Exploring Computational Techniques in Pre-processing Neonatal Physiological Signals for Detecting Adverse Outcomes: A Scoping Review

Jessica Rahman, Aida Brankovic, Mark Tracy, Sankalp Khanna

Submitted to: Interactive Journal of Medical Research
on: March 02, 2023

Table of Contents

Exploring Computational Techniques in Pre-processing Neonatal Physiological Signals for Detecting Adverse Outcomes: A Scoping Review

Jessica Rahman[1] BSc, PhD; Aida Brankovic[2] PhD; Mark Tracy[3] PhD; Sankalp Khanna[2] PhD

[1]CSIRO Australian e-Health Research Centre, Australia Brisbane AU
[2]Neonatal Intensive Care Unit, Westmead, Sydney, Australia Sydney AU

Corresponding Author:
Jessica Rahman BSc, PhD

Abstract

Background: Computational signal pre-processing is a prerequisite for developing data-driven predictive models for clinical decision support. Thus, identifying the best practices that adhere to clinical principles is critical to ensure transparency and reproducibility which will drive clinical adoption.

Objective: This review focuses on the Neonatal intensive care unit (NICU) setting and summarises the state-of-the-art computational methods used for pre-processing neonatal clinical physiological signals for the development of machine learning models in predicting the risk of adverse outcomes.

Methods: Five databases (Pubmed, Web of Science, Scopus, IEEE, ACM Digital Library) were searched using a combination of keywords and MeSH terms. 3,585 papers from the year 2013 to January 2023 were identified based on the defined search terms and inclusion criteria. After removing duplicates, 2,994 papers were screened by title and abstract, and 81 were selected for full-text review. Of these, 52 were eligible for inclusion in the detailed analysis.

Results: The papers included in the review were heterogeneous in design and the selection of adverse outcomes modelled. We found a partial or complete lack of transparency in reporting the setting and the methods used for signal pre-processing. This includes reporting methods to handle missing data, segment size for considered analysis, and details regarding the modification of the state-of-the-art methods for physiological signal processing to align with the clinical principles for neonates.

Conclusions: The review found heterogeneity in techniques and inconsistent reporting of parameters and procedures used for pre-processing neonatal physiological signals, which is necessary to confirm their adherence to clinical practices, usefulness and choice of the best practices. Improving this aspect will ensure transparent reporting and hence facilitate the interpretation and reproducibility of the studies as well as accelerate their clinical adoption. Clinical Trial: N/A

(JMIR Preprints 02/03/2023:46946)
DOI: https://doi.org/10.2196/preprints.46946

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Review

Jessica Rahman[1], Aida Brankovic[1], Mark Tracy[1,2,3], Sankalp Khanna[1]

[1]The Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation, Australia
[2]Neonatal Intensive Care Unit, Westmead, Sydney, Australia
[3]University of Sydney, Australia

Corresponding Author:
Jessica Rahman
Email: jessica.rahman@csiro.au

Exploring Computational Techniques in Pre-processing Neonatal Physiological Signals for Detecting Adverse Outcomes: A Scoping Review

Abstract

Background: Computational signal pre-processing is a prerequisite for developing data-driven predictive models for clinical decision support. Thus, identifying the best practices that adhere to clinical principles is critical to ensure transparency and reproducibility that drives clinical adoption. It further fosters reproducible, ethical and reliable conduct of studies. This procedure is also crucial for setting up a software quality management system (QMS) to ensure regulatory compliance in developing software as a medical device (SaMD) aimed at early preclinical detection of clinical deterioration.

Objective: This scoping review focuses on the Neonatal intensive care unit (NICU) setting and summarises the state-of-the-art computational methods used for pre-processing neonatal clinical physiological signals for the development of machine learning models in predicting the risk of adverse outcomes.

Methods: Five databases (Pubmed, Web of Science, Scopus, IEEE, ACM Digital Library) were searched using a combination of keywords and MeSH terms. 3,585 papers from the year 2013 to January 2023 were identified based on the defined search terms and inclusion criteria. After removing duplicates, 2,994 papers were screened by title and abstract, and 81 were selected for full-text review. Of these, 52 were eligible for inclusion in the detailed analysis.

Results: Of the 52 articles reviewed, 24 studies focused on diagnostic models, while the remainder focused on prognostic models. The analysis conducted in these studies involved various physiological signals, with ECG being the most prevalent. Different programming languages were utilised, with Matlab and Python being notable. The monitoring and capturing of physiological data employed diverse systems, impacting data quality and introducing study heterogeneity. Outcomes of interest included sepsis, apnea, bradycardia, mortality, necrotizing enterocolitis (NEC) and hypoxic ischemic encephalopathy (HIE), with some studies analysing combinations of adverse outcomes. We found a partial or complete lack of transparency in reporting the setting and the methods used for

signal pre-processing. This includes reporting methods to handle missing data, segment size for considered analysis, and details regarding the modification of the state-of-the-art methods for physiological signal processing to align with the clinical principles for neonates. Only seven out of the 52 reviewed studies reported all the recommended pre-processing steps, which could have impacts on the downstream analysis.

Conclusions: The review found heterogeneity in techniques and inconsistent reporting of parameters and procedures used for pre-processing neonatal physiological signals, which is necessary to confirm their adherence to clinical and software QMS practices, usefulness and choice of the best practices. Enhancing transparency in reporting and standardising procedures will boost study interpretation, reproducibility, and expedite clinical adoption. This will instill confidence in the research findings and streamline the translation of research outcomes into clinical practices, ultimately contributing to the advancement of neonatal care and patient outcomes.

Keywords: physiological signals; preterm; neonatal intensive care unit; morbidity; signal processing; signal analysis; adverse outcomes, predictive and diagnostic models

Introduction

Background

Premature infants are those born at less than 37 weeks gestational age, ranging from extreme preterm (23 weeks gestation) to late preterm (37 weeks gestation) and are defined as having very low birth weight (VLBW) of <1500 grams. These extremely premature infants have a higher risk of death, and surviving infants are highly prone to physical, cognitive and emotional impairment [1]. The patients usually have a long length of stay (LOS), ranging from <10 days to >120 days [2] in the neonatal intensive care unit (NICU), where high-fidelity physiological changes are monitored to observe their health status and signs of deterioration. During this long LOS, a large amount of data from infants are generated and not typically electronically aggregated for permanent storage [3]. With the advent of electronic health records (EHR), relevant patient information is easily available for advanced data analytics that can be used to improve health outcomes. The records contain demographics, etiology, pathology, medication, and physiology information. Physiological changes are regularly monitored in preterm infants, notably: electrocardiogram (ECG), oxygen saturation (SpO2), heart rate (HR), respiratory rate (RR), arterial blood pressure (ABP), electroencephalography (EEG) and temperature. Some advanced centres around the world have started linking the information derived from the EHR data with continuously monitored physiological information for permanent storage, more frequently in lower resolution, which facilitates various data analytics [4–6]. Continuous capturing and analysis of the physiological data from the standard bedside monitors allow for better understanding of trends and has been shown to improve outcomes of infants in the NICU, in comparison to intermittent assessment and review [5].

Clinical decision support systems (CDSS) can integrate clinical and physiological information to provide automated support in patient care planning to facilitate the diagnostic process, therapy planning, generating critical alerts and reminders and predicting the risk of patient deterioration. CDSSs have the potential for a positive impact in improving clinical and economic measures in the healthcare system [7–9]. The technological advancement that allowed storing *big data* as well as the advancement of artificial intelligence (AI) has given rise to machine learning (ML)/AI-based CDSSs

aiming to build data-driven models to predict adverse outcomes in premature infants ahead of clinical diagnosis time [10–12].

The steps of building the machine learning pipeline to predict adverse outcomes involve several intermediate computational steps using the physiological data of which data pre-processing is the first indispensable step. Namely, in the NICU physiological signals are collected using a diverse range of devices, which introduce a number of artefacts such as *environmental artefacts* (e.g. device connection failure, equipment noise, electrosurgical noise and power line interferences); e*xperimental or human error* due to patient movement during data acquisition, incorrect or poor contact of the electrodes and other contact noise; and *artefacts* due to muscle contraction, cardiac signals and blinking [13, 14]. These noises distort signals and may adversely affect model generalisation capability and predictive power [10].

Although recently a lot of progress has been made in building ML models using neonatal physiological data, there are limitations in detailed reporting of the pre-processing techniques of these signals [15] which in return hinder reproducibility of the methods and results. In AI-powered software as a medical device (SaMD), this is especially important as the implementation of a software quality management system (QMS) is only possible by following best practices adhering to relevant regulatory standards and guidelines for medical devices, such as ISO 13485, IEC 62304 and IEC 82304-1. Beyond market access considerations, the ongoing international discourse on the regulation of medical software is specifically concentrated on AI/ML. This focus is a response to their growing applications, demanding increased attention by regulatory bodies such as Therapeutic Goods Administration (TGA) and the U.S. Food and Drug Administration (FDA) [16]. Thus, it is crucial to adhere to a standardised protocol following clinical principles guided by domain experts and regulatory requirements while pre-processing the signals and reporting these techniques in detail to ensure reproducibility of the methods that allows transparency in their clinical adoption.

Aim of the Review

This review aims to identify studies that used computational methods to analyse physiological signals of premature infants used for detecting adverse outcomes as the first step in bridging the gap in their reproducibility for clinical adoption. The review describes different tools and techniques used to pre-process physiological signals and provides recommendations on what aspects need further details for clinical adoption of the techniques. The remainder of the paper is organised as follows: the Methods section explains the detailed search and screening process; the Results section begins with an overview of the reviewed studies, followed by a detailed analysis. The discussion section highlights the key reporting patterns identified in this review along with their shortcomings and provides recommendations for transparent reporting of future studies as it allows to reproduce the results accurately and make them usable in the clinical setting [17]. Summary of the work concludes the manuscript.

Methods

The database searches and study screening were conducted following the recommendations of PRISMA (Preferred Reporting in Systematic Reviews and Meta-Analyses) guidelines [18] and the Centre for Reviews and Dissemination guidance for undertaking reviews in healthcare [19].

Database and Search Strategies

A systematic database search was conducted on five databases: Pubmed, IEEE, Web of Science, Scopus and ACM Digital Library. The keywords were categorized into four concepts, which were then merged using the “AND” operator. The concepts are the following:

1. Concept 1: Neonates/Preterm infants
2. Concept 2: Vital Signs/ Physiological signals
3. Concept 3: Computational techniques/Signal processing
4. Concept 4: Outcomes relating to neonates

Within each of these concepts, a combination of keywords and Medical Subject Headings (MeSH) terms were used to conduct the search process. The keywords under each concept were combined by “OR” operator. The searches were limited to only the title and abstracts. Table 1 shows the list of keywords and MeSH terms used to search the database.

Table 1. List of keywords and MeSH terms used to conduct database search

Concept 1: Neonates/ Preterm Babies	
MeSH Terms	Infant, Premature
Keywords	premature OR preterm OR neonat* OR newborn OR infant OR nicu OR “neonatal intensive care unit”
Concept 2: Physiological Signals/ Vital Signs	
MeSH Terms	"Vital Signs" OR "Physiology"
Keywords	physiolog* OR ecg OR "heart rate " OR electrocardiography OR "vital sign*" OR physiomarker OR biomarker OR hrv
Concept 3: Computational techniques/Signal processing	
MeSH Terms	Signal Processing, Computer-Assisted
Keywords	“signal *” OR predict* OR detect* OR comput*
Concept 4: Outcomes	
MeSH Terms	None
Keywords	sepsis OR mortality OR “length of stay” OR “intraventricular hemorrhage” OR “hypoxi*” OR apnea OR “necrotising entercolitis” OR “necrotizing entercolitis”

The search was done on 09 January 2023 and publication year of the papers was limited to 2013-2023. The reason for choosing the 10-year range was to report on recent techniques and tools as the devices and computational tools used more than 10 years ago may be obsolete. Scopus, Wiley Online Library and Web of Science have an additional filter of choosing the subject area. This was used to restrict the subject areas to multidisciplinary, engineering, computing and statistics. This was done to identify more papers on multidisciplinary areas through these databases as Pubmed covers all the major medical and health informatics databases. The combination of the five databases ensured that all medical, information technology and multidisciplinary research papers were included in the database search. The search was restricted to articles in English. Finally, review articles were excluded from the search.

Screening and Study Selection

The initial screening of the databases led to a total of 3,585 papers. Of these, 590 papers were manually identified to be duplicates which were excluded from the analysis. One paper was identified as a duplicate by the automation tool and removed. The remaining 2,994 papers were subjected to title and abstract screening using the Rayyan Intelligent Systematic Review application (Qatar Computing Research Institute) [20].

Several inclusion criteria were set to select papers for full-text review. The criteria are mentioned in Table 2.

Table 2. Inclusion/Exclusion Criteria

Criteria Type	Inclusion Criteria	Exclusion Criteria
Article type	Articles must be peer-reviewed publications in a journal, conference, or workshop	Review papers are excluded
Data	Articles must conduct analysis on premature human infant data	Non-human data e.g. piglet infant data would not be considered
	Articles must employ physiological responses in some form	Videos, images that do not look at the physiological responses, and articles solely using demographic data for analysis were excluded
Outcome	Articles discuss applications relating to adverse neonatal outcomes such as mortality, length of stay, sepsis, Necrotizing Enterocolitis (NEC), Intraventricular Haemorrhage (IVH), Hypoxic Ischemic Encephalopathy (HIE), apnea, bradycardia and other poor health outcomes, also known as morbidity. The disease outcomes were chosen based on the commonly researched outcome metric using preterm infant data and the search terms used in McAdams et al. that investigated AI and ML techniques used to predict clinical outcomes in the NICU [10].	Article not focusing on these specified neonatal adverse outcomes were excluded
Analysis	Articles reported some form of computational techniques in their analysis	Articles that only reported responses in their raw format were excluded
Language	English	Any other languages except English

After screening the titles and abstracts, 81 articles were selected for full-text review. 29 papers were excluded during this stage as they did not align with the inclusion criteria, leaving 52 papers eligible for detailed synthesis and analysis. The title/abstract screening was done by one reviewer, while two reviewers independently checked paper eligibility against the inclusion criteria in the full text review stage. Where both reviewers were not in agreement on any papers, a third reviewer assessed them to provide a final decision on the inclusion/exclusion of the papers. Data charting was done using Microsoft Excel and the following variables were recorded, in line with related review papers [10, 21] – title, year, journal, authors, doi, dataset, participant number, participant demographic, signals used, dataset size, sample rate, other data (if applicable), outcome metric, device software, programming language, pre- processing methods, algorithms, other techniques, features, models, model type, results (quantified), and key findings. Data synthesis was done using a narrative approach by summarizing finding based on the similarities in the datasets and techniques used. Figure 1 shows the full process of database search and study selection using a PRISMA flow diagram.

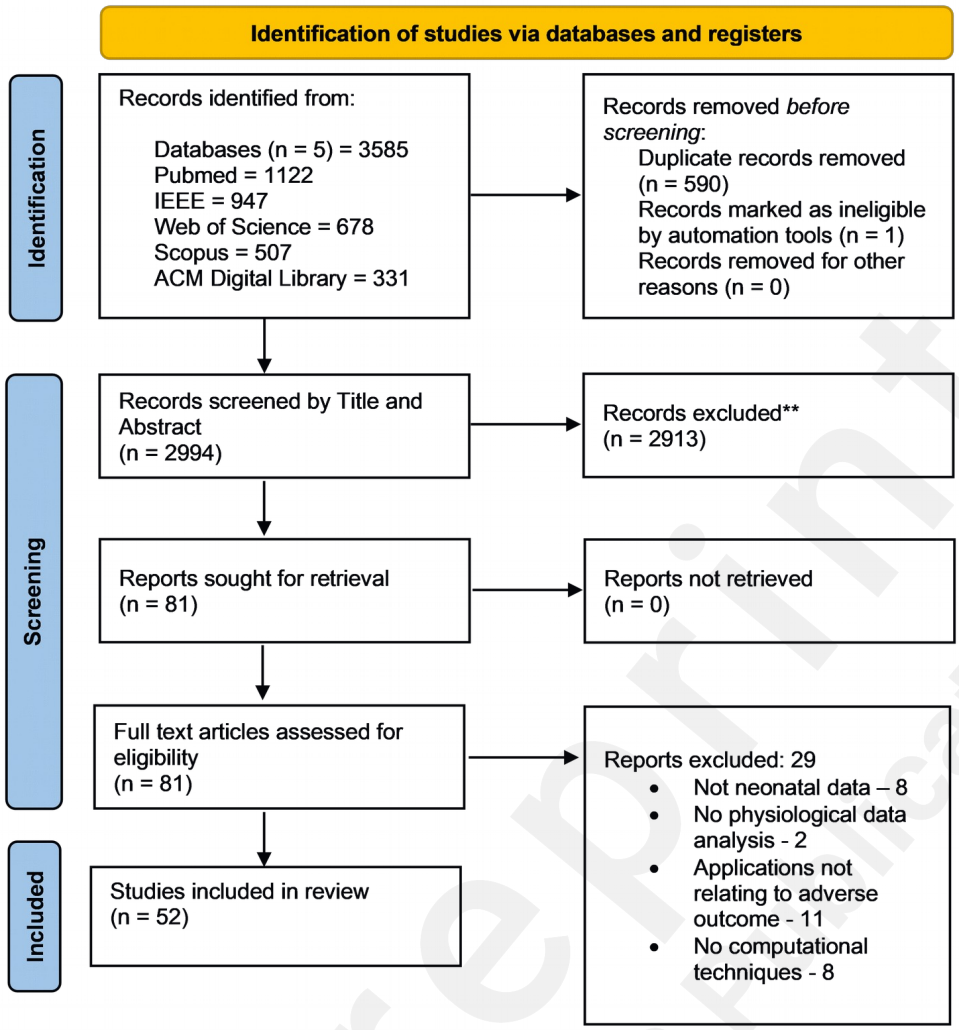


Figure 1: PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of for the database search and study selection

Results

Overview of Studies

Out of the 52 selected articles, 24 studies focused on diagnostic models, while the rest focused on prognostic models. These included journal articles (n=34), conference articles (n=17) and one workshop article (n=1). The most prominent physiological signals analysed were ECG (n=36), SpO2 (n=21), HR (n=16), Respiration (n=16), BP (n=6), EEG (n=4) and Temperature (n=3). Eight studies used a combination of programming languages, others used Matlab (n=6), Python (n=6), R (n=1), while the rest did not report what language was used (n=31). Physiological data monitoring and capturing was done using a range of systems, which subsequently impacts the sampling rate and quality of the data, thus leading to heterogeneity of the studies. The most commonly used devices for data capturing were Phillips Intellivue MP20, MP70, MP450 and MX800 machines [22] (n=14).

Some other notable devices and software were, BedMaster Ex System [23], NicoletOne EEG system [24], ixTrend [25], Phillips Data warehouse connect [26] and Vuelogger patient monitoring system. The most commonly analysed outcomes of interest were, Sepsis (n = 20), Apnea (n = 17), Bradycardia (n = 13), Mortality (n = 7) and HIE (n = 5). It should be noted that fourteen of the reviewed studies analysed a combination of adverse outcomes.

As the studies were found to be heterogeneous in their study design and analysis techniques, a narrative approach was taken to summarise the studies and their key findings. The studies were grouped according to the homogeneity in terms of the datasets used, sorted by the publication year. This approach was inspired by the review article by Mann et al. [27]

Table 2. Summary of the articles reviewed in this study, grouped according to the homogeneity in terms of the datasets used, sorted by the publication year.

Dataset Used	Year & Author	Study Settings	Physiological Signal Analysed	Signal processing/Computational techniques	Key Findings
National Collaborative Home Infant Monitoring Evaluation (CHIME) Database [28]	2013, Cohen and Chazal [29]	Participants: 288, Data Size: Not reported, Model: Diagnostic. Outcome metric: Sleep apnea	ECG from single channel at 100 Hz, SpO2 at 1 Hz	SpO2 values below 65% and change in saturation exceeding 4% per second were discarded. ECG QRS complex detected using the Pan-Tompkins algorithm [30] to generate RR intervals. QRS complexes were filtered using a technique from Chazal et al. [31]. Filtered intervals were time aligned with SpO2 using 30s epochs.	Eleven features were extracted from the signals. A combination of features from both signals resulted in 88.8% accuracy, 94.3% specificity and 73.4% sensitivity in detecting sleep apnea.
CHIME	2014, Cohen and Chazal [32]	Participants: 402, Data Size: Not reported, Model: Diagnostic. Outcome metric: Sleep apnea	ECG from single channel at 100 Hz, SpO2 at 1 Hz, actigraphy signals at 50 Hz	Actigraphy signals artefact rejection was done using technique described in Lewicke et al. [33]. SpO2 values below 65% and change in saturation exceeding 4% per second were discarded. ECG data was passed through a QRS detection algorithm (not reported) to produce RR intervals, which were filtered using method outline at [31].	Fourteen features were extracted from the signals. A linear discriminants classifier achieved an accuracy of 74.1%, a sensitivity of 82.0% and a specificity of 60.9% in detecting sleep apnea.
CHIME	2015, Cohen and Chazal [34]	Participants: 394, Data Size: Not reported, Model: Diagnostic. Outcome metric: Sleep apnea	ECG from single channel at 100 Hz, SpO2 at 1 Hz	SpO2 and ECG signals were time-aligned to 30-s epochs. SpO2 values below 65% and change in saturation exceeding 4% per second were discarded. ECG QRS complex detected using the Pan-Tompkins algorithm [30] to generate RR intervals. QRS complexes were filtered using a technique from Chazal et al. [31]	Eleven features were extracted from both signals. A linear discriminant model achieved 66.7% accuracy, 67% specificity and 58.1% sensitivity using features from both signals.
Preterm Infant Cardio-Respiratory Signals (PICS) Database [35, 36]	2016, Gee et al. [37]	Participant:10, Data size: ~20-70h each, Model: Diagnostic. Outcome metric: Bradycardia	3-lead ECG at 500 Hz, respiration signal at 50 Hz	RR intervals from ECG extracted using a modified Pan-Tompkins algorithm (modification details not reported). Analysis done on 3m window prior to each bradycardia. No processing was reported for respiration signals.	Bradycardia severity estimation accuracy was improved by an average of 11% using a point process model of heart rate and respiration.

PICS	2017, Gee et al. [35]	Participant:10, Data size: ~20-70h each, Model: Prognostic (+116s), Outcome metric: Bradycardia	3-lead ECG at 500 Hz, respiration signal at 50 Hz	RR intervals from ECG extracted using a modified Pan-Tompkins algorithm (modification details not reported). The artifacts due to movement, disconnection, or erroneous peaks were removed by visual inspection. No processing was reported for respiration signals. Additional analysis on frequency content of RR time series was done using Morlet wavelet transform [38].	A point process model-based prediction algorithm achieved mean AUROC of 0.79 for over 440 bradycardic events and able to predict bradycardic event on an average of 116 seconds prior to onset (FPR=0.15)
Dataset Used	**Year & Author**	**Study Settings**	**Physiological Signal Analysed**	**Signal processing/Computational techniques**	**Key Findings**
PICS	2019, Das et al. [39]	Participant:10, Data size: ~20-70h each, Model: Prognostic (time not reported), Outcome metric: Bradycardia	3-lead ECG at 500 Hz	Baseline wander was removed using a high-pass filter with cut-off frequency between 0.5-0.6 Hz. Motion and disconnection artifacts were removed by visual inspection. QRS complexes were detected using Pan-Tompkins algorithm [30]. Signals was segmented five minutes prior and two minutes after bradycardic event.	A nonparametric modeling using kernel density estimation achieved 5% false alarm rate in predicting the onset of bradycardia events
PICS	2019, Mahmud et al. [40]	Participant:11, Data size: ~20-70h each for 10, 10 weeks for one participant, Model: Prognostic (time not reported), Outcome metric: Bradycardia	3-lead ECG at 500 Hz	QRS complex was detected using an algorithm (not reported). RR intervals were calculated from the detected peaks.	Time and frequency domain features were extracted. An extreme gradient boosting model achieved an average AUROC of 0.867. HRV results showed a significant variation between a healthy infant and an infant prone to bradycardia.
PICS	2019, Gee et al. [41]	Participant:10, Data size: ~20-70h each, Model: Diagnostic, Outcome metric: Bradycardia, apnea of prematurity	3-lead ECG at 500 Hz, respiration signal at 50 Hz	Respiration signals were clipped into 60s segments and normalized to zero mean, unit variance. RR intervals from ECG signals were extracted using a Morlet wavelet transformation. An open-source peak finder (name not reported) was applied to the wavelet scale ranges from 0.01 to 0.04 related to the QRS complex formation in the spectrogram. ECG signals were segmented to 15s with the event in the middle. The segments were ban-passed filtered from 3 to 45 Hz, scaled to zero-mean, unit-variance, and scaled to the median QRS complex amplitude. Waveforms were visually inspected to remove segments with no distinguishable QRS complex or respiratory peaks.	An autoencoder-prototype model was proposed which achieves 93.1±0.4% accuracy in predicting bradycardia and 82.3±3.8% accuracy in classifying apnea.

MIMIC-III database from Beth Israel Deaconess Medical Center [42]	2020, Song et al. [43]	Participants: 2819 (21 sepsis, 2798 control), Data Size: Not reported, Model: Prognostic (+48h), Outcome metric: Sepsis	HR, SBP, DBP, MBP, SpO2, Respiration, TEMP, other (sampling rate not reported)	Data quality was assessed by missing value filter and three-sigma rule. Last observation carried forward was applied for vital signs not meeting data quality. Zero imputation was performed is calculation could not be performed (e.g. divided by zero).	Several statistical features were extracted on 3h, 6h, 12h and 24h window. Linear model, naïve Bayes, decision tree, ensemble method and neural network models were evaluated. The AUROC of the 48-hour prediction model achieved 0.861 and that of the onset detection model was 0.868.
Dataset Used	**Year & Author**	**Study Settings**	**Physiological Signal Analysed**	**Signal processing/Computational techniques**	**Key Findings**
MIMIC-III	2021. Baker et el. [44]	Participants: 179 for 3-day, 181, for 14-day model. Data Size: Not reported, Model: Prognostic (+3d) Outcome metric: Mortality	HR, Respiration signal, sampled hourly	Values less than zero and flatline cases were eliminated.	Several statistical features were extracted from the signals. CNN-LSTM model using 3-day scheme achieved AUROC of 0.9336±0.0337 across 5-fold cross validation.
MIMIC-III	2022, Juraev et al. [45]	Participants: 3133, Data size: 24h from each Model: Prognostic (time not reported), Outcome metric: Mortality and Length of Stay	HR, Respiration signal, SpO2, BP, Temp (sampling rate not reported)	Missing data was filled by forward and backward filling, using the mean value. For participants >24 measurements, they were reduced by taking the average of the nearest records. For <24 measurements, values were generated using filling algorithm.	A dynamic ensemble KNN method reached 0.988 ± 0.001 F1 score in mortality classification. A voting of static ensemble regression models achieved RMSE = 12.509 ± 0.079 in length of stay prediction.
University Hospitals in France	2015, Ghahjaverestan et al. [46]	Participant:32, Data size: 105 segments of ECG with 250s duration, Model: Diagnostic, Outcome metric: Apnea-Bradycardia	One lead ECG at 400 Hz and respiration signals (sampling rate not reported)	Baseline and noise of 50 Hz was removed from ECG signals, QRS complexes were detected using Pan-Tompkins algorithm [30]. RR intervals were further downsampled to 10 Hz for one prediction model.	A Kalman-filter based method achieved sensitivity and specificity 94.74% and 94.17%, respectively in predicting apnea-bradycardia episodes.
University Hospitals in France	2015, Navarro et al. [47]	Participant:51, Data size: testing cohort mean duration – 2.4h, Model: Diagnostic, Outcome metric: Sepsis	Respiration signals at 400 Hz, downsampled to 64 Hz	Frequency content over 32 Hz from breathing signals were removed using a 7^{th} order butterworth low pass filter. After rejecting artifacts due to gross movements, a 4^{th} order butterworth filter with cut-off frequency between 0.5 to 20 Hz was applied. A smoothing filtering using Savitzky-Golay (SG) filter [48] was applied. A simple extrema detector is then applied to detect respiratory cycles.	14 features, computed in 10s sliding excerpts were extracted from the breathing signals. A logistic regression classifier automatically rejects artifacts to 86% sensitivity and specificity, which is used in the proposed framework for neonatal sepsis detection.

University Hospitals in France	2016, Ghahjaverestan et al. [49]	Participant:32, Data size: Real – 236 segments Synthetic – 200 sequences of 400s, Model: Diagnostic (0.59s delay), Outcome metric: Apnea-Bradycardia	ECG at 400 Hz. Synthetic signals at 10 Hz.	Baseline and noise of 50 Hz was removed from ECG signals using a combination of low-pass and notch filter, QRS complexes were detected using Pan-Tompkins algorithm [30]. Three features were extracted using a wavelet-based beat delineator [50]. Features were transformed to 10 Hz using interpolation (technique not reported).	A coupled HMM (CHMM) achieved 95.74% sensitivity and 91.88% specificity in detecting apnea-bradycardia episodes, with a detection delay by -0.59 seconds.
Dataset Used	**Year & Author**	**Study Settings**	**Physiological Signal Analysed**	**Signal processing/Computational techniques**	**Key Findings**
University Hospitals in France	2021, Leon et al. [51]	Participant:49, Data Size: Not reported, Model: Prognostic (+6h), Outcome metric: Sepsis	ECG at 500 Hz	RR intervals were detected using modified Pan-Tompkins algorithms, filter coefficients adapted for newborns [52]. A sliding window of 30m, with no overlaps, was applied to extract HRV parameters from RR time series. 30m segments with a maxRR greater than one second, or a minRR of less than 0.19s were excluded.	Time, frequency and non-linear features were extracted from the HRV parameters. A logistic regression model using visibility graph features achieved 0.877 AUROC in predicting Sepsis six hours prior to start of antibiotics.
University Hospitals in France	2021, Leon et al. [53]	Participant:259, Data Size: Not reported, Model: Prognostic (+6h), Outcome metric: Sepsis	ECG at 500 Hz	RR intervals were detected using modified Pan-Tompkins algorithms, filter coefficients adapted for neonates [52]. RR time series were extracted and segmented into 5m segments. The 5m periods corresponding to 30 continuous minutes were grouped together by calculating the median of each corresponding HRV feature.	Time, frequency, non-linear and visibility graph features were extracted from the HRV parameters. An RNN model achieved 0.904 AUROC in predicting sepsis six hours before the time of infection, and more than 80% accuracy 24h prior to onset of infection.
University Hospitals in France	2021. Doyen et al. [54]	Participant:52, Data size: 8h recording from each, Model: Diagnostic (+2.9s delay), Outcome metric: Bradycardia	3-lead ECG at 300 Hz	QRS complexes were detected using a multi-feature probabilistic real-time detector [52].	A high rate of false alarms (64%) was observed in real life. The proposed optimal decentralized fusion of three detection methods had a significant detection delay of 2.9s, sensitivity of 97.6% and false alarm rate of 63.7%.
University Hospitals in France	2021, Sadoughi et al. [55]	Participant:32, Data size: 233 episodes with duration of 21.48 ± 16.07s, Model: Diagnostic (+5.05s delay), Outcome metric: Apnea Bradycardia	One lead ECG at 400 Hz	Same pre-processing techniques as reported in Ghahjaverestan et al. [49]. QRS complexes were identified using Pan-Tompkins method [30]. The RR time series were uniformly upsampled to 10 Hz using linear interpolation technique.	A proposed layered HMM model achieved 97.14± 0.31% accuracy in detecting AB episodes, with a detection delay of -5.04± 0.41s.

Cork University Maternity Hospital	2015, Ahmed et al. [56]	Participant - not reported, Data size: 54 1h recordings, Model: Diagnostic, Outcome metric: HIE	2-lead ECG, EEG (sampling rate not reported)	Artefacts were manually removed. R-peaks from raw ECG signals were extracted using Pan-Tompkins method [30]. Timing of the peaks were adjusted and uniformly samples to 256 Hz using hermite spline quadratic interpolation. Then, HRV features were extracted from one minute window with 30s overlap using the normalized RR interval.	Seven time and frequency domain HRV features were extracted. A gaussian supervector approach with SVM achieved 0.81 AUROC in classifying HIE.
Dataset Used	**Year & Author**	**Study Settings**	**Physiological Signal Analysed**	**Signal processing/Computational techniques**	**Key Findings**
Cork University Maternity Hospital	2015, Temko et al. [57]	Participant: 38, Data size: 1h EEG and ECG recordings from each, Model: Diagnostic, Outcome metric: HIE	ECG and Video EEG at 256 Hz	The 1h EEG segments were downsampled to 32 Hz with an anti-aliasing filter set to 16 Hz. The filtered EEG was segmented into 60s epoch with no overlap. QRS complexes from ECG signals were extracted using the algorithm reported in [58]. Resulting peaks were manually inspected to correct ectopic beats or mark artifacts. Then signals were segmented into 60s epochs.	An SVM classifier using a subset of 9 EEG, 2h and 1 clinical feature achieved 87% AUROC and 84% accuracy in predicting HIE.
Cork University Maternity Hospital	2016, Lloyd et al. [59]	Participant: 43, Data size: Mean recording duration 41h 40m, Model: Diagnostic, Outcome metric: future adverse outcome in infants	EEG at 256 Hz, SpO2 and HR at 1 Hz	EEG recordings were visually checked for quality and poor-quality data were discarded. One-hour epochs of EEG at 12h and 2h of age were then extracted from each recording. One-hour epochs of HR and SpO2 were extracted at 12h and 24h time point.	A logistic regression model predicted 2-y poor outcome with an AUROC of 0.83.
Cork University Maternity Hospital	2018, Semenova et al. [60]	Participant: 35, 23 used, Data size: 824h, Prognostic (time not reported), Outcome metric: Short-term adverse outcome	ECG at 256 or 1024 Hz, BP at 1 Hz	Diastolic and systolic pressures every second were used to calculate MAP. ECG signals were segmented to non-overlapping 5-minute epochs. QRS complexes were extracted by Pan-Tompkins method [30]. Abnormal RR intervals were corrected by moving average. Periods of clear movement artefact were automatically discarded (method not reported).	Fifteen time, frequency and non-linear features were extracted from HRV. An XGBoost decision tree using all features achieved an AUROC of 0.97 in predicting short-term outcomes in infants.

Cork University Maternity Hospital	2019, Semenova et al. [61]	Participant: 43, 23 used, Data size: Total 831h, Prognostic (time not reported), Outcome metric: Five adverse outcomess	ECG at 256 or 1024 Hz, BP at 1 Hz	Diastolic and systolic pressures every second were used to calculate MAP. Segments with MAP<10mmHg were discarded due to disconnection of the pressure transducer or movements. The MAP was segmented into 1h windows. Values outside ±3 SD were discarded. ECG signals were segmented to non-overlapping 5m epochs. QRS complexes were extracted using Pan-Tompkins method [30]. ECG signal was bandpass filtered with 4-30 Hz cut-off frequency. Abnormal values of RR intervals were corrected by the moving average filter.	Time, frequency, and non-linear features were extracted from HRV. An XGBoost decision tree using a single HRV feature achieved 0.87 AUROC, while multiple features reached 0.97 AUROC in predicting adverse outcomes.
Dataset Used	**Year & Author**	**Study Settings**	**Physiological Signal Analysed**	**Signal processing/Computational techniques**	**Key Findings**
Máxima Medical Center NICU	2020, Joshi et al. [62]	Participant: 49, Data size: ~144h each, Model: Prognostic (+0-24h), Outcome metric: Sepsis	ECG at 250 Hz, CI at 62.5 Hz	Respiration waveforms were band-pass filtered between 0.45-1.45 Hz. QRS complexes from ECG were extracted using a DT-CWT based method described in [63]. IBIs were detected from the CI signal peaks using an algorithm (not reported). Features were extracted from every 3h data.	Twenty-two features were extracted from the signals. A naïve bayes classifier reached upto 0.78 AUROC, 3h leading up to sepsis.
Máxima Medical Center NICU	2021, Varisco et al. [64]	Participant: 20, Data size: ~570h, Model: Prognostic (+6h), Outcome metric: Central apnea preceding late onset sepsis	ECG at 240 or 250 Hz, CI at 60 or 62.5 Hz, SpO2 at 0.5 or 1 Hz	A filtered respiration signal without cardiac artifacts was generated using Lee's algorithm [65–67]. Steps include- Fourier transformation and integer frequencies filtered out, then resampled to 60 Hz and high-pass filtered with a cut-off frequency of 0.4 Hz, and a low-pass filter with a very low cut-off frequency optimized to fit apnea annotations by clinical experts (value not reported).	An optimisation of Lee's algorithm was proposed to detect central apnea which achieved 90.5% recall, 19.7% precision and 30.8% F1 score.
Máxima Medical Center NICU	2021, Cabrera-Quiros et al. [68]	Participant:64, Data Size: Not reported, Model: Prognostic (+3h), Outcome metric: Sepsis	ECG at 250 Hz, CI at 62.5 Hz	QRS complexes from ECG were extracted using a DT-CWT based method described same as Joshi et al. [62]. CI signal was filtered to remove cardiac artifacts and peaks were detected using methods similar to their previous works (not reported). Features were extracted from every 1h signals.	Time domain features were extracted from HRV. Classification using a combination of all features and logistic regression model reached a mean accuracy of 0.79 ± 0.12 and mean precision of 0.82 ± 0.18, 3h before the onset of sepsis

Máxima Medical Center NICU	2022, Varisco et al. [69]	Participants: 20, Data size: 960h of data from 20 infants, 7818 event extracted, Model: Diagnostic, Outcome metric: Central Apnea	ECG at 250Hz, CI at 62.5Hz, SpO2 at 1Hz	QRS complexes were detected using same method as [62, 68]. From ECG, signal instability index (SII) was calculated by applying a band-pass filter (0.001–0.40 Hz) using 10s segments and then computing a kernel density estimate to return patient motion measurement every second. RR intervals were resampled at 250 Hz. CI signal was processed using method by Redmond et al. [70] to calculate ribcage respiratory effort (RRE). No pre-processing was done on SpO2. Each feature was extracted using 30s windows. Z score normalization was applied to the feature matrix.	47 features were extracted from the vitals. A logistic regression model achieved 0.9 AUROC in detecting central apnea
Dataset Used	**Year & Author**	**Study Settings**	**Physiological Signal Analysed**	**Signal processing/Computational techniques**	**Key Findings**
Máxima Medical Center NICU	2022, Peng et al. [71]	Participants: 128, Data size: ~24h each, Model: Prognostic (+24h), Outcome metric: Sepsis	ECG at 250 Hz	QRS complexes from ECG were extracted using a DT-CWT based method described in [63]. RR intervals from the complexes were divided into non-overlapping 1h segments. The segments were centered, missing values in the segments were filled by zero padding on the two ends.	A ResNet based neural network DeepLOS was proposed which achieved 0.72 in F1-score in predicting late onset sepsis.
Máxima Medical Center NICU	2022, Peng et al. [72]	Participants: 127, Data size: ~48h each, Model: Prognostic (+6h), Outcome metric: Sepsis	ECG at 250Hz, CI at 62.5Hz	QRS complexes from ECG were extracted using a DT-CWT based method described in [63]. CI signal was filtered to remove cardiac artifacts (method not reported). Peaks were detected using method reported in Lee et al. [65]. SII was calculated from ECG and CI waveforms using a CWT based method reported in [73]. Signals were divided into 1h long non-overlapping segments. Features were calculated in both 1h segments and 5m subsegments.	60 Features were extracted from the signals. An XBG model using the features achieved an AUROC of 0.88 in predicting late onset sepsis 6h preceding the onset.
Royal Infirmary of Edinburgh NICU	2014, Stanculescu et al. [74]	Participant: 24, Data size: 30h each, Model: Prognostic (+3-6h), Outcome metric: Sepsis	ECG derived HR, PR (sampling rate not reported)	An extension of the forward-backward algorithm [75] is developed for missing data inference.	An autoregressive HMM model achieved upto 0.80 AUROC in predicting sepsis.

Royal Infirmary of Edinburgh NICU	2014, Stanculescu et al. [76]	Participant: 24, Data size: 540h, Model: Diagnostic, Outcome metric: Sepsis	ECG derived HR, PR core and peripheral TEMP and SpO2 at 1 Hz	Automated oximeter error detection algorithm applied based on the method described in Stanculescu et al. [74]. Rows containing missing data on the observation matrix are set to zero.	A Hierarchical Switching Linear Dynamical System (HSLDS) was able to predict sepsis with up to 0.65 F1 score
Kasturba hospital NICU, Manipal, India	2016, Shirwaikar et al. [77]	Participant: not reported, Data size: 229 examples, Model: Diagnostic, Outcome metric: Apnea	HR (sampling rate not reported)	Visualisation technique was applied to identify issues in data. Missing values were not treated due to low percentage. For categorical features, 0 was added for missing values. Min-max normalisation and Z score normalisation was done.	An RF model using HR features achieved 0.88 accuracy and 0.72 kappa in detecting apnea.
Dataset Used	**Year & Author**	**Study Settings**	**Physiological Signal Analysed**	**Signal processing/Computational techniques**	**Key Findings**
Kasturba hospital NICU, Manipal, India	2019, Shirwaikar et al. [78]	Participant: 367, 315 used, Data Size: Not reported, Model: Diagnostic, Outcome metric: Apnea	ECG (sampling rate not reported)	No pre-processing techniques were reported on the raw signals. Observations with missing features were discarded. Other features (continuous values) that had missing values were converted to discrete with addition of group name ‘not known’	Statistical features were extracted from the signals. A Multilayer Perceptron model and a deep auto-encoder model reached 0.82 and 0.83 AUROC respectively in detecting apnea.
University of Massachusetts Memorial Healthcare NICU	2013, Williamson et al. [79]	Participant: 6, Data size: ~5-8h each patient, Model: Prognostic (+5.5m), Outcome metric: Apnea	ECG, SpO2, respirator signal, pulse plethysmogram (sampling rate not reported)	IBIs were extracted from abdominal respiratory movements (method not reported) and RR intervals were extracted from ECG signals (method not reported). Physically implausible IBI and RR interval values were automatically removed (range not reported). Values were resampled to 10 Hz using shape-preserving piecewise cubic interpolation. Signals were then log-transformed and converted to zero mean, unit variance.	Features were extracted from all signals. A GMM model reached 0.8 AUROC in predicting apnea.

Jackson Memorial Hospital NICU	2013, Schiavenato et al. [80]	Participant: 20, Data size: 1186m, Model: Diagnostic, Outcome metric: Periods of high distress/ pain	ECG at 1000 Hz	Pan-Tompkins algorithm [30] was modified to detect QRS complexes. ECG was filtered using a band-pass filter with 16-26 Hz cut-off frequency. A low-pass filter by an order 120 FIR filter with corner frequency of 25 Hz, and a high pass filter by an order 160 FIR filter, with a corner frequency of 25 Hz was applied. Then, a polynomial filter of order 21 was applied as the differentiator filter. Finally, a 111-order moving average filter was used and QRS complex was detected using an adaptive threshold. Lomb-Scargle LMS spectral estimation [81] was used for missing and irregular RR intervals.	The proposed framework provided real-time analysis and HRV extraction to identify characteristics correlated to periods of high distress/pain
Montreal Children's Hospital	2014, Rubles-Rubio et al. [82]	Participant: 24, Data size: 9.0 ± 2.2h for each, Model: Diagnostic, Outcome metric: Apnea	SpO2, RIP (sampling rate not reported)	Signals were low-pass filtered with cut-off frequency of 10 Hz, with an 8-pole Bessel anti-aliasing filter, digitized, sampled at 50 Hz.	A linear Gaussian discriminant classifier detected the episodes with a 0.73 probability of detection and 0.22 probability of false alarm.
Dataset Used	**Year & Author**	**Study Settings**	**Physiological Signal Analysed**	**Signal processing/Computational techniques**	**Key Findings**
University of Alabama at Birmingham	2017, Amperayani et al. [83]	Participant: 18, Data size: 24h each, Model: Prognostic (+23h), Outcome metric: Bradycardia, Hypoxemia	ECG at 500 Hz and HR at 1 Hz	HR data was converted to inter beat RR intervals using RR = 60/HR. No processing on ECG signals were reported.	A point process model using RR intervals showed strong correlation with bradycardia events and modest correlation with hypoxemia events.
Monash Children's Hospital NICU, Australia	2018, Hu et al. [84]	Participant: <80, Data size: 407 patient-day, Model: Prognostic (+24h), Outcome metric: Sepsis	HR, SpO2, Respiration signal at 1 Hz	Data was scaled down to one record per minute. Data block with invalid values were deleted. Then sliding window was set to 60m to feed to the ML models.	Features were extracted from all signals. A gradient boosting decision tree achieved upto 0.97 AUROC and 0.92 weighted F1 in patient-based cross validation in predicting sepsis

University of Virginia and Columbia University NICU	2018, Sullivan et al. [85]	Participant: 778, Data Size: Not reported, Model: Prognostic (+12h), Outcome metric: Death, sIVH (severe), BPD, treated ROP, late-onset sepsis and NEC	HR, SpO2 at 0.5 Hz	Infants with less than 6 hours of data within 12 hours of birth were discarded. Cross correlation of HR and SpO2 was calculated over 10m windows using the XCORR function of Matlab with a lag time of -30 to +30s.	A pulse oximetry predictive score (POPS) was developed and fit to a multivariate logistic regression model which performed well in predicting death, sIVH and BPD, but not tROP, sepsis and NEC
9 NICUs in the USAs	2020, Zimmet et al. [86]	Participant: 2989, Data size: 121 data points per infant, Model: Prognostic (+2d), Outcome metric: Mortality, sepsis	HRC index from ECG	Infants with missing data on either end of the total duration were extrapolated to the window edge by repeating the most proximal HRC index values. Interior mussing values were updated using the linear interpolation. A fifth order B-splines with equally spaced knots was used to capture information from independent samples (HRC indices 12 samples apart).	An unsupervised ensemble clustering techniques was proposed to cluster infants to different levels of risk.
Dataset Used	**Year & Author**	**Study Settings**	**Physiological Signal Analysed**	**Signal processing/Computational techniques**	**Key Findings**
Children's National Hospital, Washington	2020, Kota et al. [87]	Participant: 95, Data size: Median recording duration 75.78h, Model: Diagnostic, Outcome metric: HIE	EEG at 200 or 256 Hz	ECG contamination from EEG was detected using method described in Govindan et al. [88]. EEG signals with amplitude > 500 µV or SD < 0.01 µV were discarded as artifacts. The volume conduction was attenuated by calculating the global average of EEG voltages from all electrodes and subtracting the global average from EEG value of every electrode in the frequency domain [89]. The values were then transformed to time domain for spectral analysis. EEG was segmented into 10m non-overlapping artifact and seizure free epochs. Spectral analysis was done using a Welch periodogram approach [90, 91] using 3s epochs.	EEG delta power was identified to be a crucial biomarker for predicting neonates with HIE who died with those who survived.
Akbar Abadi Hospital NICU, Iran	2021, Mirnia et al. [92]	Participant: 5, Data size: ~24h each, Model: Diagnostic, Outcome metric: Sepsis	ECG at 200 Hz	RR intervals were calculated from ECG using HeRO model.	Features were extracted from HRV. HeRO model was tested using this dataset. HeRO score was able to distinguish between healthy and septic newborns.

St. Louis Children's Hospital NICU	2021, Lee et al. [93]	Participant: 275, Data size: 4,01,33,460 data points, Model: Prognostic (+6h), Outcome metric: Mortality	HR, Respiration Signal and SpO2 at 1 Hz	Missing or out-of-range values were replaced with NaN, and then imputed using mean values for that variable across all training and testing data. Data was downsampled to every ten seconds to extract features. Dynamic variables were calculated as rolling means, SD and absolute z-score on 5m and 30m windows to reduce influence of outliers.	Thirty-four features were extracted from the signals. An RF model achieved 88% sensitivity and 0.93 AUROC in predicting mortality.
University of Virginia Children's Hospital, Morgan Stanley Children's Hospital and St. Louis Children's Hospital	2021, Sullivan et al. [94]	Participant: 408, 266 used, Data Size: Not reported, Model: Diagnostic, Outcome metric: Sepsis	HR and SpO2 at 0.5 Hz	HR and SpO2 values of zero were removed. Eight features were extracted in 10m windows and averaged hourly. Cross-correlation between HR and SpO2 was calculated in 10m windows of signals normalized to have zero mean and SD of 1. Cross correlation was done using XCORR function of Matlab with a lag time of -30 to +30s.	A logistic regression model using clinical and physiological features achieved the AUC of 0.821 in predicting late-onset sepsis.
Dataset Used	**Year & Author**	**Study Settings**	**Physiological Signal Analysed**	**Signal processing/Computational techniques**	**Key Findings**
St. Louis Children's Hospital NICU	2021, Feng et al. [95]	Participant: 285, Data size: ~80h each, Model: Prognostic (+6h), Outcome metric: Mortality	HR, Respiration, SPO2, and ART-M or NIBP-M at 1 Hz	Infants with data >80h were truncated, and <80h were padded with zeros. Mean, median, mode and Bayesian ridge data imputation techniques were explored. Bayesian ridge was used to sample five datasets by sampling difference posteriors each time. Then average was reported using 4-fold cross-validation. The rolling mean of each vital sign with a range of 5m was used to reduce noise. Finally, the end of each sample was padded with one segment where all features equaled to zero. Features were extracted from 5m segments.	A deep learning model using LSTM named DeepPBSMonitor was developed to predict mortality with 0.888 accuracy, 0.78 recall and 0.897 AUC

University of Massachusetts Memorial Healthcare	2021, Zuzarte et al. [96]	Participant: 10, Data size: 241.34h, Model: Prognostic (+310s), Outcome metric: ABH	ECG at 500 Hz, PPG at 125 Hz, SpO2, HR, Respiration signals from pneumogram at 50 Hz	PPG signals were filtered using a wavelet-based algorithm to remove gross body movements. A binary marker sampled at 25 Hz was obtained to indicate presence or absence of movement. QRS complexes were detected using modified Pan-Tompkins algorithm (modification not reported). IBIs were detected using automated peak-detection from Labchart Software RR intervals and IBI values were then interpolated at 10 Hz.	Prediction framework using GMM and logistic regression model achieved 75% accuracy in predicting bradycardia severity during the ABH event.
University of Virginia NICU	2022, Niestroy et al. [97]	Participant: 5957, Data size: Random daily 10m segments from each, Model: Prognostic (+1-7d), Outcome metric: Mortality	HR and SpO2 at 0.5 Hz	No pre-processing was reported on the vitals. They were grouped to calculate average in 10m non-overlapping windows.	Features were extracted from all signals. A multivariable logistic regression model using five features achieved the AUROC of 0.83 in predicting mortality.
University of Virginia Children's Hospital, Morgan Stanley Children's Hospital and St. Louis Children's Hospital	2023, Kausch et al. [98]	Participants: 2494, Data Size: Not reported, Model: Prognostic (+24h), Outcome metric: Sepsis	HR and SpO2 at 0.5 Hz	HR and SpO2 was pre-processed by removing values containing zero. Features were calculated in 10m non-overlapping windows. Windows with more than 50% missing data were excluded from subsequent analysis.	Several features were extracted from the vitals. An XGB model achieved training AUROC of 0.834 using data from NICU 1, and 0.792 and 0.807 testing AUROC using data from NICU 2 and NICU 3 respectively.
Dataset Used	**Year & Author**	**Study Settings**	**Physiological Signal Analysed**	**Signal processing/Computational techniques**	**Key Findings**
Karolinska University Hospital Solna and Huddinge NICU, Stockholm, Sweden	2023, Honore et al. [99]	Participants: 325, Data Size: 2866 hospitalisation days, Model: Prognostic (+24h), Outcome metric: Sepsis	IBI from ECG, Respiration from CI, SpO2 (sampling rate 1-500 Hz)	All signals were resampled to 1 Hz. Segments with at most 15s missing were linearly interpolated. All signals were filtered with a moving mean filter of width 3. IBI signals were further filtered to remove ectopic beats and strong nonlinearities with a moving median filter of width 3 and Butterworth band-pass filter of order 6 with low-cut and high-cut frequency of 0.0021 Hz and 0.43 Hz. Signals were divided into 45m segments. Features were calculated using a sliding time frame with 50% overlap.	A naïve bayes classifier achieved AUROC of 0.82 up to 24h prior to clinical suspicion of sepsis. Adding respiratory signals improved performance compared to only using heart rate features.

Simulated and real data (NICU name not reported)	2013, Masoudi et al. [100]	Participant: 32, Data size: 233 episodes, ~7s each, Model: Diagnostic (+2.32s delay), Outcome metric: AB	Two-channel ECG	No pre-processing techniques were reported. Signals were sampled in 7s intervals.	A coupled HMM model achieved 84.92% sensitivity, 94.17% specificity with a time detection delay of 2.32±4.82s in detection AB episodes
Simulated and real signals from NICU (NICU name not reported)	2015, Altuve et al. [101]	Participant: 32, Data size: 148 RR intervals with mean duration of 26.25 ± 11.37m, Model: Diagnostic (+1.73s delay), Outcome metric: AB	ECG (sampling rate not reported)	Hidden semi-markov models to represent the temporal evolution of RR intervals. A pre-processing method is proposed that includes quantisation and delayed version of the observation vector. RR time series was resampled at 10 Hz and segmented at 7s interval	Proposed model achieved up to 93.84 ± 0.79 in specificity and 89.66 ± 0.71 in sensitivity with a detection delay of 1.59 ± 0.24s
NICU (name not reported)	2020, Honore et al. [102]	Participant: 22, Data size: 3501 time series, 1200 samples in each, Model: Prognostic (+72h), Outcome metric: Sepsis	SpO2, respiratory frequency and RR interval from ECG at 1 Hz	Data was segmented into 20m time frames. Time frames with missing data were discarded.	Features were extracted from all signals. A combined GMM-HMM model achieved 0.74% ± 0.05 accuracy in detecting sepsis. Model was compared with HeRO model which under-performed using this dataset.

One of the noticeable patterns identified through the results reported in Table 2 is that the groups publishing studies using the same dataset followed similar pre-processing techniques, although not at every step. For instance, studies using the ECG data from Cork University Maternity Hospital all used the same algorithm for QRS complex detection. However, they were diverse in their selection of filtering techniques and segmentation duration. Furthermore, they systematically lacked in reporting detailed parameter settings for the QRS complex detection. While the approach of using similar pre-processing techniques helps maintain consistency to some extent, they do not confirm adhering to clinical practices identified from domain expert knowledge.

QRS complex characteristics and RR intervals for neonates are different from adults and as such require an appropriate adjustment for QRS detection algorithms. This is a necessary first step for HRV analysis in neonates. However, a review published on neonatal HRV by Latremouille et al. [15] revealed that given a lack of a clear guidelines on neonatal vital signs and HRV analysis, several studies followed HRV analysis guidelines for adults published by the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology [103]. Our review found that 16 out of the 36 studies analysing ECG signals used the Pan-Tompkins algorithms for QRS complex detection. The original implementation of the algorithm was based on the ECG characteristics of the adult population, and therefore was pre-processed accordingly. Only four of those 16 studies reported adjustment of the original algorithm to adapt to neonates, of which only two provided specific modification details. In absence of detailed reporting on the parameter settings, it is difficult to determine whether the settings adhered to neonatal waveform morphology. Incomplete reporting and lack of transparency hinders understanding the strengths and weaknesses of a study and limits its reproducibility and usability. Moreover, transparent and detailed reporting is required to confirm the adherence to regulatory compliance and is crucial for clinical adoption of these methods.

Like QRS complex in ECG signals, the acceptable ranges of physiological signals for neonates are also different from the adult population. This review found that no studies reviewed the acceptable ranges of the analysed signals against any published guidelines, which could pose several limitations in the clinical adoption of the methods. This is consistent with another review looking into physiological vital sign ranges from 34 weeks gestational age which identified that several studies reported the means of vital signs, instead of ranges, which makes the interpretation into clinical practice difficult [104]. Here we recommend clear reporting and the use of physiological signal ranges that are clinically validated through published studies and textbooks [105–107].

Pre-processing of physiological data typically involves several steps, including the handling of missing data, filtering, segmentation, and waveform analysis for feature extraction. Here we define five required pre-processing steps (based on the steps outlined in Berkaya et al [13]) and identify the steps reported by each of the studies in this review (see Table 3). Below are the definition of each of the steps,

1. **Handling of Missing Data -** During neonatal physiological monitoring, instances of missing data may arise due to sensor disconnection, improper placements, or signal dropouts. To tackle this issue, methodologies like data imputation or interpolation are applied. For example, if gaps exist in a neonate's heart rate monitoring data, interpolation methods can estimate the missing values by considering neighbouring data points. Widely used interpolation techniques include linear interpolation, spline interpolation, and time-based interpolation. Additionally, common data imputation methods involve forward fill, backward fill, and imputation using mean or median values. Methods such as forward fill [43], moving average [61], mean imputation [93, 95] and interpolation [96] were used by some studies reviewed in this paper.
2. **Artefact Removal** - Neonatal signals can be affected by artefacts, like those from muscle movements or electrical interference. Commonly employed techniques, such as bandpass or notch filters, along with moving averages, are used to effectively eliminate these disturbances. For instance, in neonatal electroencephalogram (EEG) signals, adaptive filters prove beneficial in eliminating artefacts caused by muscle movements, resulting in a clearer representation of the baby's brain activity. Some methods used by the reviewed papers were, high-pass filter [39, 64], band-pass filter [41, 46, 61, 62, 80].
3. **Resampling, Normalisation** - Resampling is a technique that standardises data intervals, involving either upsampling (increasing data point frequency) or downsampling (decreasing frequency) to create a regular time-series. This aligns signals from different devices or physiological sources. Normalisation ensures uniformity and reliability across these standardised sampling rates. For instance, if neonatal heart rate signals from different devices have varied sampling rates, resampling achieves a common rate, while normalisation, using techniques like min-max, z-score, or log scale, ensures consistent amplitude scaling for accurate comparative analysis. In the reviewed studies, normalization techniques such as min-max [77] and zero mean normalization [41, 84] were used. In terms of resampling, both downsampling [46, 47, 57] and upsampling [55] techniques were used.
4. **Waveform Feature Extraction** - Extracting relevant features from a signal's waveform is a fundamental step in signal pre-processing. This involves identifying key characteristics such as peaks, troughs, or other significant points in the signal. In the context of neonatal electrocardiogram (ECG), feature extraction may involve identifying key points such as R-peaks to analyse HRV, providing valuable insights into the infant's autonomic nervous system development. Pan-Tompkins algorithm is a popular method chosen by multiple papers reviewed in this study that conducted R peak detection from the QRS complex [29, 34, 39, 46, 49, 55].
5. **Data Segmentation** - Segmenting data is the process of breaking down a continuous signal into smaller, more manageable sections to enable targeted analysis. This practice is especially beneficial when dealing with lengthy signals. Data segmentation is a common pre-processing step in machine learning workflows. For instance, in the analysis of neonatal sleep patterns using electroencephalography (EEG), data segmentation can involve dividing the

continuous EEG signal into epochs, allowing for the identification and study of sleep stages in shorter, more manageable segments. Commonly used segmentation techniques include fixed-length, sliding window, threshold based and feature-based segmentation. Some of the data segmentation size used in the reviewed studies were, 30-second [29, 32, 62, 108] and 1-minute [57] epochs, sliding window of varied sizes [49, 56, 79, 84, 93].

In neonatal physiological signal processing, these pre-processing techniques contribute to the accurate interpretation of signals, aiding healthcare professionals in monitoring and providing appropriate care in the NICU or other clinical settings.

Table 3: Required physiological signal pre-processing steps reported by each of the studies in this review

Study	Required pre-processing step reported				
	Handling of missing data	**Artefact removal**	**Resampling, normalization**	**Waveform feature extraction**	**Data segmentation**
Cohen and Chazal, 2013 [29]	ü	ü	û	ü	ü
Cohen and Chazal, 2014 [32]	ü	ü	ü	ü	ü
Cohen and Chazal, 2015 [34]	ü	ü	û	ü	ü
Gee et al., 2016 [37]	û	û	û	ü	ü
Gee et al., 2017 [35]	û	ü	û	ü	ü
Das et al. 2019 [39]	û	ü	û	ü	ü
Mahmud et al. 2019 [40]	û	û	û	û	û
Gee et al., 2019 [41]	û	ü	ü	ü	ü
Song et al. 2020 [43]	ü	ü	û	N/A	û
Baker et el. 2021 [44]	û	ü	ü	N/A	ü
Juraev et al. 2022 [45]	ü	û	ü	N/A	û
Ghahjaverestan et al. 2015 [46]	û	ü	ü	ü	û
Navarro et al. 2015 [47]	û	ü	ü	ü	ü
Ghahjaverestan et al. 2016 [49]	û	ü	ü	ü	ü
Leon et al. 2021 [51]	û	û	û	ü	ü
Leon et al. 2021 [53]	û	û	û	ü	ü
Doyen et al. 2021 [54]	û	û	û	ü	ü
Sadoughi et al. 2021 [55]	û	ü	ü	ü	û
Ahmed et al. 2015 [56]	ü	ü	ü	ü	ü
Temko et al. 2015 [57]	û	ü	ü	ü	ü
Lloyd et al. 2016 [59]	û	ü	û	N/A	ü
Semenova et al. 2018 [60]	û	ü	û	ü	ü
Semenova et al. 2019 [61]	ü	ü	ü	ü	ü
Joshi et al. 2020 [62]	û	ü	û	ü	ü
Varisco et al. 2021 [64]	û	ü	ü	ü	û
Cabrera-Quiros et al. 2021 [68]	û	û	û	ü	ü
Varisco et al. 2022 [69]	û	ü	ü	ü	ü

Peng et al. 2022 [71]	û	û	û	ü	ü
Peng et al. 2022 [72]	û	ü	û	ü	ü
Stanculescu et al. 2014 [74]	ü	ü	û	N/A	û
Stanculescu et al. 2014 [76]	ü	ü	û	N/A	û
Shirwaikar et al. 2016 [77]	ü	ü	ü	N/A	ü
Shirwaikar et al. 2019 [78]	ü	û	û	N/A	ü
Williamson et al. 2013 [79]	û	û	ü	û	ü
Schiavenato et al. 2013 [80]	ü	ü	û	ü	û
Robles-Rubio et al. 2014 [82]	û	ü	ü	N/A	û
Amperayani et al. 2017 [83]	û	û	û	û	û
Hu et al. 2018 [84]	ü	ü	ü	N/A	ü
Sullivan et al. 2018 [85]	ü	û	û	N/A	ü
Zimmet et al. 2020 [86]	ü	û	û	N/A	ü
Kota et al. 2020 [89]	û	ü	û	ü	ü
Mirnia et al. 2021 [92]	û	û	û	N/A	û
Lee et al. 2021 [93]	ü	ü	ü	N/A	ü
Sullivan et al. 2021 [94]	ü	ü	û	N/A	ü
Feng et al. 2021 [95]	ü	ü	ü	N/A	ü
Zuzarte et al. 2021 [96]	û	ü	ü	ü	ü
Niestroy et al. 2022 [97]	û	û	û	N/A	ü
Kausch et al. 2023 [98]	ü	ü	û	N/A	ü
Honore et al. 2023 [99]	ü	ü	ü	N/A	ü
Masoudi et al. 2013 [100]	û	û	û	û	ü
Altuve et al. 2015 [101]	û	ü	ü	û	ü
Honore et al. 2020 [102]	ü	û	û	N/A	ü

It can be seen from Table 3 that only seven out of the 52 reviewed studies reported all the recommended pre-processing steps. This could have several impacts on the downstream analysis. For instance, several papers missed reporting on how they segmented the data for feature extraction and classification though it is essential for clinical validation in cases where the segment duration is dependent on the adverse outcome prediction performance. In HRV analysis, it is important to indicate whether it is a short-term (~5 minutes) or a long-term (>=24 hours) analysis as they reflect different underlying physiological process and thus, demonstrate different predictive power [109]. Along with the segment duration, additional information such as sampling rate of the signals will provide a clear reflection of the dataset size. Downsampling the data to a low sampling rate (e.g. 50 Hz) has also shown significant impact in HRV analysis [110]. Although all the reviewed studies mentioned the participant number, and majority of them reported the sampling rate of the signals, very few provided details on the sample size or dataset duration, or whether the dataset was resampled for subsequent analysis. These elements provide a clearer picture on the computational time and resources required for clinical validation and adoption. Though physiological recordings collected in the NICU environment suffer greatly from missing data due to similar factors that introduce artefacts [111], reporting how missing data are handled is scarce. Different methods for

dealing with missing values could cause different results and not all might be suitable for a particular problem. Therefore, it is important to report all the details related to adopted approach.

The incomplete or partial reporting found in these studies has significant implications for the implementation of QMS in utilising these techniques for clinical adoption. A well implementation of QMS requires comprehensive reporting of each intermediary step involved in constructing an AI/ML pipeline. The International Medical Device Regulators Forum (IMDRF) offers guidance on the clinical evaluation required for any product intended for use as a medical device [112]. According to the IMDRF guidelines, during clinical evaluation, relevant research articles are reviewed to identify clinical evidence supporting the product [113] . The guideline encourages manufacturers to follow these recognised standards and best practices in the development, validation, and manufacturing processes. Clinical evaluations are required by the EU medical device regulation, and it is also mentioned in the ISO 13485 – the quality management standard for medical devices. Thus, detailed reporting is crucial as it can be utilised for the clinical evaluation of future SaMD products by regulatory bodies. Steps such as the missing data handling procedures are also required by the TRIPOD checklist for model development and validation, that assesses risk of bias and clinical usefulness of the prediction model [114]. As another example, a questionnaire prepared by the German Notified Body Interest Group has been adopted in assessing some AI-powered medical products in the EU. This questionnaire includes inquiries about data management that includes data collection, labeling, pre-processing procedures, and relevant documentation. Transparent and detailed reporting of these steps is essential to ensure the safety, efficacy, and reliability of software as a medical device.

Discussion

Principal Results

The aim of this review was to summarise computational methods used for pre-processing preterm infants' physiological data as a first step in developing data-driven predictive models for adverse outcomes related to clinical decision support. This is an important step, especially from a clinician's perspective because it increases the trustworthiness of the developed models by allowing for the verification and reproduction of the results. In addition, it aids in achieving regulatory compliance and ensures the safety, efficacy, and ethical use of AI-based healthcare devices. Furthermore, it allows us to recognise shortcomings in current state of the art studies and recommend guidelines for transparent reporting. The review found that studies were heterogeneous in terms of their methods and applications. Therefore, a narrative approach of reporting the results was taken instead of a quantitative approach. Through the analysis we identified several key components incomplete or partially reported by the included studies, which were summarised in Table 3. To ensure transparent reporting for any future studies in this area we recommend detailed reporting of all pre-processing steps listed in Table 3 which will

allow revealing their strengths, weaknesses and ultimately making them usable and reproducible. Reproducible research allows clinicians to make more informed decisions about patient care and treatment based on evidence that has been thoroughly assessed.

Comparison with Prior Works

Reviews published in recent years have highlighted the potential of big data and artificial intelligence in supporting clinical decision making in the neonatal healthcare domain [10, 15, 21, 115, 116], particularly in using the physiological data for detecting or predicting neonatal health outcomes. However, appropriate pre-processing of this data is a prerequisite for developing clinically deployable models. A systematic review by McAdams et al [10] reported different ML models used to predict different clinical outcomes in neonates. However, their primary focus was on five neonatal morbidities and they did not focus on reporting the pre-processing methods applied prior to building the ML models. Furthermore, they did not include studies using real-time continuous physiological data; 28 out of their 68 studies were based on physiological data (not continuous) and the rest were based on electronic medical records (EMR) and imaging data. The authors in [15] performed a review on HRV analysis for neonates. The primary limitation of the work was the lack of reporting in details about the pre-processing steps of ECG signals prior to HRV analysis, such as ECG handling and segmentation, R-wave (QRS complex) identification technique, software and parameters, and ranges of all HRV features. They identified these components incomplete or missing in the studies they reviewed thus recommended clear reporting of these aspects for future studies in this area. These limitations served as a motivation for our review to focus on the pre-processing techniques of the neonatal physiological signals in broader sense, which serves as the preliminary step for any big data-based approaches.

Limitations

There are several limitations of this review. While all of the included studies were independently reviewed by two reviewers and conflicts were resolved by a third reviewer, the initial searching and screening was conducted by one reviewer, which may introduce bias. Also, this review did not include a quantitative or comparative analysis of the reviewed studies as the techniques used to analyse the physiological signals were diverse. Future work could include a quantitative evaluation of the studies that were homogeneous in design.

Conclusion

This review explores the computational methods employed by the current state of the art machine learning driven clinical decision support approaches to pre-process physiological signals collected from infants treated in the neonatal setting. A summary of the studies identified heterogeneity in techniques used for analysis and revealed a lack of consistent and detailed reporting, which is important for building robust, transparent and clinically deployable prediction models. The availability of powerful hardware and software

resources in the NICU environment and growing interest in big data and AI is driving strong demand for clinical decision support applications. We recommend clear reporting of the different steps in the pre-processing of the neonatal physiological signals to ensure transparency in clinical validation and accelerate the adoption of developed models in the clinical setting. This will further enhance delivery and adoption of reliable, regulatory-compliant safe, and effective products in healthcare.

Acknowledgements

The work is supported by Cerebral Palsy Alliance and University of Sydney (Ethics approval number 2021/ETH00824).

Data Availability

The paper is accompanied by supplementary files.

Conflicts of Interest

None Declared

Abbreviations

AB: Apnea Bradycardia
ABH: Apnea Bradycardia Hypoxia
ABP: Arterial Blood Pressure
AI: Artificial Intelligence
ART-M: Arterial Mean Blood Pressure
AUC: Area Under Curve
AUROC: Area Under Receiver Operating Characteristic curve
BPD: Bronchopulmonary dysplasia
CDSS: Clinical Decision Support System
CI: Chest Impedance
CNN: Convolutional Neural Network
CWT: Continuous Wavelet Transform
DBP: Diastolic Blood Pressure
DT-CWT: Discrete Time Continuous Wavelet Transform
ECG: Electrocardiogram
FDA: Food and Drug Administration
FPR: False Positive Rate
GMM: Gaussian Mixture Model
HIE: Hypoxic Ischemic Encephalopathy
HMM: Hidden Markov Model
HR: Heart Rate
HRV: Heart Rate Variability
IBI: Interbreath Intervals
IMDRF: The International Medical Device Regulators Forum
IVH: Intraventricular Hemorrhage

LSTM: Long Short-Term Memory
MAP: Mean Arterial Pressure
MBP: Mean Blood Pressure
MeSH: Medical Subject Headings
ML: Machine Learning
NEC: Necrotizing Enterocolitis
NIBP-M: Non-invasive Mean Blood Pressure
NICU: Neonatal Intensive Care Unit
PR: Pulse Oximeter
PRISMA: Preferred Reporting in Systematic Reviews and Meta-Analyses
RF: Random Forest
RIP: Respiratory Inductive Plethysmograph
RNN: recurrent Neural Network
ROP: Retinopathy of prematurity
RRE: Ribcage Respiratory Effort
SBP: Systolic Blood Pressure
SD: Standard Deviation
SII: Signal Instability Index
SVM: Support Vector Machines
TEMP: Temperature
TGA: Therapeutic Goods Administration
XGBoost: Extreme Gradient Boosting

References

1. Platt MJ. Outcomes in preterm infants. Public Health. 2014 May 1;1285:399–403. https://doi.org/10.1016/j.puhe.2014.03.010.
2. Seaton SE, Barker L, Draper ES, Abrams KR, Modi N, Manktelow BN. Estimating neonatal length of stay for babies born very preterm. Arch Dis Child Fetal Neonatal Ed. 2019 March;1042:F182–6. PMID: 29588296. https://doi.org/10.1136/archdischild-2017-314405.
3. De Georgia MA, Kaffashi F, Jacono FJ, Loparo KA. Information Technology in Critical Care: Review of Monitoring and Data Acquisition Systems for Patient Care and Research. Sci World J. 2015 February 4;2015:e727694. https://doi.org/10.1155/2015/727694.
4. Andreu-Perez J, Poon CCY, Merrifield RD, Wong STC, Yang G-Z. Big Data for Health. IEEE J Biomed Health Inform. 2015 July;194:1193–208. https://doi.org/10.1109/JBHI.2015.2450362.
5. Kumar N, Akangire G, Sullivan B, Fairchild K, Sampath V. Continuous vital sign analysis for predicting and preventing neonatal diseases in the twenty-first century: big data to the forefront. Pediatr Res. 2020 January;872:210–20. https://doi.org/10.1038/s41390-019-0527-0.
6. Baljak V, Ljubovic A, Michel J, Montgomery M, Salaway R. A scalable realtime analytics pipeline and storage architecture for physiological monitoring big data. Smart Health. 2018 December 1;9–10:275–86. https://doi.org/10.1016/j.smhl.2018.07.013.

7. Bright TJ, Wong A, Dhurjati R, et al. Effect of Clinical Decision-Support Systems. Ann Intern Med. 2012 July 3;1571:29–43. https://doi.org/10.7326/0003-4819-157-1-201207030-00450.
8. Comito C, Falcone D, Forestiero A. Current Trends And Practices In Smart Health Monitoring And Clinical Decision Support. 2020 IEEE Int. Conf. Bioinforma. Biomed. BIBM, 2020 December, p. 2577–84.
9. Yu K-H, Kohane IS. Framing the challenges of artificial intelligence in medicine. BMJ Qual Saf. 2019 March 1;283:238–41. PMID: 30291179. https://doi.org/10.1136/bmjqs-2018-008551.
10. McAdams RM, Kaur R, Sun Y, Bindra H, Cho SJ, Singh H. Predicting clinical outcomes using artificial intelligence and machine learning in neonatal intensive care units: a systematic review. J Perinatol. 2022 December;4212:1561–75. https://doi.org/10.1038/s41372-022-01392-8.
11. Mahadevaiah G, Rv P, Bermejo I, Jaffray D, Dekker A, Wee L. Artificial intelligence-based clinical decision support in modern medical physics: Selection, acceptance, commissioning, and quality assurance. Med Phys. 2020;475:e228–35. https://doi.org/10.1002/mp.13562.
12. Sloane EB, J. Silva R. Chapter 83 - Artificial intelligence in medical devices and clinical decision support systems. In: Iadanza E, editor. Clin. Eng. Handb. Second Ed., Academic Press; 2020 January 1, p. 556–68.
13. Kaplan Berkaya S, Uysal AK, Sora Gunal E, Ergin S, Gunal S, Gulmezoglu MB. A survey on ECG analysis. Biomed Signal Process Control. 2018 May 1;43:216–35. https://doi.org/10.1016/j.bspc.2018.03.003.
14. Sweeney KT, Ward TE, McLoone SF. Artifact Removal in Physiological Signals—Practices and Possibilities. IEEE Trans Inf Technol Biomed. 2012 May;163:488–500. https://doi.org/10.1109/TITB.2012.2188536.
15. Latremouille S, Lam J, Shalish W, Sant'Anna G. Neonatal heart rate variability: a contemporary scoping review of analysis methods and clinical applications. BMJ Open. 2021 December;1112:e055209. https://doi.org/10.1136/bmjopen-2021-055209.
16. Baird P, Cobbaert K. Medical Device White Paper Series. Software as a medical device – A comparison of the EU's approach with the US's approach. BSI. 2020.
17. Simera I. Get the content right: following reporting guidelines will make your research paper more complete, transparent and usable. JPMA J Pak Med Assoc. 2013 February 1;632:283–5. PMID: 23894916.
18. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev. 2021 March 29;101:89. https://doi.org/10.1186/s13643-021-01626-4.
19. Centre for Reviews and Dissemination. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Healthcare. York, England: University of York NHS Centre for Reviews & Dissemination; 2008 December 1.
20. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev. 2016 December 5;51:210. https://doi.org/10.1186/s13643-016-0384-4.
21. Mangold C, Zoretic S, Thallapureddy K, Moreira A, Chorath K, Moreira A. Machine Learning Models for Predicting Neonatal Mortality: A Systematic Review. Neonatology. 2021;1184:394–405. PMID: 34261070. https://doi.org/10.1159/000516891.

22. Phillips Continuous Patient Monitoring Systems n.d. https://www.philips.com.au/healthcare/solutions/patient-monitoring/continuous-patient-monitoring-systems.
23. BedMasterEx. Real time clinical data acquisition n.d. https://www.anandic.com/produkte/produkt-detail/bedmasterex/.
24. NicoletOne™ EEG System n.d. https://natus.com/products-services/nicoletone-eeg-system.
25. ixTrend – recording medical signals from patient monitors n.d. https://www.ixellence.com/index.php/en/home/17-default-en/products.
26. Patient Information Center iX (PIC iX) n.d. https://www.philips.com.au/healthcare/product/HCNOCTN171/patient-information-center-ix-pic-ix.
27. Mann KD, Good NM, Fatehi F, et al. Predicting Patient Deterioration: A Review of Tools in the Digital Hospital Setting. J Med Internet Res. 2021 September 30;239:e28209. https://doi.org/10.2196/28209.
28. Crowell DH, Brooks LJ, Colton T, et al. Infant Polysomnography: Reliability. Sleep. 1997 July 1;207:553–60. https://doi.org/10.1093/sleep/20.7.553.
29. G. Cohen, P. de Chazal. Detection of sleep apnoea in infants using ECG and oximetry signals. Comput. Cardiol. 2013, 2013, p. 859–62.
30. Pan J, Tompkins WJ. A Real-Time QRS Detection Algorithm. IEEE Trans Biomed Eng. 1985 March;BME-323:230–6. https://doi.org/10.1109/TBME.1985.325532.
31. de Chazal P, Heneghan C, Sheridan E, Reilly R, Nolan P, O'Malley M. Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea. IEEE Trans Biomed Eng. 2003 June;506:686–96. https://doi.org/10.1109/TBME.2003.812203.
32. G. Cohen, P. de Chazal. A multi-modal approach to sleep-wake classification in infants using minimally invasive sensors. Comput. Cardiol. 2014, 2014, p. 149–52.
33. Lewicke AT, Sazonov ES, Schuckers SAC. Sleep-wake identification in infants: heart rate variability compared to actigraphy. 26th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., vol. 1, 2004 September, p. 442–5.
34. Cohen G, de Chazal P. Automated detection of sleep apnea in infants: A multi-modal approach. Comput Biol Med. 2015;63:118–23. https://doi.org/10.1016/j.compbiomed.2015.05.007.
35. Gee AH, Barbieri R, Paydarfar D, Indic P. Predicting Bradycardia in Preterm Infants Using Point Process Analysis of Heart Rate. IEEE Trans Biomed Eng. 2017 September;649:2300–8. https://doi.org/10.1109/TBME.2016.2632746.
36. PhysioBank, PhysioToolkit, and PhysioNet n.d. https://www.ahajournals.org/doi/epub/10.1161/01.CIR.101.23.e215 (accessed December 15, 2022).
37. A. H. Gee, R. Barbieri, D. Paydarfar, P. Indic. Improving heart rate estimation in preterm infants with bivariate point process analysis of heart rate and respiration. 2016 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC, 2016, p. 920–3.
38. Torrence C, Compo GP. A Practical Guide to Wavelet Analysis. Bull Am Meteorol Soc. 1998 January 1;791:61–78. https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2.
39. S. Das, B. Moraffah, A. Banerjee, S. K. S. Gupta, A. Papandreou-Suppappola.

Bradycardia Prediction in Preterm Infants Using Nonparametric Kernel Density Estimation. 2019 53rd Asilomar Conf. Signals Syst. Comput., 2019, p. 1309–13.
40. Md Mahmud S, Wang H, Kim Y. Accelerated Prediction of Bradycardia in Preterm Infants Using Time-Frequency Analysis, ["Electrical and Computer Engineering, University of New Hampshire, Dover, NH, United States", “Electrical and Computer Engineering, University of Massachusetts Dartmouth, Dartmouth, MA, United States”, “Biomedical Engineering, University of Massachusetts Dartmouth, Dartmouth, MA, United States”]: 2019, p. 468–72.
41. Gee AH, Garcia-Olano D, Ghosh J, Paydarfar D. Explaining deep classification of time-series data with learned prototypes. vol. 2429, ["Electrical and Computer Engineering, University of Texas at Austin, United States", “Neurology, Dell Medical School, University of Texas at Austin, United States”]: 2019, p. 15–22.
42. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. Sci Data. 2016 May 24;31:160035. https://doi.org/10.1038/sdata.2016.35.
43. Song W, Jung S, Baek H, Choi C, Jung Y, Yoo S. A Predictive Model Based on Machine Learning for the Early Detection of Late-Onset Neonatal Sepsis: Development and Observational Study. JMIR Med Inform. 2020;87. https://doi.org/10.2196/15965.
44. Baker S, Xiang W, Atkinson I. Hybridized neural networks for non-invasive and continuous mortality risk assessment in neonates. Comput Biol Med. 2021;134:104521. https://doi.org/10.1016/j.compbiomed.2021.104521.
45. Juraev F, El-Sappagh S, Abdukhamidov E, Ali F, Abuhmed T. Multilayer dynamic ensemble model for intensive care unit mortality prediction of neonate patients. J Biomed Inform. 2022 November;135:104216. PMID: 36208833. https://doi.org/10.1016/j.jbi.2022.104216.
46. Ghahjaverestan NM, Shamsollahi MB, Ge D, Hernández AI. Switching Kalman filter based methods for apnea bradycardia detection from ECG signals. Physiol Meas. 2015;369:1763–83. https://doi.org/10.1088/0967-3334/36/9/1763.
47. Navarro X, Poree F, Beuchee A, Carrault G. Artifact rejection and cycle detection in immature breathing: Application to the early detection of neonatal sepsis. Biomed SIGNAL Process CONTROL. 2015;16:9–16. https://doi.org/10.1016/j.bspc.2014.10.007.
48. Savitzky Abraham, Golay MJE. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal Chem. 1964 July 1;368:1627–39. https://doi.org/10.1021/ac60214a047.
49. Ghahjaverestan N, Masoudi S, Shamsollahi M, et al. Coupled Hidden Markov Model-Based Method for Apnea Bradycardia Detection. IEEE J Biomed Health Inform. 2016;202:527–38. https://doi.org/10.1109/JBHI.2015.2405075.
50. Dumont J, Hernandez AI, Fleureau J, Carrault G. Modelling temporal evolution of cardiac electrophysiological features using Hidden Semi-Markov Models. 2008 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 2008 August, p. 165–8.
51. C. León, G. Carrault, P. Pladys, A. Beuchée. Early Detection of Late Onset Sepsis in Premature Infants Using Visibility Graph Analysis of Heart Rate Variability. IEEE J Biomed Health Inform. 2021;254:1006–17. https://doi.org/10.1109/JBHI.2020.3021662.
52. Doyen M, Ge D, Beuchée A, Carrault G, Hernández AI. Robust, real-time generic detector based on a multi-feature probabilistic method. PLOS ONE. 2019 October 29;1410:e0223785. https://doi.org/10.1371/journal.pone.0223785.
53. C. León, P. Pladys, A. Beuchée, G. Carrault. Recurrent Neural Networks for Early

Detection of Late Onset Sepsis in Premature Infants Using Heart Rate Variability. 2021 Comput. Cardiol. CinC, vol. 48, 2021, p. 1–4.
54. Doyen M, Hernández AI, Flamant C, et al. Early bradycardia detection and therapeutic interventions in preterm infant monitoring. Sci Rep. 2021;111:10486. https://doi.org/10.1038/s41598-021-89468-x.
55. Sadoughi A, Shamsollahi M, Fatemizadeh E, Beuche A, Hernndez A, Ghahjaverestan N. Detection of Apnea Bradycardia from ECG Signals of Preterm Infants Using Layered Hidden Markov Model. Ann Biomed Eng. 2021;499:2159–69. https://doi.org/10.1007/s10439-021-02732-z.
56. Ahmed R, Temko A, Marnane WP, Boylan G, Lightbody G. Classification of hypoxic-ischemic encephalopathy using long term heart rate variability based features. 2015 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC, 2015 August, p. 2355–8.
57. Temko A, Doyle O, Murray D, Lightbody G, Boylan G, Marnane W. Multimodal predictor of neurodevelopmental outcome in newborns with hypoxic-ischaemic encephalopathy. Comput Biol Med. 2015;63:169–77. https://doi.org/10.1016/j.compbiomed.2015.05.017.
58. Hamilton PS, Tompkins WJ. Quantitative Investigation of QRS Detection Rules Using the MIT/BIH Arrhythmia Database. IEEE Trans Biomed Eng. 1986 December;BME-3312:1157–65. https://doi.org/10.1109/TBME.1986.325695.
59. Lloyd R, O'Toole J, Livingstone V, et al. Predicting 2-y outcome in preterm infants using early multimodal physiological monitoring. Pediatr Res. 2016;803:382–8. https://doi.org/10.1038/pr.2016.92.
60. O. Semenova, G. Carra, G. Lightbody, G. Boylan, E. Dempsey, A. Temko. Heart Rate Variability during Periods of Low Blood Pressure as a Predictor of Short-Term Outcome in Preterms. 2018 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC, 2018, p. 5614–5517.
61. Semenova O, Carra G, Lightbody G, Boylan G, Dempsey E, Temko A. Prediction of short-term health outcomes in preterm neonates from heart-rate variability and blood pressure using boosted decision trees. Comput Methods Programs Biomed. 2019 October;180:104996. https://doi.org/10.1016/j.cmpb.2019.104996.
62. Joshi R, Kommers D, Oosterwijk L, Feijs L, van Pul C, Andriessen P. Predicting Neonatal Sepsis Using Features of Heart Rate Variability, Respiratory Characteristics, and ECG-Derived Estimates of Infant Motion. IEEE J Biomed Health Inform. 2020 March;243:681–92. https://doi.org/10.1109/JBHI.2019.2927463.
63. Rooijakkers MJ, Rabotti C, Oei SG, Mischi M. Low-complexity R-peak detection for ambulatory fetal monitoring. Physiol Meas. 2012 June;337:1135. https://doi.org/10.1088/0967-3334/33/7/1135.
64. G. Varisco, D. Kommers, X. Long, et al. Optimized Detection of Central Apneas Preceding Late-Onset Sepsis in Premature Infants. 2021 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC, 2021, p. 5463–8.
65. Lee H, Rusin CG, Lake DE, et al. A new algorithm for detecting central apnea in neonates. Physiol Meas. 2011 December;331:1. https://doi.org/10.1088/0967-3334/33/1/1.
66. Mohr MA, Vergales BD, Lee H, et al. Very long apnea events in preterm infants. J Appl Physiol. 2015 March 1;1185:558–68. PMID: 25549762. https://doi.org/10.1152/japplphysiol.00144.2014.

67. Vergales BD, Paget-Brown AO, Lee H, et al. Accurate Automated Apnea Analysis in Preterm Infants. Am J Perinatol. 2014 February;312:157–62. PMID: 23592319. https://doi.org/10.1055/s-0033-1343769.
68. Cabrera-Quiros L, Kommers D, Wolvers MK, et al. Prediction of Late-Onset Sepsis in Preterm Infants Using Monitoring Signals and Machine Learning. Crit Care Explor. 2021;31:e0302. https://doi.org/10.1097/CCE.0000000000000302.
69. Varisco G, Peng Z, Kommers D, et al. Central apnea detection in premature infants using machine learning. Comput Methods Programs Biomed. 2022 November;226:107155. PMID: 36215858. https://doi.org/10.1016/j.cmpb.2022.107155.
70. Redmond SJ, Heneghan C. Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea. IEEE Trans Biomed Eng. 2006 March;533:485–96. https://doi.org/10.1109/TBME.2005.869773.
71. Peng Z, Varisco G, Liang R-H, et al. DeepLOS: Deep learning for late-onset sepsis prediction in preterm infants using heart rate variability. Smart Health. 2022 December 1;26:100335. https://doi.org/10.1016/j.smhl.2022.100335.
72. Peng Z, Varisco G, Long X, et al. A Continuous Late-Onset Sepsis Prediction Algorithm for Preterm Infants Using Multi-Channel Physiological Signals From a Patient Monitor. IEEE J Biomed Health Inform. 2022 October 20;PP. PMID: 36264730. https://doi.org/10.1109/JBHI.2022.3216055.
73. Zuzarte I, Indic P, Sternad D, Paydarfar D. Quantifying Movement in Preterm Infants Using Photoplethysmography. Ann Biomed Eng. 2019 February;472:646–58. PMID: 30255214. https://doi.org/10.1007/s10439-018-02135-7.
74. Stanculescu I, Williams CKI, Freer Y. Autoregressive Hidden Markov Models for the Early Detection of Neonatal Sepsis. IEEE J Biomed Health Inform. 2014;185:1560–70. https://doi.org/10.1109/JBHI.2013.2294692.
75. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE. 1989 February;772:257–86. https://doi.org/10.1109/5.18626.
76. Stanculescu I, Williams CKI, Freer Y. A hierarchical switching linear dynamical system applied to the detection of sepsis in neonatal condition monitoring, ["School of Informatics, University of Edinburgh, Edinburgh, EH8 9AB, United Kingdom", “Simpson Centre for Reproductive Health, Royal Infirmary of Edinburgh, Edinburgh, EH16 4SA, United Kingdom”]: 2014, p. 752–61.
77. Shirwaikar RD, Acharya UD, Makkithaya K, Surulivelrajan M, Lewis LES. Machine learning techniques for neonatal apnea prediction. J Artif Intell. 2016;91:33–8. https://doi.org/10.3923/jai.2016.33.38.
78. Shirwaikar RD, Acharya U D, Makkithaya K, M S, Srivastava S, Lewis U LES. Optimizing neural networks for medical data sets: A case study on neonatal apnea prediction. Artif Intell Med. 2019;98:59–76. https://doi.org/10.1016/j.artmed.2019.07.008.
79. Williamson JR, Bliss DW, Browne DW, Indic P, Bloch-Salisbury E, Paydarfar D. Individualized apnea prediction in preterm infants using cardio-respiratory and movement signals, ["Massachusetts Institute of Technology, Lincoln Laboratory, Lexington, MA, United States", “School of Electrical, Computer and Energy Engineering, Arizona State University, Glendale, AZ, United States”, “Department of Neurology, University of Massachusetts Medical School, Worcester, MA, United States”, “Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, United States”]:

2013.
80. Schiavenato M, Oliu C, Bello E, Bohorquez J, Claure N. Development of a System for the Assessment of Heart Rate Variability in the Neonatal Intensive Care Unit (NICU), 2013, p. 25-+.
81. Lomb NR. Least-squares frequency analysis of unequally spaced data. Astrophys Space Sci. 1976 February 1;392:447–62. https://doi.org/10.1007/BF00648343.
82. C. A. Robles-Rubio, K. A. Brown, G. Bertolizio, R. E. Kearney. Automated analysis of respiratory behavior for the prediction of apnea in infants following general anesthesia. 2014 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 2014, p. 262–5.
83. Apurupa Amperayani V, Indic P, Travers CP, Barbieri R, Paydarfar D, Ambalavanan N. An algorithm for risk stratification of preterm infants. vol. 44, ["Department of Electrical Engineering, University of Texas at Tyler, 3900 University Blvd., Tyler, TX 75799, United States", “Department of Pediatrics, University of Alabama at Birmingham, Birmingham, United States”, “Politecnico Di Milano, I-Milano, Italy”, “Dell Medical School, University of Texas at Austin, Austin, United States”]: 2017, p. 1–4.
84. Y. Hu, V. C. S. Lee, K. Tan. Prediction of clinicians’ treatment in preterm infants with suspected late-onset sepsis — An ML approach. 2018 13th IEEE Conf. Ind. Electron. Appl. ICIEA, 2018, p. 1177–82.
85. Sullivan B, Wallman-Stokes A, Isler J, et al. Early Pulse Oximetry Data Improves Prediction of Death and Adverse Outcomes in a Two-Center Cohort of Very Low Birth Weight Infants. Am J Perinatol. 2018 November;3513:1331–8. https://doi.org/10.1055/s-0038-1654712.
86. Zimmet AM, Sullivan BA, Moorman JR, Lake DE, Ratcliffe SJ. Trajectories of the heart rate characteristics index, a physiomarker of sepsis in premature infants, predict Neonatal ICU mortality. JRSM Cardiovasc Dis. 2020;9:2048004020945142. https://doi.org/10.1177/2048004020945142.
87. Kota S, Massaro A, Chang T, et al. Prognostic Value of Continuous Electroencephalogram Delta Power in Neonates With Hypoxic-Ischemic Encephalopathy. J CHILD Neurol. 2020;358:517–25. https://doi.org/10.1177/0883073820915323.
88. Govindan RB, Kota S, Al-Shargabi T, Massaro AN, Chang T, du Plessis A. Effect of electrocardiogram interference on cortico-cortical connectivity analysis and a possible solution. J Neurosci Methods. 2016 September 1;270:76–84. https://doi.org/10.1016/j.jneumeth.2016.06.009.
89. Kota S, du Plessis A, Massaro AN, Chang T, Al-Shargabi T, Govindan RB. A frequency based spatial filter to mitigate volume conduction in electroencephalogram signals. 2016 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC, 2016 August, p. 4001–4.
90. Govindan RB, Massaro A, Vezina G, Tsuchida T, Cristante C, du Plessis A. Does relative or absolute EEG power have prognostic value in HIE setting? Clin Neurophysiol Off J Int Fed Clin Neurophysiol. 2017 January;1281:14–5. PMID: 27866114. https://doi.org/10.1016/j.clinph.2016.10.094.
91. Halliday DM, Rosenberg JR, Amjad AM, Breeze P, Conway BA, Farmer SF. A framework for the analysis of mixed time series/point process data—Theory and application to the study of physiological tremor, single motor unit discharges and electromyograms. Prog Biophys Mol Biol. 1995 January 1;642:237–78. https://doi.org/10.1016/S0079-6107(96)00009-0.

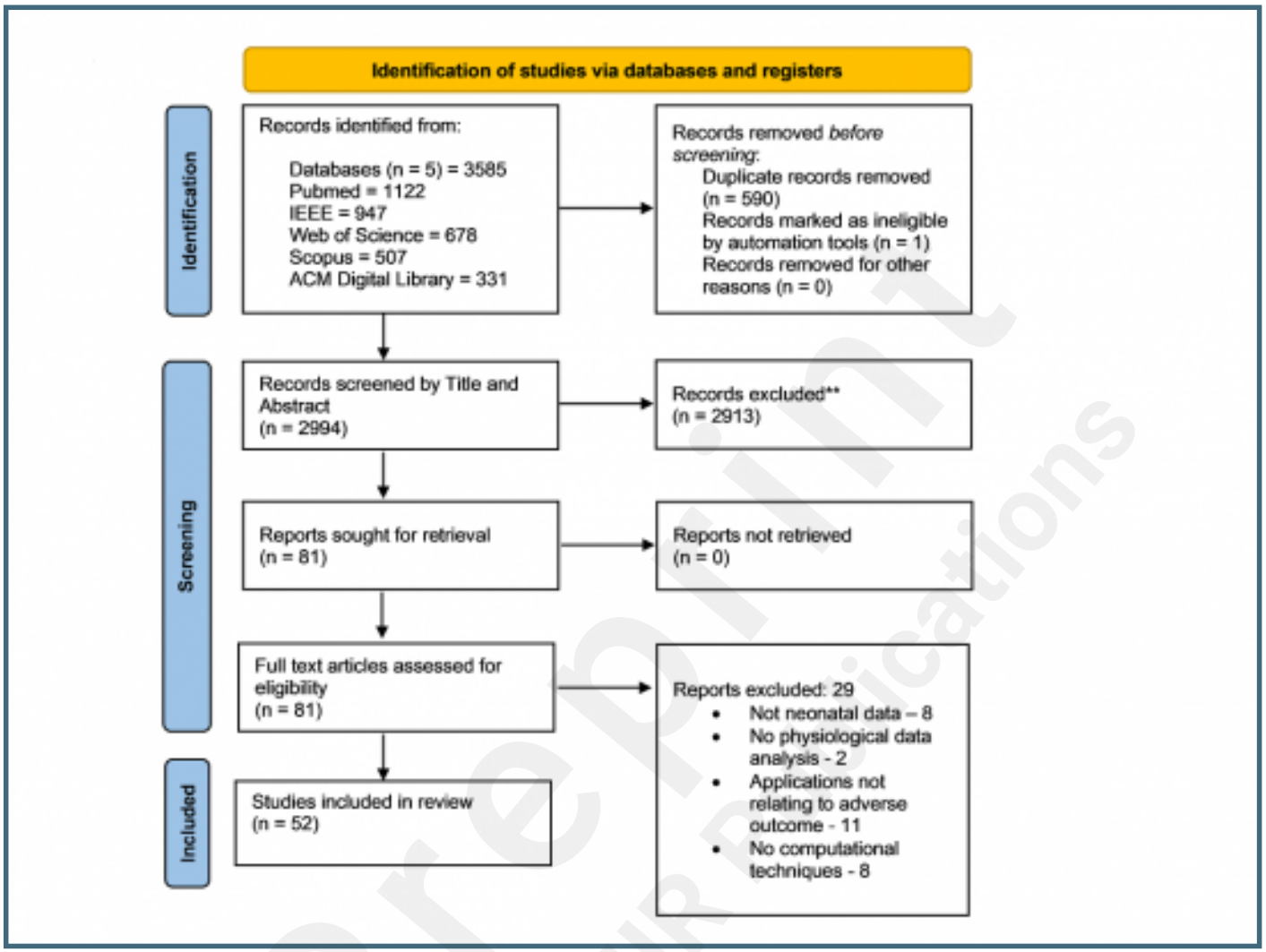
92. Mirnia K, Heidarzadeh M, Afjeh SA, et al. Signal Processing of Heart Rate for Predicting Sepsis in Premature Neonates. J Med Signals Sens. 2021 July 21;113:222–6. PMID: 34466402. https://doi.org/10.4103/jmss.JMSS_30_20.
93. Lee J, Cai J, Li F, Vesoulis ZA. Predicting mortality risk for preterm infants using random forest. Sci Rep. 2021;111:7308. https://doi.org/10.1038/s41598-021-86748-4.
94. Sullivan BA, Nagraj VP, Berry KL, et al. Clinical and vital sign changes associated with late-onset sepsis in very low birth weight infants at 3 NICUs. J Neonatal-Perinat Med. 2021;144:553–61. PMID: 33523025. https://doi.org/10.3233/NPM-200578.
95. Feng J, Lee J, Vesoulis ZA, Li F. Predicting mortality risk for preterm infants using deep learning models with time-series vital sign data. NPJ Digit Med. 2021;41:108. https://doi.org/10.1038/s41746-021-00479-4.
96. Zuzarte I, Sternad D, Paydarfar D. Predicting apneic events in preterm infants using cardio-respiratory and movement features. Comput Methods Programs Biomed. 2021;209. https://doi.org/10.1016/j.cmpb.2021.106321.
97. Niestroy JC, Moorman JR, Levinson MA, et al. Discovery of signatures of fatal neonatal illness in vital signs using highly comparative time-series analysis. Npj Digit Med. 2022 January 17;51:1–10. https://doi.org/10.1038/s41746-021-00551-z.
98. Kausch SL, Brandberg JG, Qiu J, et al. Cardiorespiratory signature of neonatal sepsis: development and validation of prediction models in 3 NICUs. Pediatr Res. 2023 January 2:1–9. https://doi.org/10.1038/s41390-022-02444-7.
99. Honoré A, Forsberg D, Adolphson K, Chatterjee S, Jost K, Herlenius E. Vital sign-based detection of sepsis in neonates using machine learning. Acta Paediatr. 2023;n/an/a. https://doi.org/10.1111/apa.16660.
100. S. Masoudi, N. Montazeri, M. B. Shamsollahi, et al. Early detection of apnea-bradycardia episodes in preterm infants based on coupled hidden Markov model. IEEE Int. Symp. Signal Process. Inf. Technol., 2013, p. 000243–8.
101. Altuve M, Carrault G, Beuchée A, Pladys P, Hernández AI. Online apnea–bradycardia detection based on hidden semi-Markov models. Med Biol Eng Comput. 2015;531:1–13. https://doi.org/10.1007/s11517-014-1207-1.
102. Honoré A, Liu D, Forsberg D, et al. Hidden Markov Models for Sepsis Detection in Preterm Infants. ICASSP 2020 - 2020 IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP, 2020 May, p. 1130–4.
103. Malik M, Bigger JT, Camm AJ, et al. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. Eur Heart J. 1996 March 1;173:354–81. https://doi.org/10.1093/oxfordjournals.eurheartj.a014868.
104. Paliwoda M, New K, Davies M, Bogossian F. Physiological vital sign ranges in newborns from 34 weeks gestation: A systematic review. Int J Nurs Stud. 2018 January 1;77:81–90. https://doi.org/10.1016/j.ijnurstu.2017.10.004.
105. Gardner SL, Carter BS, Enzman-Hines MI, Niermeyer S. Merenstein & Gardner's Handbook of Neonatal Intensive Care - E-Book: An Interprofessional Approach. Elsevier Health Sciences; 2020 February 5.
106. Sinha S, Miall L, Jardine L. Essential Neonatal Medicine. John Wiley & Sons; 2017 August 14.
107. Paliwoda M, New K, Bogossian F, Ballard E. Physiological vital sign reference ranges for well late preterm newborns calculated during a typical 2 h newborn period between 2 h and 7 days of life. Physiol Meas. 2021 September 1;429:095001.

https://doi.org/10.1088/1361-6579/ac155b.
108. Cohen G, de Chazal P. Automated detection of sleep apnea in infants: A multi-modal approach. Comput Biol Med. 2015;63:118–23. https://doi.org/10.1016/j.compbiomed.2015.05.007.
109. Shaffer F, Meehan ZM, Zerr CL. A Critical Review of Ultra-Short-Term Heart Rate Variability Norms Research. Front Neurosci. 2020;14.
110. Kwon O, Jeong J, Kim HB, et al. Electrocardiogram Sampling Frequency Range Acceptable for Heart Rate Variability Analysis. Healthc Inform Res. 2018 July;243:198–206. PMID: 30109153. https://doi.org/10.4258/hir.2018.24.3.198.
111. Gravel J, Opatrny L, Gouin S. High rate of missing vital signs data at triage in a paediatric emergency department. Paediatr Child Health. 2006 April 1;114:211–5. https://doi.org/10.1093/pch/11.4.211.
112. Software as a medical device (SAMD): clinical evaluation : guidance for industry and Food and Drug Administration staff - Digital Collections - National Library of Medicine n.d. https://collections.nlm.nih.gov/catalog/nlm:nlmuid-101720008-pdf (accessed February 27, 2024).
113. Software design – AI/ML – Part 1. Introd. SaMD IEC 62304 IEC 82304-1, Medical Device HQ; n.d.
114. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. BMC Med. 2015 January 6;131:1. https://doi.org/10.1186/s12916-014-0241-z.
115. Ramakrishnan R, Rao S, He J-R. Perinatal health predictors using artificial intelligence: A review. Womens Health. 2021 January 1;17:17455065211046132. https://doi.org/10.1177/17455065211046132.
116. Shen D, Wu G, Suk H-I. Deep Learning in Medical Image Analysis. Annu Rev Biomed Eng. 2017 June 21;19:221–48. PMID: 28301734. https://doi.org/10.1146/annurev-bioeng-071516-044442.

Supplementary Files

Figures

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of for the database search and study selection.



Multimedia Appendixes

Detailed search queries.
URL: http://asset.jmir.pub/assets/3710820a300e48868bc8bf5e0e4a63f2.docx

Bibliography files for all databases.
URL: http://asset.jmir.pub/assets/ea61e71a10f87186a49d40e7d3191516.zip

All included papers.
URL: http://asset.jmir.pub/assets/a1d2f5ff8667f33fbe79897c9463bfb3.zip

Metadata of all papers.
URL: http://asset.jmir.pub/assets/8d40a0201c00a8e87adf9c78741617ee.xlsx

Metadata of inlcuded papers.
URL: http://asset.jmir.pub/assets/37af921b5eb36388efdfe27865437ac3.xlsx