Comparative Study of Informer, Prophet, and SARIMA Time Series Forecasting Models for Predicting Pneumonia-Related Hospitalizations and Emergency Room Visits in Elderly Patients Using OMOP-CDM

Seonghwan Shin, PharmD[1], Junhyuk Chang, PharmD[1],
Min-Gyu Kim, MD[2], Byungjin Choi, MD[2], Rae Woong Park, MD, Ph.D.[1,2]
[1]Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, South Korea
[2]Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea

Background

Pneumonia is an infection of the lung due to bacterial, viral, or fungal and a significant cause of mortality worldwide and prevalent in elderly patients. Pneumonia patients aged over 65 often exhibits fewer symptoms, making diagnosis and timely treatment challenging, which can lead to many deaths.[1, 2] Also, respiratory infections show seasonal variations and outbreak tendencies. Therefore, it is essential to predict the number of pneumonia hospitalizations and ER visits among the elderly.

Effective forecasting of these events can be achieved through time series forecasting, which specifically analyzes temporal sequences to predict future values. Specifically, traditional statistical models such as univariate Seasonal Autoregressive Integrated Moving Average (SARIMA), have been used to predict the next "lag" of time-series data. Prophet, developed by Facebook, is an additive model designed for ease of use with minimal parameter tuning.[3,4] Moreover, Informer leverages self-attention mechanisms to capture long-range dependencies in time series data efficiently, offering superior accuracy in long-term forecasting. Despite the promising capabilities of these models, there is a lack of comprehensive studies comparing their performances in forecasting pneumonia hospitalizations and ER visits in the elderly.

The Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM) is a standardized and structured database used across multiple institutions. This model allows researchers to easily extract patient information that meets their specific criteria. This capability is particularly efficient for extracting and preprocessing input data for time series forecasting studies. Therefore, this study aims to compare forecasting models for elderly pneumonia hospitalizations using these three models using OMOP-CDM.

Methods

This study utilized the OMOP-CDM database of Ajou University Hospital. Data on patients aged over 65 who were hospitalized or visited the emergency room due to pneumonia were extracted from the OMOP-CDM database for the years 2018 to 2023. The dataset was then split into training and testing sets with an 80:20 ratio. We developed three time series forecasting models, Informer, SARIMA, and Prophet using the daily patient count data.

Informer model is a deep learning model based on transformer architecture, designed for long-term time series forecasting. It uses a ProbSparse self-attention mechanism to handle long-range dependencies efficiently. The training process involves preprocessing the data into input sequences, using a stacked encoder-decoder architecture with self-attention layers and sliding window. The optimization of parameters is performed using gradient descent. SARIMA model extends ARIMA by incorporating seasonality. It models the time series using autoregressive terms, differencing, and moving averages, along with seasonal components. The training process involves selecting the appropriate parameters through techniques like ACF/PACF plots and optimizing them using maximum likelihood estimation. Prophet model decomposes time series into trend, seasonality, and holiday effects. The training process

involves fitting piecewise linear or logistic growth curves for the trend, Fourier series for seasonality, and incorporating user-defined holidays. We set the same forecast horizon of 2-weeks for each model.

The objective was to compare the performances of the three developed models, Informer, SARIMA, and Prophet, each with a forecast horizon of 2 weeks, in predicting daily patient counts of pneumonia hospitalizations and ER visits among the elderly. Model performance was evaluated using mean absolute error (MAE), and root mean square error (RMSE). All analyses were conducted using Python v3.12.

Results

The data extracted for training and testing the models consists of daily patient counts over a span of 1,825 days. The Informer model performed the best in predicting the number of hospitalizations over the 2-week forecast horizon, achieving an RMSE of 1.089 and an MAE of 0.778. This was followed by the SARIMA model, while the Prophet model showed the worst performance (**Figure 1, Table 1**).

Table 1. Performance metrics of the models

Models	MAE	RMSE
Informer	**0.778**	**1.089**
SARIMA	2.227	2.595
Prophet	4.489	4.776

***Note: Bold values indicate the best performance for each metric.**

Conclusion

We compared the performance of Informer, SARIMA, and Prophet models in predicting daily hospitalizations and emergency room visits for pneumonia among the elderly. The Informer model demonstrated superior predictive capabilities with lower MAE and RMSE values. In an environment where there is a lack of time series forecasting research on pneumonia patient counts using deep learning models, these findings highlight the potential of deep learning models in forecasting the daily pneumonia patients for the elderly. To reflect real clinical settings and improve the accuracy of model predictions, further research is needed on the application of time series forecasting models utilizing multivariate input data, which can be facilitated by utilizing OMOP-CDM.

References

1. Tichopad A, Roberts C, Gembula I, Hajek P, Skoczynska A, Hryniewicz W, et al. Clinical and Economic Burden of Community-Acquired Pneumonia among Adults in the Czech Republic, Hungary, Poland and Slovakia. PLoS ONE. 2013;8(8)
2. Yoshikawa TT, Marrie TJ. Community-Acquired Pneumonia in the Elderly. Clin Infect Dis. 2000;31(4):1066-1078.
3. Luo Z, Jia X, Bao J, Song Z, Zhu H, Liu M, et al. A Combined Model of SARIMA and Prophet Models in Forecasting AIDS Incidence in Henan Province, China. Int J Environ Res Public Health. 2022;19(10):5910.

Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, et al. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. Proceedings of the AAAI Conference on Artificial Intelligence. 2021;35(12):11106-11115.

Figure 1.(a) Daily count forecast using the Informer model

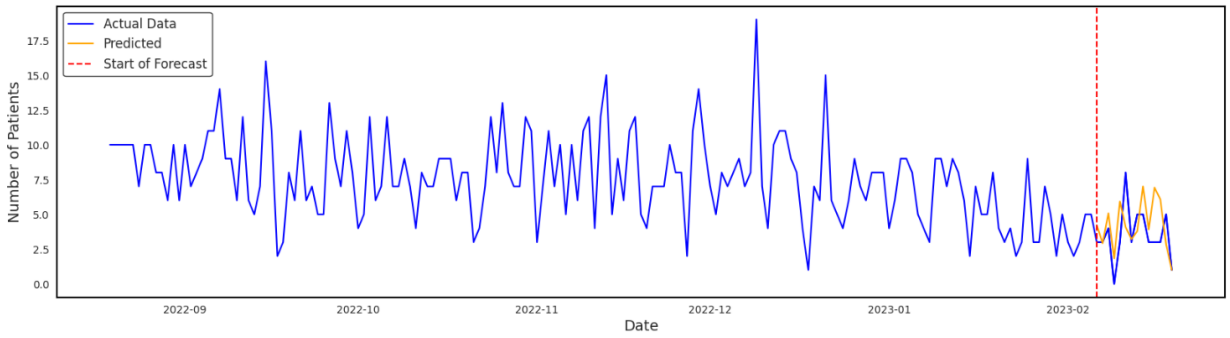

Figure 1.(b) Daily count forecast using the SARIMA model

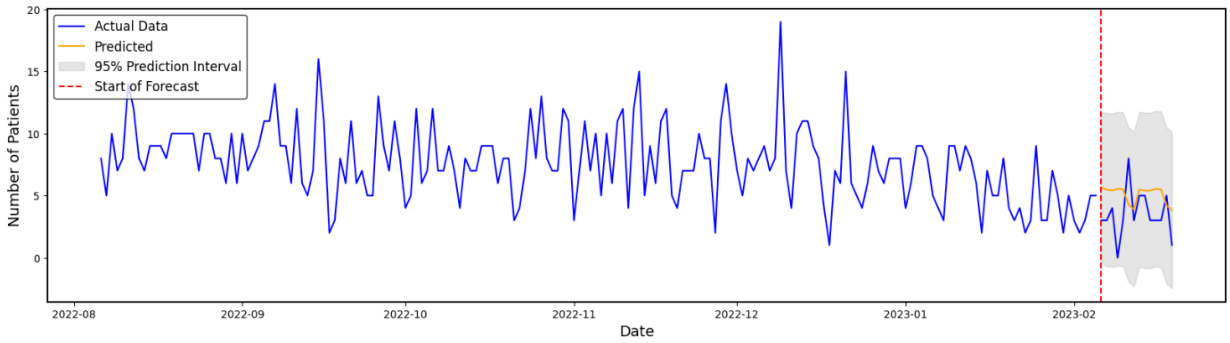

Figure 1.(c) Daily count forecast using the Prophet model

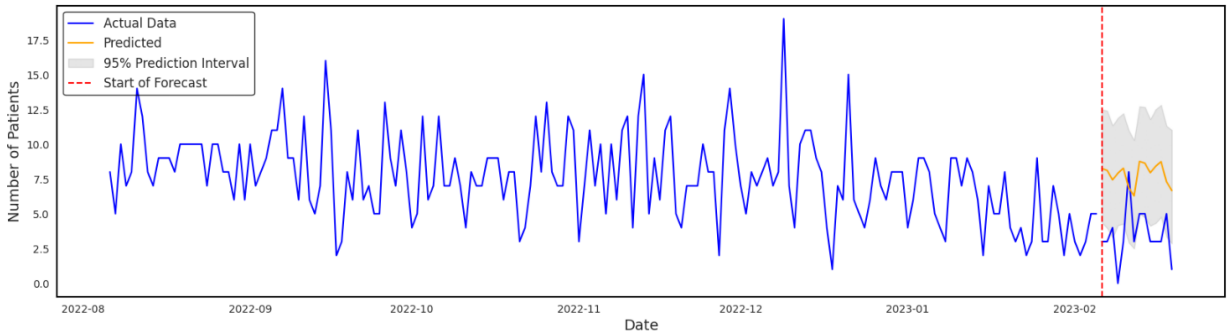

Figure 1. Daily count forecast using models