Engineered Science
DOI: https://dx.doi.org/10.30919/es1476

Federated Learning for Early Detection of Neonatal Sepsis with Enhanced Data Interoperability

Roshan David Jathanna,[1] Abhilash C B,[2] Dinesh Acharya U.,[1,*] Leslie Edward Lewis[3] and Krishnamoorthi Makkithaya[1]

Abstract

Healthcare systems are increasingly dependent on data-driven approaches in patient care. However, heterogeneity and fragmentation of healthcare data cause problems for interoperability and seamless data integration across various healthcare sources. This study uses a synthetic dataset of late-onset neonatal sepsis and a public dataset of early-onset neonatal sepsis to discuss the potential influence of federated learning on healthcare interoperability. Federated learning provides a decentralized approach for training machine learning models across multiple institutions while maintaining the privacy and security of the data. This study demonstrates that integrating disparate healthcare sources using federated learning resulted in an accuracy of 65% for the late-onset neonatal sepsis dataset and 92% for the early-onset neonatal sepsis dataset, compared to 67% and 94% achieved without federated learning. Our findings will, therefore, be instrumental in deepening the understanding of the potential of federated learning to solve the interoperability problems of healthcare systems, paving the way for more efficient and effective data-driven solutions for healthcare.

Keywords: Federated learning; Neonatal sepsis; Healthcare data; Healthcare interoperability; Machine learning; Neonatal healthcare; Data privacy; Data security.
Received: 17 November 2024; Revised: 21 March 2025; Accepted: 29 March 2025.
Article type: Research article.

1. Introduction

A baby in its first 28 days of life is called a neonate. According to a World Health Organization (WHO) report of 2024, out of the 5 million under-five-year-old deaths in the year 2022, 2.3 million were neonates. The primary causes of neonatal death are prematurity, birth asphyxia, and sepsis.[1] Based on the timing of the infection, neonatal sepsis is termed early-onset or late-onset neonatal sepsis.[2]

Neonatal sepsis, particularly gram-negative late-onset sepsis, remains a significant challenge in tertiary care settings. Effective management of this condition requires a thorough understanding of bacterial profiles and drug susceptibility patterns to guide the formulation of effective prevention and treatment methods.[3] Blood culture is the gold standard for identifying neonatal sepsis, but it usually takes 24-48 hours for microbial growth, which can delay critical intervention. Studies have explored alternative approaches, including physiological parameters such as temperature instability, heart rate variability, and respiratory distress, combined with other biomarkers to predict sepsis much before clinical evidence.[2, 4-16]

Machine learning involves the development of algorithms and models that can learn from data to make predictions or informed decisions.[17] It is a collection of techniques concerned with training models on datasets such that the patterns are generalized to yield accurate results. It has been observed that machine learning classifiers like extreme gradient boosting have high accuracy in predicting late-onset neonatal sepsis (LONS) even up to 6 hours before onset, with an area under the receiver operating characteristic curve (AUC) of 0.88. Besides improving the performance of the prediction, this approach provides significant insight into how various physiological signals impact the algorithm's decision-making process, improving clinical decision support.[18,19]

[1] *Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, 576104, India*
[2] *Department of Computer Science and Engineering, JSS Academy of Technical Education, Bengaluru, 560060, India*
[3] *Department of Paediatrics, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, 576104, India*
**Email:* dinesh.acharya@manipal.edu (D. Acharya)

The success of any data mining algorithm depends on the size of the data used for mining. Since the number of cases in the neonatal intensive care unit (NICU) with Sepsis is small, there is a need for collaborative research with other NICUs. Fast Healthcare Interoperability Resources (FHIR) is a popular standard for exchanging digital health records. FHIR uses the hypertext transfer protocol (HTTP)-based representational state transfer (REST) architecture, with resources as the main entities of information exchange. Each resource is an entity that has a clear boundary and a unique identifier. Resources will also have a version that changes whenever the structure of the resource changes. Resources include patients, practitioners, medications, diagnostics, *etc.*,[20] Kasthurirathne *et al*.[21] built an add-on module that allowed users to access open medical record system (OpenMRS) data regarding FHIR resources. They also demonstrated the ease of FHIR implementation for exchanging data between disparate healthcare applications.

Another essential aspect that must be addressed in interoperability is ensuring that the patient's personal and sensitive data does not flow outside the hospital without the patient's consent. Personal data involves names, addresses, or any other data that can be used to identify a patient. Sensitive data involves monetary, physical, psychological, and mental health information, passwords, health history, or biometrics. The most widely used approach for privacy preservation is patient de-identification. The HIPAA (Health Insurance Portability and Accountability Act of 1996) privacy policy mentions two de-identification methods: safe harbor or expert. In the safe harbour method, specific identifiers of the person and details about the employer, family, and household members are removed.[22]

Federated learning is a decentralized machine learning framework wherein multiple devices or institutions collaboratively build a global model without sharing individual datasets. This approach efficiently addresses the growing concerns of data security and privacy in the traditional machine learning procedures, which are based on centralized data for training. Local training is done across various healthcare centres, and only the model weights are updated to the central server, thus preserving patient confidentiality and restricting the possibility of data leaks. This new paradigm enables collaboration and knowledge sharing across healthcare systems, expediting research and innovation in analytics for healthcare and personalized medicine.[23,24]

Federated learning has the potential to revolutionize many critical domains that use edge devices, including finance, transportation, healthcare, and many more. The most prominent applications are the Internet of Things, wearable gadgets, autonomous cars, sensors in smart homes, e-commerce, and healthcare. While it offers more benefits than traditional Machine Learning approaches, like real-time continual learning, data security, and diversity, it also faces a few challenges, which can be categorized into training-related and security-related. Training challenges include improving communication efficiency throughout the network and managing the heterogeneity of devices and data. Security-related challenges include privacy concerns and data protection from attacks.[25-31]

Data privacy and security in the healthcare domain are critical, and the most common applications of federated learning. While centralized training methods are traditionally used to secure patient data, federated learning offers a privacy-preserving alternative that enhances data interoperability and facilitates seamless data exchange between nodes or systems. There is variation in performance, with centralized models exhibiting increased overfitting risks, while federated learning demonstrates enhanced predictive accuracy and robustness by leveraging decentralized and heterogeneous data sources without breaching data privacy.[24]

The changing healthcare landscape with newer paradigms, such as federated learning and patchwork learning, has overcome fundamental healthcare interoperability challenges. Federated learning helps train deep models with distributed healthcare data and ensures data privacy and security, achieving performance compared to traditional methods. On the other hand, patchwork learning is a way to stitch together information across disparate datasets at secure sites, enabling the private simultaneous use of complementary data sources and, thus, genuinely generalizing machine learning models for clinical applications. Both approaches stress the critical idea of seamlessly integrating data from heterogeneous sources to improve patient care, population health, and healthcare workflows with more holistic and generalized models in healthcare applications.[31,32]

The Flower framework is designed to facilitate federated learning. The framework is flexible and versatile as it is agnostic to the underlying machine learning library, such as TensorFlow or PyTorch. This compatibility ensures that researchers can implement federated learning in their projects with minimal changes. Flower emphasizes scalability and ease of use by offering features that simplify the deployment and management of federated learning tasks across diverse and distributed environments.[33]

Neonatal sepsis prediction generally employs small, localized datasets, compromising generalizability. This study addresses the gap by using federated learning to achieve data interoperability for predicting neonatal sepsis, thus enhancing collaborative learning results.

In this study, we have used two datasets to demonstrate the federated learning approach for healthcare interoperability. The outcomes of federated learning are also compared with the traditional method from the recent state-of-the-art. Overall, our contributions are as follows:

- A practical approach for healthcare interoperability using federated learning is proposed.
- A synthetic data generation process and procedure for late-onset sepsis is proposed.
- Evaluated the feasibility of seamlessly integrating generated data with data from literature.
- An effective evaluation method that involves domain experts in assessing the outcomes is proposed.

This study aims to contribute to advancing healthcare data analytics and enhance patient care concerning early and late-onset neonatal sepsis using the federated learning approach. The paper is structured as follows: Section 2 discusses the process and procedure for generating synthetic data. Section 3 discusses the methodology for achieving interoperability using a federated approach. Section 4 presents the results and discussion concerning the implementation. The results are compared with the recent state-of-the-art and discussed accordingly. Finally, the paper concludes in Section 5.

2. Data description

In this study, we have used two sets of data to demonstrate the federated learning approach for healthcare interoperability. The outcomes of federated learning are also compared with the traditional method from the recent state-of-the-art.

2.1 Lopez dataset

The first dataset used in our experiment was initially collected from neonates at a private institution in Cartagena, Colombia, between 2016 and 2017 by López-Martínez *et al*.[34] to predict early-onset neonatal sepsis (EONS) using artificial neural networks (ANN). This dataset comprises 46 independent variables alongside a diagnosis of neonatal sepsis determined by clinical criteria and laboratory blood culture results. Alvi *et al*.[35] conducted subsequent research using the same dataset. However, they narrowed the features to 27 independent variables, demonstrating the highest significance in univariate analysis, employing the chi-square test for variable selection. The Lopez dataset encompasses attributes crucial for the early identification of neonatal sepsis. These attributes include:

- Sociodemographic Variables: Age, teen mother status, health regimen, origin (rural/urban), marital status, and level of education.
- Neonatal Variables: Weight in grams, APGAR (Appearance, Pulse, Grimace, Activity, and Respiration) score after 1 minute, less than 1500 g, less than 2500 g, prematurity, respiratory distress, and gender.
- Obstetric Variables: Premature rupture of membranes (>18 hours and >6 hours), number of pregnancies, number of births, number of C-sections, type of birth (vaginal/C-section), assistance for prenatal control, and gestational age at birth.
- Maternal Infectious Pathology: Maternal fever, yeast infections, urinary tract infections, and sexually transmitted disease history.

2.2 *GenSynData* dataset

Given the scarcity of publicly available neonatal datasets, the second dataset used in our study was generated synthetically. Fig. 1 represents the method for synthetic data generation incorporating key attributes about LONS obtained from the literature. The dataset generated will be referred to as *GenSynData* throughout the rest of this chapter.

2.2.1 Process of attribute selection

Attribute selection was performed after an exhaustive literature review on LONS risk factors and diagnostic markers. Additionally, domain experts, such as senior neonatologists, were consulted to ensure that all clinically important features were included. Ranges and distributions of each attribute to be included were determined after consulting domain experts. The medical experts verified and certified attribute ranges based on clinical guidelines and experience. The description of attributes generated is depicted in Table 1. For instance, attributes like vital signs, laboratory tests, and demographic data were given ranges based on clinical practice and available medical literature. Iterative assessments were conducted to optimize data ranges and distributions so that the synthetic data would closely resemble real-world scenarios encountered in neonatal care units.

2.2.2 Exploratory data analysis

Exploratory data analysis (EDA) on the *GenSynData* provided several insights regarding the relationship between various features and the outcome variable. Summary statistics of two medical scoring systems, sequential organ failure assessment (SOFA) and systemic inflammatory response syndrome (SIRS), are shown in Table 2.

Fig. 2 depicts the distributions and central tendencies of critical features such as temperature and heart rate. The pairplot plot shown in Fig. 3 revealed distinct patterns in feature distributions concerning the outcome variable, indicating potential predictive relationships. Additionally,

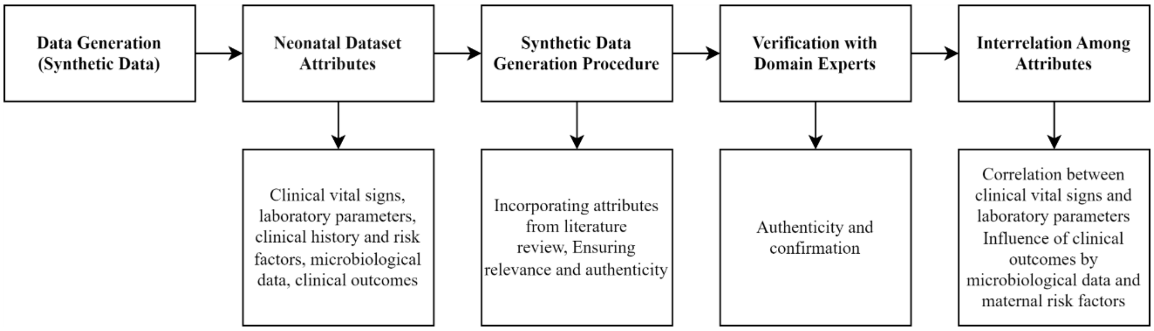


Fig. 1: Data generation methodology for *GenSynData*.

Table 1: Description of attributes generated for LONS.

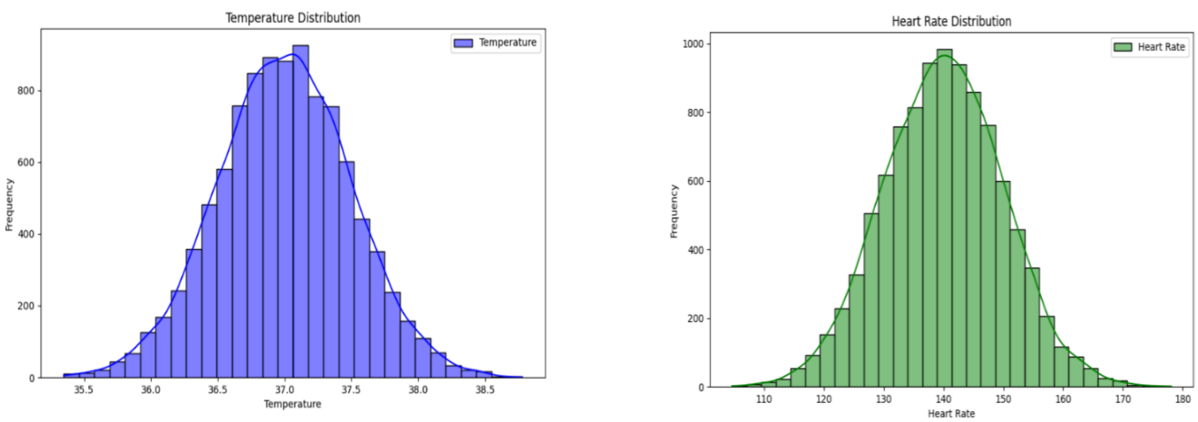
Attribute Name	Range	Description
Temperature	Normal distribution with mean 37 °C and standard deviation 0.5 °C	Body temperature of newborns with slight variation
Heart rate	Normal distribution with mean 140 bpm and standard deviation 10 bpm	Heart rate of newborns with slight variation
Respiratory rate	Normal distribution with mean 40 breaths/min and standard deviation 5 breaths/min	Respiratory rate of newborns with slight variation
Blood pressure	Random integers between 50 mmHg and 80 mmHg	Blood pressure range for newborns
White blood cell (WBC) count	Random integers between 5000/mm³ and 15000/mm³	White blood cell count range
I/T ratio	Uniform distribution between 0.2 and 0.8	Immature to total neutrophil ratio
C-reactive protein (CRP) level	Uniform distribution between 0 and 10 mg/L	C-reactive protein level
Procalcitonin level	Uniform distribution between 0 and 1 ng/mL	Procalcitonin level

Table 2: Summary statistics on the *GenSynData*

#	SOFA score	SIRS score
Count	10000.00	10000.00
Mean	7.521	1.983
Std	4.617	1.414
Min	0.00	0.00
25%	3.00	1.00
50%	7.50	2.00
75%	12.00	3.00
Max	15.00	4.00

boxplots illustrated the variation in scores, such as the Neonatal SOFA Score, across different outcome categories. Overall, the EDA provided valuable insights into the dataset's characteristics and potential predictive factors associated with the variable, informing further analysis and model development for predicting clinical outcomes in LONS.

Fig. 4 illustrates the correlation heatmap of critical physiological and laboratory parameters related to neonatal sepsis. The findings indicate that clinical individual characteristics are marked by low direct correlations, highlighting the multifactorial pathophysiological nature of



(a) (b)

Fig. 2: Distribution of *GenSynData* (a) temperature and (b) heart rate.

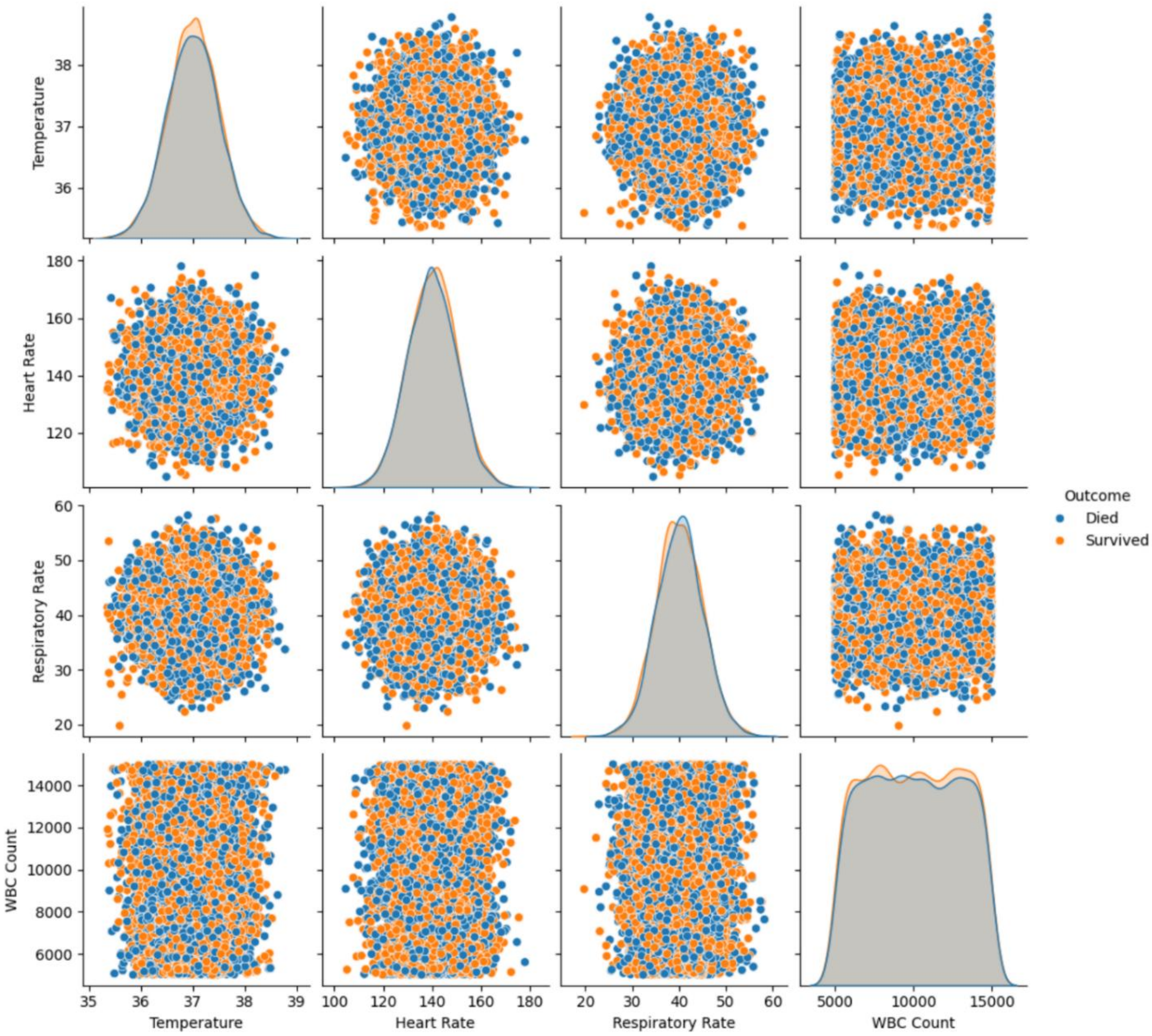


Fig. 3: Pairplot visualization of vital signs for neonatal sepsis prediction.

neonatal sepsis. It was observed that inflammatory markers (CRP, Procalcitonin, I/T Ratio) are weakly correlated with vital signs, highlighting their complementary role in clinical diagnostics. The correlations we observed indicate that the development of neonatal sepsis is driven by complex interactions, but not by one dominant factor.

The inferences drawn from these observations influenced our strategy in model development and prompted us to use machine learning techniques that could detect non-linear relationships and multi-variable interactions, and thus improve prediction accuracy in sepsis detection at an early stage. Domain experts validated the generated synthetic data, and exploratory data analysis was conducted for second-level conformance.

3. Methods

3.1 Centralized machine learning model

The ANN model, initially developed by Alvi *et al.*,[35] was implemented and trained using the PyTorch deep learning library for the benchmark centralized machine learning model. The training process was conducted within Google Colab, a cloud-based interactive development environment that offers free access to computational resources, including graphics processing units (GPUs) and tensor processing units (TPUs).

3.2 Federated learning framework

Federated learning is a pioneering machine learning approach, enabling collaborative model training across decentralized data sources without centralized data aggregation.[36] We explored the application of federated learning in the healthcare domain, focusing on analyzing patient demographic data across multiple healthcare institutions. A comprehensive pipeline was devised for federated learning implementation by leveraging Python libraries such as pandas and scikit-learn, as shown in Fig. 5.

The first step of our approach was rigorous preprocessing

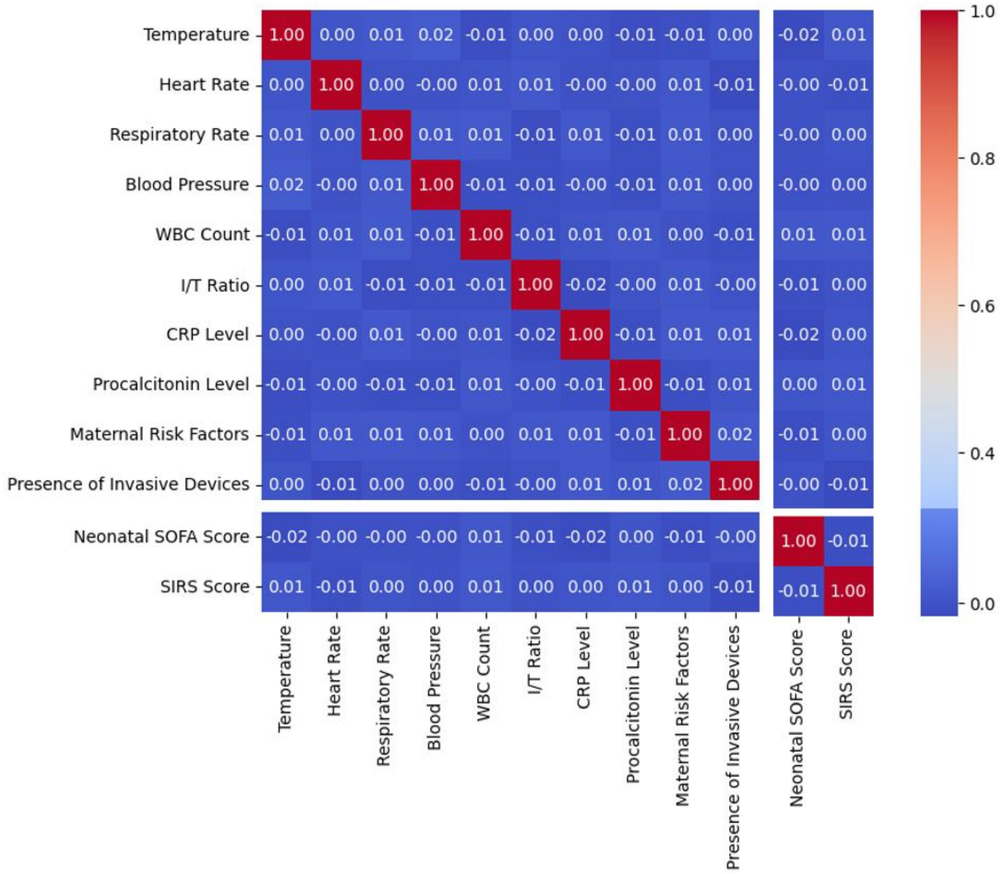


Fig. 4: Feature correlation heatmap for neonatal sepsis.

of the data. We read a synthetically generated dataset using the pandas library. It included patients with different demographics to integrate the categorical variables as smoothly as possible using encoding methods provided by LabelEncoder from the scikit-learn library. It gave a standard format for all data sources while preserving privacy and integrity.

3.2.1 Algorithms

This section proposes an effective algorithm with a procedure for federated learning to achieve interoperability in the healthcare domain. Algorithm 1 outlines the federated learning process for attaining interoperability and accurate predictions across heterogeneous healthcare data sources. The main steps are data preprocessing (outlined in Procedure 2), local model training (outlined in Procedure 3), aggregation of the trained models (explained in Procedure 4), and result evaluation based on interoperability and accuracy.

Each step of the algorithm cumulatively enables the achievement of interoperability and accurate predictions with decentralized data from heterogeneous sources. The data preprocessing step normalizes and cleanses the data for successful training, and the local model training recognizes unique patterns from every data source without exposing sensitive data. The aggregation step consolidates these heterogeneous bits of information into a unified global model, thus enhancing its generalizability. Finally, interoperability testing and outcome analysis validate the model's smooth functionality across different systems and establish its effectiveness in making accurate predictions in healthcare applications. The following procedures outline the steps involved in the proposed algorithms:

(1) Procedure 1 (Preprocessing): Facilitates consistent data quality across different sources, which is a prerequisite for interoperability. Also, it eliminates learning biases by standardizing feature levels and handling missing values.

(2) Procedure 2 (Federated model): Local models are trained on decentralized data sources without revealing sensitive data, maintaining privacy and data compliance (*e.g.*, Health Insurance Portability and Accountability Act (HIPAA) or general data protection regulation (GDPR)). Each operation (*e.g.*, gradient descent) ensures that the local models are adapted to the specific features of their own data.

(3) Procedure 3 (Model aggregation): Aggregates the output of local models into a single global model, allowing it to generalize well across all data sources. This approach ensures interoperability by simultaneously aggregating knowledge from heterogeneous healthcare systems.

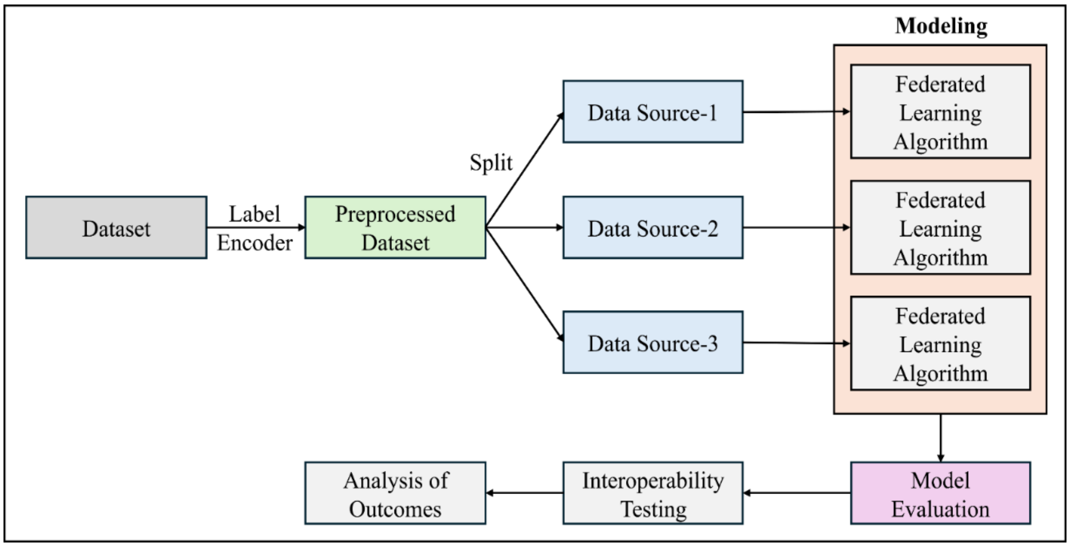


Fig. 5: Proposed federated learning methodology for interoperability.

Algorithm 1 Federated Learning for Healthcare Interoperability

Input: Dataset D, number of data sources n, Federated Learning Algorithm $\mathcal{F}$
Preprocess the dataset D using PREPROCESSINGPROCEDURE(D)
Split the dataset into n subsets: $\{D_1, D_2, \ldots, D_n\}$
for $i = 1$ to n **do**
    Train local model $\mathcal{M}_i$ on data source D_i using FEDERATEDMODEL($\mathcal{M}_i$, D_i)
end for
Aggregate local models $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_n$ using the MODELAGGREGATION function
Evaluate the global model $\mathcal{M}_{global}$ using MODELEVALUATION on test data
Test model interoperability using INTEROPERABILITYTESTING()
Analyze outcomes using OUTCOMEANALYSIS()

Procedure 1 PreprocessingProcedure

procedure PREPROCESSINGPROCEDURE(D)
    Apply label encoding to categorical features of D
    Normalize continuous features to a common scale
    Impute missing values using mean imputation for continuous data and mode imputation for categorical data
    Return the preprocessed dataset $\tilde{D}$
end procedure

Procedure 2 FederatedModel

procedure FEDERATEDMODEL($\mathcal{M}_i$, D_i)
    Initialize model parameters θ_i for local model $\mathcal{M}_i$
    for each local epoch $t = 1$ to T **do**
        Update model parameters: $\theta_i \leftarrow \theta_i - \eta \nabla_{\theta_i} \mathcal{L}(\mathcal{M}_i, D_i)$ ▷ Use gradient descent to minimize loss $\mathcal{L}$
    end for
    Send model updates θ_i to the central server for aggregation
    Return local model $\mathcal{M}_i$
end procedure

Procedure 3 ModelAggregation

procedure MODELAGGREGATION($\theta_1, \theta_2, \ldots, \theta_n$)
    Initialize global model parameters θ_{global}
    Aggregate updates:

$$\theta_{global} \leftarrow \frac{1}{n} \sum_{i=1}^{n} \theta_i$$

    Return global model parameters θ_{global}
end procedure

Several sources were simulated such that heterogeneity was reflected as is usually present in healthcare ecosystems. Then, each source simulated one or another healthcare institution. Partitioning of data samples was realized accordingly. Balanced splitting of a dataset into a training set and a test set was realized using the train_test_split function for robust model evaluation.

In our study, we trained a local model based on locally available data subsets for each data source. Therefore, the data was in a sovereign and private status. These models were then aggregated to provide their collective performance and interoperability. A comprehensive evaluation framework was developed that additionally included accuracy scores for the performance of each of the models. We quantified to what degree good models trained from a single data source using the sci-kit-learn accuracy_score function. Further, we investigated interoperability testing that rates the accuracy of predictions between the combined models when trained from different data sources.

The work was implemented in Python, using essential libraries like pandas, scikit-learn, PyTorch, and Flower. Individual data clusters were formed to imitate different health institutions or data sources. In this case, such clustering lets the data remain decentralized, close to real scenarios where data are siloed across various entities. Then, interoperability was implemented in a Flower framework for collaborative model training and evaluation across these disparate data clusters. The Flower framework abstracts the complexities involved in client-server communication, model weight aggregation, and the orchestration of training and evaluation phases. The reason for using Flower was its flexibility, ease of use, and compatibility with PyTorch, allowing us to focus on model design and training logic.

Data privacy, security, and integrity were stressed during the implementation phase. Strong encryption techniques were applied to secure the sensitive patient information used in training and model evaluation. More importantly, strict access controls and authentication mechanisms were in place to ensure that only authorized personnel had access to the data and model outputs.

3.2.2 Model architecture

In this study, the ANN model used is a fully connected neural network with four layers. The first three layers are dense layers equipped with ReLU (Rectified Linear Unit) activation functions to introduce non-linearity, followed by a final layer using SoftMax activation to produce probability distributions over the two classes. The architecture is as follows:

- An input layer that accepts 27 features.
- Three hidden layers with 550, 300, and 120 neurons, respectively.
- A final output layer with two neurons representing the class probabilities.

3.2.3 Model training procedure

Training was conducted on each client’s local dataset using stochastic gradient descent (SGD) as depicted in equation 1 with a learning rate of 0.01, momentum of 0.9, and a weight decay of 0.0001 to minimize overfitting. The negative log likelihood loss (NLLLoss) was used in conjunction with the log softmax output of the model, providing a stable training procedure in Eq. (1):

$$L(\theta) = \frac{1}{N}\sum_{i=1}^{N} l\big(y_i, f(X_i;\theta)\big) \tag{1}$$

where $L(\theta)$ is the loss function (*e.g.*, cross-entropy for classification), N is the number of training samples, y_i is the true label, $f(X_i;\theta)$ is the model prediction, and θ represents model parameters.

Each client updates θ using the gradient descent step as depicted in Eq. (2).

$$\theta_{t+1} = \theta_t - \eta\nabla L(\theta_t) \tag{2}$$

where η is the learning rate.

We used the Flower’s federated average (FedAvg) strategy as depicted in Eq. (3), which aggregates model updates from all clients by averaging their weights, ensuring that each client contributes equally to the global model, regardless of volume or diversity.

$$\theta = \sum_{i=1}^{K} \frac{n_i}{N}\theta_i \tag{3}$$

where K is the total number of clients, n_i is the number of samples at client i, and $N = \sum n_i$ is the total dataset size across all clients.

3.2.4 Evaluation and metrics

After the training rounds, the global model’s performance was evaluated on the unseen test dataset to assess its generalization ability. Metrics such as accuracy and loss were computed to evaluate the model's effectiveness in classifying new examples, as shown in Eqs. (4) and (5), respectively. Additionally, we explored the impact of federated learning on model performance by comparing it to a centrally trained equivalent, highlighting the benefits and challenges of adopting a federated learning approach.

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}} \tag{4}$$

$$\text{Loss} = -\sum y \log(\hat{y}) \tag{5}$$

where y is the ground truth and $\hat{y}$ is the predicted probability.

4. Results and discussion

The experiments with federated learning demonstrated promising results in both model accuracy and interoperability. Specifically, the study achieved high accuracy in predicting LONS using aggregated models trained on individual data clusters. Besides, the interoperability framework collaborates seamlessly between different healthcare institutions to share insights and best practices while showing respect for data sovereignty and privacy regulations.

In Figs. 6 and 7, the X-axis represents the number of training epochs completed during the training process on the *GenSynData* and Lopez datasets, respectively. The Y-axis represents the model's performance, plotting accuracy, and loss values for the training and validation datasets. Figs. 6 and 7 suggest that the model trained using federated learning exhibits variability in performance across rounds, with the centralized approach generally marginally outperforming the distributed approach in terms of loss and accuracy. Fig. 8 represents the test accuracy and loss trends over training rounds with varying numbers of clients (2, 3, 4, and 5). The results indicate that increasing the number of clients enhances generalization. However, optimal client selection is essential for ensuring stable convergence. These findings emphasize the importance of regularization techniques and strategic client distribution to improve model robustness and overall performance. Overall, the gap between the training and validation accuracy curves indicates overfitting, where the model performs well on the training data but is less expected to generalize to new data. On the other hand, if both curves are close and increasing together, the model learns effectively from the training data and generalizes well to unseen data.

Referring to Table 3, the evaluation of interoperability accuracy using federated learning, we observed varying results across different datasets. For *GenSynData*, the accuracy achieved with federated learning was 0.65, slightly lower than the accuracy of 0.67 achieved without using federated learning. Similarly, for the Lopez dataset, the accuracy with and without federated learning was 0.92 and 0.94, respectively. These results suggest that the effectiveness of improving interoperability accuracy may depend on the dataset's characteristics, with different datasets exhibiting different performance outcomes.

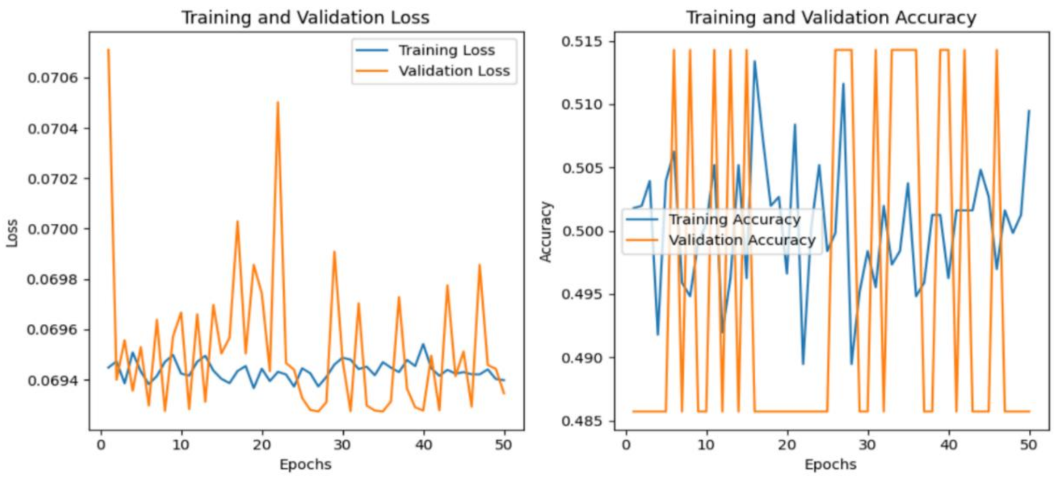


Fig. 6: Training and validation loss and accuracy of *GensynData*.

Fig. 7: Training and validation loss and accuracy of the Lopez dataset.

Table 3: Comparison of accuracy with and without federated learning for interoperability.

Dataset	Accuracy (with federated)	Accuracy (without federated)
Lopez dataset[34]	0.92	0.94
GenSynData	0.65	0.67

4.1 Global interoperable testing

The results obtained for the accuracy of the three local models and the global model are shown in Table 4. At first glance, it can be clearly observed that the accuracy of the global model ranged between 61.3% and 87.7%, which is almost similar to that of the centralized local models. However, to find the statistical significance of the variances in predictive accuracy of the global model with that of the local models, analysis of variance (ANOVA) was performed, and the results are shown in Table 5. Overall, there is no significant difference in the predictive accuracies based on the four models under comparison ($F = 0.01$, $p = 0.998$). Moreover, the S (standard deviation of the residuals) equals 10.67, indicating that predictions may vary by this amount.

Table 4. Accuracies of the global interoperable model vs local models

Rounds	Local model 1	Local model 2	Local model 3	Global model
1	62.1	60.5	61.2	61.3
2	68.4	66.9	67.1	67.5
3	75.2	73.8	74.3	74.5
4	82.7	81.5	81.9	82.0
5	88.3	87.2	87.6	87.7

The R-square value as low as 0.26% indicates that only this much variance in accuracy can be explained by differences in the models (centralized vs. federated learning models). Therefore, the choice of the model selected has no difference in terms of accuracy variances. This point is further strengthened by the R-square (predicted) = 0.001%, indicating that adjusting the number of predictors does not explain the variation in accuracy. To summarize, there is no statistically significant difference in performance between the global model and the local models. This statistical analysis confirms that the global federated model can predict as accurately as the centralized NICU models.

Table 5: The ANOVA for the overall comparison

Source	Degrees of freedom (DF)	Sum of squares (SS)	Mean square (MS)	F-statistic	p-value
Models	3	5	2	0.01	0.998
Error	16	1821	114		
Total	19	1825			
Standard error of the regression (S) = 10.67 R-squared = 0.26% R-squared (adj) = 0.001%					

Further, it was necessary to compare the accuracies of the global model individually with the three local models. Thus, the Mann-Whitney U (Wilcoxon rank sum) test was conducted to see if the accuracy varies individually. The results are shown in Table 6. The results indicate no statistically significant difference between the global model and any of the three local models at a 5% significance level (all the p values are greater than 0.05). Thus, the validity of federated learning as an alternative to centralized NICU models (local models) in maintaining similar predictive accuracy while ensuring data privacy is guaranteed empirically.

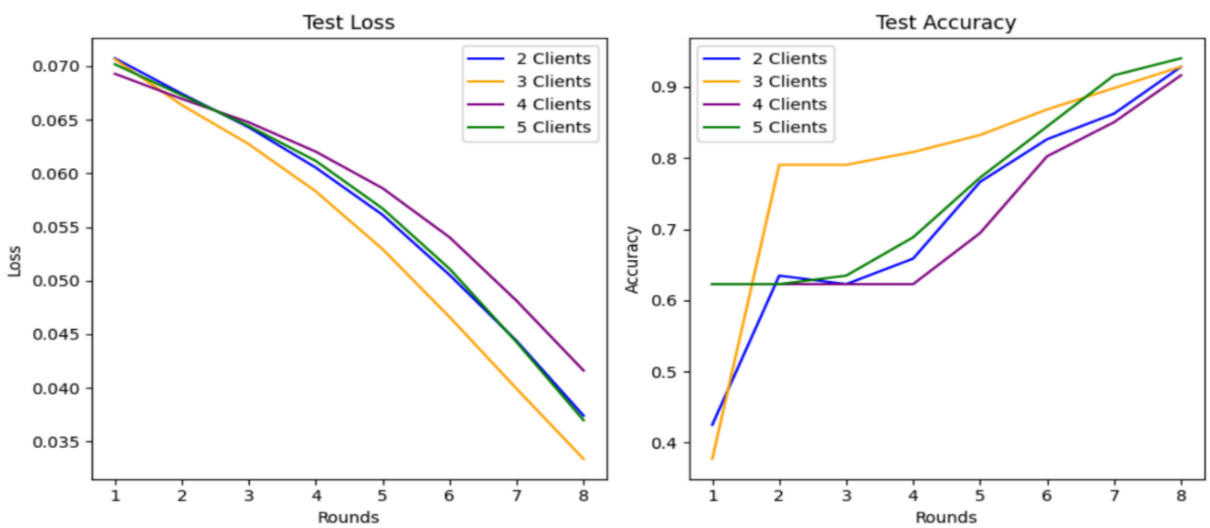


Fig. 8: Test loss and accuracy trends over training rounds with varying numbers of clients (2, 3, 4, and 5).

Table 6: Mann-Whitney U test for individual accuracy comparison.

Comparison	U-statistic	p-value
Global model vs local model 1	7	0.841
Global model vs local model 2	6	0.69
Global model vs local model 3	6	0.69

4.2 Implications and practical considerations for healthcare applications

Federated learning enhances neonatal sepsis research by enabling cooperative model training across heterogeneous healthcare systems without requiring standardized data formats or centralization. By preserving local data privacy, federated learning improves interoperability, manages diverse datasets, and facilitates real-time model updates for enhanced predictive accuracy in NICUs. To mitigate training difficulties in federated learning for neonatal sepsis prediction, federated averaging and adaptive optimization techniques were utilized to address data heterogeneity and hardware differences between institutions. Cross-validation over heterogeneous datasets was performed to avoid risks of performance degradation and provide model robustness and reliability. The global interoperable testing results demonstrate the viability of applying federated learning to predict neonatal sepsis across multiple healthcare institutions accurately. The observed improvements in accuracy, reaching 87.7% in the fifth round, suggest that collective learning from heterogeneous data sources enhances diagnostic precision while maintaining patient privacy.

4.2.1 Clinical implications

The ability to train on distributed datasets ensures that the model learns diverse clinical patterns and, thus, has the potential of making more accurate and timely predictions of neonatal sepsis. This can aid clinicians in decision-making and enhance patient outcomes in critical care settings.

4.2.2 Practical considerations

• Data Privacy and Security: Federated learning decentralizes sensitive patient data but requires robust security protocols for model updates to prevent breaches.

• Interoperability: Real-world deployment involves overcoming heterogeneity in data format and having standardized pre-processing pipelines within institutions.

• Computational Constraints: For smaller healthcare facilities to contribute meaningfully to federated networks, edge computing or lightweight models may be required.

• Clinical Validation: Intensive testing on real-world patient data is critical to evaluate model generalizability and clinical feasibility.

4.3 Addressing overfitting and enhancing generalization

We have employed regularization techniques such as weight decay and dropout in local training and adaptive model aggregation to mitigate overfitting and enhance model generalization. Data augmentation approaches, such as feature perturbations, can be employed to enhance robustness further. Also, early stopping based on validation loss will prevent overfitting in training. Finally, these combined techniques yield a generalized, robust, and reliable model covering various heterogeneous healthcare datasets.

4.4 Analysis of performance differences between federated and non-federated learning

Performance differences between federated and non-federated learning are attributed to differences in the nature of datasets. Considering the case of *GenSynData*, in LONS, the lower accuracy that was observed (0.65 using federated learning versus 0.67 using non-federated) can be attributed to higher data heterogeneity and lower samples per node available, which make it challenging for models to converge. In the case of the Lopez dataset, however, it exhibited better performance (0.92 using federated learning and 0.94 without federated) due to its more homogenous data distributions and a higher sample size that allowed effective training and generalization. These results indicate the need for federated learning to adopt adaptive approaches and strict data pre-processing to reduce variations due to heterogeneous data distributions and feature complexities in healthcare organizations.

5. Conclusion

Utilizing advancements in predictive modelling to enhance clinical decision-making and patient outcomes demonstrates the transformative impact of combining monitoring data with machine learning algorithms, especially in predicting late-onset sepsis in preterm neonates. As we address the hurdles of healthcare data integration and predictive modelling, it becomes evident that adopting innovative strategies like federated learning and patchwork learning offers the potential for revolutionizing healthcare workflows and ultimately enhancing patient care and population health. This research demonstrates the effectiveness of federated learning in neonatal sepsis diagnosis through the overcoming of essential issues of privacy, heterogeneity of data, and model generalization among different healthcare centers. The use of federated averaging with adaptive optimization methods ensured convergence and robustness despite varying data distribution and hardware capabilities.

Our work has shown positive results regarding using federated learning methodologies to improve interoperability and increase the effectiveness of treatments and preventive measures in neonatal sepsis care. Innovative methods, such as federated learning, may help tackle barriers to interoperability in patient care in complex health environment service delivery. Moreover, though the federated approach is slightly less accurate than the conventional distributed means, one must pay attention to the fact that both the models of federated learning and the distributed method have almost similar performances. However, the global interoperable testing accuracy is 88.3%, showing the importance of adopting federated learning in the healthcare domain. Future research can explore innovations in communication protocols and resource-efficient model updates to minimize latency and computational overhead in federated learning designs. Additionally, combining diverse aggregation methods with differential privacy techniques can further enhance model security and robustness against adversarial attacks.

Acknowledgements

We want to thank the early onset neonatal sepsis dataset creators and contributors for their efforts and commitment to making this important resource openly accessible to the scientific community. We also acknowledge the cooperation from our institutions, Manipal Academy of Higher Education, Manipal and JSS Academy of Technical Education Bangalore.

Conflict of Interest

There is no conflict of interest.

Supporting Information

Not applicable.

References

[1] A. Lokangaka, M. Ramani, M. Bauserman, J. Patterson, C. Engmann, A. Tshefu, S. Cousens, S. A. Qazi, A. I. Ayede, E. A. Adejuyigbe, F. Esamai, Incidence of possible serious bacterial infection in young infants in the three high-burden countries of the Democratic Republic of the Congo, Kenya, and Nigeria: A secondary analysis of a large, multi-country, multi-centre clinical trial, *Journal of Global Health*, 2024, **14**, 04009, doi: 10.7189/jogh.14.

[2] R. D. Jathanna, D. Acharya, L. E. Lewis, K. M. Makkithaya, Early detection of late-onset Neonatal Sepsis Using Machine Learning Algorithms, *Engineered Science*, 2023, **26**, 976, doi: 10.30919/es976.

[3] F. Iqbal, A. Barche, P.A. Shenoy, L.E.S. Lewis, J. Purkayastha, K. Vandana, Gram-Negative Colonization and Bacterial Translocation Drive Neonatal Sepsis in the Indian Setting, *Journal of Epidemiology and Global Health*, 2024, **14**, 1-11, doi: 10.1007/s44197-024-00303-8.

[4] R. S. Procianoy, R. C. Silveira, The Challenges of Neonatal Sepsis Management, *Jornal de Pediatria*, 2020, **96**, 80-86, doi: 10.1016/j.jped.2019.10.004.

[5] C. O'Sullivan, D. H. T. Tsai, I. C. Y. Wu, E. Boselli, C. Hughes, D. Padmanabhan, Y. Hsia, Machine learning applications on neonatal sepsis treatment: a scoping review, *BMC Infectious Diseases*, 2023, **23**, 441, doi: 10.1186/s12879-023-08409-3.

[6] M. P. Griffin, D. E. Lake, J. R. Moorman, Heart rate characteristics and laboratory tests in Neonatal sepsis, *Pediatrics*, 2005, **115**, 937-941, doi: 10.1542/peds.2004-1393.

[7] 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.

[8] C. McGregor, C. Catley, A. James, Variability analysis with analytics applied to physiological data streams from the neonatal intensive care unit, 25th IEEE International Symposium on Computer-Based Medical Systems, Italy, 2012, 1-5, doi: 10.1109/CBMS.2012.6266385.

[9] 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.

[10] 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, Septicemia mortality reduction in neonates in a heart rate characteristic monitoring trial, *Pediatric Research*, 2013, **74**, 570-575, doi: 10.1038/pr.2013.136.

[11] L. Mithal, R. Yogev, H. Palac, I. Gur, K. Mestan, Computerized vital signs analysis and late onset infections in extremely low gestational age infants, *Journal of Perinatal Medicine*, 2016, **44**, 491-497, doi: 10.1515/jpm-2015-0362.

[12] B. Y. H. Wee, J. H. Lee, Y. H. Mok, S. L. Chong, A narrative review of heart rate and variability in sepsis, *Annals of Translational Medicine*, 2020, **8**, 768-768, doi: 10.21037/atm-20-148.

[13] V. M. Quinten, M. V. Meurs, M. H. Renes, J. J. Ligtenberg, J. C. Maaten, Protocol of the SEPSIVIT Study: A prospective observational study to determine whether continuous heart rate variability measurement during the first 48 hours of hospitalisation provides an early warning for deterioration in patients presenting with infection, *BMJ Open*, 2017, **7**, 11, doi: 10.1136/bmjopen-2017-018259.

[14] F. J. Bohanon, A. A. Mrazek, M. T. Shabana, S. Mims, G. L. Radhakrishnan, G. C. Kramer, R. S. Radhakrishnan, Heart Rate Variability Analysis is More Sensitive at Identifying Neonatal Sepsis than Conventional Vital Signs, *The American Journal of Surgery*, 2015, **210**, 661-667, doi: 10.1016/j.amjsurg.2015.06.002.

[15] H. J. Koppens, A. Onland, D. H. Visser, N. P. Denswil, A. H. van Kaam, C. A. Lutterman, Heart rate characteristics monitoring for late-onset sepsis in preterm infants: A systematic review,

Neonatology, 2023, **120**, 548-557, doi: 10.1159/000531118.
[16] M. Chiera, F. Cerritelli, A. Casini, N. Barsotti, D. Boschiero, F. Cavigioli, C.G. Corti, A. Manzotti, Heart rate variability in the perinatal period: A critical and conceptual review, *Frontiers in Neuroscience*, 2020, **14**, 1-23, doi: 10.3389/fnins.2020.561186.
[17] V. Aithal, R. D. Jathanna, Credit risk assessment using machine learning techniques, *International Journal of Innovative Technology and Exploring Engineering*, 2019, **9**, 3482-3486, doi: 10.35940/ijitee.A4936.119119.
[18] L. Cabrera-Quiros, D. Kommers, M. K. Wolvers, L. Oosterwijk, N. Arents, J. van der Sluijs-Bens, E. J. Cottaar, P. Andriessen, C. van Pul, Prediction of late-onset sepsis in preterm infants using monitoring signals and machine learning, *Critical Care Explorations*, 2021, **3**, e0302, doi: 10.1097/CCE.0000000000000302.
[19] Z. Peng, G. Varisco, X. Long, R. Liang, D. Kommers, W. Cottaar, P. Andriessen, C. V. Pul, A continuous late-onset sepsis prediction algorithm for preterm infants using multi-channel physiological signals from a patient monitor, *IEEE Journal of Biomedical and Health Informatics*, 2022, **27**, 550-561, doi: 10.1109/JBHI.2022.3216055.
[20] D. Bender, K. Sartipi, HL7 FHIR: An Agile and RESTful Approach to Healthcare Information Exchange, 26th IEEE International Symposium on Computer-Based Medical Systems, Porto, Portugal, June 20-22, 2013, 326-331, doi: 10.1109/CBMS.2013.6627810.
[21] S. N. Kasthurirathne, B. Mamlin, H. Kumara, G. Grieve, P. Biondich, Enabling better interoperability for healthcare: lessons in developing a standards-based application programming interface for electronic medical record systems, *Journal of Medical Systems*, 2015, **39**, 1-8, doi: 10.1007/s10916-015-0356-6.
[22] W. Moore, S. Frye, Review of HIPAA, Part 1: History, Protected Health Information, and Privacy and Security Rules, *Journal of Nuclear Medicine Technology*, 2019, **47**, 269-272, doi: 10.2967/jnmt.119.227819.
[23] T. Li, A. K. Sahu, A. Talwalkar, V. Smith, Federated learning: challenges, methods, and future directions, *IEEE Signal Processing Magazine*, 2020, **37**, 50-60, doi: 10.1109/MSP.2020.2975749.
[24] C. Abhilash, K. Deepak, R. Hegadi, K. Mahesh, Common data interface for sustainable healthcare system, Machine Learning and Deep Learning in Medical Data Analytics and Healthcare Applications, CRC Press, Boca Raton, 2022, 1-22, ISBN: 9781003226147.
[25] C. Campolo, G. Genovese, G. Singh, A. Molinaro, Scalable and interoperable edge-based federated learning in IoT Contexts, *Computer Networks*, 2023, **223**, 109576, doi: 10.1016/j.comnet.2023.109576.
[26] S. Gupta, S. Katiyar, Communication-efficient federated learning in wireless-edge architecture, Federated Learning for IoT Applications, Springer, Cham, 2022, 117-134, ISBN: 9783030855581.
[27] L. Su, V. K. Lau, Distributed edge learning with inter-type and intra-type over-the-air collaboration, ICC 2021 - IEEE International Conference on Communications, Montreal, QC, Canada, 14-23 June, 2021, 1-6, doi: 10.1109/ICC42927.2021.9500991.
[28] T. Zhang, L. Gao, C. He, M. Zhang, B. Krishnamachari, A.S. Avestimehr, Federated Learning for the Internet of Things: Applications, Challenges, and Opportunities, *IEEE Internet of Things Magazine*, 2022, **5**, 24-29, doi: 10.1109/IOTM.004.2100182.
[29] R. Khan, M. Gupta, P. Kumari, N. Kumar, A prospective study of federated machine learning in medical science, Federated Learning for IoT Applications, Springer, Cham, 2022, 105-116, ISBN: 9783030855581.
[30] D. Ilie, Achieving interoperability in a federated environment and in the current security context, Romanian Military Thinking International Scientific Conference Proceedings, Centrul tehnic-editorial al armatei, Bucharest, 2022, 222-235, ISBN: 26687984.
[31] F. Zhang, D. Kreuter, Y. Chen, S. Dittmer, S. Tull, T. Shadbahr, M. Schut, F. Asselbergs, S. Kar, S. Sivapalaratnam, S. Williams, Recent methodological advances in federated learning for healthcare, *Patterns*, 2024, **5**, 6. doi: 10.1016/j.patter.2024.101006.
[32] S. Rajendran, W. Pan, M. R. Sabuncu, Y. Chen, J. Zhou, F. Wang, Patchwork learning: A paradigm towards integrative analysis across diverse biomedical data sources, arXiv preprint arXiv: 2305.06217, 2023, doi: 10.48550/arXiv.2305.06217.
[33] D. J. Beutel, T. Topal, A. Mathur, X. Qiu, J. Fernandez-Marques, Y. Gao, L. Sani, K. Hei Li, T. Parcollet, P. Porto Buarque de Gusmão, N. D. Lane, Flower: a friendly federated learning research framework, arXiv preprint arXiv: 2007.14390, 2022, doi: 10.48550/arXiv.2007.14390.
[34] F. López-Martínez, E. R. Núñez-Valdez, J. L. Gomez, V. García-Díaz, A neural network approach to predict early neonatal sepsis, *Computers & Electrical Engineering*, 2019, **76**, 379-388, doi: 10.1016/j.compeleceng.2019.04.015.
[35] R. H. Alvi, M. H. Rahman, A. A. S. Khan, R. M. Rahman, Predicting Early Neonatal Sepsis Using Neural Networks and Other Classifiers, 2020 IEEE 10th International Conference on Intelligent Systems, Varna, Bulgaria, 28-30 Aug, 2020, 443-450, doi: 10.1109/IS48319.2020.9199937.
[36] D. Roschewitz, M. A. Hartley, L. Corinzia, M. Jaggi, iFedAvg: Interpretable data-interoperability for federated learning, arXiv preprint arXiv: 2107.06580, 2021, doi: 10.48550/arXiv.2107.06580.

Publisher's Note: Engineered Science Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access