Interpreting Forecasted Vital Signs Using N-BEATS in Sepsis Patients[*]

Anubhav Bhatti[1,*], Naveen Thangavelu[2], Marium Hassan[1], Choongmin Kim[3], San Lee[1,*], Yonghwan Kim[3] and Jang Yong Kim[4]

[1]*SpassMed Inc., Ontario, Canada*

[2]*University of Toronto, Ontario, Canada*

[3]*Spass Inc., Seoul, South Korea*

[4]*St. Mary's Hospital, Seoul, South Korea*

Abstract

Detecting and predicting septic shock early is crucial for the best possible outcome for patients. Accurately forecasting the vital signs of patients with sepsis provides valuable insights to clinicians for timely interventions, such as administering stabilizing drugs or optimizing infusion strategies. Our research examines N-BEATS, an interpretable deep-learning forecasting model that can forecast 3 hours of vital signs for sepsis patients in intensive care units (ICUs). In this work, we use the N-BEATS interpretable configuration to forecast the vital sign trends and compare them with the actual trend to understand better the patient's changing condition and the effects of infused drugs on their vital signs. We evaluate our approach using the publicly available eICU Collaborative Research Database dataset and rigorously evaluate the vital sign forecasts using out-of-sample evaluation criteria. We present the performance of our model using error metrics, including mean squared error (MSE), mean average percentage error (MAPE), and dynamic time warping (DTW), where the best scores achieved are 18.52e-4, 7.60, and 17.63e-3, respectively. We analyze the samples where the forecasted trend does not match the actual trend and study the impact of infused drugs on changing the actual vital signs compared to the forecasted trend. Additionally, we examined the mortality rates of patients where the actual trend and the forecasted trend did not match. We observed that the mortality rate was higher (92%) when the actual and forecasted trends closely matched, compared to when they were not similar (84%).

Keywords

Time Series Forecasting, Septic Shock Detection, Interpretable Forecasting, Explainable AI

CEUR Workshop Proceedings (CEUR-WS.org)

1. Introduction

Septic shock, the most severe form of sepsis, is characterized by profound circulatory and cellular abnormalities and is associated with a high mortality rate [1, 2]. Early detection and treatment of sepsis and septic shock are critical for improving patient outcomes, as delays in intervention can lead to a rapid decline in a patient's condition [3]. While various methods have shown promise in predictive accuracy [4, 5], their interpretability remains a significant concern. Understanding and explaining a model's underlying mechanisms is crucial for gaining trust [6, 7, 8], facilitating model debugging, and informing clinical decision-making [9, 10, 11]. Interpretability and explainability methods for time series forecasting models can provide valuable insights into the relationships between vital signs and the risk of septic shock [12, 13, 14, 15]. Several studies have been undertaken to incorporate model-agnostic explainability by leveraging rule-based [6, 7] methods and argumentation [16, 17, 18, 8, 19].

This paper explores the interpretability of vital sign forecasting models for patients with Sepsis and septic shock condition in critical care settings. Based on our knowledge, this work is one of the first to explore deep learning models to forecast the vital signs in the eICU dataset [20]. Further, we investigate the interpretability and explainability of patients' forecasted signals in conjunction with drug infusion. Our goal is to contribute to developing more interpretable and trustworthy models for septic shock prediction, ultimately improving patient outcomes.

2. Methodology

2.1. Method

In this paper, we utilize the N-BEATS [21] interpretable architecture to forecast the seasonality and trend components. By accurately forecasting the trend component, we aim to provide valuable support to clinicians in monitoring the vital signs trend of patients and making informed decisions regarding medication administration. The interpretable configuration of N-BEATS consists of two distinct stacks that enable the direct incorporation of trend and seasonality decomposition within the model's architecture. This integration allows for a clearer interpretation of the stack outputs, as the model explicitly captures and represents the trend and seasonality components of the time series. As described in Equation 1 [21], the trend model is constrained to have a polynomial of small degree p, a function slowly varying across the forecast window.

$$\widehat{\mathbf{y}} = \sum_{i=0}^{p} \theta_{tr,i} t^i \tag{1}$$

Here time vector $\mathbf{t} = [0, 1, 2, \ldots, H-2, H-1]^T / H$ is defined on a discrete grid running from 0 to $(H-1)/H$, forecasting H steps ahead and θ_{tr} are polynomial coefficients predicted by a fully connected network. Similarly, the seasonality model, which captures the regular, cyclic and recurring fluctuation, is constrained with the Fourier basis as shown in Equation 2 [21], where θ_{sea} are Fourier coefficients predicted by a fully connected network.

$$\widehat{\mathbf{y}} = \sum_{i=0}^{\lfloor H/2-1 \rfloor} \theta_{sea,i} \cos(2\pi it) + \theta_{sea,i+\lfloor H/2 \rfloor} \sin(2\pi it), \tag{2}$$

2.2. Dataset

To evaluate our approach using N-BEATS, we conducted experiments on the publicly available eICU Collaborative Research Database [20]. The dataset includes information on 139,367 unique patients (53.96% male, 45.95% female, 0.09% other/unknown). Our experiments focused on forecasting the mean blood pressure (MBP) of the patients diagnosed with sepsis or septic shock. We perform data cleaning and preprocessing on our samples for the deep learning model. Missing values in the MBP are imputed using fill forward technique. We extract 9 hours of MBP before diagnosing sepsis or septic shock and split them into a 6-hour lookback horizon and a 3-hour forecasting horizon. Moving average smoothing with a window length of 3 is applied to remove noise. We remove samples with a standard deviation of ≤ 0.025, as they exhibit minimal variability. Additionally, we apply min-max scaling in the range of 0 to 190 to normalize the samples [22]. After preprocessing, we obtain 4020 samples from 1442 patients.

3. Results and Interpretability

We assess the effectiveness of N-BEATS Generic and Interpretable configurations by analyzing their performance using three error metrics: Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and Dynamic Time Warping (DTW). To establish a baseline, we compare the forecasting outcomes with a simple persistence model. This model predicts future values of the time series by replicating the last observed value within the lookback horizon. The performance of the models on the test set are presented in Table 1.

Table 1
Performance of N-BEATS with Generic and Interpretable Configurations.

Model Configuration	MSE (1e-4)	MAPE	DTW (1e-3)
Presistence	24.55	8.47	34.50
N-BEATS - Generic	**18.52**	**7.60**	**17.63**
N-BEATS - Interpretable	21.49	8.60	18.20

Additionally, we investigate the extracted trend of the forecasts obtained using the N-BEATS interpretable configuration and analyze cases where the forecasted trend does not align with the actual trend. To better understand this discrepancy, we calculate the DTW score between the actual and the forecasted trends. We select the top 25% of samples with the highest DTW scores, indicating a significant trend mismatch. We observe that, in several samples, a noticeable deviation between actual and forecasted trends occurs when drugs are administered after the training cut-off, as seen with patient ID 261982, who experienced an increased MBP trend due to the administration of vasoactive drugs like norepinephrine and vasopressin.
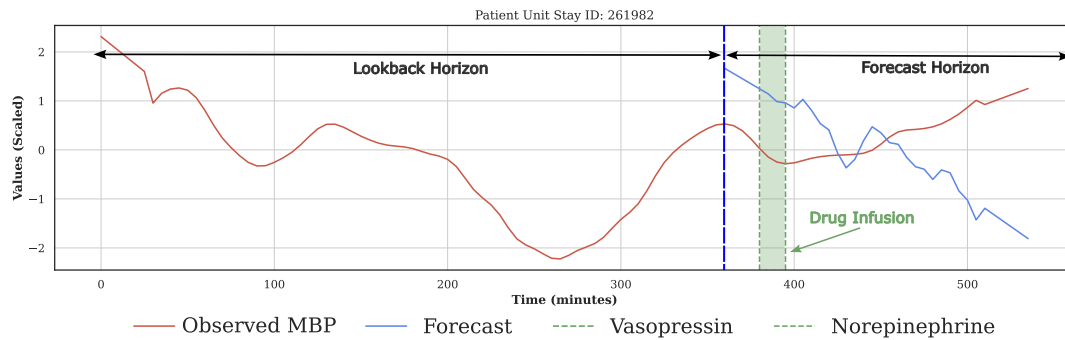
Figure 1: The observed and forecasted trends of mean blood pressure (MBP) for a sample drawn from the subset with the highest DTW scores (top 25%).

Since the drug infusion information was an unobserved variable during the training of the forecasting model, the discrepancy in the actual and the forecasted trend can be attributed to the fact that the model is trained on historical vital sign data, which does not include the effects of drugs introduced after the training cut-off. Figure 1 shows the observed MBP trend, the forecasted trend, the training cut-off, and the interval of drug infusion for a specific patient. Furthermore, our findings suggest that cases where the actual and forecasted trends matched had a higher mortality rate (92%) compared to cases where the trends were dissimilar (84%). Although not directly related, conducting further experiments could provide more insights into this aspect and explore its implications more extensively.

Exceptions to observed patterns can occur when drugs are administered before the specified cut-off, possibly due to drug-to-drug interactions. Conducting additional studies is crucial for understanding the causal inferences behind these interactions and better comprehending the relationship between medication administration, vital signs, and their subsequent effects.

4. Conclusion

In conclusion, our study utilized the eICU dataset to evaluate the forecasting performance of the interpretable N-BEATS model, highlighting the significance of accounting for drug infusion's influence on trends in ICU patients' vital signs. Future research will focus on developing approaches that integrate drug infusion information and investigate drug-to-drug interactions within the ICU context, aiming to enhance the overall performance of forecasting models in critical care. Addressing these challenges effectively can advance the understanding and application of forecasting methods, leading to improved patient care and better outcomes.

References

[1] A. Rhodes, L. E. Evans, W. Alhazzani, M. M. Levy, M. Antonelli, R. Ferrer, A. Kumar, J. E. Sevransky, C. L. Sprung, M. E. Nunnally, et al., Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016, Intensive care medicine 43 (2017) 304–377.

[2] M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J.-D. Chiche, C. M. Coopersmith, et al., The third international consensus definitions for sepsis and septic shock (sepsis-3), Jama 315 (2016) 801–810.

[3] A. Kumar, D. Roberts, K. E. Wood, B. Light, J. E. Parrillo, S. Sharma, R. Suppes, D. Feinstein, S. Zanotti, L. Taiberg, et al., Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock, Critical care medicine 34 (2006) 1589–1596.

[4] T. Desautels, J. Calvert, J. Hoffman, M. Jay, Y. Kerem, L. Shieh, D. Shimabukuro, U. Chettipally, M. D. Feldman, C. Barton, et al., Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach, JMIR medical informatics 4 (2016) e5909.

[5] M. Komorowski, L. A. Celi, O. Badawi, A. C. Gordon, A. A. Faisal, The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care, Nature medicine 24 (2018) 1716–1720.

[6] G. Vilone, L. Longo, A quantitative evaluation of global, rule-based explanations of post-hoc, model agnostic methods, Frontiers in artificial intelligence 4 (2021) 717899.

[7] G. Vilone, L. Rizzo, L. Longo, A comparative analysis of rule-based, model-agnostic methods for explainable artificial intelligence (2020).

[8] L. Rizzo, L. Longo, An empirical evaluation of the inferential capacity of defeasible argumentation, non-monotonic fuzzy reasoning and expert systems, Expert Systems with Applications 147 (2020) 113220.

[9] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al., Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai, Information fusion 58 (2020) 82–115.

[10] G. Vilone, L. Longo, Notions of explainability and evaluation approaches for explainable artificial intelligence, Information Fusion 76 (2021) 89–106.

[11] G. Vilone, L. Longo, A novel human-centred evaluation approach and an argument-based method for explainable artificial intelligence, in: Artificial Intelligence Applications and Innovations: 18th IFIP WG 12.5 International Conference, AIAI 2022, Hersonissos, Crete, Greece, June 17–20, 2022, Proceedings, Part I, Springer, 2022, pp. 447–460.

[12] M. T. Ribeiro, S. Singh, C. Guestrin, " why should i trust you?" explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135–1144.

[13] L. Rizzo, L. Majnaric, L. Longo, A comparative study of defeasible argumentation and non-monotonic fuzzy reasoning for elderly survival prediction using biomarkers, in: AI* IA 2018–Advances in Artificial Intelligence: XVIIth International Conference of the Italian Association for Artificial Intelligence, Trento, Italy, November 20–23, 2018, Proceedings

17, Springer, 2018, pp. 197–209.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).

[15] S. Wachter, B. Mittelstadt, C. Russell, Counterfactual explanations without opening the black box: Automated decisions and the gdpr, Harv. JL & Tech. 31 (2017) 841.

[16] L. Longo, L. Rizzo, P. Dondio, Examining the modelling capabilities of defeasible argumentation and non-monotonic fuzzy reasoning, Knowledge-Based Systems 211 (2021) 106514.

[17] L. Rizzo, L. Longo, A qualitative investigation of the degree of explainability of defeasible argumentation and non-monotonic fuzzy reasoning, in: Proceedings of the 26th AIAI Irish conference on artificial intelligence and cognitive science, 2018.

[18] L. Rizzo, L. Longo, Inferential models of mental workload with defeasible argumentation and non-monotonic fuzzy reasoning: a comparative study (2019).

[19] L. Rizzo, L. Longo, Comparing and extending the use of defeasible argumentation with quantitative data in real-world contexts, Information Fusion 89 (2023) 537–566.

[20] T. J. Pollard, A. E. Johnson, J. D. Raffa, L. A. Celi, R. G. Mark, O. Badawi, The eicu collaborative research database, a freely available multi-center database for critical care research, Scientific data 5 (2018) 1–13.

[21] B. N. Oreshkin, D. Carpov, N. Chapados, Y. Bengio, N-beats: Neural basis expansion analysis for interpretable time series forecasting, 2020. arXiv:1905.10437.

[22] H. M. O'Halloran, K. Kwong, R. A. Veldhoen, D. M. Maslove, Characterizing the patients, hospitals, and data quality of the eicu collaborative research database, Critical Care Medicine 48 (2020) 1737–1743.