

CORRESPONDENCE

Predicting vital sign deviations during surgery from patient monitoring data: developing and validating single-stream deep learning models

Alina Dubatovka¹, Christoph B. Nöthiger², Donat R. Spahn²,
Joachim M. Buhmann¹, Tadzio R. Roche^{2,*†} and David W. Tscholl^{2,†}

¹Department of Computer Science, ETH Zurich, Zurich, Switzerland and ²Institute of Anaesthesiology, University of Zurich, University Hospital Zurich, Zurich, Switzerland

*Corresponding author. E-mail: Tadzioraoul.roche@usz.ch

†Contributed equally to this work and share last authorship.

Keywords: algorithms; artificial intelligence; biosignals; clinical decision support systems; digital medicine; machine learning; perioperative medicine, vital signs

Editor—In a joint project between two Swiss universities, we developed machine learning models to predict deviations from upper and lower thresholds across 13 distinct vital signs, as outlined in [Table 1](#). This project had two primary goals: to demonstrate the feasibility of short-term predictions (20 s–20 min) across a wider range of vital signs than previously researched; and to develop models for a future Artificial Intelligence Clinical Decision Support System (AI-CDSS) to be used in patient monitors, relying only on the continuous vital sign data available in these monitors. Such a system could proactively alert clinicians to impending vital sign changes, enabling adjustments based on forecasts rather than alarms.¹ We envision a system that, similar to the common trend displays for the recent past of each vital sign, would provide predictive trend indications.

This correspondence provides an overview of our models, with more detailed information in the [Supplementary Appendix Sections Supplement 1](#) (Detailed Description of the Model Development) and [Supplement 2](#) (Model Characteristics and Performance). We developed single-stream deep learning models for each vital sign.^{2,3} Each model continuously analyses its respective vital sign data stream and generates predictions. Our model architecture includes a Long Short-Term Memory layer to process raw input,⁴ compute an internal representation, and a dense output layer for making predictions. These models can predict as far into the future as they have observed in the past. For instance, after 20 min of data analysis, the models can make predictions for the next 20

min, checking if the computed value falls within the vital sign threshold.

We used 37,069 anaesthesia records from the University Hospital Zurich, Switzerland, a tertiary care university hospital. Most cases had a 1-min resolution, with 653 cases at a 1-s resolution. Characteristics of the dataset are outlined in [Supplementary Table S1](#). The dataset covered a wide range of procedures and patients for more than 2 yr, including elective, emergency, ambulatory, cardiac, noncardiac, general, and regional anaesthesia cases during all perioperative phases. Before developing the models, we excluded vital sign signals outside predefined plausible ranges ([Supplementary Table S2](#)).

Our strategic decision to use a large real-life dataset containing only vital sign data and minimal data cleaning was guided by our intent to create models that would work in a patient monitor using exclusively high-quality vital sign data. This approach addressed data source and quality issues, substantial obstacles to the practical implementation of AI-CDSS.⁵

Next, we internally validated the models and used the eICU⁶ and MIMIC-III⁷ datasets for external validation. Internal validation results are shown in [Table 1](#) and [Supplementary Table S4](#), and external validation results are presented in [Supplementary Tables S5 and S6](#). We used the balanced accuracy metric to present model performance because it is easy to understand. Balanced accuracy evaluates the correct or incorrect prediction of deviations, considering the varying number of actual deviations and non-deviations in the

Table 1 Performance of the vital sign prediction models (classification task). Numbers are balanced accuracy ($=[\text{Sensitivity}+\text{Specificity}]/2$). Predictions with a forecast horizon of seconds use the high-resolution dataset (1 s^{-1}), and predictions in minutes use the 1 min^{-1} dataset. Values in bold: balanced accuracy 0.50–0.70; values in italic: balanced accuracy 0.71–0.84; values in regular font: balanced accuracy 0.85–1. The vital sign thresholds, for which the models achieved the balanced accuracies shown here, are listed in [Supplementary Table S2](#). BIS, Bispectral Index®; CVP, central venous pressure; etCO₂, end-tidal CO₂ concentration; HR, heart rate; IABP, invasive arterial blood pressure; N/A, not available; NIABP, noninvasive arterial blood pressure; PR, pulse rate; RR, respiratory rate; SpO₂, oxygen saturation; Temp, temperature; TOF%, train-of-four ratio; TV, tidal volume.

Variable name	Prediction horizon											
	10 s		20 s		40 s		5 min		10 min		20 min	
	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High
Temp	0.984	0.991	0.990	0.992	0.965	0.988	0.967	0.961	0.954	0.952	0.916	0.940
HR	0.965	0.954	0.955	0.950	0.952	0.937	0.921	0.890	0.904	0.875	0.874	0.853
PR	0.742	0.925	0.717	0.924	0.679	0.920	0.819	0.896	0.807	0.875	0.786	0.844
ST-segment deviation	0.929	0.940	0.908	0.902	0.901	0.920	0.879	0.912	0.875	0.906	0.847	0.879
IABP	0.942	0.940	0.975	0.926	0.912	0.912	0.753	0.746	0.727	0.720	0.715	0.695
CVP	0.924	0.940	0.936	0.938	0.913	0.930	0.864	0.887	0.851	0.870	0.813	0.838
RR	0.825	0.885	0.876	0.851	0.788	0.781	0.920	0.770	0.907	0.762	0.890	0.734
TV	0.900	0.895	0.883	0.890	0.868	0.852	0.850	0.877	0.843	0.862	0.829	0.831
etCO ₂	0.949	0.889	0.936	0.875	0.938	0.825	0.881	0.824	0.857	0.767	0.821	0.731
TOF%	N/A	0.994	N/A	0.998	N/A	0.997	N/A	0.894	N/A	0.858	N/A	0.849
BIS	0.931	0.927	0.911	0.907	0.913	0.899	0.838	0.833	0.810	0.806	0.774	0.772
SpO ₂	0.953	0.957	0.948	0.950	0.939	0.940	0.868	0.869	0.838	0.840	0.809	0.810
NIABP	No data	No data	No data	No data	No data	No data	0.685	0.668	0.650	0.618	0.607	0.585

dataset. A balanced accuracy of >0.85 indicates the models' predictions would have been correct for 8.5 out of 10 instances with the tested thresholds. For 19 of 23 (82%) vital sign deviations, predictions could be made 40 s in advance with a balanced accuracy of >0.85 . In anaesthesia, this could allow an anaesthetist to act just before an event such as low blood pressure by administering a vasopressor or using the Trendelenburg position. For 16 of 25 (64%) vital sign deviations, predictions with a balanced accuracy >0.85 could be made as

early as 5 min ahead. For 7 of 25 (28%) vital sign deviations, including temperature and ST-segment changes, balanced accuracy was >0.85 , 20 min ahead. Low temperature raises bleeding risk,⁸ and ST-segment changes can indicate developing coronary hypoperfusion.⁹

Our models have strengths and limitations. They were trained and validated using data from a broad range of patients and cases, encompassing all perioperative phases for more than 2 yr. These data were recorded under continuous

vital sign assessment by clinicians who actively intervened to correct impending vital sign deviations. Because our goal was to create models that could function effectively with only vital sign data in real-life clinical settings, we did not include additional data at this stage and do not have access to such data for the dataset used to train our models. We recognise that providing demographic data for our dataset could further improve the models' applicability and interpretation. Model performance could also be improved by incorporating additional data such as age, ASA physical status, phases of anaesthesia, and interventions by clinicians. Moreover, applying artifact removal strategies beyond merely excluding implausible values could enhance model accuracy. Our models were trained and developed using perioperative patient data, whereas external validation was performed using intensive care unit patient data. In addition, not all vital signs studied were available in the external validation datasets. Both training and validation datasets originated from high-income countries. Patient groups from countries with different healthcare settings might exhibit substantial differences, potentially affecting model performance. The balanced accuracy values provided in Table 1 apply to the vital sign thresholds listed in Supplementary Table S2. Different thresholds would alter model performance. These limitations highlight areas for future improvement and underscore the need for additional research to ensure the robustness and generalisability of our models across different patient populations and healthcare settings.

With this correspondence, we aim to raise awareness of the feasibility of developing deep learning models to predict a wider range of vital sign deviations than commonly studied. To explore how these predictions can be integrated with a user-friendly interface to provide a clinically meaningful AI-CDSS for patient monitoring requires further research and validation, both regarding hypothetical and clinical outcomes. We cannot make the software code for the models publicly available at this time because it would reduce our chances of developing them into a product. The code is intellectual property of ETH Zurich, and Philips (Amsterdam, The Netherlands) holds a nonexclusive license. Specific strengths of our models lie in their tailored development for making vital sign predictions using exclusively vital sign data and their surprisingly good performance in validation considering the single-stream input data.

Declarations of interest

CBN, TRR, and DWT are designated inventors of VisualPatient-Predictive, and CBN and DWT are designated inventors of Visual Patient, for which the University of Zurich and Koninklijke Philips N.V. hold various patents, design protections, and trademarks. There are cooperation and licensing agreements with Philips Medizin Systeme Böblingen GmbH, Böblingen, Germany; Koninklijke Philips N.V., Amsterdam, The Netherlands; Philips Research/Philips Electronics Nederland BV, Eindhoven, The Netherlands; Philips USA, Cambridge, MA, USA. Under these agreements, CBN, TRR and DWT receive research funding travel support, lecturing honoraria and royalties. DWT receives honoraria/travel support for consulting or lecturing from Instrumentation Laboratory, Bedford, MA, USA; Swiss Foundation for Anesthesia Research, Zurich, Switzerland; and ISICEM, Brussels, Belgium. CBN receives

honoraria/travel support for consulting or lecturing from Instrumentation Laboratory, Bedford, MA, USA. DRS's academic department is receiving grant support from the Swiss National Science Foundation, Berne, Switzerland; the Swiss Society of Anesthesiology and Perioperative Medicine (SSAPM), Berne, Switzerland; the Swiss Foundation for Anesthesia Research, Zurich, Switzerland; Vifor SA, Villars-sur-Glâne, Switzerland; and Vifor (International) AG, St. Gallen, Switzerland. DRS is co-chair of the ABC-Trauma Faculty, sponsored by unrestricted educational grants from Novo Nordisk Health Care AG, Zurich, Switzerland; CSL Behring GmbH, Marburg, Germany; LFB Biomédicaments, Courtaboeuf Cedex, France; and Octapharma AG, Lachen, Switzerland. DRS received honoraria/travel support for consulting or lecturing from Danube University of Krems, Austria; US Department of Defense, Washington, USA; European Society of Anesthesiology and Intensive Care, Brussels, BE; Korean Society for Patient Blood Management, Seoul, Korea; Korean Society of Anesthesiologists, Seoul, Korea; Network for the Advancement of Patient Blood Management, Haemostasis and Thrombosis, Paris, France; Alexion Pharmaceuticals Inc., Boston, MA; Bayer AG, Zurich, Switzerland; B. Braun Melsungen AG, Melsungen, Germany; CSL Behring GmbH, Hattersheim am Main, Germany and Berne, Switzerland; Celgene International II Sàrl, Couvet, Switzerland; Daiichi Sankyo AG, Thalwil, Switzerland; Haemonetics, Braintree, MA, USA; Instrumentation Laboratory (Werfen), Bedford, MA, USA; LFB Biomédicaments, Courtaboeuf Cedex, France; Merck Sharp & Dohme, Kenilworth, New Jersey, USA; Novo Nordisk Health Care AG, Zurich, Switzerland; PAION Deutschland GmbH, Aachen, Germany; Pharmacosmos A/S, Holbaek, Denmark; Pfizer AG, Zurich, Switzerland; Pierre Fabre Pharma, Alschwil, Switzerland; Portola Schweiz GmbH, Aarau, Switzerland; Roche Diagnostics International Ltd., Reinach, Switzerland; Sarstedt AG & Co., Sevelen, Switzerland and Nümbrecht, Germany; Shire Switzerland GmbH, Zug, Switzerland; Tem International GmbH, Munich, Germany; Vifor Pharma, Munich, Germany; Neuilly sur Seine, France and Villars-sur-Glâne, Switzerland; Vifor (International) AG, St. Gallen, Switzerland; and Zuellig Pharma Holdings, Singapore. The other authors do not have any conflicts of interest, except that this project was funded in part by Philips.

Funding

Partially funded by Philips Research North America (A-ETA-0004).

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Open AI Chat GPT Version 4o for grammar checking. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bja.2024.06.030>.

References

1. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med* 2019; **380**: 1347–58
2. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**: 436–44
3. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med* 2019; **25**: 24–9
4. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw* 2005; **18**: 602–10
5. Pinsky MR, Bedoya A, Bihorac A, et al. Use of artificial intelligence in critical care: opportunities and obstacles. *Crit Care* 2024; **28**: 113
6. Pollard TJ, Johnson AEW, Raffa JD, et al. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Sci Data* 2018; **5**, 180178
7. Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data* 2016; **3**, 160035
8. Sessler DI. Perioperative thermoregulation and heat balance. *Lancet* 2016; **387**: 2655–64
9. Thygesen K, Alpert JS, Jaffe AS, et al. Fourth Universal Definition of Myocardial Infarction (2018). *Circulation* 2018; **138**: e618–51

doi: 10.1016/j.bja.2024.06.030