

Evaluation of screening parameters and machine learning models for the prediction of neonatal sepsis: A systematic review

Dennis Peace Ezeobi (✉ 2018mhit004@std.must.ac.ug)

Mbarara University of Science and Technology <https://orcid.org/0000-0003-2573-3414>

Dr. William Wasswa

Mbarara University of Science and Technology

Dr. Angella Musimenta

Mbarara University of Science and Technology

Dr. Stella Kyoyagala

Mbarara Regional Referral Hospital

Systematic Review

Keywords: Neonatal sepsis, screening parameters, algorithms, models

Posted Date: February 18th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1354764/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

About 2.9 million neonates die every year worldwide, and most of these deaths occur in low-resource settings. Neonatal sepsis occurs when there is a bacterial invasion in the bloodstream; the immune system begins a systemic inflammatory response syndrome (SIRS) damaging to the body and can quickly advance to severe sepsis, multi-organ failure, and finally, death. Sepsis in neonates can progress more rapidly than in adults; therefore, a timely diagnosis is critical. The standard gold test for diagnosing neonatal sepsis is blood culture, which takes at least 72 hours. Hence, identifying key predictor variables and models that work best can help reduce neonatal morbidity and mortality.

The matching articles were identified by searching the PubMed, IEEE, and Cochrane bibliography databases. For the inclusion of articles, the abstract and titles were first screened based on some predetermined criteria and then, the full-text articles were screened. Thirty-one studies met the full inclusion criteria. The duration of ROM was found to be more significant than other maternal risk factors. Heart rate and heart rate variability were found to be more significant than other neonatal clinical signs. C reactive protein and I/T ratio were found to be more significant than other laboratory tests. The main limitation is the variation in the performance measures used in the studies, which made it difficult to perform a quantitative assessment.

A combination of predictor variables has been shown to strengthen neonatal sepsis prediction, as shown by some of the reviewed studies. Predictive algorithms that combine multiple variables are urgently needed to improve models for early detection, prognosis, and treatment of neonatal sepsis.

Introduction

About 2.4 million neonates die every year worldwide, and most of these deaths occur in low resource settings [1] [2]. The third Sustainable Development Goal (SDG) for child health aims to end the mortality of newborns and children under five years of age, which is preventable by 2030. However, this may not be achieved if there is no significant reduction of neonatal deaths directly related to infection in developing countries [3]. Sepsis is a significant cause of neonatal mortality and morbidity around the world [4] [5] [6] and most of the morbidity and mortality from sepsis is preventable.

Neonatal sepsis is classified as early-onset (< 48–72h) and late-onset sepsis (> 48–72h), and this depends on the age at onset [7] [8]. About 30–50% of neonatal sepsis survivors obtain significant long-term impairments, including prolonged hospitalization, chronic lung disease, and neurodevelopmental disabilities [9] [10] [11]. Sepsis remains one of the most expensive causes of hospitalization, as recent data highlight its costs and burdens [12] [13] [14] [15]. Physicians caring for infected neonates are faced with multiple challenges in diagnostic and treatment decisions. Despite the increased understanding of its pathophysiology and efforts to improve clinical decision support in intensive care, there have been just fair improvements in neonatal sepsis outcomes [16]. Neonatal sepsis occurs when there is a bacterial invasion in the bloodstream; the immune system begins a systemic inflammatory response syndrome

(SIRS), which is damaging to the body and can quickly advance to severe sepsis, multi-organ failure, and finally, death [17] [18]. However, early recognition and prompt treatment have been predicted to improve the clinical management of sepsis and serve as the key to reducing morbidity and mortality [19] [20] [21] [22] [23].

Delays in recognition and treatment of sepsis is still a challenge despite the explored importance of early intervention [6] [16] [24] [25] [26] [27] [28]. Neonatal clinical presentation is non-specific and overlaps with other newborn disease processes. The laboratory tests have limited diagnostic accuracy, which makes rapid diagnosis for neonatal sepsis difficult. The standard gold test for neonatal sepsis diagnosis, blood culture, faces the challenge of insufficient blood volume for blood culture and low amount of invading microorganisms in the blood, which usually generates false-negative results [29] [30]. Infants suspected of having sepsis are subjected to prolonged antibiotic therapy despite negative cultures. In other to tackle the challenges associated with sepsis recognition and care management studies are making use of machine learning and statistical modeling approaches [31] [32] [33] [34].

Compared to other significant conditions, neonatal sepsis receives less substantial international investment as a public health priority despite the heavy burden of newborn deaths related to neonatal sepsis [3]. Knowledge of neonatal sepsis's predictor variables, early identification, and early interventions can reduce neonatal mortality and morbidity rates. This study aims to review the existing screening parameters and models based on their diagnostic performance, strength, and weaknesses to better understand the algorithm development process.

Materials And Methods

Selection of screening parameters for analysis

A preliminary examination of the available literature was carried out, after which a list of parameters was consolidated for further review. These parameters were selected based on their publication and their potential for diagnosing and prognosis of neonatal sepsis. The parameters include;

- Maternal risk factors (which include; intrapartum fever, chorioamnionitis, postnatal distress, duration of ROM, GBS colonization, and intrapartum antibiotics).
- Neonatal clinical signs (which include; gestational age, birth weight, heart rate, and feeding difficulty).
- Laboratory tests (which include; absolute neutrophil count, C reactive protein, I/T ratio, micro-ESR, platelet count, and total leukocyte count).

Search Strategies

In order to carry out a landscape analysis to identify studies with the diagnostic performance of the previously mentioned parameters, PubMed, IEEE, and Cochrane's bibliography database were searched. The search strategies for the databases were carefully made to give maximum output. A combination of text words was used to develop the search strategy, which includes; "neonatal sepsis" AND "prediction" AND "machine learning", "neonatal sepsis" AND "prediction" AND "EHR", "neonatal sepsis" AND "prediction" AND "model", "neonatal sepsis" AND "prediction" AND "algorithm", "neonatal sepsis" AND "diagnostic algorithm" AND "machine learning", "neonatal sepsis" AND "screening parameters" AND "models", "neonatal sepsis" AND "screen" AND "models". The search strategy was restricted to the subject (humans) and the time period (January 2000 to April 2020). A total of 463 PubMed, 305 citations from IEEE, and 86 Cochrane citations from Cochrane were retrieved. These references were imported as separate files into an excel sheet except for the Cochrane database's references; it was imported only as CSV file. The duplicates were removed, the titles and abstracts of the retrieved citations were screened to find the articles relevant to the study. Additional relevant studies were retrieved by scrutinizing the bibliography of searched studies.

Inclusion Criteria

For the inclusion of articles, the abstract and titles were screened based on the following predetermined criteria:

- The subject population are neonates.
- Subjects have culture-proved sepsis or suspected sepsis based on a clinical algorithm.
- The article evaluated any of the consolidated screening parameters and algorithms/models for neonatal sepsis diagnosis or prognosis.

The exhaustive search based on the titles and abstracts returned a broad spectrum of infection-related studies from which only cases of neonatal sepsis were considered. Finally, full-text articles with the following criteria were included for analysis:

- The subject population are neonates.
- The study provided a clear definition of neonatal sepsis.
- The study provides neonatal sepsis onset definition (i.e., time of onset).
- The study clearly described the predictor variables used.
- The study clearly described the machine learning models used or evaluated in any of the consolidated screening parameters.
- The study must have provided diagnostic performance results (i.e., AUROC results).

Exclusion Criteria

It was a great challenge trying to select the relevant articles for this review from the large number of papers retrieved (n=854) based on the selection criteria. To make a comprehensive list of appropriate

papers, articles that did not deal with neonates, duplicates, reviews, meta-analyses, abstracts, editorials, and commentaries were excluded.

Data Extraction

The available full papers were downloaded from PubMed, IEEE, and Cochrane sources. The data was extracted and compiled in an Excel spreadsheet. The following information was extracted from all the studies:

- a. Publication characteristics (author's name, year of publication).
- b. Study design (retrospective, prospective data collection and analysis).
- c. Selection of cohorts (sex, age, number of patients with sepsis, prevalence of sepsis).
- d. Neonatal sepsis definition.
- e. Neonatal sepsis onset definition.
- f. Specifics on analyzed data (the type of variables, number of screening parameters).
- g. Model selection (ML algorithm, platforms, software, packages, and hyperparameters).
- h. Statistics for the performance model (methods for evaluating the model, statistical significance, handling of missing data).
- i. Methods to avoid overfitting and also any additional external validation approaches.

For a point of reference, the leading hospital in Mbarara, Uganda, was contacted to learn what tests/algorithms are currently being used in their clinical settings.

Quality Assessment of the Included Studies

Table 1: Quality assessment of the included studies.

Categories	Items	Description	Reported
Unmet needs	Limits in current machine learning or non-machine learning Applications	Low diagnostic accuracy, low human-level prediction accuracy, or prolonged diagnostic procedure.	Yes/No
Reproducibility	Prevalence of Neonatal sepsis	The proportion of neonates who suffered sepsis out of the entire study cohort.	Yes/No
	Data availability	Is the data used in the study publicly available?	Yes/No
	Feature engineering methods	How features were generated before model training	Yes/No
	Code for data wrangling and analysis	Code describing the details of the cleaning, preprocessing, and analysis of the data.	Yes/No
	Code of label	Code describing neonatal sepsis label generation	Yes/No
	Platforms/packages	Both platforms and packages should be reported	Yes/No
	Hyperparameters	All hyperparameters which are needed for study replication	Yes/No
Robustness	Sample size > 50	Neonatal sepsis case sample size >50 is required for the interpretation, power, and validity of machine learning methods.	Yes/No
	Valid methods for over-fitting	Valid methods for unbiased performance assessment (or methods "against" overfitting)	Yes/No
	Stability of results	Calculated variation in the validation statistics	Yes/No
Generalizability	External data validation	Validation in settings different from the research framework	Yes/No
Clinical significance	Predictor's explanation	Explanation (biological or quantification) of the importance of each predictor	Yes/No
	Suggested clinical use	Clinical usability and requirements (e.g. what are still necessary for making deployment possible)	Yes/No

The quality of the selected ML studies was assessed based on 14 criteria relevant to the objectives of the review, which was adopted from [35]. The assessment consists of five categories described in table 1 above. A quality assessment table was provided by listing "yes" or "no" for each category's items using the provided criteria.

Results

Out of 854 studies, 31 studies met the inclusion criteria. The literature search results with reasons for exclusions at each stage are presented in figure 1 above.

Study Characteristics

Of the 31 included studies, 16 employed solely prospective analyses, 13 employed solely retrospective analyses, while 2 studies used both retrospective and prospective analyses [37, 38]. The most frequent data sources used in the studies were the University of Virginia Hospital (n = 8; 26%), followed by MIMIC-III (n = 3; 10%). In terms of neonatal sepsis definition, the majority of the studies employed Blood culture (n = 26; 84%) or Observational condition (use of clinical signs) (n = 16; 52%) or combination of Blood culture and Observational condition (n = 12; 39%). The studies modified observational and Laboratory definitions based on available data and the predetermined neonatal sepsis onset time; this is mainly due to the absence of a consensus definition of neonatal sepsis. The prevalence of neonates with sepsis ranged between 0.27% and 87.0%. Five studies did not report the prevalence [39, 40, 41, 42, 43]. Regarding the category of neonatal sepsis of interest, the category with a high focus is late-onset (n = 18; 58%) and early-onset (n = 4; 13%). While 9 studies [39, 44, 45, 46, 47, 48, 49, 50, 51] did not report the category of focus. In demographics, 6 studies reported the median or mean age of the neonates, 11 reported the prevalence of male neonates, 2 reported the prevalence of female neonates, and only 3 reported the investigated cohorts' ethnicity (see supplementary table 1).

Overview of Machine Learning Algorithms and Variables

A wide range of ML algorithms has been employed to build models for the early detection of neonatal sepsis, with some models being specific to the study population. Regression was the most used model of which various types (n = 25; 81%) were used. This includes Logistic Regression or Linear Regression [52]. Furthermore, boosted tree models were the second most used model (n = 6; 19%), including gradient boosting [42], or random forest [43]. And lastly SVM [42] (n = 5; 16%). Most of the studies (n = 24; 77%) arbitrarily chose one- or two-ML models without arguing the reasons. Seven studies (23%) [53, 54, 55, 42, 43, 50, 56] compared several models and identified the one with the best performance.

As for the analyzed variables, the most common variable used was neonatal clinical signs (n = 28; 90%), followed by laboratory tests (n = 10; 32%), and maternal risk characteristics (n = 5; 19%). Sixteen studies (52%) were found to use one variable, while the remaining fifteen studies (48%) were found to combine these variables, which includes neonatal clinical signs and laboratory tests (n = 10; 32%), maternal risk characteristics, and neonatal clinical signs (n = 5; 16%). None of the studies were found to explore the combination of maternal risk characteristics, neonatal clinical signs, and laboratory tests. The number of screening parameters included in the respective models ranged between 2 [57] and 22 [50]. Concerning the features for detecting neonatal sepsis, the reviewed studies show that duration of ROM was found to be more significant than other maternal risk factors [46, 58, 50, 59]. Heart rate and heart rate variability were found to be more significant than other neonatal clinical signs [40, 43]. C reactive protein and I/T ratio were found to be more significant than other laboratory tests [45, 57].

Table 2: Pseudo code for the HRV monitoring algorithm

HRV Monitoring Algorithm for Neonatal sepsis [43]

Step 1: Create a set H of 17 heart rate variability, $H = (v_0 \dots v_d), 1 \leq d \leq 17$
Step 2: Initialize elements of set H; $T = (c_0 \dots c_i), 1 \leq i \leq 17$
Step 3: Check result of the time, frequency and non-linear analysis in elements of set T
 IF i is defined as "Absolute" THEN
 RETURN True
 ELSE
 RETURN False
 END IF
Step 4: IF number of Absolute is defined as "High" THEN
 RETURN "Neonatal Sepsis"
 ELSE
 RETURN "Normal"
 END IF

Table 3: Phase I: Pseudo code for the observational condition

Medical Decision Support Algorithm for Neonatal sepsis [54]

The algorithm consists of three phases: observational condition, laboratory condition, and neonatal sepsis.

Step 1: Create a tuple Z of 4 neonatal clinical signs, $Z = (w_0 \dots w_x), 1 \leq x \leq 4$
Step 2: Initialize elements of tuple Z; $U = (d_0 \dots d_j), 1 \leq j \leq 4$
Step 3: FOR each j in U DO
 IF $j = =$ condition THEN
 RETURN True
 ELSE
 RETURN False
 END IF
 END FOR
Step 4: IF True \geq 1
 RETURN "Observational Condition"
 ELSE
 RETURN "No Observational Condition"
 END IF

Table 4: Phase II: Pseudo code for the laboratory condition

Step 1: Create a tuple Q of 5 laboratory tests, $Q = (p_0 \dots p_c), 1 \leq c \leq 5$
Step 2: Initialize elements of tuple Q; $X = (f_0 \dots f_k), 1 \leq k \leq 5$
Step 3: FOR each k in X DO
 IF k == condition THEN
 RETURN True
 ELSE
 RETURN False
 END IF
END FOR
Step 4: IF True ≥ 1
 RETURN "Laboratory Condition"
ELSE
 RETURN "No Laboratory Condition"
END IF

Table 5: Phase III: Pseudo code for the neonatal sepsis

Step 1: Create a set R of 3 neonatal sepsis variables, $R = (d_0 \dots d_a), 1 \leq a \leq 3$
Step 2: Initialize elements of set R; $L = (g_0 \dots g_y), 1 \leq y \leq 3$
Step 3: FOR each y in L DO
 IF y == "Yes" THEN
 RETURN True
 ELSE
 RETURN False
 END IF
END FOR
Step 4: IF True == 3
 RETURN "Septic"
ELSE
 RETURN "Not Septic"
END IF

Some of the existing neonatal sepsis prediction algorithms using neonatal clinical signs and maternal risk factors are shown below in tables 2, 3, 4, and 5.

Model Validation

Approximately 58% of the studies did not report what valid methods were used to prevent overfitting, while 29% employed cross-validation technique (e.g., 4-fold, 5-fold, 10-fold, or leave-one-out cross-validation), and 19% employed bootstrap to avoid overfitting. Concerning the models' limitations, 13% of the studies recommend that the models require additional variables to optimize their performance. Additional external validation of the models was only performed in seven studies [60, 44, 40, 58, 61, 48, 38]. Particularly, Fairchild & O'Shea (2010) used datasets from University of Virginia NICU and Wake Forest University NICU to train, test, and validate the use of neonatal heart rate characteristics (HRC) to detect late-onset (LOS) neonatal sepsis. In another study, Fairchild et al. (2017) explored the use of vital

signs to build models that predict neonatal sepsis using datasets from the University of Virginia, UVA, and Columbia University. Gur et al. (2015) trained, tested, and validated the RALIS algorithm's ability to detect LOS before clinical suspicion with datasets from neonatal intensive care units (NICU) of three hospitals in Israel. Aiming to develop and validate a nomogram for assessing the individual prior probability of LOS based on maternal risk factors and neonatal clinical signs in preterm infants, Huang et al. (2020) created a validation cohort using data from three neonatal critical care centers in Guangdong province of China. Lastly, the study by Popowski et al. (2011) investigated the predictive value of maternal risk factors for early-onset (EOS) neonatal sepsis using a dataset from two French tertiary university referral centers.

Quality Assessment of Included Studies

[Table 6 is in the supplementary files section.]

Table 6 above shows the results of the quality assessment of the studies. The 31 studies' quality ranged from poor (meeting $\leq 40\%$ of the criteria) to very good (meeting $\geq 90\%$ of the criteria). None of the studies fulfilled all 14 criteria as none of the studies met $\geq 90\%$ of the criteria. Few studies made the data used in their study available ($n = 3$; 10%). Only ten studies (32%) explained how features were generated before model training. Only two studies (6%) provided the code used for data cleaning and analysis. Only one study (3%) provided code to reproduce the exact sepsis labels [58]. Few studies reported the hyperparameters needed for study replication ($n = 5$; 16%). Finally, only seven studies (23%) validated their study result on an external data set. With the exception of two studies [64, 53], all other studies had sample sizes larger than 50, which is a requirement for the interpretation, power, and validity of machine learning methods.

[Tables 7-10 are in the supplementary files section.]

Table 11: Strength and weaknesses of the existing screening parameters

Screening Parameters	Strength	Weakness
Duration of ROM	Has strong and nearly linear association with neonatal sepsis.	Association is stronger with EOS than LOS.
Maternal age	Neonatal sepsis is common among infants of older mothers and maternal age < 20 can be associated with EOS risk factors.	Not validated or considered as a determining risk factor for neonatal sepsis.
Parity	Has strong association with neonatal sepsis.	Association with neonatal sepsis is controversial.
Antibiotic treatment	Reduces the risk of infection to a mother and neonate.	Increases other health risks to newborn infants.
Maternal CRP	It is associated with neonatal sepsis and a significant risk factor for neonatal sepsis.	CRP values are also affected by other factors.
GBS status	Strongly associated with neonatal sepsis.	Even though GBS remains the most frequent pathogen for EOS, there has been a shift in this as Escherichia coli (E. coli) becomes the most important pathogen causing EOS in preterm and very low birth weight infants.
Intrapartum fever	It's generally considered a major risk factor for EOS.	The risk of neonatal sepsis in newborns delivered by mothers with intrapartum fever is low.
Heart rate variability	It is a significant risk factor for neonatal sepsis, and neonates have reduced heart rate variability (HRV) before clinical signs of sepsis.	Its main drawback for early diagnosis of neonatal sepsis is the high false-positive rate.
Birth weight	It is one of the determining factors for neonatal sepsis as newborns with less than 2.5 kg are 1.42 times more likely to develop neonatal sepsis than newborns born with 2.5 kg and above.	Infants with low birth weights are at increased risk for other forms of infection and infection-related mortality.
Respiratory rate	Its variability can be an indicator of sepsis.	The variability in respiratory rate is also associated with other respiratory problems.
Heart rate	It is one of the most important clinical indicators to evaluate sepsis.	An elevated score is not specific for sepsis and may occur in other conditions associated with nonspecific inflammation.
SpO2	Performs well for preclinical detection of sepsis.	High altitudes and other factors may affect what is considered normal for a given neonate.
Poor feeding	It appears to be crucial in a diagnosis of sepsis.	It is a nonspecific symptom seen in newborn
Temperature	Its variability can be an indicator of sepsis.	Newborns cannot regulate their body temperature well, causing instability.
Apnea	It can be a clinical sign of neonatal sepsis	It is common in infant breathing and can be caused by other factors.
Lethargy	It can be a sign of infection	It can be a sign of other conditions.
Duration of umbilical venous catheters	A long duration of use is associated with bloodstream infection in newborn	It is not significant in the diagnosis of sepsis.
Use of antibiotics on newborns	Are administered to target most types of bacteria that cause an infection	Because infants have a higher risk of complications, pediatricians often prescribe antibiotics even if they aren't positive that it's a bacterial infection.
Blood	It can be associated with newborn	It can be affected by other conditions.

pressure	infection.	
Gestational age	Preterm babies are more likely to develop neonatal sepsis than term babies.	It is a risk factor for other conditions.
Gender	It influences both the incidence and the outcomes of sepsis.	It is not a strong indicator of sepsis.
Platelets	It's beneficial to predict mortality or to diagnose the sepsis	It is not very sensitive for the diagnosis of neonatal sepsis and is not very helpful in monitoring the response to therapy.
WBC	It is highly predictive of infection.	Multiple variables can affect the various components of WBC.
CRP	It increases significantly in cases with infection.	It may not be elevated in the early stages of infection due to the time taken for its synthesis in the liver and, eventually, appear in the blood.
Leucocyte count	It can aid in clinical decision-making in cases where a low-to-moderate clinical suspicion for sepsis is present.	It has low sensitivity in diagnosing neonatal sepsis.
I/T ratio	It is highly predictive of infection.	It is not particularly useful as an independent test in identifying the majority of septic infants.
pH	It can be significantly lower in newborns with sepsis	It can be caused by other conditions.
Glucose	Its level can be significantly affected by neonatal sepsis.	Low blood sugar can happen for many reasons
HCO ₃	It can be significantly lower in newborns with sepsis	It can be caused by other conditions.

[Table 12 is in the supplementary files section.]

Discussion

The review summarized studies on neonatal sepsis with ML algorithms to facilitate early prediction. Looking at ML methods, which includes cohort selections, predictor variables, outcomes, the building of models, and validation methods. A wide range of ML algorithms was chosen for analysis in the studies to leverage neonates' digital health data to predict sepsis. Based on the findings from the reviewed studies, this section outlines three major challenges that studies on neonatal sepsis prediction leveraging machine learning are currently facing: (i) asynchronicity, (ii) comparability, and (iii) reproducibility.

Asynchronicity

Studies focused on predicting neonatal sepsis with ML have shown to have the advantage of increasing the prediction power and have promising results [43, 50]. But so far, the reports on which of the open challenges are the most pressing challenges that need to be addressed are diverging, which poses difficulty in achieving the goal of early neonatal sepsis detection. On one hand, the blood culture test, which is the standard gold test, has been stated as the most reliable test for confirming neonatal sepsis [62]. While on the other hand, recent findings have cast doubt on the validity and meaningfulness

of the blood culture test. As it has been stated to be unreliable due to the longer time (48-72 hours), it takes to obtain the result and the insufficient amount of blood obtained from neonates, which produces false-negative results [54, 50]. Also, it was stated that neonatal clinical signs (e.g., the use of heart rate variability) alone are sufficient in detecting neonatal sepsis [40, 61]. However, recent studies are posing doubts to this as they state that a combination of predictor variables yield better results in the detection of neonatal sepsis [48, 57, 43, 50]. The developed ML models need to be explored in clinical trials to ascertain their clinical settings usage as most of the models are developed retrospectively, facing multiple obstacles.

Comparability

In terms of comparability of the reviewed studies, several challenges were identified that are yet to be overcome; (i) neonatal sepsis definition, (ii) implementation of a given neonatal sepsis definition, and (iii) performance measures of the models. Each of these challenges is discussed below.

Defining and Implementing Neonatal Sepsis

The choice of neonatal sepsis definition is an obstacle that affects the comparison of studies in terms of septic neonates' prevalence. A various set of neonatal sepsis definitions (and modifications) were used in the reviewed studies. Having a large set of septic neonates is anticipated to be useful in training ML models (most especially the deep neural networks). However, having a high number of septic neonates could make it difficult to differentiate the septic neonates from the non-septic neonates. Neonatal sepsis is inherently hard to define as, over the years, there has not been a consensus definition for it. The previous study shows that the use of different sepsis definitions on the same dataset gives a largely dissimilar cohort [68]. This study found that blood culture is less inclusive, leading to a small cohort showing severe symptoms, which has been reported in several studies [54, 43, 50]. It was also seen that even the use of the same definition on the same dataset gives dissimilar cohorts. This can be confirmed from studies carried out at the University of Virginia and studies that used the MIMIC-III dataset (see Table 9 above). The underlying problem cannot be easily discovered, as the code for assigning the labels are not available in 30 studies out of 31 (97%) studies. The diversity of neonatal sepsis prevalence is another factor that is increasing the problem of comparability. Some studies balance their datasets to improve the training of the ML models, but this training setup can partly affect the study [56]. While other studies keep the observed case counts to see how their approach will work in clinical settings. From this study findings, it has been identified that the neonatal sepsis definition used and the data pre-processing steps affect the prediction of sepsis and also the prevalence [68]. The maximum prevalence reported is 87.0% [64].

Performance Measures of the Models

The choice of performance measures is the last obstacle to be discussed that is obstructing comparability. This obstacle is largely affected by the prevalence of neonatal sepsis in the study. Accuracy is a simple performance metric directly influenced by class prevalence; comparing two studies with different prevalence values is problematic. Some studies report the area under the receiver operating characteristic curve (AUROC, also known as AUC) to improve the performance metrics report. However, AUROC also depends on class prevalence and can be less informative on highly imbalanced classes [69]. The area under the precision-recall curve (AUPRC, also known as average precision) is preferable in such a situation. Both AUPRC and AUROC are affected by prevalence. However, AUPRC allows comparison with a random baseline that just "guesses" the neonate label, and it's useful when considering the positive class. While AUROC can be high even for classifiers that could not classify the minority class of septic neonates. The effect of the choice of performance metrics is greatly seen with highly imbalanced classes. Recent research recommends reporting the AUPRC of models, particularly in clinical studies [70], which is a good recommendation.

Comparing Studies of Low Comparability

Based on this review study's findings, comparing the reviewed studies quantitatively is currently a challenging task to accomplish, which was also seen in a study by (Moor, et al., 2020). The studies were assessed qualitatively to identify underlying biases that could lead to unduly optimistic results. This was done as the best-performing methods could not be ascertained by just evaluating the performance measures' numeric values. A meta-analysis will be preferable to sum up, an overall trend in the performance of the models.

Reproducibility

Reproducibility, which is the ability to obtain consistent results using the same data and code as the original experiment, is the means for scientific accountability. There have been failures of this accountability in several disciplines, including ML [35]. The use of sensitive data makes it difficult to make available the dataset used in studies, which is one of the challenges digital medicine poses to reproducibility. Another challenge is the failure to provide detailed preprocessing methods used in ML papers. Based on the quality assessment carried out, important areas were outlined that need to be improved. As it was seen, only two studies [50, 66] made available their analysis code. Only one study [58] made available their code for generating a "label." Both cases amount to less than 10% of the eligible studies. In addition, only three studies [57, 50, 66] made available the dataset used for their study. Only eight studies were found to share the hyperparameters used in their studies. However, a positive finding of this analysis is that a considerable number of studies (n = 10) shared the preprocessing methods used in their studies, which is useful information in the reproducibility of computational experiments.

This review focused on publications that studied the prediction of neonatal sepsis implementing ML algorithms. The majority of the reviewed studies investigating neonatal sepsis prediction defined neonatal sepsis as having positive blood culture or observational condition (i.e., the use of neonatal clinical signs). None of the 31 included studies reflects an African cohort, which shows a significant dataset bias in the publications and insufficient research in Africa (see Supplemental table 1 for an overview of demographical information). The review found a lot of room for improvement, which will benefit the comparability of different models, most importantly, when ML models are going to be evaluated prospectively.

Limitations

This review was carried out with some shortcomings. The reviewed studies had certain inherent limitations, as previously mentioned. The diagnostic performance evaluation report of the models was suboptimal. The studies were assessed qualitatively due to the variation in the performance measures used. A meta-analysis will be preferable to evaluate the performance of the models. Some studies may have been omitted from the review as English language restrictions were applied.

Conclusion

Combination of these variables have been predicted to strengthen the prediction of neonatal sepsis which was shown in some of the studies above. This study seeks to inform researchers on what predictor variables are required to develop algorithms/models with better diagnostic performance which will improve the detection of neonatal sepsis. The parameters and machine learning models used in the reviewed studies were largely different, so diagnostic performance was different. It should be considered that neonatal sepsis is consistent with other symptoms as well as underlying conditions. What is important here is the weight assigned to a variable. Suggestions for risk stratification based on maternal risk factors (such as; intrapartum fever, chorioamnionitis, duration of ROM, GBS colonization and intrapartum antibiotics), neonatal clinical signs (such as; gestational age, birth weight, heart rate, postnatal distress and feeding difficulty), and laboratory tests (such as; absolute neutrophil count, C reactive protein, I/T ratio, micro-ESR, platelet count and total leukocyte count) could be considered for future studies.

List Of Abbreviations

ANC	Absolute Neutrophil Count
AR-HMM	Autoregressive hidden Markov model
AUROC	Area Under the Receiver Operating Characteristics
CNNs	Convolutional neural networks
CRP	C-reactive protein
CSF	Cerebrospinal Fluid
ECG	Electrocardiogram
EMR	Electronic Medical Record
EOS	Early onset sepsis
GA	Gestational age
GBS	Group B Streptococcus
HELLP syndrome	Hemolysis elevated liver enzymes low platelet count
HRC	Heart rate characteristics
HR	Heart rate
HRV	Heart Rate Variability
I/T ratio	Immature to Total Neutrophil Ratio
LNS	Late-neonatal sepsis
LOCF	Last observation carried forward
LOS	Late-onset sepsis
M-ESR	Micro Erythrocyte Sedimentation Rate
MIMIC	Medical Information Mart for Intensive care III
ML	Machine Learning
NICUs	Neonatal Intensive Care Units
NPV	Negative Predictive Value
PPROM	Preterm Premature Rupture of the Membranes
PPROMEXIL	Preterm pre-labor rupture of the membrane expectant management or induction of labor study
Pro	Prospective
PR	Pulse rate

PPV	Positive Predictive Value
Retro	Retrospective
ROC curve	Receiver-operating characteristic curve
ROM	Rupture of Membranes
RR	Respiratory rate
SDG	Sustainable Development Goals
SIRS	Systemic Inflammatory Response Syndrome
SO	Oxygen saturation
SpO2	Blood oxygen level
SSA	Sub-Saharan Africa
TC	Core temperature
TLC	Total Leukocyte Count
TP	Peripheral temperature
WBC	White blood cell

Declarations

The work presented in this Manuscript is the result of our original research work. Where we have used the works of other persons, due acknowledgements are clearly stated. This work has not been submitted for publication in any journal before.

Ethical approval and consent to participate

Not applicable; as the study reviewed only published data.

Consent for publication

Not applicable

Availability of data and materials

Not applicable.

Competing Interests

The authors declare that they have no competing interests.

Funding

No external funding was obtained for this study.

Author's contributions

EDP, WW and AM carried out the preliminary literature search following the PRISMA guidelines, tabulated and analyzed the collected data and developed the first draft of the manuscript. KS contributed the neonatology expertise and edited the manuscript accordingly. All the authors reviewed and approved the final manuscript.

Acknowledgement

Not applicable.

Authors information

EDP

Department: Information Technology

Course: Masters in Health Information Technology

Authors' qualifications: BS.c, MS.c

Institution: Mbarara University of Science and Technology (MUST)

Position: Post-Graduate Student

WW

Authors' qualifications: BS.c, MS.c, PhD

Department: Biomedical Sciences and Engineering

Institution: Mbarara University of Science and Technology (MUST)

Position: Head of Department (HOD)

AM

Authors' qualifications: BS.c, MS.c, PhD

Department: Information Technology

Institution: Mbarara University of Science and Technology (MUST)

Position: Deputy Dean of the Faculty

Location: Mbarara, Uganda

KS

Authors' qualifications: MBChB, MMed

Department: Pediatrics and Child Health MRRH

Hospital: Mbarara Regional Referral Hospital (MRRH)

Position: Head of Department (HOD), Neonatology division

Supplementary Material

The supplementary file contains the filled PRISMA statement checklist for the study.

Table 1: The table contain an overview of trial designs, demographical information and machine learning details of the studies.

References

- [1] WHO, "Newborn death and illness," Partnership for Maternal, Newborn & Child Health, Geneva, Switzerland, 2011.
-
- [2] UNICEF, "Levels and Trends in Child Mortality Report 2017.," United Nations Children's Fund, New York, 2017.
-
- [3] S. W. B. Ranjeva and S. Schiff, "Economic burden of neonatal sepsis in sub-Saharan Africa.," *BMJ global health*, vol. 3, no. 1, p. e000347, 2018.
-
- [4] L. Liu, S. Oza, D. Hogan, J. Perin, I. Rudan, J. Lawn, S. Cousens, C. Mathers and R. Black, "Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis.," *Lancet*, vol. 385, pp. 430-440, 2015.
-
- [5] A. Sarkar, G. Dhar, M. Sarkar, T. Ghosh and S. Ghosh, "Early diagnosis of neonatal sepsis in primary health care unit.," *Bangladesh J. Med. Sci.*, vol. 14, no. 2, pp. 5-8, 2015.
-
- [6] M. V. M. Stephannie and J. M. M. Patrick, "Neonatal Sepsis.," *Clin. Pediatr. Emerg. Med.*, vol. 17, pp. 129-133, 2016.
-
- [7] Z.-V. Alonso and J. O. Theresa, "Challenges in the diagnosis and management of neonatal sepsis," *Journal of Tropical Pediatrics*, vol. 61, p. 1–13, 2015.
-
- [8] A. Shane, P. Sánchez and B. Stoll, "Neonatal sepsis.," *Lancet*, 2017.
-
- [9] J. S. M. Barbara, I. H. M. Nellie, A.-C. M. Ira, A. F. M. Avroy, R. H. M. Susan, V. M. Betty and D. H. M. Rosemary, "Neurodevelopmental and Growth Impairment Among Extremely Low-Birth-Weight Infants With Neonatal Infection," *JAMA*, vol. 292, no. 19, pp. 2357-2365, 2004.
-
- [10] E. S. M. Bonnie and R. M. Betty, "Neurodevelopmental Outcome of the Premature Infant," *Pediatric Clinics of North America*, vol. 56, no. 3, p. 631–646., 2009.
-
- [11] J. S. M. Barbara, I. H. M. Nellie, F. B. M. Edward, S. M. Seetha, R. L. M. Abbot, C. W. M. M. Michele and B. R. Nancy S. Newman, "Neonatal Outcomes of Extremely Preterm Infants From the NICHD Neonatal Research Network," *Pediatrics*, vol. 126, no. 3, p. 443–456, 2010.
-
- [12] C. Jonathan, V. Jean-Louis, K. J. A. Neill, R. M. Flavia, C. A. Derek, C. Thierry, J. Katia, G. Stefano, D. Julie, O. Steven, T. Kevin, v. d. P. Tom and P. Eric, "Sepsis: a roadmap for future research," *The Lancet Infectious Diseases Commission*, vol. 15, no. 5, p. 581–614., 2015.
-
- [13] G. M. M. Elsie, S. M. P. Lisa, F. R. P. Lynn and M. M. M. Paul, "A Nurse-Driven Screening Tool for the Early Identification of Sepsis in an Intermediate Care Unit Setting," *J Hosp Med*, vol. 10, no. 2, p. 97–103, 2015.
-
- [14] J. P. P. M. Carly, P. Mark A. Reynolds, M. Meenal Sinha, P. Matthew Gitlin and M. Elliott Crouser, "Epidemiology and Costs of Sepsis in the United States—An Analysis Based on Timing of Diagnosis and Severity Level," *Critical Care Medicine*, vol. 46, no. 12, 2018.
-
- [15] H. Jihane, B. Natalie, S. Jola and J. Douglas, "The "Centrality of Sepsis": A Review on Incidence, Mortality, and Cost of Care," *Healthcare*, vol. 6, no. 3, 2018.
-
- [16] L. W. James, "Defining Neonatal Sepsis," *Curr Opin Pediatr.*, vol. 28, no. 2, p. 135–140, 2016.
-
- [17] R. Konrad, B. Michael, C. R. Niels and S. H. Christiane, "New Approaches to Sepsis: Molecular Diagnostics and Biomarkers," *Clinical Microbiology Reviews*, vol. 25, no. 4, p. 609 – 634, 2012.
-
- [18] B. M. P. M. Fran, R. A. M. M. Elizabeth, K. A. M. R. Mary, H. B. Katie, S. P. Aileen, L. M. Jane, M. P. Julie C. Fitzgerald, L. W. M. M. Scott and J. Z. M. M. Joseph, "Improving Recognition of Pediatric Severe Sepsis in the Emergency Department: Contributions of a Vital Sign Based

Electronic Alert and Bedside Clinician Identification," *Ann Emerg Med.* , vol. 70, no. 6, p. 759–768., 2017.

-
- [19] L. J. M. M. Stephen, M. A. M. M. Carol, K. M. Lisa, G. N. Elizabeth, B.-G. D. R. F.-C. C. N.-B. Charyl, D. R. C. Maureen, M. M. F. F. Faisal, A. S. P. Beverly and M. M. Nelda P. Wray, "Reductions in Sepsis Mortality and Costs After Design and Implementation of a Nurse-Based Early Recognition and Response Program," *Jt Comm J Qual Patient Saf*, vol. 41, no. 11, p. 483–491, 2015.
-
- [20] O. Lori, "Early Recognition and Management of Maternal Sepsis," *Nursing for Women's Health* , 2016.
-
- [21] A. S. Michael, J. B.-M. Samantha and D. P. Gavin, "Identification of adults with sepsis in the prehospital environment: a systematic review," *BMJ Open* , 2016.
-
- [22] B. F. Deshayne, H. Steven, A. W. Coralie, A. W. Lindsay, S. Q. M. Malia, C. Pranesh, L.-M. Thierry, K. P. Beth and W. Kumanan, "Using newborn screening analytes to identify cases of neonatal sepsis," *SCIENTIFIC REPORTS* , 2017.
-
- [23] B. Mercedes, P. S. Joao, A. Edgardo, F. Bukola, K. Marian, K. Seni, L. Pisake, N. Ashraf, N. Ruta, B. Vanessa and G. A. Metin, "The global maternal sepsis study and awareness campaign (GLOSS): study protocol," *Reproductive Health*, 2018.
-
- [24] A. S. Birju and F. P. James, "Neonatal sepsis An old problem with new insights," *Virulence*, vol. 5, no. 1, p. 170–178, 2014.
-
- [25] D. M. Clarissa, S. M. Karen and K. J. M. Sunil, "Biomarkers of Neonatal Sepsis.," *Neoreviews*, vol. 16, no. 5, p. e297–e308, 2015.
-
- [26] F. Aline, B. Julia, M. Shrey, S. Mike and N. V. D. A. Johannes, "Antibiotic Use for Sepsis in Neonates and Children: 2016 Evidence Update," *WHO*, 2016.
-
- [27] K. Claus, F. K. René, B. Giuseppe, F. M. Rolf and S. Martin, "Culture-Negative Early-Onset Neonatal Sepsis – At the Crossroad Sepsis Care and Antimicrobial Stewardship," *Front. Pediatr.*, 2018.
-
- [28] J. T. Priti, O. Anjumanara and M. Rachel, "Audit of Antibiotic Prescribing Practices for Neonatal Sepsis and Measurement of Outcome in New Born Unit at Kenyatta National Hospital," *International Journal of Pediatrics*, 2019.
-
- [29] R. Kartik, M. Sahana, D. Prathiba and J. Malavika, "Evaluation of screening of neonatal sepsis," *International Journal of Contemporary Pediatrics*, vol. 5, no. 2, pp. 580-583, 2018.
-
- [30] J. M. Aaron, C. H. Mary, F. Daniel, O. Svetlana, S. Lakshmi, P. B. Christopher, B. Fran, S. Melissa and W. G. Robert, "Machine learning models for early sepsis recognition in the neonatal intensive care unit using readily available electronic health record data," *PLoS ONE* , vol. 14, no. 2, 2019.
-
- [31] E. H. Katharine, N. H. David, J. P. Peter and S. Suchi, "A targeted real-time early warning score (TREWScore) for septic shock," *Science Translational Medicine*, vol. 7 , no. 299, 2015.
-
- [32] R. Taylor, J. Pare and A. Venkatesh, "Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data–Driven, Machine Learning Approach," *Acad Emerg Med.* , vol. 23, no. 3, p. 269–278., 2016.
-
- [33] T. Desautels, J. Calvert and J. Hoffman, "Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach," *JMIR Med Inform.* , vol. 4, no. 3, 2016.
-

- [34] J. K. Hye and Y. K. Ha, "Learning representations for the early detection of sepsis with deep neural networks," *Computers in Biology and Medicine*, 2017.
-
- [35] M. Moor, B. Rieck, M. Horn, C. Jutzeler and K. Borgwardt, "Early Prediction of Sepsis in the ICU using Machine Learning: A Systematic Review," *medRxiv*, vol. 8, no. 7, pp. 952-960, 2020.
-
- [36] A. A. D. T. J. M. C. G. P. I. J. C. M. Liberati and P. Devereaux, "The PRISMA statement for reporting systematic and meta-analyses of studies that evaluate interventions: explanation and elaboration.," *PLoS Medicine*, vol. 6, no. 7, pp. 1-28, 2009.
-
- [37] L. Mahieu, A. De Muynck, J. De Dooy, S. Laroche and K. Van Acker, "Prediction of nosocomial sepsis in neonates by means of a computer-weighted bedside scoring system (NOSEP score)," *Critical care medicine*, vol. 28, no. 6, pp. 2026-2033, 2000.
-
- [38] Y. Huang, X. Yu, W. Li, Y. Li, J. Yang, Z. Hu, Y. Wang, P. Chen, W. Li and Y. Chen, "Development and validation of a nomogram for predicting late-onset sepsis in preterm infants on the basis of thyroid function and other risk factors: Mixed retrospective and prospective cohort study," *Journal of Advanced Research*, vol. 24, pp. 43-51, 2020.
-
- [39] D. Lake, J. Richman, M. Griffin and J. Moorman, "Sample entropy analysis of neonatal heart rate variability," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 283, no. 3, pp. R789-R797, 2002.
-
- [40] K. Fairchild and T. O'Shea, "Heart rate characteristics: physiometers for detection of late-onset neonatal sepsis," *Clinics in perinatology*, vol. 37, no. 3, pp. 581-598, 2010.
-
- [41] Y. Xiao, M. Griffin, D. Lake and J. Moorman, "Nearest-neighbor and logistic regression analyses of clinical and heart rate characteristics in the early diagnosis of neonatal sepsis," *Medical Decision Making*, vol. 30, no. 2, pp. 258-266, 2010.
-
- [42] Y. Hu, V. Lee and K. Tan, "Prediction of clinicians' treatment in preterm infants with suspected late-onset sepsis—An ML approach. In 2018 13th IEEE conference on industrial electronics and applications (ICIEA)," *IEEE*, pp. 1177-1182, 2018.
-
- [43] R. Gómez, N. García, G. Collantes, F. Ponce and P. Redon, "Development of a non-invasive procedure to early detect neonatal sepsis using HRV monitoring and machine learning algorithms. In 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)," *IEEE*, pp. 132-137, 2019.
-
- [44] M. Griffin, D. Lake and J. Moorman, "Heart rate characteristics and laboratory tests in neonatal sepsis," *Pediatrics*, vol. 115, no. 4, pp. 937-941, 2005.
-
- [45] M. Griffin, D. Lake, T. O'Shea and J. Moorman, "Heart rate characteristics and clinical signs in neonatal sepsis," *Pediatric research*, vol. 61, no. 2, pp. 222-227, 2007.
-
- [46] A. Herbst and K. Källén, "Time between membrane rupture and delivery and septicemia in term neonates," *Obstetrics & gynecology*, vol. 110, no. 3, pp. 612-618, 2007.
-
- [47] D. H. D. Van, S. Van Kuijk, B. Opmeer, C. Willekes, J. Van Beek, A. Mulder, A. Van Loon, M. Groenewout, G. Mantel, K. Bloemenkamp and M. Porath, "Can neonatal sepsis be predicted in late preterm premature rupture of membranes? Development of a prediction model," *European Journal of Obstetrics & Gynecology and Reproductive Biology*, vol. 176, pp. 90-95, 2014.
-
- [48] K. Fairchild, D. Lake, J. Kattwinkel, J. Moorman, D. Bateman, P. Grieve, J. Isler and R. Sahni, "Vital signs and their cross-correlation in sepsis and NEC: a study of 1,065 very-low-birth-weight infants in two NICUs," *Pediatric research*, vol. 81, no. 2, pp. 315-321, 2017.
-
- [49] J. Thakur, S. Pahuja and R. Pahuja, "Neonatal sepsis prediction model for resource-poor developing countries. In 2018 2nd International Conference on Electronics, Materials

-
- [50] A. Masino, M. Harris, D. Forsyth, S. Ostapenko, L. Srinivasan, C. Bonafide, F. Balamuth, M. Schmatz and R. Grundmeier, "Machine learning models for early sepsis recognition in the neonatal intensive care unit using readily available electronic health record data," *PLoS one*, vol. 14, no. 2, p. e0212665, 2019.
-
- [51] J. Thakur, S. Pahuja and R. Pahuja, "Non-invasive prediction model for developing countries to predict SEPSIS in neonates," *Biomedical Engineering: Applications, Basis and Communications*, vol. 31, no. 1, p. 1950001, 2019.
-
- [52] L. Mithal, R. Yogev, H. Palac, D. Kaminsky, I. Gur and K. Mestan, "Vital signs analysis algorithm detects inflammatory response in premature infants with late onset sepsis and necrotizing enterocolitis," *Early human development*, vol. 117, pp. 83-89, 2018.
-
- [53] I. Stanculescu, C. Williams and Y. Freer, "Autoregressive hidden Markov models for the early detection of neonatal sepsis," *IEEE journal of biomedical and health informatics*, vol. 18, no. 5, pp. 1560-1570, 2013.
-
- [54] S. Mani, A. Ozdas, C. Aliferis, H. Varol, Q. Chen, R. Carnevale, Y. Chen, J. Romano-Keeler, H. Nian and J. Weitkamp, "Medical decision support using machine learning for early detection of late-onset neonatal sepsis.," *Journal of the American Medical Informatics Association*, vol. 2, no. 21, pp. 326-336, 2014.
-
- [55] I. Güzey and Ö. Uçar, "Machine Learning Based Non-Invasive Vital Signs Analysis for Preterm Sepsis Risk Prediction," *International Journal of Research in Engineering and Applied Sciences (IJREAS)*, vol. 8, no. 12, 2018.
-
- [56] W. Song, S. Jung, H. Baek, C. Choi, Y. Jung and S. Yoo, "A Predictive Model Based on Machine Learning for the Early Detection of Late-Onset Neonatal Sepsis: Development and Observational Study.," *JMIR Medical Informatics*, vol. 8, no. 7, p. 15965, 2020.
-
- [57] M. Beltempo, I. Viel-Thériault, R. Thibeault, A. Julien and B. Piedboeuf, "C-reactive protein for late-onset sepsis diagnosis in very low birth weight infants," *BMC pediatrics*, vol. 18, no. 1, p. 16, 2018.
-
- [58] T. Popowski, F. Goffinet, F. Maillard, T. Schmitz, S. Leroy and G. Kayem, "Maternal markers for detecting early-onset neonatal infection and chorioamnionitis in cases of premature rupture of membranes at or after 34 weeks of gestation: a two-center prospective study," *BMC pregnancy and childbirth*, vol. 11, no. 1, p. 26, 2011.
-
- [59] A. Palatnik, L. Liu, A. Lee and L. Yee, "Predictors of early-onset neonatal sepsis or death among newborns born at < 32 weeks of gestation," *Journal of Perinatology*, vol. 39, no. 7, pp. 949-955, 2019.
-
- [60] M. Griffin, T. O'Shea, E. Bissonette, F. Harrell, D. Lake and J. Moorman, "Abnormal heart rate characteristics preceding neonatal sepsis and sepsis-like illness," *Pediatric research*, vol. 53, no. 6, pp. 920-926, 2003.
-
- [61] I. Gur, A. Riskin, G. Markel, D. Bader, Y. Nave, B. Barzilay, F. Eyal and A. Eisenkraft, "Pilot study of a new mathematical algorithm for early detection of late-onset sepsis in very low-birth-weight infants," *American journal of perinatology*, vol. 32, no. 04, pp. 321-330, 2015.
-
- [62] J. Bekhof, J. Reitsma, J. Kok and I. Van Straaten, "Clinical signs to identify late-onset sepsis in preterm infants," *European journal of pediatrics*, vol. 172, no. 4, pp. 501-508, 2013.
-
- [63] S. Dutta, R. Reddy, S. Sheikh, J. Kalra, P. Ray and A. Narang, "Intrapartum antibiotics and risk factors for early onset sepsis," *Archives of Disease in Childhood-Fetal and Neonatal Edition*, vol. 95, no. 2, pp. 99-103, 2010.
-

- [64] M. Griffin and J. Moorman, "Toward the early diagnosis of neonatal sepsis and sepsis-like illness using novel heart rate analysis," *Pediatrics*, vol. 107, no. 1, pp. 97-104, 2001.
-
- [65] Y. Hu, V. Lee and K. Tan, "An application of convolutional neural networks for the early detection of late-onset neonatal sepsis. In 2019 International Joint Conference on Neural Networks (IJCNN)," *IEEE*, pp. 1-8, 2019.
-
- [66] D. Husada, P. Chanthavanich, U. Chotigeat, P. Sunttarattiwong, C. Sirivichayakul, K. Pengsaa, W. Chokejindachai and J. Kaewkungwal, "Predictive model for bacterial late-onset neonatal sepsis in a tertiary care hospital in Thailand," *BMC infectious diseases*, vol. 20, no. 1, pp. 1-11, 2020.
-
- [67] R. Rosenberg, A. Ahmed, S. Saha, M. Chowdhury, S. Ahmed, P. Law, R. Black, M. Santosham and G. Darmstadt, "Nosocomial sepsis risk score for preterm infants in low-resource settings," *Journal of tropical pediatrics*, vol. 56, no. 2, pp. 82-89, 2010.
-
- [68] A. Johnson, J. Aboab, J. Raffa, T. Pollard, R. Deliberato, L. Celi and D. Stone, "A comparative analysis of sepsis identification methods in an electronic database," *Critical care medicine*, vol. 46, no. 4, p. 494, 2018.
-
- [69] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PloS one*, vol. 10, no. 3, p. e0118432, 2015.
-
- [70] E. Pinker, "Reporting accuracy of rare event classifiers," *NPJ digital medicine*, vol. 1, no. 1, pp. 1-2, 2018.

Tables

Tables 6-10 and 12 are in the supplementary files section.

Figures

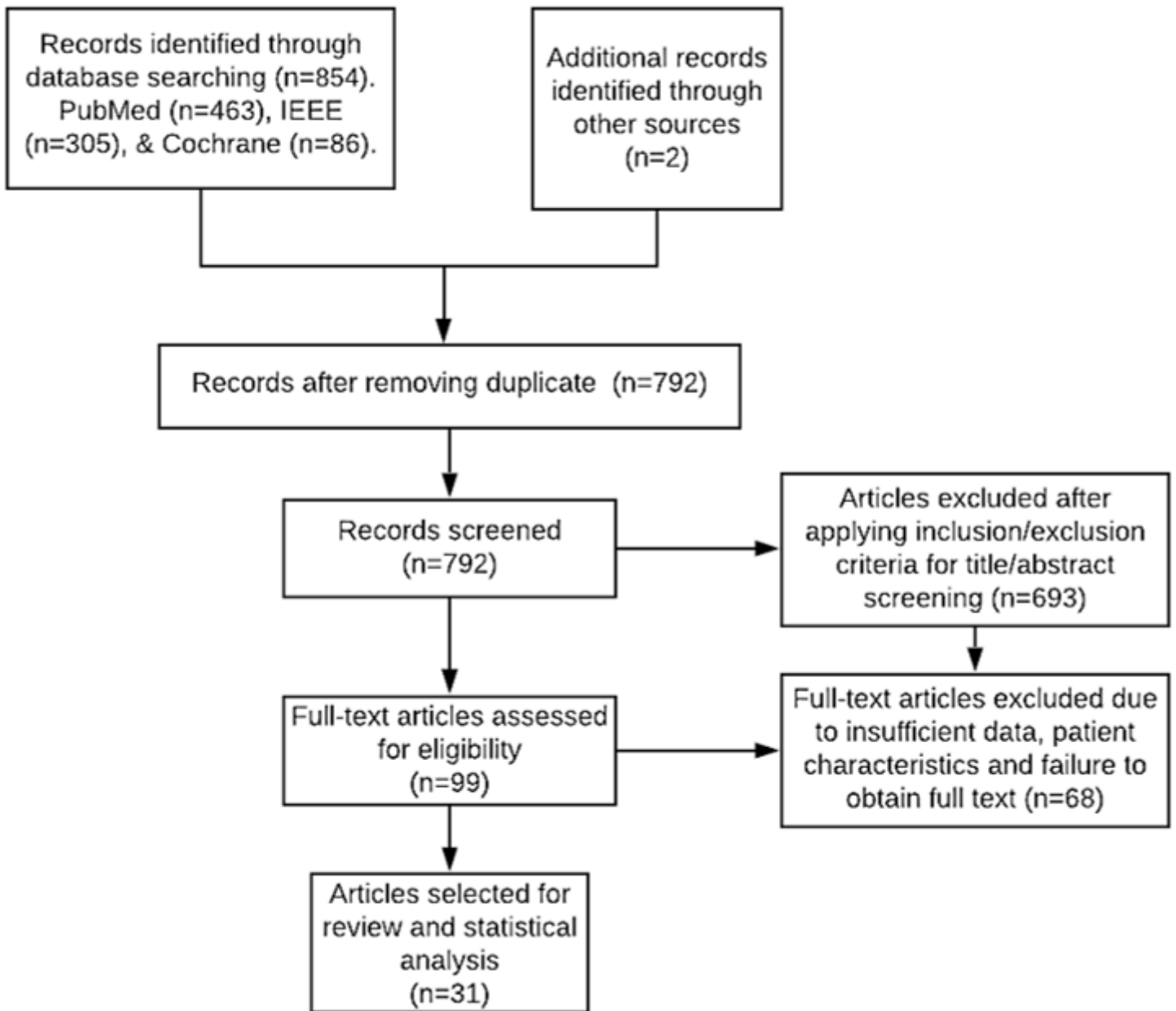


Figure 1

PRISMA flow diagram showing the search strategy and identified articles [36]

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Tables.docx](#)