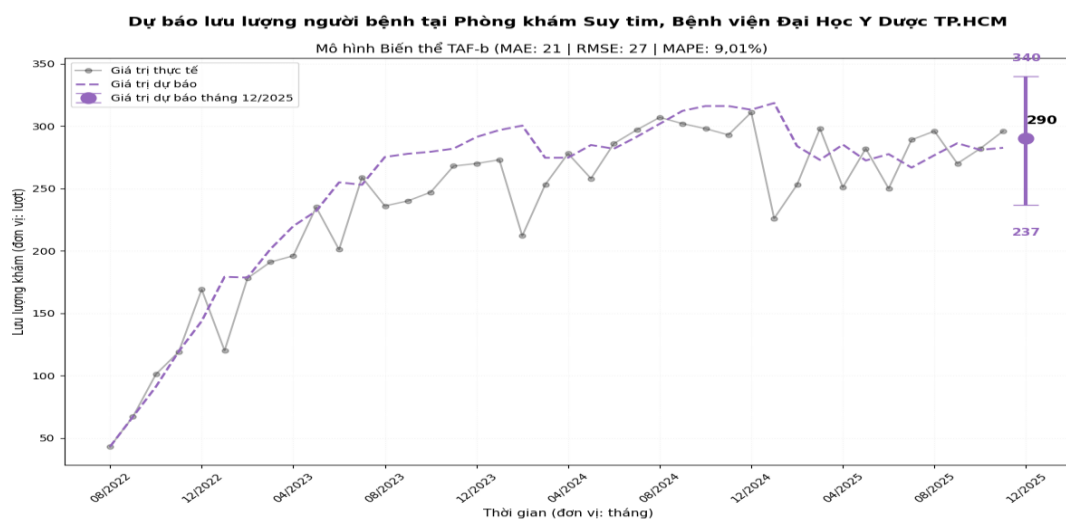
Hình 2: Biểu đồ dự báo lưu lượng người bệnh đến khám ngoại trú tại Phòng khám Suy tim tháng 12/2025 bằng mô hình Hồi quy tuyến tính đơn giản (Nguồn: Kết quả phân tích dữ liệu của nhóm tác giả, 2025)



Hình 3: Biểu đồ dự báo lưu lượng người bệnh đến khám ngoại trú tại Phòng khám Suy tim tháng 12/2025 bằng mô hình ARIMA (1,1,1)
(Nguồn: Kết quả phân tích dữ liệu của nhóm tác giả, 2025)



Hình 4: Biểu đồ dự báo lưu lượng người bệnh đến khám ngoại trú tại Phòng khám Suy tim tháng 12/2025 bằng mô hình biến thể TAF-a
(Nguồn: Kết quả phân tích dữ liệu của nhóm tác giả, 2025)



Hình 5: Biểu đồ dự báo lưu lượng người bệnh đến khám ngoại trú tại Phòng khám Suy tim tháng 12/2025 bằng mô hình biến thể TAF-b

(Nguồn: Kết quả phân tích dữ liệu của nhóm tác giả, 2025)

5. Thảo luận

Kết quả hồi quy tuyến tính đơn giản cho thấy khoảng 60% biến động của lưu lượng khám ngoại trú được giải thích bởi biến thời gian (R^2 hiệu chỉnh = 0,60; p -value < 0,05), cho thấy sự tồn tại của xu hướng theo thời gian. Tuy nhiên, mô hình này còn hạn chế trong khả năng giải thích khi sai số dự báo có giá trị tương đối cao (MAE = 36, RMSE = 43, MAPE = 23,15%).

Đối với các mô hình ARIMA, các cấu trúc tham số cao như ARIMA (5,1,1), ARIMA (1,1,5) và ARIMA (5,1,5) cho giá trị AIC và BIC cao hơn, cho thấy mức độ phù hợp kém hơn trong bối cảnh dữ liệu ngắn (bảng 3). Kết quả này nhấn mạnh sự phức tạp trong mô hình không thực sự cần thiết khi xây dựng dự báo trong bối cảnh hạn chế dữ liệu. Ngược lại, mô hình ARIMA (1,1,1) đạt kết quả tốt hơn (MAE = 22; RMSE = 28; MAPE = 10,85%) và đáp ứng các kiểm định chẩn đoán phần dư, với phần dư có đặc tính nhiễu trắng. Việc lựa chọn mô hình ARIMA (1,1,1) cho thấy khả năng nắm bắt cấu trúc động cơ bản của chuỗi thời gian phù hợp với những phát hiện của Vieira và cộng sự (2023).

Trong các mô hình được so sánh, biến thể TAF-c đạt hiệu suất tốt nhất với sai số thấp nhất (MAE = 20, RMSE = 26, MAPE = 8,73%), cho thấy khả năng ứng dụng hiệu quả trong dự báo ngắn hạn. Với MAPE < 10%, kết quả gợi ý rằng việc tích hợp thành phần xu hướng giúp cải thiện độ chính xác dự báo trong bối cảnh dữ liệu ngắn. Một cách lí giải cho kết quả này có thể liên quan đến đặc điểm của chuỗi dữ liệu trong nghiên cứu. Cụ thể, chuỗi quan sát thể hiện xu hướng rõ ràng theo thời gian nhưng chưa đủ dài để nhận diện chắc chắn mùa vụ hay cấu trúc tự tương quan phức tạp. Trong bối cảnh này, các mô hình tham số như ARIMA phụ thuộc vào việc ước lượng cấu trúc tự tương quan từ dữ liệu lịch sử nên có thể gặp hạn chế khi thông tin chưa đủ rõ ràng. Ngược lại, TAF-c thuộc nhóm mô hình bán thực nghiệm, trong đó cơ chế làm mịn cho phép cập nhật dự báo dựa trên thông tin gần nhất của chuỗi. Điều này giúp mô hình phản ứng linh hoạt hơn với các biến động ngắn hạn mà không phụ thuộc mạnh vào việc xác định cấu trúc toàn cục của dữ liệu. Kết quả này gợi ý rằng các phương pháp dựa

trên làm mịn xu hướng có thể phù hợp hơn trong các bối cảnh dữ liệu ngắn, đồng thời cung cấp cơ sở thực nghiệm hỗ trợ hoạch định và ra quyết định vận hành tại phòng khám.

Các kết quả của nghiên cứu này bổ sung bằng chứng thực nghiệm về nguyên tắc lựa chọn mô hình trong bối cảnh dữ liệu ngắn, trong đó việc kiểm soát độ phức tạp mô hình đóng vai trò quan trọng nhằm hạn chế hiện tượng quá khớp và cải thiện khả năng khái quát hóa trong dự báo ngắn hạn. So với các nghiên cứu trước đây tại Việt Nam chủ yếu tập trung trên chuỗi dữ liệu dài (Nguyễn Tấn Phát & Dương Tuấn Anh, 2022; Nguyễn Duy Long & cộng sự, 2024), nghiên cứu này mở rộng bằng chứng thực nghiệm trên dữ liệu vận hành thực tế ở cấp phòng khám. Kết quả cho thấy các phương pháp dự báo có cấu trúc đơn giản vẫn có thể đạt hiệu quả dự báo tốt khi được lựa chọn phù hợp với đặc điểm của chuỗi dữ liệu. Những phát hiện này gợi ý tiềm năng ứng dụng của các mô hình dự báo trong hỗ trợ ra quyết định vận hành, đặc biệt trong bối cảnh chuyển đổi số y tế tại Việt Nam.

6. Kết luận

Nghiên cứu đã đánh giá ba phương pháp dự báo chuỗi thời gian bao gồm hồi quy tuyến tính đơn, san bằng hàm mũ có điều chỉnh xu hướng và ARIMA trên chuỗi dữ liệu ngắn của phòng khám chuyên khoa trong bối cảnh chuyển đổi số y tế Việt Nam. Năm mô hình dự báo đã được phát triển xây dựng dựa trên ba phương pháp trọng tâm bao gồm mô hình hồi quy tuyến tính đơn, biến thể TAF-a, biến thể TAF-b, biến thể TAF-c và mô hình ARIMA (1,1,1), với hiệu quả dự báo được đánh giá thông qua các chỉ số MAE, RMSE và MAPE. Kết quả cho thấy biến thể TAF-c đạt hiệu suất tốt nhất $MAE = 20$, $RMSE = 26$, $MAPE =$

8,73%), gợi ý rằng việc tích hợp thành phần xu hướng và cơ chế làm mịn giúp cải thiện độ chính xác dự báo trong điều kiện dữ liệu ngắn. Phát hiện này nhấn mạnh vai trò của việc lựa chọn mô hình phù hợp với đặc điểm dữ liệu, trong đó các phương pháp có cấu trúc đơn giản vẫn có thể đạt hiệu quả dự báo cao khi được thiết kế hợp lý.

Các kết quả dự báo được chuyển hóa thành các hành động quản trị mang tính thực tiễn, cụ thể cho các hoạt động quản trị vận hành phòng khám trong hoạch định nguồn lực phòng khám, xây dựng định mức bác sĩ/điều dưỡng phù hợp, thông qua cơ cấu xây dựng ca làm việc luân phiên hoặc kế hoạch tăng cường cộng tác hợp lý và cải thiện dòng chảy công việc. Để tăng tính khả thi trong triển khai, các ngưỡng dự báo có thể được chuyển đổi thành các chỉ tiêu vận hành. Khi lưu lượng dự báo tiệm cận ngưỡng cận trên (khoảng 338 lượt/tháng), phòng khám có thể kích hoạt các biện pháp tăng cường như mở thêm bàn khám, bổ sung nhân sự hỗ trợ hoặc mở rộng thời gian khám nhằm phân tán áp lực. Đồng thời, việc thiết lập ngưỡng cảnh báo vận hành tại mức này cho phép chủ động kích hoạt sớm các cơ chế điều phối. Ngược lại, khi lưu lượng giảm về gần ngưỡng cận dưới (khoảng 238 lượt/tháng), phòng khám có thể tối ưu công suất bằng cách giảm số buổi khám, gộp ca hoặc tái phân bổ nhân sự sang các hoạt động hỗ trợ như chuẩn hóa quy trình, đào tạo nội bộ. Cách tiếp cận dựa trên ngưỡng này cho phép hình thành một cơ chế ra quyết định dựa trên dữ liệu, góp phần nâng cao hiệu quả sử dụng nguồn lực và khả năng thích ứng trong vận hành.

Các kỹ thuật dự báo có thể triển khai trên các nền tảng phổ biến như Microsoft Excel hoặc Power BI, cũng như tích hợp

vào hệ thống thông tin bệnh viện (HIS), qua đó hỗ trợ nhà quản trị theo dõi và ra quyết định kịp thời trong môi trường vận hành năng động. Nhìn chung, nghiên cứu cung cấp bằng chứng thực nghiệm về khả năng ứng dụng của các phương pháp dự báo chuỗi thời gian trong bối cảnh dữ liệu hạn chế, đồng thời góp phần thúc đẩy cách tiếp cận quản trị dựa trên dữ liệu trong hệ thống y tế.

7. Hạn chế nghiên cứu và hướng nghiên cứu tiếp theo

Nghiên cứu còn tồn tại một số hạn chế nhất định cần xem xét cải thiện trong các nghiên cứu tiếp nối. Thứ nhất, cách thức đánh giá mô hình trong nghiên cứu thừa nhận tiến hành đánh giá trong mẫu xuất phát từ điều kiện hạn chế chuỗi dữ

liệu còn ngắn. Thứ hai, nghiên cứu thừa nhận chưa xem xét đến các yếu tố có khả năng tác động đến sự biến động lưu lượng khám như các yếu tố mùa vụ, do hạn chế về độ dài của chuỗi mẫu dữ liệu nghiên cứu. Thứ ba, nghiên cứu được thực hiện tại một địa điểm, chưa thể khái quát hóa kết quả nghiên cứu.

Trong các nghiên cứu tiếp theo, nhóm tác giả dự kiến sử dụng các kỹ thuật đánh giá ngoài mẫu như kiểm định chéo chuỗi thời gian và tiến hành xây dựng các mô hình dự báo có yếu tố mùa vụ như SARIMA hay Holt-Winters hoặc mô hình hybrid. Việc mở rộng nghiên cứu sang đa trung tâm cũng được cân nhắc nhằm nâng cao khả năng diễn giải kết quả nghiên cứu ở mức độ khái quát hơn.

Lời cảm ơn: Nghiên cứu này là kết quả của đề tài nghiên cứu cấp cơ sở Đại học Kinh tế Thành phố Hồ Chí Minh, mã số CS-COB-2025-34.

TÀI LIỆU THAM KHẢO

- Arora, S., Taylor, J. W., & Mak, H.-Y. (2023). Probabilistic forecasting of patient waiting times in an emergency department. *Manufacturing & Service Operations Management*, 25(4), 1489–1508. <https://doi.org/10.1287/msom.2023.1210>.
- Basha, S. M., Zhenning, Y., Rajput, D. S., Caytiles, R. D., & Iyengar, N. C. S. (2017). Comparative study on performance analysis of time series predictive models. *International Journal of Grid and Distributed Computing*, 10(8), 37–48. <http://dx.doi.org/10.14257/ijgcd.2017.10.8.04>.
- Bệnh viện Đại học Y Dược Thành phố Hồ Chí Minh. (2022). *Phòng khám Suy tim* [Tài liệu nội bộ].
- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.
- Bowerman, B. L., O'Connell, R. T., & Koehler, A. B. (2005). *Forecasting, time series, and regression: An applied approach* (4th ed.). Thomson Brooks/Cole.
- Cipra, T. (2020). Box-Jenkins Methodology. In *Time Series in Economics and Finance*, 123–173. Springer International Publishing. https://doi.org/10.1007/978-3-030-46347-2_6.
- Feng, D., Mo, Y., Tang, Z., Chen, Q., Zhang, H., Akerkar, R., & Song, X. (2021). Data-driven hospital personnel scheduling optimization through patients prediction. *CCF Transactions on Pervasive Computing and Interaction*, 3(1), 40–56. <https://doi.org/10.1007/s42486-020-00052-0>.

- Gopakumar, S., Tran, T., Luo, W., Phung, D., & Venkatesh, S. (2016). Forecasting daily patient outflow from a ward having no real-time clinical data. *JMIR Medical Informatics*, 4(3). <https://doi.org/10.2196/medinform.5650>.
- Gul, M., & Celik, E. (2020). An exhaustive review and analysis on applications of statistical forecasting in hospital emergency departments. *Health Systems*, 9(4), 263–284. <https://doi.org/10.1080/20476965.2018.1547348>.
- Hu, Y., Chan, C. W., & Dong, J. (2025). Prediction-driven surge planning with application to emergency department nurse staffing. *Management Science*, 71(3), 2079–2126. <https://doi.org/10.1287/mnsc.2021.02781>.
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- Kadri, F., Harrou, F., Chaabane, S., & Tahon, C. (2014). Time series modelling and forecasting of emergency department overcrowding. *Journal of Medical Systems*, 38(9), 1–20. <https://doi.org/10.1007/s10916-014-0107-0>.
- Khalidi, R., El Afia, A., & Chiheb, R. (2019). Forecasting of weekly patient visits to emergency department: real case study. *Procedia Computer Science*, 148, 532–541. <https://doi.org/10.1016/j.procs.2019.01.026>.
- Mohamed, B., & Mohamad, M. (2020). Using autoregressive integrated moving average (ARIMA) models to predict the number of patient admission in paediatric clinic at a private hospital in Kuantan. *Journal of Sciences and Management Research*.
- Nguyễn Duy Long, Huỳnh Cẩm Nhi, Nguyễn Hồng Yến, Nguyễn Thắng Nhật Tuệ & Phạm Đình Quyết. (2024). Đặc điểm nhu cầu cấp cứu ngoại viện từ năm 2014-2020 và dự báo nhu cầu của người dân tại thành phố Hồ Chí Minh từ năm 2021-2025. *Tạp chí Y học Thành phố Hồ Chí Minh*, 27(5), 165–177. doi: 10.32895/hcjm.m.2024.05.20.
- Nguyễn Tấn Phát & Dương Tuấn Anh. (2022). Dự báo lượng bệnh nhân cấp cứu hàng ngày tại bệnh viện: Một nghiên cứu tình huống. *Tạp chí Khoa học Trường Đại học Ngoại ngữ - Tin học Thành phố Hồ Chí Minh*, 7(2), 13–21.
- Ordu, M., Demir, E., & Tofallis, C. (2019). A comprehensive modelling framework to forecast the demand for all hospital services. *The International Journal of Health Planning and Management*, 34(2), e1257–e1271. <https://doi.org/10.1002/hpm.2771>.
- Stevenson, W. J. (2018). *Operations Management* (13th ed.). McGraw-Hill Education.
- Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37–45. <https://doi.org/10.7287/peerj.preprints.3190v2>.
- Vieira, A., Sousa, I., & Dória-Nóbrega, S. (2023). Forecasting daily admissions to an emergency department considering single and multiple seasonal patterns. *Healthcare Analytics*, 3. <https://doi.org/10.1016/j.health.2023.100146>.
- Wang, G. C. (2008). A guide to Box-Jenkins modeling. *The Journal of Business Forecasting*, 27(1), 19. <https://www.proquest.com/scholarly-journals/guide-box-jenkins-modeling/docview/226917915/se-2>.
- Wilson, J. H., & Keating, B. (2009). *Business forecasting with ForecastX*. McGraw-Hill.

Zhang, Y., & Meng, G. (2023). Simulation of an adaptive model based on AIC and BIC ARIMA predictions. In *Journal of Physics: Conference Series*. doi:10.1088/1742-6596/2449/1/012027.

**TIME SERIES FORECASTING METHODS UNDER LIMITED DATA
CONDITIONS: A CASE STUDY OF A SPECIALIZED
CLINIC IN VIETNAM**

Dang Huu Phuc¹

Vo Nguyen Kim Thao^{1,2}

Nguyen Van Bac³

¹University of Economics Ho Chi Minh City (UEH), Vietnam

²University Medical Center Ho Chi Minh City (UMC), Vietnam

³Military Hospital 175, Vietnam

*Corresponding author: Dang Huu Phuc – Email: phucdh@ueh.edu.vn

(Received: 10/3/2026, Revised: 18/4/2026, Accepted for publication: 23/4/2026)

ABSTRACT

This study aims to identify a time series forecasting approach with practical applicability under short data conditions, supporting clinic managers in short-term operational decision-making. Three methods, including simple linear regression (SLR), trend-adjusted forecasting (TAF) and ARIMA, were applied to outpatient volume data from a specialized clinic. Forecasting performance was evaluated using MAE, RMSE, and MAPE. The results indicate that the TAF-c model, a variant of trend-adjusted exponential smoothing, achieved the lowest forecasting error (MAE = 20, RMSE = 26, and MAPE = 8.73%), demonstrating strong applicability for short-term forecasting (MAPE < 10%). The study provides quantitative evidence to support clinic resource planning, operational flow optimization, and strategic decision-making, while highlighting the role of data-driven management in the context of digital transformation in healthcare in Vietnam.

Keywords: ARIMA; time series forecasting; outpatient volume; trend-adjusted forecasting (TAF)