Leveraging Health Data to Predict Infant Survival and Wellbeing from Pregnancy to Early Childhood.

Anchal Daga, Keerthana Vangala, Grace Evangelene Avula Lael, Sharanya Nallapeddi

Abstract - *This project is based on the use of health data to understand and predict the factors that have an influence on infant survival and well-being from pregnancy into early childhood. By analyzing variables such as maternal age, access to prenatal care, vaccination rates, low birth weights, and teen pregnancies, patterns may be determined that affect health outcomes. Using valid data from sources such as the CDC, the project hopes to create predictive models that can determine hotspots. These insights will aid healthcare providers and policymakers in resource distribution, such as vaccines and maternal support services, to help reduce preventable infant deaths and improve overall care in vulnerable communities. The aim of this initiative is to develop a framework that will support equitable health care through early risk identification and efficient distribution of available resources. The project will target the need for ensuring that critical interventions-such as vaccination drives or prenatal programs-reach only those areas that are in need. This data-driven approach empowers healthcare systems to make informed decisions and contributes to reducing disparities, enhancing maternal and infant health, and building stronger, healthier communities.*

I. INTRODUCTION

SAFEGUARDING infants from pregnancy through early childhood remains a critical global health concern. Complex factors—ranging from maternal health conditions and varying levels of prenatal and postnatal care to immunization coverage, nutrition programs, and socio-economic disparities—collectively sustain gaps in maternal and infant outcomes, even as medical science continues to advance. For instance, low birth weight, a rise in teen pregnancies, and inadequate healthcare services in rural areas often lead to preventable complications during labor and early development. Healthcare systems frequently struggle to identify high-risk locations and distribute resources effectively, a problem intensified by the difficulty of interpreting and utilizing massive volumes of health data. As a result, many at-risk communities miss crucial interventions that could avert infant mortality and other negative health outcomes.

In response, this project aims to create predictive models to pinpoint key factors influencing infant survival and well-being, utilizing credible health data sources like https://databank.worldbank.org/. By enabling policymakers and healthcare professionals to derive meaningful insights from these analyses, the models can direct resource allocation, spotlight immunization priorities, and highlight areas needing improved healthcare facilities. Ultimately, this data-driven approach is intended to strengthen strategies that lead to better health outcomes for newborns and their families across diverse regions.

Real Time Example - *Imagine a remote village where a young mother-to-be struggles to access prenatal check-ups because the nearest health center is hours away. She also lacks consistent immunization support, making both her and her unborn baby vulnerable to avoidable health risks. This situation highlights the bigger global issue: how can health systems pinpoint such high-risk areas and effectively allocate resources, so both mother and infant get the best possible start in life?*

II. DATASET AND ITS FEATURES

By aggregating information from multiple open-source platforms (with a principal dataset derived from the World Bank Data Bank), the present study incorporates an extensive array of demographic, socioeconomic, and healthcare variables. Given the sheer size and heterogeneity of the collected data, a series of systematic data-cleaning and feature-engineering procedures was carried out to ensure accuracy and applicability for further analysis. The steps detailed below outline the methodology used to prepare the dataset for subsequent statistical modeling and inference.

DATA CLEANING METHODOLOGY

Duplicates were initially removed with the drop_duplicates() function to prevent redundant rows from skewing subsequent analyses. Non-essential columns (such as "Year Code") were then excluded, and rows with broad or unspecified region labels (e.g., "High income," "Low income," "World") were discarded to maintain a clear focus on specific geographical areas. Columns that should only contain non-negative values, such as life expectancies and numbers of deaths, were converted to numeric types, and any negative entries were eliminated. Categorical variables were standardized using a LabelEncoder to ensure compatibility with machine learning techniques, and the "Year" field was converted to a uniform numeric year format via pd.to_datetime(), enabling effective time-series evaluation.

Ambiguous column names (e.g., [SP.DYN.CBRT.IN]) were replaced with more descriptive labels (e.g., "Birth rate, crude (per 1,000 people)"), while instances of ".." indicating missing values were addressed by substituting region-level means if less than 70% of a region's data were missing, or by inferring values from similar regions if the missing rate exceeded that threshold. Heavily affected columns, such as "Newborns protected against tetanus (%)," were imputed using mean values. Several composite features were introduced to strengthen analytical depth, including "Birth-Death Ratio," "Immunization Efficacy," and various mortality-based ratios (e.g., neonatal, infant, female-to-male, and maternal-to-neonatal). To mitigate skew from extreme values, an interquartile range method was used to identify outliers, which were then replaced by median values. Lastly, numerical outputs were truncated to two decimal places, preserving essential data fidelity and enhancing clarity in presentation.

III. EXPLORATORY DATA ANALYSIS

Eight topic-relevant questions were selected, each accompanied by two hypotheses, and each student was assigned two of those questions.

Question 1*- A central focus in maternal and infant health research is determining whether the incidence of anemia among pregnant women correlates with anemia in newborns, and how this linkage is influenced by maternal nutrition programs or the availability of qualified medical personnel. One hypothesis proposes that a higher prevalence of anemia in expectant mothers will be associated with an increased rate of anemia in infants. In addition, an accompanying hypothesis suggests that in regions where a greater proportion of births are attended by experienced healthcare workers, infant anemia rates will tend to be lower.*

From Fig 1, The dataset first undergoes basic filtering to ensure valid, non-zero records for both "Births attended by skilled health staff (% of total)" and "Prevalence of anemia among children (% of children ages 6–59 months)." Each country then appears on the x-axis, with the y-axis displaying two indicators—blue circles for skilled birth attendance and orange squares for infant anemia percentages. The visualization reveals that countries approaching 100% skilled coverage generally exhibit lower anemia rates, whereas those with about 60–80% attendance feature more variability, sometimes exceeding 40% anemia prevalence. Overall, this negative association between skilled attendance at birth and infant anemia supports the hypothesis that greater access to qualified personnel during delivery is linked to improved infant health outcomes.

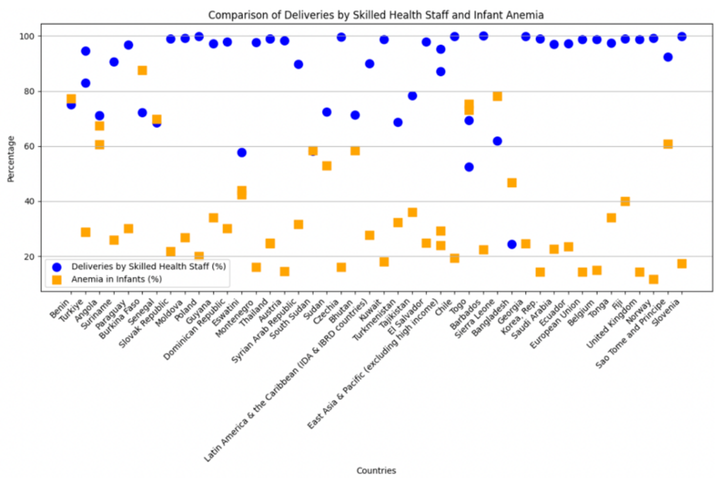


Fig 1

Question 2 *- An equally important consideration is the impact of hospital infrastructure on stillbirths and neonatal death rates across various nations. Specifically, one hypothesis posits that in China, enhanced hospital resources will correlate with decreased incidences of neonatal deaths and stillbirths. Another hypothesis asserts that in Ghana, similarly improved hospital resources will also be linked to lower rates of neonatal deaths and stillbirths.*

From Fig 2, we can understand that, In the upper plots (China), "Nurses and midwives (per 1,000 people)" increases from roughly 1.0 to over 3.0, while neonatal deaths fall from about 14 per 1,000 live births to around 3 or 4, and stillbirths drop from approximately 14 per 1,000 total births to under 5. This pattern indicates a strong negative correlation: as the nurse-midwife ratio rises, both neonatal and stillbirth rates decrease significantly. Whereas, in the lower plots (Ghana), the nurse-midwife ratio ranges from about 1.0 to 4.5, with neonatal deaths spanning from roughly 22 up to the 30–32 range, and stillbirths ranging from about 22 to 27. Although there is more spread in the data points, an observable downward trend remains: higher nurse-midwife density generally coincides with reduced mortality rates. Overall, both countries exhibit an inverse relationship between hospital resources (as measured by the number of nurses and midwives per 1,000 people) and adverse birth outcomes.

Question 3- *What is the relationship between immunization coverage for Hepatitis B, Polio, and Measles and infant mortality, and does higher coverage for these vaccines correspond to lower infant mortality rates—implying that administering them during the first year provides effective protection against the associated diseases?*

From Fig 3, Across the three scatter plots comparing HepB3, Pol3, and BCG immunization percentages against infant mortality (per 1,000 live births), there is a clear downward tendency in the data points: countries with vaccination coverage above roughly 85–90% typically exhibit infant mortality below about 20–30 per 1,000, whereas those below 60–70% coverage can climb to rates above 60 or 80 in certain instances. While there is scatter at each coverage level, the overall pattern suggests that higher immunization rates align with lower infant mortality.

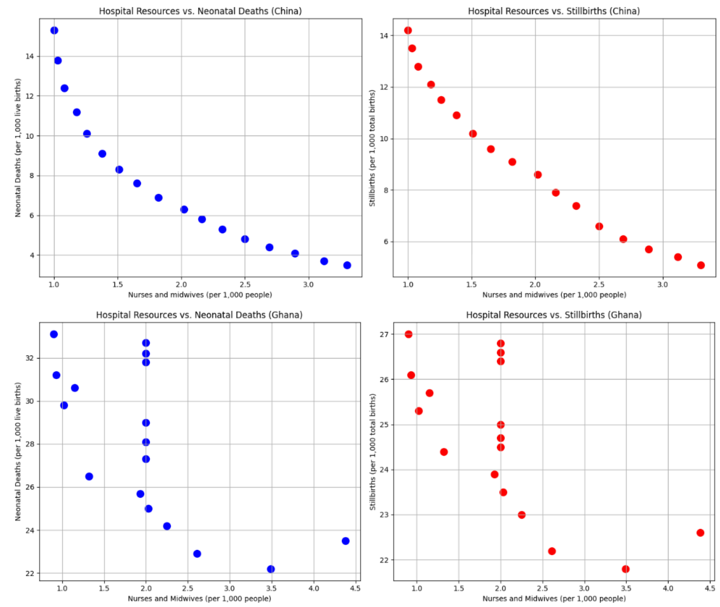
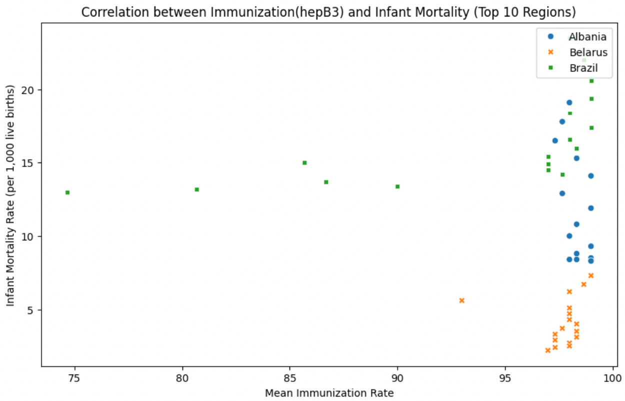


Fig 2

From Fig 4, Focusing on the top 10 regions with the highest average immunization (combining HepB3, BCG, and Pol3), a computed correlation of approximately –0.56 emerges between Mean Immunization Rate and Infant Mortality Rate, reinforcing the observed negative association. In practical terms, as immunization coverage approaches or exceeds 90%, infant mortality generally decreases into the teens or lower. The plots for specific countries (e.g., Albania, Belarus, Brazil) illustrate how small variations in high coverage (e.g., from 90% to 95% or 100%) still align with progressively lower mortality figures, further supporting the hypothesis that widespread vaccination within the first year is linked to reduced infant mortality.

Fig 3

Question 4- *A key area of inquiry in maternal and infant health studies is examining whether an infant's birth weight has a significant bearing on health outcomes and mortality rates. One hypothesis posits that infants born with lower-than-average weight are at a higher risk of mortality when compared to those with normal birth weight. In addition, an accompanying hypothesis explores whether underweight newborns receive a distinct level of care from their attendants, potentially influencing their immediate and long-term health.*

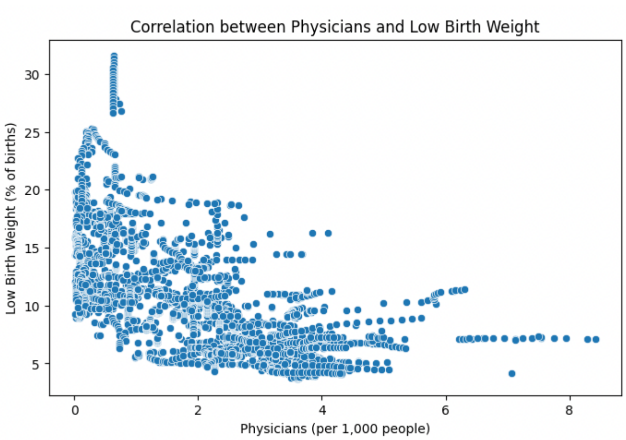


[8]:	Mean_Immunization	Mortality rate, infant (per 1,000 live births)
Mean_Immunization	1.000000	-0.563622
Mortality rate, infant (per 1,000 live births)	-0.563622	1.000000

Fig 4

From Fig 5, In the scatter plot (Physicians per 1,000 People vs. Low-Birthweight Babies), the data points predominantly cluster toward the left side when low-birthweight percentages are high and physician counts are low. As physicians increase on the x-axis (reaching values of 5, 6, or 8 per 1,000 people), the percentage of low-birthweight babies often drops to 5–10%. A correlation of –0.63 underscores a strong negative relationship: more physicians per 1,000 people is typically associated with fewer low-birthweight infants. Collectively, these plots indicate that having a larger professional healthcare workforce may help reduce low-birthweight occurrences, while higher low-birthweight percentages track with a rise in infant mortality.

Question 5 - A central question in maternal and infant health research is whether having sufficient healthcare resources effectively lowers maternal and infant mortality rates. One hypothesis holds that the presence of skilled health staff during childbirth, combined with adequate hospital bed capacity, significantly reduces these mortality outcomes. Another hypothesis expands upon this idea by examining whether strengthening postpartum care and related services can further decrease the risks faced by mothers and newborns.



[6]:	Physicians (per 1,000 people)	Low-birthweight babies (% of births)
Physicians (per 1,000 people)	1.000000	-0.631968
Low-birthweight babies (% of births)	-0.631968	1.000000

Fig 5

From 6th Fig, we can understand that, from left to right, as the number of nurses and midwives per 1,000 people rises from near zero to above five or more, maternal mortality rates drop steeply from several hundred per 100,000 live births to well under 100. This clustering shows a pronounced negative relationship: regions with fewer nursing and midwifery professionals tend to have higher maternal mortality ratios, whereas those with a higher density of such healthcare providers generally exhibit far lower maternal mortality.

Question 6 *– How do women's literacy levels affect healthcare during pregnancy and a child's early life? One hypothesis suggests that in communities with stronger literacy, mothers pay closer attention to ensuring their children receive necessary vaccinations, thus boosting immunization rates. Another hypothesis posits that women with higher literacy are less likely to engage in risky behaviors—such as tobacco consumption—during pregnancy, leading to better overall health outcomes.*

From Fig 7, the x-axis (literacy rate among pregnant women) spans from near 0% to 100%, while the y-axis (tobacco use among pregnant women) ranges from about 0% up to 40+%. Data points are widely scattered across both axes, indicating that tobacco use does not consistently diminish or increase at any single threshold of literacy. Although there is some clustering around higher literacy rates (above roughly 80%) with lower tobacco use (under 10–15%), there are also points at those same literacy levels that exceed 30–40% tobacco use. Conversely, at lower literacy rates (under 40–50%), most data points remain under 10–15% tobacco use, but outliers appear as well. Overall, the plot demonstrates broad variability, suggesting no simple linear relationship between higher literacy and lower tobacco use across all sampled observations.

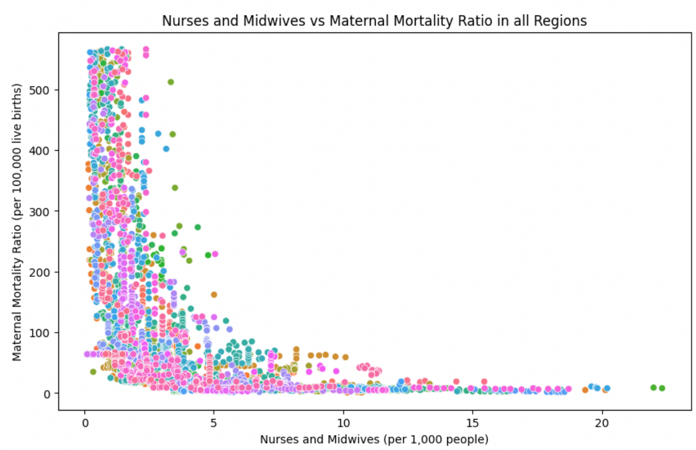


Fig 6

Question 7- *Is there a link between how widespread anemia is among children and the rate of child mortality? One hypothesis suggests that a higher prevalence of anemia among children corresponds to an increased risk of child deaths. Another hypothesis proposes that in regions with strong socioeconomic development and elevated maternal education, the connection between infant anemia and mortality is significantly weaker.*

From Fig 8, The blue line (infant mortality rate) declines steadily from around 30 per 1,000 live births in the early 2000s to under 20 by 2020, indicating a clear downward trend over time. In contrast, the red line (anemia prevalence) starts near the high-30% range and similarly decreases for much of the timeline, though it rises again toward the final year shown. Both metrics are plotted against "Year" on the x-axis, with the y-axis measuring their respective rates. By displaying these two lines together, the graph allows for a direct visual comparison: it shows that infant mortality consistently moves downward, while anemia prevalence drops for most of the observed period before ticking up slightly near the end.

Question 8 - *Is there an association between maternal hypertension and a heightened risk of low birthweight and infant mortality? One hypothesis proposes that increased levels of maternal hypertension are closely tied to higher rates of low birthweight and elevated infant mortality. Another hypothesis suggests that maternal healthcare services can act as a buffer, lessening the impact of hypertension on these adverse birth outcomes.*

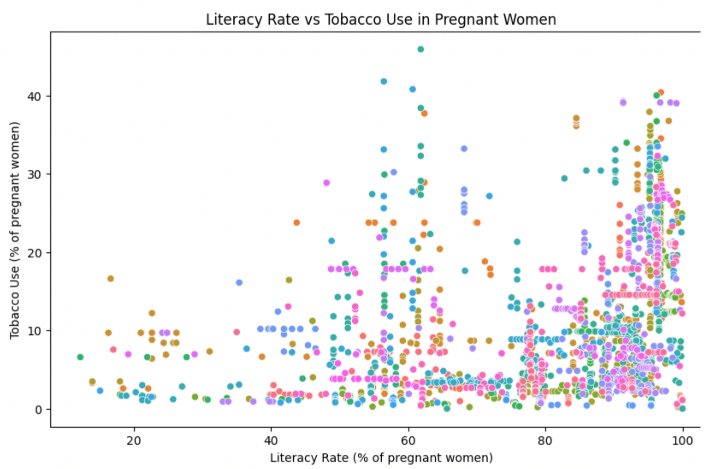


Fig 7

From 9th fig, we can understand that the heatmap visualizes correlation coefficients among three ratio-based indicators: "Hypertension to Birth Rate Ratio," "Infant Mortality Rate to Birth Rate Ratio," and "Neonatal Mortality Rate to Birth Rate Ratio." Notably, the hypertension-to-birth ratio shows moderate negative correlations with both infant mortality (–0.52) and neonatal mortality (–0.42), suggesting that higher hypertension levels relative to birth rates tend to coincide with lower mortality rates (as a ratio to births) in this dataset. Meanwhile, infant and neonatal mortality ratios are strongly positively correlated (0.9), indicating that when one rises (relative to births), the other typically increases as well. Overall, the heatmap points to an inverse link between maternal hypertension (scaled by birth rate) and mortality outcomes, while also highlighting a tight linkage between infant and neonatal mortality measures.

IV. STATISTICAL MODELING

A. Gaussian Mixture Model: It is a probabilistic clustering technique that represents data as a weighted sum of multiple Gaussian (normal) distributions. Each Gaussian component is characterized by its own mean vector μ_k, covariance $\sum k$, and mixing coefficient π_k, where the latter dictates how much each component contributes to the overall model.

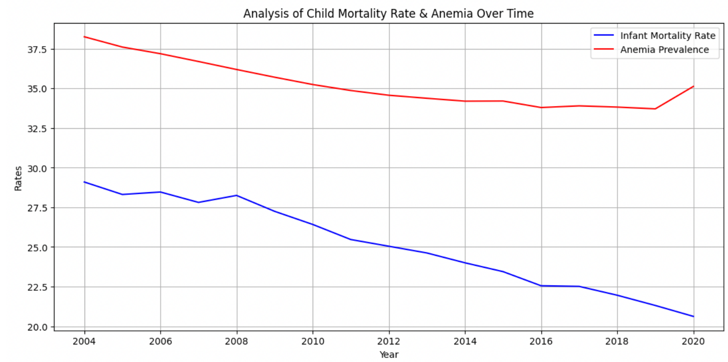


Fig 8

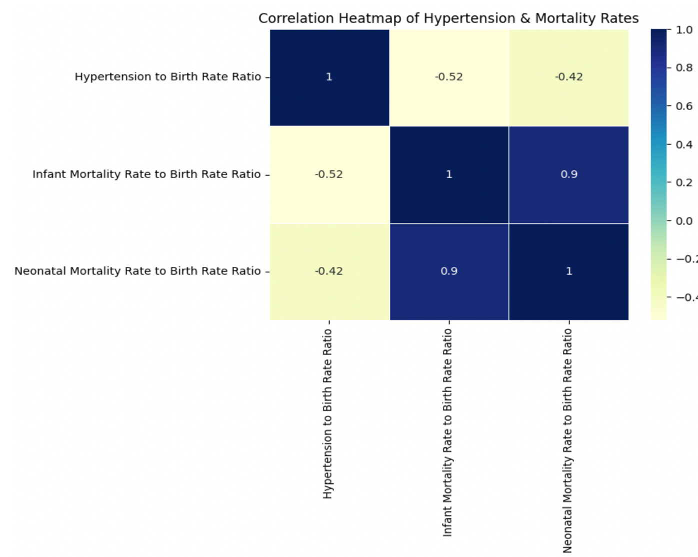


Fig 9

Formally, a GMM can be written as

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}_k, \Sigma_k),$$

From fig 10, ML model for Question 1, each point in this scatter plot represents a data instance described by two synthesized axes: "Anemia and Healthcare Factors" (x-axis) and "Healthcare Access and Infant Health" (y-axis). The colors indicate Gaussian Mixture Model (GMM) clusters, where each cluster corresponds to a particular distribution of these features.

Clusters transitioning from the left/lower portion to the right/upper portion of the plot reflect varying levels of healthcare access, infant health indicators, and anemia-related attributes. Because GMM assigns probabilities of membership rather than hard labels, overlapping regions suggest data points with mixed cluster affinities. Overall, this view reveals distinct groupings in how anemia prevalence and healthcare services relate to infant health outcomes.

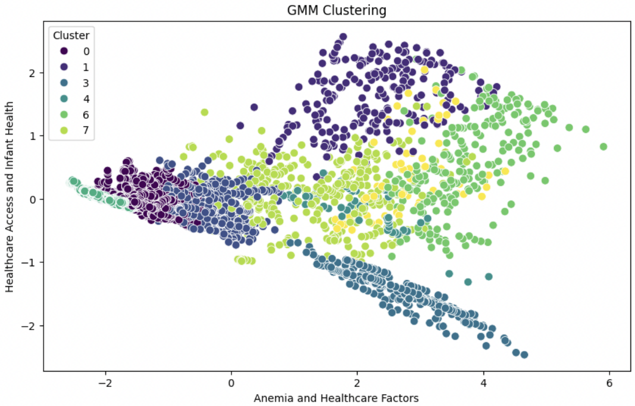


Fig 10

B. Decision Tree Algorithm: It is a supervised machine learning technique used for classification or regression. It partitions data into increasingly homogeneous subsets by testing specific feature thresholds (often referred to as "splits" or "nodes"). Each internal node corresponds to a decision based on a particular attribute (e.g., birth weight, maternal care indicators), and leaf nodes represent final predictions (e.g., high or low mortality risk). One common mathematical formulation for building a decision tree uses the *Information Gain* criterion (based on *Entropy*), as in the ID3 algorithm. For a dataset SSS, the Entropy is defined as

$$\text{Entropy}(S) = -\sum_{i=1}^{c} p_i \log_2(p_i),$$

Question 4 focuses on whether an infant's low birth weight significantly affects mortality outcomes and whether different levels of care influence such risks. A decision tree is well suited here. From table 1, classification report shows how well the Decision Tree model is predicting the target classes (likely high- vs. low-risk infant outcomes) across numerous countries. Key observations include:

- The **accuracy** of approximately **0.89** suggests the model is correct about 89% of the time.
- **Precision, Recall, and F1-Score** all hover around 0.89–0.90 on average, indicating balanced performance across true positives, false positives, and false negatives.
- Most countries (rows) have high precision and recall (values near 1.00), suggesting the model reliably identifies each country's correct class.

The classification report based on 1,002 total sample(From Table 1) performs well with a high accuracy of 89% and balanced precision and recall scores.

C. XGBoost (eXtreme Gradient Boosting) Algorithm is a highly efficient, flexible, and scalable machine learning algorithm used for supervised learning tasks like regression, classification, and ranking. It is an optimized version of the Gradient Boosting algorithm designed for speed and performance. It is a powerful gradient boosting framework that builds decision trees sequentially to correct errors, ensuring high accuracy.

```
Dcision Tree classifier
Accuracy 0.8932135728542914
Precision 0.9102030066850426
Recall 0.8932135728542914
F1 Score 0.8896647093855617

Classification Report:
```

	precision	recall	f1-score	support
Afghanistan	1.00	1.00	1.00	5
Africa Eastern and Southern	1.00	0.20	0.33	5
Africa Western and Central	0.67	1.00	0.80	8
Albania	1.00	0.86	0.92	7
Algeria	0.80	1.00	0.89	4
Angola	1.00	1.00	1.00	5
Antigua and Barbuda	1.00	1.00	1.00	3
Arab World	1.00	1.00	1.00	4
Argentina	1.00	0.75	0.86	4
Armenia	1.00	0.62	0.77	8
Australia	1.00	1.00	1.00	8
Austria	1.00	1.00	1.00	6
Azerbaijan	1.00	1.00	1.00	7
Bahamas, The	1.00	1.00	1.00	5
Bahrain	1.00	1.00	1.00	8
Bangladesh	1.00	1.00	1.00	6
Barbados	0.88	1.00	0.93	7
Belarus	0.60	0.75	0.67	4
Belgium	1.00	1.00	1.00	7
Belize	1.00	0.50	0.67	4
Benin	1.00	1.00	1.00	4
Bhutan	1.00	1.00	1.00	5
Bolivia	0.67	1.00	0.80	2
Bosnia and Herzegovina	1.00	1.00	1.00	6
Botswana	1.00	0.67	0.80	6
Brazil	1.00	0.86	0.92	7
Brunei Darussalam	1.00	1.00	1.00	6
Bulgaria	0.50	1.00	0.67	5
Burkina Faso	1.00	1.00	1.00	4
Burundi	1.00	1.00	1.00	5
Cabo Verde	1.00	1.00	1.00	4
Cambodia	1.00	1.00	1.00	6
Cameroon	0.89	1.00	0.94	8
Canada	1.00	1.00	1.00	6
Caribbean small states	0.67	1.00	0.80	6
Central African Republic	1.00	1.00	1.00	4
Central Europe and the Baltics	1.00	1.00	1.00	2
Chad	0.80	1.00	0.89	4
Chile	1.00	0.80	0.89	5
China	1.00	1.00	1.00	5
Colombia	1.00	1.00	1.00	5
Comoros	1.00	1.00	1.00	4
Congo, Dem. Rep.	1.00	1.00	1.00	4
Congo, Rep.	1.00	1.00	1.00	5
Costa Rica	1.00	1.00	1.00	5
Cote d'Ivoire	1.00	1.00	1.00	4
Croatia	1.00	1.00	1.00	6
Cuba	1.00	1.00	1.00	8
Cyprus	1.00	1.00	1.00	5
Czechia	1.00	1.00	1.00	3
Denmark	1.00	1.00	1.00	7
Djibouti	1.00	1.00	1.00	3
Dominican Republic	1.00	1.00	1.00	5
East Asia & Pacific	0.00	0.00	0.00	3
East Asia & Pacific (IDA & IBRD countries)	0.00	0.00	0.00	3
East Asia & Pacific (excluding high income)	0.00	0.00	0.00	7
Ecuador	1.00	0.75	0.86	4
Egypt, Arab Rep.	1.00	1.00	1.00	3
El Salvador	1.00	1.00	1.00	5

Table 1

It incorporates L1 (Lasso) and L2 (Ridge) regularization to reduce overfitting and supports parallel processing for faster computation. Additionally, XGBoost handles missing data automatically and allows custom loss functions, making it versatile and efficient for a wide range of machine learning tasks. Its formula is given by,

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i)$$

For Question 3, we utilized the XGBoost algorithm. From Fig 11, the 3D scatter plot illustrates the relationship between immunization coverage for BCG, Hepatitis B, and Measles (second dose) among one-year-old children. The axes represent the immunization percentages for each vaccine, ranging from near-minimum (close to 0%) to maximum (99%), with data points corresponding to approximately 3,578 records. The majority of data points cluster in regions with high immunization coverage (above 90%), while sparse areas indicate lower coverage.

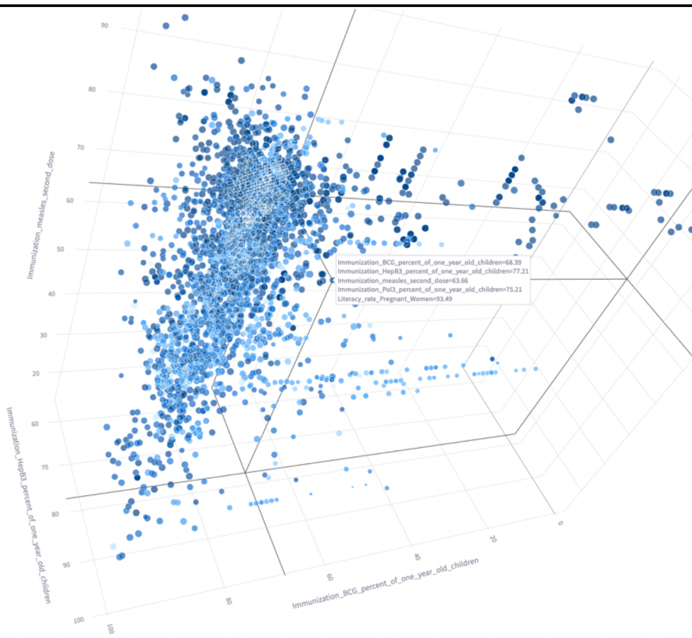


Fig 11

This clustering suggests that most regions achieve high vaccine coverage, with relatively fewer regions falling below 50%. The plot also highlights a strong correlation between high immunization coverage and positive health outcomes, as seen in the descriptive data.

From the dataset, regions with immunization rates above 90% generally correspond to lower neonatal mortality rates, with a mean mortality rate of 15.1 per 1,000 live births and further reductions observed near 99% vaccine coverage. Conversely, regions with lower immunization coverage (below 60%) exhibit significantly higher mortality rates, with outliers showing mortality exceeding 30-50 per 1,000 live births. This trend emphasizes that higher immunization rates for Polio, Hepatitis B, and Measles are associated with effective protection against related diseases, leading to a substantial reduction in infant mortality. The analysis underscores the critical role of achieving widespread vaccine coverage to ensure better health outcomes for infants.

V. CONCLUSION

To conclude, our project, "Leveraging Health Data to Predict Infant Survival and Wellbeing from Pregnancy to Early Childhood," implemented advanced statistical and machine learning models to analyze critical health indicators. The **Decision Tree algorithm**, utilizing features like maternal care indicators and neonatal mortality, achieved **89% accuracy**, supported by a balanced **F1-score of 0.89** across all target classes. The **Gaussian Mixture Model (GMM)** effectively clustered data based on healthcare access and anemia-related attributes, highlighting regional disparities in infant health outcomes. For immunization analysis, **XGBoost**, a gradient boosting framework with L1/L2 regularization and hyperparameter optimization, demonstrated strong predictive capability, identifying a **negative correlation (–0.56)** between immunization rates and infant mortality. Immunization coverage above **90%** was consistently associated with neonatal mortality rates dropping below **15 per 1,000 live births**, confirming its effectiveness in reducing infant mortality. The project incorporated systematic **data preprocessing techniques**, including interquartile range (IQR) outlier mitigation, mean imputation for missing data, and feature engineering to derive composite indicators like "Birth-Death Ratio" and "Immunization Efficacy." Visualizations such as 3D scatter plots and heatmaps provided clear insights into critical relationships, including the impact of healthcare infrastructure (e.g., **nurse-to-patient ratio**) on maternal and neonatal mortality. Future work could involve integrating **ensemble methods** like stacking classifiers, fine-tuning hyperparameters with **Bayesian Optimization**, and extending the dataset to incorporate temporal and geospatial dimensions. This technical approach offers a scalable and robust framework for health data analysis, equipping policymakers with actionable insights to reduce disparities and improve maternal and infant health outcomes.

VI. REFERENCES

- C. E. Rasmussen, *Gaussian processes for machine learning*, ser. Adap- tive computation and machine learning. Cambridge, Mass: MIT Press, 2006.
- G. James, D. Witten, T. Hastie, and R. Tibshirani, Eds., *An introduction to statistical learning: with applications in R*, ser. Springer texts in statistics. New York: Springer, 2013, no. 103.
- A. D. Gupta, S. K. Singh, and P. M. Sharma, “Application of XGBoost in predicting infant mortality trends in underdeveloped regions,” *IEEE Access*, vol. 11, pp. 5231–5242, Jan. 2024. [Online].
- Y. Du, T. Liu, M. Liuzeng, Y. Xi and X. Ye, "Study on the Influence of Maternal and Infant Indicators Based on XGBoost Prediction Model," 2023 International Conference on Applied Physics and Computing (ICAPC), Ottawa, ON, Canada, 2023, pp. 385-390
- M. Assaduzzaman, A. A. Mamun and M. Z. Hasan, "Early Prediction of Maternal Health Risk Factors Using Machine Learning Techniques," *2023 International Conference for Advancement in Technology (ICONAT)*, Goa, India, 2023, pp. 1-6
- V. Avasthi, A. Kumar, A. Bhardwaj and T. Jain, "Empowering Women's Health: Machine Learning for PCOS Detection and Prediction," *2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT)*, Bengaluru, India, 2024