International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 04 | April – 2026 DOI:10.55041/ISJEM06026
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Predicting Child Mortality using Ensemble Machine Learning Models

V.Tejaswi, P.Harshitha, Noorunnisha, N.Sindhil Raj, Dr Marada Srinivasa Rao
Department of Information Technology, Maharaj Vijayaram Gajapathi Raj College, Vizianagaram, Andhra Pradesh, India

Abstract—Child mortality, which means the death of children under the age of five, is a serious global issue, especially in low and middle income countries, where proper healthcare and nutrition are not easily available. Early prediction of the risk can help doctors and healthcare workers to take necessary actions and save lives. In this work, multiple machine learning models including traditional classifiers, tree-based algorithms, boosting techniques, and ensemble methods are implemented and compared. A custom ensemble model is developed by combining the outputs of multiple models . To make the system practical and user-friendly, the trained model is deployed using a flask-based application. The proposed system not only improves the prediction accuracy but also enhances the usability and interpretability. Present work is carried out during the period of eight months(Aug 2025 to Mar 2026). In this work, we focus on developing a predictive system using machine learning models to estimate the risk of child mortality and also it highlights which features influence child mortality the most.

Keywords–Child Mortality, Machine Learning, Ensemble Model, Healthcare, Risk Score***, Flask.***

I.INTRODUCTION

Child mortality is still a serious issue, especially in developing countries where healthcare and nutrition are not easily available[2], [13]. It refers to the death of children under the age of five due to various factors such as poor healthcare facilities, malnutrition, lack of awareness, and socio-economic conditions[7], [9]. Identifying high-risk cases at an early stage is essential for reducing mortality rates and improving child health outcomes[1], [6].

This work mainly focuses on developing a machine learning-based system for predicting child mortality risk using various demographic, maternal, and healthcare-related features. The system implements multiple machine learning modelsPal, K. et al., “Predictive Modeling of Neonatal MortalityTrendsusing Machine Learning,” 2025. Including traditional classifiers, tree-based algorithms, and boosting techniques [3], [4]. We are using ensemble machine learning methods to combine the outputs of different models to improve prediction performance [5], [8], [11]. The traditional studies mostly depend on the surveys and statistical methods, which show general trends but cannot predict the risk for each child. With the growth of data and technology, machine learning provides a better way to study these factors and identify hidden patterns. Using past data, it can predict whether a child is at risk and also highlight which features influence the most.

In this study , we used a complex dataset that includes a wide range of demographic, health care and socio-economic features. These features are carefully selected based on their association with child mortality outcomes[13], [15]. To evaluate the predictive capability of machine learning in this domain, the study systematically analyzes and compares 13 different classifiers, including traditional models (Support Vector Machine, Logistic Regression, K-Nearest Neighbors,Naive Bayes), tree-based models (Decision Tree, Random Forest), boosting algorithms (AdaBoost, Gradient Boosting, XGBoost), and ensemble methods (Soft Voting, Hard Voting, and stacking) [3], [5], [8]. Each classifier is trained and tested on the given dataset, and their performance is calculated using evaluation metrics that are accuracy, precision, recall and F1-score. The study integrates predictive modelling with Explainable AI(XAI) techniques, and also here we are integrating a new model known as a custom ensemble model ,this model is developed based on the combination of existing models. And the combination is drawn by comparing the existing models with evolutionary metrics like Accuracy, Precision, Recall, and F1 Score based on this it takes the top two best models and also here particularly utilizing SHAP (SHapley Additive Explanations) for the interpretation of model predictions. It offers a detailed comparative analysis of traditional and ensemble models for child mortality prediction to enhance the reliability of the model outcomes.

II. LITERATURE SURVEY

Satty, A. et al. [1] used statistical analysis to develop a machine learning model to predict child mortality and tested various models using statistical methods to determine the effective model. The study was limited to a specific dataset. The absence of explainability reduces its applicability in real world health care scenarios.

Naznin, S. et al.[2] focused on predicting child mortality using regression and time series models. The study achieved strong performance with R2 values, indicating good predictive capability. The approach was limited to a specific geographic region only for Bangladesh.

Pal et al.[3] developed a predictive model for neonatal mortality using Random forest WHO datasets. The model achieved high prediction accuracy and reliability. The study focused on identifying key factors which affect neonatal mortality. The scope was limited only to neonatal data and did not consider broader demographic variables. This limits the effectiveness of the prediction of child mortality.

Das et al. [4] Applied machine learning models such as logistic regression, gradient boosting and random forest to predict mortality in ICU children. This study achieved better performance across multiple evaluation metrics. It is mainly focused on healthcare environments. The dataset was limited to ICU patients only lacking diversity.

Iqbal et al. [5] Proposed a smart health care system using machine learning models including SVM, KNN, and random forest. The study used SMOTE to handle class imbalance and achieved good prediction accuracy. The study lacked feature interpretability and transparency.

Samuel et al. [6] Used random forest and artificial neural networks to predict child mortality using DHS datasets. The models achieved high accuracy and AUROC values. This study demonstrated both traditional and deep learning approaches. The issues are overfitting and limited explainability were observed.

Saroj et al. [7] Compared multiple machine learning models including decision trees, naive bayes, KNN, and neural networks. The study achieved high accuracy in predicting mortality risk factors. The results were dataset specific and lacked generalization. Also did not focus on model interpretability.

Mfateneza et al. [8] Applied machine learning techniques such as random forest, XG boost and stacking methods to predict infant mortality in Rwanda. The study showed good results with improved accuracy using ensemble methods. The dataset suffered form data imbalance issues. Additionally, it was limited to a specific region.

Islam et al. [9] predictive analytics in IoT enabled smart health care systems using models like logistic regression, decision trees and SVM. The study demonstrated good prediction accuracy. It lacked real time deployment and scalability. This is a partial implementation.

Tesfaye et al. [10] developed a web based child mortality estimation using decision trees and statistical techniques. The system provided basic prediction capabilities through the web application. The model relied on simple techniques and assumed linear relationships, it lacked advanced machine learning methods and accuracy improvements.

Bogale et al. [11] Developed a machine learning model to predict perinatal mortality using ensemble methods such as random forest, gradient boosting and cataboost. The gradient boosting model achieved the highest accuracy and was selected for the deployment.

Khan et al. [12] Proposed an ensemble machine learning approach for predicting cesarean child birth using models like XG boost, ada boost and catboost. The study achieved high accuracy and identified most important medical features influencing child birth decisions. The study was focused on child birth prediction rather than prediction tasks.

Bizzego et al. [13] Used machine learning techniques on a large UNICEF dataset to identify predictors of under five mortality. The study analyzed 37 variables and ranked the most influential factors using random forest. The study focused more on factor identification rather than real time prediction systems.

Sheakh et al. [14] Analyzed child and mortality risk factors using multiple machine learning algorithms. This study applied 7 classification models and achieved high accuracy after handling imbalance data. It focused mainly on accuracy and lacked detailed model interpretability.

Nguyen [15] conducted a study to evaluate statistical and machine learning methods for predicting child mortality using data from hospitals in Uganda. Multiple models such as Logistic regression, Random forest and gradient boosting were compared using cross validation. The models achieved high AUC values.

III.METHODOLOGY

The proposed system follows a systematic machine learning process for predicting child mortality risk. The methodology consists of multiple stages including data preprocessing, feature selection, model training, evaluation, new custom ensemble model proposal, 5 fold cross validation, SHAP explainability, risk scores percentages and web deployment. This is how the project flow looks like. Initially, the dataset containing approximately 5,000 records is collected and analyzed. The data includes demographic, maternal and health care related features. The features are mother_age, father_age, birth_order, birth_weight, mother_education, wealth_index, residence, antenatal_visits, institutional_delivery, vaccination_status, access_to_water, toilet_facility, low_birth_interval and mortality. These are the total 14 features in this mortality is the target feature.Data preprocessing is performed to handle missing values, null values, encode categorical data and scaling of numerical data. Exploratory data analysis (EDA) is conducted to understand the feature distributions and relationships.

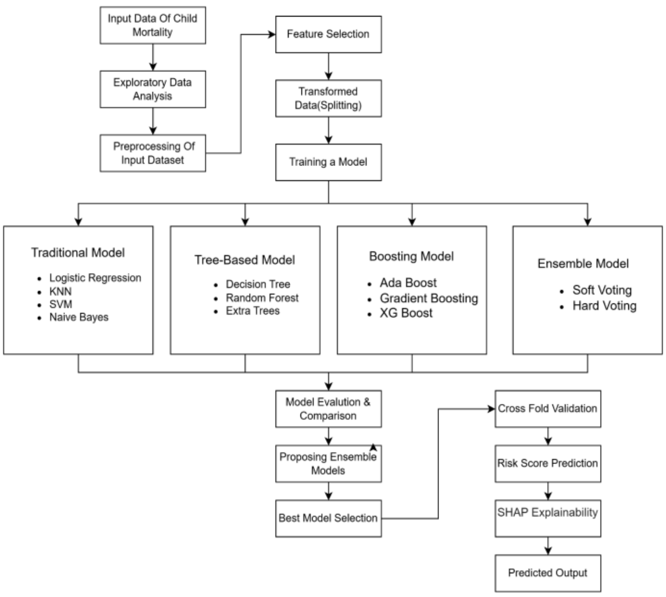

Fig.1 Overall Workflow of Child Mortality Prediction System

Fig.1 shows the overall work flow of the proposed child mortality risk prediction system. The system follows this architecture to get the best model for the predictions.

In the next stage, feature selection techniques are applied to reduce the dimensionality and select the most relevant features among them for model training. Then multiple machine learning models are implemented including traditional, tree based, boosting and ensemble algorithms. These models are trained and evaluated using standard performance metrics such as accuracy, precision, recall and F1 score. The custom ensemble model is developed by combining the top top models among those 4 types of algorithms. By using weighted averages technique the new custom ensemble model developed.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 04 | April – 2026 DOI:10.55041/ISJEM06026
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Cross validation is used to ensure the stability and generalization of the model across different data subsets. SHAP (SHapley additive explanations) is applied to identify the contribution of each feature providing better understanding and transparency of the model's predictions. Finally, the trained custom ensemble model is integrated into a flask based web application to enable real time prediction and user interactions.

IV.RESULTS

Table-1: Performance comparison of machine learning models

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Gradient Boosting	0.963	0.793	1	0.885
Random Forest	0.96	0.783	0.993	0.858
XGBoost	0.956	0.769	0.986	0.864
Extra Trees	0.95	0.752	0.965	0.846
AdaBoost	0.938	0.702	0.979	0.817
SVM	0.938	0.706	0.965	0.815
Decision Tree	0.936	0.693	0.986	0.814
Logistic Regression	0.924	0.66	0.957	0.786
KNN	0.912	0.625	0.951	0.754
Naive Bayes	0.865	0.514	0.887	0.651

Table-1 shows the performance comparison of different machine learning models using standard evaluation metrics. It is observed that gradient boosting and random forest achieves the highest score among all models This shows that boosting models perform better than traditional models.

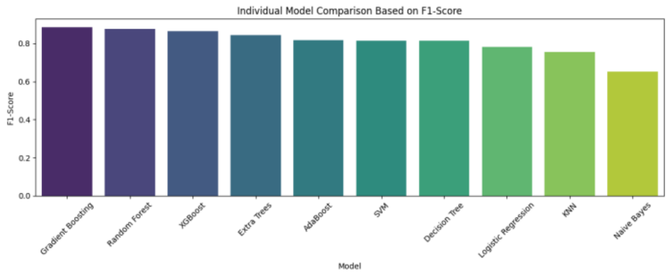

Fig.2 Final comparison of machine learning models based on F1-score

Fig.2 shows the final comparison of different machine learning models based on F1-score. F1-score is used as the primary evaluation metric as it balances precision and recall, making it suitable for medical datasets. In this representation also gradient boosting and random forest achieve high scores. This helps in identifying the best performing model visually.

Table-2: Performance comparison of ensemble models

Model	Accuracy	Precision	Recall	F1-Score
Hard Voting	0.963	0.7966101695	0.9929577465	0.8840125392
Soft Voting	0.956	0.7692307692	0.9859154930	0.8641975309
Custom Performance-Weighted Ensemble	0.988	0.9276315789	0.9929577465	0.9591836735

Table-2 shows the performance comparison of different ensemble models including hard voting, soft voting and custom ensemble model. It can be observed that the custom ensemble model achieves the highest accuracy and F1-score among all. The results demonstrate that the custom ensemble model is more effective and reliable for predicting child mortality risk.

Hence, the custom ensemble model is selected as the final model for predicting child mortality.

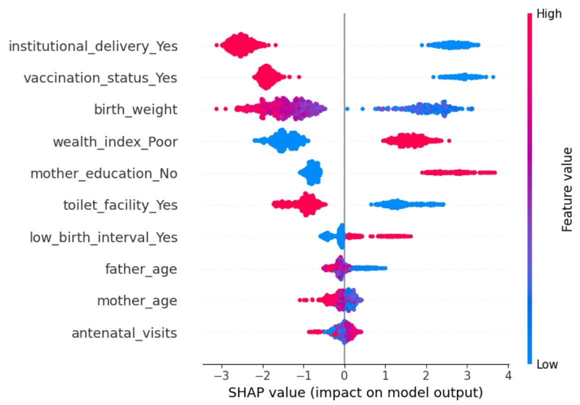

Fig.3 SHAP summary plot showing feature impact on child mortality prediction

Fig.3 shows the SHAP summary plot, which explains how different features influence the model's prediction Features

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 04 | April – 2026 DOI:10.55041/ISJEM06026
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

such as institutional delivery, vaccination status and birth weight have the highest impact on the model output.

Table-3: Sample prediction results with risk probability and risk score

Actual_Class	Predicted_Class	Risk_Probability	Risk_Score (%)
1	1	0.9952	99.52
0	0	0	0
1	1	0.8715	87.15
0	0	0.0039	0.39
0	1	0.7239	72.39

Table-3 shows sample prediction results generated by the model, including actual class, predicted class, risk probability and risk score. The risk score is derived from the probability and expressed in percentage format for easy interpretation. The model correctly predicts most cases demonstrating good performance. This helps in understanding both prediction accuracy and the severity of risk.

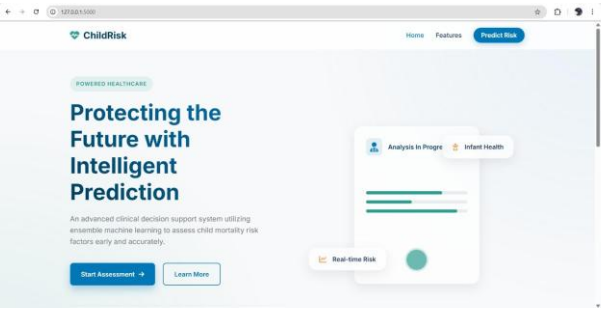

Fig. 4 Web application Home page interface

Fig.4 shows the homepage of the developed web application. It provides a simple and user-friendly interface for accessing the system. The design ensures easy interaction and smooth user experience.

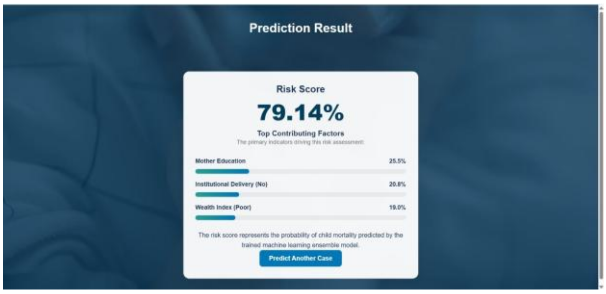

Fig. 5 Risk score prediction output with top contributing factors

Fig. 5 shows the prediction result generated by the web application. The system calculates and displays the risk score in percentage format along with the top contributing factors.

V.CONCLUSION

In this paper, Machine learning based approach is used for predicting child mortality risk using features like demographic and healthcare related features. Multiple classification models are used which includes traditional, tree-based, boosting and ensemble models. Those models are trained and tested using a dataset which has 5000 rows and 13 features. These models are implemented and evaluated using the metrics accuracy, precision, recall and F1score. In the comparison of the traditional, tree based, boosting and ensemble models, Gradient boosting and random forest are the top models among them. A custom ensemble model is proposed by integrating the top 2 models using weighted averaging, which improves prediction performance The new custom ensemble model is implemented and evaluated with the same dataset. The model is further validated using 5 fold cross-validation to insure its stability and generalization capability. And also it uses SHAP explainability to identify the most influential features that are affecting the predictions which provides transparency and interpretability. The system also generates the risk score percentage which allows users better understanding of child mortality risk. The integration of the trained custom ensemble model with the web application enables realtime prediction and user interaction. The results confirm that the proposed model is accurate, reliable and practical for real world health care applications.

Conflict of Interest: There is no conflict of interest.

VI.REFERENCES

[1] ALI SATTY (2025) Statistical Insights into Machine Learning Models for Predicting Under-Five Mortality, Arar, Saudi Arabia Vol-13 45312-45320.
[2] Naznin S, Uddin MJ, Ahamad I, Kabir A (2025) Analyzing and forecasting under-5 mortality trends in Bangladesh using machine learning techniques. PLoS ONE 20(2):e0317715
[3] Dr. Kaushika Pal and Dr. Jitendra K Pal (2025) Predictive Modeling of Neonatal Mortality Trends Using Machine Learning Journal of Neonatal Surgery, 14 (22s), 1033-1039.
[4] Das S, Erdman L, Brals D, et al. (2024), Development of machine learning predicting mortality using routinely collected observational health data from 0-59 months old children admitted to an intensive care unit in Bangladesh.
[5] Farrukh Iqbal and Muhammad Islam Satti, (2023) Predictive analytics in smart healthcare for child mortality prediction using a machine learning approach. vol-18.
[6] Oduse Samuel, Temesgen Zewotir and Delia North (2024) Application of machine learning methods for predicting under-five mortality. vol-24
[7] Rakesh Kumar Saroj,Pawan Kumar Yadav, Rajneesh Singh3 (2022) Machine learning algorithms for understanding the determinants of under-five mortality, vol-15.
[8] Emmanuel Mfateneza, Pierre Claver Rutayisire, Emmanuel Biracyaza (2022)Application of machine learning methods for predicting infant mortality in Rwanda, vol-22.
[9] Muhammad Islam, Muhammad Usman , Azhar Mahmood (2020) Predictive analytics framework for accurate estimation of child mortality rates for internet of Things enabled smart healthcare systems. Vol. 16 (5).
[10] Brook Tesfayea, Suleman Atique b, Noah Elias a (2017) Determinants and development of a web-based child mortality prediction model in resource-limited settings, vol-14.
[11] Grant Nguyen (2016) Evaluating statistical and machine learning methods to predict risk of in-hospital child mortality in Uganda, University of Washington,vol-22.
[12] Md. Alif Sheakh, Taminul Islam, Mst. Sazia (2023) Child and maternal mortality risk factor analysis using machine learning approaches. IEEE Conference proceedings, pp.1-6.
[13] Andrea Bizzego, Giulio Gabrieli (2021) Predictors of contemporary under-5 child mortality in low and middle income countries , vol-18
[14] Nafiz Imtiaz Khan, Tahasin Mhumud (2020) Prediction of cesarean child birth using ensemble machine learning methods, pp,1-9
[15] Dawit S. Bogale, Tesfamaria, M. Abduhay (2022) Predicting perinatal mortality using homogenous ensemble machine learning methods, vol-22.