# Deployment of an On-the-Edge Clinical Decision Support System in Neonatal Intensive Care Units

Meng Chen[1], Alain Beuchée[1], Fabrice Tudoret[1], Arnaud Coursin[1], Pheng Ho[2], Alfredo I Hernández[1]

[1] Univ Rennes, CHU Rennes, Inserm, LTSI-UMR 1099, Rennes, France
[2] Philips Medizin Systeme Böblingen GmbH, Böblingen, Germany

## Abstract

*Preterm infants require continuous monitoring prior to discharge from neonatal intensive care units (NICU). Conditions such as icterus and sepsis are life-threatening and may cause long-term consequences for this population. The deployment of an AI-based clinical decision support system (CDSS) facilitating early prediction, diagnosis, and intervention in the NICU setting is challenging yet promising. The objective of this work is to design, implement, deploy, and technically evaluate a CDSS that integrates quasi-real-time signal processing chains and AI models on the edge in a clinical context.*

*The proposed system consisting of data transmission, pseudonymization, data fusion, processing, and inference was deployed at the University Hospital of Rennes in Jan 2023. During the first six months of deployment, the service continuously received monitoring signal data from 138 neonates, processing live data and generating bilirubin level estimations at a temporal resolution of 15 minutes. Results on the stability and robustness of the service are presented. To our knowledge, this is the first description of a multi-source, on-the-edge CDSS deployed in a NICU scenario for patient-specific early detection of high-risk events. This proof-of-concept is particularly encouraging and further prospective evaluations on clinical performance are warranted.*

## 1. Introduction

Neonates, especially preterm infants who are born before a gestational age (GA) of 37 weeks, are highly vulnerable, requiring continuous monitoring prior to discharge from neonatal intensive care units (NICU). Conditions such as icterus, apnea, bradycardia, and sepsis pose life-threatening challenges with potential long-term consequences for this population. Although difficult to identify, these pathologies could greatly benefit from early prediction and intervention [1, 2].

In recent decades, clinical decision support systems (CDSS) have emerged as valuable technologies for managing complex medical conditions. A range of CDSS applications are focused in neonatal care, including the management of hyperbilirubinemia, medication, nutrition optimization, and risk estimation for morbidity, mortality, and sepsis [3], such as Artemis [4], Baby Steps [5], Etiometry [6, 7] and iNICU [8, 9]. With the progress in research and the evolution of technical resources within NICUs, establishing an AI-based clinical decision support system that facilitates early intervention for preterm neonates is challenging yet promising.

Furthermore, the development and deployment of on-the-edge CDSS integrated into hospitals and ICUs gained increasing attention. Given the surge in medical data, resource limitations, and the need for real-time, low-latency, and bandwidth-efficient processing, working on the edge is becoming an original and appealing direction.

The objective of this work is to design, implement, deploy, and technically evaluate a CDSS integrating signal processing and machine learning inference models, allowing for quasi-real-time processing and fusion of high-resolution monitoring time series and hospital information system (HIS) data. As a pilot use case, data from 138 neonates born in the first six months of 2023 in an authorized NICU of Rennes University Hospital (CHU) were processed, and the technical performance of the system in terms of stability and resource consumption was quantitatively evaluated.

## 2. Methods

### 2.1. System architecture

As illustrated in Fig. 1, the proposed on-the-edge CDSS architecture consists of three main components: 1) NICU monitors connected to Philips' Data Warehouse Connect (DWC), 2) Electronic Health Records (EHR) system acquiring data from other devices and the HIS and 3) a virtual machine (VM) deployed into a restricted network for holding our CDSS.

The flow of data goes with two parallel pathways (blue

 ISSN: 2325-887X DOI: 10.22489/CinC.2023.061

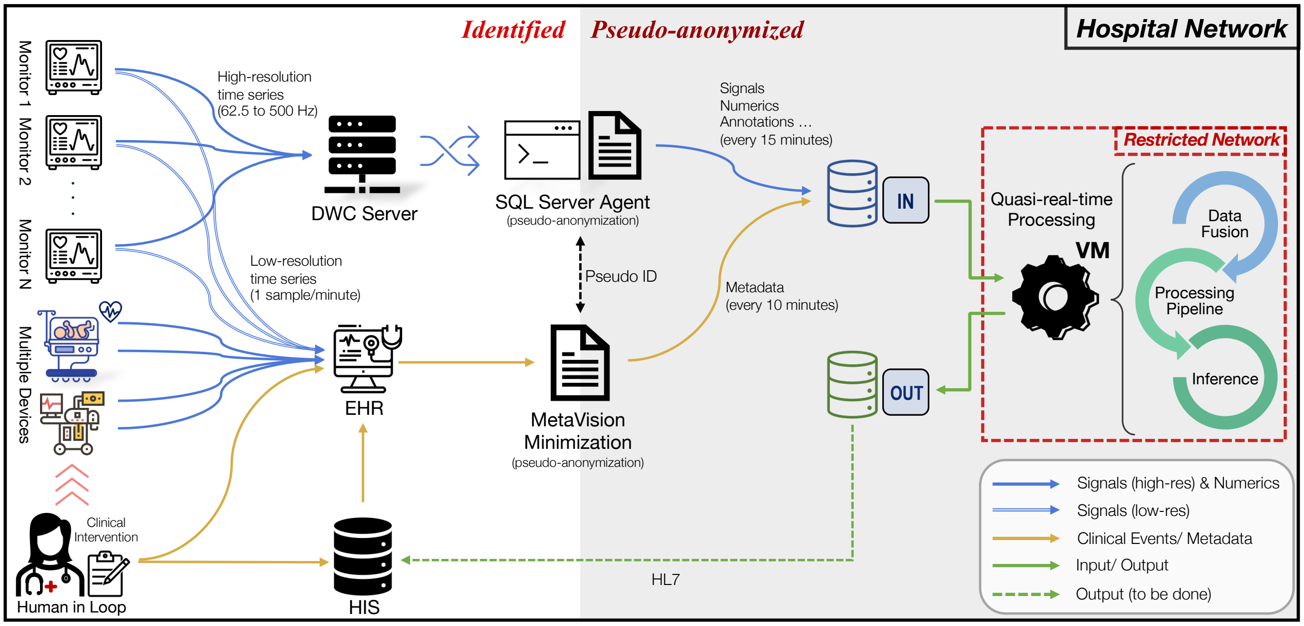


Figure 1: Architecture of the proposed on-the-edge system for quasi-real-time clinical decision support in NICU.

and yellow arrows in Fig. 1). On one hand, high-resolution time series continuously monitored by bedside monitors as well as the structured data (such as instant heart rates, annotations, alerts, etc.) computed by the monitors' built-in algorithms are stored at the DWC server. A Philips-implemented custom SQL agent is configured to automatically export the data between the last 20 minutes to the last 5 minutes, by every 15 minutes. On the other hand, clinical metadata including demographics and laboratory test results from other devices and HIS, stored at the EHR, are selectively exported every 10 minutes. Both branches perform data pseudonymization with a common ID and then copy current data segments to a shared mount point (IN).

The processing unit is activated on a first-in-first-out (FIFO) basis once a new batch of data segments arrives at the IN (i.e., 15 minutes). The processed data segments are erased to free up storage before the output is written to another shared point (OUT) at the end of the workflow.

## 2.2. Core processing unit

Fig. 2 shows the system's core processing engine comprising data fusion, processing, and model inference, with two mount points as input and output, i.e., IN and OUT.

First, data from two sources (DWC and EHR) are fused by ID matching and time synchronization steps, once the current data segment successfully passes the readiness check, it will proceed to the next module. Next comes the signal processing pipeline that transforms multi-source structured data into clinically relevant features. In this first version of deployment, instead of detecting each cardiac beat (QRS complexes) from the acquired ECG, such as performed in our previous works [10], we directly exploit the instantaneous heart rates generated by the monitors to construct the RR series. For RR correction, a multi-step approach of logic rules based on pathology and rhythm correction is then applied to the RR time series to automatically reject or correct possible artifacts and errors. Then alterations over time of the mean and variance of the corrected series are estimated and signal stationarity is analyzed [11]. Afterward, a set of heart rate variability (HRV) [12] describing cardiovascular functions modulated by the autonomic nervous system are extracted from the segments, including eight time-domain features [mean; median; SD; RMSSD; skewness; kurtosis; deceleration capacity and acceleration capacity], five frequency-domain features [power in the low-frequency spectral band (LF, 0.02-0.2 Hz); normalized LF, LFnu; power in the high-frequency band (HF, 0.2-2 Hz); normalized HF, HFnu; and LF/HF ratio]. In addition, we also utilized five features from nonlinear analyses including SD1 and SD2 from the Poincaré plot, sample entropy, and the detrended fluctuation analysis ($\alpha 1$ and $\alpha 2$).

The final block is carried out by a previously trained ML estimator, which is activated in inference mode to make predictions as a plug-and-play plug-in in the PICKLE format, allowing for easy generalization of the tool to other inference applications.

It is worth mentioning that we configured additional data quality validation along the processing chain, which could guarantee a high signal-to-noise ratio and reliability of data analyzed from such complex and heterogeneous data sources in the context of long-term monitoring in NICU.

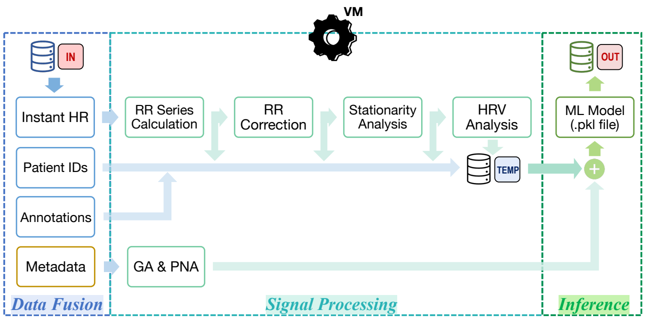


Figure 2: Diagram of the core data processing unit.

## 3. Results

The proposed system was successfully deployed and implemented at Rennes University Hospital in January 2023. To date, the service has been running continuously for several months, constantly receiving data from a total of 138 authorized neonates [GA (median, IQR): 35.5 (30.3; 39.4) weeks; Birth Weight (median, IQR): 2240 (1341; 3211) grams] born between the first six months of 2023 during their stay in NICU. As a first use case, we embedded a baseline random forest regressor trained for bilirubin level estimation as the inference model, and the system processes live data and then generates estimations for each baby with a time resolution of 15 minutes while minimizing computation time and memory usage.

Table 1: System technical performance.

| | Memory Usage* | Execution Time† | System Latency† |
|---|---|---|---|
| Max. | 206.4 MB | 5.56 s | 29.4 min |
| Min. | 168.5 MB | 0.04 s | 20.1 min |
| **Mean** | 187.0 MB | 2.36 s | 26.5 min |
| **Std.** | 8.84 MB | 0.45 s | 1.95 min |
| Median | 187.5 MB | 2.40 s | 26.8 min |

*Statistics are calculated from June 9, 2023 to June 19, 2023.
†Statistics are calculated for the 24 hours from the noon of June 18, 2023 to the noon of June 19, 2023.

Table 1 summarizes the technical performance of the system measured in given periods, in terms of memory usage, execution time, and system latency. As shown, the proposed VM-based system has a low memory footprint, using merely 187.0±8.84 MB of memory during execution.

Execution time refers to the time it takes for one patient's qualified data segment to completely undergo quasi-real-time processing, starting with the arrival of required files at the IN, going through the processing chain, outputting the results, and then being released. During the 24 hours counted, a total of 3293 data segments were qualified and the mean execution time is 2.36±0.45 seconds.

System latency consists of acquisition delay (the time required to enter data from the monitors to the DWC database), database query delay (inherent 20 minutes due to SQL agent configuration), transmission delay (from the DWC database to the IN mount point), processing delay (execution time), and waiting time (the time that data waits for resources in the system to be processed or responded to). The minimum overall latency in the considered period is 20.1 minutes, indicating the possibility of the whole processing being finalized immediately after the data arrival.

Fig. 3 visualizes the execution time and latency over a 24-hour time span, where the x-axis is the timeline divided into every 15 min and the y-axis is the response time in seconds for one single data batch. Generally, the majority of data segments that are output to the OUT (red scatter points) exhibit processing times ranging from 2 to 3 seconds. Observing the timeline, data arrives every 15 minutes, and each batch of data transmission and processing is typically concluded within five minutes. This indicates that the system possesses sufficient capacity to achieve higher concurrency and throughput to accommodate more resource-intensive processing tasks.

## 4. Discussion

To our knowledge, this is the first description of a multi-source, on-the-edge CDSS deployed in a NICU scenario. The quantitative technical performance estimations of this prototype show the feasibility of the deployment and real-time execution of the proposed architecture in a hospital. The proposed VM-based architecture highlights the features of usability, adaptability, and scalability, as well as its ability to be implemented effectively. In addition, the pipeline can be easily generalized by replacing the ".pkl" object model for inference. Integration of our previously published signal processing chains and ML models in this field [13] can be performed in this way.

The main limitation of the proposed system is related to the inference model embedded in the current version, which was trained by data from another project without fine-tuning. The evaluation of the generalization of this model to this particular application has not yet been performed. Further prospective evaluations on clinical performance are thus warranted (protocol currently running). Besides, alert and intervention mechanisms can be added, by finalizing the close-loop from OUT to HIS, such as advising phototherapy when estimated bilirubin levels exceed a quantitative threshold for a given postnatal age.

## 5. Conclusion

The proposed system consisting of data transmission, pseudonymization, data fusion, processing and inference was deployed at the University Hospital of Rennes in Jan 2023. The technical performance of the system in terms

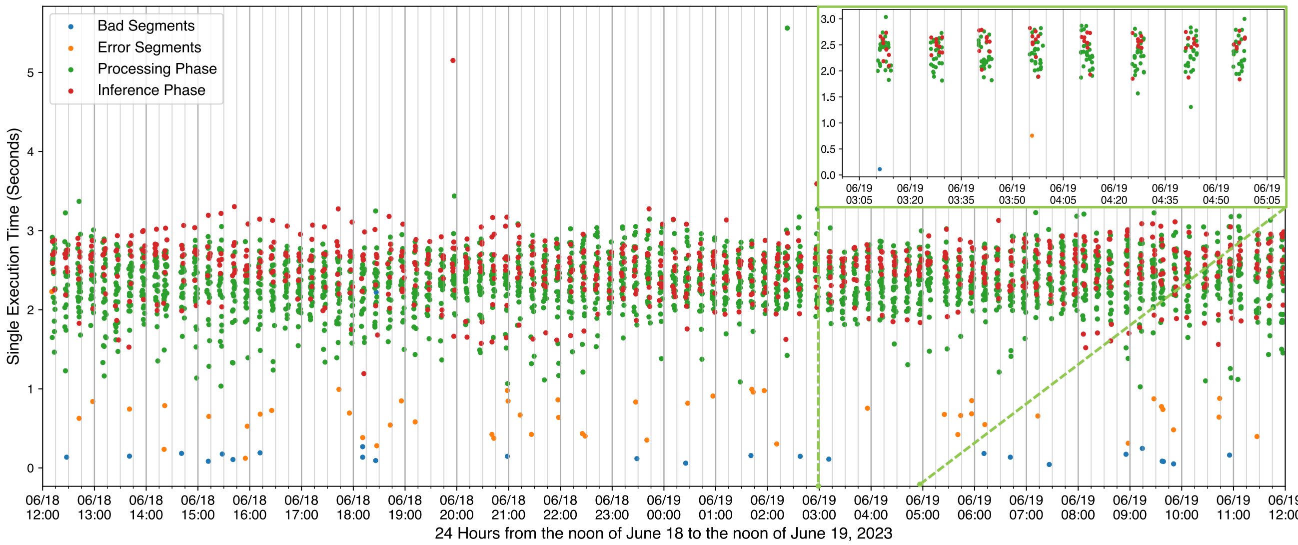


Figure 3: Visualization of execution time and system latency.

of stability and resource consumption has been quantitatively evaluated and a robust and satisfactory configuration was obtained. This proof-of-concept is a solid first step to initiate concrete on-the-edge clinical applications in the proposed platform, which can accommodate AI methods for patient-specific early detection of high-risk events by exploiting the dynamical properties of multivariate and multi-source longitudinal health data.

## Acknowledgments

This work was partly funded by the China Scholarship Council.

## References

[1] Qattea I, Farghaly MAA, Elgendy M, Mohamed MA, Aly H. Neonatal hyperbilirubinemia and bilirubin neurotoxicity in hospitalized neonates: Analysis of the US Database. Pediatr Res 2022;91(7):1662–1668.

[2] Poets CF. Intermittent hypoxia and long-term neurological outcome: How are they related? Seminars in Fetal and Neonatal Medicine 2020;25(2):101072.

[3] Rao A, Palma J. Clinical decision support in the neonatal ICU. Seminars in Fetal and Neonatal Medicine 2022; 27(5):101332.

[4] Bressan N, James A, McGregor C. Trends and opportunities for integrated real time neonatal clinical decision support. In Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics. ISSN 2168-2208, 2012; 687–690.

[5] Spitzer AR, Ellsbury D, Clark RH. The Pediatrix BabySteps® Data Warehouse — A Unique National Resource for Improving Outcomes for Neonates. The Indian Journal of Pediatrics 2015;82(1):71–79.

[6] Meyeroff E, Tremoulet P. Etiometry's t3 heuristic evaluation. Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care 2021; 10(1):37–41.

[7] Asfari A. Artificial intelligence role and clinical decision support system extubation readiness trail and etiometry scoring system. Biomedical Journal of Scientific Technical Research 2021;35(1):27291–27293.

[8] Singh H, Mallaiah R, Yadav G, Verma N, Sawhney A, Brahmachari SK. ichrcloud: web & mobile based child health imprints for smart healthcare. Journal of Medical Systems 2018;42(1):14.

[9] Singh H, Kaur R, Gangadharan A, Pandey AK, Manur A, Sun Y, Saluja S, Gupta S, Palma JP, Kumar P. Neo-bedside monitoring device for integrated neonatal intensive care unit (inicu). IEEE Access 2018;7:7803–7813.

[10] Doyen M, Ge D, Beuchée A, Carrault G, I. Hernández A. Robust, real-time generic detector based on a multi-feature probabilistic method. PLoS ONE 2019;14(10):e0223785.

[11] Beuchée A, Hernández AI, Duvareille C, Daniel D, Samson N, Pladys P, Praud JP. Influence of Hypoxia and Hypercapnia on Sleep State-Dependent Heart Rate Variability Behavior in Newborn Lambs. SLEEP 2012;.

[12] Thu TNP, Hernández AI, Costet N, Patural H, Pichot V, Carrault G, Beuchée A. Improving methodology in heart rate variability analysis for the premature infants: Impact of the time length. PLOS ONE 2019;14(8):e0220692.

[13] Montazeri Ghahjaverestan N, Shamsollahi MB, Ge D, Beuchée A, Hernández AI. Apnea bradycardia detection based on new coupled hidden semi Markov model. Med Biol Eng Comput 2021;59(1):1–11.

Address for correspondence:

Alfredo I. Hernández
Univ Rennes, Inserm, LTSI-UMR 1099, F-35000 Rennes, France
alfredo.hernandez@univ-rennes.fr