COMBINING HOSPITAL-GRADE CLINICAL DATA AND WEARABLE VITAL SIGN MONITORING TO PREDICT SURGICAL COMPLICATIONS

Robin van de Water[1]*, **Axel Winter**[2], **Max M. Maurer**[2], **Felix A, Treykorn**[1], **Daniela Zuluaga**[2], **Bjarne Pfitzner**[1], **Igor M. Sauer**[2], **Bert Arnrich**[1]
[1]Hasso Plattner Institute, University of Potsdam, Germany
[2]Charité Universitätsmedizin Berlin, Germany

ABSTRACT

Access to the right data at the right time is a substantial hurdle for developing effective machine learning (ML) in healthcare. Many hospitals maintain several isolated patient databases, often leading to incomplete datasets that severely impact the clinical usability of ML prediction systems. Moreover, in visceral surgery, postoperative complications often occur in the nursing ward, where real-time monitoring is non-existent. ML-powered predictive systems are increasingly ineffective as they leave the Intensive Care Unit (ICU) due to data shortage. However, the patient is still at risk of various complications. Our work addresses both issues using a collection framework combining pre-operative, intra-operative, ICU, and general patient parameters. We add a high-resolution continuous vital sign measurement modality collected on a non-intrusive hybrid nursing ward. Using the wearable data, we observe improved prediction accuracy for Surgical Site Infection (SSI). Our work suggests a need for hybrid monitoring after a patient's ICU stay to further ML modeling in clinical settings.

1 INTRODUCTION

The Electronic Health Record (EHR) provides considerable administrative benefits. In addition to efficient administration, the data can be repurposed to improve patient treatment and predict clinical outcomes using ML. Using this retrospective data, several datasets have been developed (e.g., Johnson et al. (2023); Thoral et al. (2021); Hyland (2020); Rodemund et al. (2023)). Often, these datasets mostly contain heavily monitored Intensive Care Unit (ICU) data due to its abundance and ease of extraction (Shillan et al., 2019). Despite the increased popularity and usage of EHR for ML (Stam et al., 2022; Muralitharan et al., 2021; Souwer et al., 2020), there are major limits to retrospective data collection. Due to bias towards ICU data, the monitoring timespan is severely limited, and cohorts often include patients with considerably different treatment histories, which can influence model usability. Additionally, many clinically relevant endpoints are often not available. This means machine learning researchers are restricted to predicting retrospectively defined endpoints (van de Water et al., 2023) that are not always clinically relevant and, therefore, have no impact on improving treatment outcomes.

We, instead, perform a prospective data collection study in visceral surgery to develop an ML-prediction system to target clinically relevant complications. After an operation, patients are stabilized in the ICU for about 24 hours and transferred to the nursing ward afterward. However, complications caused by the surgery often do not occur immediately, but days or even weeks later (Feld et al., 2016; Chiarello et al., 2022). At this stage, the patient is in the nursing ward where, unlike the ICU, continuous monitoring of vital signs is not available. There are several reasons for this seemingly early discharge from the ICU: patient discomfort, such as high noise levels, invasive monitoring, and poorer mobilization — which impedes recovery (Jovanović et al., 2018; Morgan, 2021). In addition, ICU resources are costly, limited, and reserved for critically ill patients (Bagshaw et al., 2020).

*Corresponding author email: robin.vandewater@hpi.de

Our study implements a *hybrid* nursing ward where patients are monitored with minimally intrusive wearable vital signs equipment to address the lack of monitoring. This approach does not have the cost and quality of life disadvantages of prolonged ICU stays. Moreover, we use the monitoring data to improve ML-based risk stratification for complications and create insights into patient evolution.

2 Data collection and processing

Figure 1 contains a high-level data collection and processing pipeline overview. If patients give informed consent, we enroll them in our study. We recognize four stages in the patient care process: patient intake, surgery, ICU stay, and nursing ward stay.

Intake information is recorded with a RedCAP (Patridge & Bardyn, 2018) database clinical trial application. After the surgery and ICU stay, study subjects enter the nursing ward. They are monitored using the Corsano Cardio Watch 1 or 2 (Blok et al., 2021), which monitors photoplethysmography (PPG) at 25 and 32 Hz, respectively. From PPG signals, we can compute features such as heart pulse rate, blood oxygen saturation, and respiration rate (Almarshad et al., 2022). Additionally, we attach the GreenTEG Core body temperature sensor (Verdel et al., 2021) on their chest. This data is fed into the Devicehub database (Appendix B.1). We hypothesize that this non-intrusive, continuous monitoring can provide more information at a critical stage of the patient's stay.

The hospital Patient Data Management System (PDMS) infrastructure uses COPRA6 (Appendix B.2), which stores intra- and post-operative ICU data. General patient characteristics, operations, ward movements, and lab values are collected in i.s.h.med (Appendix B.3). Data is then ingested into an on-premise datalake (Appendix B.4) with Apache Spark (Zaharia et al., 2016) for further processing: matching wearable device streams with patients at the right point in time. We perform further preprocessing and feature extraction on the High-Performance Compute Cluster (HPC). Here, we can directly use the data for ML, analysis, and visualizations.

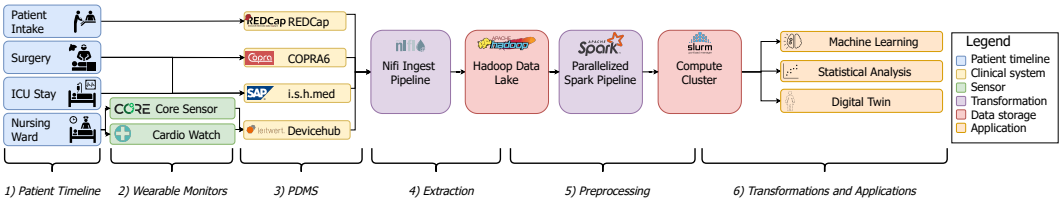

Figure 1: *Data collection and processing pipeline.* From left to right: **1)** simplified time steps of patient treatment process **2)** patient monitors worn on nursing ward, **3)** the respective PDMS involved with data collection and storage, **4)** the extraction pipeline that ingests data from the PDMS using proprietary APIs, **5)** on-premises Hadoop-based data lake **6)** Spark-based transformation pipeline and further preprocessing on compute cluster, **6)** High-Performance Computing infrastructure used to construct ML pipelines, perform patient analytics, and produce visualizations.

Table 1 describes aspects of the currently included patient cohort. The collected dataset contains wearable and clinical data, allowing us to perform real-time risk prediction modeling outside of critical care. To our knowledge, it is the biggest dataset that contains the combination of wearable and clinical modalities (Clifton et al., 2014). We note that our dataset contains considerable amounts of data points per subject. The dataset is, however, sparse over time as we collect different types of data in each stage of the patient's stay. It contains a mix of high-frequency time series, such as ICU and wearable data, as well as events that encapsulate a large amount of information, such as lab results; Figure 2 demonstrates this. Clinicians can use this application to gain further insights into the causes of adverse outcomes and discover spurious data.

3 Experimental results

As the study is still running, we are expanding our cohort dynamically. The primary endpoint of our study is the clinically relevant SSI-III (McLean et al., 2023), a deep organ space infection caused by visceral surgery. As the time series from the included subjects is very sparse, e.g., ICU segments

Table 1: *Description of the collected data.* **Numeric** variables are summarized by *median [IQR]*. **Categorical** variables are summarized by *incidence (%)*.

Years collected	2022-2024	**Surgery System**†	
Average data points per patient	≈ 500,000	Esophagus	42
Number of patients	780* proj. 1,000	Stomach	24
Age at admission (years)	62.0 [53.0, 70.5]	Colorectal	395
Female	235 (30.1%)	Liver	435
Complications		Pancreas	187
Deep Surgical Site Infections (SSI-III)	138 (19.5%)		
Of which occurred on nursing ward	108 (78.3%)		
Time to SSI-III* (days)	8 [5, 10]	**Wearable data**‡	
Hospital death	24 (3.9%)	Corsano Band 1	318
Hospital length of stay (days)	8 [6,12]	Corsano Band 2	324
No. of variables	≈ 200	Core Device	660

*From the moment of intake.
†Patients can receive surgery for multiple systems.
‡ Due to logistic issues in data collection, not every patient has usable wearable data, see discussion.

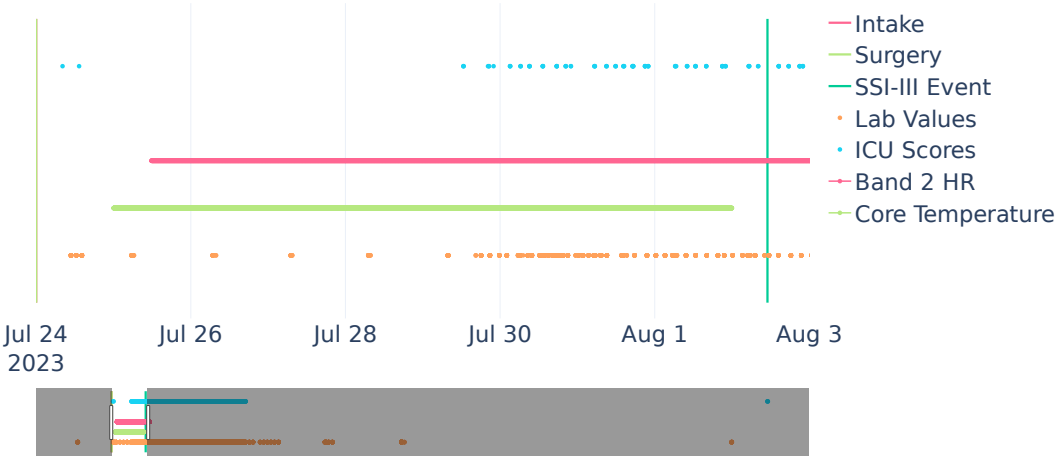

Figure 2: *Interactive digital twin produced using our extraction pipeline and Plotly*[1]*. Individual data points can be selected and inspected.*

contain medical scores such as SOFA, we decided to use a feature extraction pipeline with a non-deep model as these still tend to outperform DL models (Grinsztajn et al., 2022). Features are extracted with clinical assistance to facilitate potential clinical procedure improvements. They are extracted using TSFresh (Christ et al., 2018). Data from the entire registered stay was used for a stay-level prediction: we predict SSI once per stay at a predefined point. Stays that resulted in complications were cut 12 hours before complication time. To prevent the model from associating stay duration with complication, remaining stays were cut at a random timepoint using a Gaussian distribution modeled on the complication length of stay: $\mathcal{N}(\mu_{LoS\,SSI}, \sigma^2_{LoS\,SSI})$, where $LoS\,SSI$ is the period from intake to the occurrence of SSI-III. Moreover, we excluded generated features based on signal length to minimize model dependence on this feature.

XGBoost (Chen & Guestrin, 2016) resulted in the most accurate model using the extracted features when compared to LightGBM (Ke et al., 2017), Logistic Regression, Random Forest, Gradient Boosting (Pedregosa et al., 2011). We compared the inclusion of the wearable data over a nested 5-fold cross-validation with 10 unique seeds. Due to the potential performance decrease, we also tested SMOTE as an oversampling technique, which did not positively impact the results.

Table 2 shows a feature ablation with combinations of three sets of features (F_{ICU}, $F_{General}$, and $F_{Wearable}$; they are separated based on the origin PDMS (respectively, COPRA, i.s.h.med, and Devicehub, see Figure 1), which simulates EHR-based ML modeling with partial information.

[1]`https://plotly.com/graphing-libraries/`, version 5.19

Table 2: *The performance impact of feature sets.* The full feature set was used in every experiment.

Feature set	AUPRC	AUROC
F_{ICU}	26.4±10.0	66.3±8.0
$F_{General}$	41.6±12.6	73.2±6.8
$F_{Wearable}$	28.2±7.5	63.6±7.3
$F_{ICU \cup General}$	39.1±11.4	71.9±7.0
$F_{ICU \cup Wearable}$	29.8±9.0	65.2±6.5
$F_{General \cup Wearable}$	44.0±11.0	74.5±6.1
$F_{ICU \cup General \cup Wearable}$	**46.7±13.1**	**74.7±7.1**

Table 3: *Performance benchmarked per cut-off length*

Hours to SSI-III	AUPRC	AUROC
0	45.5±12.3	78.2±6.1
12	46.7±13.1	74.7±7.1
24	42.0±12.0	75.1±6.2

Table 4 contains the exact feature sets used. Notable is the poor result of just using ICU-based scoring. Moreover, the wearable and lab features demonstrate relatively high performance compared to the combined feature set (a paired t-test over the 25 model results indicated p-values of 1.01×10^{-4} and 8.17×10^{-5} for AUROC and AUPRC, respectively). Combining ICU and Lab features leads to poorer performance than just using Lab features, suggesting that ICU data insufficiently reflects the patient's state around the time of complication and the need for feature selection.

Table 3 shows the influence of the cut-off length before a complication occurs. With our current feature composition and extraction, there are no large differences in performance within the first 24 hours. We hypothesize that this could be due to several reasons; it might indicate that our feature extraction can be improved, and we should harmonize features to reduce the standard deviation.

Feature importance SHAP (Lundberg & Lee, 2017)-based aggregated feature importance is shown in Figure 9. A high value for lab features, such as C Reactive Protein and Gamma Glutamlytransferase, seems to indicate the likelihood of SSI. Features associated with the Corsano band, and to a lesser extent the Core temperature sensor, also indicate some improvement to predictiveness.

4 Discussion

Despite promising results, we recognize the limitations of our work and see opportunities for improvement. Prediction accuracy could likely be improved by extracting more features from available data. The PPG signal in particular lends itself well to various feature extraction techniques (Almarshad et al., 2022). To further improve prediction accuracy, we aim to use grouped time-step generation such as that used in Kuznetsova et al. (2023); this should particularly highlight the importance of wearable data as we collect significant periods of data. Moreover, an online (hourly) prediction model could be tested and potentially used in a future clinical study. During our study, we also recorded wearable signals in the ICU, allowing us to compare the accuracy of noninvasive wearable and invasive clinical monitoring. Finally, process improvements, such as an increased capacity of fully-charged wearables, that ease the process of wearing wearable devices can lead to higher-quality sensor data and minimal loss of data.

Conclusion Our work shows that, although access to digitalized EHR data has drastically increased, we might still be missing data at the right place and time. A hybrid ward that combines EHR information with continuous monitoring of vital signs addresses this and benefits ML in clinical settings such as early detection of complications. This subsequently improves postoperative outcomes. More broadly, medical ML is in need of specialized data collection techniques to bring more reliable systems to practice. We recommend that ML research focuses on problem modeling and data collection to develop accurate prediction systems.

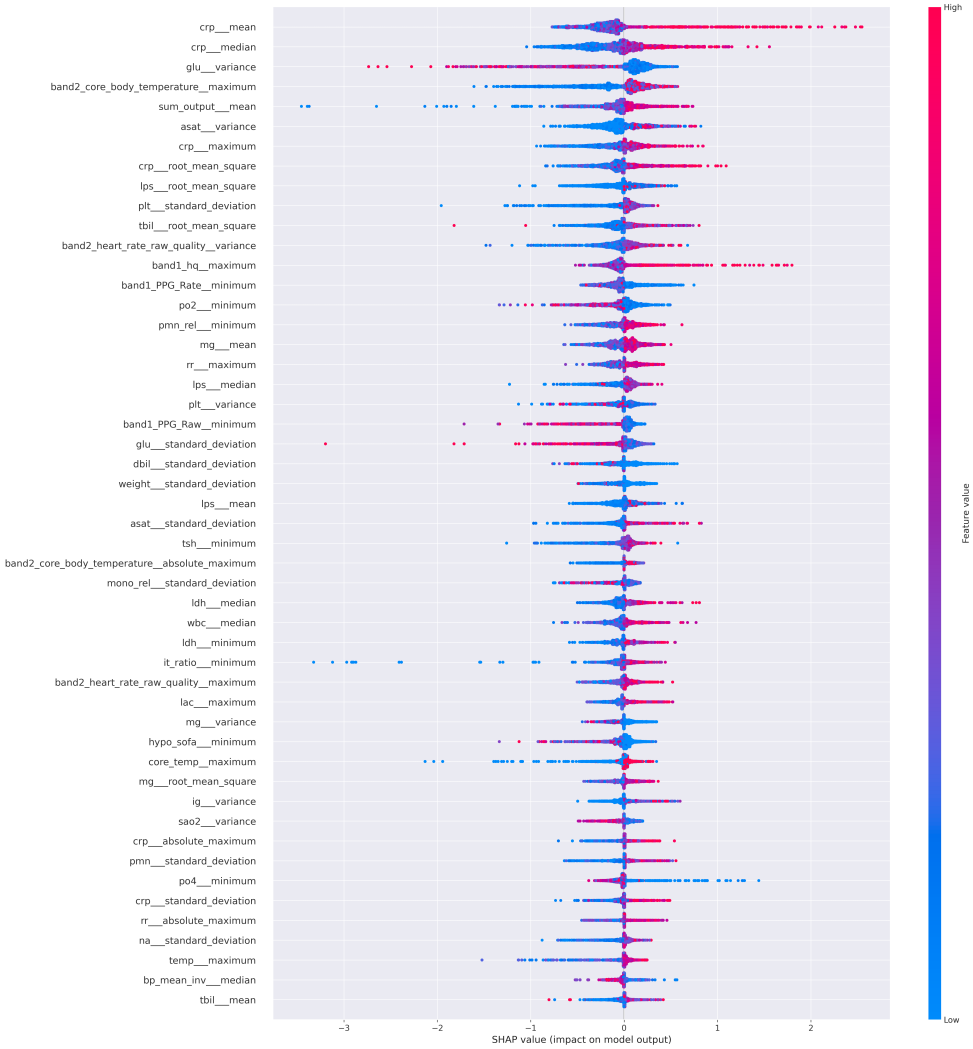

Figure 3: *SHAP values for 12-hour prediction aggregated over all folds and seeds. We show only the 50 most influential features (according to the SHAP calculation).* Appendix E contains additional SHAP visualizations.

5 REPRODUCIBILITY STATEMENT

The dataset used in this study is currently restricted to individuals involved in the ongoing study as these are the conditions signed by participants under informed consent as we recognize. We are working on an agreement to be able to publish a version of the dataset as we recognize the potential value to the community. More information can be found in Appendix C. Code to preprocess the data on an HPC can be found at `https://github.com/rvandewater/cass-preprocessing`. Code to train the models can be found at `https://github.com/rvandewater/CASS-PROPEL` More details about model evaluation are found in Appendix D.

6 ACKNOWLEDGEMENTS

This work and some of the authors (R. van de Water, B. Pfitzner) are funded by the "Gemeinsamer Bundesausschuss (G-BA) Innovationsausschuss" in the framework of "CASSANDRA - Clinical ASSist AND aleRt Algorithms" (project number 01VSF20015).

REFERENCES

Malak Abdullah Almarshad, Md Saiful Islam, Saad Al-Ahmadi, and Ahmed S. BaHammam. Diagnostic Features and Potential Applications of PPG Signal in Healthcare: A Systematic Review. *Healthcare*, 10(3):547, March 2022. ISSN 2227-9032. doi: 10.3390/healthcare10030547. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8950880/.

Sean M. Bagshaw, Dat T. Tran, Dawn Opgenorth, Xiaoming Wang, Danny J. Zuege, Armann Ingolfsson, Henry T. Stelfox, and Nguyen X. Thanh. Assessment of Costs of Avoidable Delays in Intensive Care Unit Discharge. *JAMA Network Open*, 3(8):e2013913, August 2020. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2020.13913. URL https://doi.org/10.1001/jamanetworkopen.2020.13913.

S. Blok, M. A. Piek, I. I. Tulevski, G. A. Somsen, and M. M. Winter. The accuracy of heartbeat detection using photoplethysmography technology in cardiac patients. *Journal of Electrocardiology*, 67:148–157, July 2021. ISSN 0022-0736. doi: 10.1016/j.jelectrocard.2021.06.009. URL https://www.sciencedirect.com/science/article/pii/S0022073621001291.

Noé Brasier, Lukas Geissmann, Miro Käch, Markus Mutke, Bianca Hoelz, Fiorangelo De Ieso, and Jens Eckstein. Device- and Analytics-Agnostic Infrastructure for Continuous Inpatient Monitoring: A Technical Note. *Digital Biomarkers*, 4(2):62–68, August 2020. ISSN 2504-110X. doi: 10.1159/000509279. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7548886/.

Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, KDD '16, pp. 785–794. Association for Computing Machinery, August 2016. ISBN 978-1-4503-4232-2. doi: 10.1145/2939672.2939785. URL https://doi.org/10.1145/2939672.2939785. Place: New York, NY, USA.

Maria Michela Chiarello, Pietro Fransvea, Maria Cariati, Neill James Adams, Valentina Bianchi, and Giuseppe Brisinda. Anastomotic leakage in colorectal cancer surgery. *Surgical Oncology*, 40: 101708, March 2022. ISSN 1879-3320. doi: 10.1016/j.suronc.2022.101708.

Maximilian Christ, Nils Braun, Julius Neuffer, and Andreas W. Kempa-Liehr. Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). *Neurocomputing*, 307:72–77, September 2018. ISSN 0925-2312. doi: 10.1016/j.neucom.2018.03.067. URL https://www.sciencedirect.com/science/article/pii/S0925231218304843.

Lei Clifton, David A. Clifton, Marco A. F. Pimentel, Peter J. Watkinson, and Lionel Tarassenko. Predictive Monitoring of Mobile Patients by Combining Clinical Observations With Data From Wearable Sensors. *IEEE Journal of Biomedical and Health Informatics*, 18(3):722–730, May 2014. ISSN 2168-2194, 2168-2208. doi: 10.1109/JBHI.2013.2293059. URL http://ieeexplore.ieee.org/document/6675775/.

Shara I. Feld, Sarah E. Tevis, Alexander G. Cobian, Mark W. Craven, and Gregory D. Kennedy. Multiple postoperative complications: Making sense of the trajectories. *Surgery*, 160(6):1666–1674, December 2016. ISSN 1532-7361. doi: 10.1016/j.surg.2016.08.047.

G. Gell, P. Schmücker, M. Pedevilla, H. Leitner, J. Naumann, H. Fuchs, H. Pitz, and W. Köle. SAP and Partners: IS-H™ and IS-H*MED™. *Methods of Information in Medicine*, 42(1):16–24, February 2003. ISSN 0026-1270, 2511-705X. doi: 10.1055/s-0038-1634205. URL http://www.thieme-connect.de/DOI/DOI?10.1055/s-0038-1634205. Publisher: Schattauer GmbH.

Léo Grinsztajn, Edouard Oyallon, and Gaël Varoquaux. Why do tree-based models still outperform deep learning on tabular data?, July 2022. URL http://arxiv.org/abs/2207.08815. arXiv:2207.08815 [cs, stat].

Stephanie L Hyland. Early prediction of circulatory failure in the intensive care unit using machine learning. *Nature MedIcIne*, 26:28, 2020.

Alistair E. W. Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J. Pollard, Benjamin Moody, Brian Gow, Li-wei H. Lehman, Leo A. Celi, and Roger G. Mark. MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data*, 10 (1):1, January 2023. ISSN 2052-4463. doi: 10.1038/s41597-022-01899-x. URL https://www.nature.com/articles/s41597-022-01899-x.

Gordana Jovanović, Dea Karaba Jakovljević, and Mirka Lukić-Šarkanović. Enhanced Recovery in Surgical Intensive Care: A Review. *Frontiers in Medicine*, 5, 2018. ISSN 2296-858X. URL https://www.frontiersin.org/articles/10.3389/fmed.2018.00256.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A highly efficient gradient boosting decision tree. In *Advances in neural information processing systems*, volume 30. Curran Associates, Inc., 2017. URL https://papers.nips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html.

Rita Kuznetsova, Alizée Pace, Manuel Burger, Hugo Yèche, and Gunnar Rätsch. On the Importance of Step-wise Embeddings for Heterogeneous Clinical Time-Series, November 2023. URL http://arxiv.org/abs/2311.08902. arXiv:2311.08902 [cs].

Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in neural information processing systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

Kenneth A McLean, Tanvi Goel, Samuel Lawday, Aya Riad, Joana Simoes, Stephen R Knight, Dhruva Ghosh, James C Glasbey, Aneel Bhangu, Ewen M Harrison, and NIHR Global Health Research Unit on Global Surgery. Prognostic models for surgical-site infection in gastrointestinal surgery: systematic review. *British Journal of Surgery*, 110(11):1441–1450, November 2023. ISSN 1365-2168. doi: 10.1093/bjs/znad187. URL https://doi.org/10.1093/bjs/znad187.

Amy Morgan. Long-term outcomes from critical care. *Surgery (Oxford)*, 39(1):53–57, January 2021. ISSN 0263-9319. doi: 10.1016/j.mpsur.2020.11.005. URL https://www.sciencedirect.com/science/article/pii/S0263931920302386.

Sankavi Muralitharan, Walter Nelson, Shuang Di, Michael McGillion, Pj Devereaux, Neil Grant Barr, and Jeremy Petch. Machine Learning–Based Early Warning Systems for Clinical Deterioration: Systematic Scoping Review. *Journal of Medical Internet Research*, 23(2):e25187, February 2021. ISSN 1438-8871. doi: 10.2196/25187. URL https://www.jmir.org/2021/2/e25187.

Alexander Nelde, Markus G. Klammer, Christian H. Nolte, Helena Stengl, Michael Krämer, Regina von Rennenberg, Andreas Meisel, Franziska Scheibe, Matthias Endres, Jan F. Scheitz, and Christian Meisel. Data lake-driven analytics identify nocturnal non-dipping of heart rate as predictor of unfavorable stroke outcome at discharge. *Journal of Neurology*, 270(8):3810–3820, August 2023. ISSN 1432-1459. doi: 10.1007/s00415-023-11718-x. URL https://doi.org/10.1007/s00415-023-11718-x.

Emily F. Patridge and Tania P. Bardyn. Research Electronic Data Capture (REDCap). *Journal of the Medical Library Association : JMLA*, 106(1):142–144, January 2018. ISSN 1536-5050. doi: 10.5195/jmla.2018.319. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5764586/.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011. ISSN 1533-7928. URL http://jmlr.org/papers/v12/pedregosa11a.html.

Niklas Rodemund, Bernhard Wernly, Christian Jung, Crispiana Cozowicz, and Andreas Koköfer. The Salzburg Intensive Care database (SICdb): an openly available critical care dataset. *Intensive Care Medicine*, 49(6):700–702, June 2023. ISSN 1432-1238. doi: 10.1007/s00134-023-07046-3. URL https://doi.org/10.1007/s00134-023-07046-3.

Duncan Shillan, Jonathan A. C. Sterne, Alan Champneys, and Ben Gibbison. Use of machine learning to analyse routinely collected intensive care unit data: a systematic review. *Critical Care*, 23 (1):284, December 2019. ISSN 1364-8535. doi: 10.1186/s13054-019-2564-9. URL https://ccforum.biomedcentral.com/articles/10.1186/s13054-019-2564-9.

Esteban T. D. Souwer, Esther Bastiaannet, Ewout W. Steyerberg, Jan-Willem T. Dekker, Frederiek van den Bos, and Johanna E. A. Portielje. Risk prediction models for postoperative outcomes of colorectal cancer surgery in the older population - a systematic review. *JOURNAL OF GERIATRIC ONCOLOGY*, 11(8):1217–1228, November 2020. ISSN 1879-4068. doi: 10.1016/j.jgo.2020.04.006. tex.eissn: 1879-4076 tex.unique-id: WOS:000583348000005.

Wessel T Stam, Lotte K Goedknegt, Erik W Ingwersen, Linda J Schoonmade, Emma R J Bruns, and Freek Daams. The prediction of surgical complications using artificial intelligence in patients undergoing major abdominal surgery: A systematic review. *Surgery*, 171(4):1014–1021, 2022. ISSN 1532-7361. doi: 10.1016/j.surg.2021.10.002. Type: Journal article.

Patrick J. Thoral, Jan M. Peppink, Ronald H. Driessen, Eric J. G. Sijbrands, Erwin J. O. Kompanje, Lewis Kaplan, Heatherlee Bailey, Jozef Kesecioglu, Maurizio Cecconi, Matthew Churpek, Gilles Clermont, Mihaela van der Schaar, Ari Ercole, Armand R. J. Girbes, and Paul W. G. Elbers. Sharing ICU Patient Data Responsibly Under the Society of Critical Care Medicine/European Society of Intensive Care Medicine Joint Data Science Collaboration: The Amsterdam University Medical Centers Database (AmsterdamUMCdb) Example*. *Critical Care Medicine*, 49(6):e563–e577, June 2021. ISSN 0090-3493. doi: 10.1097/CCM.0000000000004916. URL https://journals.lww.com/10.1097/CCM.0000000000004916.

Robin van de Water, Hendrik Schmidt, Paul Elbers, Patrick Thoral, Bert Arnrich, and Patrick Rockenschaub. Yet Another ICU Benchmark: A Flexible Multi-Center Framework for Clinical ML, June 2023. URL http://arxiv.org/abs/2306.05109. arXiv:2306.05109 [cs].

Nina Verdel, Tim Podlogar, Urša Ciuha, Hans-Christer Holmberg, Tadej Debevec, and Matej Supej. Reliability and Validity of the CORE Sensor to Assess Core Body Temperature during Cycling Exercise. *Sensors*, 21(17):5932, January 2021. ISSN 1424-8220. doi: 10.3390/s21175932. URL https://www.mdpi.com/1424-8220/21/17/5932. Number: 17 Publisher: Multidisciplinary Digital Publishing Institute.

Matei Zaharia, Reynold S. Xin, Patrick Wendell, Tathagata Das, Michael Armbrust, Ankur Dave, Xiangrui Meng, Josh Rosen, Shivaram Venkataraman, Michael J. Franklin, Ali Ghodsi, Joseph Gonzalez, Scott Shenker, and Ion Stoica. Apache Spark: a unified engine for big data processing. *Communications of the ACM*, 59(11):56–65, October 2016. ISSN 0001-0782. doi: 10.1145/2934664. URL https://dl.acm.org/doi/10.1145/2934664.

A Detailed inclusion criteria

Patients were included in the study with informed consent. Their age was 18 years or more. The groups included in the study are: Visceral Surgical Major Resections in the Organ Systems of Liver, Pancreas, Upper and Lower Gastrointestinal Tract according to several OPS groups. Intestine

- Small Intestine Resections: 5-454.0 to 5-545.y
- Reversal of a Double-Barreled Enterostomy: 5-465.0 to 5-465.y
- Restoration of Intestinal Continuity in Terminal Enterostomies: 5-466.0 to 5-466.y
- Partial Resection of the Colon: 5-455.0 to 5-455.y
- Total Colectomy, Variations: 5-456.0 to 5-456.y
- Rectum Resection with Sphincter Preservation: 5-484.0 to 5-484.y
- Rectum Resection without Sphincter Preservation: 5-485.0 to 5-485.y

Upper Gastrointestinal Tract:

- Partial Esophagectomy without Restoration of Continuity: 5-423.0 to 5-423.y
- Partial Esophagectomy with Restoration of Continuity: 5-424.0 to 5-424.y
- (Total) Esophagectomy without Restoration of Continuity: 5-425.0 to 5-425.y
- (Total) Esophagectomy with Restoration of Continuity: 5-426.0 to 5-426.y
- Atypical Partial Gastrectomy: 5-434.0 to 5-434.y
- Partial Gastrectomy (2/3 Resection): 5-435.0 to 5-435.y
- Subtotal Gastrectomy (4/5 Resection): 5-436.0 to 5-436.y
- (Total) Gastrectomy: With Esophagojejunostomy: 5-437.0 to 5-437.y
- (Total) Gastrectomy with Subtotal Esophagectomy: 5-438.0 to 5-438.y

Pancreas Surgery:

- Partial Resection of the Pancreas: 5-524.0 to 5-524.y
- (Total) Pancreatectomy: 5-525.0 to 5-525.y

Liver Surgery:

- Anatomical (Typical) Liver Resection: 5-502.0 to 5-502.y
- Atypical Resections: 5-501.0 to 5-501.y

The following groups of patients are excluded:

- Patients not capable of giving consent
- Patients who have undergone an organ transplant during the same stay
- Performance of an additional intraoperative hyperthermic chemotherapy (HiPEC)

B Systems

B.1 Devicehub

Devicehub is a device-agnostic vital sign recording system (Brasier et al., 2020). It runs on gateways that connect directly to wearable devices using bluetooth. It is developed by the company Leitwert[2] which provides Software As A Service support for the product. In our study, it is used to persist data collected using the Corsano[3] Band 1 and 2 and GreenTEG Core[4] devices

[2] `https://www.leitwert.ch/technology/device-hub/` (accessed 16-02-2024)

[3] `https://corsano.com/` (accessed 16-02-2024)

[4] `https://shop.greenteg.com/core-body-temperature-monitor`(accessed 16-02-2024)

B.2 COPRA

COPRA PDMS (patient data management system)[5] is an Electronic Health Record (EHR) system solution. It enables users to analyze changes and repeated logins, automate the transfer of device values, record billing-relevant data and nursing services, determine approval of the documentation, and contrive drug administration and automated export transition. Features include medical history analysis, prescription management, worklist optimization, case management, scalability, and support SQL server and the .NET framework. The latest version is COPRA 6.

B.3 I.S.H.MED

i.s.h.med (old spelling IS-H*med)(Gell et al., 2003) is a clinical information system produced by Oracle Cerner. SAP IS-H industry solution for healthcare facilities and thus results in a hospital information system. The name comprises "IS" for SAP Industry Solution and "H" for Healthcare. (Gell et al., 2003). According to the manufacturer, the software is in use in more than 500 hospitals and operators.

B.4 HEALTH DATA PLATFORM

The Health Data Platform (HDP) (Nelde et al., 2023) is a central service that enables regulated access to routine and research data from hospital patients. As of Q2 2022, the most important systems developed for HDP include the hospital information system, laboratory information system, radiology information system, and patient data management system for intensive care medicine. As a result of this indexing, around 39 million diagnoses, 17 million procedures, and 666 million laboratory values for 4 million patients with 14 million outpatient and inpatient cases, as well as other clinical parameters, are available for evaluation by the HDP. The HDP harmonizes this data based on common data models and international interoperability standards.[6] .

C DETAILED DATA CHARACTERISTICS

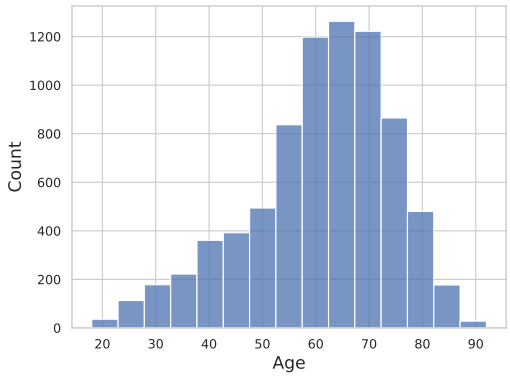

Figure 4: Age distribution of cases

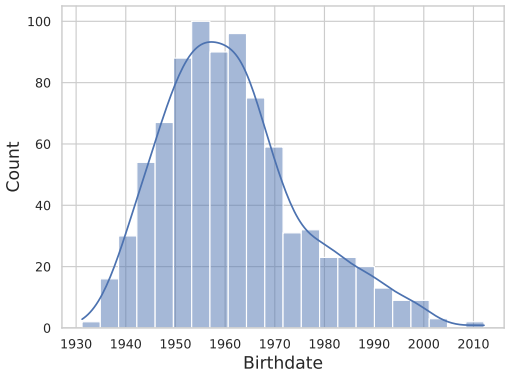

Figure 5: Birthdate of patients

[5]information adapted from https://discovery.hgdata.com/product/copra-pdms(accessed 16-02-2024)

[6]information adapted from https://health-data.charite.de/faq/start (accessed 16-02-2024)

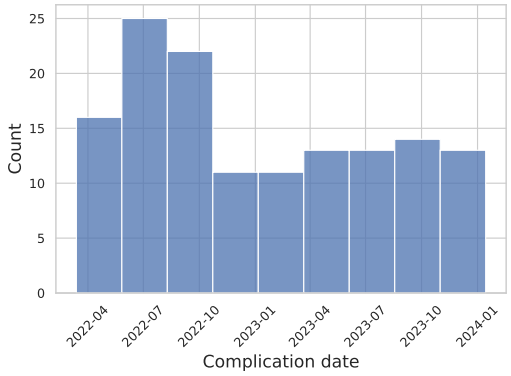

Figure 6: Complication distribution of cases

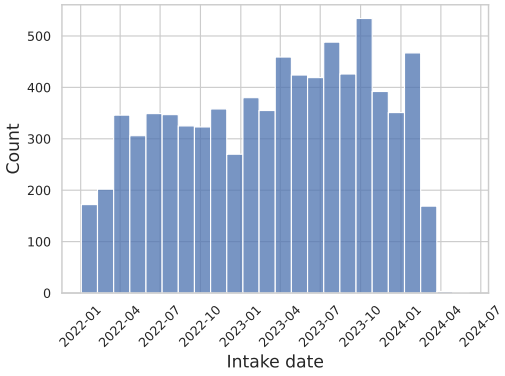

Figure 7: Intake distribution for cases

Table 4: Origin of the modalities recorded per patient. S: static data; T: time-series data

PDMS	Table	Description	Type	Pre-OP	OP	Post-OP	Nursing ward	Integrated*
REDCAP[1]	complications	Complications including SSI.	S		✓	✓	✓	✓
	register_export	Study details and wearable intervals	S		✓	✓	✓	✓
ISHMED[2]	patient_details	Basic statistics for patient	S	✓				✓
	case	Intake data and release dates for patients	S	✓				✓
	lab_values	Various lab measurements	T	✓	✓	✓	✓	✓
	procedure	(Surgical) procedures	S	✓	✓			✓
	diagnosis	Diagnosis details	S	✓				✓
	movement	Movement between wards	T	✓	✓	✓	✓	✗
COPRA6[3]	scores	Common ICU scores and composite measurements	T		✓			✓
	observation	Vital sign observation and scores	T		✓	✓		✓
	medication	Detailed medication history	T	✓	✓	✓	✓	✗
	fluid_balance	Fluids going into/out of the patient	T		✓	✓		✓
	therapy	Machines that the patient is connected to	T		✓	✓		✗
DEVICEHUB[4]	corsano v1/v2	Real-time vital sign monitoring	T				✓	✓
	greenteg core	Real-time temperature monitoring	T				✓	✓

* Indicates the current
[1] Study information
[2] Features denoted as $F_{General}$
[3] Features denoted as F_{ICU}
[4] Features denoted as $F_{Wearable}$

Table 5: Lab values and their counts

	Full Name	Name	Unit	Count
0	Alanin Aminotransferase	alat	U/l	20,571
1	Albumin	alb	g/l	7,512
2	Alkaline Phosphatase	alp	U/l	17,644
3	Amylase	ams	U/l	1,278
4	Partial Thromboplastin Time	aptt	sec	1,988
5	Aspartate Aminotransferase	asat	U/l	20,543
6	Antithrombin	at	%	2,777
7	Basophils	baso	/nl	6,287
8	Basophils Relative	baso_rel	%	5,869
9	Base Excess	be	mmol/l	62,886
10	Calcium	ca	mmol/l	35,945
11	Creatine Kinase	ck	U/l	1,248
12	Creatine Kinase MB	ck_mb	U/l	485
13	Chloride	cl	mmol/l	32,946
14	Carboxyhemoglobin	cohb	%	31,647
15	Creatinine	cr	mg/dl	1,184
16	C Reactive Protein	crp	mg/l	18,747
17	Cystatin C	cys_c	mg/l	171
18	D Dimer	d_dim	mg/l	90
19	Direct Bilirubin	dbil	mg/dl	2,053
20	Erythroblasts	ebl	/nl	4,239
21	Erythroblasts Relative	ebl_rel	%	159
22	Eosinophils	eos	/nl	6,296
23	Eosinophils Relative	eos_rel	%	5,886
24	Iron	fe	µmol/l	1,744
25	Ferritin	fer	µg/l	1,492
26	Fibrinogen	fg	g/l	3,239
27	Fraction of Inspired Oxygen	fio2	%	32,257
28	Glutamate Dehydrogenase	gdh	U/l	1,094
29	Gamma Glutamyltransferase	ggt	U/l	19,848
30	Glucose	glu	mg/dl	44,628
31	Hemoglobin	hb	g/dl	31,776
32	Glycated Hemoglobin	hba1c	%	990
33	Bicarbonate	hco3	mmol/l	63,286
34	Hematocrit	hct	%	31,641
35	High Density Lipoprotein	hdl	mg/dl	948
36	Deoxyhemoglobin	hhb	%	31,342
37	Haptoglobin	hp	g/l	207
38	Indirect Bilirubin	ibil	mg/dl	410
39	Immature Granulocytes	ig	/nl	6,258
40	Immature Granulocytes Relative	ig_rel	%	5,498
41	Immunoglobulin A	iga	g/l	245
42	Immunoglobulin E	ige	kU/l	30
43	Immunoglobulin G	igg	g/l	254
44	Immunoglobulin M	igm	g/l	242
45	International Normalized Ratio	inr	-	18,652
46	I/T Ratio	it_ratio	-	5,708
47	Potassium	k	mmol/l	49,699
48	Lactate	lac	mg/dl	31,151
49	Lactate Dehydrogenase	ldh	U/l	6,332
50	Low Density Lipoprotein	ldl	mg/dl	927
51	Lipase	lps	U/l	15,326
52	Lymphocytes	lym	/nl	6,302

Continued on next page

Table 5: Lab values and their counts

	Full Name	Name	Unit	Count
53	Lymphocytes Relative	lym_rel	%	5,906
54	Mean Corpuscular Hemoglobin	mch	pg	25,594
55	Mean Corpuscular Hemoglobin Concentration	mchc	g/dl	25,560
56	Mean Corpuscular Volume	mcv	fl	25,573
57	Methemoglobin	methb	%	31,654
58	Magnesium	mg	mmol/l	3,719
59	Monocytes	mono	/nl	6,293
60	Monocytes Relative	mono_rel	%	5,898
61	Mean Platelet Volume	mpv	fl	25,087
62	Myelocytes	myelo	%	194
63	Sodium	na	mmol/l	49,369
64	Ammonia	nh3	µmol/l	117
65	N Terminal Pro B Type Natriuretic Peptide	nt_probnp	ng/l	449
66	Oxyhemoglobin	o2hb	%	31,707
67	Phosphorus	p	mmol/l	92
68	Pseudocholinesterase	pche	kU/l	823
69	Carbon Dioxide Partial Pressure	pco2	mmHg	31,590
70	Procalcitonin	pct	µg/l	2,738
71	Potential of Hydrogen	ph	-	33,107
72	Platelets	plt	/nl	25,568
73	Neutrophils	pmn	/nl	6,279
74	Neutrophils Relative	pmn_rel	%	5,460
75	Oxygen Partial Pressure	po2	mmHg	31,672
76	Phosphate	po4	mmol/l	5,277
77	Protein	pro	g/l	2,014
78	Prothrombin Time	pt	NaN	362
79	Quick Value	quick	%	18,237
80	Erythrocytes	rbc	/pl	25,582
81	Red Cell Distribution Width	rdw	%	25,517
82	Reticulocytes	rtic	/nl	1,998
83	Schistocytes	schisto	%	67
84	Oxygen Saturation	so2	%	31,723
85	Temperature	t	°C	32,661
86	Total Bilirubin	tbil	mg/dl	50,847
87	Total Cholesterol	tc	mg/dl	1,275
88	Total Triglycerides	tg	mg/dl	1,396
89	Transferrin	trans	g/l	1,639
90	Thyroid Stimulating Hormone	tsh	mU/l	5,601
91	Urea	urea	mg/dl	19,535
92	25 OH Vitamin D3	vd25	nmol/l	310
93	Leukocytes	wbc	/nl	25,586

Table 6: ICU observations and and the amount

	Name	Count
0	bi_sofa	5,684
1	crea_sofa	5,684
2	fio2	452,963
3	gcs_sofa	5,684
4	height	1,273
5	hr	1,137,862

Continued on next page

Table 6: ICU observations and and the amount

	Name	Count
6	hypo_sofa	5,684
7	pao2_sofa	5,684
8	rr	308,152
9	sao2	570,031
10	sofa	5,687
11	temp	373,436
12	thromb_sofa	5,684
13	vital_IBP	1,305,951
14	vital_NBP	750,909
15	weight	3,511

Table 7: ICU scores and the amount

	Name	Count
0	APACHE2	9,319
1	CO_Score_CAM_ICU	24,579
2	E-NRS	448
3	Frailty	575
4	GCS	52,327
5	NIHSS	208
6	P-RISK	30,928
7	Patient_Score_BSAS	8
8	Patient_Score_DDS8	4,518
9	Patient_Score_ICDSC	5,949
10	RASS	30,805
11	SAPS2	52,827
12	SOFA	21,595
13	Score_BPS	6,510
14	Score_Delir	13,073
15	Score_NAS	6,161
16	Score_SU_Neurostatus	6
17	TISS10	9,974

D MODEL EVALUATION

A double-nested cross-validation setup was used for all experiments. The hyperparameters can be seen in Table 8. The model evaluation code is available at: `https://github.com/rvandewater/CASS-PROPEL`.

Table 8: Used hyperparameter ranges for the XGBoost model used for all experiments.

Hyperparameter	Range
learning_rate	0.005, 0.01, 0.1, 0.3, 0.5, 0.7, 1
colsample_bytree	0.1, 0.25, 0.5, 0.75, 1.0
n_estimators	50,100,250, 500, 750
min_child_weight	1,0.5
max_depth	3, 5, 10, 15

E EXTENDED SHAP VISUALIZATIONS

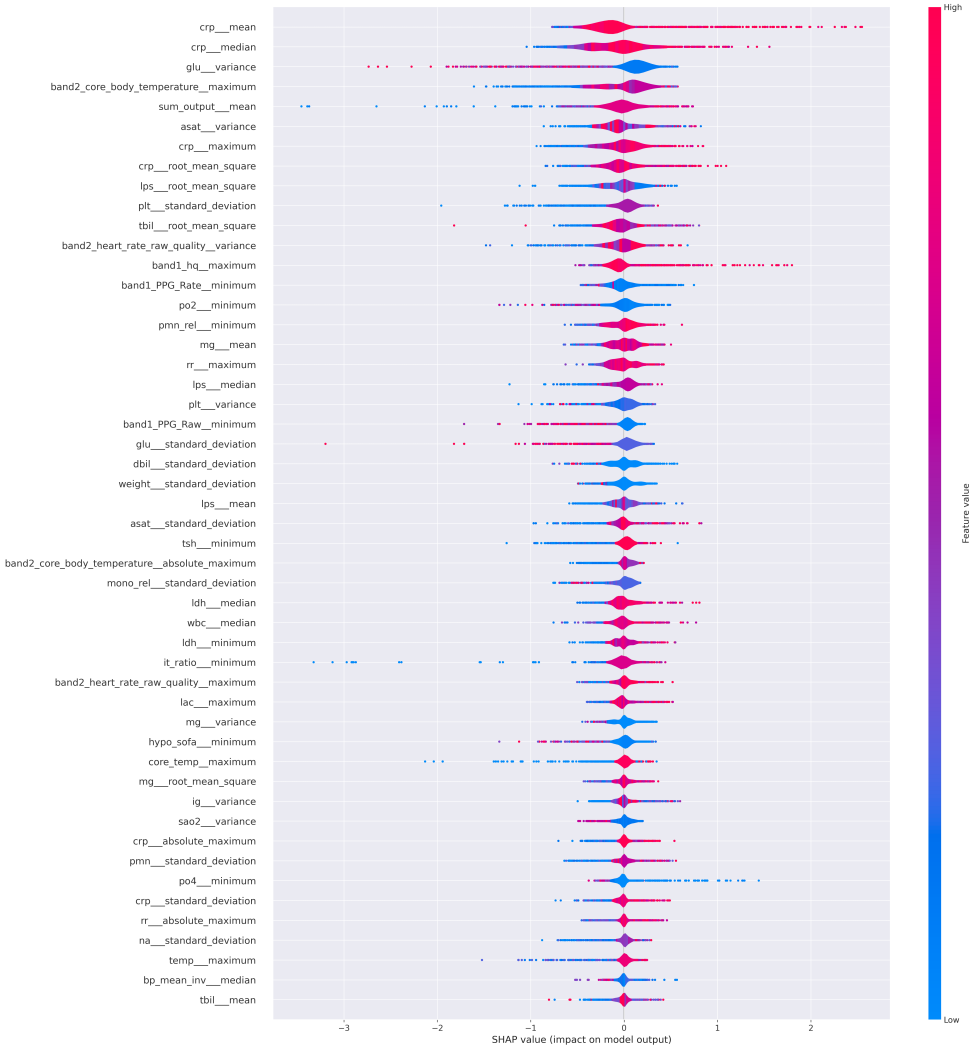

Figure 8: *SHAP values for 12-hour prediction aggregated over all folds and seeds. We show only the 50 most influential features (according to the SHAP calculation).*

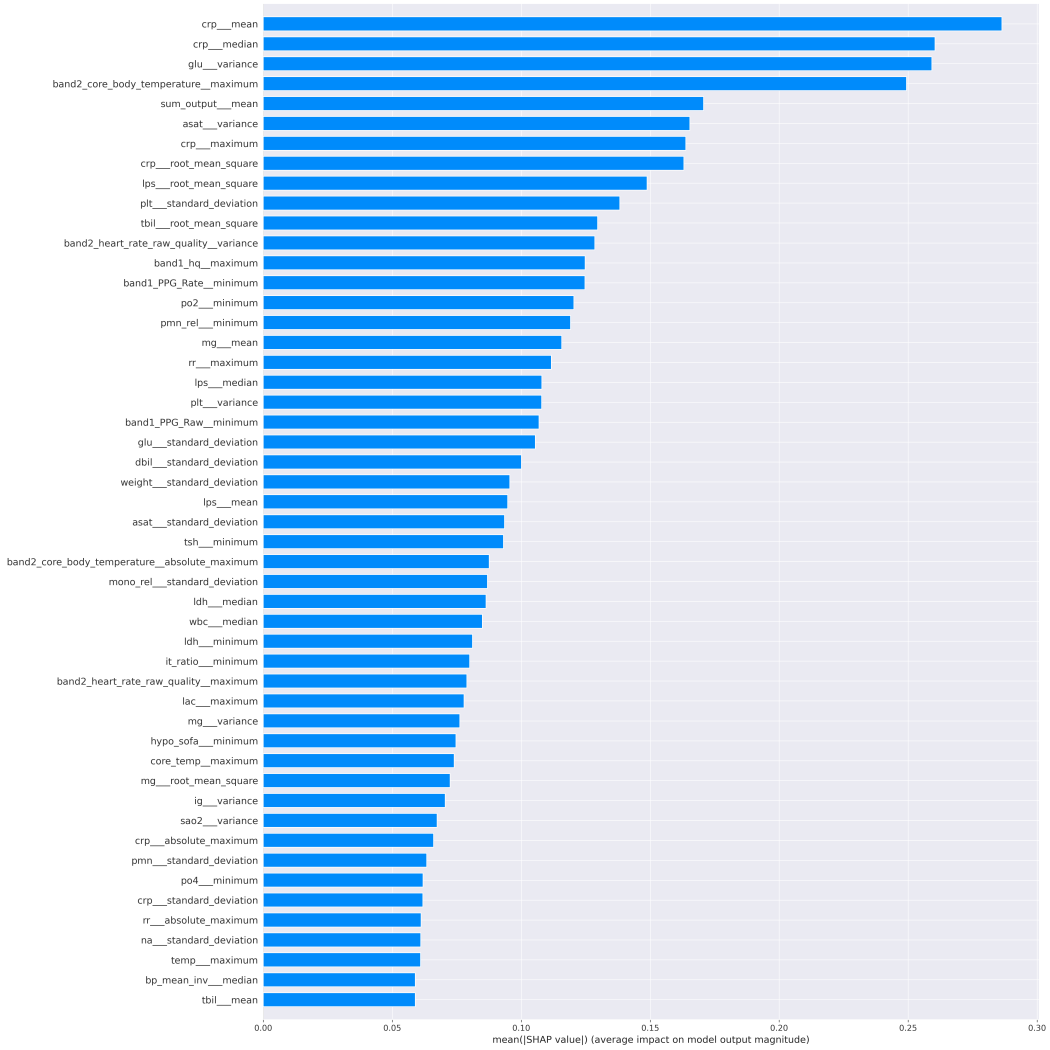

Figure 9: *SHAP values for 12-hour prediction aggregated over all folds and seeds. We show only the 50 most influential features (according to the SHAP calculation).*