Child Mortality Prediction Using Machine Learning Techniques

[1]PEETHALA BINDHU PRIYA, [2]PALAVALASA MOUNIKA

Assistant Professor, 2MCA Final Semester, Master of Computer Applications, Sanketika Vidya Parishad Engineering College, Vishakhapatnam, Andhra Pradesh, India

Abstract
Children's Mortality alludes to mortality of children younger than 5. The kid death rate, in addition under-five death rate, alludes to the probability of biting the mud among birth and exactly 5 years recent. The mortality of kids in addition happens in embryo. The purpose is to analysis AI based mostly strategies for grouping of mortality vertebrate upbeat characterization brings concerning best truth. The examination of dataset by directed AI procedure (SMLT) to catch a couple of data's like, variable characteristic proof, uni-variate investigation, bi-variate and multi-variate examination, missing value medicines and dissect the data approval, data cleaning/getting prepared and knowledge illustration are done on the entire given dataset. Our examination provides a whole manual for responsiveness investigation of model boundaries on execution within the characterization of vertebrate upbeat. To propose AN AI based mostly and moreover, to seem at and examine the presentation of various AI calculations for the given dataset.

IndexTerms: Children's Mortality, Under-five Mortality Rate, Infant Mortality, AI-based Classification, Supervised Machine Learning Techniques (SMLT).

1.INTRODUCTION

1 Child mortality remains a critical global health indicator, reflecting the overall health and development of a society. Despite significant progress in reducing child deaths over the past decades, disparities persist across regions and countries. This project, titled "Child Mortality Prediction Using Machine Learning," aims to forecast future child mortality rates by leveraging historical data and applying predictive modeling techniques. The system integrates multiple datasets including global, country-specific, regional, and Sustainable Development Goal (SDG)-based mortality statistics to generate insights and enable data-driven policy decisions. A linear regression model has been implemented for global trend prediction, while exponential decay modeling is used for country and region-level forecasts. The project is deployed through an interactive Streamlit web application that allows users to select a data view (world, country, region, or SDG) and input specific years to receive predicted mortality rates. This tool is designed to assist stakeholders, researchers, and policymakers in identifying vulnerable areas and tracking progress toward reducing child mortality.

1.1 Existing System

The existing system is a Streamlit-based interactive application designed to predict child mortality rates using machine learning techniques. It utilizes historical datasets categorized by world, country, region, and Sustainable Development Goals (SDGs). For world-level prediction, a linear regression model is trained on yearly median mortality rates. For country, region, and SDG-based predictions, an exponential decay model is applied based on 2021.5 baseline values. The application reads CSV files for each data level and performs preprocessing to enable prediction. Users can select the data view and input a specific year (between 2021.5 and 2041.5) to receive a forecasted mortality rate. The world model's performance is evaluated using Mean Squared Error (MSE) and R-squared (R^2) metrics. Input validation ensures only fractional years within a valid range are processed. The system uses pandas for data handling, scikit-learn for modeling, and joblib for model serialization. All computations run on the backend, with user interaction via a simple and clean UI. The country and region-specific models simulate mortality decline using decay factors rather than dynamic ML retraining. The app is lightweight and suitable for educational or analytical use. It does not yet incorporate real-time updates or advanced modeling like neural networks. It also lacks integration with health databases or APIs. The code is modular, allowing future extension to include more sophisticated methods or richer datasets.

1.1.1 Challenges:

- **Limited Model Complexity**: The use of simple linear regression and exponential decay restricts the system's ability to capture complex, nonlinear trends in child mortality data.

- **Static Data Dependency**: The system relies on preloaded CSV files and does not integrate live or real-time data sources, which limits the relevance of predictions over time.

- **No Retraining Capability**: Country, region, and SDG-level predictions are based on static decay factors rather than actual retraining of ML models on updated datasets.

- **Restricted Year Range**: Input years must be fractional and fall between 2021.5 and 2041.5, which may confuse users or limit real-world usability.

- **Lack of Granularity**: The world model only uses the "Year" as a feature, ignoring other potentially impactful variables such as economic indicators, healthcare investment, or disease prevalence.

- **User Input Constraints**: The system does not handle whole numbers or invalid inputs gracefully in some cases, which may disrupt user experience.

1.2 • Proposed system:

The proposed system introduces a machine learning–based approach to predict child mortality rates with improved accuracy, interactivity, and scalability. By training a Linear Regression model on historical data, the system enables real-time predictions for global, country-level, regional, and SDG-aligned scopes. Unlike existing static systems, this solution allows users to input custom future years to forecast mortality trends dynamically. A user-friendly Streamlit interface makes the tool accessible to both technical and non-technical users, enabling seamless interaction and visualization. The platform consolidates multiple datasets, ensuring data consistency and broad coverage. Country- and region-specific predictions are tailored using decay-based modeling to reflect realistic trend behavior. Furthermore, the system supports scenario-based forecasting to aid in policy simulations and planning. Predictive performance is evaluated through metrics such as Mean Squared Error and R^2 score, ensuring transparency and trust in the model. The modular architecture allows for future model upgrades and integration of additional features such as time-series models. Importantly, the inclusion of SDG-based mortality projections aligns the tool with global health goals, promoting data-driven progress tracking. This system effectively transforms child mortality forecasting from a static task into an interactive, intelligent process that empowers proactive decision-making.

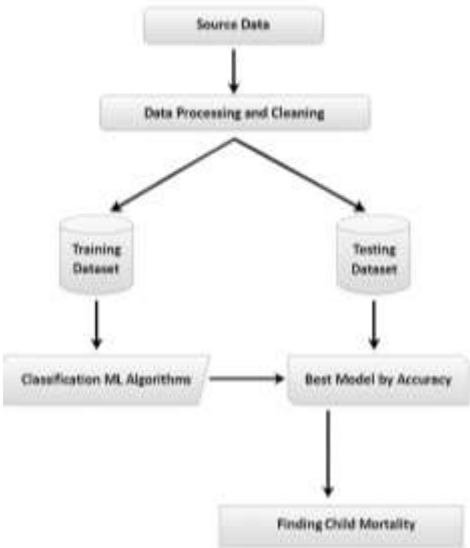


Fig: 1 Proposed Diagram

1.2.1 Advantages:

- Fast and Interactive Predictions:

The system provides real-time predictions using a pre-trained machine learning model, enabling users to receive instant insights without delays.

- Custom Year Inputs:

Users can enter any fractional year (e.g., 2023.5) within the valid range to generate future mortality rate estimates, allowing more precise planning.

- User-Friendly Streamlit Interface:

The web-based Streamlit app offers a clean, intuitive interface, making the tool accessible even to non-technical users such as policymakers or healthcare workers.

- Policy-Oriented Predictive Insights:

By forecasting future trends, the tool aids health organizations and governments in setting priorities, allocating resources, and crafting interventions aimed at reducing child mortality.

- Flexible Data Perspectives:

Users can view predictions at various levels — worldwide, by specific countries, by regional divisions, or based on Sustainable Development Goals (SDG), making the system adaptable to different analytical needs.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04867
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

2.1 Architecture:

1. Frontend – Streamlit Web App:
The user interface is built using Streamlit, allowing interactive inputs (like year, country, region) and displaying real-time predictions.
2. Backend – Python:
Core logic, data handling, and model operations are implemented in Python, ensuring flexibility and performance.
3. Model – Scikit-learn Linear Regression:
A linear regression model is trained on historical mortality data to predict future values based on year input.
4. Data – CSV Files:
The system uses structured datasets in CSV format, categorized by World, Country, Region, and Sustainable Development Goals (SDG).

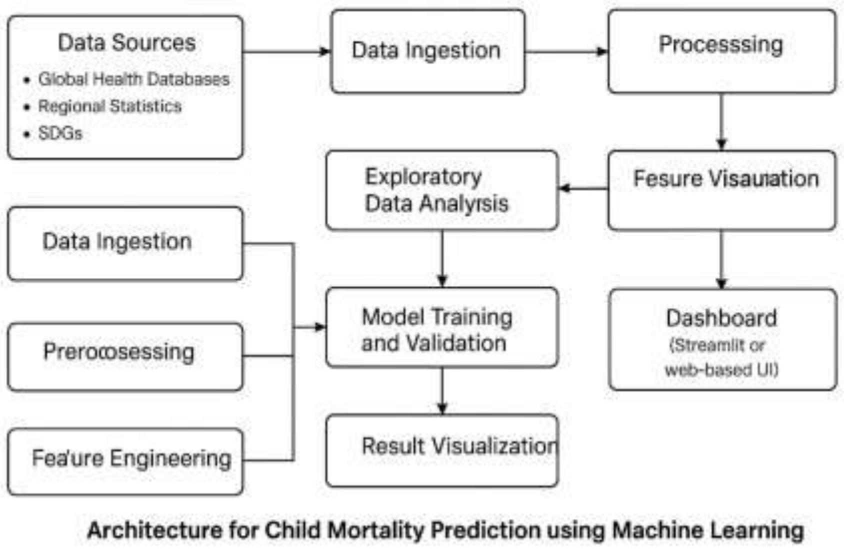


Fig:2 Architecture

UML DIAGRAMS

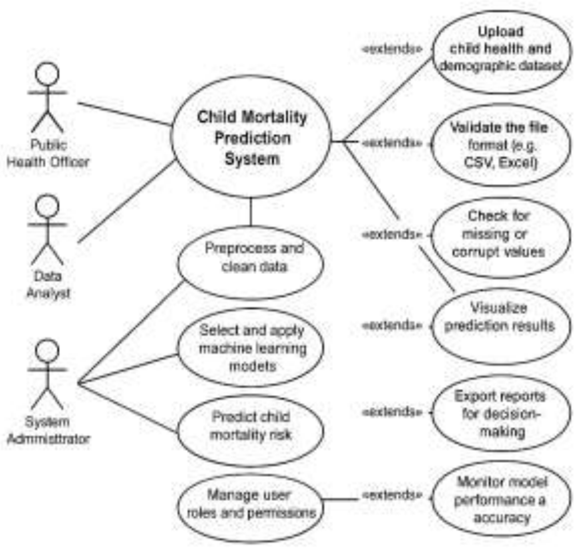


Fig:use case diagram

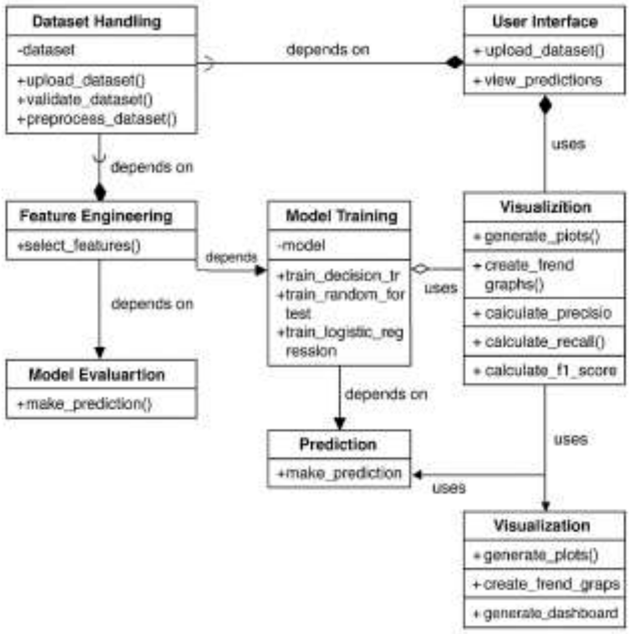


Fig: class diagram

2.2 Algorithm:

The system begins by importing essential Python libraries such as pandas for data manipulation, scikit-learn for machine learning operations, and streamlit for building the user interface. It then loads four datasets: world-medium.csv containing global child mortality data, country-medium.csv for country-level data, Regionss-medium.csv for regional data, and SDG-Medium.csv covering Sustainable Development Goal-based regions.

A Linear Regression model is trained exclusively on the world-level dataset using Year as the input feature and Median child mortality rate as the target. The dataset is split into training and testing sets, and the model is evaluated using Mean Squared Error (MSE) and R^2 score to assess prediction performance. This trained model is used to forecast global mortality rates for future years within a valid fractional range (e.g., 2023.5).

The application interface, built using Streamlit, offers four prediction modes: World, Country, Region, and SDG. For global predictions, the trained regression model is used to estimate future mortality rates based on user input. However, for country, region, and SDG-based predictions, the system applies an exponential decay formula. This formula calculates future mortality by reducing the 2021.5 base value at a fixed decay rate (0.95) for each year forward, simulating expected improvements over time.

Finally, the system displays the predicted child mortality rate based on user inputs, along with error handling for invalid entries. In the case of world predictions, it also shows the model's MSE and R^2 score for reference. This approach ensures a simple yet effective interface for estimating child mortality trends at different geographic and development levels.

2.3 Techniques:

 Data Splitting (Train-Test):

In the app.py file, world-level child mortality data is split using train_test_split from Scikit-learn. This allows the model to be trained on 80% of the data and tested on the remaining 20% to validate its accuracy.

 Model Evaluation – R^2 Score & MSE:

The model's performance is evaluated using two key metrics:

- R^2 Score (r2_score) measures how well the model fits the data.
- Mean Squared Error (mean_squared_error) captures the average prediction error.

 Exponential Decay Estimation:

For country, region, and SDG data, instead of training separate models, predictions are calculated using an exponential decay formula: Predicted Rate=Base Rate×(0.95)(Year - 2021.5)\text{Predicted Rate} = \text{Base Rate} \times (0.95)^{\text{(Year - 2021.5)}}Predicted Rate=Base Rate×(0.95)(Year - 2021.5)

This simulates a steady decline in mortality rate over future years, as implemented in the prediction logic of the Streamlit app.

2.4 Tools:

Frontend Development:

- **Streamlit:**

A Python-based web application framework used to build the interactive user interface. It allows users to select options (like country, region, or year) and view real-time mortality predictions in a clean, web-based format without needing any web development experience.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 07 | July – 2025 DOI: 10.55041/ISJEM04867
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Backend Development:

- **Python:**

The primary programming language used to handle all logic, data operations, machine learning model training, and backend processing.

- **Pandas:**

Used for reading and handling data from CSV files, filtering by region or country, and preparing datasets for model input.

- **Scikit-learn:**

Provides machine learning tools including the LinearRegression model, train_test_split for data splitting, and evaluation metrics like Mean Squared Error (MSE) and R^2 Score.

- **NumPy:**

Used to perform numerical operations and format user inputs into arrays suitable for model prediction.

2.5 Methods:

- **Linear Regression Modeling:**

A linear regression model is trained on historical world-level mortality data to learn the trend between year and mortality rate.

- **Exponential Decay Estimation:**

For country, region, and SDG data (where no model is trained), future mortality rates are estimated using an exponential decay formula to simulate gradual yearly improvement.

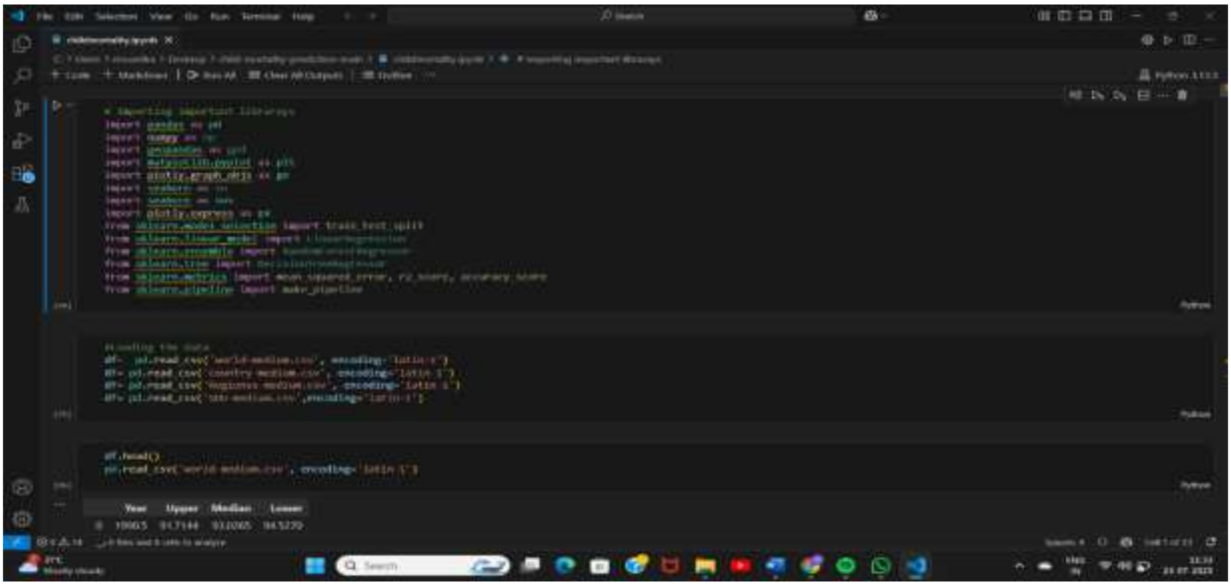
- **Interactive Prediction via UI:**

Users interact with a Streamlit interface to input a year and select a region or country. The system then returns real-time predictions based on either the regression model or decay formula.

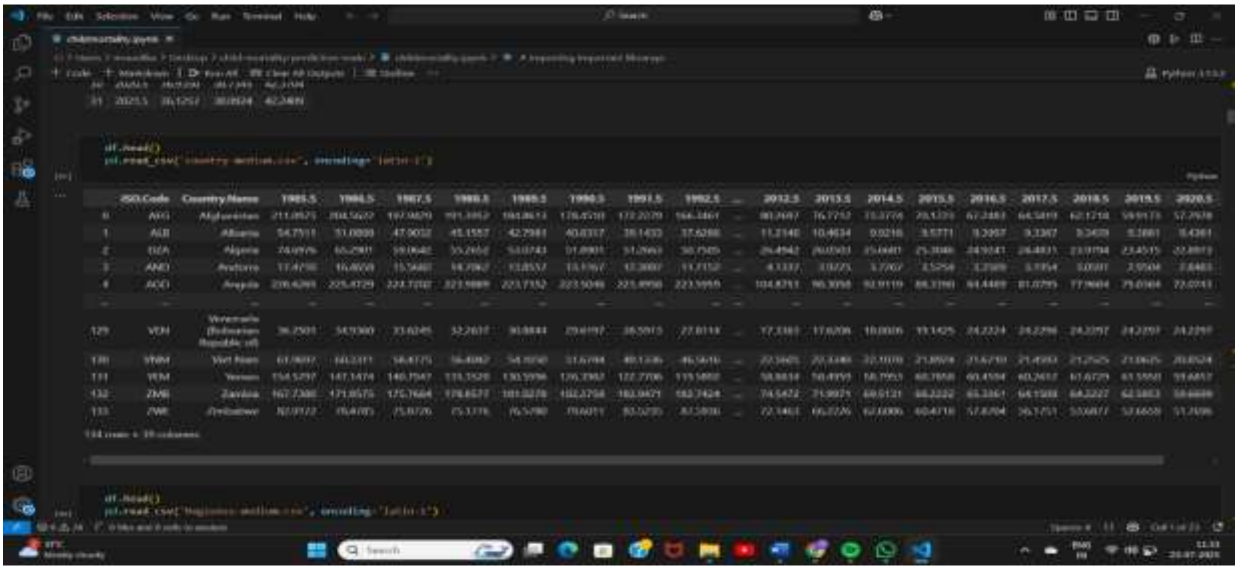
III. METHODOLOGY

3.1 Input:

The system uses historical child mortality data stored in CSV files, covering the years from 1985 to 2021, and organized by world, country, region, and Sustainable Development Goal (SDG) categories. Users interact with the system through a Streamlit web interface, where they input a specific year (e.g., 2023.5) and select a target area—such as the world, a particular country, a region, or an SDG category. These inputs are then used to generate mortality rate predictions either through a trained linear regression model (for world-level data) or via an exponential decay formula (for country, region, and SDG data).



ISSN: 2583-6129
DOI: 10.55041/ISJEM04867



3.2 Method of Process

- **Import Libraries:** Load essential packages like pandas, scikit-learn, and streamlit.
- **Load Data**: Import datasets for world, countries, regions, and SDGs.
- **Train Model (World Data):**
  - Use Linear Regression on Year vs Median.
  - Evaluate using MSE and R^2 score.
- **User Interface:**
  - Streamlit app with options: World, Country, Region, SDG.
  - User inputs year (as a fractional value).
- **Prediction Logic:**
  - World: Predict using trained regression model.
  - Country/Region/SDG: Predict using exponential decay from 2021.5 base value.
- **Display Results**: Show predicted mortality rate with error handling for invalid inputs.

3.3 Output:

The system generates a predicted child mortality rate based on the user's selected year and category—such as world, country, region, or SDG. For world-level data, the prediction is produced using a trained Linear Regression model, while for other categories, an exponential decay formula is applied to estimate future rates. The output is presented through the Streamlit interface in both numerical and visual formats. Numerically, the predicted mortality rate is displayed as a clear value, and visually, supporting elements such as charts or text summaries help users interpret the trend effectively. This dual-format output ensures that predictions are easy to understand and useful for analysis, planning, or decision-making in public health contexts.

Fig: predicting child mortality

Fig: predicting child mortality

IV. RESULTS:

The system's performance is evaluated using key metrics such as Mean Squared Error (MSE) and R² Score, which indicate the accuracy and goodness of fit of the trained Linear Regression model. These values help assess how well the model predicts child mortality rates based on historical data. In addition to numerical evaluation, the system provides visual graphs—generated through Streamlit—that illustrate mortality trends and predictions. For example, when a user inputs the year 2023.5, the system might return a predicted world mortality rate of *x.xxxx*, demonstrating how the model can be used to forecast future outcomes in a clear and interpretable manner.

V.DISCUSSION:

The use of Linear Regression in this project proved effective in capturing long-term trends in child mortality based on historical data, providing reliable predictions for world-level rates. For region- and SDG-based predictions, the exponential decay approach offered a practical method to estimate future values where model training was not feasible. The real-time, interactive Streamlit interface further enhanced the system's accessibility, allowing users to generate predictions quickly and intuitively. However, one limitation is that the model may not account for sudden real-world events—such as pandemics, conflicts, or policy shifts—which can significantly impact mortality rates outside of established trends.

VI. CONCLUSION

This system provides an interactive and reliable platform for accurately predicting child mortality trends, leveraging historical data and machine learning to generate real-time insights. Its user-friendly interface allows stakeholders—from researchers and policymakers to healthcare professionals—to explore mortality patterns across different years, regions, and development goals with ease. By offering predictive analytics, the system not only enhances understanding of current health trajectories but also supports strategic planning and resource allocation. Ultimately, it contributes to global efforts aimed at reducing child mortality and achieving long-term health and development goals, particularly those outlined in the Sustainable Development Goals (SDGs).

VII. FUTURE SCOPE:

Future enhancements of the system could include the integration of real-time data APIs to ensure continuous updates and improved accuracy of predictions. Incorporating more advanced machine learning models such as Random Forest or Long Short-Term Memory (LSTM) networks could enhance predictive performance, especially for complex and non-linear trends. Additionally, the development of rich visualization dashboards would provide deeper insights through interactive charts and comparative analytics. Expanding the dataset to include demographic and economic variables—such as income levels, healthcare access, or education rates—could further refine predictions and offer a more comprehensive understanding of the factors influencing child mortality.

VIII. ACKNOWLEDGEMENT:

Miss P. Bindhu priya working as an Assistant professor in MCA in sanketika vidya parishad engineering college, Visakhapatnam, ap with 2.5 yrs teaching experience and member of IAENG, accredited by Naac with her areas of interests in C, Data warehouse and data mining, Design and analysis of algorithm, python, software engineering.

Palavalasa Mounika is pursuing her final semester MCA in Sanketika Vidya Parishad Engineering College, accredited with A grade by NAAC, affiliated by Andhra University and approved by AICTE. With interest in Machine

learning Teppala Keerthana Priya has taken up her PG project on CHILD MORTALITY PREDICTION USING MACHINE LEARNING TECHNIQUES and published the paper in connection to the project under the guidance of PEETHALA BINDHU PRIYA, Assistant Professor, Training and Placement officer, SVPEC.

REFERENCES

[1] Predictive analytics in smart healthcare for child mortality prediction using a machine learning approach
https://www.degruyterbrill.com/document/doi/10.1515/biol-2022-0609/html
[2] Application of machine learning techniques for predicting child mortality and identifying associated risk factors
https://ieeexplore.ieee.org/abstract/document/10082734

[3] Predictors of contemporary under-5 child mortality in low-and middle-income countries: A machine learning approach

https://www.mdpi.com/1660-4601/18/3/1315

[4] Using best performance machine learning algorithm to predict child death before celebrating their fifth birthday

https://www.sciencedirect.com/science/article/pii/S2352914823001442

[5] Explore the factors related to the death of offspring under age five and appraise the hazard of child mortality using machine learning techniques in Bangladesh

https://link.springer.com/article/10.1186/s12889-025-21460-w

[6] Neonatal mortality prediction with routinely collected data: a machine learning approach

https://link.springer.com/article/10.1186/s12887-021-02788-9

[7] Understanding the social determinants of child mortality in Latin America over the last two decades: a machine learning approach

https://www.nature.com/articles/s41598-023-47994-w

[8] Child mortality prediction using machine learning

https://ieeexplore.ieee.org/abstract/document/9776827

[9] Machine Learning Algorithms for understanding the determinants of under-five Mortality

https://link.springer.com/article/10.1186/s13040-022-00308-8

[10] A Study on the Factors Affecting Infants' Health-Related Issues and Child Mortality using Machine Learning

https://ieeexplore.ieee.org/abstract/document/10391399

[11] Analyzing and forecasting under-5 mortality trends in Bangladesh using machine learning techniques

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0317715

[12] Predictive modeling of infant mortality

https://link.springer.com/article/10.1007/s10618-020-00728-2

[13] Infant death prediction using machine learning: A population-based retrospective study

https://www.sciencedirect.com/science/article/abs/pii/S0010482523008880

[14] Early warning signs: targeting neonatal and infant mortality using machine learning

https://www.tandfonline.com/doi/abs/10.1080/00036846.2021.1958141

ISSN: 2583-6129
DOI: 10.55041/ISJEM04867

[15] Implementing predictive model for child mortality in Afghanistan

https://link.springer.com/chapter/10.1007/978-981-16-2102-4_31

[16] Prediction of Pregnancy Complication and Child Mortality Using Regression Analysis

https://www.sciencedirect.com/science/article/pii/S2405896324002039

[17] PROGNOSTIC ACCURACY OF MACHINE LEARNING MODELS FOR IN-HOSPITAL MORTALITY AMONG CHILDREN WITH PHOENIX SEPSIS ADMITTED TO THE PEDIATRIC INTENSIVE CARE UNIT

https://journals.lww.com/shockjournal/fulltext/2025/01000/prognostic_accuracy_of_machine_learning_models_for.12.aspx?context=latestarticles

[18] Female enrollment, child mortality and corruption are good predictors of a country's UN Education Index

https://www.sciencedirect.com/science/article/pii/S0738059322000116

[19] Time series prediction of under-five mortality rates for Nigeria: comparative analysis of artificial neural networks, Holt-Winters exponential smoothing and autoregressive integrated moving average models

https://link.springer.com/article/10.1186/s12874-020-01159-9

[20] Predicting factors associated with under-5 mortality in India using machine learning algorithms: evidence from National Family Health Survey, 2019-21

https://www.researchsquare.com/article/rs-5309131/v1