

Proceeding Paper

# Forecasting Vital Signs in Human-Robot Collaboration Using Sequence-to-Sequence Models with Bidirectional LSTM: A Comparative Analysis of Uni- and Multivariate Approaches <sup>†</sup>

Oliver Chojnowski <sup>\*</sup>, Dario Luipers, Caterina Neef and Anja Richert

Cologne Cobots Lab, TH Köln—University of Applied Sciences, Betzdorfer Str. 2, 50679 Köln, Germany; dario.luipers@th-koeln.de; caterina.neef@th-koeln.de

<sup>\*</sup> Correspondence: oliver.chojnowski@th-koeln.de (O.C.); anja.richert@th-koeln.de (A.R.)

<sup>†</sup> Presented at the 10th International Electronic Conference on Sensors and Applications (ECSA-10), 15–30 November 2023; Available online: <https://ecsa-10.sciforum.net/>.

**Abstract:** Our research investigates an approach to forecasting human vital signs by formulating the problem as a sequence-to-sequence (seq2seq) task, utilizing bidirectional long short-term memory models (BiLSTM). The study aims to compare the forecasting accuracy of uni- and multivariate modeling strategies over different forecasting horizons ranging from 1 s to 10 s. The dataset comprises sensor data collected during a lab study in which thirteen participants engaged in a collaborative assembly scenario with a robot. Our results show that univariate models outperform multivariate ones in terms of forecasting accuracy, offering valuable insights into accurate forecasting of human physiological parameters, with potential implications for human-robot collaboration, personalized medical monitoring, and healthcare applications.

**Keywords:** human-robot collaboration; forecasting; vital signs; deep learning; collaborative assembly

**Citation:** Chojnowski, O.; Luipers, D.; Neef, C.; Richert, A. Forecasting Vital Signs in Human-Robot Collaboration Using Sequence-to-Sequence Models with Bidirectional LSTM: A Comparative Analysis of Uni- and Multivariate Approaches. *2023*, *56*, x.

<https://doi.org/10.3390/xxxxx>

Academic Editor(s): Name

Published: 15 November 2023



**Copyright:** © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In the dynamic realm of human-robot collaboration (HRC), a significant challenge lies in equipping robotic systems with the ability to seamlessly adapt to users' internal states, such as stress or relaxation. Ongoing research in this field has shown that stress can be indirectly assessed through the integration of diverse sensors that monitor various physiological indicators, including Electrocardiograph (ECG), Pupil Dilation (PD), Electromyograph (EMG), Electroencephalograph (EEG), Heart Rate Variation (HRV), skin temperature, respiratory rate and Electrodermal activity (EDA) or Galvanic Skin Response (GSR) [1–4]. Machine Learning classification techniques have made noteworthy advancements in stress detection [5–8]. In diverse environments, such as academic, driving, or office-like settings, accuracy rates exceeding 90% have been achieved [5]. By going beyond simply recognizing emotions in real-time, to anticipatory modelling, robotic systems can adjust their behavior proactively, leading to more natural, productive collaborations. However, despite the promising developments in stress detection, the exploration of forecasting future states remains limited. Some research has been conducted on forecasting vital signs in intensive care patients [9], postoperative complications [10] or in health monitoring [11]. In [11], the authors compared different models, evaluating their accuracy in univariate forecasts of pulse, oxygen level percentage (SpO<sub>2</sub>) and blood pressure. Notably, deep learning models such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) outperformed classical forecasting strategies like Autoregressive (AR) and Autoregressive Integrated Moving Average (ARIMA) models, with GRUs performing the best. Earlier work also revealed in different use cases that Bidirectional Long Short-Term Memory (BiLSTM) models lead to a significant improvement in average time

series prediction accuracy of 37.78% [12] compared to classical LSTMs. It was observed that training the bidirectional variant was slower, suggesting that it extracts unique features inaccessible to other models [13]. In the field of mental state and vital sign forecasting the performance of BiLSTMs is unknown. Given the current state of the research, an intriguing avenue for further investigation pertains to the exploration of the intricate interplay between diverse sensor modalities, which may hold the potential to enhance vital sign forecasting. Specifically, there is an opportunity to explore whether the simultaneous utilization of multiple modalities in a multivariate forecasting framework can yield improved forecasting accuracies by leveraging information that remains latent to univariate models. This study significantly contributes by highlighting the impacts of multivariate forecasting strategies versus univariate approaches. It also provides insights into vital sign forecasting, particularly through the integration of BiLSTMs with collaborative robotics, thus advancing the existing knowledge in this field.

## 2. Materials and Methods

### 2.1. Dataset

The dataset used in this study consists of vital signs from 13 subjects recorded in the context of a collaborative assembly. In this assembly, a human worker collaboratively assembles a component with a collaborative robot (cobot). To capture the influence of the cobot on the human's vital signs, six different scenarios, differing in various factors such as the degree of collaboration or the working speed of the robot, were executed. Between every configuration the recording was stopped. As a result, each of the 13 subjects contributes 6 individual sequences, each lasting approximately 2 min, culminating in a total of 76 sequences. The utilized sensor modalities are the Interbeat Intervals of the heart (IBI) measured via ECG and the EDA of the skin, both using the BITalino (r)evolution Plugged Kit BLE/BT (PLUX Wireless Biosignals, Portugal) as well as the Pupil Dilation (PD), measured with pupil core eye tracking glasses (Pupil Labs, Berlin, Germany).

### 2.2. Bidirectional Long Short-Term Memory Model

Bidirectional Long Short-Term Memory networks (BiLSTM) are a type of recurrent neural network (RNN) architecture used in natural language processing and sequential data tasks, like time series data. Introduced to overcome the limitations of regular RNN they enhance traditional LSTMs by processing input data in both forward and backward directions, capturing context from both past and future [14]. BiLSTMs were introduced to address the vanishing gradient problem and improve the modeling of long-range dependencies in sequential data.

### 2.3. Preprocessing

The data preprocessing involved three steps. First, each modality was handled independently. For IBI, no direct measures were needed. For PD, blink removal was essential using the procedure outlined in [15,16]. EDA-Signal involved extracting the Skin Conductance Response, as described in [17]. In the second step, all modalities underwent uniform processing, which included resampling, smoothing, and data normalization to enhance quality and ensure consistency. In the final phase, individual modalities were synchronized to create a multivariate dataset. Extensive feature engineering was then performed on this dataset, yielding both static features (e.g., means, minimums, and maximums of time series) and dynamic features (e.g., moving averages and lag features).

### 2.4. Stationarity

Stationarity signifies that statistical parameters such as the mean and variance exhibit relative constancy throughout the observed time span [18]. This property holds significant importance, particularly in forecasting applications. To assess stationarity, we employed the Augmented Dickey-Fuller Test (ADF-Test), which is one of the most commonly used

measures of stationarity [19–21]. To induce stationarity a differentiation procedure was implemented, resulting in stationarity in 99% of all sequences.

### 2.5. Sequence-to-Sequence Modeling

In the context of time series forecasting, sequence-to-sequence modeling is a technique wherein a learner maps a sequence of past values to a sequence of future values [22]. To adapt the dataset into a format suitable for input and output sequences, we employed the sliding window method presented in [23]. Three variations of each dataset were created for one-second, five-second, and ten-second forecasting horizons, with consistent look-back window lengths.

### 2.6. Measures of Evaluation

To assess the forecasting accuracy of the models, we employ the Symmetric Mean Absolute Percentage Error (sMAPE). The formula for calculating sMAPE is presented below [24].

$$sMAPE = \left( \frac{2}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{|y_t| + |\hat{y}_t|} \right) * 100\% \quad (1)$$

To establish a baseline for assessing the model's performance and to ensure the robustness of our results, we employ a simple benchmark known as the Naïve Forecast as recommended by [25]. In this approach, the prediction for the next time step is generated by using the value from the previous time step, which makes it simple to calculate but nonetheless an effective benchmark method. This basic forecasting method is mathematically represented by Equation (2) [26].

$$\hat{y}_{t+k} = y_t \quad (2)$$

## 3. Results

### 3.1. Univariate Forecast

Table 1 illustrates the superior performance of the BiLSTM model compared to the baseline across all forecasting horizons for univariate IBI.

**Table 1.** sMAPE of the univariate forecast of the interbeat intervals.

Forecasting Horizon	Naïve Forecast	BiLSTM
1 s	123.79%	2.1%
5 s	144.79%	16.01%
10 s	146.37%	17.36%

Table 2 displays results for univariate PD forecasting. The Naïve Method consistently shows higher prediction errors than the BiLSTM model across all horizons.

**Table 2.** sMAPE of the univariate forecast of the pupil dilation.

Forecasting Horizon	Naïve Forecast	BiLSTM
1 s	142.92%	2.07%
5 s	154.14%	5.39%
10 s	157.00%	5.66%

Particularly noteworthy is the fact that when extending the forecasting horizon from 1 to 5 s, a marked increase in sMAPE is observed, amounting to 13.91% for univariate IBI and 3.32% for univariate PD. In contrast, extending the forecasting horizon from 5 to 10 s only results in an increase of 1.35% for univariate IBI and 0.27% for univariate PD.

### 3.2. Multivariate Forecast

Table 3 compares forecasting accuracy for univariate and multivariate models across different horizons. The multivariate approach consistently yields slightly higher sMAPE, outperforming univariate IBI by just 0.24% at the 5-s horizon.

**Table 3.** sMAPE of the multivariate compared to the univariate forecasts.

Forecasting Horizon	Univariate IBI	Univariate PD	Multivariate
1 s	2.1%	2.07%	16.12%
5 s	16.01%	5.39%	15.77%
10 s	17.36%	5.66%	22.48%

## 4. Discussion

The results presented in this work reveal a substantial disparity in performance between univariate and multivariate models. Despite the potential for multivariate models to leverage relationships among individual parameters, generated features, and additional skin conductance data, the incorporation of this supplementary input does not yield an improvement in forecasting accuracy. Several possible explanations for this phenomenon can be considered. Firstly, it is possible that no meaningful relationships exist among the various parameters under investigation. This lack of inherent correlations may limit the capacity of multivariate models to extract valuable predictive insights, rendering the inclusion of additional input variables ineffective. Secondly, the quality of the supplementary skin conductance data may be a contributing factor. It is conceivable that these data introduce noise into the prediction process, thereby diminishing overall accuracy. Further investigation into the reliability and relevance of the additional data may clarify its impact on model performance. Thirdly, the selected features for the multivariate models may either have no significant influence on the prediction accuracy or, in some cases, exert a detrimental effect. The inclusion of irrelevant or potentially confounding features can hinder the model's ability to discern meaningful patterns in the data, leading to suboptimal forecasting outcomes. These findings underscore the importance for a thorough understanding of the underlying relationships within the data and the potential consequences of incorporating additional variables.

## 5. Conclusions

The research findings presented in this study shed light on the predictive performance of univariate versus multivariate deep learning models in the context of forecasting vital signs. Notably, the univariate prediction of IBI and pupil diameter yields superior results when compared to the multivariate approach, which incorporates additional variables such as skin conductance and generated features. This suggests that the univariate models excel in capturing the intricate patterns and relationships within these physiological signals. Interestingly, as the forecasting horizon increases from one to five seconds, a significant decrease in accuracy is observed. However, this decline in accuracy remains relatively stable when extending the forecasting horizon from five to ten seconds. These findings have important implications for predictive modeling in physiological signal analysis where high precision is required, such as assessing cognitive load or attention levels. The observed stability in forecasting accuracy for longer horizons indicates that the univariate approach may offer a reliable foundation for longer-term physiological forecasting tasks. This work contributes valuable insights into the selection of modeling approaches for vital sign forecasting, underscoring the significance of considering the specific predictive goals and horizons in such applications. Future research in this domain should explore alternative feature engineering strategies, data preprocessing techniques, and model architectures to unlock the latent predictive potential of multivariate approaches. Overall, this study contributes valuable insights into the complexities of multivariate modeling in physiological signal analysis and paves the way for further advancements in this field.

**Author Contributions:** Conceptualization, O.C. and D.L.; methodology, O.C., D.L. and C.N.; software, O.C.; validation, O.C.; formal analysis, O.C.; investigation, O.C.; resources, A.R.; data curation, D.L.; writing—original draft preparation, O.C.; writing—review and editing, C.N., D.L. and A.R.; visualization, O.C.; supervision, A.R.; project administration, A.R.; funding acquisition, A.R. and D.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Ministry of Innovation, Science and Research of North Rhine-Westphalia Germany.

**Institutional Review Board Statement:** Institutional Review Board approval was not sought for this study as the data utilized in our research originated from a pre-existing dataset. Furthermore, all data employed in our analysis had been anonymized to protect the privacy and confidentiality of the individuals involved. This rigorous anonymization process ensured that no personally identifiable information was accessible or discernible in the dataset, thereby mitigating any ethical concerns associated with the use of human subjects' data. Consequently, the research conducted in this study adheres to established ethical guidelines and does not require additional IRB approval.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy reasons.

**Acknowledgments:** The authors would like to thank all subjects that participated in the study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Villani, V.; Sabattini, L.; Secchi, C.; Fantuzzi, C. A Framework for Affect-Based Natural Human-Robot Interaction. In Proceedings of the 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Nanjing, China, 27–31 August 2018; pp. 1038–1044. <https://doi.org/10.1109/ROMAN.2018.8525658>.
2. Arai, T.; Kato, R.; Fujita, M. Assessment of operator stress induced by robot collaboration in assembly. *CIRP Ann.* **2010**, *59*, 5–8. <https://doi.org/10.1016/j.cirp.2010.03.043>.
3. Lu, L.; Xie, Z.; Wang, H.; Li, L.; Xu, X. Mental stress and safety awareness during human-robot collaboration—Review. *Appl. Ergon.* **2022**, *105*, 103832. <https://doi.org/10.1016/j.apergo.2022.103832>.
4. Peternel, L.; Tsagarakis, N.; Caldwell, D.; Ajoudani, A. Robot adaptation to human physical fatigue in human-robot co-manipulation. *Auton. Robot.* **2018**, *42*, 1011–1021. <https://doi.org/10.1007/s10514-017-9678-1>.
5. Gedam, S.; Paul, S. A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques. *IEEE Access* **2021**, *9*, 84045–84066. <https://doi.org/10.1109/ACCESS.2021.3085502>.
6. Iqbal, T.; Elahi, A.; Shahzad, A.; Wijns, W. Review on Classification Techniques used in Biophysiological Stress Monitoring. *arXiv* **2022**, arXiv:2210.16040. <https://doi.org/10.48550/arXiv.2210.16040>.
7. Baltaci, S.; Gokcay, D. Stress Detection in Human-Computer Interaction: Fusion of Pupil Dilation and Facial Temperature Features. *Int. J. Hum. Comput. Interact.* **2016**, *32*, 956–966. <https://doi.org/10.1080/10447318.2016.1220069>.
8. TuerxunWaili; Alshebly, Y.S.; Sidek, K.A.; Johar, M.G.M. Stress recognition using Electroencephalogram (EEG) signal. *J. Phys. Conf. Ser.* **2020**, *1502*, 012052. <https://doi.org/10.1088/1742-6596/1502/1/012052>.
9. Phetrittikun, R.; Suvirat, K.; Pattalung, T.N.; Kongkamol, C.; Ingviya, T.; Chaichulee, S. Temporal Fusion Transformer for forecasting vital sign trajectories in intensive care patients. In Proceedings of the 2021 13th Biomedical Engineering International Conference (BMEiCON), Ayutthaya, Thailand, 19–21 November 2021; pp. 1–5. <https://doi.org/10.1109/BMEiCON53485.2021.9745215>.
10. Fritz, B.A.; Chen, Y.; Murray-Torres, T.M.; Gregory, S.; Ben Abdallah, A.; Kronzer, A.; McKinnon, S.L.; Budelier, T.; Helsten, D.L.; Wildes, T.S.; et al. Using machine learning techniques to develop forecasting algorithms for postoperative complications: Protocol for a retrospective study. *BMJ Open* **2018**, *8*, e020124. <https://doi.org/10.1136/bmjopen-2017-020124>.
11. Bhavani, T.; VamseeKrishna, P.; Chakraborty, C.; Dwivedi, P. Stress Classification and Vital Signs Forecasting for IoT-Health Monitoring. *IEEE/ACM Trans. Computat. Biol. Bioinform.* **2022**, 1–8. <https://doi.org/10.1109/TCBB.2022.3196151>.
12. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. The Performance of LSTM and BiLSTM in Forecasting Time Series. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 3285–3292. <https://doi.org/10.1109/BigData47090.2019.9005997>.
13. Jia, M.; Huang, J.; Pang, L.; Zhao, Q. Analysis and Research on Stock Price of LSTM and Bidirectional LSTM Neural Network. In Proceedings of the 3rd International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2019), Chongqing, China, 30–31 May 2019; Atlantis Press: Amsterdam, the Netherland, 2019; pp. 467–473. <https://doi.org/10.2991/iccia-19.2019.72>.

14. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. <https://doi.org/10.1109/78.650093>.
15. Pedrotti, M.; Lei, S.; Dzaack, J.; Rötting, M. A data-driven algorithm for offline pupil signal preprocessing and eyeblink detection in low-speed eye-tracking protocols. *Behav. Res.* **2011**, *43*, 372–383. <https://doi.org/10.3758/s13428-010-0055-7>.
16. Pedrotti, M.; Mirzaei, M.A.; Tedescho, A.; Chardonnet, J.-R.; Merienne, F.; Benedetto, S.; Baccino, T. Automatic Stress Classification With Pupil Diameter Analysis. *Int. J. Hum. Comput. Interact.* **2014**, *30*, 220–236. <https://doi.org/10.1080/10447318.2013.848320>.
17. Zangróniz, R.; Martínez-Rodrigo, A.; Pastor, J.M.; López, M.T.; Fernández-Caballero, A. Electrodermal Activity Sensor for Classification of Calm/Distress Condition. *Sensors* **2017**, *17*, 10. <https://doi.org/10.3390/s17102324>.
18. Livieris, I.E.; Stavroyiannis, S.; Iliadis, L.; Pintelas, P. Smoothing and stationarity enforcement framework for deep learning time-series forecasting. *Neural Comput. Applic* **2021**, *33*, 14021–14035. <https://doi.org/10.1007/s00521-021-06043-1>.
19. Worden, K.; Iakovidis, I.; Cross, E.J. On Stationarity and the Interpretation of the ADF Statistic. In *Dynamics of Civil Structures, Volume 2*; Pakzad, S., Ed.; Conference Proceedings of the Society for Experimental Mechanics Series; Springer International Publishing: Cham, Switzerland, 2019; pp. 29–38. [https://doi.org/10.1007/978-3-319-74421-6\\_5](https://doi.org/10.1007/978-3-319-74421-6_5).
20. Dickey, D.A.; Fuller, W.A. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *J. Am. Stat. Assoc.* **1979**, *74*, 427–431. <https://doi.org/10.1080/01621459.1979.10482531>.
21. Dickey, D.A.; Fuller, W.A. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica* **1981**, *49*, 1057–1072. <https://doi.org/10.2307/1912517>.
22. Mariet, Z.; Kuznetsov, V. Foundations of Sequence-to-Sequence Modeling for Time Series. In Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics, Naha, Japan, 16–18 April 2019; pp. 408–417. Available online: <https://proceedings.mlr.press/v89/mariet19a.html> (accessed on 22 September 2023).
23. Shi, J.; Jain, M.; Narasimhan, G. Time Series Forecasting (TSF) Using Various Deep Learning Models. *arXiv* **2022**, arXiv:2204.11115. <https://doi.org/10.48550/arXiv.2204.11115>.
24. Abbasimehr, H.; Paki, R. Improving time series forecasting using LSTM and attention models. *J. Ambient. Intell. Hum. Human. Comput.* **2022**, *13*, 673–691. <https://doi.org/10.1007/s12652-020-02761-x>.
25. Dhakal, C. A Naïve Approach for Comparing a Forecast Model. August 2018. Available online: [https://www.researchgate.net/publication/326972994\\_A\\_Naive\\_Approach\\_for\\_Comparing\\_a\\_Forecast\\_Model](https://www.researchgate.net/publication/326972994_A_Naive_Approach_for_Comparing_a_Forecast_Model) (accessed on).
26. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 3rd ed. Available online: <https://otexts.com/fpp3/index.html> (accessed on 22 September 2023).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.