2024 IEEE First International Conference on Artificial Intelligence for Medicine, Health and Care (AIMHC) | 979-8-3503-7198-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/AIMHC59811.2024.00018

Vital Sign Forecasting for Sepsis Patients in ICUs

Anubhav Bhatti[1], Yuwei Liu[1,2], Chen Dan[1,2], Bingjie Shen[1,2], San Lee[1], Yonghwan Kim[3], Jang Yong Kim[4]
[1]*AI Engineering Team, SpassMed Inc.,* [2]*University of Toronto*, Canada,
[3]*Spass Inc.,* [4]*St. Mary's Hospital*, South Korea
(anubhav.bhatti, yuwei.liu, chen.dan, bingjie.shen, sanlee)@spassmed.ca, kyh@spass.ai, vasculakim@catholic.ac.kr

Abstract—**Sepsis and septic shock are a critical medical condition affecting millions globally, with a substantial mortality rate. This paper uses state-of-the-art deep learning (DL) architectures to introduce a multi-step forecasting system to predict vital signs indicative of septic shock progression in Intensive Care Units (ICUs). Our approach utilizes a short window of historical vital sign data to forecast future physiological conditions. We introduce a DL-based vital sign forecasting system that predicts up to 3 hours of future vital signs from 6 hours of past data. We further adopt the DILATE loss function to capture better the shape and temporal dynamics of vital signs, which are critical for clinical decision-making. We compare three DL models, N-BEATS, N-HiTS, and Temporal Fusion Transformer (TFT), using the publicly available eICU Collaborative Research Database (eICU-CRD), highlighting their forecasting capabilities in a critical care setting. We evaluate the performance of our models using mean squared error (MSE) and dynamic time warping (DTW) metrics. Our findings show that while TFT excels in capturing overall trends, N-HiTS is superior in retaining short-term fluctuations within a predefined range. This paper demonstrates the potential of DL in transforming the monitoring systems in ICUs, potentially leading to significant improvements in patient care and outcomes by accurately forecasting vital signs to assist healthcare providers in detecting early signs of physiological instability and anticipating septic shock.**

Index Terms—**Time Series Forecasting, Deep Learning, N-HiTS, N-BEATS, Temporal Fusion Transformer, Dynamic Time Warping.**

I. Introduction

Sepsis is a severe medical condition that can significantly threaten one's life. It occurs when the body's immune system responds to an infection by releasing chemicals in the bloodstream, leading to inflammation and potential damage to tissues and organs [1]–[3]. In North America alone, around 1.7 million people are affected by sepsis yearly, and roughly 270,000 cases result in death [3], [4]. Globally, sepsis claims the lives of approximately 6 million out of the 30 million who develop the condition [5]. Monitoring vital signs is crucial in healthcare settings to detect early signs of physiological deterioration and take necessary actions to improve the patient outcomes [6]. Traditional scoring frameworks such as the Acute Physiology, Age, Chronic Health Evaluation (APACHE) [7], Simplified Acute Physiology Score (SAPS) [8], and Sequential Organ Failure Assessment (SOFA) [9], and qSOFA [10], are predicated upon physiological metrics to ascertain the severity of sepsis in patients. Nevertheless, these systems are not configured to detect sepsis or septic shock at an early stage, highlighting the urgent necessity to create more sophisticated monitoring methods that can enable swift medical action. Moreover, the need for clinical input to calculate SOFA and qSOFA scores makes them impractical for real-time detection systems, limiting their effectiveness in proactive sepsis management.

Recent studies have shown that machine learning and deep learning (DL) forecasting models have great potential in classification and forecasting time series data [11]–[15]. These models have demonstrated the ability to learn complex patterns and relationships within time series data and enable accurate predictions of future values. In recent years, several state-of-the-art DL forecasting models, including N-BEATS [16], N-HiTS [17], and Temporal Fusion Transformer (TFT) [18], have emerged as promising approaches for time series forecasting tasks. These architectures have demonstrated superior performance in various applications, including energy consumption forecasting, financial time series prediction, and weather forecasting, and have enabledtransfer learning in time series forecasting. However, these techniques have not been applied in forecasting vital signs for critical care patients.

In this work, we introduce a deep learning-based multi-step forecasting system for forecasting vital signs using previously observed vital signs of patients in the Intensive Care Units (ICUs) suffering from sepsis in critical care. Accurately forecasting the vital signs has the potential to assist healthcare providers in clinical settings such as ICUs in detecting early signs of physiological instability and anticipating changes in a patient's condition e.g., septic shock that is defined as a subset of sepsis. Our contribution in this paper can be summarized as follows: **(1)** We introduce a deep learning vital sign forecasting system to forecast 3 hours of the vital signs of patients in critical care using only 6 hours of previous vital sign data and evaluate the forecasted results using evaluation metrics: mean squared error (MSE), and dynamic time warping (DTW). **(2)** We evaluate the performance of three state-of-the-art forecasting models, N-BEATS, N-HiTS, and TFT, on forecasting using a publicly available dataset eICU Collaborative Research Database (eICU-CRD) [19]. **(3)** Since the shape and temporal changes in the vital signs are essential for healthcare providers in making clinical decisions, we use a DILATE [20] loss function to capture the spatial and temporal variations in the forecasted vital signs. **(4)** We carried out a comprehensive qualitative analysis by comparing the performance of different forecasting models and assessing the performance of models on different forecasting horizons.

We would like to acknowledge Spass Inc., South Korea for funding this research.

979-8-3503-7198-7/24/$31.00 ©2024 IEEE
DOI 10.1109/AIMHC59811.2024.00018

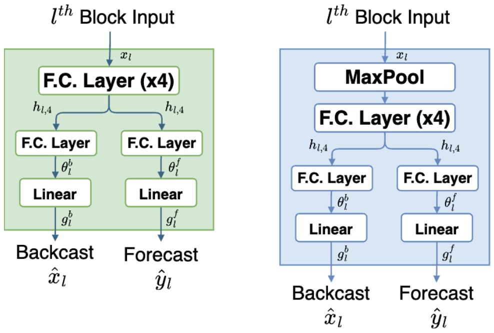


Fig. 1: The figure shows the l^{th} basic block of the N-BEATS (left) and the N-HiTS architecture.

The results of our model and experiments provide insights into data performance, potentially leading to improved patient care and outcomes in real-world scenarios.

II. Method

A. Dataset Description and Data Preprocessing

In our study, we employed the eICU-CRD [19], focusing on the *vitalPeriodic* and *diagnosis* tables. We gathered 5-minute interval time series data on mean blood pressure (MBP), heart rate (HR), and respiration rate (RR) for patients with sepsis or septic shock. MBP was derived using the formula MBP = DBP + 1/3 [SBP – DBP] [21], where DBP and SBP are diastolic and systolic blood pressure, respectively. Missing data were imputed via forward fill, excluding cases with over 25 minutes of missing data pre-diagnosis. To impute missing vital sign data, we utilized forward fill and excluded patients with over 25 minutes of missing data before diagnosis offset. For each patient, we retained data up to 9 hours prior to diagnosis offset and arranged time series data into groups that concluded with sepsis or septic shock diagnosis. This permitted the model to recognize the pattern of vital signs that led to sepsis or septic shock diagnosis [22]. To remove outliers, we implemented a low pass filter on the vital signs and eliminated patients with less than 0.0025 standard deviations in their vital signs. The final dataset contained 4020 groups across 1442 patients. For scaling the time series data, we used min-max scaling. In the normalization of the vital signs data, we eschewed the conventional approach of using the dataset's own minimum and maximum values for scaling. Instead, we applied domain expertise to determine clinically sensible scale ranges for each vital parameter: HR from 0 to 300 bpm, MBP from 0 to 190 mmHg, and RR from 0 to 100 mmHg [22], [23].

B. Experiment Setup and Deep Learning Pipeline

Our process involves randomly splitting the dataset into mutually exclusive training, validation, and testing groups in an 80:10:10 ratio. For the forecasting model, we used the vital signs of the training and validation groups to create a 6 hour input trajectory (72 steps) and a 3 hour future trajectory (36 steps) for prediction. We also conducted experiments with and without covariates (i.e., HR, MBP, and RR) to gauge their impact on model performance. We use standard metrics like MSE and DTW to compare the models' performance against a naive persistence model that predicted future values by repeating the last observed value in the input window for all forecasted time steps. See Figure 2 for a visual representation of our pipeline.

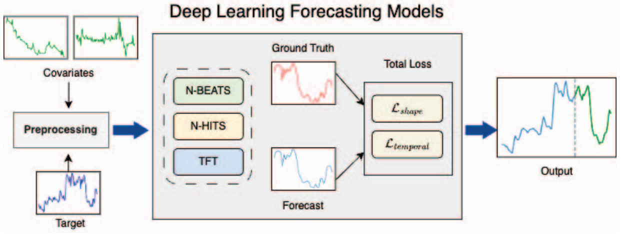


Fig. 2: The figure shows our deep learning pipeline for forecasting vital signs of patients suffering from Sepsis or Septic Shock condition in the ICUs.

C. Deep Learning Forecasting Models

We utilize state-of-the-art forecasting techniques, namely N-BEATS [16], N-HiTS [17], and TFT [18], to predict a 3 hour time series of vital signs: MBP and HR.

N-BEATS [16] is a DL architecture designed for both univariate and multivariate time series forecasting across multiple horizons. The architecture consists of three core components, namely a basic block, a stack (a combination of blocks), and the final model. As depicted in Figure 1, each basic block in the architecture receives a corresponding lookback window x_l as input and outputs two vectors, backcast ($\widehat{x}_l$) and forecast ($\widehat{y}_l$). The basic block comprises a stack of fully connected layers that generate backward (θ_l^b) and forward (θ_l^f) expansion coefficients (Eq. 1) [16]:

$$\theta_l^b = \text{Linear}_l^b(h_{l,4}), \quad \theta_l^f = \text{Linear}_l^f(h_{l,4}), \tag{1}$$

where $h_{l,4}$ is the output of the fourth fully connected layer in the basic block and Linear layer is a linear projection layer [16]. And the second part consists of backward (g_l^b) and forward (g_l^f) basis layers that produce backcast and forecast outputs (Eq. 2) [16]:

$$\widehat{y}_l = \sum_{i=1}^{\dim(\theta_l^f)} \theta_{l,i}^f \text{v}_{l,i}^f, \quad \widehat{x}_l = \sum_{i=1}^{\dim(\theta_l^b)} \theta_{l,i}^b \text{v}_{l,i}^b. \tag{2}$$

Here, $v_{l,i}^f$ and $v_{l,i}^b$ are forecast and backcast basis vectors.

N-HiTS [17], a DL model tailored for long-horizon forecasting, adeptly addresses prediction volatility and computational challenges. It employs multi-rate time series sampling and a novel hierarchical interpolation method, allowing for the integration of short-term and long-term temporal effects across various time scales. The Figure 1 highlights the architectural distinctions between N-HiTS and its counterpart, N-BEATS. The Eq. 3 is used to achieve multi-rate signal sampling for an l^{th} basic block [17].

$$y^{(p)}_{t-L:t,l} = \text{MaxPool}\left(y_{t-L:t,l}, k_l\right), \tag{3}$$

Here, k_l is the kernel size of the MaxPool layer.

TFT [18] is an attention-based DL model designed for multi-horizon time series forecasting. The architecture of TFT comprises the following building blocks: 1. Gated Residual Networks (GRNs) allow for non-linear processing to be applied only when necessary, resulting in a concentration on crucial components while suppressing extraneous ones [24]. The GRNs are formulated based on the Eq. 4 outlined in [18].

$$\text{GRN}_\omega(a, c) = \text{LayerNorm}\left(a + \text{GLU}_\omega(\eta_1)\right) \tag{4}$$

where a and c are the primary input and optional context vector inputs to the GRN, GLU is Gated Linear Unit [24], $\eta_1 = W_{1,\omega}\eta_2 + b_{1,\omega}$ and $\eta_2 = \text{ELU}(W_{2,\omega}a + W_{3,\omega}c + b_{2,\omega})$ are intermediate layers, ELU is Exponential Linear Unit activation function [25], LayerNorm is a standard layer normalization [25], ω is an index to denote weight sharing, $W_{(.)}$ and $b_{(.)}$ are weights and biases. 2. A sequence-to-sequence encoders/decoders block that enables the identification of relationships between time steps and their surrounding values, summarizing smaller patterns. 3. A temporal multi-head attention block to identify long-term dependencies in time series data and prioritize essential patterns. 4. A variable selection network that performs instance-wise selection of variables for both static and time-dependent covariates based on the importance of their features.

D. Training with DILATE Loss Function

In critical care, a patient's vital signs are crucial in determining their current and future condition. To accurately predict critical events, it's important for a DL forecasting model to recognize sudden changes in the time series' shape and temporal features. The loss function proposed in [20] is specifically designed to address these aspects. The DILATE Loss function is composed of two distinct terms: the shape term [20], [26], [27] and the temporal term [20], which both aim to capture changes in the time series spatial and temporal characteristics, as shown in Eq. 5 [20].

$$\begin{aligned}\mathcal{L}_{\text{DILATE}}\left(\hat{\mathbf{y}}_i, \overset{*}{\mathbf{y}}_i\right) &= \alpha \mathcal{L}_{\text{shape}}\left(\hat{\mathbf{y}}_i, \overset{*}{\mathbf{y}}_i\right) \\ &\quad + (1-\alpha)\mathcal{L}_{\text{temporal}}\left(\hat{\mathbf{y}}_i, \overset{*}{\mathbf{y}}_i\right)\end{aligned} \tag{5}$$

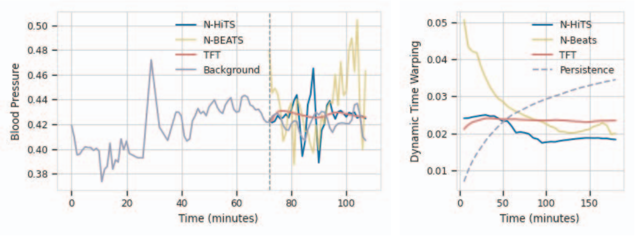
Here, the DILATE objective function consists of the shape term ($\mathcal{L}_{shape}$) and temporal term $\mathcal{L}_{temporal}$ that compare the predictions $\hat{y}_i$ with the ground truth future trajectory $\overset{*}{y}_i$. The shape and temporal terms are balanced by a hyperparameter $\alpha \in [0, 1]$.

Shape Term: The DILATE loss function's shape term was based on DTW and made differentiable using a smooth minimum operator for the differentiable shape term as per [20], [26], [27]. Here, $\mathbf{\Delta}\left(\hat{y}_i, \overset{*}{y}_i\right)$ is the pair-wise cost matrix, the warping path is defined as a binary matrix $\mathbf{A} \subset \{0,1\}^{k \times k}$ and $\gamma > 0$.

TABLE I: Performance of forecasting models on forecasting MBP and HR. Here, L1 is the MSE loss function and L2 is the DILATE loss function, covariates (W C) for MBP are HR & RR and covariates for HR are MBP & RR.

Models	Cov.	Mean Blood Pressure				Heart Rate			
		MSE*		DTW		MSE*		DTW	
		L-1	L-2	L-1	L-2	L-1	L-2	L-1	L-2
Persistence	-	24.55	24.55	34.50	34.50	7.35	7.35	17.52	17.52
N-HiTS	W C	**18.78**	19.99	20.44	**16.73**	**7.37**	7.57	13.12	**10.05**
	W/o C	**18.02**	19.81	20.46	**16.32**	7.22	**7.18**	13.97	**7.92**
N-BEATS	W C	**19.79**	24.40	19.37	**18.59**	**8.73**	10.95	14.36	**14.20**
	W/o C	**18.52**	27.42	**17.63**	18.60	**7.48**	12.98	**10.71**	17.90
TFT	W C	**18.89**	19.10	25.93	**23.51**	7.71	**7.19**	16.12	**15.16**
	W/o C	19.45	**19.00**	25.65	**23.46**	8.12	**7.57**	16.65	**15.79**

*The MSE values are scaled by $1e^{-4}$ for better representation.



(a) Forecasting results

(b) Error analysis

Fig. 3: N-HiTS, N-BEATS and TFT without covariates using DILATE loss function.

$$\begin{aligned}\mathcal{L}_{\text{shape}}\left(\hat{\mathbf{y}}_i, \overset{*}{\mathbf{y}}_i\right) &= \text{DTW}_\gamma\left(\hat{\mathbf{y}}_i, \overset{*}{\mathbf{y}}_i\right) \\ &= -\gamma \log\left(\sum_{\mathbf{A} \in \mathcal{A}_{k,k}} \exp\left(-\frac{\left\langle \mathbf{A}, \mathbf{\Delta}\left(\hat{\mathbf{y}}_i, \overset{*}{\mathbf{y}}_i\right)\right\rangle}{\gamma}\right)\right)\end{aligned} \tag{6}$$

Temporal Term: The temporal loss function is inspired from the time distortion index [28], [29], where $Z = \sum_{\mathbf{A} \in \mathcal{A}_{k,k}} \exp^{-\frac{\langle \mathbf{A}, \mathbf{\Delta}(\hat{\mathbf{y}}_i, \overset{*}{\mathbf{y}}_i)\rangle}{\gamma}}$ and $\mathbf{\Omega}$ is a square matrix of size $k \times k$ penalizing each element $\hat{y}_i^h$ being associated to an $\overset{*}{y}{}_i^j$, for $h \neq j$. Similar to [20], we have used the squared penalization for our experiments, i.e., $\Omega(h, j) = \frac{1}{k^2}(h-j)^2$. The differentiable loss used for the temporal term is given as:

$$\begin{aligned}\mathcal{L}_{\text{temporal}}\left(\hat{\mathbf{y}}_i, \overset{*}{\mathbf{y}}_i\right) &:= \left\langle \mathbf{A}^*_\gamma, \mathbf{\Omega}\right\rangle \\ &= \frac{1}{Z} \sum_{\mathbf{A} \in \mathcal{A}_{k,k}} \langle \mathbf{A}, \mathbf{\Omega}\rangle \exp^{-\frac{\left\langle \mathbf{A}, \mathbf{\Delta}\left(\hat{\mathbf{y}}_i, \overset{*}{\mathbf{y}}_i\right)\right\rangle}{\gamma}}\end{aligned} \tag{7}$$

III. Results and Discussion

We evaluate the performance of N-HiTS, N-BEATS, and TFT using test samples, summarized in Table I through MSE and DTW error scores. Figure 3a indicates that TFT captures the trends well, while N-HiTS better retains fluctuations within a certain range. Interestingly, N-HiTS performs better without covariates, whereas TFT benefits from them. Our evaluation of trained models assesses their performance over a range of forecasting horizons, from 5 minutes to 3 hours, using the DTW metric. To accomplish this, we generate horizon windows of increasing length, starting at one timestep (5 minutes) and ending with 36 timesteps (180 minutes). This enables a thorough evaluation of each model's performance compared to the Persistent model. Figure 3b displays the DTW errors for the models N-HiTS, N-BEATS, TFT, and Persistence for MBP. Deep learning models initially have higher DTW errors than the persistence model near the 50-minute mark but later demonstrate a downward trend and eventually outperform the persistence model, a pattern anticipated due to the gradual change in vital signs that tend to stay near prior values.

IV. Conclusion and Future Work

Our research has introduced a DL system that accurately forecasts vital signs in ICU patients with sepsis. We tested three sophisticated models and applied a DILATE loss function to capture the vital sign fluctuations crucial for timely clinical intervention. The promising results indicate that this approach could greatly assist in the early detection of conditions like septic shock. However, future efforts should look into integrating more clinical parameters and multimodal data to enhance forecasting accuracy. This refinement and broader data incorporation have the potential to further improve the system's utility in real-world clinical settings, ultimately benefiting patient care in critical situations.

References

[1] H. I. Kim and S. Park, "Sepsis: early recognition and optimized treatment," *Tuberculosis and respiratory diseases*, vol. 82, no. 1, pp. 6–14, 2019.

[2] J. S. Calvert, D. A. Price, U. K. Chettipally, C. W. Barton, M. D. Feldman, J. L. Hoffman, M. Jay, and R. Das, "A computational approach to early sepsis detection," *Computers in biology and medicine*, vol. 74, pp. 69–73, 2016.

[3] C. Wacker, A. Prkno, F. M. Brunkhorst, and P. Schlattmann, "Procalcitonin as a diagnostic marker for sepsis: a systematic review and meta-analysis," *The Lancet infectious diseases*, vol. 13, no. 5, pp. 426–435, 2013.

[4] J. M. Ferreras, D. Judez, G. Tirado, C. Aspiroz, R. Martínez-Álvarez, P. Dorado, A. Ezpeleta, R. Marrón, B. Gargallo, and C. Herranz, "Implementación de un sistema de alarmas automático para la detección precoz de los pacientes con sepsis grave," *Enfermedades Infecciosas y Microbiología Clínica*, vol. 33, no. 8, pp. 508–515, 2015.

[5] M. J. Hall, S. N. Williams, C. J. DeFrances, and A. Golosinskiy, "Inpatient care for septicemia or sepsis: a challenge for patients and hospitals," *NCHS data brief*, 2011.

[6] 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*, vol. 34, no. 6, pp. 1589–1596, 2006.

[7] W. A. Knaus, J. E. Zimmerman, D. P. Wagner, E. A. Draper, and D. E. Lawrence, "Apache—acute physiology and chronic health evaluation: a physiologically based classification system," *Critical care medicine*, vol. 9, no. 8, pp. 591–597, 1981.

[8] J.-R. Le Gall, P. Loirat, A. Alperovitch, P. Glaser, C. Granthil, D. Mathieu, P. Mercier, R. Thomas, and D. Villers, "A simplified acute physiology score for icu patients." *Critical care medicine*, vol. 12, no. 11, pp. 975–977, 1984.

[9] A. E. Jones, S. Trzeciak, and J. A. Kline, "The sequential organ failure assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation," *Critical care medicine*, vol. 37, no. 5, p. 1649, 2009.

[10] P. E. Marik and A. M. Taeb, "Sirs, qsofa and new sepsis definition," *Journal of thoracic disease*, vol. 9, no. 4, p. 943, 2017.

[11] M. J. Kane, N. Price, M. Scotch, and P. Rabinowitz, "Comparison of arima and random forest time series models for prediction of avian influenza h5n1 outbreaks," *BMC bioinformatics*, vol. 15, no. 1, pp. 1–9, 2014.

[12] G. P. Zhang, "Time series forecasting using a hybrid arima and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[13] X. Wang, H. Liu, J. Du, X. Dong, and Z. Yang, "A long-term multivariate time series forecasting network combining series decomposition and convolutional neural networks," *Applied Soft Computing*, vol. 139, p. 110214, 2023.

[14] B. Behinaein, A. Bhatti, D. Rodenburg, P. Hungler, and A. Etemad, "A transformer architecture for stress detection from ecg," in *Proceedings of the 2021 ACM International Symposium on Wearable Computers*, 2021, pp. 132–134.

[15] A. Bhatti, B. Behinaein, D. Rodenburg, P. Hungler, and A. Etemad, "Attentive cross-modal connections for deep multimodal wearable-based emotion recognition," in *2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 2021, pp. 01–05.

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

[17] C. Challu, K. G. Olivares, B. N. Oreshkin, F. G. Ramirez, M. M. Canseco, and A. Dubrawski, "Nhits: Neural hierarchical interpolation for time series forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 6, 2023, pp. 6989–6997.

[18] B. Lim, S. Arik, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting. arxiv," *arXiv preprint arXiv:1912.09363*, 2019.

[19] T. J. Pollard, A. E. Johnson, J. D. Raffa, L. A. Celi, R. G. Mark, and O. Badawi, "The eicu collaborative research database, a freely available multi-center database for critical care research," *Scientific data*, vol. 5, no. 1, pp. 1–13, 2018.

[20] V. Le Guen and N. Thome, "Shape and time distortion loss for training deep time series forecasting models," *Advances in neural information processing systems*, vol. 32, 2019.

[21] G. Sainas, R. Milia, G. Palazzolo, G. Ibba, E. Marongiu, S. Roberto, V. Pinna, G. Ghiani, F. Tocco, and A. Crisafulli, "Mean blood pressure assessment during post-exercise: result from two different methods of calculation," *Journal of Sports Science & Medicine*, vol. 15, no. 3, p. 424, 2016.

[22] A. Bhatti, N. Thangavelu, M. Hassan, C. Kim, S. Lee, Y. Kim, and J. Y. Kim, "Interpreting forecasted vital signs using n-beats in sepsis patients," *arXiv preprint arXiv:2306.14016*, 2023.

[23] H. M. O'Halloran, K. Kwong, R. A. Veldhoen, and D. M. Maslove, "Characterizing the patients, hospitals, and data quality of the eicu collaborative research database," *Critical Care Medicine*, vol. 48, no. 12, pp. 1737–1743, 2020.

[24] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language modeling with gated convolutional networks. arxiv," *Computation and Language*, 2016.

[25] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *arXiv preprint arXiv:1607.06450*, 2016.

[26] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE transactions on acoustics, speech, and signal processing*, vol. 26, no. 1, pp. 43–49, 1978.

[27] M. Cuturi and M. Blondel, "Soft-dtw: a differentiable loss function for time-series," in *International conference on machine learning*. PMLR, 2017, pp. 894–903.

[28] L. Frías-Paredes, F. Mallor, M. Gastón-Romeo, and T. León, "Assessing energy forecasting inaccuracy by simultaneously considering temporal and absolute errors," *Energy Conversion and Management*, vol. 142, pp. 533–546, 2017.

[29] L. Vallance, B. Charbonnier, N. Paul, S. Dubost, and P. Blanc, "Towards a standardized procedure to assess solar forecast accuracy: A new ramp and time alignment metric," *Solar Energy*, vol. 150, pp. 408–422, 2017.