Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 16s (2025)
https://www.jneonatalsurg.com

Early Diagnosis of Neonatal Sepsis Through Predictive Analytics and Feature Selection Techniques

Dr. Lilly Sheeba S[1], Ms. Rachel Evelyn R[2], Ms. Faritha Begum M[3], Ms. Preethi Parameswari S[4], Ms. Sandhiyaa S[5], Ms. Sangeetha N[6]

[1]Professor, Department: Cyber Security, Jerusalem College of Engineering, Chennai, Tamil Nadu, India
ORC ID: 0000-0002-9465-1300
Email ID: lillysheeba1@gmail.com
[2]Assistant Professor, Department: Cybersecurity, Jerusalem College of Engineering, Chennai, Tamil Nadu, India
Email ID: rachelevelyn12@gmail.com
[3]Assistant Professor, Department: Cyber Security, Jerusalem College of Engineering, Chennai, Tamil Nadu, India
Email ID: farithait14@gmail.com
[4]Assistant Professor, Department: Cyber Security, Jerusalem College of Engineering, Chennai, Tamil Nadu, India
ORC id: 0000-0002-7142-3700
Email ID: preethiparameswari8@gmail.com
[5]Assistant Professor, Department: Cyber Security, Jerusalem College of Engineering, Chennai, Tamil Nadu, India
Email ID: sandhiyaascs@jerusalemengg.ac.in
[6]Assistant Professor, Department: Cyber Security, Jerusalem College of Engineering, Chennai, Tamil Nadu, India
Email ID: sanrai2799@gmail.com

Cite this paper as: Dr. Lilly Sheeba S, Ms. Rachel Evelyn R, Ms. Faritha Begum M, Ms. Preethi Parameswari S, Ms. Sandhiyaa S, Ms. Sangeetha N, (2025) Early Diagnosis of Neonatal Sepsis Through Predictive Analytics and Feature Selection Techniques. *Journal of Neonatal Surgery*, 14 (16s), 251-258.

ABSTRACT

Goal 3.2 of the Sustainable Development Agenda aims to reduce the infant death rate by the year 2030. The leading causes of mortality in neonates are preterm and birth asphyxia, followed by neonatal infections. It is more probable for new-borns to get late-onset neonatal sepsis (LOS) from their surroundings than from their mothers. This kind of sepsis often manifests between 3 and 28 days of age. A difficult aspect of early LOS diagnosis is the lack of obvious clinical signs during the early stages of infection. Predicting LOS before obvious clinical signs is possible using physiological factors, according to studies. These metrics may be used as warning indications by clinicians to keep a careful eye on infants and act quickly to avoid problems and provide them good treatment.This research examines several machine learning algorithms that can forecast when new-born sepsis will start by analysing the MIMIC III dataset, which includes vital signs, laboratory results, and observations taken during the first 24 hours of arrival. Out of all the algorithms tested using 10-fold stratified cross-validation, the ones with the highest area under the receiver operating characteristic (AUROC) values were adaptive boosting (0.9248), light gradient boosting (0.9245), and random forest with Synthetic Minority Oversampling Technique (0.9238). With an AUROC of 0.9266, an accuracy of 0.8553, F1 score of 0.7829, and Matthew's correlation score of 0.6995, the soft voting classifier trained on an ensemble of the most effective three models identified the beginning of newborn sepsis.

Keywords: *Machine Learning, MIMIC III, SDG 3.2, Neonatal Sepsis and Vital signs.*

1. INTRODUCTION

Any baby less than four weeks old is considered a neonate. Among the 5 million deaths of children less than five in 2020, the World Health Organisation reported2.3 million of the world's children are new-borns. More over half of India's infant mortality rate happened in the first four weeks of life. Goal 3.2 of the Sustainable Development Agenda aims to zero infant mortality by the year 2030, with a goal of 12 per 1,000 live births. Premature or preterm babies are those born before the average gestational age of 37 weeks. Babies that are born prematurely or with serious health problems are treated by the neonatal intensive care unit (NICU). Medical professionals in neonatal intensive care units have been instrumental in saving the lives of many preterm and sick infants.The leading causes of mortality in neonates are preterm and complications linked to birth asphyxia, followed by neonatal infection.When the immune system releases chemicals into the bloodstream in response to an infection, it causes inflammation and blood clotting, which in turn reduces blood supply to vital organs and limbs, leading to the potentially fatal medical condition known as sepsis. Several organs, including the lungs, kidneys, and liver, might fail or even die in extreme instances of infection-induced hypotension, which is known as septic shock. There are two forms of newborn sepsis that are determined by when the infection begins: early onset sepsis (EOS) and late onset neonatal sepsis (LOS). Signs of infection in EOS usually start showing up no later than three days after the baby is born.

As a result of bacterial vertical transmission in the maternal vaginal canal, EOS is most often caused. The main cause of LOS is the slow and subtle horizontal transfer of infections that a baby contracts after delivery; this process begins about day three of life but have devastating consequences It is difficult to diagnose LOS sepsis early on because clinical symptoms are not noticeable in the early stages of infection. Physiological variables, such as respiratory characteristics and heart rate variability, may foretell LOS before overt clinical symptoms become apparent, according to studieshealth problemson pages 5 to 9 The conventional approach is shown in Fig. 1.learning where LOS is located.One publicly accessible dataset that researchers extensively use is MIMICIII, or Medical Information Mart for Intensive Care III used to construct models for the prediction of sepsis, mortality, duration of stay, and other outcomes.On the basis of patient data, researchers have developed clinical decision systemsthat propose suitable therapy. The dataset has been used by researchers to assess the security of medications administered in the ICU.Fornew-borns admitted to the NICU, a scoring system called SNAPPE-II (Score for Neonatal Acute Physiology with Perinatal Extension-II) is used to determine the severity of disease and the risk of death. The score estimates the new-born's health by combining physiological and perinatal variables. Because of their weakened immune systems, neonates with high SNAPPE-II scores are at increased risk of developing sepsis.

A scoring system called the Clinical Risk Index for Babies (CRIB) II is used to estimate the risk of death in children with a very low birth weight. It considers characteristics such as gestational age, birth weight, and other clinical data collected during the first 12 hours after admission to the neonatal intensive care unit (NICU).The study of computers that learn to do tasks automatically is known as machine learning. Machine learning is extensively used in several fields to reveal previously unseen correlations, patterns, and insights, including healthcare, education, production engineering, intrusion detection, fraud detection, consumer segmentation, and bioinformatics. when it comes to supervised machine learning methods for newborn illness diagnosis, the ensemble strategy outperforms Support Vector Machine (SVM), decision trees, and neural networks in terms of predictive powerTheFor their analysis of machine learning and deep learning algorithms' efficacy in predicting sepsis and neonatal sepsis, Parvin et al. methodically evaluated Scopus, Web of Science, and PubMed databases from 2015 to 2022. Area under the receiver operating characteristic (AUROC) values varied from 0.68 to 0.95 across eleven articles chosen from various medical care units. Several methods were tested in the research, such as neural networks, logistic regression, support vector machines, and random forests.The four main newborn disorders that Robi and Sitote have identified using the classification stacking approach are sepsis, birth asphyxia, necrotising enterocolitis (NEC), and respiratory distress syndrome. These four diseases are responsible for seventy-five percent of neonatal mortality. Asella Comprehensive Hospital provided the data set between 2018 and 2021. The constructed stacking model outperformed three alternative ML models: Xtreme Gradient Boosting, Random Forest, and Support Vector Machines The results indicate that the machine learning method may greatly aid in the correct identification of newborn illnesses, especially in healthcare institutions with limited resources.

This research compares several machine learning algorithms that may use vital signs, laboratory measures, and observations taken within 24 hours of arrival to predict when newborn sepsis would start. The voting classifier had the best results in terms of accuracy, recall, MCC, and F1 score after being trained on an ensemble of adaptive boosting, light gradient boosting without Synthetic Minority Oversampling Technique (SMOTE), and random forest with SMOTE. When medical complications are caught early on, therapy becomes more easier and more successful.

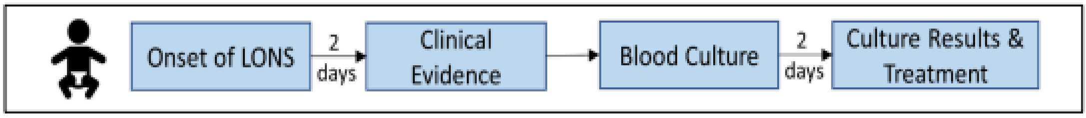

Figure 1: Conventional approach to LOS detection

2. METHODOLOGY

The methodical approach used to identify LOS early using the MIMIC-III dataset is shown in Figure 2. When data is collected, it goes through a series of steps called a data preparation pipeline. These steps include imputation to fill in missing values, normalisation to make the data more consistent, and feature selection to make the dataset more useful and high-quality. The next step in guaranteeing reliable model performance is to partition the dataset into training and testing sets.

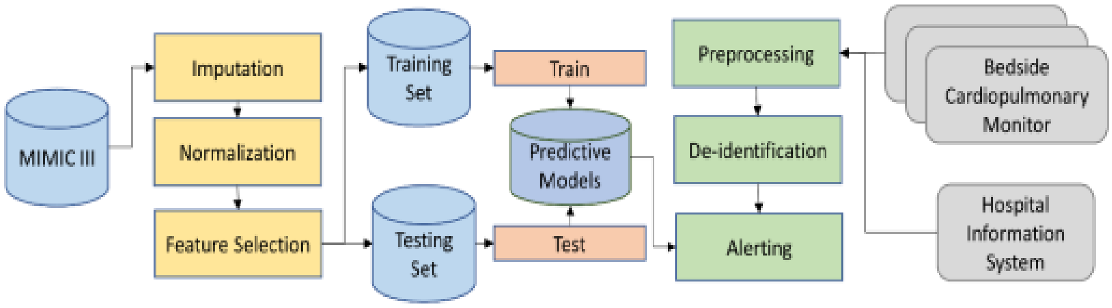

Figure 2: Overall Methodology

A variety of supervised ML algorithms are heavily trained on the training set within the context of model creation. To provide a thorough evaluation of performance and easy selection of the best model, the following model evaluation uses measures that have been cross-validated using the training set. Testing the generalisability of the trained models using a testing set is the last stage. In order to forecast the incidence of LOS, the model is fed preprocessed, deidentified data from the first twenty-four hours after admission, collected from bedside monitors and the hospital information system. Stakeholders are alerted when it is projected that a neonate may develop sepsis, allowing them to make well-informed choices.

2.1 Dataset

From 2001–2012, 61,532 patients were admitted to the intensive care unit at Beth Israel Deaconess Medical Centre as part of MIMIC-III[10–12]. It includes information about 7,870 newborns. After finishing an approved course on protecting human research subjects, which includes following HIPAA regulations and signing a data use agreement outlining acceptable data usage procedures, maintaining stringent security requirements, and prohibiting any effort to identify individual patients, access to MIMIC III was granted.

2.2 Preparing Data

To make data input into database systems easier, the MIMIC-III dataset is given as comma separated value (CSV) files with accompanying Structured Query Language (SQL) scripts. A backup of the full database—approximately 49 G.B. in size—was downloaded to a secure server running PostgreSQL 13.1 for easy access. We chose PostgreSQL because it is an open-source RDBMS that guarantees data integrity and stability and offers powerful optimisation features for queries. Through the utilisation of the Jupter notebook, sepsis-3-getdata. piny, and the entire MIMIC III database, a dataset was generated by querying numerous MIMIC III tables. These tables include admissions, chart events, isostasy, input events, output events, lab events, microbiology events, note events, patients, prescriptions, procedure events, and services.The Patients hospitalised after five days of birth had 53550 records removed from consideration for possible adulthood or infection-related admissions. Adult data will not be included for this research since it only covers new-borns. In order to focus on the initial 24 hours after hospitalisation for the purpose of predicting new-born sepsis, this research does not include secondary admissions. Due to concerns about data dependencies, biased findings, and over-estimation of the model's performance, 138 instances of second admission for the same person were excluded. In order to maintain data completeness and reliability in the study, seven entries were removed from the chart events table since they did not contain matching plotted data. The five records of infants diagnosed with endocrinopathy of the stomach (EOS) were not included in the analysis of LOS risk variables as EOS is associated with distinct risk factors and clinical presentations from LOS.The likelihood that the neonate will get sepsis served as the response variable in this investigation. One of the most common ways to identify a neonate as septic is by entering an ICD-9 number, such as 995.92 for severe sepsis or 785.52 for septic shock.2) Organ dysfunction considered to be caused by an infection (SOFA ≥ 2) score determined at the time of suspected infection (Sepsis-3).3) The ICD-9 codes . The amount of records found in MIMIC III for each criteria is shown in Table 1. Out of 7832 new-borns, 2184 (27.88%) were classified as septic according to the sepsis-3 criteria employed in this research. Sepsis-3 criteria is a valid and trustworthy way to diagnose new-borns with sepsis, according to studies.

Table 1:Distribution of information according to categories

Criteria	**Not Septic**	**Septic**
Explicit	8050	3
Sepsis-3	5500	2325
Martin et al.	7650	58

In order to get the values from the tables for 118 characteristics within the first 24 hours after admission, the database was searched using Jupyter notebook. The 35 traits were removed since they were not useful for any newborn. For neonates, MIMIC III data showed several missing values. The precision and dependability of diagnostic and predictive models are directly impacted by how missing information are handled. Misdiagnoses or incorrect predictions might have a devastating effect on patient safety and treatment choices if missing data is incorrectly input into models, which in turn introduces noise, biases, and reduces model performance. A reliable model can only be achieved with a strong imputation strategyphysician with more accurate forecasts and useful information. Various approaches were used to estimate the missing values in nominal features in our research. These methods included iterative imputation using 10 iterations of Light Gradient Boosting Machine (LightGBM), mean, median, mode, and k closest neighbours.Data normalisation makes sure that a highly valued characteristic won't affect the outcome just because it's there. The input data was preprocessed using three popular normalisation techniques: z-score, Min-Max, and Maximum Absolute Value Scalar. The objective was to find out how they affected the model's ability to predict the outcome variable.With the greatest area under the receiver operating characteristic curve (AUROC) of the three approaches, our findings demonstrated that Maximum Absolute Value Scalar normalisation performed the best. Because it had the greatest AUROC, which indicates greater discriminating and prediction accuracy, the Maximum Absolute Value Scalar was chosen as the normalisation approach for this investigation.

3. EVALUATION AND OUTCOME

Building and evaluating models utilising cross-validated findings from the training set was done using all the methods covered in the Methodology section. Table S1 displays the results of all models using the testing set without SMOTE, whereas Table S2 displays the results with SMOTE, ranked according to AUROC.The 10 best machine learning algorithms, ordered by AUROC, are shown in Table 2.Using features retrieved during the first 24 hours after admission, adaptive boosting without SMOTE achieved a highest AUROC of 0.9248, an accuracy of 0.8494, and an F1 score of 0.7277 on the testing set for predicting sepsis.The area under the receiver operating characteristic curve (AUROC) for light gradient boosting without SMOTE was 0.9245, whereas for random forest with SMOTE it was 0.9238. With an F1 score of 0.7774 and a recall of 0.9252, random forest using SMOTE performed the best.Figures 3(a-f) show the ROC curve and feature significance plot of the top three models chosen according to the AUROC score. To filter out features, we look for a minimum absolute Pearson correlation greater than 0.95. An essential part of all three models is the generation of urine. Blood tests may also reveal other characteristics, such as bilirubin, platelet count, salt, haemoglobin, white blood cell count, and heart rate. The classifiers' heatmaps may be seen in Figures 4, S1, and S2.A combination of the three best models chosen using AUROC was used to train the soft voting classifier. After all three models have produced their probabilities for each class, an average is determined. We take the class with the best average and use it as our output. Equation (1) is used by the soft voting classifier:

$$\hat{y} = arg\max_{i} \sum_{j=1}^{3} p_{ij} (1)$$

A soft voting classifier's main benefit is that it may enhance predictions by integrating the best features of the top three models and decreasing prediction variance, resulting in more reliable models.

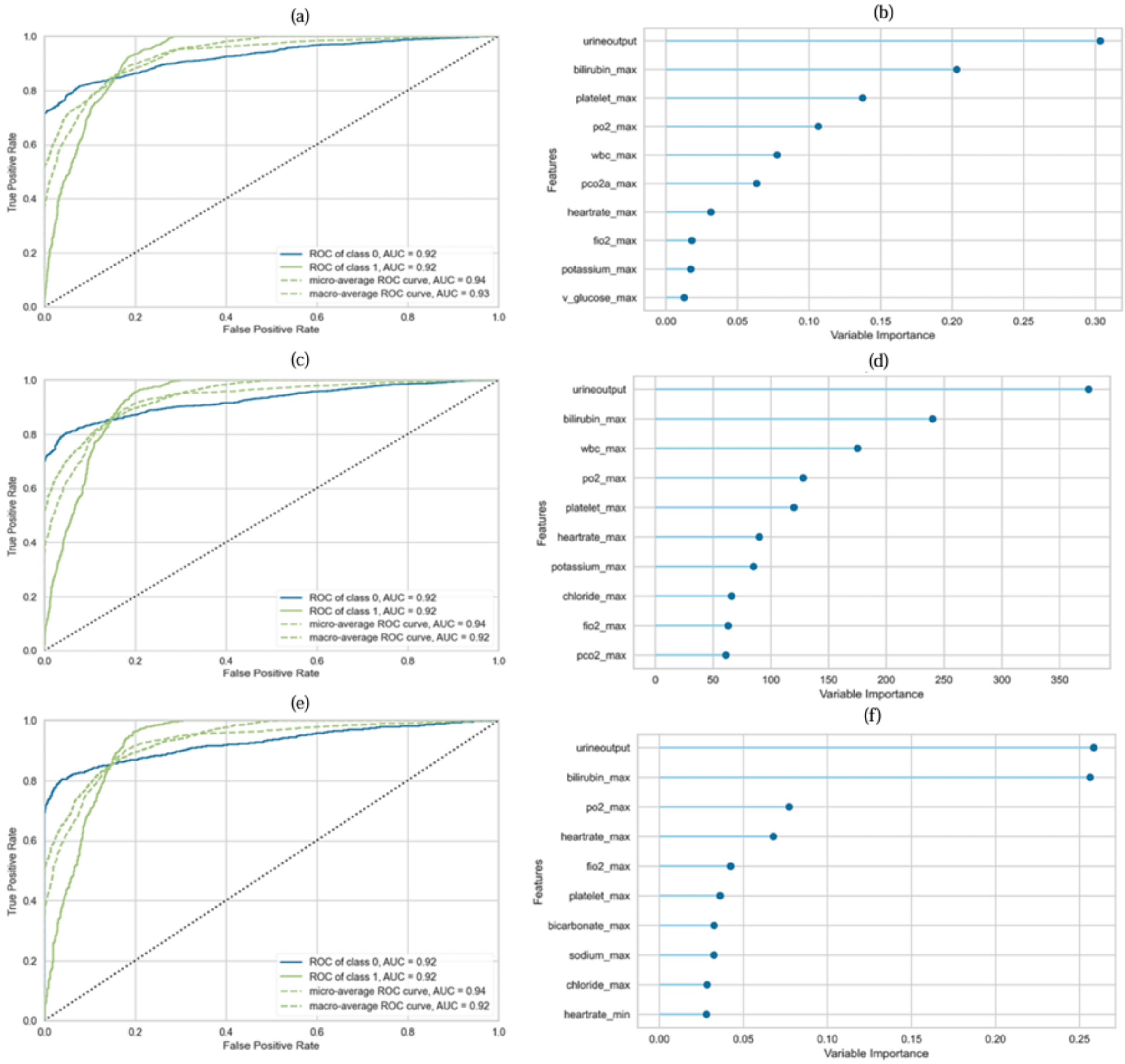

Figure 3 shows the ROC curves and importance of features plots for 3 different classification methods: adaptive boosting, the light gradient enhancement machine, and random forest (SMOTE).

A logistic regression classifier was used to determine the outcome of stacking the top three models chosen based on AUROC. This process resulted in a stacking classifier. Figure 5 shows the receiver operating characteristic (ROC) curve for voting and stacking classifiers. You can see the results of the voting and stacking classifier's data analysis in Table 3. With a top AUROC of 0.9266, an accuracy of 0.8553, an F1 score of 0.7829, and a recall of 0.9359 on holdout data, the voting classifier clearly shows great promise as a classifier. The classifier's excellent overall accuracy, balanced precision and recall, and great discriminating power allow it to distinguish between septic and non-septic cases, helping healthcare providers to prioritise patients for quick intervention. Working with unbalanced datasets requires a classifier that can effectively detect underlying data patterns; the MCC highlights this ability.

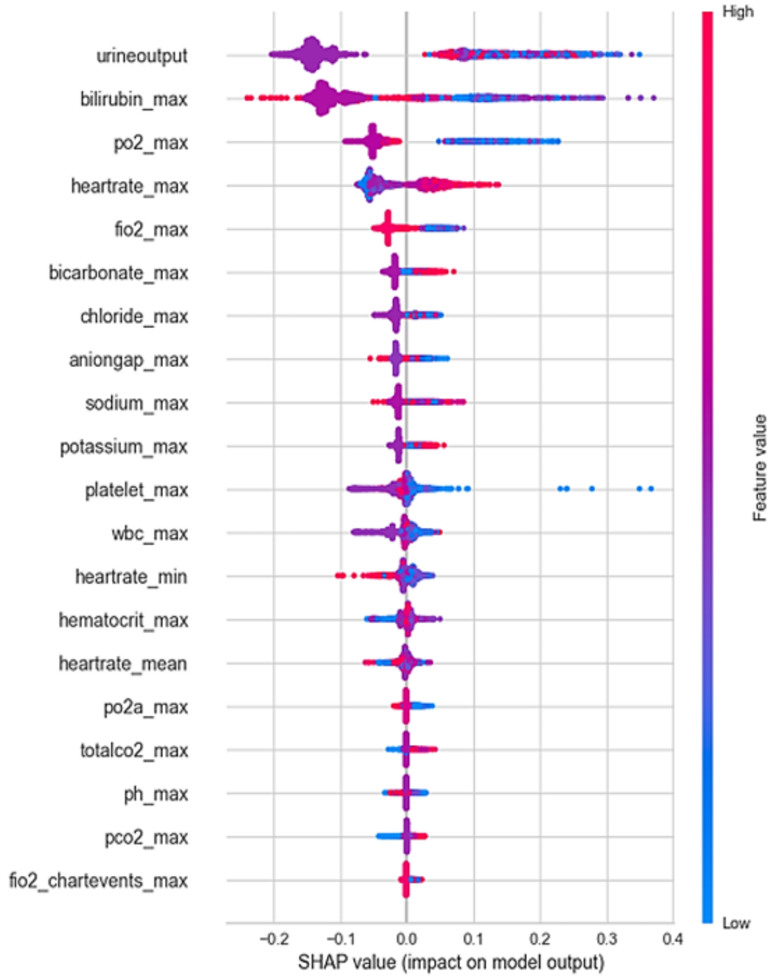

Figure 4: An AdaBoost Classification heatmap that does not use SMOTE

Table 2: Findings from the Evaluation of the Data: AdaBoost, Light Gradient Boosting Machine, Random Forests, Gradient Boosting, and Extra Trees.

Model	TP	FP	FN	TN	Acc.	AUROC	Recall	Prec.	F1	Kappa	MCC
AdaBoost	460	160	195	1530	0.8421	0.9183	0.7025	0.7419	0.7217	0.6104	0.6109
LightGBM	495	190	160	1500	0.8472	0.9199	0.7558	0.7226	0.7388	0.6285	0.6351
RF (SMOTE)	610	310	45	1385	0.8507	0.9201	0.9313	0.6635	0.7767	0.6642	0.6848
AdaBoost (SMOTE)	530	250	125	1445	0.8394	0.9187	0.8092	0.6795	0.7385	0.6241	0.6334
RF	510	200	145	1495	0.8538	0.9181	0.7786	0.7185	0.7472	0.6432	0.6505
GBC	498	205	157	1490	0.8465	0.9175	0.7602	0.7083	0.7332	0.6268	0.6317
LightGBM (SMOTE)	578	270	77	1425	0.8541	0.9162	0.8825	0.6818	0.7689	0.6604	0.6717
GBC (SMOTE)	585	290	70	1405	0.8447	0.9148	0.8931	0.6686	0.7658	0.6527	0.6703
ET	555	280	100	1415	0.8362	0.9113	0.8471	0.6650	0.7435	0.6257	0.6371
ET (SMOTE)	538	255	117	1440	0.8401	0.9102	0.8213	0.6783	0.7432	0.6282	0.6346

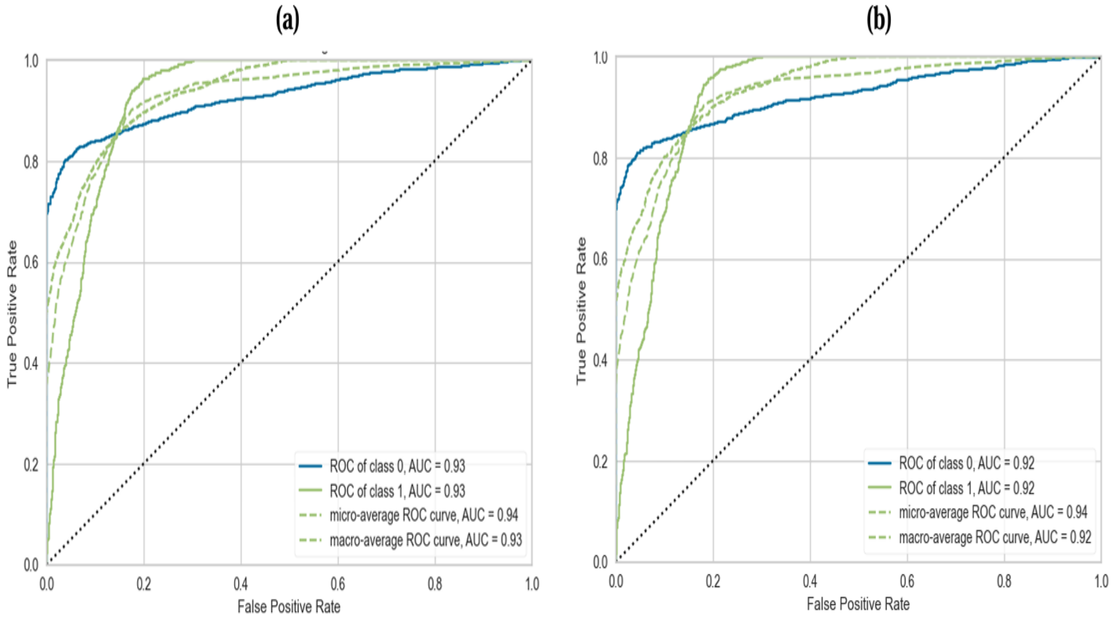

Figure 5: ROC curves for the stacking classification and the ROC for the vote-based classification.

Table 3: Evaluation of Data Result

Model	TP	FP	FN	TN	Acc.	AUROC	Recall	Prec.	F1	Kappa	MCC
Voting Classifier	605	285	50	1410	0.8572	0.9251	0.9237	0.6798	0.7832	0.6764	0.6951
Stacking Classifier	552	250	103	1445	0.8490	0.9210	0.8429	0.6883	0.7572	0.6507	0.6602

There are a number of limitations and possible biases in the MIMIC III dataset that make it difficult to draw broad conclusions. The dataset may not accurately represent the variety of patient demographics and healthcare procedures in various contexts since, first, it is mostly composed of patients from one institution. The results may also not be applicable to modern medicine as the dataset only included patients hospitalised between 2001 and 2012.Accurately correlating patient records may also be hindered by privacy concerns and data anonymisation techniques, which restrict the results' generalisability.

Researchers and practitioners should use caution when applying the MIMIC III dataset's results to larger populations and situations, despite the dataset's unique insights.The evidence base and the ability to make educated medical decisions are both necessitated by the need for prospective investigations and randomised controlled trials.For the voting classifier with the greatest AUROC, an interactive web interface was developed using the open-source python software Gradio (Fig. S3(a)). The result is a score that predicts whether a sample is septic or not, as seen in Figure S3(b).To ensure the concept is applicable and can be used to different situations, the web interface will be tested in an Indian NICU.There are significant ethical concerns and possible difficulties in incorporating machine learning algorithms into healthcare decision-making. First and first, protecting the confidentiality of patients' personal information and preventing its exploitation is of the utmost importance.Another issue is that machine learning models aren't always easy to grasp and interpret because to their complicated algorithms, which might make it difficult for doctors to use them.

have faith in the way they make decisions. Unchecked bias in data used to train algorithms has the potential to exacerbate healthcare disparities. Responsible and useful integration of machine learning algorithms into clinical decision making requires careful assessment of these ethical issues and overcoming possible obstacles.

4. CONCLUSION

Preventing complications or deaths in newborns requires prompt action during the early stages of infection, since sepsis is the third leading cause of mortality in this age group. This research used data from the MIMIC III dataset, which includes information on the first 24 hours after admission, to assess the effectiveness of several machine learning algorithms in predicting when newborn sepsis will start.Accuracy, recall, MCC, and F1 score were all top-notch in the voting classifier's performance, which was trained using an ensemble of adaptive boosting, light gradient boosting without SMOTE, and random forest with SMOTE. During newborn admission, the online interface that was built as a consequence of this research may be used to evaluate the risk of sepsis. Neonatal intensive care unit (NICU) practitioners may use the tool's data to better treat their patients. Achieving SDG 3.2 and lowering infant death rates are greatly affected by the early identification of neonatal sepsis. Optimal resource allocation, increased survival rates, and the avoidance of long-term problems are all outcomes of early detection and rapid treatment. Our long-term goal is to investigate how the model's findings may influence infant care and clinical decision-making in Indian NICUs.

REFERENCES

[1] Pamulaparthyvenkata, S., Sharma, J., Dattangire, R., Vishwanath, M., Mulukuntla, S., Preethi, P., & Indhumathi, N. (2024, June). Deep Learning and EHR-Driven Image Processing Framework for Lung Infection Detection in Healthcare Applications. In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE.

[2] Raj, R. R. M., Saravanan, T., Preethi, P., & Ezhilarasi, I. (2022). Comparative evaluation of efficacy of therapeutic ultrasound and phonophoresis in myofascial pain dysfunction syndrome. Journal of Indian Academy of Oral Medicine and Radiology, 34(3), 242-245.

[3] B. A. Shah, J. F. Padbury, Neonatal sepsis, Virulence, 2014, 5, 170-178, doi: 10.4161/viru.26906.

[4] A. L. Shane, P. J. Sánchez, B. J. Stoll, Neonatal sepsis, The Lancet, 2017, 390, 1770-1780, doi: 10.1016/s0140- 6736(17)31002-4.

[5] J. R. Moorman, W. A. Carlo, J. Kattwinkel, R. L. Schelonka, P. J. Porcelli, C. T. Navarrete, E. Bancalari, J. L. Aschner, M. W. Walker, J. A. Perez, C. Palmer, G. J. Stukenborg, D. E. Lake, T. M. O'Shea, Mortality reduction by heart rate characteristic monitoring in very low birth weight neonates: A randomized trial, The Journal of Pediatrics, 2011, 159, 900-906, doi: 10.1016/j.jpeds.2011.06.044.

[6] Vivek Yadav. (2021). AI and Economics of Mental Health: Analyzing how AI can be used to improve the cost-effectiveness of mental health treatments and interventions. Journal of Scientific and Engineering Research, 8(7), 274–284. https://doi.org/10.5281/zenodo.13600238.

[7] H. Khazaei, N. Mench-Bressan, C. McGregor, J. E. Pugh, Health informatics for neonatal intensive care units: an analytical modeling perspective, IEEE Journal of Translational Engineering in Health and Medicine, 2015, 3, 1-9, doi: 10.1109/JTEHM.2015.2485268.

[8] K. D. Fairchild, R. L. Schelonka, D. A. Kaufman, W. A. Carlo, J. Kattwinkel, P. J. Porcelli, C. T. Navarrete, E. Bancalari, J. L. Aschner, M. W. Walker, J. A. Perez, C. Palmer, D. E. Lake, T. M. O'Shea, J. R. Moorman, Septicemia mortality reduction in neonates in a heart rate characteristics monitoring trial, Pediatric Research, 2013, 74, 570-575, doi: 10.1038/pr.2013.136.

[9] Yadav, V. (2022). AI-Driven Predictive Models for Healthcare Supply Chains: Developing AI Models to Predict and Optimize Healthcare Supply Chains, especially during Global Health Emergencies. Progress in Medical Sciences, 6(1), 1–8. https://doi.org/10.47363/pms/2022(6)211.

[10] [10] A. Johnson, T. Pollard, R. Mark, MIMIC-III Clinical Database (version 1.4), PhysioNet, 2016, doi: 10.13026/C2XW26.

[11] A. E. W. Johnson, T. J. Pollard, L. Shen, L.-W H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, R. G. Mark, MIMIC-III, a freely accessible critical care database, Scientific Data, 2016, 3, 160035, doi: 10.1038/sdata.2016.35.

[12] Korada, L., Sikha, V. K., & Somepalli, S. (2024). Digital transformation balanced with sustainability goals. Journal of Engineering and Applied Sciences Technology, 6(5), 2–7. https://doi.org/10.47363/JEAST/2024(6)E106.

[13] K. E. Henry, D. N. Hager, P. J. Pronovost, S. Saria, A targeted real-time early warning score (TREWScore) for septic shock, Science Translational Medicine, 2015, 7, eaab3719, doi: 10.1126/scitranslmed.aab3719.

[14] T. Desautels, J. Calvert, J. Hoffman, M. Jay, Y. Kerem, L. Shieh, D. Shimabukuro, U. Chettipally, M. D. Feldman, C. Barton, D. J. Wales, R. Das, Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach, JMIR Medical Informatics, 2016, 4, e28, doi: 10.2196/medinform.5909.

[15] Sikha, V. K. (2019). Affordable incident response using cloud-based open-source data pipelines with integrated threat intelligence platforms. International Journal of Intelligent Systems and Applications in Engineering, 7(4).