Computer Systems Science & Engineering
DOI: 10.32604/csse.2023.025281
Article

Tech Science Press

Explainable AI Enabled Infant Mortality Prediction Based on Neonatal Sepsis

Priti Shaw[1], Kaustubh Pachpor[2] and Suresh Sankaranarayanan[3,*]

[1]Barclays Bank, Bund Garden Road, Pune, 411001, India
[2]University of Illinois, 60607, Illinois, USA
[3]SRM Institute of Science and Technology, Chennai, 603203, India
*Corresponding Author: Suresh Sankaranarayanan. Email: pessuresh@hotmail.com
Received: 18 November 2021; Accepted: 07 January 2022

Abstract: Neonatal sepsis is the third most common cause of neonatal mortality and a serious public health problem, especially in developing countries. There have been researches on human sepsis, vaccine response, and immunity. Also, machine learning methodologies were used for predicting infant mortality based on certain features like age, birth weight, gestational weeks, and Appearance, Pulse, Grimace, Activity and Respiration (APGAR) score. Sepsis, which is considered the most determining condition towards infant mortality, has never been considered for mortality prediction. So, we have deployed a deep neural model which is the state of art and performed a comparative analysis of machine learning models to predict the mortality among infants based on the most important features including sepsis. Also, for assessing the prediction reliability of deep neural model which is a black box, Explainable AI models like Dalex and Lime have been deployed. This would help any non-technical personnel like doctors and practitioners to understand and accordingly make decisions.

Keywords: APGAR; sepsis; explainable AI; machine learning

1 Introduction

Infant mortality has reduced from 5 million in 1990 to 2.4 million in 2019 where newborns face the greatest risk in their first 28 days. It has been seen that neonatal deaths within the first 28 days endure from harmful pathogenic diseases, especially viruses, in the first days of life. Poor immunity conditions are one of the major reasons that cause infant mortality. Machine Learning algorithms such as Auto Regressive Integrative Moving Average (ARIMA), Support Vector Machine (SVM), and Extreme Gradient Boost (XG-Boost) were used to make predictions of infant mortality [1]. In addition to infant mortality prediction using Machine learning, there have been researches done about sepsis in humans [2–6] and vaccine response towards evolving immunity. The powerful sepsis has the power of affecting human health potentially and the overall financial condition of the whole world.

The drawbacks in earlier systems are that only basic machine learning models have been deployed. There has been no usage of the advanced deep learning model which is the state of art for mortality prediction among infants. Also, most infant mortality predictions did not involve any human sepsis which

is the third most cause of neonatal mortality. They have considered features like birth weight, age, gestational weeks, and so forth. Even works related to vaccine response did not consider neonatal sepsis severity. Finally, it has become a matter of question on the reliability of machine learning prediction, and explaining those models towards mortality prediction, is important for the advancement of the medical field. So based on the above-mentioned drawbacks, we have deployed an Explainable AI-based models like Dalex and Lime for giving the interpretation of the Deep Neural Network (DNN) Model towards infant mortality prediction which is the state of art. The reason behind applying Explainable Artificial Intelligence (AI) for Deep Neural Network is that DNN is a black box and users are not aware of network complexity and computation happening inside towards the prediction. The traditional machine learning gives insight on model prediction as they are based on regression, probabilistic, decision rules and it clearly shows as how model predicts. So based on these motivations, Explainable AI models like Dalex and Local Interpretable Model-agnostic Explanations (LIME) are applied to DNN for giving interpretation of model prediction which has resulted in higher accuracy. The major contribution of the paper is as follows:

- Feature Selection contributing to Mortality prediction by applying feature selection algorithms.
- Validation and evaluation of Deep Neural Model against Machine learning model based on evaluation metrics.
- Explainable AI model for the interpretation of Deep Neural Model towards model prediction reliability.

The rest of the paper is organized as follows. Section 2 gives an in-depth literature review of Infant mortality prediction. Section 3 talks about the proposed work on the Explainable AI model for infant mortality prediction. Section 4 gives the result and analyses about Mortality prediction using Deep learning followed by the Explainable AI model towards the interpretation of prediction. Section 5 gives the conclusion and future work.

2 Literature Review

There has been a good amount of research work done in applying machine learning towards infant mortality prediction which are explained below.

Birth certificates of all registered United States Births were used, for three years (2000–2002). Four ML classifiers–Logistic Regression, Gaussian Naïve Bayes, SVM, and Boosted Trees were used to predict mortality before the primary birthday was enforced. Also, age at death (early neonatal, late neonatal, and post neonatal) and explanation for death were included in this research [7]. The work resulted in Logistic Regression and SVM outperforming the others with an accuracy of 99%. The challenge in this work is that advanced deep learning models are not used and only basic machine learning models have been used. Also, the most important features contributing to infant immunity like sepsis, blood culture are not used in the work for mortality prediction.

A trend was discovered wherever the mortality rate was reduced by 200 between 2015–2016 as compared to 2005–2006. The ARIMA model was used to forecast the mortality rate between 2017 to 2025. The outcome of the precision during 1971–2016 was depicted by the ARIMA model. The analysis showed that the mortality rate decreased from 33 per thousand recent births in 2017 to 15 per thousand recent births in 2025. This work involved infant mortality rate forecasting using ARIMA and did not involve infant mortality prediction based on immunity features of infants.

To find patterns and predict the risk of child mortality in South Africa, both supervised and unsupervised ML algorithms were used. The work involved a data set about demographic, socio-economic, and health data of children and adult members of surveyed households. Different clustering algorithms like heuristic and non-heuristic methods have been used for identifying the risk among different groups. The clustering

results were used to group individuals having similar features to identify the different risk groups contributing towards child mortality. In terms of supervised algorithms, Logistic Regression and XGBoost yielded the highest accuracy of 58.8% [8]. The challenge in this work is that features based on infant immunity like sepsis, blood culture, etc. are not used towards infant mortality prediction. Also, no advanced deep learning model is used for infant mortality prediction for achieving higher accuracy. The accuracy achieved in this work is way too less.

A study in [9] utilized ML algorithms like Multiple Linear Regression, Random Forest, and Gradient Boosting for the prediction of child mortality in India. An accuracy of 96.09% was achieved for the Random Forest Algorithm on comparing the performance of the above-mentioned algorithms on different error metrics like the Mean Absolute Error, Mean Square Error, and the Root Mean Square (RMS) error. Random Forest showcased the least error value for each of these metrics. The drawback of this work is that features like birth rate, below poverty line index, fertility rate, mother literacy, Female primary education, and undernourishment were only taken and features like sepsis, blood culture, etc were not taken for mortality prediction. Also, the work involved only a basic machine learning model for predicting child mortality where an accuracy of 96% was achieved.

In another study to predict infant mortality, the health data, available publically, from Sao Paulo, Brazil was referred and three important methodologies were proposed. Data from 2000 to 2017 were extracted and a suitable target variable was created, after removing all the inconsistencies from the data. Standard Exploratory Data Analysis (EDA) techniques were used in the data preprocessing stage, followed by different supervised ML algorithms like SVM, Logistic Regression, XGBoost, and Random Forest. Out of these, the XGBoost model outperformed the other algorithms [10]. The challenge in this work is that features like age, sex, prenatal age consultations, birth weight, APGAR index, gestational weeks are taken into consideration. Features like sepsis, blood culture which contribute to infant mortality are not considered. Also, no advanced deep learning model is used for achieving higher accuracy as basic machine models alone have been employed.

Research in preterm infant survival has focused on developing a tool for forecasting, and the use of machine learning methods. A test set of 2015–2016 was used where N= 5810. Among several machine learning methods, artificial neural networks (NN) have been used. In conclusion, the Neural Network had a small but significant advantage over the current approaches [11]. The drawback is that gestational age and birth weight are only taken into consideration and other immunity features like sepsis, blood culture, etc are not taken for preterm infant survival. Also, advanced deep learning models are not used for discriminating against other machine learning models for preterm infant survival.

So, from the literature review, it has been found that features about neonatal sepsis have never been considered for infant mortality prediction which is the biggest drawback when sepsis is the third most common cause of mortality in neonates. Also, when machine learning models are considered for infant mortality prediction, there is always a question on the reliability of the output. That is how medical practitioners can rely on neonatal mortality prediction based on the features selected. Hence, there is a need to explain the model which brings in Explainable AI for giving the interpretation of the machine learning models deployed. There has been no work where Explainable AI is used based on the literature reviewed towards neonatal mortality prediction based on neonatal sepsis features. All the above-mentioned drawbacks worked as a motivation towards developing an Explainable AI-based model for predicting infant mortality which will be discussed in the forthcoming sections.

3 Explainable AI Based Infant Mortality Prediction

We, in this work, have aimed at analyzing the neonatal sepsis features of newborn infants by employing deep neural model towards infant mortality prediction. For validating the model, top-ranked features are

selected by applying feature selection algorithms. Deep Neural Network which is an advanced form of Artificial Neural Network been validated based on different metrics and analyzed in terms of accuracy and error against standard machine learning models like "Naive Bayes, Logistic Regression, SVM, K Nearest Neighbor (KNN), and Decision Tree", and ensemble machine learning algorithms [12] like the Gradient Boosting Algorithms–"Extreme Gradient Boosting, and Light Gradient Boosting Machine, and Random Forest". Lastly Explainable AI [13] models like Dalex and Lime have been deployed to give an interpretation of the best performing model which is Deep Neural Network towards the prediction reliability of the model. The complete system architecture of the proposed system is shown below in Fig. 1.

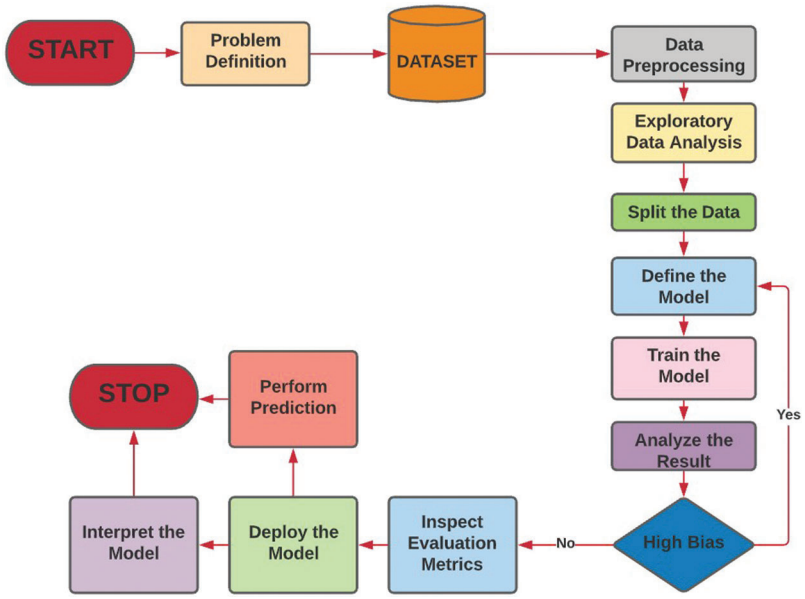

Figure 1: System architecture

In the proposed system, Data pertaining to infant mortality predicting sought with different feature including neonatal sepsis for prediction from children hospital. So towards building an efficient prediction model and testing the reliability of prediction model using Explainable AI, the first step is to preprocess the dataset obtained. Based on the data set obtained, exploratory analysis performed followed by splitting the data set into training and testing in the ratio of 70:30. Data set taken for training are modeled using different machine learning model including deep Neural Model. Model Trained are analyzed for Bias. If there is high bias, model classified as unsuitable and redefined. If the model resulted in low bias, the model validated using different evaluation metrics using training data set. Then the model is deployed for testing using real life data which is unknown to model and prediction performed. This model performance evaluated based on testing data to validate best performing model. The model performing best are deployed towards infant mortality prediction to give explanation of model using explainable AI like Dalex and Lime. This gives insight only how model working towards giving prediction. This clearly show as model prediction is being carried out properly rather than just relying on model evaluation metrics.

The theoretical background behind the Deep Neural Network and explainable AI models deployed, for the proposed work, are discussed below. As our main focus was on Deep Neural and Explainable AI for giving prediction reliability based on DNN which is black box, we have given insight into those models. Details on standard machine learning algorithm are not explained as it has been used for comparative analysis only.

3.1 Deep Neural Network

In 1986, authors Geoffrey Hinton, Rumelhart [14] discussed the importance of error backpropagation which was initially anticipated in the 1970s. Neural Network learns from backpropagation algorithms. During the learning phase, the right biases and weights are learned. Cost change is the most important factor to understand in the model given by ∂C/∂w at the cost of work C as for any weight w (or predisposition b) in the system. This articulation tells how quickly biases and weights change. So, in summary, error-backpropagation is not only a learning concept but gives in-depth knowledge as to how metrics change for the behavior of a system. In "deep learning" models, the Rectified Linear unit is one of the most common activation functions used in hidden layers. It is the most rapid learning function, delivering advanced success and excellent outcomes.

One main technique in deep learning is deep neural network. A typical DNN model is based on hierarchy of composition of linear functions and a given nonlinear activation function. We propose using deep neural network or deep learning. This type of DNN is called a DNN (k þ 1) layer, and is said to have k hidden layers. Unless otherwise stated, all layers mean hidden layers. The size of this DNN is n1 +· · · + nk. In this research work, we consider RELU as a special activation function (rectified linear unit).

$$\mathrm{RELU}(x) = \max(0, x),\ x \in R \tag{1}$$

For altering the characteristics of a neural network, such as weights and learning rate, to minimize losses, Optimizers are used. Adam is the best-performing optimizer which falls under Adaptive optimizer. Using Adam to train a neural network effectively and in less time would be more effective. Optimizers with complex learning speeds are used for sparse data.

3.2 Explainable AI

There is a need to decode the black box AI models to gain insights on model interpretations, working based on features, in a format understandable by any layman or user. This brings in the need for Explainable AI. There are numerous algorithms, under the umbrella of XAI, to explain and explore complex models. Some of the important ones are discussed below which have been deployed for our research work.

3.2.1 DALEX

There are already many great tools for Explainable AI (XAI) in Python, but the main advantage of using dalex lies in the fact that it is based on the expandable grammar of the Explanatory Model Analysis process. While exploring a model using dalex, to understand its behavior completely, some of the important lookouts are:

- Variable importance: Assess how different features are contributing to the decision-making for the final prediction.
- Partial Dependency Profile: Show the effect of the selected feature on the predicted outcome.
- Ceteris Paribus Profile: Ceteris Paribus in Latin means, "other things held constant". It shows the model response towards the individual profiles, around a single point in the feature space.
- Residual: This diagnosis helps in analyzing and comparing the errors

3.2.2 LIME

LIME is an acronym for Local Interpretable Model-agnostic Explanations. It is a visual approach that helps define each prediction. It is an agnostic model and therefore can be used in any controlled retrospective or isolation models. The output from LIME gives an insight into the inner workings of machine learning algorithms, as well as the features that are used to make a prediction. LIME attempts to measure a simple model in a single view that closely resembles how the global model behaves in that

context. The projections of a more complicated model in the field can be explained using a simpler model. LIME attempts to measure a simple model in a single view that closely resembles how the global model behaves in that context. The projections of a more complicated model in the field can be explained using a simpler model. The explanation produced by LIME at a local point **x** is obtained by the following generic formula:

$$\xi(x) = \text{argmin } L(f, g, \pi x) + \Omega(g) g \in G \tag{2}$$

where f is our real function (aka ground truth), g—is a surrogate function we use to approximate f in the proximity of x and πx defines the locality. This formulation can be used with different explanation families G, fidelity functions L, and complexity measures Ω.

4 Results and Discussions

For this research work about validating the Deep Neural Network against different machine learning models towards mortality prediction, 1946 time-stamped data was gathered between September 2014 and February 2018. The dataset used for prediction contains some of the very important features such as blood culture, time to antibiotics, sepsis group, birth weight, incubation period, and many more such features about 30 attributes that are the key deciding factors to determine an infant's immunity towards mortality prediction [15]. This system was used at the Children's Hospital of Philadelphia to record the sepsis evaluation of infants. After finalizing the dataset, a retrospective analysis of the dataset was done to determine the outcome of the attribute. Data set was split into the training and the testing sets with around 70% of the data in the training set and 30% of the data in the test set. That is 1362 samples for training and 583 samples for testing Fig. 2 shows the sample dataset used for our research.

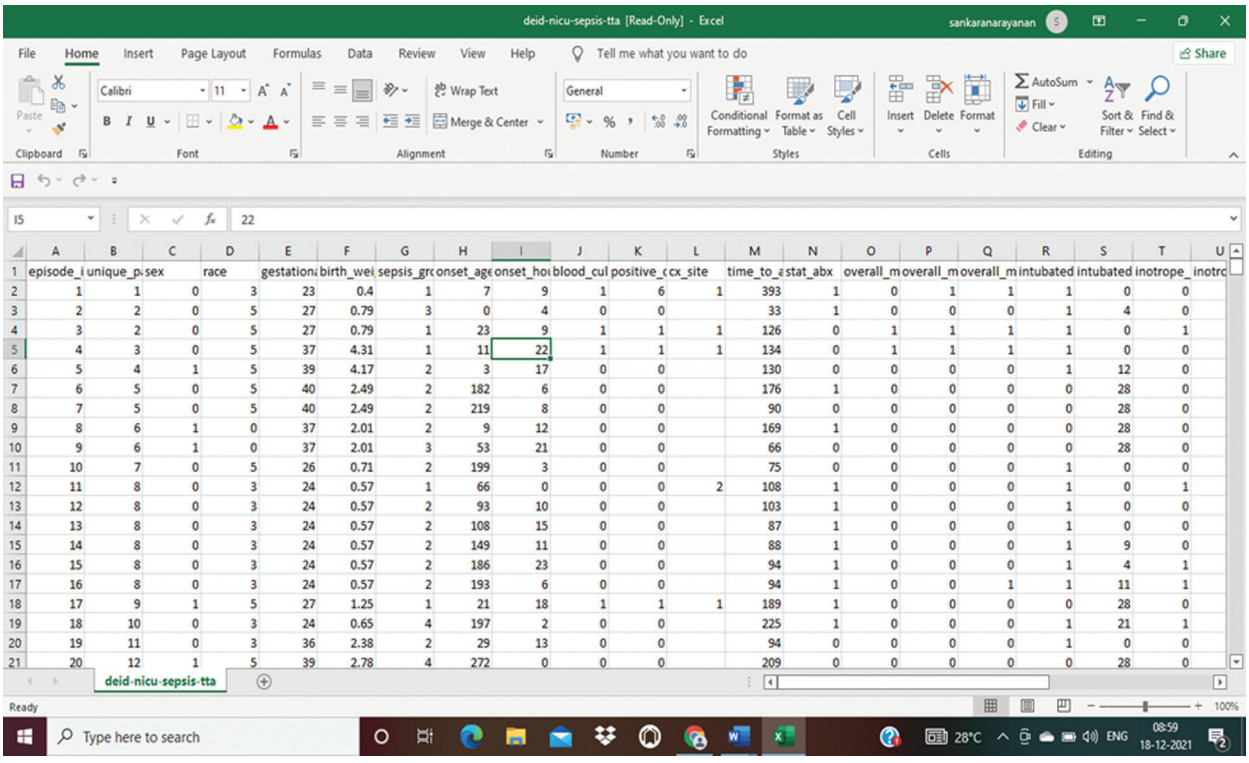

Figure 2: Sample dataset

4.1 Feature Selection

The Feature Engineering technique is considered very important to improve the performances of the machine learning algorithms. Selecting only the important features, helps in Dimensionality Reduction, thus improving the overall computational performance. A heat map was drawn to get the correlation between different features as shown in Fig. 3. All the independent features were eliminated i.e., the ones with a correlation score above 0.9. This was done to remove all the highly correlated features. But as the model performance was not good enough, other feature selection methods were applied. Different Feature selection methodologies like Pearson Coefficient, Chi-Square, Recursive Feature Elimination, Random Forest, Logistics, Light Gradient Boosting Method were implemented, and the features that got the highest ranking in all were selected which are shown in Fig. 4. From Fig. 4 it is clear that some of the features that are closely related to an infant's immunity, such as incubated_free_days, time_to_antibiotics, and sepsis_group are having the highest overall score. We used, "overall_mortality_within_30_days", as the target variable for our dataset. Its nature is dichotomous, as it has only two possible classes.

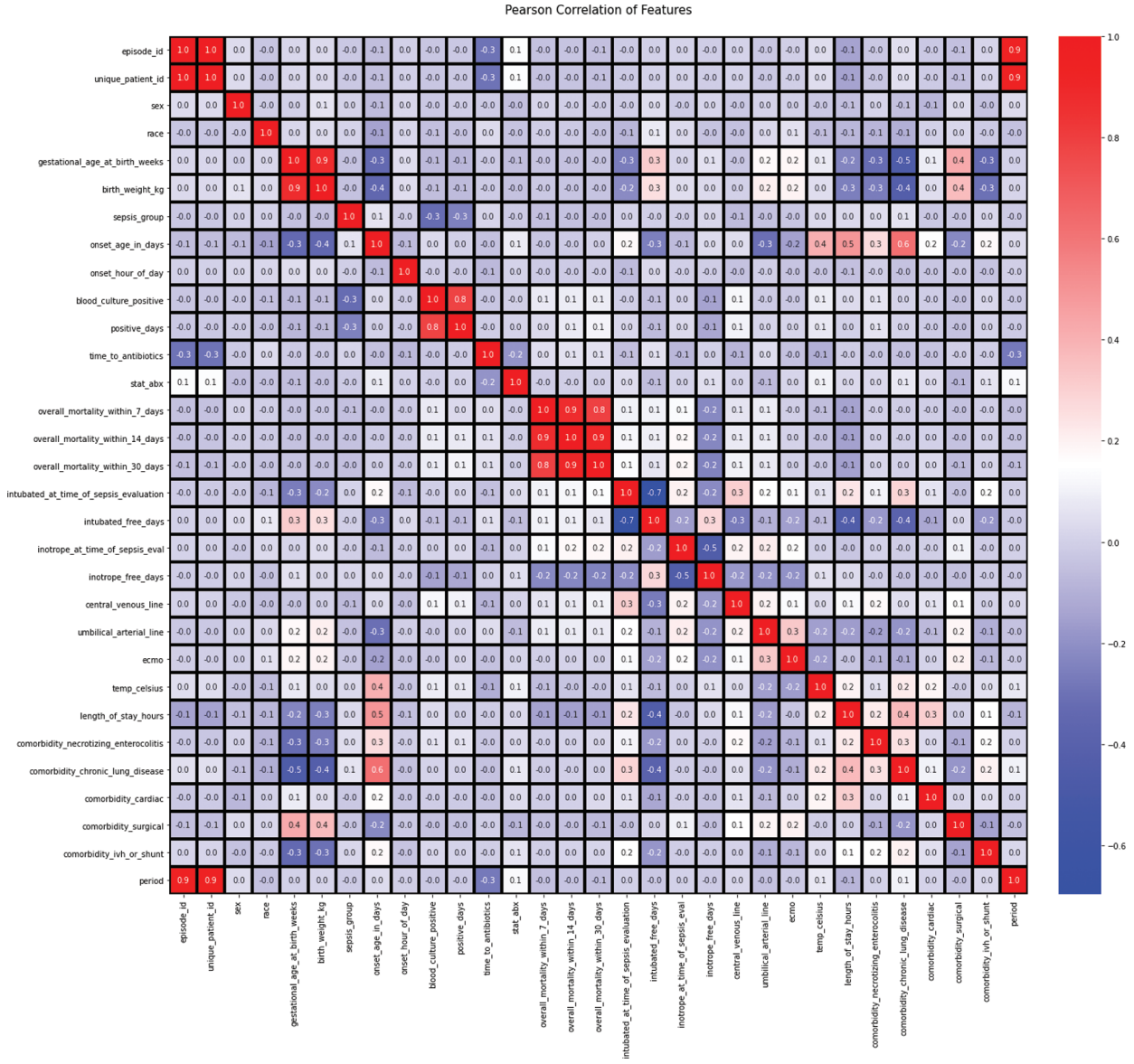

Figure 3: Heatmap of features

	Feature	Pearson	Chi-2	RFE	Logistics	Random Forest	LightGBM	Total
1	intubated_free_days	True	True	True	True	True	True	6
2	time_to_antibiotics	True	True	True	False	True	True	5
3	length_of_stay_hours	True	True	True	False	True	True	5
4	inotrope_free_days	True	True	True	True	True	False	5
5	unique_patient_id	True	True	False	False	True	True	4
6	temp_celsius	True	False	True	True	True	False	4
7	sepsis_group	True	True	True	True	False	False	4
8	race	True	True	True	True	False	False	4
9	intubated_at_time_of_sepsis_evaluation	True	True	True	True	False	False	4
10	inotrope_at_time_of_sepsis_eval	True	True	True	True	False	False	4
11	gestational_age_at_birth_weeks	True	True	False	True	False	True	4
12	central_venous_line	True	True	True	True	False	False	4
13	birth_weight_kg	True	False	True	False	True	True	4
14	umbilical_arterial_line	True	True	True	False	False	False	3
15	positive_days	True	True	True	False	False	False	3
16	period	True	True	True	False	False	False	3
17	onset_age_in_days	False	False	True	False	True	True	3
18	episode_id	True	True	False	False	True	False	3
19	comorbidity_surgical	True	True	True	False	False	False	3
20	comorbidity_cardiac	True	True	True	False	False	False	3
21	blood_culture_positive	True	True	True	False	False	False	3
22	onset_hour_of_day	False	False	False	False	True	True	2
23	stat_abx	False	False	True	False	False	False	1
24	comorbidity_necrotizing_enterocolitis	False	True	False	False	False	False	1
25	comorbidity_ivh_or_shunt	False	True	False	False	False	False	1
26	comorbidity_chronic_lung_disease	False	False	True	False	False	False	1
27	sex	False	False	False	False	False	False	0
28	ecmo	False	False	False	False	False	False	0

Figure 4: Top feature summarization

4.2 Performance of Deep Neural Model

In terms of Deep Neural Networks, a total of 6 hidden layers were used in the Deep Neural Network model. The activation function used in the hidden layer was Leaky ReLU, with an alpha value of 0.01, and in the output layer was Sigmoid. The He Normal Initializer was used to initialize the kernel and Dropout Regularization with a dropout score of 0.2 was used to avoid the problem of overfitting. The type of model created is sequential. The shape of the input is restricted to (28,), and was provided into the first hidden layer. To compile the model created, Adam was used as the optimizer, and binary_cross entropy as the loss function, on which the model was tested upon as the problem is binary classification in nature, along with accuracy used as the metric to evaluate the model performance. The reason for choosing ADAM optimizer is that it it achieves good results very fastly. In addition it updates the weights iteratively during training The model was then fitted on the training dataset, with the batch size of 32 and the epochs value was set to 70. After several try-on, 70 epochs turned out to be the best epoch value, giving the best performance. Increasing the hidden layers would have been expensive and prone to overfitting, but using Dropout Regularization helped prevent the issue of overfitting. This was compared with other standard Machine learning algorithms like Logistic Regression, SVM, Naïve Bayes, KNN, Decision Tree, Random Forest, Gradient Boosting, XGBoost, Light GBM and validated in terms of precision, recall, F1 score, losses, and accuracy. The DNN model gave an incredible accuracy of 98.54%,

on being trained for 70 epochs, with a very small loss, in both the training and the testing sets. It also gave an excellent precision, recall, and F1 score. These are shown in Tab. 1.

Table 1: Model evaluation results

Models	Accuracy	Precision	Recall	F1 Score	Log loss
Logistic regression	84.53	83.04	89.05	0.86	5.34
SVM	67.76	64.39	87.80	0.74	11.13
Naïve Bayes	79.60	81.83	79.13	0.80	7.04
KNN	97.03	95.42	99.17	0.97	1.02
Decision tree	96.71	96.13	97.72	0.96	1.13
Random forest	96.27	95.18	97.93	0.96	1.28
Gradient boosting	94.84	93.78	96.69	0.95	1.78
XGBoost	95.06	92.29	98.96	0.95	1.70
Light gradient boosting	94.72	92.08	98.55	0.95	1.81
Deep neural network	98.54	98.58	98.78	0.98	0.50

From Fig. 5, it can be inferred that the Deep Neural Network model outperformed the other machine learning models giving the highest accuracy. Also, in terms of precision, DNN outperforms the other machine learning algorithms with a precision score of 98.58. This is a clear indication that the model can correctly predict 98.78% times the mortality of infants as alive or dead based on neonatal sepsis features. In terms of Recall, DNN has a recall score of 98.78. This shows that the DNN model is highly capable of identifying the relevant data accurately. In terms of F1 score, DNN outperforms all the models with a score of 0.98. F1 is the harmonic mean between precision and recall, having a good F1 Score, close to 1, is an important metric for evaluating a model's performance. In terms of log loss, DNN is the best performing model as it suffers from a very low loss of 0.5 only.

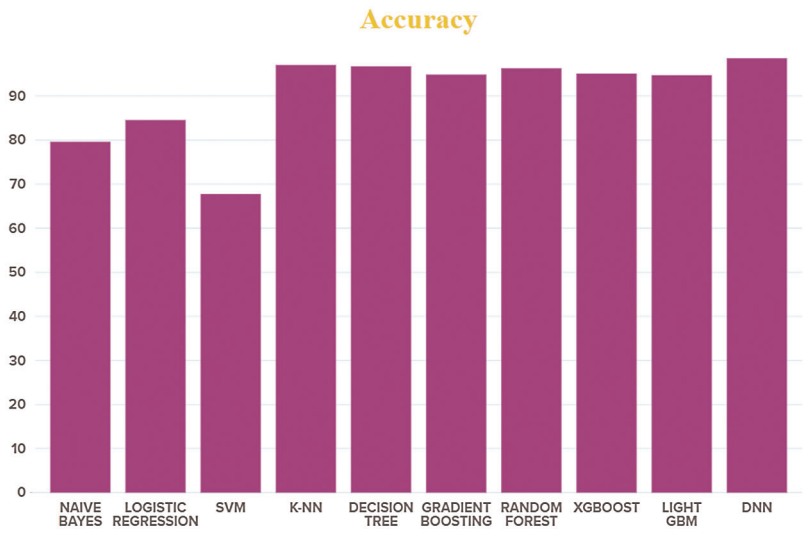

Figure 5: Accuracy of models

4.3 Performance of Explainable AI Models

Having validated the performance of machine learning and deep neural models for mortality prediction, it is very much imperative to validate the reliability of predicted results. Deep learning is a black box and there is no indication as to how the model predicts. In Traditional Machine learning, we can get insight as how model function which is based on probability values, Euclidean distance and regression method. This section gives insight into explaining the Deep Neural models that give insight into features contributing the most towards prediction thereby validating the predicted results.

4.3.1 DALEX

Model Level Explanations

Higher the loss, lower the performance. From Fig. 6, we can interpret that the performance of the model created using DNN is very high as the loss values are quite less. Error-values on mean squared error, Root Mean Square Error, Mean Absolute Error, r2, and Mean Absolute Deviation Error are quite less which indicates loss values are quite less for the model to perform well.

	mse	rmse	r2	mae	mad
Mortality	0.003912	0.062546	0.984258	0.005352	0.000002

Figure 6: Loss estimation

Variable Importance

From Fig. 7, we can see that the top-most features used for model prediction are sepsis_group-outcome of sepsis evaluation that gives information on positive culture and no positive culture after a specific period of antibiotic treatment followed by incubated_free_days-number of days that were not incubated after sepsis evaluation in the first 28 days, onset_age_in_days-infant's age in days of life and many more that hold a strong relationship with the immunity of an infant.

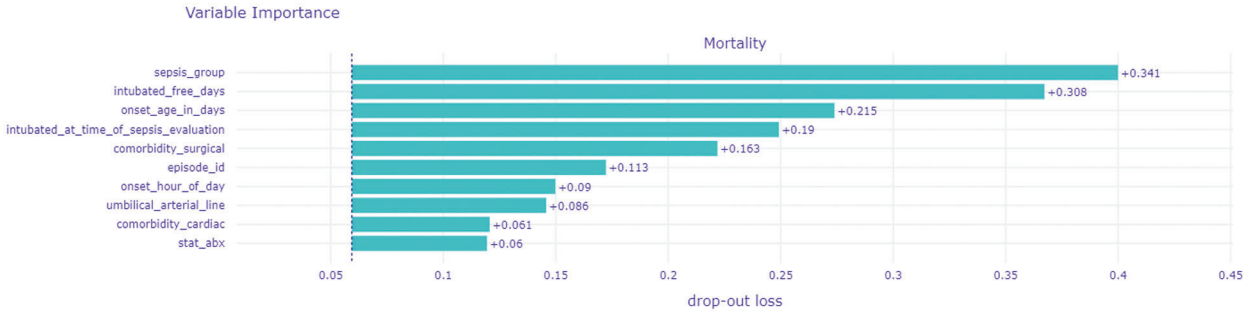

Figure 7: Variable importance

Understanding the continuous relationships between variables and predictions

This step was performed to understand the trend that the topmost–important features follow throughout the dataset, to build an appropriate and highly accurate predictive model. At any point, it gives the difference between the prediction at that point in comparison with the mean prediction, which is based on the target variable, 'Mortality'. It also calculates the ceteris paribus for explaining the complexity of the model tested upon every single observation.

Residuals

Residuals are the measured values from a sample, which helps in understanding errors. It is calculated and plotted against predictions, the target variable. It makes use of the "LOWESS" trendline, to help us see the relationship between the variables and foresee trends. It is a model diagnostic performance as it diagnoses the residuals that depend on the model performance. The recursive cumulative distribution of residual curves is plotted to check the stability of the variable in the long run. They follow an independent distribution and do not fall into the problem of falling into deficiencies when any particular part of data is smudged over all the residuals.

Partial Dependency profiles

Partial Dependency (PD) shows that the probability of the mortality rate increases with an increase in the feature as it is tested across. It helps in depicting the functional relationships between small numbers of inputs and their corresponding predictions. We can interpret that the prediction of mortality is partially dependent on the values of other input features that are of high importance.

Model_parts function

This function is used to calculate the variable importance explanations. It is an object of the class model_diagnostic_explainer. It calculates the feature importance based on the loss functions calculated for every feature as a function of the dropout loss. We can see that the overall loss suffered by the complete model is around 0.059 as shown in Fig. 8 when the target of prediction is mortality. This is a very good sign that explains the adaptable and high-performance nature of the model.

model_profile function

This function explores model responses as a function of selected variables. The explanations can be calculated on the Partial Dependence Profile (PDP) and the Accumulated Local Dependence Profile (ALDP). But the PDP calculates on the ceteris paribus whereas the ALDP calculates both the ceteris paribus and the accumulated dependency and helps in giving a visual comparison between them both. ALDP describes how different features affect the prediction, and is a faster alternative to PDP, which shows the marginal effect that one or two features have on the prediction results.

Predict Level Explanations

The predict function of the Dalex is nothing more than a normal model prediction, except for the fact that now it uses the Explainer interface. Through these prediction output results, we can conclude that the feature importance results were true as we can see that sepsis_group stands out to be the feature that is having a great impact on the chances of survival of any particular infant, being tested upon.

Surrogate Model

A surrogate model is an interpretable model that is built upon some prediction model, intending to approximate the predictions of black-box AI models. It is used to explain the feature importance along with its other models, which are used as a proxy. It is also called the approximation model. The main intention behind the creation of a Surrogate model, upon the explainer model, was to comprehend the path followed and the decisions made by the explainer model, internally, to decode the deep neural model.

Visualization of the Local Explanations

Fig. 9 gives a dataset-specific, instance-level analysis of different features, based on the surrogate model created. It shows how impactful different features are on making predictions. This was done for understanding the feature contribution from both ends which clearly shows that sepsis_group and intubated_Free_days are having a major impact on predictions.

	variable	dropout_loss	label
0	_full_model_	0.059327	Mortality
1	unique_patient_id	0.061752	Mortality
2	birth_weight_kg	0.070214	Mortality
3	period	0.070287	Mortality
4	blood_culture_positive	0.077633	Mortality
5	comorbidity_ivh_or_shunt	0.078205	Mortality
6	inotrope_at_time_of_sepsis_eval	0.080486	Mortality
7	comorbidity_necrotizing_enterocolitis	0.088712	Mortality
8	gestational_age_at_birth_weeks	0.093480	Mortality
9	central_venous_line	0.096297	Mortality
10	comorbidity_chronic_lung_disease	0.098027	Mortality
11	race	0.098151	Mortality
12	positive_days	0.105840	Mortality
13	ecmo	0.108802	Mortality
14	length_of_stay_hours	0.109764	Mortality
15	inotrope_free_days	0.111074	Mortality
16	sex	0.113365	Mortality
17	time_to_antibiotics	0.115301	Mortality
18	comorbidity_cardiac	0.120519	Mortality
19	stat_abx	0.120909	Mortality
20	temp_celsius	0.121830	Mortality
21	umbilical_arterial_line	0.143367	Mortality
22	onset_hour_of_day	0.144354	Mortality
23	episode_id	0.179250	Mortality
24	comorbidity_surgical	0.222128	Mortality
25	intubated_at_time_of_sepsis_evaluation	0.248871	Mortality

Figure 8: To study variable importance as a change in loss functions

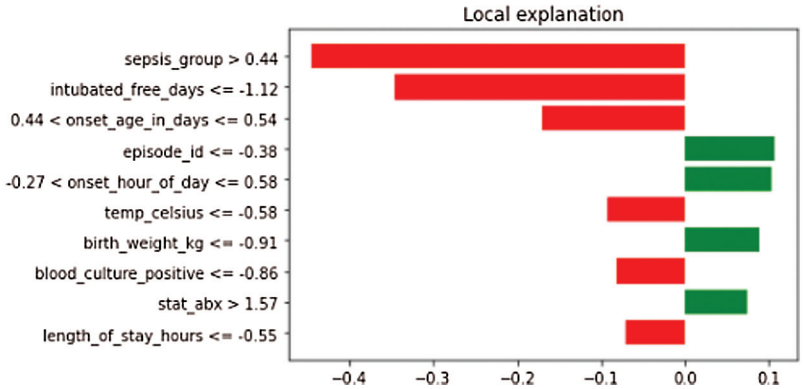

Figure 9: Plot for instance level surrogate model

The decision tree regressor shows the interconnection between several important features in the dataset, and their overall contribution towards the final predicted results, by breaking the analysis into deeper and more simplified metrics as shown in Fig. 10.

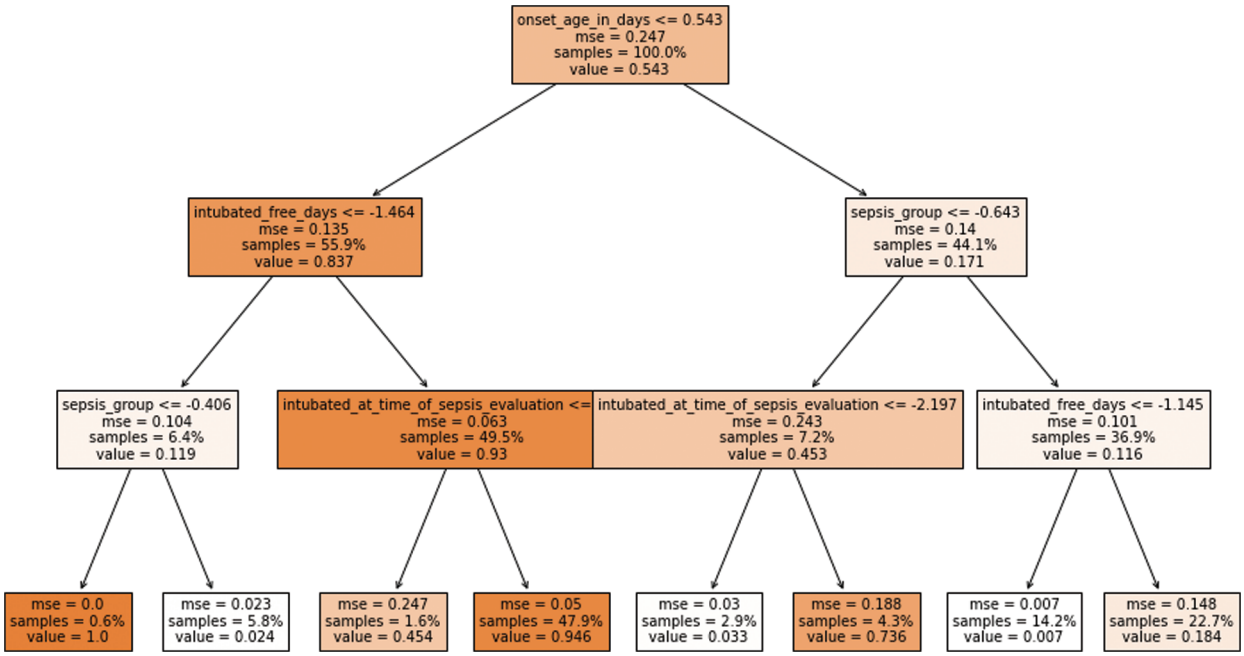

Figure 10: Decision tee plot for the surrogate model

4.3.2 LIME

The LIME module has been used to interpret the predictions made by deep neural network models. This was done to draw insights into the decisions made by prediction models locally using LIME for Deep Neural Network. This model focused only on predict level explanation and not like dalex on model level. So, we have taken DNN which outperformed other models for an explanation using LIME.

Fig. 11 for DNN shows the feature orientation in the positive-negative bar along with the feature value as displayed on the table for a particular infant from a dataset. The blue color (0) represents the positive and the orange color (1) represents the negative orientation. This shows the child will survive in the neonatal phase based on sepsis_group on the positive side. The output of explanations shows the effect and the contribution of every feature to the prediction made. It provides local illustrations which help to determine those features whose changes can severely affect the prediction results using LIME.

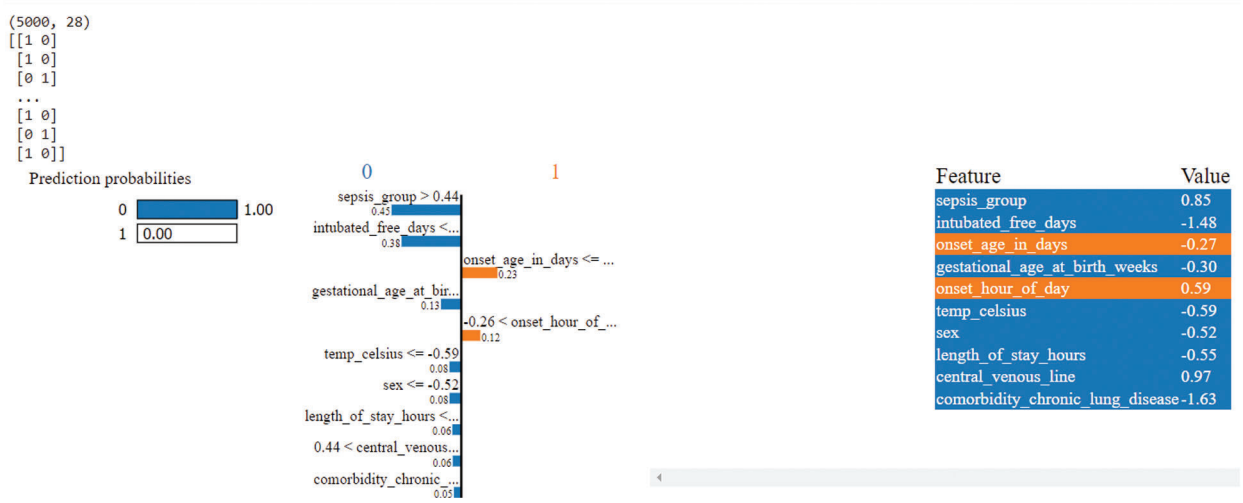

Figure 11: Feature values using LIME

5 Conclusion

Neonatal sepsis is the third most common cause of neonatal mortality and a serious public health problem, especially in developing countries. We have deployed deep neural networks for predicting infant mortality and compared them against other machine learning models in terms of standard metrics like Precision, Recall, F1 Score, Accuracy, and log loss. From all the models deployed, it was found that the deep neural model turned out to be the best performing model, with very high accuracy and a low log loss. Though the Deep Neural model gave the best prediction accuracy with reduced loss, it is highly important how to rely on the predicted results. So, the Explainable AI model DALEX deployed a deep neural for giving model and prediction level explanations. LIME model deployed for giving prediction level explanation locally for deep neural models. Thus, this information can be used to work on and improve an infant's immunity and make it resistant to viral infections. Consequently, saving the infant from neonatal mortality and reducing the overall infant mortality rate. The limitation of the research is limited data set. So, with more larger data set, advanced deep learning models like RNN, LSTM etc could be deployed and evaluated leading to better validation results in future. In addition, research work can be extended for the prediction of the immunity for all age groups for different kinds of viruses emanating. Also, the immunity of older people could be studied in predicting against different viruses in specific with different health conditions too. These Machine learning and Explainable AI models could play a major role in this.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] K. M. Amit, S. Chandar and S. M. Mani, "Forecasting Indian infant mortality rate: An application of autoregressive integrated moving average model," *Journal of Family Community*, vol. 23, no. 2, pp. 123–126, 2019.

[2] A. Zea-Vera and T. J. Ochoa, "Challenges in the diagnosis and management of neonatal sepsis," *Journal of Trop Pediatr*, vol. 61, no. 1, pp. 1–13, 2015.

[3] R. F. Hamdy and R. L. DeBiasi, "Every minute counts: The urgency of identifying infants with sepsis," *Journal of Pediatr*, vol. 217, no. 1, pp. 10–12, 2020.

[4] M. Schmatz, L. Srinivsan, R. W. Grundmeir, O. U. Elci, S. L. Weiss *et al.,* "Surviving sepsis in a referral neonatal intensive care unit: Association between time to antibiotic administration and in-hospital outcomes," *The Journal of Pediatrics*, vol. 217, no. 1, pp. 59–65, 2020.

[5] J. Phua, W. Ngerng, K. See, C. Tay, T. Kiong *et al.,* "Characteristics and outcomes of culture-negative versus culture positive severe sepsis," *Journal of Crit Care*, vol. 17, no. 5, pp. 1–12, 2013.

[6] A. Dierig, C. Berger, P. K. A. Agyeman, S. B. Stirnemann, E. Giannoni *et al.,* "Time-to-positivity of blood cultures in children with sepsis," *Frontiers in Pediatrics*, vol. 6, no. 1, pp. 1–9, 2018.

[7] A. Saravanou, C. Noelke, N. Huntington, D. A. Garcia and D. Gunopulos, "Infant mortality risk from information available at the time of birth," in *Proc. Population Association of America Annual Meeting*, Austin, Texas, USA, pp. 1–18, 2019.

[8] C. Kabudula, R. Kara, H. Wandera, F. A. A. Dake, J. Dansou *et al.,* "Evaluation of machine learning methods for predicting the risk of child mortality in South Africa," in *Proc. 8th African Population Conf. (APC)*, Kampala, Uganda, pp. 1–6, 2019.

[9] R. Gwande, S. Indulkar, H. Keswani, M. Khatri and P. Saindane, "Analysis and prediction of child mortality in India," *International Research Journal of Engineering and Technology*, vol. 6, no. 3, pp. 5071–5074, 2019.

[10] C. E. Beluzo, L. C. Alves, E. Silva, R. Bresan, N. Arruda *et al.,* "Machine learning to predict neonatal mortality using public health data from sao paulo-Brazil," medRxiv, 2020.

[11] M. Poddo, D. Bacciu, A. Micheli, R. Bellu, G. Placidi *et al.,* "A machine learning approach to estimating preterm infants survival: Development of the preterm infants survival assessment (PISA) predictor," *Scientific Reports*, vol. 8, pp. 1–8, 2018.

[12] G. Ke, Qi. Meng, T. Finley, T. Wang, W. Chen *et al.,* "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. 31st Conf. on Neural Information Processing Systems*, CA, USA, pp. 1–9, 2017.

[13] B. Przemyslaw, "DALEX: Explainers for complex predictive models in R," *Journal of Machine Learning Research*, vol. 19, no. 84, pp. 1–5, 2018.

[14] D. Rumelhart, G. Hinton and R. Williams, "Learning representations by backpropagating errors," *Nature*, vol. 323, no. 1, 533–536, pp. 533–536, 1986.

[15] S. Ostapenko, M. Schmatz, L. Srinivasan, O. U. Elci, S. L. Weiss *et al.,* "Neonatal sepsis registry: Time to antibiotic dataset," *Data in Brief*, vol. 27, no. 1, pp. 1–6, 2019.