Prediction of mortality among neonates with sepsis in the neonatal intensive care unit: A machine learning approach

Faiza Iqbal a, Prashant Chandra b, Aakif Ashar Khan c, Leslie Edward S Lewis a, 1,*, Dinesh Acharya d, K.E. Vandana e, P. Jayashree e, Padmaja A. Shenoy e

a Department of Pediatrics, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal, Karnataka, India
b Department of Pharmacy Practice, Manipal College of Pharmaceutical Sciences, Manipal Academy of Higher Education, Manipal, Karnataka, India
c Department of Artificial Intelligence and Machine Learning, Manipal School of Information Science, Manipal Academy of Higher Education, Manipal, Karnataka, India
d Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India
e Department of Microbiology, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal, Karnataka, India

* Corresponding author.
E-mail addresses: faiza.iqbal1@learner.manipal.edu (F. Iqbal), prashantpharmacy@yahoo.com (P. Chandra), Akifkhan892@gmail.com (A.A. Khan), leslie.lewis@manipal.edu (L. Edward S Lewis), dinesh.acharya@manipal.edu (D. Acharya), Vandana.ke@manipal.edu (K.E. Vandana), jaya.p@manipal.edu (P. Jayashree), padmaja.shenoy@manipal.edu (P.A. Shenoy).

https://doi.org/10.1016/j.cegh.2023.101414

Received 29 May 2023; Received in revised form 28 August 2023; Accepted 12 September 2023

ABSTRACT

Introduction: The use of machine learning (ML) methods can help clinicians predict neonatal sepsis better. Predicting mortality due to sepsis is essential for benchmarking and assessing NICU healthcare services.

Methodology: The newborn records of those diagnosed with neonatal bacterial sepsis were reviewed retrospectively over five years. For feature selection and model development, the WEKA v-3.8.6 tool was employed. Numerous ML models, including Naïve Bayes, Random Forest, Bagging, Logistic Regression, and J48 models, were created after identifying significant risk factors for newborn sepsis. Based on these models’ reliability, we used them to predict sepsis and mortality in the NICU.

Result: Records of 388 sepsis patients were used to build the model using training and test data sets. Mortality was best predicted using the feature selection method, OneR attribute evaluation + Ranker method, and logistic regression performed better (A = 88.4; ROC = 0.906) than others.

Conclusion: These effective ML models can assist clinicians in forecasting mortality in neonates admitted to NICUs with sepsis.

1. Introduction

With the worldwide increase in the rate of sepsis in neonatal intensive care units (NICUs), the mortality rates in neonates have significantly increased. Global Burden of Disease study report (2016–17) showed a global incidence of 1.3 million cases of neonatal sepsis (NS) per year with more than two hundred thousand sepsis-attributable deaths. NS is a blood infection in newborns less than 28 days old. 2, 3 Risk factors such as rupture of membrane, very low birth weight (VLBW), prematurity, invasive medical procedures, poor intra- and postpartum hygiene, maternal pyrexia, and prolonged NICU stay are critically associated with NS. 2, 3, 8

Neonatal morbidities is defined as the risk of death during the newborn period - the first 28 days of life. Severe neonatal morbidities, included: respiratory distress syndrome (RDS), neonatal sepsis, hypoxic ischemic encephalopathy (HIE), convulsions of new-born, brachial plexus palsy, persistent foetal circulation/neonatal hypertension, grade III or IV intraventricular haemorrhage (IVH), gastrochisis, diaphragmatic hernia, omphalocele, and congenital malformations of the circulatory system listed under International Classification of Diseases (ICD) –10.4 The major causes of neonatal mortality in India are prematurity, birth asphyxia, and sepsis. 5 Delays in illness recognition and care seeking, a dearth of primary health care providers, and limited access to facility care contribute to the deaths. A blood culture test is required for the conclusive diagnosis of culture-positive sepsis, and it typically takes 3–4 days to get the findings. Antibiotics are continued when sepsis-related clinical symptoms are present despite any negative results from blood cultures due to the likelihood of culture-negative sepsis. 6 In 2019, the Centre for Disease Dynamics, Economics and Policy (CDDEP), U.S.A., reported nearly one million neonatal deaths annually within the
first four weeks of life, of which 190,000 deaths were caused by sepsis in India.\textsuperscript{7} NS, especially in VLBW populations, is significantly associated with increased complications of prematurity and adverse neuro-developmental outcomes, which underlie the need for immediate recognition of NS and the commencement of antibiotic therapy.\textsuperscript{5,6}

An effective early warning system combined with the best clinical approach offers the best chance of lowering mortality. Several clinical prognostic tools, including the neonatal Sequential Organ Failure Assessment (nSOFA), Modified Early Warning Score (MEWS), Systemic Inflammatory Response Syndrome (SIRS), and quick SOFA, have been used to predict the risk of death following the onset of sepsis.\textsuperscript{11,12} But the conventional scoring system has a low prediction accuracy of mortality among neonates with sepsis. Machine learning (ML) algorithms can improve prediction accuracy in predicting sepsis in neonates admitted to NICU.

ML algorithms utilizing routinely recorded information and electronic data records have been shown to predict mortality, and morbidity, assisting clinicians in making more appropriate treatment decisions. The ML methods can leverage predictive power for early diagnosis of mortality by comparing accuracy (A\textsuperscript{2}), precision (P), recall (R) and receiver operative curve (ROC) values. The primary aim of this study is to develop an ML model using the WEKA tool that predicts mortality by using non-invasive vital sign data and other maternal and neonatal attributes. The secondary outcome compared the ML method in this study to predict sepsis-related death with various standard scoring systems to forecast sepsis-related neonatal mortality.

2. Methodology

2.1. Patients, setting and study design

A retrospective study was carried out on neonates diagnosed with NS between January 2017 and November 2021 in a South Indian tertiary care teaching hospital, with prior approval of the institutional ethics committee (IEC: 40/2022). Neonates with sepsis (both culture-positive and culture-negative) were included in the study. Fungal or viral sepsis was excluded. Data concerning neonates diagnosed with bacterial sepsis (classified under ICD Code No P36) was exclusively collected from the medical record department (MRD). Following this, cases that resulted in mortality were then identified and analysed separately. Deaths caused due to other causes were excluded. Demographic details of neonates (gender, gestational age (GA), birth weight, and other clinical features) were collected from case records. Antimicrobial susceptibility data were taken from ‘Laboratory Information Services’ maintained by the hospital. Following data collection, 90 patient attributes/features were discovered.

2.2. Study variables

Baseline demographics, clinical features, maternal predictors and laboratory findings were used as variables to predict the mortality among neonates with sepsis. Baseline demographics, including birthweight, GA, gender, mode of delivery, parity, and prenatal history of the neonate, were documented. Clinical features included apnoea, tachypnea, Metabolic acidosis, leucocytosis, hyperglycaemia, bradycardia, desaturation, tachycardia, respiratory distress etc. Laboratory data such as white blood cell count, haemoglobin, platelet count, C-reactive protein, bilirubin, and renal and hepatic function results were measured at the onset of the infection. On the basis of previously published literature,\textsuperscript{5,11,14} newborn and maternal clinical variables were retrospectively recorded and gathered from neonatal and maternal case sheets received from the MRD. The onset of sepsis was defined when a microbiological blood culture report showed the presence of bacteria. If the culture is negative, however, sepsis is clinically suspected based on the systemic inflammatory response syndrome (SIRS), which is the presence of clinical signs, symptoms and laboratory evidence suggesting sepsis and prompting the start of antibiotic therapy.\textsuperscript{15} Early onset of sepsis and late onset of sepsis was defined as sepsis occurring within first 72 h of life and at or 72 h of life, respectively.\textsuperscript{16} Data were extracted independently by two investigators (FI and PC) from the case records and compiled into shared sheets, and the data was validated by a third author (LEL and DAU). The outcomes of neonates recruited were transferred to a datasheet which has 90 variables which was then generated for the feature selection process.

2.3. Sample size and missing data points

This research was conducted retrospectively, utilizing data collected from January 2017 to November 2021. Due to the nature of this study, there was no need for sample size calculations. The investigation involved the use of all available data drawn from the case records. The study used a “complete case analysis” approach to handle missing data. This means that any cases with more than four missing values for any of the variables of interest were likely excluded from the analysis. Complete case analysis was also used to conduct analyses, create models, and validate them using participants and data for every predictor.

2.4. Machine learning (ML) models

Waikato Environmental Knowledge Analysis-Machine Learning (WEKA-ML) version 3.8.6 tool was used for developing models. Different attribute or feature selection methods such as “cfs (Correlation based Feature Selection) attribute evaluation + Bestftr”, “correlation attribute evaluation + Ranker”, “OneR attribute evaluation”, and “Releif” attribute evaluation” were used and checked for the performance of various algorithms to predict the features associated with the mortality.\textsuperscript{17} This method provided a correlation value and assigned ranking to each attribute. The required number of attributes were selected based on the difference between the two ranks. At some point, if the difference between the two rankings are larger than the previous one, that was fixed as a threshold value (TV).

The selected attributes were used in training (supervised learning method) using the classification algorithm.

- Naive bayes (base line model)
- Logistic regression
- Sequential minimal optimization (SMO)
- Classification and Regression tree (CART) models such as J48 and Random Forest

2.5. Data pre-processing

The data was pre-processed by loading the dataset in CSV format followed by cleaning data, handling missing values, transforming data, and selecting attributes. The “Filter” option was used for data cleaning, missing value replacement, and data transformation. Lastly, for attribute selection the “Select attributes” panel was used for reducing the dimensionality of the dataset.

2.6. Cross-validation

We also employed the k-fold cross-validation approach (for example, the 10-fold cross-validation technique), which works on the same principle as k-1 and categorize the dataset into two parts: training and test sets. The first set is chosen as the test set in the first iteration, and the remaining k-1 sets are used to train the model. The second set is chosen as a test set, while the remaining k-1 sets are utilized to train the data in the second iteration. This method is repeated for all k sets.

2.7. New or test data set

At the same time, we employed a 66/34 and an 80/20% split, with
34% and 20% being deemed test data. The model may be used for prediction if it performs well on test data and is highly accurate, indicating that it hasn’t overfit the training data.1,4,19

It is determined whether the recommended framework is stable and reliable by comparing the performance of the training and testing sets. We assessed the prediction error of the training and tested data sets to examine if the ML algorithm had issues with overfitting or underfitting. If the prediction error is low for the training data set but high for the tested data set, the technique outputs suffered from excessive variance due to the overfitting issue. The technique was underfitting, though, and the outcome is highly biased if the prediction error was considerable for both the training and testing data sets.20

In binary and multiclass classification tasks, a classifier’s performance is evaluated using four metrics: A, P, R, and ROC values, based on these evaluation metrics, the best feature selection method and the respective model was selected and reported. The model’s performance was considered best and accurate based on higher the values of ‘A’, P, R and ROC.

‘A’ = Percentage of accurate predictions made by a classifier across all instances in the test set.

\[ P = \frac{TP}{TP + FP} \]

where, \( P=\)Precision, \( TP = \) True positive and \( FP = \)False positive.

Precision enables to visualize how consistently the ML model classifies the model as being positive.21

\[ R = \frac{TP}{TP + FN} \]

where, \( R=\)Recall \( TP = \) True positive, \( FN = \)False positive.

The degree or measure of separability is represented by area under the curve (AUC), and ROC is a probability curve. It indicates how well the model can distinguish between classes. The more accurate the model is in classifying 0 classes as 0, and 1 classes as 1, the higher the AUC.22

3. Results

3.1. Results

3.1.1. Demographics

The records of 388 sepsis patients were used to build and validate the model using training and test data sets. Among 388 neonates, 62% were term and 38% preterm. Males were more predominant (n = 247) than females (n = 141).47% (n = 184) of those with clinically suspected sepsis had blood culture-positive confirmed NS, while 52% (n = 204) had blood culture-negative clinical sepsis. The neonatal mortality rate was 39.6% (n = 154). These mortality cases occurred at the age of 19.0 (13.0–39.0) (median interquartile range, IQR). Clinical characteristics, neonatal morbidities and mortality are reported in Table 1.

4. Development of ML models

Using different attribute or feature selection methods such as “cfs attribute evaluation + Bestfit”, “correlation attribute evaluation + Ranker”, “OneR attribute evaluation”, and “Relieff” attribute evaluation method and the model that provides good ‘A’ and ROC. The attributes and its correlation value are provided in Fig. 1.

4.1. Mortality

Of the 90 attributes, we have selected 34 attributes using the OneR attribute evaluation + Ranker method with a TV = 63.1443. The obtained features were applied on different classification algorithms such as Naïve Bayes, Random Forest, SMO, Logistic Regression, and J48 models, however, Logistic Regression algorithm performed well over the others with ‘A’ = 88.4 ROC = 0.906 on 20% of the tests data set (Table 2). Since the root mean square error (RMSE) for the training set and test set (percentage split of 80/20) was very low, the model didn’t show any variance or bias, ruling out the problem of overfitting and underfitting. Predictors of mortality due to sepsis are shown in Fig. 3. ROC of different algorithms is produced in Fig. 2. Similarly, we have also shown the results of SMO (Table 3) that was obtained as the second best model compared to Logistic Regression.

Attributes highly associated with neonatal mortality among neonates with sepsis were divided into four domains: maternal predictors, demographic attributes, clinical attributes and laboratory attributes. Premature rupture of membrane, absent end diastolic flow, and chorioamnionitis were shown to be the maternal predictors; demographic characteristics associated with mortality due to sepsis were preterm birth, LBW, birthweight (>2500 g) and appearance, pulse, grimace, activity, and respiration (APGAR) at 5 min. Clinical and laboratory attributes related to neonatal death are summarised in Fig. 2.

5. Discussion

In this research, we devised predictive models for assessing mortality.
rates among neonates afflicted with sepsis. The recruited neonates were admitted to the NICU of Kasturba Hospital, which is a private healthcare institution. The hospital is equipped with 45 beds in total, with 30 dedicated for intensive care, and 15 for intermediate care beds. The ratio of nurses to new-borns in intensive care and emergency care wards is 1:2 and 1:3, respectively, and the average bed occupancy on a daily basis fluctuates between 75% and 80%. Among the admissions to the NICU, 60%–65% are inborn infants, while 30%–35%, are from other healthcare facilities from neighbouring districts. Even though the NICU’s capacity to sustain a controlled environment has significantly helped in diminishing the mortality related to sepsis, but there are certain limitations regarding birth weight and gestational age that determine whether a neonate will be resuscitated or not. Specifically, babies with a birth weight of less than 500 g or with a gestational age of less than 23 weeks are not resuscitated in this unit after discussing with parents and treating obstetrician.

The current study documented the incidence of bacterial sepsis in around 2.4% of live newborns annually for five years in a row. The ‘Burden of Antibiotic Resistance in Neonates from Developing Societies’ (BARNARDS) study showed the incidence of bacterial sepsis as 0.04%. Faster and more effective detection of NS is required to target interventions to reduce the incidence and subsequent mortality. ML-based predictive tools may aid medical decision-making by providing novel elements to improve the correct and early identification of sepsis. Early detection of sepsis with targeted antibiotic therapy allows clinicians to implement supportive treatments, lowering mortality and reducing antibiotic resistance and healthcare costs.

The current study assessed the demographics of all the neonates. In our study, the mean GA was 34.12 ± 4.58 weeks, and the median birthweight was 2155 g. Sixty-two percent of all neonates were preterm, and 19% were small for gestational age (SGA), neonatal morbidities like apnoea, respiratory distress syndrome (RDS) and bradycardia were seen in 32%, 27% and 27% of the neonates, respectively. A study from Iran by Sheikhtaheri et al. (2021) from Iran reported mean birthweight and GA as 2323.2 g and 240.49 days, respectively. Fifty-seven per cent of all neonates were preterm, and among them, 0.5% were SGA. Additionally, 31.3% of neonates had RDS, and 84.6% needed ventilation.

WEKA software was used as a database to predict mortality using different feature selection methods. This study compares different ML models like Naïve Bayes, SMO, Logistic regression, and Random Forest to determine the best model to deploy in hospitals. The model is trained using features chosen from the dataset. The attribute selected were characterized under four domains (Maternal predictors, Demographic characteristics, Clinical attributes and Laboratory variables) for

![Fig. 1. Attribute selection for prediction of mortality.](image-url)
The study showed Logistic Regression (‘A’ = 88.4; ROC = 0.906) performed well in predicting neonatal mortality due to sepsis with a 80/20% split using the feature selection method correlation attribute evaluation. The majority of studies are focused on predicting mortality in NICU. However, Sheikhtaheri et al. (2021) developed a model for the prediction of neonatal mortality in NICU and concluded that the highest ROC was obtained for Support Vector Machine (SVM) (‘A’ = 94%; ROC = 0.98) and Ensemble models (‘A’ = 95%; ROC = 0.98). We observed that the model built with ML algorithms has much better predictive power than the traditional neonatal severity scoring systems, which are limited by the inability to consider integrated variables fully. Muk presents (2019) validated the SNAPPE-II score (Neonatal Acute Physiology with Perinatal Extension-II) to predict neonatal mortality in NICU. The study revealed that a cut-off score of 38 with a sensitivity of 84.4%, specificity of 91%, positive predictive value of 66.7%, and negative predictive value of 88.4% was found.

### Table 3

Performance evaluation of the SMO algorithm over the mortality with its confusion matrix.

<table>
<thead>
<tr>
<th>Model</th>
<th>Method</th>
<th>Accuracy (A) (%)</th>
<th>TP</th>
<th>FP</th>
<th>P</th>
<th>Recall (R)</th>
<th>F-measure</th>
<th>RMSE</th>
<th>ROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMO</td>
<td>Training set</td>
<td>86.3</td>
<td>0.863</td>
<td>0.160</td>
<td>0.863</td>
<td>0.863</td>
<td>0.863</td>
<td>0.3696</td>
<td>0.809</td>
</tr>
<tr>
<td>10-fold-CV</td>
<td>82.9</td>
<td>0.830</td>
<td>0.207</td>
<td>0.829</td>
<td>0.830</td>
<td>0.828</td>
<td>0.4124</td>
<td>0.767</td>
<td></td>
</tr>
<tr>
<td>Percentage split</td>
<td>86.1</td>
<td>0.818</td>
<td>0.234</td>
<td>0.817</td>
<td>0.818</td>
<td>0.817</td>
<td>0.591</td>
<td>0.761</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>80.7</td>
<td>0.872</td>
<td>0.146</td>
<td>0.872</td>
<td>0.872</td>
<td>0.872</td>
<td>0.3581</td>
<td>0.625</td>
</tr>
</tbody>
</table>

SMO = Sequential minimal optimization, TP = True positive, FP = False positive, CV = Cross-validation, P = Precision, ROC = Receiver Operating Characteristic Curve, RMSE = Root mean square error.

**Fig. 2.** R.O.C. of mortality.

**Fig. 3.** Variables for prediction of mortality due to sepsis.
predictive value of 96.5% was the best for predicting mortality.\textsuperscript{12}

Hsu et al. (2021) applied ML algorithms such as Deep Neural Network (DNN), k-nearest neighbours, SVM, Random Forest, and extreme gradient boost for the prediction of mortality in clinically suspected sepsis. He concluded that DNN exhibited the highest AUC (0.923, 95% confidence interval [CI] 0.953–0.893) and the greatest accuracy (95.64%, 95% CI 96.76–94.52%) for prediction of mortality. We used the AUC, accuracy, F1 score, Kappa, and Matthews correlation coefficient (MCC) to compare all ML models. With the best AUC, accuracy, F1 score, Kappa, and MCC out of seven models, in this study the Logistic regression model is the most persuasive with 88.4% Accuracy in predicting mortality in neonates with sepsis with AUC.

6. Evidence before this study

There is sufficient evidence to predict neonatal deaths using artificial intelligence. Moreover, there is a lack of literature on neonatal mortality associated with sepsis.

7. Implication of evidence

Neonatal sepsis and its associated mortality are major global health concerns that are difficult to manage, particularly in low- and middle-income countries. With prospective validation, such predictive models can significantly benefit clinical exploitation. Future research should concentrate on external validation, calibration, and application deployment that healthcare professionals can easily access. Note that healthcare professionals can easily access such forecasting of future mortality can assist healthcare organisations in identifying high-risk neonates and avoiding crises. To concentrate on external validation, calibration, and application deployment that healthcare professionals can easily access. Such research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

No conflicts of interest are involved with this review. This present review is an original work and has never been published elsewhere.

Acknowledgements

The authors would like to acknowledge Kasturba Medical College and Manipal Academy of Higher Education for providing technical support.

References

17 Chowdhury MZL, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. Fam Med Community Health. 2020;8(1), e00262.


