

DỰ BÁO LƯỢNG BỆNH NHÂN NGOẠI TRÚ TẠI BỆNH VIỆN SỬ DỤNG MÔ HÌNH HOLT-WINTERS, MẠNG NƠ RON NHÂN TẠO ANN VÀ LSTM

Dương Tuấn Anh¹, Lý Thanh Tịnh²

¹ Khoa Công Nghệ Thông Tin, Trường Đại Học Ngoại Ngữ - Tin Học, Thành Phố Hồ Chí Minh

² Bệnh Viện Đa Khoa Khu Vực Củ Chi, Thành Phố Hồ Chí Minh

anhdt@huflit.edu.vn, lythanhthinh@gmail.com

TÓM TẮT— Dự báo lượng bệnh nhân ngoại trú tại bệnh viện là một công việc quan trọng đối với bệnh viện thời hiện đại để thực hiện sự quản lý thông minh những nguồn tài nguyên y tế của bệnh viện. Vì dữ liệu lượng bệnh nhân ngoại trú thường phi tuyến, có tính mùa và hay biến động, chúng tôi sẽ khảo sát so sánh bằng thực nghiệm ba phương pháp dự báo chuỗi thời gian tiêu biểu: mạng nơ ron nhân tạo (ANN), mô hình làm trơn hàm mũ Holt-Winters và mạng Bộ nhớ ngắn dài hạn (LSTM) nhằm tìm ra phương pháp hiệu quả nhất đối với bài toán dự báo này. Mô hình Holt-Winters là một phương pháp xác xuất thống kê, trong khi ANN là phương pháp học máy và LSTM là phương pháp học sâu. Chúng tôi sử dụng ba phương pháp dự báo nêu trên để dự báo lượng bệnh nhân ngoại trú hàng ngày tại Bệnh viện đa khoa khu vực Củ Chi thu thập trong 3 năm từ năm 2018 đến năm 2020. Kết quả thực nghiệm cho thấy LSTM và mô hình Holt-Winters đem lại kết quả dự báo tốt hơn nhiều so với phương pháp ANN. Tiêu chí sai số dự báo MAPE của mô hình LSTM đạt 11.24%.

Từ khóa— lượng bệnh nhân ngoại trú, mô hình Holt-Winters, mạng nơ ron nhân tạo (ANN), LSTM, chuỗi thời gian có tính mùa.

I. GIỚI THIỆU

Dự báo chính xác nhu cầu chăm sóc sức khỏe và tính sẵn sàng của nguồn tài nguyên y tế là một công việc rất cần thiết và quan trọng đối với bệnh viện. Các phòng khám bệnh nhân ngoại trú của bệnh viện đóng một vai trò quan trọng và hiện nay chịu một áp lực về lượng bệnh nhân ngày càng cao (Lou và các cộng sự, 2017 [1]). Dự báo chính xác lượng bệnh nhân ngoại trú một cách hiệu quả sẽ rất hữu ích cho việc lập kế hoạch và phân bổ nhiệm vụ chăm sóc sức khỏe đáp ứng được những nhu cầu y tế.

Do những đặc điểm riêng của dữ liệu lượng bệnh nhân ngoại trú, thí dụ như tính ngẫu nhiên, tính chu kỳ và tính xu hướng [1], trong nhiều năm qua, các nhà nghiên cứu trên thế giới đã đề xuất nhiều phương pháp khác nhau cho bài toán dự báo này, như các mô hình xác xuất thống kê và các phương pháp học máy. Một số những công trình tiêu biểu được liệt kê như sau: Li và các cộng sự, năm 2014 [2] đề xuất mô hình ARIMA (AutoRegressive Integrated Moving Average) để dự báo lượng bệnh nhân ngoại trú hàng tháng tại một bệnh viện hạng 3 ở Trung Quốc; Kim và các cộng sự, năm 2020 [3], so sánh hiệu quả của hai phương pháp: ARIMA và SARIMA (Seasonal ARIMA) dùng để dự báo lượng bệnh nhân ngoại trú với dữ liệu thu thập tại một số bệnh viện ở Gangnam-gu, Seoul và kết quả cho thấy mô hình SARIMA đem lại hiệu quả dự báo tốt hơn mô hình ARIMA; Sumitra và Basri, năm 2020 [4], so sánh hiệu quả của ba phương pháp: ARIMA, Làm trơn hàm mũ đơn giản (Simple Exponential Smoothing – SES) và Holt-Winters dùng để dự báo lượng bệnh nhân ngoại trú với dữ liệu thu thập tại Trung tâm y tế cộng đồng của Indonesia và kết quả cho thấy mô hình Holt-Winters đem lại hiệu quả dự báo tốt nhất trong ba mô hình; Guan và Elgelhardt năm 2019 [5] so sánh hiệu quả của ba phương pháp: Hồi quy tuyến tính (Linear Regression), ANN và mạng nơ ron học sâu Bộ nhớ ngắn dài hạn (Long Short Term Memory - LSTM) dự báo lượng bệnh nhân ngoại trú của một bệnh nhi đồng và kết quả cho thấy mô hình LSTM đem lại hiệu quả dự báo tốt nhất trong ba mô hình. Thapa và Timalisina năm 2023 [6] đề xuất mạng nơ ron học sâu GRU (Gated Recurrent Unit), một biến thể cải tiến của mạng nơ ron LSTM để dự báo lượng bệnh nhân ngoại trú.

Cũng có một số phương pháp lai ghép (hybrid method) được đề xuất để dự báo lượng bệnh nhân ngoại trú, được liệt kê như sau: Wang và các cộng sự năm 2015 [7] đề xuất một phương pháp lai ghép kết hợp một kỹ thuật phân rã chuỗi thời gian (EEMDN) với tổ hợp bộ dự báo cục bộ để dự báo lượng bệnh nhân bệnh tiêu chảy ngoại trú tại một số bệnh viện tại Shanghai, Trung Quốc; Lou và các cộng sự năm 2017 [1] đề xuất một phương pháp lai ghép giữa mô hình ARIMA và mô hình Làm trơn hàm mũ đơn giản (SES) để dự báo lượng bệnh nhân ngoại trú tại một số bệnh viện lớn Trung Quốc; Huang và Wu (2017) [8] đề xuất một phương pháp lai ghép kết hợp một kỹ thuật phân rã chuỗi thời gian (EMD) với một tổ hợp các bộ dự báo ANN để dự báo lượng bệnh nhân ngoại trú; Yu và các cộng sự năm 2017 [9] đề xuất một phương pháp lai ghép kết hợp một kỹ thuật phân rã chuỗi thời gian dựa vào biến đổi Wavelet và tổ hợp các bộ dự báo ANN để dự báo lượng bệnh nhân ngoại trú tại một số bệnh viện ở Trung Quốc; Jiang và các cộng sự năm 2019 [10] đề xuất một phương pháp lai ghép kết hợp mạng nơ ron học sâu và các mô hình hồi quy cổ điển để dự báo lượng bệnh nhân ngoại trú tại Hongkong; Deng và các cộng sự năm 2023 [11] đề xuất một phương pháp lai ghép kết hợp mô hình ARIMA và mô hình học sâu LSTM để dự báo lượng bệnh nhân ngoại trú

tại Bệnh viện Đại học Y khoa Shanxi, Trung Quốc.; Mô hình lai ghép của công trình [11] kết hợp khả năng nắm bắt các thành phần tuyến tính của mô hình ARIMA với khả năng nắm bắt các thành phần phi tuyến của mô hình LSTM.

Một điều đáng ngạc nhiên là trong số những công trình đi trước về dự báo lượng bệnh nhân ngoại trú tại bệnh viện, không hề có công trình nào so sánh hiệu quả dự báo của ba phương pháp dự báo tiêu biểu: mạng nơ ron nhân tạo (ANN), mô hình Làm trơn hàm mũ Holt-Winters và Long-Short-Term Memory (LSTM) mà trong đó Holt-Winters là một phương pháp xác suất thống kê, trong khi ANN là một phương pháp học máy và LSTM là một phương pháp học sâu. Được gợi cảm hứng từ sự thiếu sót vừa nêu, trong nghiên cứu này, chúng tôi nhằm đến mục tiêu so sánh hiệu quả dự báo của ba phương pháp dự báo: mạng nơ ron nhân tạo (ANN), mô hình Holt-Winters và LSTM đối với bài toán dự báo lượng bệnh nhân ngoại trú tại bệnh viện.

Chúng tôi sử dụng ba phương pháp dự báo nêu trên để dự báo lượng bệnh nhân ngoại trú hàng ngày tại Bệnh viện đa khoa khu vực Củ Chi, Thành Phố Hồ Chí Minh với dữ liệu thu thập trong 3 năm từ tháng giêng năm 2018 đến tháng 12 năm 2020. Bộ dữ liệu chuỗi thời gian này có tính phi tuyến, tính xu hướng và tính mùa. Kết quả thực nghiệm trên bộ dữ liệu này cho thấy LSTM và mô hình Làm trơn hàm mũ Holt-Winters đem lại kết quả dự báo tốt hơn rất nhiều so với phương pháp ANN. Tiêu chí sai số dự báo MAPE của mô hình LSTM đạt được một kết quả đầy ấn tượng: 11.24%.

Phần tiếp theo của bài báo được tổ chức như sau: Mục II giới thiệu về các phương pháp dự báo liên quan và kỹ thuật hỗ trợ; Mục III mô tả và phân tích bộ dữ liệu nghiên cứu; Mục IV tường thuật kết quả thực nghiệm đánh giá hiệu quả của ba phương pháp dự báo đối sánh, Mục VI nêu một vài kết luận và các hướng phát triển của nghiên cứu này.

II. CÁC MÔ HÌNH DỰ BÁO VÀ KỸ THUẬT HỖ TRỢ

A. MÔ HÌNH HOLT-WINTERS

Làm trơn hàm mũ (exponential smoothing) là một loại mô hình tuyến tính có thể nắm bắt các đặc trưng tuyến tính trong một chuỗi thời gian. Một trong những ý tưởng căn bản của mô hình làm trơn hàm mũ là tạo ra những giá trị tương lai như là những giá trị trung bình có trọng số của những giá trị trong quá khứ theo đó những giá trị quan sát gần đây được đánh trọng số cao hơn những giá trị quan sát rất xa trong quá khứ. Bằng cách hình thành những giá trị dự báo dựa vào những giá trị trung bình có trọng số, chúng ta đang sử dụng một phương pháp “làm trơn” (smoothing). Thuật ngữ “hàm mũ” (exponential) xuất phát từ sự kiện các mô hình làm trơn hàm mũ không chỉ đánh trọng số giảm dần theo thời gian mà còn giảm dần theo kiểu hàm mũ.

Để ứng dụng mô hình làm trơn hàm mũ, có ba dạng của mô hình này mà được áp dụng rộng rãi cho những chuỗi thời gian khác nhau ([12],[13]). Làm trơn hàm mũ đơn giản (Simple exponential smoothing) (Loại I) được dùng khi chuỗi thời gian không có xu hướng và không có tính mùa. Giả sử chúng ta có chuỗi thời gian Y_t , đo đạc tại những thời điểm $t = 1, \dots, T$. Mô hình làm trơn hàm mũ đơn giản được định nghĩa bằng công thức đệ quy như sau:

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t \quad (1)$$

với $\hat{Y}_{t+1}$ là giá trị dự báo tại thời điểm $t+1$, α là hằng số làm trơn (smoothing constant), Y_t là giá trị dữ liệu thực sự tại thời điểm t , $\hat{Y}_t$: là giá trị dự báo tại thời điểm t .

Khi có tồn tại xu hướng (trend) trong chuỗi thời gian, mô hình hàm mũ Holt (Loại II) có thể được dùng. Phương pháp Holt đòi hỏi ước lượng *độ dốc* (slope) hiện hành và *biên độ* (level) hiện hành, do đó phương pháp dùng hai hằng số làm trơn cho mỗi công thức ước lượng. Hai hằng số làm trơn này giúp tính giá trị ước lượng của biên độ và độ dốc mà biến đổi theo thời gian khi có những quan sát mới được đưa vào. Ba phương trình của phương pháp Holt như sau:

$$\hat{Y}_{t+p} = L_t + pT_t \quad (2)$$

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (3)$$

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

với $\hat{Y}_{t+p}$: giá trị dự báo tại p thời đoạn sau thời điểm t , L_t : giá trị ước lượng của biên độ tại thời điểm hiện hành, T_t : giá trị ước lượng của độ dốc tại thời điểm hiện hành, α : hằng số làm trơn cho biên độ ($0 < \alpha < 1$) và β : hằng số làm trơn cho độ dốc, tức xu hướng ($0 < \beta < 1$).

Trong phương trình (3), biên độ hiện hành (L_t) được tính bằng cách lấy trung bình có đánh trọng số của hai giá trị ước lượng: một ước lượng là giá trị quan sát hiện hành (Y_t) và giá trị ước lượng thứ hai là bằng tổng giá trị xu hướng trước đó một thời đoạn (T_{t-1}) với giá trị biên độ ước lượng trước đó một thời đoạn (L_{t-1}). Phương trình (2) cho thấy giá trị dự báo p thời đoạn về phía tương lai ($\hat{Y}_{t+p}$) được tính bằng cách lấy giá trị ước lượng xu hướng hiện

hành (T_t) nhân với số thời đoạn hướng về tương lai để dự báo (p) và tích số này cộng với giá trị ước lượng của biên độ hiện hành (L_t).

Holt-Winters là một phương pháp làm trơn hàm mũ dùng để dự báo chuỗi thời gian có thể hiện cả tính xu hướng và tính mùa (Loại III). Phương pháp này có hai phiên bản: Holt-Winters dạng nhân (multiplicative Holt-Winters) và Holt-Winters dạng cộng (additive Holt-Winters). Mô hình Holt-Winters dạng nhân được sử dụng phổ biến hơn mô hình Holt-Winters dạng cộng. Bốn phương trình trong lược đồ đệ quy của phương pháp Holt-Winters dạng nhân như sau:

$$\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p} \quad (5)$$

$$L_t = \alpha Y_t / S_{t-s} + (1-\alpha)(L_{t-1} + T_{t-1}) \quad (6)$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \quad (7)$$

$$S_t = \gamma Y_t / L_t + (1-\gamma)S_{t-s} \quad (8)$$

với $\hat{Y}_{t+p}$: giá trị dự báo tại p thời đoạn sau thời điểm t , L_t : giá trị ước lượng của biên độ tại thời điểm hiện hành, T_t : giá trị ước lượng của xu hướng tại thời điểm hiện hành, S_t : giá trị ước lượng của thành phần mùa, α : hằng số làm trơn cho biên độ ($0 < \alpha < 1$), β : hằng số làm trơn cho xu hướng ($0 < \beta < 1$), γ : hằng số làm trơn để ước lượng tính mùa ($0 < \gamma < 1$) và s : độ dài mùa.

Nếu γ bằng 0, nghĩa là không có thành phần mùa trong chuỗi thời gian và phương pháp Holt-Winters trở thành phương pháp Holt. Nếu cả γ và β bằng 0, mô hình trở thành phương pháp làm trơn hàm mũ đơn giản. Như vậy phương pháp Holt-Winters là mô hình tổng quát nhất trong ba mô hình làm trơn hàm mũ.

Trong phiên bản nhân của phương pháp Holt-Winters, ước lượng tính mùa được thực hiện bằng *chỉ số mùa* (seasonal index) và được tính bằng phương trình Eq. (8). Phương trình (8) cho thấy thành phần mùa hiện hành, S_t bằng hằng số γ nhân với chỉ số mùa được ước lượng bằng đại lượng Y_t/L_t cộng với $(1-\gamma)$ nhân thành phần mùa tại thời điểm trước đó S_{t-s} .

B. MÔ HÌNH ANN

Trong nghiên cứu này, một trong những mô hình *mạng nơ ron nhân tạo* (ANN) thông dụng, mạng truyền thẳng một tầng ẩn được dùng để dự báo chuỗi thời gian. Mô hình này được biểu diễn bằng một mạng gồm ba tầng chứa các đơn vị xử lý nối với nhau bằng những đường nối không có chu trình (acyclic link).

Về tổng quát, mô hình có thể được huấn luyện bằng những quan sát trong quá khứ của một chuỗi thời gian để nắm bắt những đặc tính phi tuyến của chuỗi thời gian đó. Các thông số của mô hình (trọng số của các đường nối và độ lệch của các nút) sẽ được cập nhật theo kiểu lặp nhiều lần bởi một quá trình cực tiểu hóa các sai số dự báo. Với công tác dự báo chuỗi thời gian, dựa vào quá trình tính toán trước đó, mối liên hệ giữa giá trị đầu ra (y_t) và giá trị đầu vào ($y_{t-1}, y_{t-2}, \dots, y_{t-p}$) được biểu diễn bằng công thức toán học sau đây để dự báo chuỗi thời gian.

$$y_t = a_0 + \sum_{j=1}^q a_j f\left(w_{0j} + \sum_{i=1}^p w_{ij} y_{t-i}\right) + e_t \quad (9)$$

với a_j ($j = 0, 1, 2, \dots, q$) là độ lệch (bias) tại đơn vị thứ j của tầng ẩn, và w_{ij} ($i = 0, 1, 2, \dots, p; j = 0, 1, 2, \dots, q$) là các trọng số đường nối giữa các tầng của mạng ANN, $f(\cdot)$ là hàm truyền (transfer function) tại tầng ẩn, thí dụ như hàm sigmoid $f(x) = 1/(1+\exp(-x))$, p là số nút ở tầng nhập và q là số nút ở tầng ẩn. Thực tế, mô hình ANN với công thức (9) thực hiện một ánh xạ phi tuyến từ những quan sát trong quá khứ ($y_{t-1}, y_{t-2}, \dots, y_{t-p}$) thành một trị dự báo trong tương lai (y_t), tức là

$$y_t = \varphi(y_{t-1}, y_{t-2}, \dots, y_{t-p}, w) + e_t \quad (10)$$

với w là vector chứa tất cả các thông số của mạng, φ là hàm xấp xỉ được xác định bởi cấu trúc mạng và các trọng số của các đường nối và e_t là sai số dự báo tại thời điểm t .

Việc sử dụng ANN để dự báo chuỗi thời gian hàm ý rằng số đơn vị đầu ra là 1, số đơn vị tầng nhập gắn liền với số giá trị trong quá khứ mà đủ để nhận dạng ra diễn tiến của các bước tương lai. Về số lượng nút ở tầng nhập vẫn chưa có lý thuyết nào giúp xác định được số nút này một cách thích hợp. Nhưng trong công trình [14], Hamzaçebi khuyến nghị rằng số nút ở tầng nhập nên bằng với độ dài mùa s để dự báo tốt hơn đối với chuỗi thời gian có tính mùa, thí dụ, 12 đối với dữ liệu hàng tháng có độ dài mùa là 12 tháng, là 4 đối với dữ liệu hàng quý có độ dài mùa là 4 quý. Hamzaçebi đặt tên mô hình ANN đặc biệt này là *ANN có tính mùa* (Seasonal ANN), viết tắt là SANN. Trong bài báo [14], Hamzaçebi đã so sánh hiệu quả của SANN với mô hình ARIMA có tính mùa trên bốn bộ dữ liệu chuỗi thời gian mẫu và tìm thấy SANN đem lại hiệu quả dự báo tốt hơn mô hình ARIMA có tính mùa trên đa số của bốn bộ dữ liệu này.

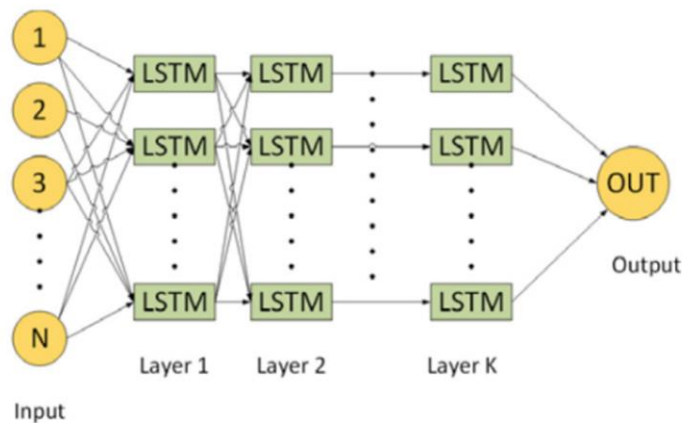
C. MÔ HÌNH LSTM

Mạng nơ ron hồi quy (Recurrent Neural Networks - RNN) là một biến thể của ANN đối với dữ liệu lệ thuộc vào thời gian (time-dependent-data). Để huấn luyện mạng nơ ron hồi quy, giải thuật lan truyền ngược được cải biên thành giải thuật Lan Truyền Ngược Theo Thời Gian (Backpropagation Through Time - BPTT). Tuy nhiên, việc ứng dụng giải thuật BPTT có thể gây ra hai vấn đề: các trọng số có thể dao động quá mạnh, được gọi là vấn đề bùng nổ độ dốc (exploding gradient problem), hoặc thời gian tính toán gia tăng cực cao, được gọi là vấn đề triệt tiêu độ dốc (vanishing gradient problem). Mạng Bộ nhớ ngắn hạn-dài hạn (Long Short-Term Memory - LSTM) là một biến thể cải tiến của mạng RNN, được đề xuất bởi Hochreiter và Schmidhuber (1997) [15] để giải quyết những nhược điểm của mạng RNN khi đối phó với sự phụ thuộc dài hạn (long-term dependencies) của dữ liệu, tức là hai vấn đề bùng nổ độ dốc và triệt tiêu độ dốc. Mỗi đơn vị LSTM là sự tổng quát hóa của đơn vị RNN, sao cho phần thông tin nào liên quan đến những điểm thời gian trước đó sẽ được lưu vào mạng.

Mỗi đơn vị LSTM (còn được gọi là tế bào LSTM) bao gồm ba cổng (gate):

- Cổng quên (Forget gate) đảm nhiệm việc ra quyết định phần thông tin nào từ trạng thái trước đó nên được lưu trữ hay bị loại bỏ.
- Cổng xuất (Output gate) đảm nhiệm việc lựa chọn bao nhiêu thông tin nên được xuất ra.
- Cổng nhập (Input gate), đảm nhiệm việc tiếp nhận thông tin mới.

Mạng nơ ron LSTM học sâu bao gồm nhiều hơn một tầng ẩn. Mạng bao gồm nhiều tầng mà những tế bào LSTM là những đầu ra của tầng trước sẽ trở thành những đầu vào của tầng kế tiếp. Hình 1 mô tả một mạng LSTM xếp chồng mà có thể được dùng để dự báo chuỗi thời gian.



Hình 1. Mạng LSTM xếp chồng

D. CÁCH ƯỚC LƯỢNG ĐỘ DÀI MÙA CHO CHUỖI THỜI GIAN CÓ TÍNH MÙA

Để kiểm tra có sự hiện diện của tính mùa cũng như ước lượng độ dài mùa trên một chuỗi thời gian, chúng ta phải xem xét các vị trí ứng với các độ trễ (lag numbers) của những hệ số tương quan (correlation values) trên biểu đồ hàm tự tương quan (Auto-correlation Function - ACF). Ý tưởng chính là chúng ta tính những khoảng cách Δd giữa những vị trí kế tiếp nhau của các giá trị có tần số cao trong biểu đồ ACF. Nếu có nhiều lần xuất hiện với cùng khoảng cách, khoảng cách Δd xuất hiện thường xuyên nhất sẽ được chọn làm độ dài mùa, dựa vào nguyên tắc đa số.

Phương pháp ước lượng độ dài mùa mà dựa vào biểu đồ ACF nêu trên được đề xuất bởi Ngọc Tran năm 2003 [16]. Phương pháp này bao gồm các bước sau:

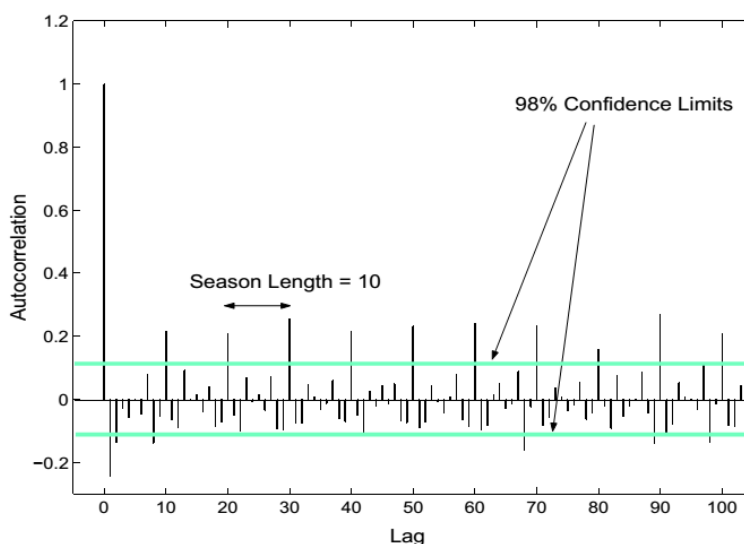
- Thực hiện tính sai phân trên chuỗi thời gian.
- Tính và vẽ biểu đồ ACF cho chuỗi thời gian đã lấy sai phân.
- Lập đường bao với phạm vi độ tin cậy (confidence limit) 95% hay 98%.
- Xác định những vị trí có tần số xuất hiện cao.
- Tính khoảng cách giữa những vị trí kế tiếp của những điểm tần số cao.

Khoảng cách xuất hiện thường xuyên nhất sẽ được chọn là độ dài mùa. Hình 1 minh họa một thí dụ về một biểu đồ ACF với phạm vi độ tin cậy 98%. Chúng ta chọn các vị trí nằm ngoài đường bao với phạm vi độ tin cậy 98% và có được 14 vị trí đầu tiên của những điểm tần số cao, như được nêu trong Bảng 1.

Bảng 1. Thí dụ về các khoảng cách của những điểm tần số cao

Vị trí tần số cao	0	1	2	8	10	20	30	40	50	60	68	70	80	90
Khoảng cách		1	1	6	2	10	10	10	10	10	8	2	10	10

Sau khi tính khoảng cách giữa các vị trí kế tiếp của những điểm tần số cao, chúng ta có được 13 khoảng cách trong đó có 7 lần xuất hiện của khoảng cách với độ dài 10, 2 lần xuất hiện của khoảng cách với độ dài 2, 1 lần xuất hiện của khoảng cách với độ dài 6 và 1 lần xuất hiện của khoảng cách với độ dài 8. Như vậy, khoảng cách với độ dài 10 có số lần xuất hiện cao nhất và 10 sẽ được chọn như là độ dài mùa trong thí dụ này.



Hình 2. Phát hiện độ dài mùa dựa vào biểu đồ ACF

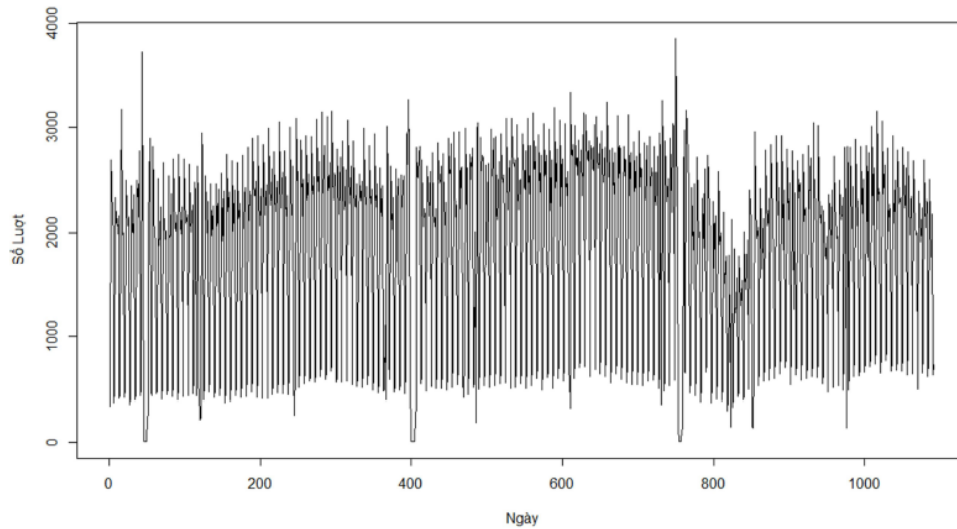
III. MÔ TẢ VÀ PHÂN TÍCH BỘ DỮ LIỆU

Dữ liệu được dùng trong nghiên cứu này là từ Bệnh viện đa khoa khu vực Củ Chi, Thành phố Hồ Chí Minh thu thập hàng ngày từ ngày 01 tháng 1 năm 2018 đến 31 tháng 12 năm 2020 với tổng số 75511 bệnh nhân. Như vậy có tất cả 1092 giá trị quan sát trên chuỗi thời gian. Giá trị trung bình của lượng bệnh nhân cấp cứu hàng ngày là 1843, giá trị tối thiểu là 4 và giá trị tối đa là 3850. Hình 3 minh họa đường biểu diễn của lượng bệnh nhân cấp cứu hàng ngày trong 3 năm từ năm 2018 đến năm 2020. Hình 4 minh họa đường biểu diễn của lượng bệnh nhân cấp cứu hàng ngày trong một năm (năm 2018).

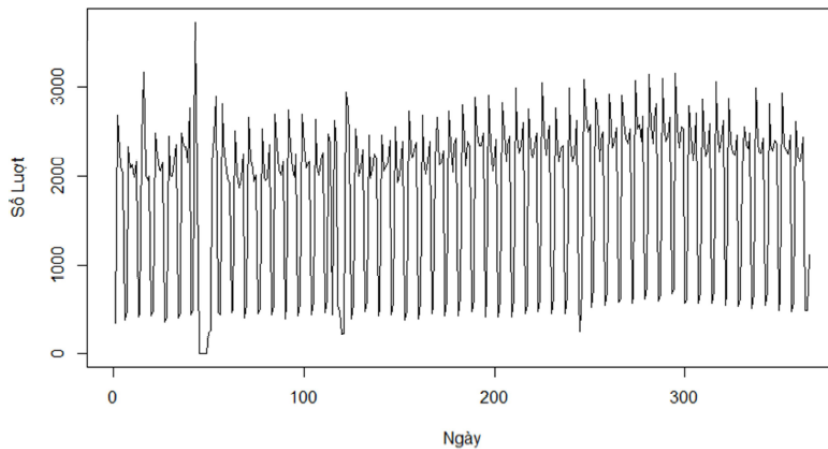
Bệnh viện đa khoa khu vực Củ Chi là bệnh viện hạng 2 trực thuộc Sở Y tế Thành Phố Hồ Chí Minh. Huyện Củ Chi nằm ở vị trí cửa ngõ, tiếp giáp với các tỉnh Tây Ninh, Long An và Bình Dương và là nơi có nhiều khu công nghiệp với qui mô công nhân lớn.

Từ Hình 3 và Hình 4, bằng trực quan, chúng ta có thể cảm nhận được tính không dừng (nonstationary) của chuỗi dữ liệu. Một chuỗi thời gian có tính không dừng khi giá trị trung bình và độ lệch chuẩn không cố định mà biến thiên khi thời gian thay đổi.

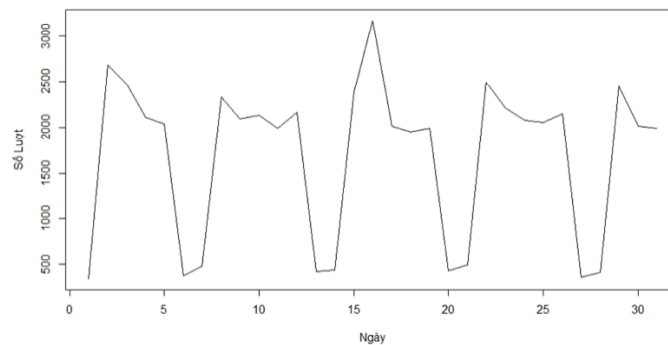
Hình 5 trình bày đường biểu diễn lượng bệnh nhân ngoại trú hàng ngày trong một tháng từ ngày 01 tháng 1 năm 2018 đến ngày 30 tháng 1 năm 2018. Từ Hình 4 và Hình 5 chúng ta có thể thấy lượng bệnh nhân ngoại trú hàng ngày biến thiên rất lớn, có khác biệt nhiều vào dịp cuối tuần và những ngày nghỉ lễ. Ngoài ra, sự biến thiên có tính mùa được nhận diện với thời đoạn là một tuần lễ.



Hình 3. Đường biểu diễn lượng bệnh nhân ngoại trú hàng ngày từ năm 2018 đến năm 2020.

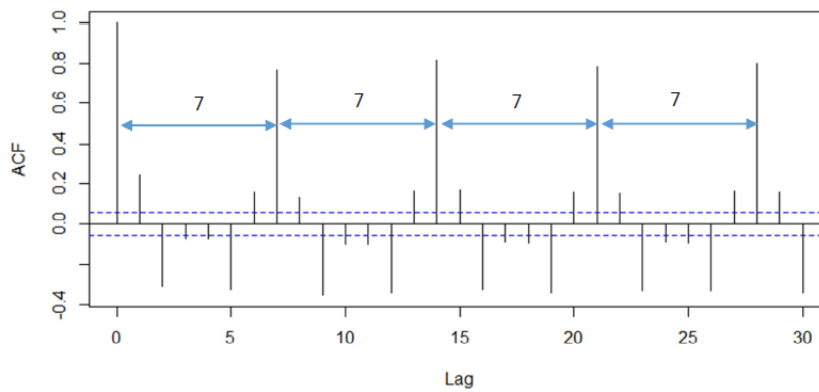


Hình 4. Đường biểu diễn lượng bệnh nhân ngoại trú hàng ngày trong năm 2018.



Hình 5. Đường biểu diễn lượng bệnh nhân ngoại trú hàng ngày trong một tháng (tháng 1- 2018)

Để phát hiện tính mùa của bộ dữ liệu này, chúng tôi áp dụng phương pháp đã được mô tả ở tiểu mục II.D và đã phát hiện chuỗi thời gian này có tính mùa với độ dài mùa là 7. Biểu đồ ACF mà trên đó chúng tôi xác định được độ dài mùa cũng được trình bày ở Hình 6. Chúng tôi xây dựng được biểu đồ ACF cho chuỗi thời gian này với sự hỗ trợ của phần mềm R.



Hình 6. Độ dài mùa của bộ dữ liệu và biểu đồ ACF tương ứng

IV. KẾT QUẢ THỰC NGHIỆM

Để so sánh bằng thực nghiệm tính hiệu quả dự báo của ba phương pháp đối sánh: mạng nơ ron ANN, mô hình Holt-Winters và mạng LSTM, chúng tôi hiện thực ba phương pháp này và thực nghiệm chúng trên bộ dữ liệu lượng bệnh nhân ngoại trú hàng ngày tại Bệnh viện đa khoa khu vực Củ Chi.

Chúng tôi hiện thực hai mô hình Holt-Winters và mạng nơ ron ANN với sự hỗ trợ của thư viện phần mềm R. R là một ngôn ngữ và môi trường lập trình cho những tính toán xác xuất, đồ họa và cung cấp nhiều công cụ phân tích dữ liệu chuỗi thời gian [17]. Chúng tôi hiện thực mô hình LSTM với sự hỗ trợ của thư viện phần mềm Keras [18]. Khi hiện thực hệ thống dự báo, chúng tôi sử dụng hai ngôn ngữ lập trình C#, Python và tiến hành thực nghiệm trên máy tính có cấu hình Intel(R) Core™ i7-1081 CPU@ 1.61GHz RAM32GB PC.

Ngoài ra, chỉ *dự báo một bước* (one-step ahead forecasting), tức là dự báo ngắn hạn, được thực hiện trong nghiên cứu này.

A. XÁC ĐỊNH THÔNG SỐ CHO CÁC MÔ HÌNH

Với sự hỗ trợ của phần mềm R và phương pháp *tìm kiếm lưới* (grid search) [19], chúng tôi xác định được ba thông số α , β và γ cho mô hình Holt-Winters đối với dữ liệu lượng bệnh nhân ngoại trú hàng ngày tại bệnh viện. Giá trị thấp của α ($=0.1$) cho thấy bộ dữ liệu có tính ngẫu nhiên (randomness) khá cao, giá trị thấp của β ($=0.1$) cho thấy xu hướng (tăng hay giảm) của bộ dữ liệu không đáng kể và giá trị trung bình của γ ($=0.3$) cho thấy bộ dữ liệu này thực sự có tính mùa. Với cả hai mô hình Holt-Winters và ANN, chiều dài mùa được gán giá trị 7 như đã nêu ở Mục III. Số đơn vị ở tầng nhập của mạng ANN cũng được gán giá trị 7. Số đơn vị ở tầng ẩn của ANN được xác định thông qua nhiều lần thực nghiệm. Để huấn luyện mạng ANN, chúng tôi sử dụng một phiên bản cải tiến của giải thuật Lan truyền ngược có tên là RPROP (**R**esilient back**P**ropagation).

Với mô hình LSTM, qua nhiều lần thực nghiệm trên bộ dữ liệu lượng bệnh nhân ngoại trú, chúng tôi xác định được các siêu tham số của mô hình học sâu này như sau. Số đơn vị của tầng nhập là 7. Số tầng ẩn là 2. Số đơn vị ở tầng ẩn thứ nhất là 128 và số đơn vị ở tầng ẩn thứ hai là 64. Các giá trị thông số thích hợp nhất cho cả ba mô hình được trình bày ở Bảng 2.

Bảng 2. Các thông số cho ba mô hình dự báo

Phương pháp	Các giá trị thông số
LSTM	Số đơn vị ở tầng ẩn thứ nhất = 128 Số đơn vị ở tầng ẩn thứ hai = 64
Holt-Winters	Mô hình nhân $\alpha=0.1, \beta=0.1, \gamma=0.3$ Chiều dài mùa = 7
ANN	Số đơn vị ở tầng nhập = 7 Số đơn vị ở tầng ẩn = 7 Số epoch tối đa = 1000 Optimizer = RPROP

Chuỗi thời gian lượng bệnh nhân ngoại trú hàng ngày tại bệnh viện được chia làm hai phần: phần huấn luyện và phần kiểm thử. Chúng tôi sử dụng 80% điểm dữ liệu đầu tiên trong chuỗi thời gian cho phần huấn luyện (xây dựng mô hình) và 20% còn lại cho phần kiểm thử. Các mô hình dự báo được kiểm thử trên phần kiểm thử và tính ra các độ đo về hiệu quả dự báo.

B. TIÊU CHÍ ĐÁNH GIÁ ĐỘ CHÍNH XÁC DỰ BÁO

Trong nghiên cứu này, sai số tuyệt đối trung bình (mean absolute error-MAE), sai số bình phương trung bình (mean squared error-MSE) và sai số phần trăm tuyệt đối trung bình (mean absolute percentage error - MAPE) được dùng làm tiêu chí đánh giá. Các công thức tính toán của các sai số dự báo MAE, MSE, MAPE được định nghĩa như sau:

$$MAE = \frac{1}{n} \sum_{t=1}^n |\hat{y}_t - y_t| \quad (11)$$

$$MSE = \frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2 \quad (12)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|\hat{y}_t - y_t|}{y_t} \times 100 \quad (13)$$

với n là số điểm dữ liệu của chuỗi thời gian, y_t là giá trị thực sự tại thời điểm t và $\hat{y}_t$ là giá trị dự báo tại thời điểm t .

Ba độ đo nêu trên biểu diễn những góc nhìn khác nhau của các phương pháp dự báo. Độ đo MAE và độ đo MSE là những độ đo tuyệt đối (absolute performance metric) trong khi độ đo MAPE là độ đo tương đối (relative performance metric). MAPE là một giá trị thống kê bất biến đối với biên độ của chuỗi thời gian mà biểu diễn sai số như là một trị phần trăm (percentage). Mô hình sẽ đạt độ chính xác dự báo càng cao khi MAE hay MSE tiến về 0 và MAPE càng tiến gần đến 0%.

C. KẾT QUẢ THỰC NGHIỆM

Kết quả thực nghiệm được trình bày ở Bảng 3. Các sai số dự báo MAE, MSE, MAPE (theo các công thức (11), (12), (13)) của ba mô hình dự báo được ghi ở cột 2, cột 3 và cột 4 của Bảng 3.

Bảng 3. Kết quả sai số dự báo của ba phương pháp

Phương pháp	MAE	MSE	MAPE
LSTM	210.80	66180.35	11.24%
Holt-Winters	220.96	69080.73	12.25%
ANN	730.49	911492.9	76.43%

Từ kết quả thực nghiệm của Bảng 3, chúng ta có thể thấy các sai số dự báo MAE, MSE, MAPE của LSTM và mô hình Holt-Winters đều thấp hơn nhiều so với mô hình ANN. Hiệu quả dự báo của LSTM tốt hơn mô hình Holt-Winters một chút và cả hai mô hình LSTM và Holt-Winters đều tốt hơn khá nhiều so với mô hình ANN cho việc dự báo lượng bệnh nhân ngoại trú hàng ngày tại bệnh viện. Lý do chính của điều này là cả hai mô hình LSTM và Holt-Winter có khả năng nắm bắt tính xu hướng và tính mùa của dữ liệu chuỗi thời gian tốt hơn nhiều so với mạng nơ ron ANN và do đó cải tiến được hiệu quả dự báo. Phát hiện này nhấn mạnh tính hữu ích của việc sử dụng mô hình LSTM và mô hình Holt-Winters để dự báo lượng bệnh nhân ngoại trú tại bệnh viện.

Mô hình dự báo LSTM chưa thể hiện được tính vượt trội so với mô hình truyền thống Holt-Winters là vì bộ dữ liệu chuỗi thời gian dự báo không thuộc loại quá phức tạp và không có kích thước thật lớn nên những ưu thế vốn có của một mạng nơ ron học sâu này chưa thể phát huy được.

Ngoài ra, kết luận thực nghiệm rút ra từ trong nghiên cứu này rằng mô hình Holt-Winters tốt hơn mô hình ANN cũng có vẻ tương thích với kết quả của một công trình nghiên cứu đi trước (Sumitra và Barsi, 2020 [4]) ghi nhận rằng Holt-Winters dự báo tốt hơn mô hình ARIMA và mô hình Làm trơn hàm mũ đơn giản (SES).

V. KẾT LUẬN

Nghiên cứu này khảo sát bằng thực nghiệm hiệu quả dự báo của ba phương pháp dự báo tiêu biểu: Holt-Winters, ANN và LSTM đối với bài toán dự báo lượng bệnh nhân ngoại trú tại bệnh viện. Nghiên cứu này cho thấy lượng bệnh nhân ngoại trú hằng ngày được đặc trưng hóa bằng những biến thiên có tính mùa và mẫu biến thiên hằng tuần. Độ chính xác dự báo tùy thuộc vào đặc điểm của bộ dữ liệu và mô hình dự báo được sử dụng. Trong phần thực nghiệm với bộ dữ liệu lượng bệnh nhân cấp cứu tại Bệnh viện đa khoa khu vực Củ Chi, chúng tôi đã đánh giá so sánh hiệu quả của ba mô hình dự báo: Holt-Winters, LSTM và mạng nơ ron ANN và rút ra một kết luận quan trọng: mô hình Holt Winters và LSTM đem lại chất lượng dự báo tốt hơn nhiều so với mô hình mạng nơ ron ANN. Tiêu chí sai số dự báo MAPE của mô hình LSTM đạt được một kết quả đầy ấn tượng: 11.24%.

Tuy nhiên, nghiên cứu này chỉ mới thực hiện việc dự báo một bước (one step ahead forecasting) tức là tầm dự báo chỉ là một ngày, như vậy những tầm dự báo khác cũng cần được thực nghiệm để xác nhận tính hiệu quả của ba mô hình dự báo nêu trên với những tầm dự báo khác nhau này [20]. Ngoài ra, chúng tôi dự định sử dụng một mô hình lai ghép kết hợp một kỹ thuật phân rã chuỗi thời gian, thí dụ như kỹ thuật EMD (Empirical Mode Decomposition), với một tổ hợp các bộ dự báo cục bộ (ensemble of local predictors) để cải tiến hơn nữa việc dự báo lượng bệnh nhân ngoại trú tại bệnh viện ([7], [8], [21]).

VI. LỜI CẢM ƠN

Các tác giả của nghiên cứu này chân thành cảm ơn Bệnh viện đa khoa khu vực Củ Chi, Tp. Hồ Chí Minh đã cho phép sử dụng bộ dữ liệu lượng bệnh nhân ngoại trú của bệnh viện.

VII. TÀI LIỆU THAM KHẢO

- [1] L. Lou, L. Lou, X. Zhang, X. He (2017). Hospital daily outpatient visits forecasting using a combinatory model based on ARIM and SES models. *BMC Health Services Research* **17**:469.
- [2] Y. Li, F. Wu, C. Zheng, K. Hou, K. Wang, N. Sun (2014). Predictive analysis of outpatient visits to a grade 3, class a hospital using ARIMA model. In: *Proceedings of the 2014 International Symposium on Information Technology (ISIT 2014)*. Dalian: CRC Press, p 285.
- [3] K.R. Kim, J. E. Park, I. T. Jang (2020). Outpatient forecasting model in spine hospital using ARIMA and SARIMA methods. *Journal of Hospital Management and Health Policy*, **4**:20.
- [4] I. D. Sumitra and I. Basri (2020). Forecasting the Number of Outpatient Patient Visits Using the ARIMA, SES And Holt-Winters Methods at XYZ Community Health Center. *IOP Conf. Ser.: Mater. Sci. Eng.* **879**:012060, 2020.
- [5] G. Guan, B. E. Engelhardt (2019). Predicting Sick Patient Volume in a Pediatric Outpatient Setting using Time Series Analysis. *Machine Learning for Healthcare* **106**:1-16.
- [6] K. Thapa, A. K. Timalisina (2023) Hospital Out Patient Visit Forecasting using Gated Recurrent Unit. In: *Proc. of 6th Int. Conf. on Information Systems and Computer Networks (ISCON)*, Mathura, India, 03-04 March.
- [7] Y. Wang, J. Gu, Z. Zhou, Z. Wang (2015). Diarrhoea outpatient visits prediction based on time series decomposition and multi-local predictor fusion. *Knowledge Base Systems* **88**, pp. 12-23.
- [8] D. Huang, Z. Wu (2017). Forecasting outpatient visits using empirical mode decomposition coupled with backpropagation artificial neural networks optimized by particle swarm optimization. *PLOS ONE* **7**, February 21.
- [9] L. Yu, G. Hang, L. Tang, Y. Zhao, K.K. Lai (2017). Forecasting Patient Visits to Hospitals using a WD&ANN-based Decomposition and Ensemble Model. *Eurasia Journal of Mathematics, Science and Technology Education*, **13**:12, pp. 7615-7627.
- [10] S. Jiang, R. Xiao, L. Wang, X. Luo, C. Huang, J. H. Wang, K. S. Chin, and X. Nie (2019). Combining Deep Neural Networks and Classical Time Series Regression Models for Forecasting Patient Flows in Hong Kong. *IEEE ACCESS* 2019. 2936550.
- [11] Y. Deng, H. Fan, S. Wu (2020). A hybrid ARIMA-LSTM model optimized by BP in the forecast of outpatient visits. *Journal of Ambient Intelligence and Humanized Computing* **14**, pp. 5517-5527.
- [12] J. E. Hanke (2014). W. Wichenrn. *Business Forecasting- 9th Edition*, Pearson New International Edition, 510 ps.
- [13] P. Goodwin (2010). The Holt-Winters Approach to Exponential Smoothing: 50 Years Old and Going Strong, *Foresight*, pp. 30-33.
- [14] C. Hamzaçebi (2008). Improving artificial neural networks' performance in seasonal time series forecasting. *Information Sciences*, **178**: 4550-4559.

- [15] S. Hochreiter, S. and J. Schmidhuber (1997). Long Short-Term Memory. *Neural Computation* **9**: 8, pp. 1735-1780.
- [16] N. Ngoc Tran (2002). *Automatic ARIMA time series modeling and forecasting for adaptive Input/Output prefetching*. Ph.D. Thesis, Department of Computer Science, University of Illinois, Urbana, Illinois.
- [17] The R project for Statistical Computing at <http://www.r-project.org/>. Accessed date: Sep. 10 – 2021.
- [18] F. Cholett, Keras, [Online]. Available: <http://keras.io>. Assessed date: Dec. 20-2024.
- [19] N, Buslim, I. L. Rahmatullah, B. A. Setyawan (2021). Comparing Bitcoin's Prediction Model Using GRU, RNN, and LSTM by Hyperparameter Optimization Grid Search and Random Search. In: *Proceedings 9th International Conference on Cyber and IT Service Management (CITSM 2021)*, Bengkulu, Indonesia, September 22-23.
- [20] S. Ben Taieb, G. Bontempi, A. F. Atiya, and A. Sorjamaa (2012). A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Syst. Appl.* **39**: 8, pp. 7067-7083.
- [21] T. Buengbongkoch, N. Nupairoj (2024). Prediction of Daily Outpatient Volume Using Ensemble Empirical Mode Decomposition and Long-Short Term Memory During the Incident. In: *Proc. of 21st Int. Joint conf. on Computer Science and Software Engineering*, 19-22 June, Phuket Thailand.

FORECASTING HOSPITAL OUTPATIENT VISITS FLOW USING HOLT-WINTERS, ANN AND LSTM

Duong Tuan Anh, Lý Thanh Tịnh

ABSTRACT— Effective hospital outpatient forecasting is an important task for modern hospitals to implement intelligent management of medical resources. As outpatient visit flow may be nonlinear, seasonal and dynamic, we empirically compare three typical forecasting methods: Holt-Winters, artificial neural network (ANN) and Long-Short-Term Memory (LSTM), to identify the most effective method in this challenging forecasting problem. Holt-Winters belongs to statistical analysis category while ANN is a machine learning method and LSTM is a deep learning method. We applied the three methods to forecast daily outpatient visit flow at the General Hospital of Cu Chi Area in Ho Chi Minh City. Experimental results reveal that Holt-Winters and LSTM outperform ANN in the hospital outpatient visits flow forecasting. The MAPE indicator of LSTM is 11.24%.

Keywords— outpatient visits flow, Holt-Winters model, artificial neural network (ANN), LSTM, seasonal time series.



Dương Tuấn Anh tốt nghiệp tiến sĩ ngành khoa học máy tính tại Học viện Công nghệ Á Châu (Asian Institute of Technology), Bangkok, Thái Lan, năm 1998 và đó cũng là nơi ông tốt nghiệp thạc sĩ với cùng chuyên ngành. Ông đã là Phó giáo sư khoa Khoa Học và Kỹ Thuật Máy tính, Trường Đại học Bách Khoa, ĐHQG Tp. Hồ Chí Minh từ năm

2007. Từ năm 2020 đến nay, ông là giảng viên Khoa Công nghệ thông tin, trường Đại học Ngoại ngữ-Tin học Tp. Hồ Chí Minh. Lĩnh vực nghiên cứu chính của ông là metaheuristics, học máy và khai phá dữ liệu chuỗi thời gian. Ông là đồng tác giả của trên 120 bài báo khoa học.



Lý Thanh Tịnh tốt nghiệp cử nhân Hệ thống thông tin kinh tế tại Trường Đại học Marketing Tp. Hồ Chí Minh năm 2008 và tốt nghiệp thạc sĩ ngành Công nghệ thông tin tại Trường Đại học Ngoại ngữ-Tin học Tp. Hồ Chí Minh năm 2022. Anh là chuyên viên công tác tại Phòng Công Nghệ Thông Tin của Bệnh Viện Đa Khoa khu vực Củ Chi

từ 2008 đến nay. Lĩnh vực nghiên cứu chính của anh là khai phá dữ liệu chuỗi thời gian.