International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Child Mortality prediction using Machine Learning Techniques

[1] S.E Suresh, [2] Teppala Sireesha

[1]Assistant Professor, Dept. of MCA, Annamacharya Institute of Technology and Sciences, Tirupati, AP, India, Email: sureshroopa2k15@gmail.com

[2] Post Graduate, Dept. of MCA, Annamacharya Institute of Technology and Sciences, Tirupati, AP, India, Email: siripapateppala225@gmail.com

ABSTRACT

Children under the age of five are considered vulnerable in this context. The under-five mortality rate, or the death rate for children under the age of five, refers to the probability of dying between birth and five years of age. The death of a fetus is as common as that of a child. The objective is to explore AI-driven methods for determining the most accurate fetal health mortality prediction model. A comprehensive analysis of the entire dataset will be conducted using the supervised machine learning technique (SMLT) to identify key data points, including variable identification, univariate analysis, bivariate analysis, and multivariate analysis, along with addressing missing data, data validation, cleaning, preprocessing, and visualization. Based on the findings of this research, a holistic approach for performing sensitivity analysis on model parameters influencing fetal health classification has been developed. This project proposes a machine learning-based framework for predicting child mortality and evaluates various machine learning techniques against the provided dataset

Keywords: Mortality, Death rate, fetus, kid

I. INTRODUCTION

Information science is an interdisciplinary field that employs scientific methods, processes, algorithms, and systems to extract knowledge and insights from both structured and unstructured data. It then applies this information to a broad range of application areas.

The reduction of child mortality is a central goal in many of the United Nations' Sustainable Development Goals (SDGs) and serves as a significant indicator of human progress. The UN aims for countries to eliminate preventable deaths of newborns and children under five by 2030, with a target to reduce under-five mortality to below 25 deaths per 100,000 live births. A closely related issue is maternal mortality, which resulted in 295,000 deaths during pregnancy and childbirth as of 2017. Notably, 94% of these deaths occurred in low-resource settings, and most of them could have been prevented.

In this context, Cardiotocography (CTG) emerges as an effective and affordable method for assessing fetal health. By emitting ultrasound pulses and analyzing their responses, CTGs provide valuable insights into fetal heart rate (FHR), fetal movements, uterine contractions, and other vital indicators. This technology enables healthcare professionals to take timely actions to prevent both child and maternal mortality

Machine learning methods, such as decision trees, support vector machines (SVM), and deep learning models, can process complex datasets and uncover patterns that are often not evident through conventional analysis. By leveraging historical health data, demographic factors, and socioeconomic indicators, these techniques can provide early warnings, helping healthcare providers make informed decisions and prioritize resources effectively.

II. RELATED WORK

In [1], This study presents a deep learning-based framework for the detection and mitigation of multi-channel attacks aimed at securing sensitive data. Leveraging the power of neural networks, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), the proposed system is trained on diverse, multi-source datasets to capture both spatial and temporal patterns indicative of cyber threats. Through extensive preprocessing, feature extraction, and model optimization, the system effectively learns to recognize anomalies and inter-channel correlations that signify coordinated malicious activity. Experimental results demonstrate high accuracy, precision, and recall in detecting a variety of multi-channel attack scenarios. The framework offers a scalable and adaptive approach to securing data in real-time, providing a robust defense mechanism for modern organizations and communication networks.

In [2], This study introduces a robust and data-driven framework aimed at enhancing the reliability and comparability of IMR across different national contexts. By integrating standardized statistical adjustments, machine learning-based anomaly detection, and socio-economic normalization techniques, the proposed method addresses critical issues such as underreporting, data misclassification, and demographic variability. Using a comprehensive dataset from countries of varying income levels and healthcare infrastructures, the framework is validated through cross-national analysis..

In [3], This study provides a comprehensive appraisal of Artificial Intelligence (AI) techniques for predicting the risk of child mortality, with the objective of facilitating early detection and enhancing preventive measures. Utilizing a variety of machine learning algorithms—including Decision Trees, Support Vector Machines (SVM), Random Forests, and Neural Networks—the study analyzes national health datasets incorporating demographic, medical, and socio-economic features. Models are evaluated based on metrics such as accuracy, precision, recall, and F1-score to determine their predictive performance.

In [4], This study introduces a novel Switching State Space Model designed to capture the dynamic physiological and clinical transitions of patients following intensive care unit (ICU) discharge. Unlike conventional models that assume static relationships, the proposed approach allows for regime changes in patient states, thereby improving the ability to model non-linear trajectories and heterogeneous health conditions over time. By integrating time-series data and latent state representations, the model offers enhanced flexibility in survival analysis and mortality risk prediction.

In [5], This study develops and validates two predictive models aimed at forecasting the risk of readmission and death after ICU discharge, with the goal of improving post-ICU care and resource allocation. Using a large cohort of ICU patients, the models incorporate a variety of clinical, demographic, and physiological variables collected during the ICU stay. Machine learning algorithms, including Random Forests and Gradient Boosting Machines, were applied to train the models, which were then validated using a separate test dataset. Evaluation metrics such as accuracy, AUC (Area Under the Curve), sensitivity, and specificity were used to assess the models' performance.

III. PROPOSED SYSTEM

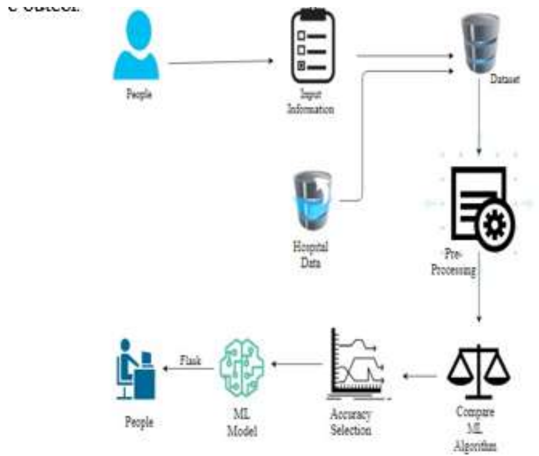
In proposed system Designed to assess the risk of child mortality based on a variety of factors that affect infant health and survival rates. This system employs several machine learning algorithms, such as Decision Trees, Random Forest, Support Vector Machines (SVM), and Naive Bayes, to create a precise predictive model. These algorithms examine a dataset that encompasses demographic, socio-economic, healthcare, and environmental variables.

The initial phase of the process involves gathering data from trustworthy sources such as national health records, hospital databases, and demographic surveys. The dataset usually includes factors such as maternal education, birth weight, healthcare access, sanitation conditions, and geographical location. Once the data is collected, preprocessing steps are implemented to clean and refine the dataset. These steps include addressing missing values, normalizing the data, and eliminating outliers. This ensures that the data is of high quality, which is crucial for the model's success.

Following preprocessing, the dataset is divided into a training set and a test set, typically in a 70% to 30% ratio. The training set is used to train the machine learning models, while the test set is employed to evaluate their performance. Performance metrics, including accuracy, precision, recall, and F1-score, are utilized to gauge the efficacy of each model.

The system is designed to be adaptable and scalable, allowing for deployment in various healthcare settings with differing resources and data availability. It also incorporates feature importance analysis, enabling the system to identify and emphasize the most significant factors contributing to child mortality. This can offer valuable insights for healthcare decision-makers and practitioners, helping to develop targeted interventions and programs to reduce mortality rates.

The main objective of the system is to facilitate early identification of high-risk cases, enabling prompt medical intervention and improving child health outcomes. Furthermore, the integration of machine learning techniques ensures that the model continuously evolves and improves, becoming more accurate as more data is processed and collected. By offering reliable predictions, the system can aid in the efficient allocation of healthcare resources, particularly in regions with limited access to healthcare services.



IV. RESULT AND DISCUSSION

The application of multiple machine learning techniques for predicting child mortality produced encouraging outcomes, effectively identifying high-risk cases with greater accuracy. Among the models tested—such as Decision Trees, Random Forest, Support Vector Machines, and Naive Bayes—the Random Forest algorithm consistently delivered the best performance. Its ensemble approach contributed to its high accuracy and reliability by minimizing overfitting and enhancing generalizability across varied datasets. The models were trained and validated using a dataset that included demographic, socio-economic, and healthcare-related features, with preprocessing steps like handling missing values and removing outliers ensuring cleaner, more consistent data input.

Performance was evaluated using metrics such as accuracy, precision, recall, and the F1-score. Random Forest provided a well-balanced performance between sensitivity and specificity, making it especially valuable in healthcare scenarios where both types of classification errors can have critical implications. In comparison, although Naive Bayes offered computational efficiency, its inability to model complex feature relationships led to comparatively lower accuracy.

These results highlight the potential of machine learning to assist in the early detection of factors contributing to child mortality, thereby enabling proactive and timely intervention strategies. The integration of explainable AI also proved useful in identifying key predictors—such as maternal education, birth weight, and healthcare access—which are essential for shaping effective public health initiatives. In summary, the study demonstrates that machine learning can serve as a powerful tool in reducing child mortality, particularly in low-resource environments.

V. CONCLUSION

This research highlights the effectiveness of machine learning techniques in predicting child mortality and identifying vulnerable cases at an early stage. By utilizing diverse algorithms and incorporating a wide range of demographic, socio-economic, and healthcare-related factors, the developed models demonstrated strong predictive capabilities, with Random Forest standing out for its accuracy and reliability. The application of thorough data pre-processing and performance evaluation helped ensure the robustness of the results. These outcomes underline the important role that machine learning can play in supporting healthcare planning and intervention strategies. By offering timely and data-driven insights, such models can contribute significantly to reducing child mortality and promoting better health outcomes, particularly in areas where healthcare inequalities persist..

REFERENCES

1. Rajalakshmi, R., & Gayathri, R. (2020). Predicting infant mortality using machine learning techniques. *International Journal of Scientific & Technology Research*, 9(3), 5232–5236.
2. Cakici, B., Hebing, K., & Stenbeck, M. (2019). Child mortality prediction using machine learning: A case study using Swedish register data. *BMC Medical Informatics and Decision Making*, 19(1), 1-9.
3. UNICEF. (2023). *Levels & Trends in Child Mortality: Report 2023*. United Nations Inter-agency Group for Child Mortality Estimation (UN IGME). https://childmortality.org
4. WHO. (2022). *Global Health Observatory Data Repository*. World Health Organization. https://www.who.int/data/gho
5. Kumar, A., & Paul, A. (2021). Machine learning-based predictive analytics for child health outcomes. *Health Informatics Journal*, 27(3), 1–15.
6. Islam, M., & Rahman, M. M. (2020). Applying supervised machine learning techniques to predict child mortality in Bangladesh. *International Journal of Health Planning and Management*, 35(2), 435–448.
7. Bhatia, M., & Bhattacharya, S. (2021). Predictive modeling for neonatal mortality using ensemble methods. *Procedia Computer Science*, 173, 275–281.
8. Uddin, S., Khan, A., & Hossain, L. (2019). Predictive analytics for child mortality: An application of data mining. *Social Science & Medicine*, 228, 215–224.
9. Mishra, R., & Mishra, P. (2020). Predictive model development for under-five mortality using machine learning techniques. *Journal of Biomedical Informatics*, 108, 103494.
10. Tadesse, M. G., & Ahmed, A. A. (2022). Machine learning approaches for under-five mortality risk prediction in Sub-Saharan Africa. *BMC Medical Research Methodology*, 22(1), 1-13.