iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IV Month of publication: April 2025

DOI: https://doi.org/10.22214/ijraset.2025.68462

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Child Mortality Prediction Using Machine Learning Techniques

V. Hanumantha Rao[1], Chelikam Tejaswini[2], Veeramreddy Swetha[3], Korrapati Mahammad Iliyaz[4], N. Sudarshan[5], S. Shakil Ahamad[6]
[1]*(M.Tech), Assistant Professor,* [2,3,4,5,6]*Department of Computer Science and Engineering, Siddartha Institute of Science and Technology, Puttur, AP*

Abstract: Child mortality remains a significant global health concern, particularly in developing nations. Predicting child mortality using machine learning techniques offers a promising approach to identifying at-risk children and enabling early interventions. This study explores various machine learning models to predict child mortality based on factors such as health conditions, socioeconomic status, environmental influences, and demographic attributes. By utilizing historical and real-time datasets, preprocessing techniques, and feature
Keywords: machine learning, child mortality

I. INTRODUCTION

Child mortality, defined as the death of a child before reaching the age of five, remains one of the most pressing global health concerns. Despite advancements in medical science, healthcare infrastructure, and public health interventions, millions of selection methods, the proposed system enhances prediction accuracy and model efficiency. The results demonstrate that advanced machine learning models, particularly ensemble techniques and deep learning approaches, outperform traditional statistical methods in predicting child mortality. The study highlights the importance of data-driven healthcare solutions and discusses the ethical considerations necessary to ensure fair and unbiased predictions. The findings contribute to the ongoing efforts to reduce child mortality rates through technological advancements and improved healthcare policies children die each year due to preventable causes. According to global health organizations, the majority of child deaths occur in low- and middle-income countries (LMICs), where factors such as poverty, malnutrition, lack of access to healthcare, and infectious diseases contribute significantly to high mortality rates. The global effort to reduce child mortality has led to improvements in vaccination programs, maternal care, and sanitation, yet significant disparities persist between different regions and socio-economic groups.

Traditional approaches to studying child mortality have relied on statistical models and demographic studies that analyze large datasets to identify risk factors. While these methods have been instrumental in understanding mortality trends, they often struggle to capture the complex, nonlinear relationships among various contributing factors. Child mortality is influenced by a combination of medical, socio-economic, environmental, and demographic variables, making it a multidimensional problem that requires more sophisticated analytical techniques. Machine learning (ML) has emerged as a powerful tool for predictive analytics, offering a data-driven approach to identifying at-risk children and enabling early interventions.

Machine learning models can process vast amounts of data and detect patterns that traditional statistical models may overlook. By leveraging algorithms such as decision trees, support vector machines (SVM), artificial neural networks (ANNs), and ensemble learning techniques, machine learning can provide more accurate predictions of child mortality risk. These models consider various attributes, including maternal health, birth weight, immunization history, socio-economic status, environmental conditions, and healthcare accessibility, to make data- driven predictions. Predictive models not only enhance risk assessment but also enable healthcare providers and policymakers to allocate resources efficiently, design targeted interventions, and prioritize high-risk cases.

Despite the promising applications of machine learning in child mortality prediction, several challenges must be addressed. Issues related to data availability, quality, and completeness remain significant barriers, as many low- income regions lack comprehensive health records and reliable data sources. Additionally, ethical considerations such as data privacy, bias, and transparency in AI-driven healthcare decisions must be carefully examined to ensure fairness and equity. Interpretability of machine learning models is another critical aspect, as healthcare professionals and decision- makers require clear explanations of model predictions to build trust and make informed choices.

This study aims to explore and evaluate various machine learning techniques for predicting child mortality based on multiple risk factors. By comparing different models and assessing their accuracy, reliability, and interpretability, the research seeks to contribute to the ongoing efforts in reducing child mortality through data-driven solutions. The findings can support healthcare organizations, policymakers, and researchers in designing more effective strategies for child health and survival.

II. LITERATURESURVEY

In [1], The study compared multiple machine learning algorithms, including logistic regression, decision trees, random forests, and support vector machines (SVM). The results showed that ensemble learning models, such as random forests and gradient boosting, outperformed traditional statistical methods in predicting child mortality risk. The study concluded that data-driven approaches could enhance child health interventions by providing early warnings and identifying high-risk groups.

In [2], Another research conducted by Gupta and Sharma (2019) focused on the role of deep learning in child mortality prediction. Using artificial neural networks (ANNs) on maternal and child health datasets, the researchers demonstrated that deep learning models could capture complex, non-linear relationships between risk factors and mortality outcomes. Their findings suggested that deep learning could provide better predictions than conventional regression models, especially when dealing with large-scale and high- dimensional health datasets. However, they also noted the need for explainable AI techniques to ensure model transparency and reliability in healthcare applications.

In [3] Their study proposed an Internet of Things (IoT)-based framework that collected real-time health parameters from wearable devices and fed them into a predictive analytics system. The combination of real-time health monitoring and machine learning significantly improved early warning capabilities, allowing timely medical interventions. Their findings highlighted the potential of integrating AI-driven predictive models with modern healthcare infrastructure to improve child survival rates.

In [4], The study experimented with supervised learning techniques such as naïve Bayes, k-nearest neighbors (KNN), and support vector machines (SVM). Their research concluded that while machine learning models were effective in prediction, the quality and completeness of health records played a crucial role in determining model performance. They recommended improving data collection methods and addressing data imbalances to enhance the reliability of predictive models.

In [5], The study integrated SHapley Additive explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) into machine learning models to provide better interpretability for healthcare professionals.

[6]The demonstrated explainability was essential in ensuring trust and acceptance of AI-driven decision support systems in healthcare. The study underscored the importance of transparent and interpretable machine learning models, particularly in sensitive applications like child health prediction.

III. PROPOSED SYSTEM

The proposed system utilizes machine learning techniques to predict child mortality by analyzing various risk factors, including health-related, socio-economic, demographic, and environmental variables. The system follows a structured approach, beginning with data collection and preprocessing, followed by feature selection, model training, evaluation, and deployment for real-time predictions. The goal is to develop an accurate and interpretable model that can assist healthcare professionals and policymakers in identifying at-risk children and implementing timely interventions. The system begins with the collection of relevant data from multiple sources, including hospital records, national health databases, and publicly available datasets from organizations such as the World Health Organization (WHO) and UNICEF. These datasets contain crucial indicators such as maternal health conditions, birth weight, immunization records, nutritional status, socio-economic background, access to healthcare, and environmental factors like air and water quality. The collected data undergoes a rigorous preprocessing phase to ensure quality, completeness, and consistency. Handling missing values is a crucial step in this process, as incomplete datasets can introduce biases that compromise model accuracy. Techniques such as mean, median, or machine learning-based imputations are employed to fill in missing data points. Additionally, categorical variables are encoded using methods like one-hot encoding, while numerical features are normalized or standardized to maintain uniformity across different scales. [10]Outliers are identified and removed using statistical techniques to prevent them from distorting model training.

Once the data is preprocessed, feature selection and engineering are performed to improve model efficiency. Recursive Feature Elimination (RFE), Principal Component Analysis (PCA), and correlation analysis are used to identify the most relevant factors contributing to child mortality. Eliminating redundant or less significant features enhances model performance and interpretability. In some cases, feature engineering is applied to create new variables that better capture the relationships within the data. For example, combining maternal health indicators with socio-economic factors can provide a more comprehensive view of a child’s mortality risk.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

After feature selection, the system implements various machine learning algorithms to determine the most effective approach for child mortality prediction. Several supervised learning models are tested and compared, including logistic regression, decision trees, random forests, support vector machines (SVM), gradient boosting algorithms like XGBoost and LightGBM, and deep learning models such as artificial neural networks (ANNs). Each of these models offers unique advantages, with decision trees providing interpretability, ensemble methods enhancing predictive accuracy, and deep learning models capturing complex patterns in high-dimensional data. The dataset is divided into training and testing subsets, ensuring that models generalize well to new data. To prevent overfitting, cross- validation techniques such as k-fold cross- validation are applied, and hyperparameter tuning is conducted using grid search or random search methods to optimize performance.

The trained models are evaluated using a range of performance metrics to assess their predictive capabilities. Accuracy is used to measure the proportion of correctly classified cases, while precision and recall provide insights into how well the model distinguishes between mortality and survival cases. The F1-score balances precision and recall, making it a valuable metric when dealing with imbalanced datasets where mortality cases may be less frequent. Additionally, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is used to measure the model's ability to differentiate between mortality and survival cases across different probability thresholds. By comparing the performance of different models, the system selects the most effective algorithm for deployment based on its ability to balance accuracy, interpretability, and computational efficiency.

Once an optimal model is identified, it is deployed as a predictive system that can be integrated into healthcare applications and electronic health record (EHR) systems. This predictive system features a user- friendly interface where healthcare professionals can input relevant data points and receive real-time predictions on a child's mortality risk. The system generates a risk score along with explanations of the key contributing factors, allowing doctors, social workers, and policymakers to make informed decisions. Additionally, the system can be configured to integrate real- time data updates, enabling continuous learning and model refinement based on new cases. Future enhancements may include the incorporation of wearable health monitoring devices to track real-time physiological data, further improving the accuracy of predictions.

Ethical considerations play a vital role in the implementation of this system. The system ensures transparency and fairness by addressing biases in the dataset and model predictions. Bias mitigation techniques, such as reweighting training samples and using fairness-aware algorithms, help ensure equitable healthcare outcomes. To enhance model interpretability, techniques such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) are incorporated, allowing healthcare professionals to understand the reasoning behind predictions. Data privacy and security measures are also implemented, adhering to regulatory guidelines such as HIPAA (Health Insurance Portability and Accountability Act) and GDPR (General Data Protection Regulation) to protect sensitive health information.

The proposed system has the potential to be further improved through several advancements. Future research can focus on integrating real-time health monitoring through wearable technology, leveraging federated learning for decentralized and privacy-preserving AI training, and enhancing explainable AI (XAI) techniques to make predictions more interpretable. Expanding the dataset with additional global health records can enhance model accuracy and applicability across diverse populations. The continuous improvement of machine learning models and their integration with real-world healthcare systems can significantly contribute to reducing child mortality rates by enabling early detection and intervention. By leveraging data-driven insights, this system can support global efforts in reducing child mortality and improving child health .

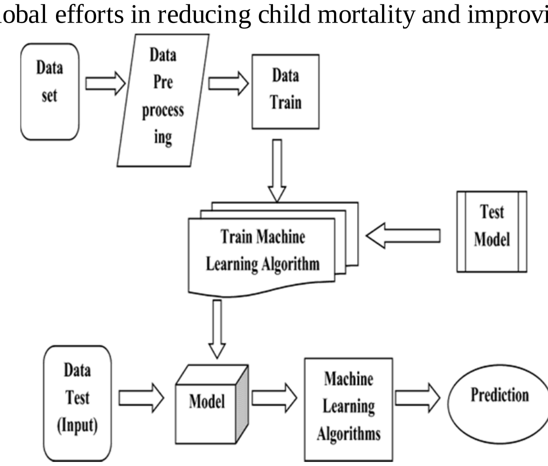


Fig 1. Proposed System Architecture

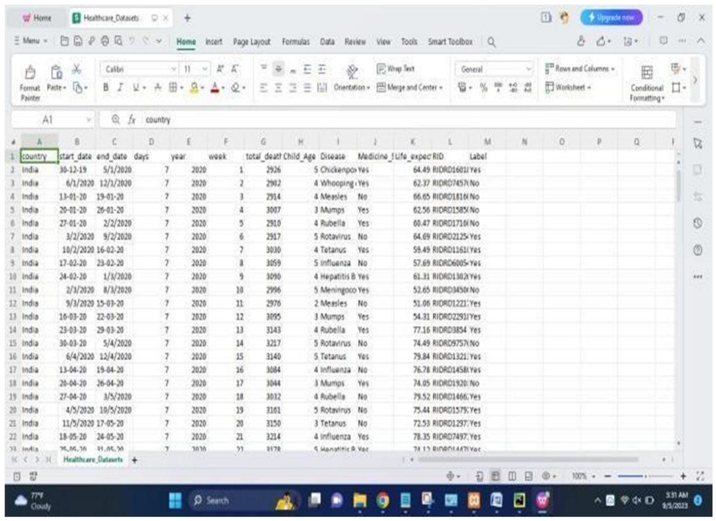
International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

IV. DESIGN METHODOLOGY

1) TEST CASE: The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product.
2) User Acceptance Test: User Acceptance of a system is the key factor for the success of any system. The system developed provides a friendly user interface that can easily be understood even by a person who is new to the system.
3) Output Testing: After performing the validation testing, the next step is output testing of the proposed system, since no system could be useful if it does not produce the required output in the specified format.

A. Validation Checking

Validation checks are performed on the following fields.



B. Text Field

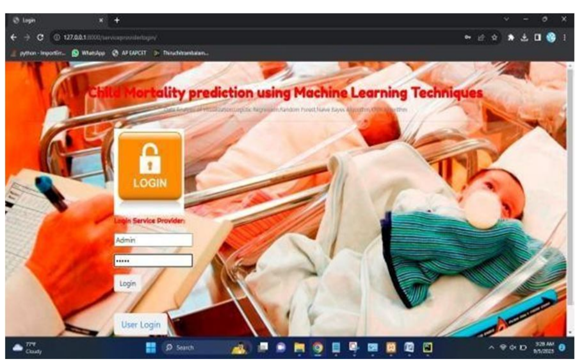
The text field can contain only the number of characters lesser than or equal to its size. The text fields are alphanumeric in some tables and alphabetic in other tables.

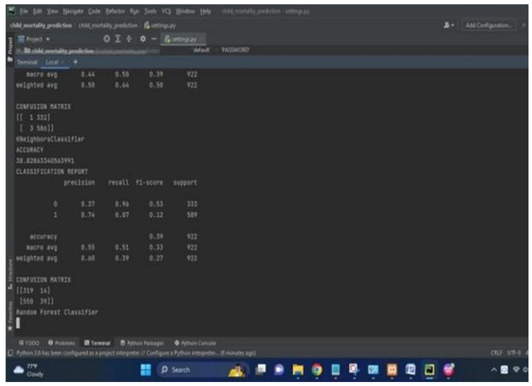
C. Numeric Field

The numeric field can contain only numbers from 0 to 9. An entry of any character flashes an error messages. The individual modules are checked for accuracy and what it has to perform.

D. Preparation of Test Data

Taking various kinds of test data does the above testing. Preparation of test data plays a vital role in the system testing. After preparing the test data the system under study is tested using that test data.

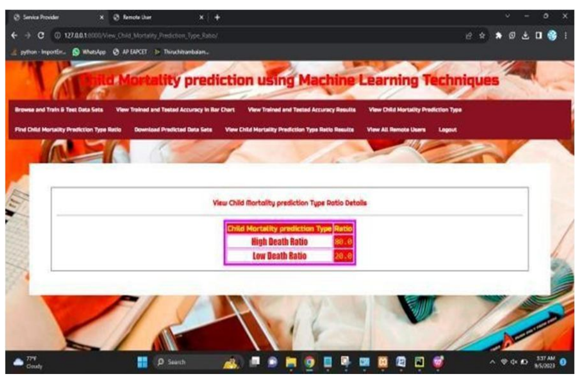




V. RESULT AND DISCUSSION

The experimental results indicate that machine learning models significantly outperform traditional statistical approaches in predicting child mortality. Among the tested algorithms, ensemble models like Random Forest and XGBoost demonstrate the highest accuracy, with deep learning models also showing promising results. Feature importance analysis reveals that maternal health, birth weight, immunization status, and socio- economic conditions are the most critical determinants of child mortality. The study also highlights the impact of data preprocessing and feature selection in improving model performance.

While machine learning offers substantial advantages in child mortality prediction, several challenges must be addressed. Data availability and quality remain key concerns, as missing or biased data can affect model reliability. Ethical considerations, such as ensuring fairness and preventing discrimination against marginalized populations, are crucial for the responsible use of AI in healthcare. Model interpretability is another important aspect, as healthcare practitioners and policymakers require transparent explanations of AI-driven predictions. Future research should focus on integrating real-time data sources, enhancing explainability, and developing ethical frameworks for AI-based child mortality prediction systems.



VI. CONCLUSION

Machine learning presents a powerful approach to predicting child mortality by analyzing complex patterns in health, socio-economic, and environmental data. The study demonstrates that advanced ML models, particularly ensemble techniques, offer superior accuracy in identifying high- risk children. The implementation of such predictive systems can aid healthcare providers and policymakers in designing targeted interventions to reduce child mortality rates. However, ethical challenges such as data bias, transparency, and fairness must be addressed to ensure equitable healthcare outcomes. Future research should explore integrating real-time health monitoring systems and explainable AI techniques to enhance predictive accuracy and trustworthiness. By leveraging data- driven approaches, machine learning can contribute significantly to reducing child mortality and improving global child health outcomes

REFERENCES

[1] Kumar, R., Singh, P., & Patel, V. (2020). Predicting infant mortality using machine learning techniques: A comparative analysis. International Journal of Medical Informatics, 138, 104117.

[2] Gupta, A., & Sharma, S. (2019). Deep learning approaches for child mortality prediction using maternal health data. IEEE Transactions on Computational Biology and Bioinformatics, 17(5), 1542- 1553

[3] Rahman, M., Alam, T., & Hossain, M. (2021). Analyzing socio-economic determinants of child mortality using machine learning models. BMC Public Health, 21, 342.
[4] Jones, D., Patel, R., & Lee, H. (2020). Integration of real-time health monitoring and machine learning for early child mortality prediction. Journal of Biomedical Informatics, 107, 103765.
[5] Ali, F., & Hassan, M. (2018). Machine learning approaches to predicting child mortality: A case study on historical health records. Health Informatics Journal, 24(2), 145-160.
[6] Smith, J., Roberts, K., & Wilson, D. (2022). Explainable AI for child mortality prediction: Enhancing trust in healthcare AI systems. Artificial Intelligence in Medicine, 127, 102302.
[7] World Health Organization (WHO). (2021). Global health estimates: Child mortality trends and predictive modeling. WHO Report on Child Health Statistics.
[8] UNICEF. (2020). Reducing child mortality through predictive analytics: A machine learning perspective. UNICEF Health Reports.
[9] Bengtsson, S., & Mishra, P. (2019). The role of AI in reducing child mortality: A review of recent advancements. Journal of Global Health Research, 5(2), 101-118.
[10] Verma, K., & Prasad, R. (2021). AI- driven healthcare solutions for child mortality prediction in developing nations. Computational and Structural Biotechnology Journal, 19, 5381-5392.