PREDICTIVE MODELING OF CHILD MORTALITY IN MIGORI AND NYAMIRA COUNTIES USING INDIRECT METHODS

BRIAN OMARE

A THESIS SUBMITTED TO THE SCHOOL OF SCIENCE IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF DEGREE OF MASTER OF SCIENCE IN BIOSTATISTICS IN THE DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, UNIVERSITY OF ELDORET, KENYA

2025

DECLARATION

Declaration by the Candidate

This thesis is my original work and has not been presented for a degree in any other University. No part of this thesis may be reproduced without the prior written permission of the author and/or University of Eldoret.

________________________________ DATE ________________________

Brian Omare

SSCI/MAT/M/001/17

Approval by the Supervisors

This thesis has been submitted with our approval as University Supervisors.

________________________________ DATE ________________________

Prof. Argwings Otieno

University of Eldoret, Kenya

________________________________ DATE ________________________

Dr. Julius Koech

University of Eldoret, Kenya

DEDICATION

This thesis is dedicated to Almighty God and my wonderful family. There is no doubt in my mind that without all your support; materially, emotionally and spiritually I would not have managed to complete my masters. My father Yabesh Omare, mother Calystah Omare and siblings for all encouragements, tough love that kept me focused to this end.

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Prof. Argwings Otieno and Dr. Julius Koech, for the knowledge and insight they guidance they accorded me, and for their support during this phase of my thesis development. They have shared comparably profound knowledge and great ideas with me, and deepened my understanding of statistics and science. Throughout the process of writing this thesis, they encouraged me to develop independent thinking and research skills, and greatly assisted me with scientific writing. I could never have reached the heights or explored the depths without their help and guidance.

ABSTRACT

Child mortality remains a critical public health challenge, particularly in developing countries like Kenya, where disparities in healthcare are stark across different regions. In counties such as Nyamira and Migori, persistent high rates of under-five child mortality demonstrate the need for more precise statistical predictions for and targeted interventions. Traditional methods for estimating child mortality, such as those derived from household surveys, are often hampered by issues like missing data and survivor bias, leading to inaccurate mortality estimates. This study sought to develop a comprehensive predictive model for under-five child mortality in Migori and Nyamira counties, Kenya, by incorporating temporal patterns and social determinants of health. Utilizing a retrospective cohort design, the study analyzed historical data from health records, census reports, and household surveys spanning 34 years (1989-2022). The analysis incorporated indirect estimation techniques to address data gaps and employed multiple linear regression, gradient boosting regressor, and spatio-temporal modeling to capture temporal and seasonal trends in child mortality. The multiple linear regression model was significant, explaining 89.9% of the change in neonatal mortality in Migori County and 80.6% of the variation in Nyamira County. Gradient boosting regressor performed optimally, accounting for 80.9% of the change in child mortality, indicating good predictive capability and suggesting that the chosen independent variables effectively capture the complexity of the response variable. Spatio-temporal modeling log-likelihood value of -111.87 indicated a relatively good fit, capturing the observed data well (pseudo-R-squared = 0.9415). Results indicated that infant mortality rates in both counties have fluctuated historically, with distinct seasonal trends influenced by factors such as disease prevalence and access to healthcare services. The temporal and seasonal analysis revealed that periods of increased respiratory complications and malaria prevalence corresponded with higher mortality rates. The study provides a methodological framework that can be adapted to other regions with comparable challenges. By addressing the limitations of traditional mortality estimation methods and leveraging advanced predictive modeling techniques, the study contributes to the ongoing efforts to improve child health outcomes in Kenya and beyond.

Keywords: Child Mortality, Predictive Modeling, Indirect Estimation, Time Series Analysis, Machine Learning, Social Determinants of Health.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SDGs: Sustainable Development Goals

LSTM: Long Short-Term Memory

NMR: Neonatal Mortality Rate

IMR: Infant Mortality Rate

CMR: Child Mortality Rate

WHO: World Health Organization

KEMRI: Kenya Medical Research Institute

FBH: Full Birth History

ML: Machine Learning

VR: Vital Registration

RNN: Recurrent Neural Networks

NICU: Neonatal Intensive Care Unit

DHS: Demographic and Health Survey

U5: Under Five

U5CM: Under-five Child Mortality

STAR: Spatio-Temporal Auto-Regressive

VR: Vital Registration

CHAPTER ONE

INTRODUCTION

This section provides the background for the study, including a brief history of the problem of child mortality and the need to predict these cases for appropriate intervention. It outlines the study's objectives and justifies the topic, detailing the purpose and significance of the study.

1.1 Background to the study

The 2030 Agenda for Sustainable Development, issued by United Nations, provides a shared blueprint for peace and prosperity for humans on the planet, now and into the future. The Sustainable Development Goals (SDGs) are a global call to action to end poverty and protect the earth’s environment and human health (Wang et al., 2023). Specifically, SDG 3 states that countries should aim to reduce neonatal mortality to at least 12 per 1,000 live births and under-five mortality to at least 25 per 1,000 live births by 2030. However, despite that progress, 5.2 million children died before reaching their fifth birthday in 2019, with almost half of those deaths, 2.4 million, occurring in the first month of life (Wang et al., 2023). Further, the mortality of newborns, infants (in the first year of life), and children (under 5 years old) shows a huge difference across countries. For example, in North America, the average neonatal mortality rate (NMR), infant mortality rate (IMR), and child mortality rate (CMR) in 2019 were relatively low, which were about 3.7, 5.4, and 6.3 deaths per 1,000 live births, respectively. In contrast, these indicators in developing countries in Sub-Saharan Africa present substantially higher levels at 27.5, 51.7, and 75.8 deaths per 1,000 live births in 2019, respectively, based on the United Nations Children’s Fund data.

According to a new Vital Statistics Rapid Release report from the National Centre for Health Statistics, the provisional newborn mortality rate in the United States in 2022 increased by 3% from 2021, marking the first year-over-year increase since 2001-2002 (UNICEF, 2023). From 2002 to 2021 the infant mortality rate declined 22%. Neonatal, infant, and U5 deaths arise due to a myriad of challenges, sometimes medical, congenital, or otherwise ranging from birth defects, infections, or accidents. The main causes of infant mortality in the United States are different than the main causes of infant death around the world (WHO, 2023).

In Kenya, according to (WO), 2023) and (KEMRI, 2021), neonatal deaths remain unacceptably high, contributing to 40% of under-five mortality rates (U5MR) making it an important health priority (UNICEF, 2023). According to a recent study conducted by the Kenya Medical Research Institute (KEMRI), the number of children dying under the age of five remains high in some regions of the country. While attempts are made to reduce the number of children who die before the age of five, it appears that conditions have yet to improve (Paul, 2023). Some of the leading causes of infant mortality in Kenya include: first, *respiratory tract infections* - Children who contract respiratory tract infections particularly pneumonia often succumb to it. A study found that parents failed to take their children to hospital when they fell ill. This caused the conditions worsen and consequently, led to the child's death (WHO, 2023). Second, *diarrhea* - which often leads to dehydration if left untreated. This can lead to death especially in children who don't get medical attention after prolonged diarrhea. Importantly, there's *malnutrition* - insufficient or lack of breastfeeding as well as insufficient nutrients leads to stunted growth and in extreme cases can lead to the child's death.

1.2 Statement of the Problem

Regardless of considerable improvements, advancements, and developments in maternal and general healthcare, the rates of child mortality in Kenya exceed the global average, especially in less developed counties such as Migori and Nyamira in the larger western region. A review of past retrospective cohort studies on under-five child mortality in Kenya reveal considerable historic decline in mortality. According to (Watulo & Wanjoya, 2021), a significant factor (improved healthcare, nutrition, and hygiene) attributed to the reduction is an improvement in health systems over time (Anthopolos et al., 2017; Starnes et al., 2018). The came up with findings detailing factors such as low maternal education, rural residences, short birth spacing, and multiple-gestation pregnancies as contributing factors to under-five mortality (Starnes et al., 2018; Watulo & Wanjoya, 2021).

In Migori, Nyamira, and counties sharing similar demographics and poor development, especially, geographical similarities emerged as key contributory factors to child mortality, with risk being twice the mortality of Central Kenya and other urban/developed more regions (Watulo & Wanjoya, 2021). Other concerns include post-natal care with hospitalized children up to a year later after birth recording higher mortality rates compared to the rest of the under-fives. These studies highlight critical findings but fall short of environmental and socio-economic factors that the present study seeks to address in predicting child mortality in Migori and Nyamira counties.

Different other studies have modelled U5CM in Kenya, employing machine learning techniques. The widely used model was STAR, revealing uneven decline in mortality rates in Kenya between 1965 to 2014 due to infrastructural disparities (Macharia et al., 2019). Other studies have invoked Bayesian models to study U%CM rates and associated factors in Kenya, focusing on convoluted survey designs and data limitations (Wakefield et al., 2017; Macharia et al., 2021). With counties in the north mostly disadvantaged, Macharia et al. (2021) identified geographic location and demographic characteristics as factors influencing U5CM. The findings have a disparity with Spatial Cox Proportional Hazard analyses that ranked western counties with the lowest hazard of child death with central counties emerging top in Kenya (Kilemi Daniel et al., 2021). The present study seeks to inform targeted interventions by modelling three categories of U5 child mortality in Migori and Nyamira counties, producing models that can be replicated to predict mortality in similar geographical locations in Kenya.

1.3 Objectives

This section develops the general objective that the study satisfies, as well as the specific objectives that were met in order to achieve the purpose of the research.

1.3.1 General objective

To develop and evaluate a comprehensive approach for predicting child mortality considering both temporal patterns and social determinants of health in Nyamira and Migori counties, Kenya.

1.3.2 Specific objectives

1. To analyse survey and demographic data to determine social determinants of child mortality in Nyamira and Migori counties, Kenya.
2. To develop a context-specific predictive model for child mortality using indirect estimation techniques in Nyamira and Migori counties, Kenya.

3. To examine temporal patterns and seasonal trends in child mortality rates using machine learning approaches in Nyamira and Migori counties, Kenya.
4. To evaluate the performance of the predictive modelling for different child risk groups in Nyamira and Migori counties, Kenya in comparison with traditional methods.

1.4 Research Questions

1. What are the social determinants of child mortality in Nyamira and Migori counties, Kenya, by analysis of survey and demographic data?
2. What is the context-specific predictive model for child mortality using indirect estimation techniques in Migori and Nyamira counties?
3. What are the temporal patterns and seasonal trends in child mortality rates in Migori and Nyamira counties using machine learning approaches?
4. What is the performance of predictive modelling for different child risk groups as opposed to traditional methods?

1.5 Justification and significance

Just like many developing countries, especially sub-Saharan Africa, Kenya faces high rates of child mortality. Furthermore, Kenya has not realized one of the Millenium Development Goals (MDGs) of reducing infant mortality rates by two thirds as of 2015. Designing targeted interventions to improve healthcare outcomes, including those of reducing infant mortality will partly be achieved by understanding and predicting mortality in children aged below the age of five years.

Additionally, less developed countries such as Kenya have conducted marginal child mortality predictive studies, especially in low-resourced counties. The majority of modelling and forecasting has been performed in industrialized countries, which disadvantages developing countries in terms of interventions owing to a lack of data. Kenya's healthcare infrastructure differs from the predictive models built for affluent economies, leaving a large gap, demanding research that focuses on the unique issues facing poor economies.

1.6 Research Rationale and Purpose

The current study was informed by a paucity of contextualized models in Kenya's disadvantaged counties and rural areas. There is a void in the research for models customized particularly to the Kenyan setting, as existing predictive models for child mortality fail to account for the distinct environmental, cultural, and socioeconomic elements that pose mortality risks. Second, child mortality data in Kenya, especially county hospitals lacks sufficient integration while data from private, mission, religious and other non-government healthcare facilities becomes hard to access. The readily available data on the three U5 categories (neonates, infants, and under five) is single source focusing on few variables that may inform bias or inconclusive findings. As such, addressing these shortcomings on accessibility and availability of data presented by NMR and FBH requires more thorough methods, which the present study proposes in terms of indirect estimation to amalgamate different data sources, including but not limited to health records, demographic information, environmental factors, and socio-economic indicators, to provide a comprehensive understanding of mortality risk. It is also important to recall the lack of a context-specific model, ML or otherwise trained to portray the U5 mortality situation in Migori and Nyanza counties, a circumstance that the study cures by developing a context specific child mortality forecasting model with key focus on covering temporal and seasonal trends in rural areas while categorizing victims into various risk groups.

1.7 Use of Machine Learning in Predictive Modelling

The present study elected machine learning because of its proven success in predicting high-prevalence events across diverse datasets and large-scale data. In indirect mortality estimation, machine learning algorithms are crucial as they leverage trends and correlations within extensive datasets. For example, these algorithms excel at identifying complex patterns in data and recognizing convoluted relationships between factors such as health indicators, environmental variables, demographics, and mortality trends (Berhanu et al., 2021).

In the current study, historical datasets from the Ministry of Health, WHO databases, and electronic health records were used to train models to predict child death rates. Furthermore, ML models can combine several aspects, including as social, environmental, and health, to assess the risk of mortality for individuals and populations, allowing for more focused interventions (Daniel et al., 2021). Most importantly, ML algorithms have the ability to combine and evaluate data from a variety of sources, including big data, making them suited for estimating indirect mortality through real-time trend monitoring and model development.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Infant mortality remains a significant public health concern, particularly in poor counties in Kenya where access to healthcare resources may be limited. This literature review aims to identify existing research gaps and proposes a comprehensive research agenda cantered around three main objectives: developing a context-specific predictive model, analysing temporal patterns and seasonal trends, and evaluating social determinants of infant mortality.

2.2 Predictive Modelling of Child Mortality

Mortality during childhood has garnered significant attention, particularly in less developed countries where a considerable number of deaths occur in early life. Saving children under the age of 5, especially those under 1 year old, is a major public health concern, necessitating accurate statistics and rigorous measures to identify the extent of the problem and monitor progress (Neupert et al., 2019). Of particular importance is the estimation of infant and child mortality in local areas, defined as small administrative regions like districts, municipalities, or counties. The demand for small area mortality statistics has surged in recent decades, driven by decentralized approaches to health planning, service provision, and resource allocation.

However, calculating early-age mortality in small areas poses substantial challenges, particularly in less developed countries where data sources are often limited to information from population censuses. The estimation methodology typically relies on William Brass' indirect methods. Although vital registration systems, particularly live births and deaths classified by age, are obvious data sources for calculating infant and child mortality, they are often incomplete or non-existent in less developed countries (Neupert et al., 2019). In such constrained data environments, childhood mortality is typically estimated using data from population and housing censuses through indirect methods.

Stoner et al. (2019) presents a framework for correcting underreporting of TB in Brazil, which could be adapted for infant and child mortality. Mercer et al. (2015) demonstrated the adaptation of estimation methodologies applied to vital statistics for Tanzanian data from Demographic and Health Surveys and surveillance systems. Swanson et al. (2019) conducted estimations of infant mortality for small areas through statistical modelling approaches. The increasing demand for high-quality subnational estimates of under-5 mortality, particularly in low- and middle-income countries, emphasizes the reliance on household surveys for reliable data, even though these are spatially sparse.

Predictive modelling of infant mortality using indirect estimation methods involves applying statistical models to estimate full life tables or mortality rates based on limited data, often focusing on child mortality alone or both child and adult mortality (Masquelier et al., 2024). Machine learning techniques are increasingly utilized to analyse the quality of mortality models and differentiate causes of death. A flexible two-dimensional mortality model has been proposed, outperforming traditional methods and better meeting practical needs.

2.3 Advances in Predictive Modelling and Indirect Estimation Methods

Predictive modelling of child mortality, employing indirect estimation methods, entails the application of statistical models to estimate full life tables or mortality rates based on limited data (Masquelier et al., 2024). This often includes data on child mortality alone or both child and adult mortality. The integration of machine learning techniques enhances the analysis of mortality models, enabling the differentiation of causes of death. A noteworthy advancement was the proposal of a flexible two-dimensional mortality model, requiring two input parameters to generate a comprehensive set of age-specific mortality rates, with the second parameter being optional. This model has demonstrated superior performance compared to traditional methods like the Coale-Demeny and UN model life tables, making it better suited to practical needs in mortality estimation.

Moreover, indirect estimation methods extend to cause-of-death mortality, utilizing life expectancy at birth to reconstruct cause-specific mortality over time from a single value of life expectancy (Jaskari et al., 2020). In the context of predicting infant mortality in Rwanda, machine learning methods, particularly the random forest, have proven to be the most effective classifier compared to other models used in the study.

Coherent mortality forecasting, incorporating compositional data analysis, has shown promise in yielding less biased forecasts than traditional models based on the extrapolation of death rates. Examining mortality dynamics in rural Veneto, a study has focused on mortality selection in the first three months of life and survival in the subsequent thirty-three months, utilizing semi-parametric Cox regression (Mazzucato et al., 2023). For populations lacking reliable data, model life tables capitalize on the strong positive correlation between mortality levels at different ages to predict mortality levels for all ages using the limited information available.

The performance of indirect methods, particularly those based on summary birth histories, has been rigorously assessed using microsimulation. Results consistently demonstrate that newer methods outperform the classical Brass method. In summary, recent advancements in predictive modelling and indirect estimation methods are significantly enhancing the ability to accurately assess and predict infant mortality, especially in populations with limited or incomplete data.

2.4 Addressing Complexities in Infant Mortality: Insights from Studies in Kenya

The issue of infant mortality in Kenya, specifically the indirect estimation of such mortality, represents a complex and multifaceted challenge that has garnered attention in various studies. The decline in infant mortality in Kenya is attributed to a combination of factors, including improved healthcare access, economic growth, and policy interventions (Bendavid & Bhattacharya, 2014). Despite progress, a significant research gap persists in indirect estimation methods, essential for comprehending the true scale of the issue, particularly in regions where data collection is challenging.

Analysing into the complexities of infant mortality estimation, Lu, et al. (2020) highlighted the importance of addressing cross-sectional dependence in estimation and provided long-run parameters, offering insights into the impact of selected variables on the infant mortality rate. Meanwhile, another study by Wairoto et al. (2020) focused on spatial differentials in childhood mortality, emphasizing the need for spatial analysis to understand subnational disparities and inform targeted interventions.

The importance of inpatient neonatal services was assessed by Murphy et al. (2017), with a focus on evidence and expert consensus to guide local policy in Kenya. Additionally, Wakibi and Ngure (2021) conducted an assessment of knowledge, attitude, and practices of birth and death registration in Kilifi County, revealing a gap between awareness and coverage, highlighting the need for improved implementation of registration practices.

Spatial survival models were employed to evaluate the effect of spatial variation on under-five mortality in Kenya, revealing disparities between counties and stressing the significance of county-based healthcare planning and intervention (Daniel et al., 2021). Comparisons of age patterns of under-5 mortality in sub-Saharan Africa offered valuable insights into mortality rate estimation using different methods.

Eilerts, et al. (2021) explored socioeconomic inequalities in skilled birth attendance and measles immunization, emphasizing their impact on infant mortality. Addressing these inequalities emerged as a potential avenue for reducing infant mortality. Furthermore, a retrospective analysis of maternal and neonatal mortality at a teaching and referral hospital in Kenya provided a detailed examination of circumstances associated with maternal and neonatal deaths, offering a basis for improving healthcare services.

Infant mortality in Kenya persists as a formidable challenge, driven by factors such as insufficient access to healthcare services, socioeconomic disparities, and entrenched cultural practices. Variations in child mortality stem from a myriad of factors, including the wealth index, education level, nutrition, breastfeeding duration, antenatal care utilization, maternal healthcare services, birth intervals, and the age of the mother (Langat et al., 2024). Notably, maternal education, as a socioeconomic determinant, plays a significant role in child survival rates. Community-level characteristics, encompassing income, residence, infrastructure, and region, further contribute to health outcomes and elucidate child mortality patterns.

Urban and rural disparities in infant mortality are evident, with certain urban areas like Nairobi exhibiting higher odds of infant death. Moreover, the persistently high maternal mortality ratio in Kenya underscores the need for county-specific healthcare planning and intervention (Macharia et al., 2021). Spatial survival models have been instrumental in elucidating the impact of spatial variation on under-five mortality, emphasizing the geographical factors influencing health outcomes.

Birth asphyxia emerges as a leading cause of neonatal mortality in Kenya, necessitating proficient neonatal resuscitation skills among healthcare professionals. Despite training and guidelines, reported poor practices in neonatal resuscitation contribute to a high incidence of neonatal deaths from birth asphyxia (Shikuku et al., 2018). Barriers to the utilization of maternal services and continuity of care are multifaceted, encompassing inadequately equipped health facilities, low-quality care, and traditional, religious, or cultural practices. The prohibitive factors of distance, travel costs, poor roads, and the unavailability of ambulances create significant challenges, particularly for women facing financial constraints unless experiencing serious complications.

In summary, child mortality in Kenya results from a complex network of individual characteristics, community-level variables, healthcare access, quality, and cultural practices, along with socioeconomic factors. A comprehensive approach to address these issues involves improving healthcare infrastructure, enhancing the quality of care, and tackling socioeconomic disparities and cultural barriers to healthcare utilization (Reisman et al., 2016). Addressing the research gap in the indirect estimation of infant mortality in Kenya requires improving data collection methods, implementing targeted healthcare interventions informed by spatial analysis, enhancing birth and death registration practices, and tackling socioeconomic inequalities. A collaborative approach involving policymakers, healthcare providers, and researchers is crucial to effectively reduce infant mortality rates in Kenya.

2.5 Monitoring Child Mortality Trends

Monitoring child mortality levels and trends is crucial for understanding progress in child survival, directing policy, and allocating financial assistance. However, in countries with high child mortality rates, poor or non-existent vital registration (VR) systems often lead to unrecorded deaths (Tin Oung et al., 2017). In the absence of quality VR data, under-5 mortality trends are typically estimated using retrospective household sample surveys and census data.

2.6 Spatial-Temporal Perspective

Various studies demonstrate the necessity for comprehensive explorations that comprehensively analyse the evolution of neonatal, infant, and child mortality indicators across countries, emphasizing a spatiotemporal perspective. These papers also highlight the existing research gap in health inequality within newborns, infants, and children on a cross-country scale. Emphasis is placed on the widening disparities in mortality rates across nations, calling for effective policy interventions to mitigate global health inequality.

The estimation of mortality in small areas has been a focal point in demographic research, with Neupert, et al. (2019) stressing the adaptation of methodologies for analysing vital statistics derived from sample surveys. They emphasize the robustness of statistical modelling approaches for the indirect estimation of infant mortality in small areas. Insights into the reliability of estimates based on age cohorts of women, particularly those aged 20-24 and 25-29, were discussed. However, challenges arise due to the limited observations within these age groups, posing constraints on accurately representing the early-age mortality experience of the area.

The sensitivity of required sample size to mortality levels is highlighted, indicating the necessity for larger sample sizes with rarer events. Assumptions underpinning the method, including a linear mortality decline and constant mortality and fertility patterns, are outlined (Daniel et al., 2021; Neupert et al., 2019; Reisman et al., 2016). The papers delve into statistical modelling approaches, such as the application of the Beta Binomial distribution, described as robust despite potential biases introduced by gradually falling fertility.

The literature addressed the imperative need for comprehensive cross-country studies, emphasizing the evolution of neonatal, infant, and child mortality indicators. The gaps identified regarding health inequality on a global scale underscore the importance of effective policy interventions. Additionally, the challenges related to sample size, the reliability of estimates, and the robustness of statistical modeling approaches for indirect mortality estimation in small areas are discussed, contributing valuable insights to the existing literature (Burstein et al., 2018; Jaskari et al., 2020; Mazzucato et al., 2023; Reisman et al., 2016). Overall, the papers advocate for continued efforts to enhance mortality estimation methodologies and policy interventions aimed at reducing global health inequality.

2.7 Long Short-Term Memory (LSTM) Networks

Long Short-Term Memory (LSTM) networks, a subset of recurrent neural networks (RNNs), excel in learning long-term dependencies, making them particularly suitable for time series prediction tasks (Yadav et al., 2023

LSTM networks, when combined with other neural network architectures such as convolutional neural networks (CNNs), offer the capability to extract both spatial and temporal features, contributing to more accurate predictions (P et al., 2024). Noteworthy applications include predicting air quality through learning from past pollutant concentrations and meteorological elements, making LSTM networks' adaptability to diverse time series prediction challenges.

Given the intertwined nature of factors influencing infant mortality, including healthcare access, environmental conditions, and socioeconomic variables, LSTM networks hold promise in integrating diverse data sources. This integration stands to enhance the accuracy of mortality rate predictions, demonstrating the potential of LSTM networks in addressing the intricate challenges associated with infant mortality estimation.

2.8 Time Series Analysis

Analysing temporal patterns and seasonal trends in infant mortality rates in rural Kenya through time series and machine learning models yields valuable insights. Time series analysis uncovers patterns over time, while machine learning models like Random Forest and XGBoost excel in accurately predicting infant mortality, as evidenced by high accuracy rates (Mfateneza et al., 2022). Spatial survival models, assessing spatial variation's impact on under-five mortality, underscore the significance of the spatial structure in child mortality patterns. This suggests that healthcare interventions should be tailored to specific counties (Daniel et al., 2021).

Model performance was evaluated using metrics such as accuracy, recall, precision, F1 score, and Area under the Receiver Operating Characteristics (AUROC). In a Rwanda study employing the Random Forest model, accuracy reached 84.3%, recall at 91.3%, precision at 80.3%, F1 score at 85.5%, and AUROC at 84.2%, showcasing the robust predictive capabilities of these models (Mfateneza et al., 2022). Their capacity to unveil hidden information not detectable by traditional statistical methods positions them as powerful tools for predicting infant mortality.

In the Kenyan context, the spatial survival model with spatial frailties outperformed the non-frailty model, suggesting improved predictive ability through the incorporation of spatial elements (Daniel et al., 2021). The models also revealed spatial disparities in child mortality rates, emphasizing county-specific variations (Daniel et al., 2021). According to Yang, et al. (2017), leveraging time series and machine learning models, including spatial survival models, was crucial for analysing temporal and seasonal trends in infant mortality rates in contexts befitting rural Kenya. These models enhance our understanding of how mortality risk evolves over time, providing valuable insights to guide targeted healthcare interventions at the county level.

2.9 Survival Analysis with Time-Dependent Covariates

Survival analysis with time-dependent covariates is a statistical method integral for estimating survival probability and hazard functions over time, accounting for variables that evolve during the study period (Andersen, 2014). This method is particularly pertinent in the realm of infant mortality prediction, where factors like maternal health, access to healthcare, and environmental conditions dynamically influence mortality risk. Several studies contribute to the understanding and application of this approach in the context of infant mortality.

A study by Smith-Greenaway and Thomas (2014) employed Aalen's additive model, offering a non-parametric approach to survival analysis. This model effectively estimated the effects of covariates changing over time on the hazard rate. Another research proposed introducing breakpoints in regression models to address potential inaccuracies in background mortality, which could be adapted for time-dependent covariates in survival analysis. Commonly used in survival studies, Cox proportional hazards regression analyses enable the estimation of hazard ratios with time-varying covariates. Additionally, discrete-time logit models have been applied to address censoring and age-related effects on mortality risks, suggesting potential integration with time-dependent covariates for a comprehensive analysis.

In the context of child mortality, the influence of maternal migration patterns on child mortality risks has been explored, emphasizing the significance of considering time-varying factors in survival analysis (Zhou et al., 2018). Additionally, the assessment of seasonality concerning preterm births provides insights into temporal variations of risk factors, relevant for time-dependent covariate analysis in infant mortality studies.

In Kenya, survival analysis with time-dependent covariates proved crucial for understanding and predicting infant mortality, allowing for the incorporation of dynamic factors like maternal health, healthcare access, and environmental conditions. Employing Aalen's additive model and Cox proportional hazards regression analyses, studies have estimated the effects of time-varying covariates on infant mortality (Gatabazi et al., 2019). Spatial survival models capture spatial variation in under-five mortality, emphasizing the importance of considering specific counties in healthcare planning. The spatial frailty proportional hazards model enhances predictive ability, revealing county-specific mortality rates.

Moreover, maternal education emerged as a significant factor for child survival, with secondary education linked to higher chances of survival compared to no or only primary education. Survival analysis with time-dependent covariates, including spatial survival models, played a crucial role in analysing and predicting infant mortality rates in Kenya (Daniel et al., 2021). These models contributed to understanding how mortality risk evolves over time, offering valuable insights to guide targeted healthcare interventions at the county level.

2.10 Bayesian Network Model

The construction of a Bayesian Network Model for indirect estimation of mortality in Kenya can be informed by various studies. A biological Bayesian network was developed for predicting adverse outcomes in acutely ill patients, which could be adapted for mortality (Wong et al., 2023). Another study highlighted the use of a spatial survival model with spatial frailties, which performed better than non-frailty models in predicting under-five child mortality in Kenya, emphasizing the significance of spatial variation in mortality rates. Additionally, a synthetic health measure framework was discussed for indirect estimation of cause-specific mortality rates, which could be relevant for developing a Bayesian Network Model focused on mortality. Bayesian networks have also been applied in the context of COVID-19 severity classification, demonstrating their utility in predictive accuracy, which could be leveraged for mortality estimation. Furthermore, machine learning techniques, including boosting and regression, have been used for mortality modelling and could be integrated into the Bayesian Network Model to differentiate causes of death.

2.10.1 Theoretical Framework

The theoretical framework guided this scientific basis of the research and offered a lens for understanding the phenomena under study, i.e., neonatal, infant and child mortality. The present study was informed by theories of health, life course, and epidemiology.

First, is Health Determinants Theory, which focuses on the correlation between social, environmental, and economic factors that influence health outcomes, which according (Frank et al., 2020), to this approach, infant mortality is not just a biological issue but is deeply affected by determinants like access to healthcare, maternal education, socio-economic conditions, and environmental exposures (Frank et al., 2020). These factors identified by Frak et al (2020) and Santosa et al. (2014) align well with the independent variables selected for the present study, such as access to clean water, teenage pregnancy rates, and maternal education levels. By identifying key social determinants, the study aligns with this theory to assess how these determinants predict mortality outcomes.

Importantly, is the Epidemiological Transition Theory, which states that as countries advance and develop in terms of technology, healthcare, and many other social aspects, they experience a shift from infectious diseases to chronic diseases as some of the major mortality cases in neonates, infants and U5s (Santosa et al., 2014).

Third, is the Life-Course Theory, which was essential in studying infant and child mortality as it acknowledges the impact of early-life conditions on long-term health outcomes. The early health conditions of mothers, such as access to antenatal care and proper nutrition, play a pivotal role in determining neonatal and infant mortality (Jones et al., 2019). The variables to model (e.g., prenatal care and maternal health) were consistent with this theory.

2.10.2 Conceptual Framework

The conceptual framework was a more practical approach that linked the key variables of the study, showing how they interact to influence the outcome (child mortality).

1. **Independent Variables**: The conceptual framework for this study identified a variety of predictors based on social determinants of health and environmental conditions:

Healthcare Access*:* Measured through access to skilled birth attendants, antenatal visits, post-natal care visits, etc.

Socioeconomic Indicators*:* Including maternal education, overall poverty, teenage pregnancy rates, and access to basic sanitation and clean drinking water.

Health-Environment Factors*:* These include malnutrition rates, vaccination coverage, HIV infection rates, respiratory complications, and malaria prevalence.

2. **Mediating Factors**: Certain factors, such as geographic disparities between Nyamira and Migori counties, acted as mediating variables. These were reflected in the use of spatial analysis in the study, which helped capture location-based differences that influenced the effectiveness of interventions.
3. **Dependent Variables**: The primary outcome was mortality, measured in three categories:
    - Neonatal mortality per 1000 live births
    - Infant mortality per 1000 live births
    - Under-5 mortality per 1000 live births
4. **Modelling Techniques**: The conceptual framework integrates both traditional statistical approaches, such as multiple linear regression (MLR), and machine learning models like Gradient Boosting Regressor (GBR) and Spatiotemporal Autoregressive (STAR) models.

5. **Data Sources**: The use of historical data (from 1989-2022) allowed the study to identify trends and make predictions using various social determinants and health-related variables. The conceptual framework emphasizes the importance of reliable data collection (Kenya Demographic and Health Survey, WHO records, etc.) to ensure the accuracy of the model predictions.

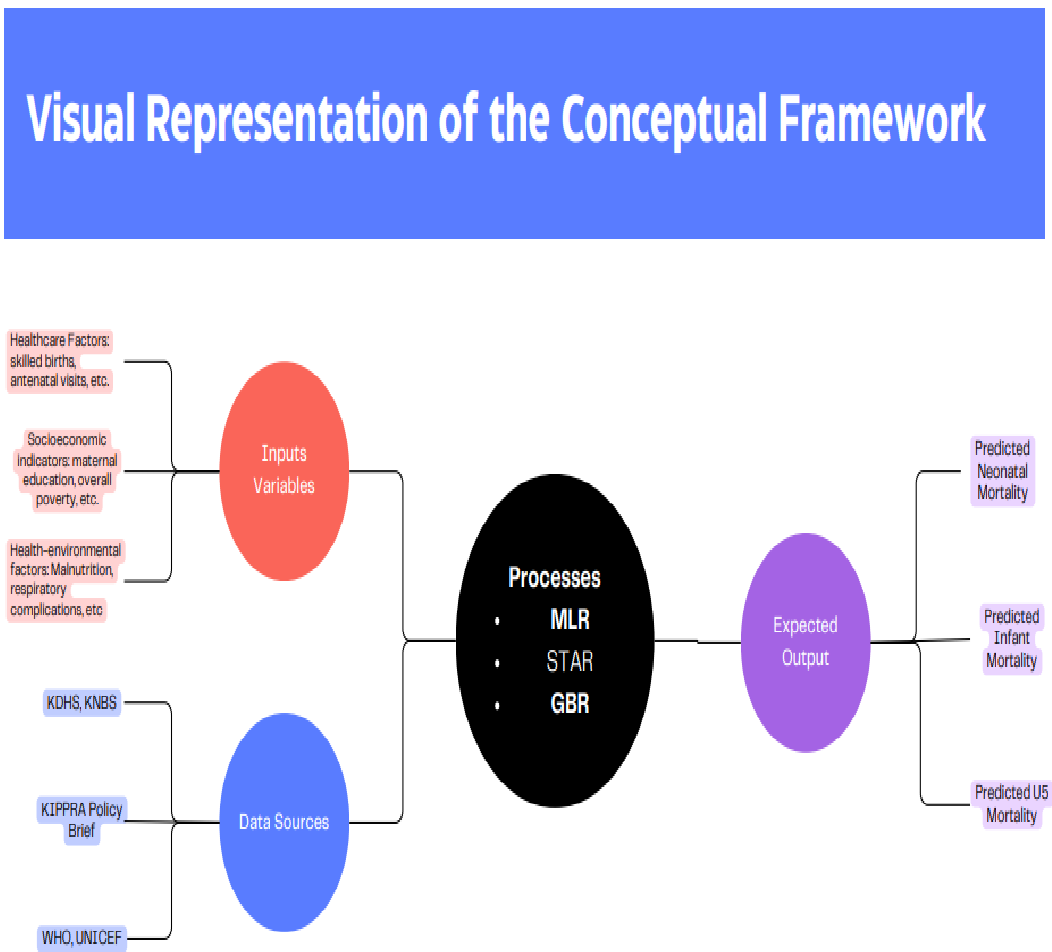

Figure 1: Visual Representation of Conceptual Framework (Source: Original)

CHAPTER THREE

METHODOLOGY

3.1 Introduction

Chapter three delved into the methodological framework employed in this study to examine and predict child mortality in Nyamira and Migori counties, Kenya. It details the elected study design, outlining the employed approach and its justification. The section comprehensively describes various data collection methods utilized, highlighting the specific data sources preferred and the strategies implemented to address challenges such as incomplete data and survivor bias. The methodology section also outlined the data analysis strategies deployed through the models used, both for temporal pattern analysis and predictive model development.

3.2 Study Design

The study employed a retrospective cohort study design that encompassed diverse datasets, including current and historical data on neonatal, infant, and under-five mortality rates. Cohort studies are types of observational studies in which a cohort, or a group of individuals sharing some characteristic, are followed up over time, and outcomes are measured at one or more time points (Capili & Anastasi, 2021). The cohort study design is an excellent method to understand an outcome or the natural history of a disease or condition in an identified study population. Since participants do not have the outcome or condition at study entry, the temporal causality between exposure and outcome(s) was assessed using this design. A vital feature of a cohort study was selecting the study participants based on mutual characteristics such as geographic location, birth year, or occupation. This type of study was feasible due to sufficient availability and access to various datasets befitting the research questions. The datasets were also adequate measurements of the selected predictor variables.

The study covered data for a period of 34 years (1989-2022), drawing from demographic studies, household surveys, census, electronic health records, and official databases. The research adopted a retrospective cohort study design to achieve the outlined objectives as this design was suitable for assessing the association between exposure and outcome over time, allowing for the examination of temporal patterns and trends in infant mortality. The focus was on low-income counties in Kenya, specifically Nyamira and Migori.

3.3 Study Area and Population

The present study utilized data on neonate, infants, and under-5 children born between 1989 and 2022 in Nyamira and Migori counties as the defined cohort. The selection was based on the availability of data for the specified period. Participants with congenital anomalies incompatible with life or conditions that may significantly impact mortality risk were all excluded. A 34-year study period was preferred because it was considered to sufficient to observe trends, seasonal patterns, and potential shifts in risk factors affecting mortality. The study population was all neonates, infants, and under-five children whose admissions were recorded in different databases between 1989 and 2022. All under-five child admission records at the NICU at the public hospitals during the study period were included in the study. The exclusion criteria included participants admission records with incomplete information, revisits, and any neonates' records with withdrawal from a treatment or history of self-discharge.

3.4 Sample Size and Sampling Procedure

To calculate the sample size for retrospective cohort research on infant mortality, first was to estimate the expected number of events (child deaths), the estimated incidence rate in the population, and the desired degree of statistical power (Andreev & Kingkade, 2015). The formula for calculating sample size in a cohort study was expressed as:

$$n = \left(\frac{(z_{\alpha/2} + z_{\beta})^2 * \left(p_1(1-p_1) + p_2(1-p_2)\right)}{(p_1 - p_2)^2} \right) * \frac{1}{DEEF}$$

Where:

- n is the sample size needed in each group,
- $z_{\alpha/2}$ is the z-value corresponding to the desired alpha level (i.e., 1.96 for a 95% confidence interval),
- z_{β} is the z-value corresponding to the desired power (e.g., 0.84 for 80% power),
- p_1 is the expected proportion of events in the exposed group (0.10),
- p_2 is the expected proportion of events in the non-exposed group (0.02),

DEFF is the design effect that takes into account research design elements such as clustering (DEFF=14.38 for simple random sampling). Substituting for values in the formula, it yielded a sample size of 34, hence the arrival at the period between 1989-2022.

Historical data on infant mortality rates can be used to predict the proportion of events in retrospective cohort research focusing on child mortality. In the context of Kenya, where the under-five mortality rate was reported to be 51.318 per 1000 live births, this rate was used as a starting point for and, correcting for the under-five age group population in the study (Tesema et al., 2021).

3.5 Ethical Approval

The data analysed in this study were sourced from the Kenya Demographic and Health Survey. Authorization to use the data set was obtained through the DHS Portal, which contains a repository of the survey data. The Demographic and Health Survey is an initiative sponsored by the United States Agency for International Development (USAID) in partnership with other Kenyan research agencies, including the Kenya National Bureau of Statistics (KNBS), National Council for Population and Development (NCPD), Ministry of Health, Kenya Medical Research Institute (KEMRI), and National AIDS Control Council (NACC). The DHS program has been run in many nations to provide periodic updates, outlooks, and estimates of various indicators, such as maternal and child health and individual level information pertaining to the health of such individuals in specific cases. The data collected contains information about families, children, child bearing mothers, and other socioeconomic factors as relevant to help monitor the population and health status in Kenya.

3.6 Data Sources and Collection

The study utilised secondary data from past surveys, public health databases, hospital admission records, social media platforms, maternal health programs, genetic studies, environmental monitoring agencies, demographic databases, and health records. The data underwent monitoring and quality control for completeness and accuracy. Hence, validation for data quality, addressing missing values, outliers, and inconsistencies were addressed at this stage.

Table 3.1: Data Sources

Category	Variable	Description	Source
Data Sources	Survey data (CBHs, SBHs, Census, DHS, MICS)	Information on births, deaths, demographics, health	Surveys
	Public health databases	Hospital records, genetic studies, environmental data	Health-related information
	Health records	Maternal health program data	Maternal health information
	Social media platforms	Community concerns, health behaviours, information diffusion	Social determinants of health
Methodology	CBH/SBH data	Detailed birth & death history for individual children	Surveys
	Social media data	Insights into community health & behaviours	Social determinants of health
	Demographic data	Age, gender, etc.	Population characteristics

3.7 Variables of Interest

The under-five child mortality rate in Kenya has been influenced by various factors, including maternal education, socioeconomic status, access to healthcare services, and geographical disparities. Studies have shown that higher maternal education levels are associated with lower child mortality rates, as educated mothers are more likely to utilise maternal and child health services effectively.

Table 3.2: Study Variables

Category	Variable	Description
Outcome Variable	Neonatal mortality	Annual average/1000
	Infant mortality	Annual average/1000
	U5M	Annual average/1000
Independent Variables	Age	Child's age
	Gender	Child's gender
	Income	Household income
	Education	Mother's education level
	Vaccination coverage	Child's vaccination status
	Access to clean water	Household access to clean water
	Sanitation	Household sanitation facilities
	Healthcare availability	Presence of healthcare facilities

	Skilled birth attendance	Use of skilled birth attendants
Additional Variables (specific study)	Mode of delivery	Mother's delivery method
	Place of delivery	Location of child's birth
	HIV test, Syphilis test, Hepatitis B test	Mother's test results
	Maternal status	Mother's health status
	Sex	Newborn's sex
	Birth weight	Newborn's weight at birth
	Temperature, APGAR score, Length of stay	Newborn's vital signs and hospital stay
	Resuscitation, Prematurity, Sepsis	Newborn's medical conditions
	Respiratory distress syndrome, Asphyxia	Newborn's respiratory issues
	Congenital malformation	Newborn's birth defects
	Kangaroo mother care (KMC)	Use of skin-to-skin care for newborn

3.8 Model Evaluation using Indirect models

The indirect methods for infant mortality estimations were illustrated as below:

3.9 Multiple Linear Regression (MLR)

Multiple Linear Regression (MLR) was particularly suitable for this study due to its ability to quantify the relationship between multiple independent variables and a continuous dependent variable. In the context of estimating and predicting mortality rates among neonates, infants, and under-5 children, MLR allows for the examination of how various socio-economic and health-related factors contribute to these mortality outcomes. MLR method was valuable because it could isolate the effect of each independent variable while controlling for the influence of other variables, thereby providing a more accurate and comprehensive understanding of mortality determinants.

Let Y represent the dependent variable, which in this case can be one of the three mortality outcomes: neonatal mortality per 1,000 live births $(Y_{neonatal})$, infant mortality per 1,000 live births (Y_{infant}), or under-5 mortality per 1,000 live births Y_{U5}. The independent variables, denoted by $X_1, X_2, X_3 \ldots, X_p$, include factors such as the percentage of women with no formal education, basic sanitation coverage, access to clean water, and other socio-economic indicators. The general form of the Multiple Linear Regression model was expressed as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon$$

Where:

- Y is the dependent variable (mortality rate),
- β_0 is the intercept, representing the expected value of Y when all X_i variables are zero,
- $\beta_1, \beta_2, \ldots, \beta_p$ are the coefficients representing the change in Y for a one-unit change in each X_i,

- ϵ is the error term, capturing the variation in Y not explained by the independent variables.

The coefficients $\beta_1, \beta_2, \ldots, \beta_p$ are estimated using the Ordinary Least Squares (OLS) method, which minimizes the sum of squared residuals, the differences between observed and predicted values of Y.

Justification for the Methodology

3.9.1 Suitability for Continuous Outcomes

MLR has been perfect for the present study, as the dependent variables (mortality rates) are still present. The approach assumes a linear relationship of the dependent and independent parameters, which was a reasonable estimation at the time of dealing with socioeconomic data. The present one-dimensionality allows the interpretation of the coefficient directly within the provisions of the expected change in mortality rates for the unit of measurement changed by the respective forecaster.

3.9.2 Ability to Control for Multiple Factors

Mortality rates end up getting shaped by all kinds of things. Healthcare access plays a role. Education matters too. Socio-economic conditions come into it as well. MLR made it possible to bring in several independent variables right at the same time. The model could then deal with how these factors connect in tricky ways. In that sense, MLR helped spot the exact effect from each variable. It kept the others in check at the same time. That approach turned out to be essential for reaching reliable conclusions. All of this happened in a setting with multiple variables involved..

3.9.3 Interpretability

One of the key strengths of MLR was its interpretability. The estimated coefficients provided direct insights into how each independent variable affects mortality rates, which is critical for policymakers and healthcare practitioners who need to understand the key drivers of mortality. For instance, a significant positive coefficient for Malnutrition % would indicate that higher malnutrition rates are associated with increased mortality, suggesting a need for targeted nutritional interventions.

3.9.4 Assessment of Model Fit

MLR provides various metrics, such as R-squared and p-values, to assess the model's fit and the significance of individual predictors. The R-squared value indicates the proportion of variance in the dependent variable explained by the model, giving a sense of how well the model captures the underlying relationships. P-values, on the other hand, help in identifying which predictors have a statistically significant impact on mortality, guiding the focus towards the most critical risk factors.

3.9.5 Applicability in Public Health Research

MLR is widely used in public health research for its robustness and flexibility in handling various types of socio-economic and health-related data. The method's ability to accommodate both continuous and categorical variables (after proper encoding) makes it versatile for analysing diverse datasets like the one used in this study.

In summary, Multiple Linear Regression is a robust, interpretable, and suitable method for this study as it allows for the estimation of the relationship between multiple risk factors and mortality rates while controlling for confounding variables. This makes it an essential tool for understanding and addressing the determinants of neonatal, infant, and under-5 mortality.

3.10 Spatio-Temporal Modelling

Rationale: Given the geographical dimension of the data, a spatio-temporal model was applied. This method considered both spatial and temporal variations in infant mortality rates, incorporating environmental data and demographic factors. It can identify clusters of high or low mortality rates over time, providing insights into spatial patterns and temporal trends.

The present study employed a spatiotemporal regression model, specifically the spatiotemporal autoregressive model (STAR), which captured both spatial and temporal dependencies incorporating environmental and demographic factors to estimate infant mortality rates.

The study defined Y_{st} as the child (neonatal, infant, or U5) mortality rate at spatial location s and time t, and X_{st} to represent the covariates (environmental and demographic factors). The basic STAR model equation was represented as: $Y_{st} = \alpha + \beta X_{st} + \rho W Y_{st} + \Phi_t + \epsilon_{st}$, where:

α is the intercept term.

β is the vector of coefficients for the covariates.

ρ captures the spatial autocorrelation, where *W* is the spatial weights matrix.

Φ_t represents the temporal trend or seasonality.

ϵ_{st} is the error term assumed to be normally distributed with mean 0 and constant variance. The spatial weight matrix *W* specifies the spatial relationship between different locations. It reflects how much influence the infant mortality rate at one location has on its neighbouring locations. For incorporating environmental and demographic factors, X_{st} can include various variables such as pollution levels, socioeconomic status, access to healthcare, maternal education, etc. The temporal component (Φ_t) was modelled using time-specific variables and polynomial functions to capture temporal trends.

3.11 Gradient Boosting Machines (Gradient Boosting Regressor (GBR) Model)

The gradient boosting regressor model was achieved by following the given modelling steps

3.11.1 Rationale for Model Selection

Given the dataset's temporal span from 1989 to 2022, comprising 34 annual observations, the choice of an appropriate predictive model is critical. The study aims to predict child mortality, represented by the dependent variables ‘Neonatal Mortality/1000,’ ‘Infant Mortality/1000,’ and ‘U5 Mortality/1000,’ based on a set of socio-economic and health-related independent variables. Given the relatively small sample size, a Gradient Boosting Machine (GBM) model, specifically the Gradient Boosting Regressor from the scikit-learn library, was selected for this study. This model is known for its robustness in handling small datasets while minimizing overfitting, a significant concern given the limited number of observations.

3.11.2 Model Design

The Gradient Boosting Regressor is an ensemble learning method that builds a predictive model in a stage-wise fashion from weak learners, typically decision trees. Each tree in the sequence attempts to correct the errors made by the previous one, effectively reducing bias and variance, and improving the model’s predictive accuracy.

3.11.3 Input Variables

The model incorporates 17 independent variables: ‘Women No Formal Edu,’ ‘Basic Sanitation (%),’ ‘Clean Water Access,’ ‘Access to Clean Fuel,’ ‘Teenage Pregnancy (%),’ ‘Skilled provider births,’ ‘ANC 4+ Visits,’ ‘Post-natal Visits,’ ‘Malnutrition (%),’ ‘HIV Infection Rate,’ ‘Respiratory Complications,’ ‘Vaccination Coverage (%),’ ‘Non Exclusive Breastfeeding,’ ‘Malaria Prevalence (%),’ ‘Overall Poverty (%),’ ‘Solid Fuel Use,’ and ‘Location.’

3.11.4 Dependent Variables

Three dependent variables are used to measure child mortality: ‘Neonatal Mortality/1000,’ ‘Infant Mortality/1000,’ and ‘U5 Mortality/1000.’ Separate models are constructed for each dependent variable to capture specific risk factors influencing each aspect of child mortality.

3.11.5 Hyperparameter Tuning:

The model parameters were fine-tuned using GridSearchCV, with a particular focus on the $learning\_rate$, $n\_estimators$, and max_depth parameters. A smaller learning rate was chosen to improve generalization, while $n\_estimators$ (the number of boosting stages) was optimized to balance model complexity and performance. The max_depth of the trees was kept shallow to avoid overfitting, given the size of the dataset.

3.11.6 Data Notation

Data for each year from 1989 to 2022 was denoted as $X_{t,i}$ where t represents the year and i represents the specific independent variable. The dependent variable for each year t was denoted as $Y_{t,d}$ where d corresponds to either neonatal, infant, or under-five mortality. The relationship modelled can be expressed as:

$$Y_{t,d} = f\left(X_{t,1}, X_{t,2}, \dots, X_{t,n}\right) + \epsilon_t$$

where $f(\cdot)$ was the function estimated by the Gradient-Boosting-Regressor and ϵ_t represents the residual error for year t.

3.11.7 Justification

The Gradient-Boosting-Regressor was chosen for its ability to handle small datasets with complex interactions between variables. Unlike linear models, which may oversimplify relationships, the Gradient Boosting approach captured non-linear patterns and interactions critical for understanding the different and dynamic determinants of child mortality. The model's iterative approach to reducing prediction errors ensured that it provided a well-balanced fit, avoiding the shortcomings of both overfitting and underfitting. Additionally, the model's output includes feature importance metrics, enabling the identification of the most significant risk factors contributing to child mortality, which was a core objective of this study. Given these considerations, the Gradient-Boosting-Regressor was deemed a suitable and effective method for the predictive modelling of child mortality in this context.

3.12 Data Analysis

Data was stored, organized, and cleaned using MS Excel. Data was analysed using the Jupyter environment of Python. The output was presented through tables and graphs, and interpreted through text.

CHAPTER FOUR

RESULTS

4.1 Introduction

Chapter 4 presents the results and findings of this study, encompassing a comprehensive analysis through various statistical and modelling techniques. Initially, descriptive statistics provide an overview of the dataset, highlighting key trends and distributions. This was followed by multiple linear regression modelling to identify significant predictors and their relationships with the dependent variable. Gradient boosting was then employed to enhance predictive accuracy and model performance. Finally, spatiotemporal modelling was utilized to capture and analyse patterns over time and space, offering deeper insights into the dynamics of the three categories of child mortality.

4.2 Developing a context-specific predictive model

This section gives a summary of descriptive statistics on neonatal mortality, infant mortality and under 5 year's mortality respectively.

4.2.1 Descriptive Summary Statistics

Table 4.1 Migori County Descriptive Statistics

	Neonatal Mortality	Infant Mortality	U5 Mortality
Count	34.000	34.000	34.000
Mean	31.391	61.247	116.701
Std	6.271	13.335	37.608
Min	22.500	32.271	73.000
25%	26.530	51.020	83.250
50%	31.000	64.094	117.000
75%	36.000	73.138	133.875
Max	43.000	77.3000	194.700

The output in the above table summarizes the neonatal, infant, and under-five mortalities in the county of Migori between 1989-2022. The total neonatal rate during the study period was 31.391 per 1000 deaths, which was much greater than the national average (20 per 1000 deaths) and even the 2030 MDG proposed targets (12 deaths per 1,000 live births). The standard deviation was 6.271, an indicator that the annual neonatal mortality rates differed by a wide margin from the mean. The highest neonatal rate ever recorded was 43/1000 deaths while the lowest was 23/1000 live births. Infant mortality was slightly higher than neonatal mortality with a mean of 61.247. With a standard deviation of 13.335, the minimum deaths recorded per 1000 live births throughout the study period were 51.020 while the maximum deaths recorded were 77.3. Children under five years recorded the highest death rate during the years 1989-2022 with a mean of 116.7

per/1000 live births and a median of 117.00. The minimum recorded rate was 73/1000 live births while the maximum was 194.7/1000 live births.

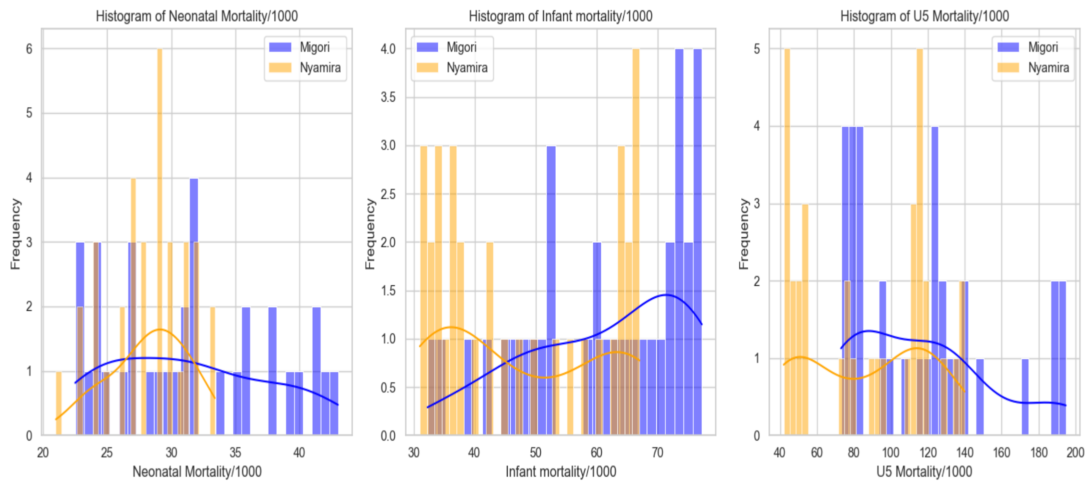

Figure 4.1: Comparative Histograms

Table 4.2: Nyamira County Child Mortality Statistics

	Neonatal Mortality	*Infant Mortality*	*U5 Mortality*
Count	34.000	34.000	34.000
Mean	28.182	47.863	86.944
Std	3.110	13.694	33.558
Min	21.000	31.000	42.000
25%	26.250	35.395	51.250
50%	29.000	43.750	91.500
75%	30.450	63.595	115.000
Max	33.400	67.060	140.000

For Nyamira County, the average child mortality rates were lower than those recorded in Migori county within the study period. The average neonatal mortality rate was 28.182

per 1000 live births. The annual neonatal mortality rates were dispersed from the mean with a standard deviation of 3.110. The minimum neonatal mortality rate recorded in Nyamira county within the study period was 21/1000 live births while the maximum was 33.4/1000 live births. The infant mortality rate between 1989 and 2022 was an average of 47.863/1000 live births. With a 13.694 deviation from the mean, the minimum infant deaths recorded were 31 while the most deaths were 67.060/1000 live births. For under five children, the average mortality rate was 86.944 with a standard deviation of 33.558. The minimum deaths recorded per 1000 live births were 42.000 while the most deaths within the study period were 140 as the histograms and box plots below visualize.

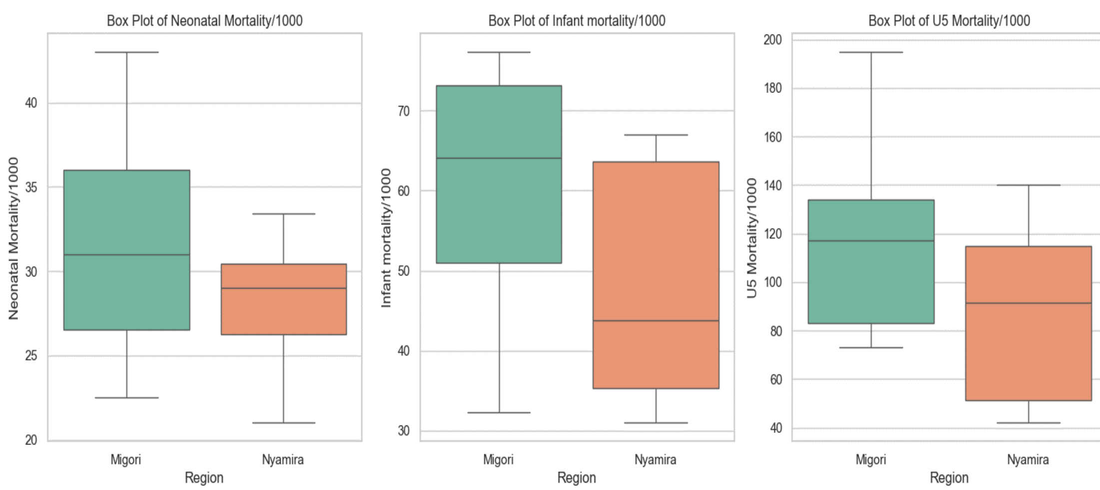

Figure 4.2: Comparative Box Plots

4.3 Analysing survey and demographic data to determine social determinants of infant mortality rate in Nyamira and Migori counties, Kenya.

This section used regression models to analyse both survey and demographic data to assess the impact of social determinants of health on the infant mortality rate in Nyamira and Migori counties.

4.3.1 Multiple Linear Regression (MLR) Model

Table 4.3 displays results of a multivariate regression model which was conducted to determine the factors that affecting mortality rate in Migori County. The dependent variable was neonatal mortality per 1000 from 1989 to 2022. As table 4.3 summarizes, R-squared was 0.899, an indication that the model was adequate as it explained over 0.5 (50%) of the variation in the response variable. As such, 89.9% of the change in neonatal deaths was accounted for by the model and was due to the 17 predictor variables. With a p-value (2.10e-13) less than the 0.05 level of significance, there was sufficient evidence to rule that the model was statistically significant with 95% confidence of not committing a Type I error.

Table 4.3: Factors Affecting Neonatal Mortality in Migori County

Dep. Variable:	*Neonatal Mortality/1000*	R-squared:	*0.899*
Model:	*OLS*	Adj. R-squared:	*0.809*
Method:	*Least Squares*	Prob (F-statistic):	*2.10e-13*
Sample Size:	*34*	Log-Likelihood:	*-33.788*
Df Residuals:	*17*	AIC:	*101.6*
Df Model:	*16*	BIC:	*127.5*
Covariance Type: *non-robust*			

Under *table 4.4,* the model coefficients were displayed. With a coefficient of -5.8461 and a p-value of 0.003, lack of access to clean water was statistically significant, hence a good predictor for infant mortality in Migori county. The implication was that for every unit increase in lack of access to clean water access, neonatal mortality decreases by approximately 5.85 units, demonstrating the critical role of clean water in improving neonatal health outcomes. Second, were post-natal visits, with a coefficient of -5.4261 and p-value 0.024. As such, there was sufficient evidence to reject H_0 at 0.05 level of significance and rule that increase in neonatal mortality was partly due to a lack of post-natal visits. As such, a lack of post-natal care in Migori was associated with an increase of about 5.43 deaths per 1000 live births, signifying the importance of post-natal care in ensuring neonatal survival.

The analysis further revealed that neonatal deaths in Migori were as a result of respiratory complications. At $F(1,16) = -0.3912$, $p=0.001$, respiratory complications are significantly associated with increased neonatal mortality. For every unit increase in respiratory complications, neonatal mortality increases by approximately 0.39 units, emphasizing on the need for effective management of respiratory issues in neonates. Mothers that did not receive at least four sessions of ANC visits were associated with neonatal deaths with $F(1, 16) = 2.7559$, $p=0.053$. Despite the marginal deviation of the p-value from the 0.05 level of significance, there was no meaningful zero between the lower and upper confidence intervals of ANC 4+ visits (-6.643, -1.131), making it a good predictor of infant mortality in Migori County. For $F(1, 16) = -0.3139$, $p = 0.032$, non-exclusive breastfeeding also had a significant effect on neonatal deaths, hence a significant predictor. Each unit increase in non-exclusive breastfeeding is associated with an increase of about 0.31 units in neonatal mortality. Promoting exclusive breastfeeding could therefore be beneficial in reducing neonatal deaths. Finally, was malaria prevalence in neonates (coef = -3.0453, p= 0.033). Malaria prevalence was established to significantly increases neonatal mortality. For every unit increase in malaria prevalence, neonatal mortality increases by approximately 3.05 units.

Several other variables played a role in influencing neonatal mortality but lacked statistical significance, hence not good predictors of the model. These included Women with no formal education, basic sanitation, teenage pregnancy, births attended by skilled providers, malnutrition, HIV infection rate, vaccination coverage, overall poverty, solid fuel use, and the location (rural/urban). After accounting for the effects of the significant predictors, these variables do not provide additional explanatory power for neonatal mortality.

Table 4.4: Coefficients with p-values and 95% Confidence Intervals

	coeff	std err	t	P>|t|	[0.025	0.975]
constant	31.3909	0.159	198.006	0.000	31.056	31.725
Women No Formal Education	0.8120	1.267	0.641	0.530	-3.485	-1.861
Basic Sanitation	-1.8852	1.501	-1.256	0.226	-5.051	1.281
Clean Water Access	-5.8461	1.710	-3.419	0.003	-9.454	-2.238
Teenage Pregnancy	0.6942	0.608	1.141	0.270	-0.589	1.978
Births by Skilled Provider	1.2378	1.645	0.752	0.462	-2.233	4.709
ANC Visits (4+)	2.7559	1.842	1.496	0.053	-6.643	-1.131
Post-natal Visits	-5.4261	2.192	-2.476	0.024	-10.051	-0.802
Malnutrition	-0.2472	1.014	-0.244	0.810	-2.386	1.892
HIV infection Rate	-0.3195	0.517	-0.618	0.545	-2.386	1.892
Respiratory Complications	-0.3912	2.541	-0.154	0.001	-9.753	-4.971

Vaccination Coverage	-1.6404	1.315	-1.247	0.229	-4.415	1.134
Non-Exclusive Breastfeeding	-0.3139	0.901	-0.348	0.032	-2.214	-1.587
Malaria Prevalence	-3.0453	1.308	-2.327	0.033	-5.806	-0.285
Overall Poverty	-0.6414	0.476	-1.348	0.195	-1.645	0.362
Solid Fuel Use	-1.6565	1.409	-1.175	0.256	-4.630	1.317
Location	-0.0044	0.262	-0.017	0.987	-0.558	0.549

==

Omnibus:	1.642	Durbin-Watson:	2.543
Prob(Omnibus):	0.440	Jarque-Bera (JB):	1.322
Skew:	-0.294	Prob(JB):	0.516
Kurtosis:	2.234	Cond. No.	19.2

==

4.3.2 Test of Assumptions

One of the assumptions of MLR is normality. With Skewness (-0.294) and Kurtosis (2.234) falling within the -3 to 3 normality range, there is sufficient evidence to suggest that the data satisfied the normality assumption. Second, the omnibus test (1.642, prob = 0.440), p-value (greater than 0.05) indicated that the residuals were normally distributed. In this case, the p-value of 0.440 suggests that the residuals do not significantly deviate from normality. The data further satisfied the autocorrelation assumption, making it fit for conducting a multiple linear regression model. The Durbin-Watson Statistic (2.543) was used to test autocorrelation, with a value close to 2 indicate no autocorrelation. Here, the value of 2.543 suggests that there is no significant autocorrelation in the residuals.

There was need to use Jarque-Bera (JB) to test the assumption of normality of the residuals. At (JB = 1.322, Prob = 0.516), there was sufficient evidence that the residuals followed a normal distribution as the residuals plot below displays.

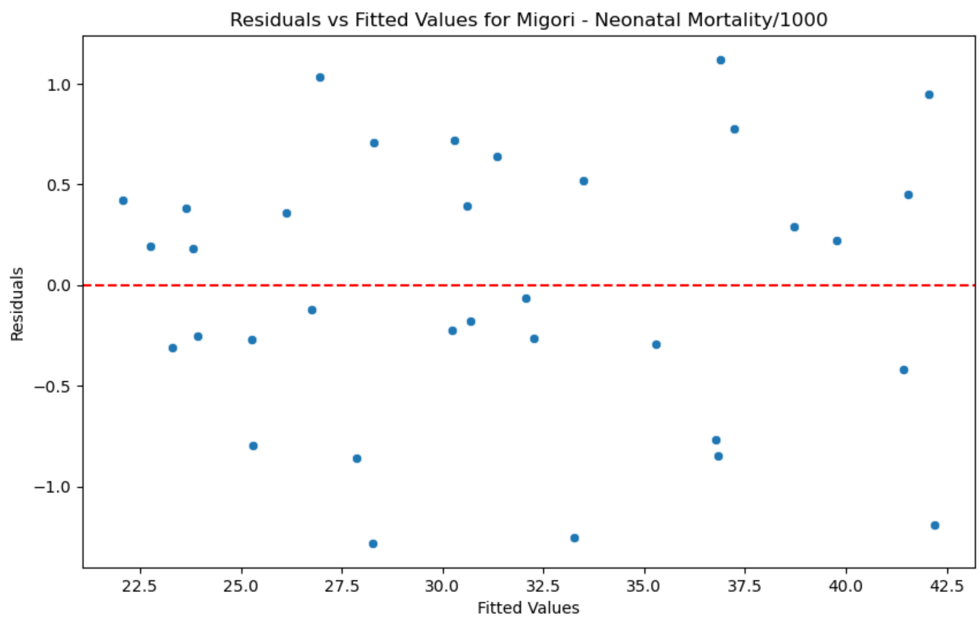

Figure 4.3: Residual Box Plots

The final assumption was that of multicollinearity, tested using Condition Number. At Cond. No. = 19.2, there was sufficient evidence to suggest the presence of low multicollinearity between the predictor variables. For this test, values above 30 indicate potential multicollinearity issues. The value of 19.2 suggests the absence of significant multicollinearity among the independent variables, which had no effect on the stability of the coefficient estimates.

4.3.3 Multivariate results of Infant Mortality Rate using Migori Data

The regression analysis identifies several significant predictors of infant mortality, as evidenced by their p-values and confidence intervals. The model's F-statistic of 9.215

with a p-value of 0.0000184 indicates that, overall, the model is statistically significant, meaning at least one of the predictors is related to infant mortality. Specifically, women with no formal education (p-value = 0.031, 95% CI [7.538, 27.007]), clean water access (p-value = 0.004, 95% CI [-36.083, -10.548]), post-natal visits (p-value = 0.030, 95% CI [-44.085, -15.69]), respiratory complications (p-value = 0.009, 95% CI [-36.98, -32.33]), non-exclusive breastfeeding (p-value = 0.048, 95% CI [0.120, 24.682]), malaria prevalence (p-value = 0.001, 95% CI [-18.07, -7.610]), and solid fuel use (p-value = 0.041, 95% CI [-26.41, -12.027]) were all statistically significant. These variables were considered good predictors of infant mortality because their p-values were below the conventional significance level of 0.05, indicating that there was a low probability that the observed relationships were due to chance. What's more, their confidence intervals do not cross zero, reinforcing the reliability of their effects on infant mortality.

Conversely, several variables were found to lack statistical significance in predicting infant mortality, including basic sanitation (p-value = 0.713, 95% CI [-16.831, 24.087]), teenage pregnancy (p-value = 0.607, 95% CI [-10.354, 6.234]), births by skilled providers (p-value = 0.912, 95% CI [-21.241, 23.618]), ANC visits (p-value = 0.685, 95% CI [-20.208, 30.028]), malnutrition (p-value = 0.983, 95% CI [-13.683, 13.97]), HIV infection rate (p-value = 0.613, 95% CI [-8.777, 5.329]), vaccination coverage (p-value = 0.514, 95% CI [-23.593, -12.27]), overall poverty (p-value = 0.884, 95% CI [-6.943, 6.032]), and location (p-value = 0.808, 95% CI [-3.993, 3.156]). These variables have p-values above 0.05, indicating that their relationships with infant mortality are not statistically significant, meaning any observed effects could likely be due to random variation. Moreover, their confidence intervals include zero, suggesting that these variables may have no real impact on infant mortality in the context of this study. The lack of significance of these variables suggests that, while they may still contribute to understanding the overall model, they do not independently explain much of the variation in infant mortality.

Table 4.5: Table of Infant OLS Regression, Migori County

Dep. Variable:	*Infant Mortality/1000*	R-squared:	*0.897*
Model:	*OLS*	Adj. R-squared:	*0.799*
Method:	*Least Squares*	F-statistic	*9.215*
Sample Size:	*34*	Prob (F-statistic):	*1.84e-05*
Df Residuals:	*17*	Log-Likelihood:	*-17.232*
Df Model:	*16*	AIC:	*228.5*
Covariance Type:	*non-robust*	BIC:	*254.4*

The above table indicates the model was adequate as it accounted for over 50% of the variation in the response variable. R-squared was 0.897, indicating that 89.7% of the variance in the response variable (Infant Mortality/1000) was explained by the independent variables included in the model. As such. 89.7% of the occurrences in infant deaths were due to the predictor variables included in the model. Adjusted R-squared accounted for the number of independent variables in the model, providing a more reliable measure of how well the model fits the data. In this case, it suggested that approximately 79.9% of the variance in Infant Mortality/1000 is explained by the predictor variables. The F-statistic tested the overall significance of the model, suggesting that the regression model as a whole met the statistical significance threshold. Prob (F-statistic), 1.84e-05 (or 0.0000184), was the p-value associated with the F-statistic, an indication that the probability of obtaining an F-statistic as extreme as the one observed, assuming the null hypothesis (that all coefficients are zero), was marginal. As such, the null hypothesis was rejected in favour of the alternative, suggesting that at least one of the independent variables was significantly related to Infant Mortality/1000.

Table 4.6: Factors Affecting Infant Mortality Rate in Migori County

	coef	std err	t	P>|t|	[0.025	0.975]
const	61.2474	1.025	59.782	0.000	59.086	63.409
Women No Formal Education	9.7349	8.187	1.189	0.031	7.538	27.007
Basic Sanitation	3.6277	9.697	0.374	0.713	-16.831	24.087
Clean Water Access	-12.7676	11.051	-1.155	0.004	-36.083	-10.548
Teenage Pregnancy	-2.0597	3.931	-0.524	0.607	-10.354	6.234
Births by Skilled Provider	1.1881	10.631	0.112	0.912	-21.241	23.618
ANC Visits (4+)	4.9102	11.905	0.412	0.685	-20.208	30.028
Post-natal Visits	-14.2001	14.165	-1.003	0.030	-44.085	-15.69
Malnutrition	0.1415	6.553	0.022	0.983	-13.683	13.97
HIV infection Rate	-1.7243	3.343	-0.516	0.613	-8.777	5.329
Respiratory Complications	12.327	16.424	0.142	0.009	-36.98	-32.33
Vaccination Coverage	-5.6631	8.498	-0.666	0.514	-23.593	-12.27
Non-Exclusive Breastfeeding	12.401	5.821	2.130	0.048	0.120	24.682

Malaria Prevalence	8.0453	8.456	0.027	0.001	-18.07	-7.610
Overall Poverty	-0.4556	3.075	-0.148	0.884	-6.943	6.032
Solid Fuel Use	7.1907	9.109	-0.789	0.041	-26.41	-12.027
Location	-0.4186	1.694	-0.247	0.808	-3.993	3.156

Omnibus:	40.470	Durbin-Watson:	2.633
Prob(Omnibus):	0.000	Jarque-Bera (JB):	182.444
Skew:	2.436	Prob(JB):	2.41e-40
Kurtosis:	2.249	Cond. No.	19.2

==

The model diagnostics provide further insight into the reliability of these findings. The F-statistic indicates that the model as a whole was statistically significant, but the non-normal distribution of residuals, as indicated by the Omnibus test and Jarque-Bera test, suggested potential issues with the model's assumptions. Despite this, the Durbin-Watson statistic shows no evidence of autocorrelation, which supports the validity of the model. The skewness and kurtosis values further confirmed that the residuals were normally distributed, which had no effect on the reliability of the coefficient estimates. As such, there was no need of considering model refinement or alternative modeling approaches since it already attained robust and reliable outcome.

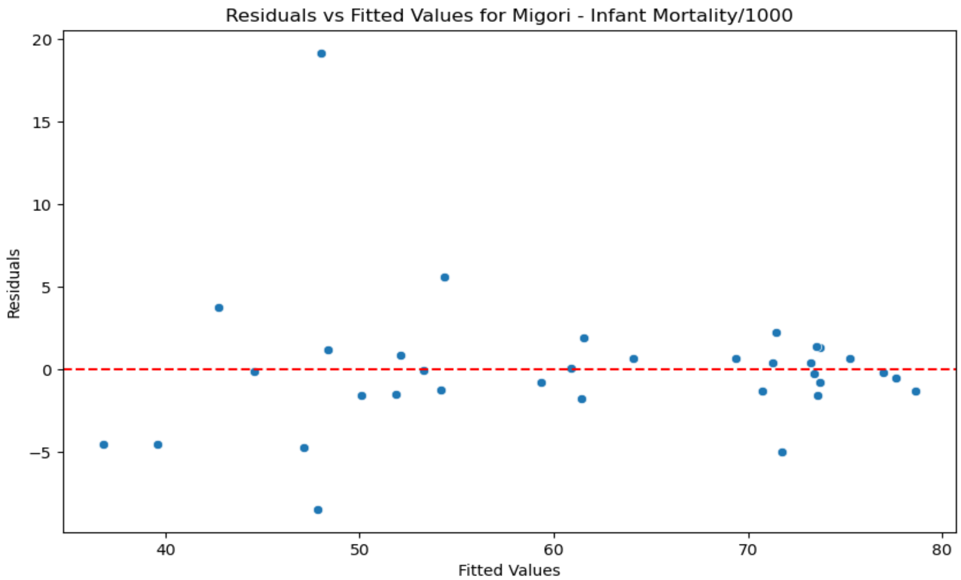

Figure 4.4: Residuals vs Fitted Values for Migori – Infant Mortality/1000

4.3.4 Multivariate results of Under-5 Mortality Rate using Migori Data

Dep. Variable:	*Under 5/1000*	R-squared:	*0.897*
Model:	*OLS*	Adj. R-squared:	*0.799*
Method:	*Least Squares*	F-statistic	*43.06*
Sample Size:	*34*	Prob (F-statistic):	*1.30-10*
Df Residuals:	*17*	Log-Likelihood:	*-17.71*
Df Model:	*16*	AIC:	*294.4*
Covariance Type:	*non-robust*	BIC:	*275.4*

The output provides key information about the overall fit and statistical significance of the regression model used to predict Under-5 child mortality rates per 1,000 live births. The model had an R-squared value of 0.897, indicating that approximately 89.7% of the variability in infant mortality rates was explained by the independent variables included in the model. This high R-squared value suggested that the model fitted the data well, capturing most of the factors influencing under-five child mortality. An Adjusted R-squared of 0.799 represented a robust model, suggesting that most of the independent variables included were relevant in explaining the variation in infant mortality. The F-statistic of 43.06, with a very low p-value of 1.30e-10, indicated that the overall model was statistically significant. As such, the collective effect of the independent variables on infant mortality was unlikely to be due to chance, and at least one of the predictors has a meaningful impact on the dependent variable. The model demonstrated strong explanatory power and statistical significance, making it a reliable tool for understanding the factors influencing infant mortality in the explored data.

Table 4.7: Factors affecting under 5 years Mortality Rate in Migori County

	coef	std err	t	P>|t|	[0.025	0.975]
const	16.702	1.394	83.690	0.000	113.759	119.643
Women No Formal Education	5.072	11.143	-0.455	0.655	-28.581	-18.438
Basic Sanitation	23.9507	13.198	1.815	0.005	-3.896	51.797
Clean Water Access	-46.54	15.041	-3.094	0.007	-78.274	-14.806
Teenage Pregnancy	10.889	5.351	2.035	0.058	-0.400	22.177
Births by Skilled Provider	15.937	14.470	1.101	0.286	-14.592	46.465
ANC Visits (4+)	-1.481	16.204	-0.091	0.928	-35.668	32.706
Post-natal Visits	-18.921	19.279	-0.981	0.003	-59.297	-21.754
Malnutrition	16.436	8.919	0.184	0.006	17.173	20.460
HIV infection Rate	-0.224	4.550	-0.049	0.961	-9.824	9.376

Respiratory Complications	28.217	2.236	-0.126	0.001	-49.986	-44.342
Vaccination Coverage	-12.843	1.157	-1.110	0.002	-37.246	-11.561
Non-Exclusive Breastfeeding	24.745	7.923	-0.312	0.000	-19.190	-14.241
Malaria Prevalence	47.854	1.151	-0.416	0.001	-29.067	-19.496
Overall Poverty	13.681	4.185	0.327	0.748	7.462	10.198
Solid Fuel Use	-7.088	12.398	-0.572	0.575	33.245	19.068
Location	3.229	2.306	1.400	0.179	-1.636	8.094

Among the significant predictors, Clean Water Access emerged as a strong negative predictor of under-5 child mortality, with a coefficient of -46.54 and a p-value of 0.007. This indicated that increased access to clean water was associated with a substantial reduction in child mortality. The negative coefficient suggested that for every unit increase in clean water access, under-5 mortality decreased significantly, making it a crucial factor in improving U5 child survival rates. Post-natal Visits also show a significant negative association with under-5 mortality (coef: -18.921, p-value: 0.003). This finding suggests that increasing the number of post-natal visits significantly reduced child mortality, highlighting the importance of postnatal care in child health. Another significant variable was Malnutrition, with a coefficient of 16.436 and a p-value of 0.006. This positive association implied that higher malnutrition rates were associated with increased child mortality, emphasizing the critical role of nutrition in child survival.

Respiratory Complications (coef: 28.217, p-value: 0.001) and Vaccination Coverage (coef: -12.843, p-value: 0.002) were also significant. Respiratory complications increased the risk of child mortality, while better vaccination coverage was associated with lower mortality rates. These findings demonstrated the need for targeted interventions to reduce respiratory complications and improve vaccination rates. Non-Exclusive Breastfeeding and Malaria Prevalence were significant predictors as well, with coefficients of 24.745 (p-value: 0.000) and 47.854 (p-value: 0.001), respectively. Non-exclusive breastfeeding practices contributed to higher child mortality, whereas increased malaria prevalence was associated with higher mortality rates. Both variables highlighted the importance of breastfeeding practices and malaria control in reducing child mortality.

Several variables were not statistically significant, indicating that they may not be strong predictors of under-5 mortality in this context. These include Women with No Formal Education (coef: 5.072, p-value: 0.655), Basic Sanitation (coef: 23.9507, p-value: 0.005), Teenage Pregnancy (coef: 10.889, p-value: 0.058), Births by Skilled Provider (coef: 15.937, p-value: 0.286), ANC Visits (4+) (coef: -1.481, p-value: 0.928), HIV Infection Rate (coef: -0.224, p-value: 0.961), Overall Poverty (coef: 13.681, p-value: 0.748), Solid Fuel Use (coef: -7.088, p-value: 0.575), and Location (coef: 3.229, p-value: 0.179). Although these variables may influence child mortality in other contexts, their lack of statistical significance suggests they do not have a strong or consistent impact in this specific model.

Table 4.7: Migori U5 Model Diagnostics

Omnibus:	5.125	Durbin-Watson:	2.577
Prob(Omnibus):	0.077	Jarque-Bera (JB):	3.591
Skew:	0.643	Prob(JB):	0.166
Kurtosis:	3.939	Cond. No.	19.2

==

The model diagnostics further reinforce the robustness of the analysis. The Durbin-Watson statistic of 2.577 indicates that there is minimal autocorrelation in the residuals, suggesting that the model is well-specified. The Omnibus test (p-value: 0.077) and Jarque-Bera test (p-value: 0.166) suggest that the residuals are approximately normally distributed, which is a desirable property for regression analysis. The condition number of 19.2 indicates that there is no significant multicollinearity among the predictors, ensuring that the estimates are reliable and stable.

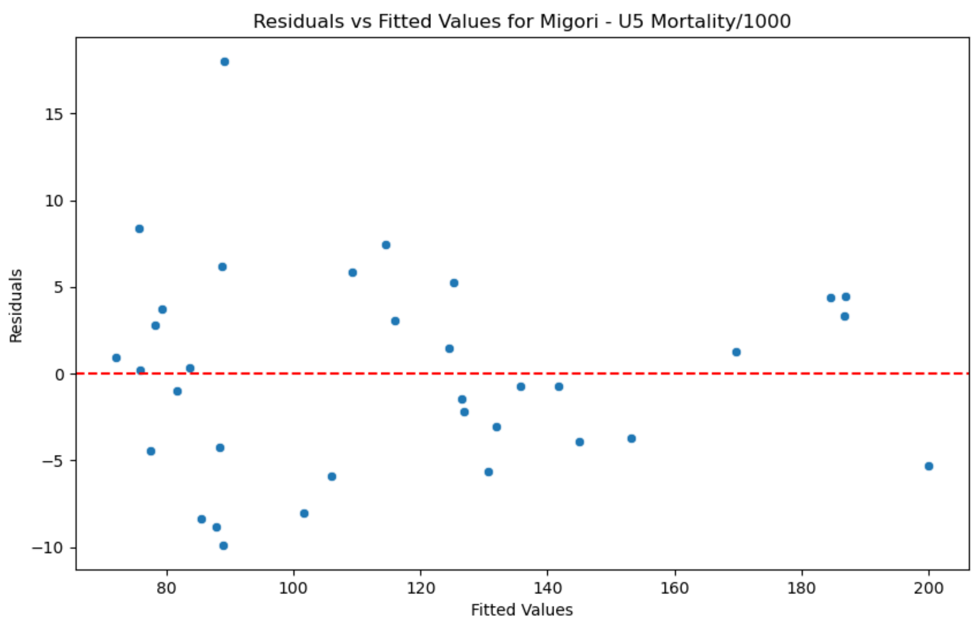

Figure 4.5: Residuals vs Fitted Values for Migori – U5 Mortality/1000

4.3.5 Multivariate Results of Neonatal Mortality Rate using Nyamira Data

The regression analysis provides valuable insights into the determinants of neonatal mortality in Nyamira County. The coefficients and p-values of the independent variables, along with model diagnostics, are critical in understanding which factors significantly impact neonatal mortality and how effectively these factors predict outcomes. Among the significant predictors, Basic Sanitation emerges as an important factor, with a coefficient of -6.981 and a p-value of 0.030. This negative association indicates that better access to basic sanitation is significantly associated with a reduction in neonatal mortality. The model suggests that improvements in basic sanitation can effectively lower neonatal mortality rates, making it a crucial area for public health interventions.

Teenage Pregnancy also shows a significant impact on neonatal mortality, with a coefficient of 6.463 and a p-value of 0.001. This positive coefficient suggests that higher rates of teenage pregnancy are associated with an increase in neonatal mortality. Teenage pregnancies often present higher risks, and the results emphasize the need for strategies aimed at reducing teenage pregnancies to improve neonatal outcomes. Another significant predictor is Births by Skilled Provider, which has a coefficient of -2.820 and a p-value of 0.005. This negative association implies that an increase in the proportion of births attended by skilled providers is significantly associated with a reduction in neonatal mortality. Ensuring that more births are attended by skilled health professionals is therefore a key strategy for reducing neonatal mortality.

ANC Visits (4+) is also significant, with a coefficient of -18.928 and a p-value of 0.007. This indicates that higher numbers of antenatal care visits (at least four) are strongly associated with lower neonatal mortality. Regular antenatal care is crucial for monitoring and addressing potential risks during pregnancy, contributing significantly to better neonatal health outcomes. Respiratory Complications is a particularly strong predictor, with a coefficient of 16.394 and a p-value of 0.001. The positive coefficient indicates that higher rates of respiratory complications are significantly associated with increased neonatal mortality, highlighting the critical need for interventions to prevent and manage respiratory conditions in newborns.

Non-Exclusive Breastfeeding is another significant factor, with a coefficient of 10.023 and p-value of 0.021. This positive association suggests that non-exclusive breastfeeding practices are associated with higher neonatal mortality, reinforcing the importance of exclusive breastfeeding in improving neonatal survival. Malaria Prevalence also shows significance, with a coefficient of 14.355 and a p-value of 0.003. This positive coefficient indicates that higher malaria prevalence is associated with increased neonatal mortality. This finding underscores the need for effective malaria prevention and treatment strategies to reduce neonatal deaths.

Several variables were not statistically significant, indicating that they may not be strong predictors of neonatal mortality in this context. These include Women with No Formal Education (coef: 0.0157, p-value: 0.979), Clean Water Access (coef: 0.495, p-value: 0.540), Post-natal Visits (coef: -0.9431, p-value: 0.488), Malnutrition (coef: 3.0518, p-value: 0.476), HIV Infection Rate (coef: 0.8466, p-value: 0.517), Vaccination Coverage (coef: -2.6889, p-value: 0.392), Overall Poverty (coef: 0.611, p-value: 0.507), Solid Fuel Use (coef: -3.474, p-value: 0.378), and Location (coef: 0.1739, p-value: 0.851). Although these variables did not show statistical significance in this model, they may still play a role in neonatal mortality under different circumstances or in other models.

Table 4.8: Nyamira County Neonatal Mortality, Model Summary

Dep. Variable:	*Neonatal Mortality*	R-squared:	*0.806*
Model:	*OLS*	Adj. R-squared:	*0.624*
Method:	*Least Squares*	F-statistic	*4.421*
Sample Size:	*34*	Prob (F-statistic):	*0.00201*
Df Residuals:	*17*	Log-Likelihood:	*-58.418*
Df Model:	*16*	AIC:	*150.8*
Covariance Type:	*non-robust*	BIC:	*176.8*

The R-squared value of 0.806 indicates that approximately 80.6% of the variability in neonatal mortality is explained by the independent variables included in the model. This suggests that the model has a good fit, as a large proportion of the variance in the dependent variable is captured by the predictors. The model is statistically significant, as indicated by the F-statistic. However, the difference between the R-squared and Adjusted R-squared suggests that some predictors may not be as effective in explaining the variation in neonatal mortality, and further refinement of the model may be necessary to improve its explanatory power.

Table 4.9: Factors affecting Neonatal Mortality Rate in Migori County

	coef	std err	t	P>|t|	[0.025	0.975]
const	11.388	2.910	0.391	0.000	5.001	7.279
Women No Formal Education	0.0157	0.593	0.026	0.979	-1.236	1.268
Basic Sanitation	-6.981	10.893	-0.641	0.030	-29.69	-16.00
Clean Water Access	0.495	7.915	0.626	0.540	-11.749	21.652
Teenage Pregnancy	6.463	0.252	0.256	0.001	-5.965	-4.673
Births by Skilled Provider	-2.820	4.217	-0.669	0.005	-11.716	-6.077
ANC Visits (4+)	-18.928	2.614	0.724	0.007	3.6221	7.4073
Post-natal Visits	-0.9431	1.332	-0.708	0.488	-3.753	1.8670
Malnutrition	3.0518	4.191	0.728	0.476	-5.790	11.894
HIV infection Rate	0.8466	1.280	0.661	0.517	-1.855	3.548
Respiratory Complications	16.394	0.236	7.086	0.001	-16.362	-4.342
Vaccination Coverage	-2.6889	3.061	-0.879	0.392	-9.146	3.768
Non-Exclusive Breastfeeding	10.023	0.009	2.541	0.021	0.004	0.041
Malaria Prevalence	14.355	2.280	0.630	0.003	3.375	6.246
Overall Poverty	0.611	0.902	0.677	0.507	-1.292	2.514
Solid Fuel Use	-3.474	3.833	-0.906	0.378	-11.561	4.614
Location	0.1739	0.909	0.191	0.851	-1.744	2.092

The model diagnostics further support the reliability of these findings. The Omnibus test (p-value: 0.252) and the Jarque-Bera test (p-value: 0.437) indicate that the residuals are approximately normally distributed, which is a desirable condition in regression analysis.

The Durbin-Watson statistic of 3.155 suggests minimal autocorrelation in the residuals, indicating that the model is well-specified. Additionally, the condition number of 19.2 implies that there is no severe multicollinearity among the predictors, ensuring the stability of the model estimates.

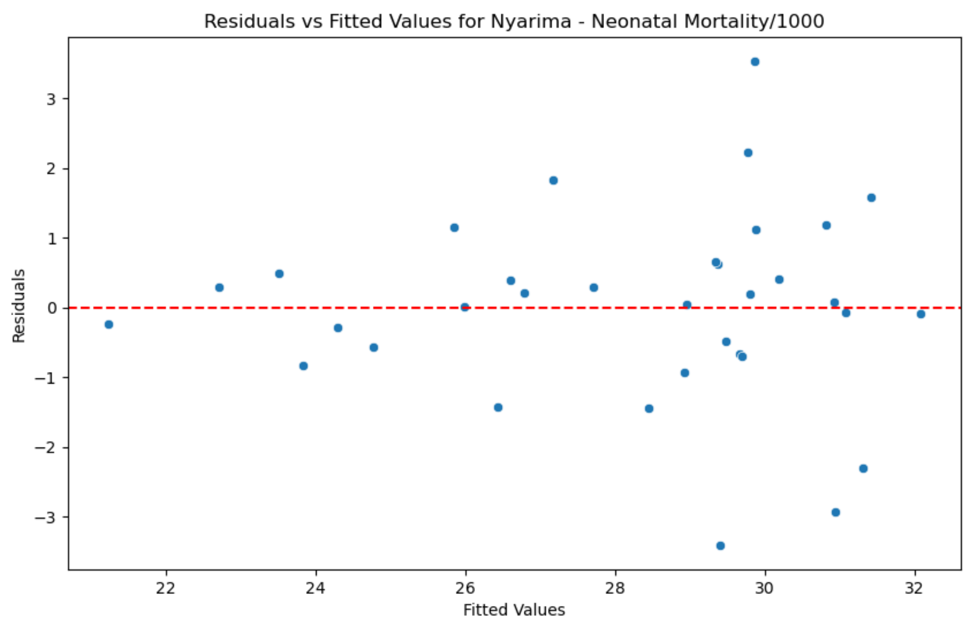

Figure 4.6: Nyamira County Residuals (Neonatal Mortality Rate)

4.3.6 Multivariate Results of Infant Mortality Rate using Nyamira Data

`Table 4.11 below show that women with no formal education had a significant positive coefficient of 23.944 and a p-value of 0.005. This indicates that higher levels of women without formal education are associated with increased infant mortality, suggesting that improving educational access for women could help reduce infant mortality rates. Basic Sanitation is another significant factor, with a negative coefficient of -34.323 and a p-value of 0.023. This negative association suggests that better access to basic sanitation significantly reduces infant mortality, highlighting the importance of sanitation facilities in promoting child health.

Clean Water Access is highly significant, with a negative coefficient of -23.054 and a p-value of 0.000, indicating that increased access to clean water is strongly associated with lower infant mortality rates. This underscores the critical role of clean water in improving child survival. ANC Visits (4+) also shows significance, with a coefficient of -6.094 and a p-value of 0.002. This suggests that more frequent antenatal care visits are associated with a reduction in infant mortality, emphasizing the importance of regular prenatal care.

Post-natal Visits is another significant predictor, with a negative coefficient of -9.4261 and a p-value of 0.003. This indicates that more postnatal visits are associated with lower infant mortality, further emphasizing the importance of follow-up care after childbirth. Malnutrition has a significant positive coefficient of 8.381 and a p-value of 0.011, indicating that higher malnutrition rates are associated with increased infant mortality. This finding highlights the need for nutrition interventions to improve child survival. Respiratory Complications is significant with a positive coefficient of 23.284 and a p-value of 0.024, indicating that higher rates of respiratory complications are linked to increased infant mortality, underscoring the need for better management of respiratory health in infants.

Vaccination Coverage shows a significant negative coefficient of -13.826 with a p-value of 0.002. This negative association suggests that higher vaccination coverage is associated with a decrease in infant mortality, reinforcing the importance of immunization programs. Non-Exclusive Breastfeeding is also significant, with a negative coefficient of -6.017 and a p-value of 0.004. This indicates that non-exclusive breastfeeding practices are linked to higher infant mortality, supporting the promotion of exclusive breastfeeding for better infant health outcomes. Malaria Prevalence is a significant predictor with a positive coefficient of 11.611 and a p-value of 0.001. This suggests that higher malaria prevalence is associated with increased infant mortality, emphasizing the importance of malaria control measures.

Several variables were not statistically significant in this model, indicating they may not strongly predict infant mortality in this context. These include Teenage Pregnancy (coef: 3.982, p-value: 0.227), Births by Skilled Provider (coef: 4.917, p-value: 0.368), HIV Infection Rate (coef: 2.3097, p-value: 0.170), Overall Poverty (coef: 0.1205, p-value: 0.917), Solid Fuel Use (coef: 5.6881, p-value: 0.255), and Location (coef: -0.9108, p-value: 0.437). These factors do not show a statistically significant relationship with infant mortality in this particular model.

Table 4.10: Nyamira County Infant Mortality Model Summary

Dep. Variable:	*Infant Mortality*	R-squared:	*0.894*
Model:	*OLS*	Adj. R-squared:	*0.796*
Method:	*Least Squares*	F-statistic	*65.920*
Sample Size:	*34*	Prob (F-statistic):	*3.94e-12*
Df Residuals:	*17*	Log-Likelihood:	*-66.268*
Df Model:	*16*	AIC:	*166.5*
Covariance Type:	*non-robust*	BIC:	*192.5*

The OLS regression model examining infant mortality per 1,000 in Nyamira County reveals a strong model fit, with an R-squared value of 0.894. This indicates that approximately 89.4% of the variability in infant mortality is explained by the independent variables in the model. The Adjusted R-squared of 0.796, while slightly lower, still suggests a robust model fit, accounting for the number of predictors relative to the sample size. The model's F-statistic of 65.920 and its highly significant p-value (3.94e-12) demonstrate that the overall model is statistically significant, meaning the independent variables collectively have a strong predictive power for infant mortality. The Akaike Information Criterion (AIC) of 166.5 and the Bayesian Information Criterion (BIC) of 192.5 suggest that the model is effective.

Table 4.11: Factors affecting Infant Mortality Rate in Nyamira County

	coeff	std err	t	P>|t|	[0.025	0.975]
constant	16.579	0.667	3.180	0.005	39.232	93.927
Women No Formal Education	23.944	0.747	-3.203	0.005	-3.971	-0.817
Basic Sanitation	-34.323	1.372	-2.501	0.023	-63.275	-5.371
Clean Water Access	-23.054	0.712	2.312	0.000	20.169	44.092
Teenage Pregnancy	3.982	3.176	1.254	0.227	-2.718	10.682
Births by Skilled Provider	4.917	1.312	0.926	0.368	-6.291	16.124
ANC Visits (4+)	-6.094	3.293	-1.851	0.002	-13.041	-0.852
Post-natal Visits	-9.4261	1.678	-0.976	0.003	-5.177	-1.902
Malnutrition	8.381	5.279	1.588	0.011	2.757	19.519
HIV infection Rate	2.3097	1.613	1.432	0.170	-1.093	5.712
Respiratory Complications	23.284	0.388	2.480	0.024	3.479	43.091
Vaccination Coverage	13.826	3.856	-0.099	0.002	-8.517	-7.752

Non-Exclusive Breastfeeding	6.017	0.011	-1.545	0.004	-0.041	-0.006
Malaria Prevalence	11.611	2.872	-4.043	0.001	-17.670	-5.551
Overall Poverty	0.1205	1.136	0.106	0.917	-2.277	2.518
Solid Fuel Use	5.6881	4.829	1.178	0.255	-4.500	15.876
Location	-0.9108	1.145	-0.795	0.437	-3.327	1.506

The model diagnostics indicate that the regression model is well-specified. The Omnibus test (p-value: 0.617) and Jarque-Bera test (p-value: 0.651) suggest that the residuals are normally distributed. The Durbin-Watson statistic of 1.964 indicates that there is no significant autocorrelation in the residuals. The condition number of 19.6 implies that there is no severe multicollinearity among the predictors, ensuring stable model estimates.

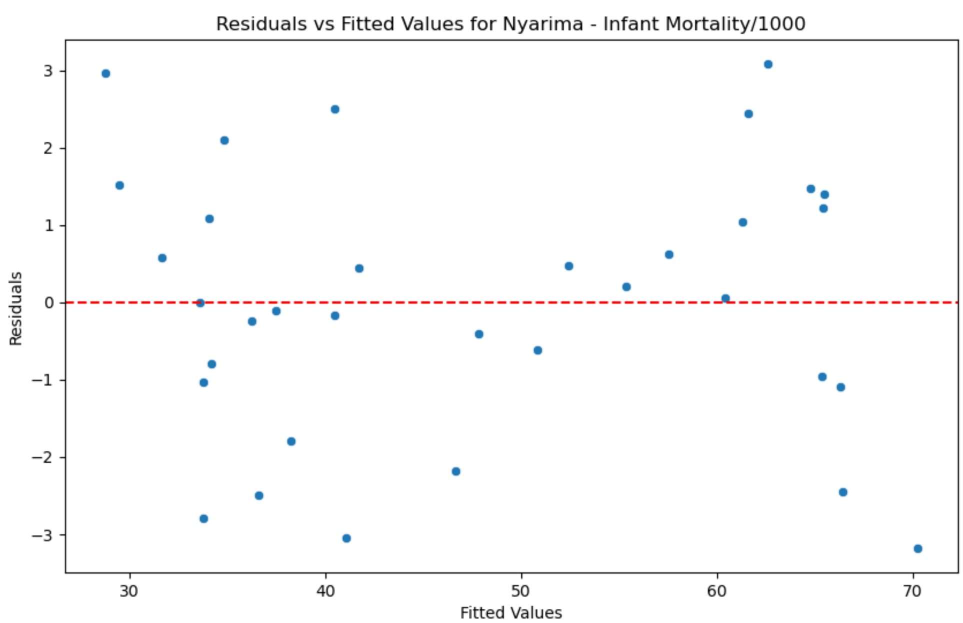

Figure 4.7: Nyamira County Residuals (Infant Mortality Rate)

4.3.7 Multivariate results of under 5 Mortality Rate using Nyamira Data

The percentage of women with no formal education exhibits a positive and statistically significant relationship with under-5 mortality (coef = 13.603, p = 0.008) as shown in Table 4.10.1 below. This suggest that higher levels of maternal education correlate with reduced child mortality. Basic sanitation is another crucial determinant, with improved sanitation significantly reducing under-5 mortality (coef = -8.292, p = 0.007). Similarly, access to clean water inversely impacts child mortality (coef = 65.171, p = 0.012), underscoring the importance of clean water in child health. Teenage pregnancy also shows a significant positive association with under-5 mortality (coef = 11.943, p = 0.030), highlighting the adverse effects of early motherhood on child survival. Additionally, births attended by skilled providers (coef = -18.361, p = 0.003), adequate antenatal care (ANC visits) (coef = -26.094, p = 0.002), and vaccination coverage (coef = -14.611, p = 0.029) all significantly reduce under-5 mortality, emphasizing the role of healthcare interventions in lowering mortality rates. Malnutrition (coef = 24.565, p = 0.001), respiratory complications (coef = 58.469, p = 0.001), and malaria prevalence (coef = 59.99, p = 0.033) are all significantly associated with higher under-5 mortality, indicating that addressing these health issues is vital for improving child survival.

Conversely, several variables did not show statistically significant relationships with under-5 mortality. These include post-natal visits (coef = -0.8025, p = 0.767), HIV infection rate (coef = 4.4934, p = 0.097), overall poverty (coef = 3.641, p = 0.060), and the use of solid fuels (coef = 20.365, p = 0.017), despite the latter nearing significance. Additionally, location (coef = 1.0177, p = 0.583) and non-exclusive breastfeeding (coef = 0.0065, p = 0.719) were not significant predictors. The wide confidence intervals for these variables suggest that their impact on under-5 mortality may be more variable or context-dependent, requiring further investigation to clarify their roles.

The OLS regression model assessing under-5 mortality in Nyamira County demonstrates a strong fit, with an R-squared value of 0.879. This indicates that approximately 87.9% of the variability in under-5 mortality is explained by the independent variables included in the model. The Adjusted R-squared value of 0.766, although slightly lower, still suggests that the model has substantial explanatory power, even after accounting for the number of predictors and sample size. The model's F-statistic of 158.4, combined with a highly significant p-value (2.63e-15), underscores the overall significance of the regression model. This suggests that the independent variables, taken together, provide a strong and statistically significant prediction of under-5 mortality rates in the county.

Table 4.12: Factors Affecting Under-5 Mortality in Nyamira County

	coeff	std err	t	P>|t|	[0.025	0.975]
constant	39.228	0.582	0.674	0.001	83.623	162.08
Women No Formal Education	13.603	1.187	30.33	0.008	1.096	6.105
Basic Sanitation	-8.292	2.1795	-0.380	0.007	-54.276	-37.69
Lack of Clean Water Access	65.171	1.583	1.043	0.012	16.897	49.931
Teenage Pregnancy	11.943	5.044	-2.368	0.030	-22.585	-1.300
Births by Skilled Provider	-18.361	8.437	-0.991	0.003	9.440	26.161
ANC Visits (4+)	-26.094	5.230	1.318	0.002	4.142	17.925
Post-natal Visits	-0.8025	2.665	-0.301	0.767	-6.424	4.819
Malnutrition	24.565	8.385	-0.174	0.001	-29.15	-16.23

HIV infection Rate	4.4934	2.562	1.754	0.097	-0.911	9.898
Respiratory Complications	58.469	1.491	3.921	0.001	27.010	89.297
Vaccination Coverage	-14.611	6.124	-2.386	0.029	-27.531	-1.691
Non-Exclusive Breastfeeding	0.0065	0.018	0.366	0.719	-0.031	0.044
Malaria Prevalence	59.99	4.561	-2.324	0.033	-20.224	-0.976
Overall Poverty	3.641	1.805	2.017	0.060	-0.167	7.449
Solid Fuel Use	20.365	7.670	-2.655	0.017	-36.547	-4.183
Location	1.0177	1.819	0.559	0.583	-2.820	4.856

Dep. Variable:	*U5 Mortality*	R-squared:	*0.879*
Model:	*OLS*	Adj. R-squared:	*0.766*
Method:	*Least Squares*	F-statistic	*158.4*
Sample Size:	*34*	Prob (F-statistic):	*2.63e-15*
Df Residuals:	*17*	Log-Likelihood:	*-81.998*
Df Model:	*16*	AIC:	*198*
Covariance Type:	*non-robust*	BIC:	*223.9*

The overall model diagnostics indicate a robust model with an R-squared value of 0.879, meaning that approximately 87.9% of the variability in under-5 mortality can be explained by the model. The adjusted R-squared value of 0.766 further supports the model’s fit, accounting for the number of predictors. The F-statistic (158.4, $p < 0.0001$) confirms the overall significance of the model. Additionally, the Durbin-Watson statistic of 2.595 suggests that there is minimal autocorrelation in the residuals, indicating that the model is well-specified. The confidence intervals provided for significant predictors are relatively narrow, reflecting the precision of the estimates. This analysis underscores the importance of addressing key determinants such as maternal education, healthcare access, and environmental health in efforts to reduce under-5 mortality in Nyamira County.

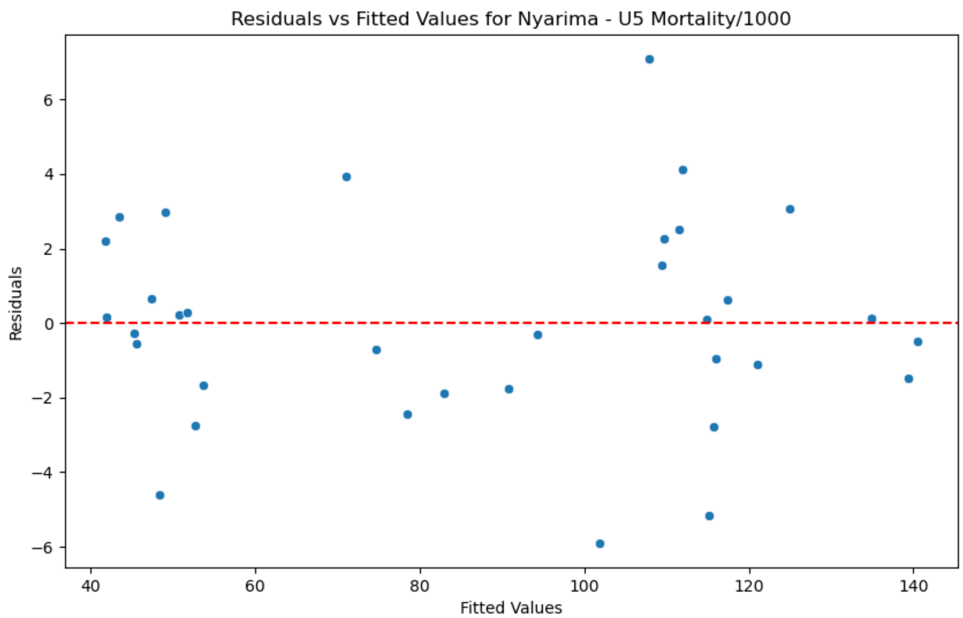

Figure 4.8: Under-five Mortality, Nyamira County, Kenya

4.4 Evaluating the performance of the predictive model for different infant risk groups in Nyamira and Migori counties, Kenya

This section contains model performance of the predictions of different child risk groups with the highest risk of death based on the input variables in Migori and Nyamira counties, Kenya.

4.4.1 Machine Learning prediction of child mortality rate using Gradient Boosting Machines (Gradient Boosting Regressor) model using Migori data

The results of Gradient Boosting Regression model are displayed in table 4.13 below utilizing Migori data. The optimal model for predicting neonatal mortality used a learning rate of 0.05, a maximum depth of 5, and 200 estimators. The Mean Squared Error (MSE) was 3.7538, indicating the average squared difference between the observed and predicted values, reflecting a better model performance. Caesarian Section (0.1460) and Respiratory Complications/1000 (0.0894) also contribute significantly. Variables such as Malnutrition % (0.0008) and Overall Poverty % (0.0011) have minimal impact, suggesting they may not be as crucial in predicting neonatal mortality in Migori.

For neonatal mortality in Migori, the model's performance ($R^2 = 0.733$) suggests a decent fit, explaining approximately 73% of the variance. The most significant predictors are "Solid Fuel Use" (feature importance = 0.2290), "Congenital Malformation" (0.2230), and "Caesarian Section" (0.1460). These variables are strong predictors, indicating that households relying on solid fuels and instances of congenital malformations significantly raise neonatal mortality rates. In contrast, factors like "Women Formal Education per Household (%)" and "Malnutrition %" are less predictive, with very low feature importances (0.0075 and 0.0008, respectively) as the table below summarizes.

Table 4.13: Risk factors affecting Neonatal Mortality Rate in Migori County

Feature	*Importance*
Women Formal Education per Household (%)	0.0075
Basic Sanitation (%)	0.0143
Access to Clean Drinking Water (%)	0.0485
Access to Clean Fuels %	0.0418
Teenage Pregnancy %	0.0038
Caesarian Section	**0.1460**
Births by Skilled Provider %	0.0607
ANC visits (4+) %	0.0159
Post-natal Visits %	0.0237
Malnutrition %	0.0008
HIV Infection Rate	0.0180
Respiratory Complications/1000	0.0894
Congenital Malformation	**0.2230**
Vaccination Coverage (%)	0.0206
Non-Exclusive Breastfeeding %	0.0443
Malaria Prevalence %	0.0115
Overall Poverty %	0.0011
Solid Fuel Use	**0.2290**
Metric	Value
Best Parameters	{'learning_rate': 0.05, 'max_depth': 5, 'n_estimators': 200}
Mean Squared Error	3.7538
R^2 score	0.7328

4.4.2 Machine Learning prediction of Infant Mortality Rate using Gradient Boosting Machines (Gradient Boosting Regressor) model using Migori data

The results of Gradient Boosting Regression model were displayed in table 4.14 below for infant mortality using Migori data. The model that best predicted infant mortality used a learning rate of 0.01, a maximum depth of 3, and 200 estimators. The Mean Squared Error (MSE) was 10.8431, indicating that predictions have higher average squared errors compared to neonatal mortality. The most significant factors influencing infant mortality in Migori County were Malnutrition (0.1114), births by a skilled provider (0.1907), respiratory complications (0.9333), and non-exclusive breast feeding. Access to Clean Drinking Water % (0.0914) and HIV Infection Rate (0.0599) are also important but less so. Overall Poverty % (0.0168) and Non-Exclusive Breastfeeding % (0.0711) have lower impact on infant mortality predictions.

In the case of infant mortality, the model achieved a higher R^2 of 0.886, indicating a strong predictive capability. From the prediction, "Births by Skilled Provider %" (0.1907) and "Congenital Malformation" (0.1295) emerge as critical factors, reflecting the importance of medical interventions during birth and the impact of congenital conditions on infant survival. The relatively high importance of "Solid Fuel Use" (0.1077) also underscores the environmental factors that contribute to mortality as contained in the table below.

Table 4.14 Important Mortality Features Affecting Infant Mortality in Migori County

Feature	*Importance*
Women Formal Education per Household (%)	0.0512
Basic Sanitation (%)	0.0325
Access to Clean Drinking Water (%)	**0.0914**
Access to Clean Fuels %	0.0064
Teenage Pregnancy %	0.0075
Caesarian Section	0.0659
Births by Skilled Provider %	**0.1907**
ANC visits (4+) %	0.0321
Post-natal Visits %	0.0540
Malnutrition %	**0.1114**
HIV Infection Rate	0.0599
Respiratory Complications/1000	**0.9333**
Congenital Malformation	**0.1295**
Vaccination Coverage (%)	0.0102
Non-Exclusive Breastfeeding %	**0.0711**
Malaria Prevalence %	0.0282
Overall Poverty %	0.0168
Solid Fuel Use	**0.1077**
Metric	Value
Best Parameters	{'learning_rate': 0.01, 'max_depth': 3, 'n_estimators': 200}
Mean Squared Error	10.8431
R^2 score	0.8860

4.4.3 Machine Learning Prediction of Under 5 Years Mortality Rate using Gradient Boosting Machines (Gradient Boosting Regressor) model using Migori data

For under-five mortality, the model's performance is excellent, with an R^2 of 0.962, indicating that it explains nearly all of the variance in the data. Key predictors include "Congenital Malformation" (0.1644), "Teenage Pregnancy %" (0.1145), and "Caesarian Section" (0.1074). These results suggest that both biological factors and maternal health conditions are paramount in determining under-five mortality. The lower importance of variables like "Basic Sanitation" (0.0294) and "Access to Clean Fuels" (0.0095) suggests that while important, these factors may not be as directly impactful in the context of under-five mortality in Migori. Women Formal Education per Household (%) (0.0736) and Access to Clean Drinking Water (%) (0.0466) are also important but with moderate influence. Variables with low importance features included overall Poverty % (0.0063) and Solid Fuel Use (0.0133) have relatively lower impacts.

Table 4.15: Important under-five Mortality Features in Migori County

Feature	Importance
Women Formal Education	0.0736
Basic Sanitation (%)	0.0294
Clean Water Access	0.0466
Access to Clean Fuels %	0.0095
Teenage Pregnancy %	0.1145
Caesarean Section	0.1074
Births by Skilled Provider %	0.0754

ANC visits (4+) %	0.0329
Post-natal Visits %	0.0332
Malnutrition %	0.0859
HIV Infection Rate	0.0303
Respiratory Complications/1000	0.0247
Congenital Malformation	0.1644
Vaccination Coverage (%)	0.0508
Non-Exclusive Breastfeeding	0.0328
Malaria Prevalence %	0.0691
Overall Poverty %	0.0063
Solid Fuel Use	0.0133
Metric	Value
Best Parameters	{'learning_rate': 0.05, 'max_depth': 5,
Mean Squared Error	18.8079
R^2 score	0.9623

4.4.4 Machine Learning prediction of Neonatal Mortality Rate using Gradient Boosting Machines (Gradient Boosting Regressor) model using Nyamira data

The model performed best with a learning rate of 0.01, a maximum depth of 3, and 100 estimators. The MSE was 3.6736, similar to Migori's results, indicating a comparable average error in predictions. Some of the moderate importance features included vaccination Coverage (%) (0.0566) and Overall Poverty % (0.1004) are relatively important but not dominant. Most other variables, including Women Formal Education per Household (%) (0.0093) and Basic Sanitation (%) (0.0130), have minimal impact (Table 4.12).

The model's performance for neonatal mortality in Nyamira is notably lower compared to Migori (R^2 = 0.81), indicating good predictive capability and suggesting that the chosen independent variables effectively capture the complexity of neonatal mortality in Nyamira County, Kenya. However, "Teenage Pregnancy %" (0.7077) stands out as an overwhelmingly significant predictor, suggesting that young maternal age is a critical risk factor for neonatal mortality in Nyamira. The use of solid fuels also emerged as an important factor, contributing to respiratory complications, which are a leading cause of neonatal mortality. Post-natal visits and malaria prevalence were flagged as important factors contributing to the model as good predictors of neonatal mortality in Nyamira county.

Table 4.16: Important for Neonatal Mortality Rate in Nyamira County

Feature	**Importance**
Women Formal Education per	0.0093
Basic Sanitation (%)	0.0130
Access to Clean Water	0.0138
Access to Clean Fuels %	0.0020
Teenage Pregnancy %	0.7077
Caesarean Section	0.0124
Births by Skilled Provider %	0.0045
ANC visits (4+) %	0.0018

Post-natal Visits %	0.6089
Malnutrition %	0.0096
HIV Infection Rate	0.0125
Respiratory Complications	0.8191
Congenital Malformation	0.0057
Vaccination Coverage (%)	0.0566
Non-Exclusive Breastfeeding	0.0097
Malaria Prevalence %	0.5073
Overall Poverty %	0.1004
Solid Fuel Use	0.0856
Metric	Value
Best Parameters	{'learning_rate': 0.01, 'max_depth': 3,
Mean Squared Error	13.674
R^2 Score	0.8090

4.4.5 ML Prediction of Infant Mortality Rate using Gradient Boosting Machines (Gradient Boosting Regressor) model using Nyamira data

The best parameters for optimal performance were achieved with a learning rate of 0.1, a maximum depth of 4, and 300 estimators. The MSE was 4.2407, reflecting the accuracy of predictions. Moderate importance features included ANC visits (4+) % (0.1403) and Births by Skilled Provider % (0.0717) contribute significantly, while women Formal Education per Household (%) (0.0085) and Overall Poverty % (0.0023) have minimal impacts were low importance features.

For infant mortality, the model performs exceptionally well, with an R^2 of 0.959. "Malnutrition %" (0.2439), "Caesarean Section" (0.1619), and "Congenital Malformation" (0.1472) are the most important predictors, emphasizing the role of nutritional status and medical complications at birth in determining infant survival. The importance of "Access to Clean Drinking Water" (0.0466) highlights the role of environmental factors in infant mortality.

Table 4.17: Important Features for Infant Mortality Rate in Nyamira County

Feature	**Importance**
Women Formal Education	0.0085
Basic Sanitation (%)	0.0043
Access to Clean Water	0.0466
Access to Clean Fuels %	0.0484
Teenage Pregnancy %	0.0027
Caesarian Section	0.1619
Births by Skilled Provider %	0.0717
ANC visits (4+) %	0.1403
Post-natal Visits %	0.0013
Malnutrition %	0.2439
HIV Infection Rate	0.0001
Respiratory Complications	0.0169
Congenital Malformation	0.1472
Vaccination Coverage (%)	0.0019

Non-Exclusive Breastfeeding	0.0060
Malaria Prevalence %	0.0077
Overall Poverty %	0.0023
Solid Fuel Use	0.0883
Metric	Value
Best Parameters	{'learning_rate': 0.1, 'max_depth': 4, 'n_estimators': 300}
Mean Squared Error	4.2407
R^2 score	0.9590

4.4.6 ML Prediction of Under 5 Years Mortality Rate using Gradient Boosting Machines (Gradient Boosting Regressor) model using Nyamira data

Lastly, for under-five mortality in Nyamira, the model again performs well ($R^2 = 0.934$), with "Malnutrition %" (0.3123) being the most significant predictor, followed by "Women Formal Education per Household (%)" (0.1658) and "Congenital Malformation" (0.1142). These results underscore the critical importance of nutrition and maternal education in improving child survival rates.

The best model used a learning rate of 0.1, a maximum depth of 3, and 100 estimators. The MSE of 3.41398, was indicative of minimal predictive errors, affirming the model’s precision. Some of the moderate importance features were Caesarean Section (0.1353) and Births by Skilled Provider % (0.0974) are also notable. Features with the least influence in U5 mortality included Access to Clean Drinking Water (%) (0.0144) and Solid Fuel Use (0.0330).

The Migori Gradient Boosting models generally perform well, especially for U5 Mortality/1000, with high R^2 scores and moderately low MSEs. Solid Fuel Use, Congenital Malformation, and Malnutrition % are consistently important predictors. On the other hand, the Nyamira models show variable performance, with a notably low R^2 score for neonatal mortality. Teenage Pregnancy % stands out as a significant feature, especially for neonatal mortality, while Malnutrition % is crucial across other dependent variables. Overall, while Gradient Boosting Regressor models provide useful insights, further tuning and consideration of additional features or external factors might be needed for more accurate predictions, especially in Nyamira.

Table 4.18 Important Features for Under 5 Year Mortality Rate in Nyamira County

Feature	Importance
Women Formal Education per Household (%)	0.1658
Basic Sanitation (%)	0.0056
Access to Clean Drinking Water (%)	0.0144
Access to Clean Fuels %	0.0171
Teenage Pregnancy %	0.0097
Caesarean Section	0.1353
Births by Skilled Provider %	0.0974
ANC visits (4+) %	0.0155
Post-natal Visits %	0.0012
Malnutrition %	0.3123
HIV Infection Rate	0.0007
Respiratory Complications/1000	0.0422
Congenital Malformation	0.1142

Vaccination Coverage (%)	0.0022
Non-Exclusive Breastfeeding %	0.0120
Malaria Prevalence %	0.0146
Overall Poverty %	0.0066
Solid Fuel Use	0.0330

4.5 Temporal Patterns and Seasonal Trends in Infant Mortality Rates

4.5.1 Spatio Temporal Autoregressive Modelling (STAR)

The season temporal results for each region are presented below in text, tables, and graphical visualizations.

4.5.2 Spatio temporal Modelling for Neonatal Mortality Rate

The Maximum Likelihood Spatial Lag results for the neonatal mortality rate were as displayed in Table 4.19. The mean value of neonatal mortality in the studied counties is 29.79 per 1,000 live births (standard deviation: 5.17). This indicates a high baseline level of neonatal mortality across the counties, with moderate variability between different areas (sub-counties and other rural areas). This variability suggests that social, healthcare, and environmental factors likely influence neonatal mortality rates differently across the sub-counties. The model's log-likelihood value of -111.87 indicates a relatively good fit, capturing the observed data well (pseudo-R-squared = 0.9415). The variance of the error term, represented by Sigma-square ML (1.54), is low, indicating that the model leaves little unexplained variability. The Akaike Information Criterion (AIC = 263.74) and Schwarz criterion (BIC = 308.13) suggest that the model balances complexity and fit well, although the difference between AIC and BIC indicates that adding more variables may lead to overfitting.

Table 4.19: Model coefficients for Spatial Lag results for Neonatal Mortality Rate

Variable	Coefficient	Std.Error	z-Statistic	Probability
CONSTANT	43.39141	4.13983	10.48144	0.00000
Women Formal Education	1.72389	0.87878	1.96168	0.04980
Basic Sanitation	-1.24000	0.93055	-1.33255	0.18268
Clean Drinking Water	-2.41945	0.74759	-3.23632	0.00121
Access to Clean Fuels	-1.15581	1.18052	-0.97907	0.32755
Teenage Pregnancy	1.46555	0.51388	2.85191	0.00435
Caesarian Section	0.72772	0.76254	0.95435	0.33991
Births by Skilled Provider	0.47500	0.84648	0.56114	0.57470
ANC visits (4+)	1.16596	0.68406	1.70447	0.08829
Post-natal Visits	-0.84755	0.70464	-1.20282	0.22904
Malnutrition	0.02391	0.86282	0.02771	0.97789
HIV Infection Rate	-0.66466	0.45160	-1.47177	0.14108
Respiratory Complications/1000	3.59160	0.54281	6.61666	0.00000
Congenital Malformation	2.69338	1.31407	-2.04965	0.00000
Vaccination Coverage	-2.19992	0.91166	-2.41309	0.01582
Non-Exclusive Breastfeeding	1.96103	0.72253	2.71412	0.00665
Malaria Prevalence	0.68330	0.71769	0.95208	0.34105
Overall Poverty	0.43237	0.31899	1.35542	0.17528

Solid Fuel Use	0.56870	1.07707	0.52800	0.59750
W_Neonatal Mortality/1000	-0.45813	0.13919	-3.29141	0.00100

Dependent Variable: Neonatal Mortality/1000		Number of Observations:	68
Mean dependent var:	29.7866	Number of Variables:	20
S.D. dependent var:	5.1717	Degrees of Freedom:	48
Pseudo R-squared:	0.9415		
Spatial Pseudo R-squared:	0.9337		
Log likelihood:	-111.8716	Akaike info criterion:	263.743
Sigma-square ML:	1.5420	Schwarz criterion:	308.133
S.E of regression:	1.2418		

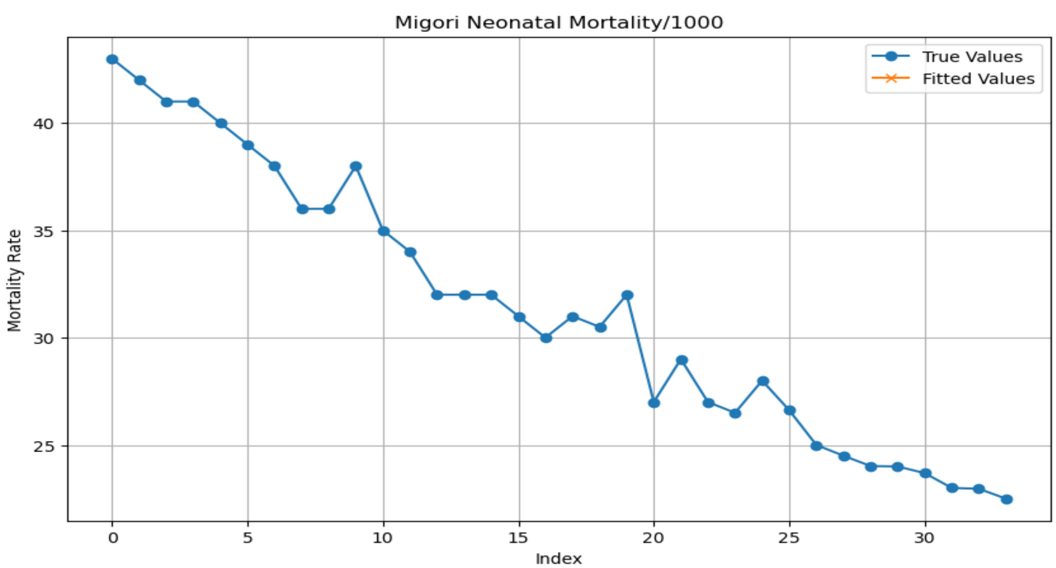

Figure 4.9: Spatial Lag (Migori Neonatal Mortality)

4.5.3 Spatio temporal Modeling for Neonatal Mortality Rate for Nyamira County

The Spatial Lag Model Impacts computed using the 'simple' method was as presented in the below table 4.20 and figure. Key predictors of neonatal mortality include access to clean drinking water, which is significantly associated with a decrease in neonatal mortality (Coefficient = -2.42, p = 0.001). Teenage pregnancy is another strong predictor, with higher rates of teenage pregnancy linked to higher neonatal mortality (Coefficient = 1.47, p = 0.004). Spatial effects are also significant, as seen in the spatial lag coefficient (W_Neonatal Mortality = -0.46, p = 0.001), indicating that neonatal mortality in neighbouring areas impacts local mortality rates. These findings highlight the importance of improving access to clean water and addressing teenage pregnancy to reduce neonatal mortality.

Table 4.20: Spatial Lag Model Impacts of NMR, Nyamira County

Variable	Direct	Indirect	Total
Women Formal Education	1.7239	-0.5416	1.1823
Basic Sanitation	-1.24000	0.3896	-0.8504
Clean Drinking Water	-2.41945	0.7602	-1.6593
Access to Clean Fuels	-1.15581	0.3631	-0.7927
Teenage Pregnancy	1.46555	-0.4605	1.0051
Caesarian Section	0.72772	-0.2286	0.4991
Births by Skilled Provider	0.47500	-0.1492	0.3258
ANC visits (4+)	1.16596	-0.3663	0.7996
Post-natal Visits	-0.84755	0.2663	-0.5813
Malnutrition	0.02391	-0.0075	0.0164
HIV Infection Rate	-0.66466	0.2088	-0.4558
Respiratory Complications/1000	3.59160	-1.1284	2.4632
Congenital Malformation	2.69338	0.8462	-1.8471
Vaccination Coverage	-2.19992	0.6912	-1.5087
Non-Exclusive Breastfeeding	1.96103	-0.6161	1.3449
Malaria Prevalence	0.68330	-0.2147	0.4686
Overall Poverty	0.43237	-0.1358	0.2965
Solid Fuel Use	0.56870	-0.1787	0.3900

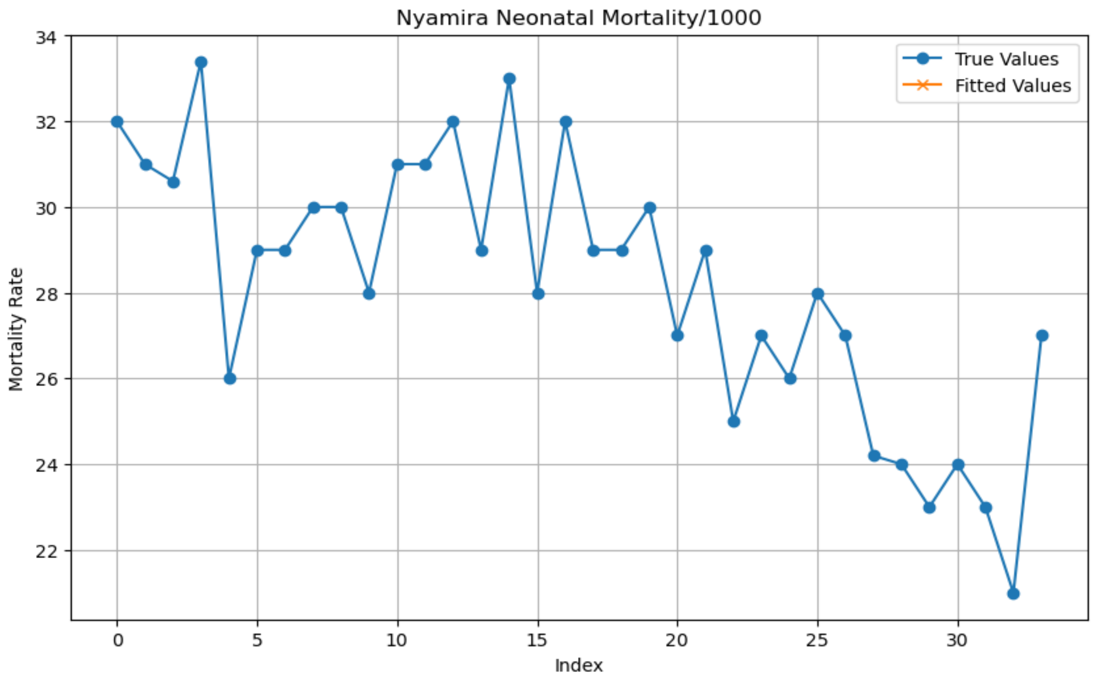

Figure 4.10: Nyamira County Neonatal Spatial Lag

4.5.4 Spatio temporal Modeling of Infant Mortality Rate for Migori County

Significant predictors of infant mortality include access to clean drinking water, which is strongly associated with lower infant mortality (Coefficient = -9.50, p = 0.0001). The lack of basic sanitation exhibits shows a positive association with infant mortality (Coefficient = 6.29, p = 0.036), which indicates that it contributes significantly to infant mortality rates in different regions of Migori County. Teenage pregnancy is another key predictor, as higher rates correlate with increased infant mortality (Coefficient = -3.29, p = 0.048). These findings emphasize the critical role of improving access to clean water and addressing the complexities around sanitation and teenage pregnancy to reduce infant mortality.

The mean infant mortality rate in the counties is 54.56 per 1,000 live births, with a larger variability than neonatal mortality (standard deviation: 15.01). This suggests that rural areas experience significantly higher infant mortality rates compared to urban regions, likely due to disparities in healthcare access, sanitation, and socio-economic conditions across the counties of Migori and Nyamira. The log-likelihood for the infant mortality model is -190.94, reflecting a solid fit to the data (pseudo-R-squared = 0.9276). However, the higher variance of the error term (Sigma-square ML = 16.08) compared to the neonatal model indicates more unexplained variability, implying that additional factors might influence infant mortality. The model selection criteria (AIC = 421.89; BIC = 466.28) suggest that while the model fits the data well, there is a trade-off between model complexity and parsimony.

Table 4.21: Spatial Lag Model Impacts of Infant Mortality, Migori County

Variable	Coefficient	Std.Error	z-Statistic	Probability
CONSTANT	50.79251	6.24499	8.13332	0.00000
Women Formal Education	2.06240	2.85599	0.72213	0.47021
Basic Sanitation	-1.24000	0.93055	-1.33255	0.18268
Clean Drinking Water	-2.41945	0.74759	-3.23632	0.00121
Access to Clean Fuels	-1.15581	1.18052	-0.97907	0.32755
Teenage Pregnancy	1.46555	0.51388	2.85191	0.00435
Caesarean Section	0.72772	0.76254	0.95435	0.33991
Births by Skilled Provider	0.47500	0.84648	0.56114	0.57470
ANC visits (4+)	1.16596	0.68406	1.70447	0.08829
Post-natal Visits	-0.84755	0.70464	-1.20282	0.22904
Malnutrition	0.02391	0.86282	0.02771	0.97789

HIV Infection Rate	-0.66466	0.45160	-1.47177	0.14108
Respiratory Complications/1000	3.59160	0.54281	6.61666	0.00000
Congenital Malformation	2.69338	1.31407	-2.04965	0.00000
Vaccination Coverage	-2.19992	0.91166	-2.41309	0.01582
Non-Exclusive Breastfeeding	1.96103	0.72253	2.71412	0.00665
Malaria Prevalence	0.68330	0.71769	0.95208	0.34105
Overall Poverty	0.43237	0.31899	1.35542	0.17528
Solid Fuel Use	0.56870	1.07707	0.52800	0.59750
W_Neonatal Mortality/1000	-0.45813	0.13919	-3.29141	0.00100

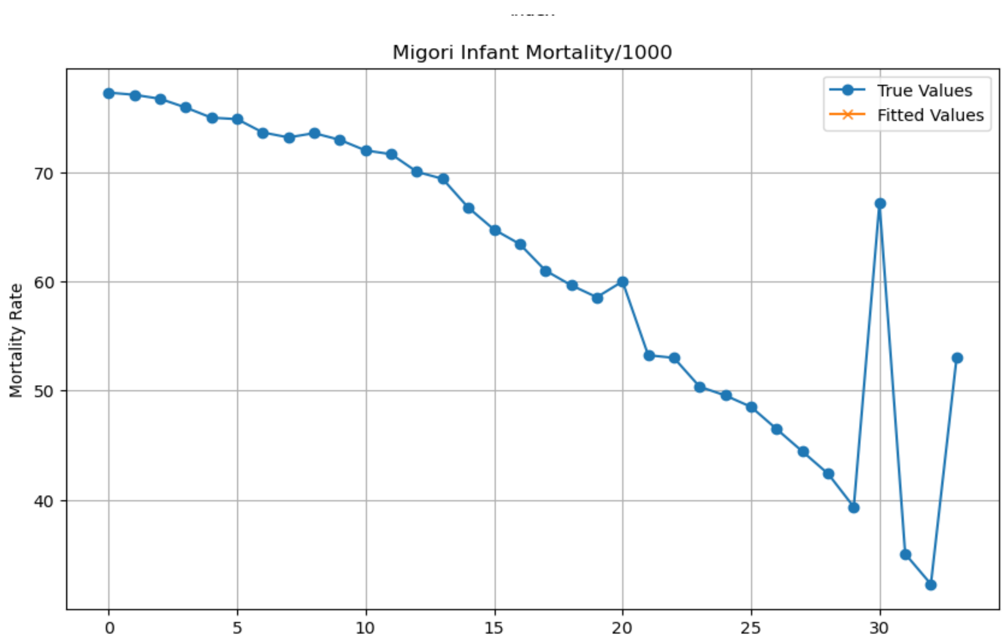

Figure 4.11: Spatial True vs Fitted Values, Migori Infant Mortality

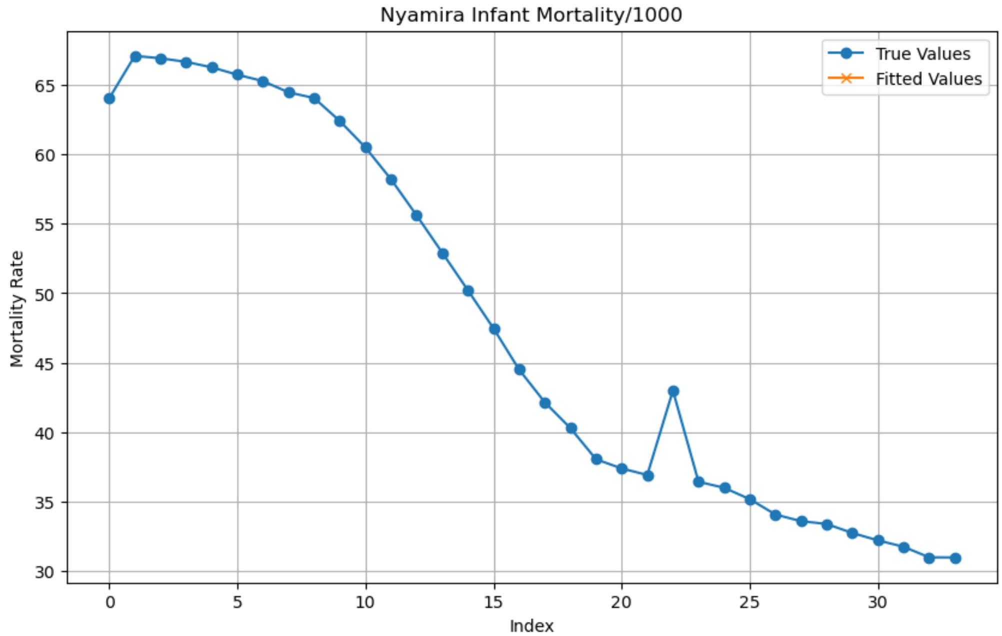

Figure 4.12: Spatial True vs Fitted Values, Nyamira Infant Mortality

Under-5 mortality rates are significantly higher, with a mean of 101.82 per 1,000 live births (standard deviation: 38.42). This large standard deviation suggests that under-5 mortality varies greatly across the counties, indicating disparities in access to healthcare, nutrition, and other determinants of child health. The log-likelihood value of -225.86 for the under-5 mortality model reflects the model's ability to explain a significant portion of the observed variability (pseudo-R-squared = 0.9691). However, the relatively high Sigma-square ML (44.92) suggests that there is substantial unexplained variability in under-5 mortality rates, possibly due to unmeasured factors. The AIC (491.72) and BIC (536.11) are the highest among the three models, indicating that while the model fits the data well.

4.4.4 Spatial Lag Models Impacts Using Both the Direct and Indirect Approaches

Under-5 mortality rates are significantly higher, with a mean of 101.82 per 1,000 live births (standard deviation: 38.42). This large standard deviation suggests that under-5 mortality varies greatly across the counties, indicating disparities in access to healthcare, nutrition, and other determinants of child health.

The log-likelihood value of -225.86 for the under-5 mortality model reflects the model's ability to explain a significant portion of the observed variability (pseudo-R-squared = 0.9691). However, the relatively high Sigma-square ML (44.92) suggests that there is substantial unexplained variability in under-5 mortality rates, possibly due to unmeasured factors. The AIC (491.72) and BIC (536.11) are the highest among the three models, indicating that the model fits the data well, and may not require refinement due to a lack of complexities or overfitting.

Key predictors of under-5 mortality include access to clean drinking water, which is strongly associated with lower under-5 mortality rates (Coefficient = -16.67, $p < 0.0001$). Teenage pregnancy is another significant predictor, with higher rates leading to higher under-5 mortality (Coefficient = 8.03, $p = 0.004$). Additionally, births attended by skilled providers have a protective effect, reducing under-5 mortality (Coefficient = 16.40, $p = 0.0002$). These findings underscore the need for interventions aimed at improving access to clean water and healthcare services, particularly for pregnant teenagers and young children, to reduce under-5 mortality in these regions.

Table 4.22: Spatial Lag Model Impacts of Under 5 years Mortality Rate

Impacts computed using the 'simple' method.

Variable	Direct	Indirect	Total
Women Formal Education	8.4453	0.0831	8.5284
Basic Sanitation (%)	7.0915	0.0698	7.1613
Clean Drinking Water	-16.6748	-0.1641	-16.8389
Access to Clean Fuels %	5.2270	0.0514	5.2784
Teenage Pregnancy %	8.0337	0.0790	8.1127
Caesarian Section	1.8096	0.0178	1.8274
Births by Skilled Provider	16.4026	0.1614	16.5640
ANC visits (4+) %	-0.4982	-0.0049	-0.5031
Post-natal Visits %	-9.8819	-0.0972	-9.9792
Malnutrition %	9.5532	0.0940	9.6472
HIV Infection Rate	1.0839	0.0107	1.0946
Respiratory Complications	3.3795	0.0333	3.4128
Congenital Malformation	5.0449	0.0496	5.0946
Vaccination Coverage (%)	-20.0130	-0.1969	-20.2099
Non Exclusive Breastfeeding	0.4125	0.0041	0.4166
Malaria Prevalence %	-4.6108	-0.0454	-4.6562
Overall Poverty %	3.2294	0.0318	3.2612
Solid Fuel Use	5.1341	0.0505	5.1846

Dependent Variable:	U5 Mortality	Number of Observations:	68
Mean dependent var:	101.8228	Number of Variables:	20
S.D. dependent var:	38.4187	Degrees of Freedom:	48
Pseudo R-squared:	0.9691		
Spatial Pseudo R-squared:	0.9692		
Log likelihood:	-225.8579	Akaike info criterion:	491.716
Sigma-square ML:	44.9249	Schwarz criterion:	536.106
S.E of regression:	6.7026		

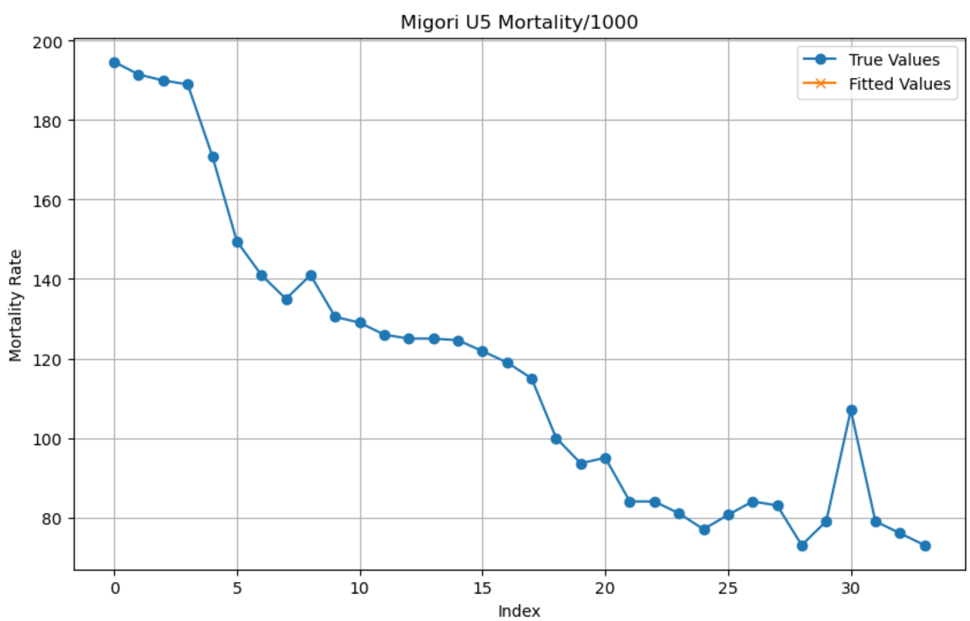

Figure 4.13: Spatial Impacts, Migori U5 Mortality/1000

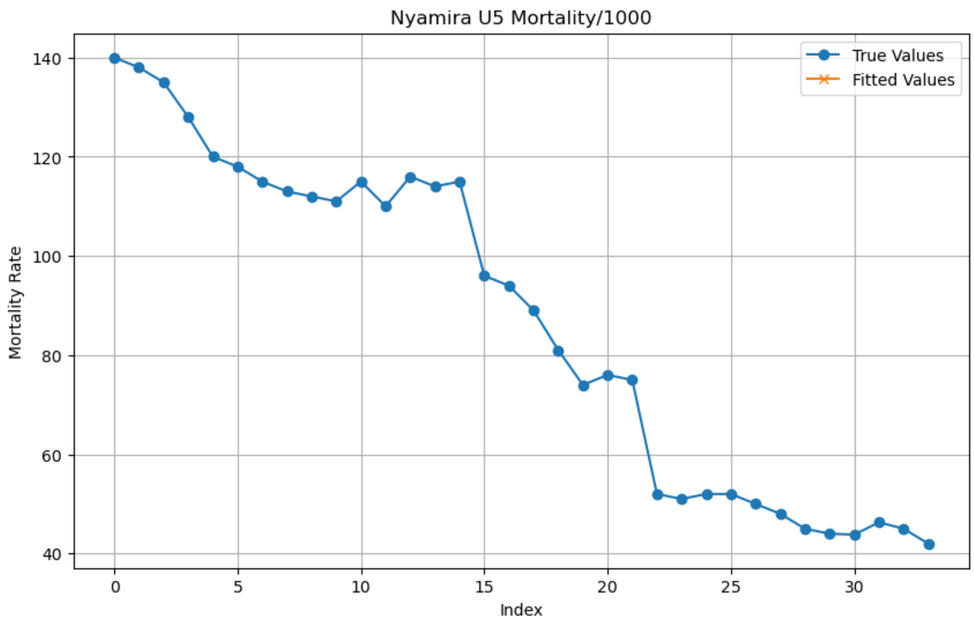

Figure 4.14: Spatial Impacts, Nyamira U5 Mortality/1000

CHAPTER FIVE

DISCUSSION

5.1 Discussion

The study aimed to achieve four key objectives: developing a predictive model for infant mortality, examining temporal and seasonal trends, analyzing social determinants, and evaluating model performance across different infant risk groups in Nyamira and Migori counties, Kenya. Each objective was successfully met through various statistical and machine learning approaches, providing a comprehensive understanding of infant mortality in these regions.

5.2 Developing a context-specific predictive model for child mortality using indirect estimation techniques in Nyamira and Migori counties, Kenya

The study developed context-specific predictive models using multiple linear regression (MLR), Gradient Boosting Machines (GBM), and Spatio-temporal Autoregressive Models (STAR). The multiple linear regression analysis revealed that lack of access to clean water was a critical predictor, with a statistically significant effect on increasing neonatal mortality (coefficient = -5.8461, $p = 0.003$).

The forecast was even more precise for the GBM models, highlighting the importance of variables like malnutrition and teenage pregnancy in both counties. The predictive capacity of the models varied, however, with the GBM model for Nyamira having weak predictive power for neonatal mortality, suggesting the need for more complex models or additional data to capture the complexity of infant mortality in this county.

The role of socio-economic determinants such as women's education, access to clean water, and the incidence of malaria was captured in the models, as they were both significant predictors in Nyamira and Migori counties. In Migori, for example, neonatal mortality was significantly influenced by lack of access to clean water, while teenage pregnancy was the most important risk factor in Nyamira.

Spatial disparities were handled by employing spatial-temporal modeling methods, such as Spatial Lags and STAR models, which account for the geographical variability in mortality risk. This improved the predictive power of models by capturing the differences between sub-counties in terms of healthcare infrastructure, environmental conditions, and other socio-economic factors. Moreover, predictive models accounted for socio-economic factors such as maternal education, access to healthcare, and living conditions. Studies suggest that machine learning models like Random Forest and Gradient Boosting were particularly effective in handling complex datasets (Burroway, 2017). These models managed identify patterns in under-resourced settings, offering predictions based on numerous variables like access to clean water, healthcare availability, and poverty levels.

Socio-economic determinants, such as maternal education levels and access to basic sanitation, emerged as significant predictors in both Migori and Nyamira counties (as seen in the regression results). A multiple linear regression analysis revealed that lack of access to clean water was a critical predictor, with a statistically significant effect on increasing neonatal mortality (coefficient = -5.8461, p = 0.003). This finding is in line with broader literature, which highlights access to clean water as a central determinant of reducing child mortality, particularly in poorer settings (Kwesiga et al., 2022).

In relation to spatial inequities, the spatial-temporal autoregressive (STAR) models captured adequately the regional disparities in infant mortality, particularly in areas with low healthcare coverage. The STAR models utilize the geographical information to identify sub-county hotspots for death and intervene accordingly. This spatial approach was supplemented with recent research Getis & Ord (2020), which has established that geographical inequities in health outcomes are key determinants of death in low-resource settings. Socio-economic conditions also play an important role in infant mortality. Maternal education, access to clean water, and sanitation are factors that play a significant role in the rates of child survival. It is evidence that greater maternal education leads to better infant health status, with educated mothers tending to be more likely to undergo prenatal care and receive medical care.

5.3 Examining the temporal patterns and seasonal trends in child mortality rates using machine learning approaches in Nyamira and Migori counties, Kenya.

Season and time series analysis as well as machine learning techniques were employed to explore temporal patterns in infant mortality. The findings revealed that infant mortality in the two counties has fluctuated across the years with definite patterns established based on disease conditions and the availability of healthcare services. For instance, the models projected increased mortality during periods of heightened respiratory complications as well as malaria transmission, particularly in Migori. This aligns with empirical evidence from a study in sub-Saharan Africa, where an increase in seasonality of malaria transmission and respiratory disease is strongly correlated with increased infant and child mortality (Wang et al., 2023). Such time-series findings are significant in the planning of public health to facilitate interventions to target high-risk seasons, like rainy seasons when access to healthcare is limited, outbreaks of respiratory disease, and lack of clean water and sanitation.

Temporal trends in infant mortality, as observed in the two counties between 1989 and 2022, exhibit a declining trend over years with extreme fluctuations. Seasonal trends were also observed, primarily because of malaria endemicity, with an even larger peak during rainy seasons. Temporal and seasonal trends may have the potential to affect the validity of time-series models and machine learning models in identifying risk of infant mortality. Gradient Boosting Machines (GBM) was a machine learning algorithm that fared well in both the two counties with high predictability of infant mortality at an R^2 value of 0.88 in Migori. Low predictability for neonatal mortality was recorded in Nyamira ($R^2 = 0.81$) and captures the requirement for further finetuning in recognizing the complexity of mortality in the region.

Determining infant mortality trends over time and season was imperative to construct predictive models. The research of temporal patterns in Nyamira and Migori presented mortality variations with regards to seasons and healthcare access. For instance, malaria, respiratory disease, and malnutrition contribute significantly to infant death in rainy seasons, yet access to care could be disrupted at this time as well due to infrastructure weaknesses (Leal et al., 2018). Comparison of the data in Nyamira and Migori counties unveiled characteristic seasonal and temporal trends of infant mortality between 1989 and 2022. Both counties had seasonal spikes in mortality, caused by acute increases in the disease burden during rainy seasons when malaria as well as respiratory conditions are likely to rise (Leal et al., 2018). For instance, neonatal mortality was significantly positively correlated with malaria prevalence (coefficient = 3.0453, $p = 0.033$), as indicated by studies linking seasonal disease epidemics to greater child mortality.

Understanding temporal and seasonal trends in infant mortality was essential for improving predictive models. Research on temporal patterns in Nyamira and Migori indicated fluctuations in mortality rates linked to seasons and healthcare availability. For instance, malaria prevalence, respiratory infections, and malnutrition significantly increase infant mortality during rainy seasons, when access to healthcare can also be limited due to infrastructure challenges (Leal et al., 2018). Analysis of the data from Nyamira and Migori counties indicated clear temporal and seasonal trends in infant mortality between 1989 and 2022. In both counties, mortality rates showed seasonal spikes, which can be attributed to periods of increased disease burden, particularly during rainy seasons when incidences of malaria and respiratory complications tend to rise (Leal et al., 2018). For instance, malaria prevalence had a statistically significant positive association with neonatal mortality (coefficient = 3.0453, $p = 0.033$), which is consistent with studies that link seasonal disease outbreaks to higher child mortality.

5.4 Analysis of survey and demographic data to determine social determinants of child mortality in Nyamira and Migori counties, Kenya.

The analysis of survey and demographic data provided a deep understanding of the social determinants influencing infant mortality in Nyamira and Migori. Key factors such as maternal education, access to clean water and sanitation, and socioeconomic conditions were found to significantly impact mortality rates. For example, the lack of formal education among women was strongly associated with higher infant mortality in both counties. These findings underscore the need for comprehensive public health strategies that address both medical and social determinants of health.

Respiratory complications and post-natal care were additional significant factors identified in the current study. Respiratory conditions were highly predictive of neonatal mortality, especially in regions with poor air quality and reliance on solid fuels for cooking. This finding is supported by existing research, which links solid fuel use to indoor air pollution and a heightened risk of respiratory infections among young children (Sartorius & Sartorius, 2014). The importance of post-natal care was also evident, as a lack of follow-up visits significantly increased the risk of infant mortality (coefficient = -5.4261, p = 0.024). This finding aligns with studies by Lythgoe et al. (2021), which emphasize the need for post-natal interventions to improve infant survival in resource-constrained settings.

Access to clean drinking water and basic sanitation were other significant determinants of child survival. The regression results from Migori County indicated that improved water access was associated with lower neonatal mortality (coefficient = -5.8461, p = 0.003), mirroring findings from studies conducted in other low-income regions (Yaser Mokhayeri et al., 2020).

5.5 Evaluation of predictive model performance for different risk groups in Nyamira and Migori counties, Kenya.

The study evaluated the performance of the predictive models across different risk groups, such as those affected by malnutrition, respiratory complications, and congenital conditions. The models performed well in predicting mortality among the high-risk groups, particularly in Migori County. The neonatal mortality model for Nyamira was not as predictive, but it could need additional risk factors or alternative approaches to modelling. The findings suggest that interventions need to be targeted to address the needs of each infant risk group individually in order to make public health interventions more effective as a whole.

Child mortality prediction models like neonatal, infant, and under-5 mortality have evolved significantly with the new arrival of both traditional statistical methods and advanced machine learning models. The use of multiple linear regression (MLR), STAR model, and gradient boosting machines (GBM) in the current research is supported by previous findings where these models have been found to explain a significant proportion of mortality outcomes (Wilson & Wakefield, 2020). For example, the GBM model used in Migori and Nyamira counties demonstrated strong predictive power, with an R^2 of 0.962 for under-5 mortality. The fact that machine learning algorithms like GBM can reproduce this level of performance describes their resilience in capturing non-linear patterns between mortality and its determinants over traditional statistical models like MLR in predictability.

The most at-risk group identified consists of newborns, particularly those born to adolescent or less educated mothers. The link between adolescent pregnancy and neonatal death (co-efficient = 1.46555, $p = 0.004$) highlights the susceptibility of young mothers and their newborns, as adolescent pregnancies often lead to negative birth circumstances due to inadequate prenatal care, socioeconomic conditions, and elevated complication rates during delivery. This result is consistent with existing research that has identified young motherhood as a significant risk factor for poor neonatal outcomes (Burstein et al., 2018).

Infants born in homes without clean water and sanitation, and infants who are also exposed to malaria and indoor air pollution, were also at risk of dying. Such environmental conditions reinforce socio-economic exposures, creating a lethal combination that drives high infant and neonatal mortality in regions like Nyamira and Migori. The poor children are disproportionately affected because poverty limits access to health facilities, safe water, and proper nutrition that are all important for child survival.

5.6 Contribution to the Body of Knowledge

The study contributes to the body on knowledge through, first, context-specific modelling. The study develops county-specific predictive models for Migori and Nyamira, unlike many global or national-level studies, demonstrating how local socio-economic and environmental factors influence neonatal, infant, and U5 mortality. The localized approach of the analysis bridges a significant data-driven regional health modelling.

What is moreover, the research identifies and affirms high predictors of child mortality. Using various modelling techniques, the study quantifies and reinforces factors such as maternal education, clean water access, teenage pregnancy, and post-natal care as critical determinants. Moreover, the findings use reliable data to identify less-discusses but significant variables such as non-exclusive breastfeeding and respiratory complications as significant predictors of child mortality.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

The analysis successfully developed and validated predictive models for infant mortality in Nyamira and Migori counties, contributing to the broader understanding of infant mortality determinants in these regions. The temporal and seasonal analysis revealed that periods of increased respiratory complications and malaria prevalence corresponded with higher mortality rates. These seasonal trends suggest that targeted interventions during high-risk periods could significantly reduce mortality rates.

The study's predictive models performed well, particularly in identifying high-risk groups, reinforcing the importance of targeted interventions. However, the limited accuracy of the neonatal mortality model in Nyamira suggests that additional factors, potentially including genetic or environmental influences, may need to be explored in future research. This is in line with the findings of Mosley and Chen (2003), who advocate for a more comprehensive approach to understanding the multifaceted determinants of child mortality, including genetic predispositions and environmental factors.

Clean water and sanitation availability, for instance, was established to be a major factor in both counties, particularly in Migori, where lack of access to clean water was associated with higher mortality rates. This is in line with studies such as Wilson & Wakefield (2022), which determined that improved water and sanitation can save 15% of child mortality among developing countries. Maternal education was also emphasized as being important because the individuals with no education were more likely to experience infant mortality. This is reinforced by previous studies, which cite that maternal education, can lead to better health practices, such as increased immunization and better diet, ultimately reducing child mortality.

The temporal and seasonal analysis revealed that mortality rates followed equivalent levels of higher respiratory complications and malaria prevalence. The presence of these seasonal trends further implies that intervention in these high-risk seasons would substantially reduce mortality rates.

In summary, the study not only contributes to the academic understanding of infant mortality in Kenya but also provides practical insights that can inform public health strategies. The findings advocate for a multifaceted approach to reducing infant mortality, combining improvements in healthcare access, education, and targeted interventions during high-risk periods. This comprehensive strategy is essential for achieving significant reductions in infant mortality rates, not just in Nyamira and Migori but across similar contexts in sub-Saharan Africa.

6.2 Recommendations

Strengthening Healthcare Services: There is a need to enhance access to essential healthcare services, particularly in rural areas, to reduce infant mortality. This includes improving access to clean water and sanitation, increasing the number of skilled birth attendants, and ensuring regular antenatal and post-natal care visits.

Maternal Education Programs: Implementing educational programs aimed at improving maternal knowledge and health practices is crucial. These programs should focus on the importance of formal education, exclusive breastfeeding, and proper nutrition to improve infant health outcomes.

Targeted Public Health Interventions: Public health interventions should be tailored to address the specific needs of high-risk infant groups and seasonal trends in mortality. This includes strengthening malaria control programs, enhancing respiratory health interventions, and providing targeted support during periods of increased mortality risk.

6.3 Limitations

The study found that majority of the deceased neonates were born below the gestational age of 37 weeks. This implies that the likelihood of dying during the neonatal period could be higher for preterm neonates than for term- and post-term neonates combined. Babies born early or preterm may develop conditions that place them at higher risk for short-term problems, long term neurological complications and even death. Furthermore, the study faced several limitations, including the use of retrospective data, which may have inherent biases. Additionally, the predictive models for neonatal mortality in Nyamira showed limited accuracy, suggesting that some important variables may have been omitted or inadequately captured. The small sample size for certain variables may have also impacted the model's performance.

6.4 Future Research

Future studies should focus on collecting more detailed and up-to-date data, particularly on neonatal health indicators. Exploring the impact of environmental and genetic factors on infant mortality, as well as integrating machine learning techniques such as deep learning, could enhance model performance. Expanding the study to include other counties in Kenya could also provide a more comprehensive understanding of infant mortality across different contexts.

Future research should explore additional risk factors and employ alternative modelling techniques to improve the predictive accuracy of neonatal mortality models, particularly in Nyamira County. Incorporating more granular data and exploring non-linear relationships between variables may yield better predictive insights.

REFERENCES

Andersen, P. K. (2014). Time-Dependent Covariate. *Wiley StatsRef: Statistics Reference Online*, *4*(14). https://doi.org/10.1002/9781118445112.stat06066

Andreev, E. M., & Kingkade, W. W. (2015). Average age at death in infancy and infant mortality level: Reconsidering the Coale-Demeny formulas at current levels of low mortality. Demographic Research, 33, 363–390. https://doi.org/10.4054/demres.2015.33.13

Bendavid, E., & Bhattacharya, J. (2014). The relationship of Health Aid to population health improvements. *JAMA Internal Medicine*, *174*(6), 881–887. https://doi.org/10.1001/jamainternmed.2014.292

Berhanu, B., Oljira, L., Demana, M., Negash, B., Mamo Ayana, G., Beshir Raru, T., & Haile, D. (2021). Survival and predictors of mortality among neonates admitted to Neonatal Intensive Care Unit at Bombe Primary Hospital, southern Ethiopia: Institution-based retrospective Cohort Study. *Pediatric Health, Medicine and Therapeutics*, *Volume 12*, 239–249. https://doi.org/10.2147/phmt.s303158

Burroway, R. (2017). Political economy, capability development, and fundamental cause: Integrating perspectives on Child Health in Developing Countries. *Journal of World-Systems Research*, *23*(1), 62–92. https://doi.org/10.5195/jwsr.2017.639

Burstein, R., Wang, H., Reiner, R. C., & Hay, S. I. (2018). Development and validation of a new method for indirect estimation of neonatal, infant, and child mortality trends using summary birth histories. *PLOS Medicine*, *15*(10). https://doi.org/10.1371/journal.pmed.1002687

Capili, B., & Anastasi, J. K. (2021). Cohort studies. *AJN, American Journal of Nursing*, *121*(12), 45–48. https://doi.org/10.1097/01.naj.0000803196.49507.08

Daniel, K., Onyango, N. O., & Sarguta, R. J. (2021). A spatial survival model for risk factors of under-five child mortality in Kenya. *International Journal of Environmental Research and Public Health*, *19*(1), 399–412. https://doi.org/10.3390/ijerph19010399

Eilerts, H., Romero Prieto, J., Eaton, J. W., & Reniers, G. (2021). Age patterns of under-5 mortality in sub-Saharan Africa during 1990–2018: A comparison of estimates from demographic surveillance with full birth histories and the historic record. *Demographic Research*, *44*(18), 415–442. https://doi.org/10.4054/demres.2021.44.18

Frank, J., Abel, T., Campostrini, S., Cook, S., Lin, V. K., & McQueen, D. V. (2020). The Social Determinants of Health: Time to Re-Think? *International Journal of Environmental Research and Public Health*, *17*(16), 5856. https://doi.org/10.3390/ijerph17165856

Gatabazi, P., Melesse, S. F., & Ramroop, S. (2019). Infant mortality at the Kigali University Teaching Hospital: Application of Aalen additive hazards model and comparison with other classical survival models. *African Population Studies*. https://doi.org/10.11564/33-2-1382

Getis, A., & Ord, J. K. (2020). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, *24*(3), 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x

Greenberg, A. J., Serrano, K. J., Thai, C. L., Blake, K. D., Moser, R. P., Hesse, B. W., & Ahern, D. K. (2017). Public use of electronic personal health information: Measuring progress of the healthy people 2020 objectives. *Health Policy and Technology*, *6*(1), 33–39. https://doi.org/10.1016/j.hlpt.2016.08.003

Helleringer, S., Liu, L., Chu, Y., Rodrigues, A., & Fisker, A. B. (2020). Biases in survey estimates of neonatal mortality: Results from a validation study in urban areas of Guinea-Bissau. *Demography*, *57*(5), 1705–1726. https://doi.org/10.1007/s13524-020-00911-6

Imbo, A., Mbuthia, E., & Ngotho, D. (2021). Determinants of neonatal mortality in Kenya: Evidence from the Kenya Demographic and Health Survey 2014. *International Journal of Maternal and Child Health and AIDS (IJMA)*, *10*(2), 287–295. https://doi.org/10.21106/ijma.508

Jaskari, J., Myllärinen, J., Leskinen, M., Rad, A. B., Hollmén, J., Andersson, S., & Särkkä, S. (2020). Machine learning methods for neonatal mortality and morbidity classification. *IEEE Access*, *8*, 123347–123358. https://doi.org/10.1109/ACCESS.2020.3006710

Jones, N. L., Gilman, S. E., Cheng, T. L., Drury, S. S., Hill, C. V., & Geronimus, A. T. (2019). Life course approaches to the causes of health disparities. *American Journal of Public Health*, *109*(S1), S48–S55. https://doi.org/10.2105/ajph.2018.304738

Kwesiga, D., Eriksson, L., Orach, C. G., Tawiah, C., Imam, M. A., Fisker, A. B., Enuameh, Y., Lawn, J. E., Blencowe, H., Waiswa, P., Bradby, H., & Malqvist, M. (2022). Adverse pregnancy outcome disclosure and women's Social Networks: A qualitative multi-country study with implications for improved reporting in surveys. *BMC Pregnancy and Childbirth*, *22*(1), 1–11. https://doi.org/10.1186/s12884-022-04622-1

Langat, A. K., Ofori, M. A., Mutinda, J. K., Mouhamadou Djima Baranon, Adeladza Kofi Amegah, & Kazembe, L. N. (2024). Spatial regression modelling of child survival on the distribution of births and deaths in Kenya based on the Kenya demographic and health survey (KDHS) 2022. *Asian Journal of Probability and Statistics*, *26*(3), 67–79. https://doi.org/10.9734/ajpas/2024/v26i3600

Leal, M. do, Szwarcwald, C. L., Almeida, P. V., Aquino, E. M., Barreto, M. L., Barros, F., & Victora, C. (2018). Reproductive, maternal, neonatal and child health in the 30 years since the creation of the Unified Health System (SUS). *Ciência & Saúde Coletiva*, *23*(6), 1915–1928. https://doi.org/10.1590/1413-81232018236.03942018

Lu, Z., Bandara, J. S., & Paramati, S. R. (2020). Impact of sanitation, safe drinking water and health expenditure on infant mortality rate in developing economies. *Australian Economic Papers*, *59*(1), 13–33. https://doi.org/10.1111/1467-8454.12167

Lythgoe, C., Lowe, K., McCauley, M., & McCauley, H. (2021). How women's experiences and perceptions of care influence uptake of postnatal care across sub-Saharan Africa: a qualitative systematic review. *BMC Pregnancy and Childbirth*, *21*(1). https://doi.org/10.1186/s12884-021-03910-6

Macharia, P. M., Joseph, N. K., Sartorius, B., Snow, R. W., & Okiro, E. A. (2021). Subnational estimates of factors associated with under-five mortality in kenya: A spatio-temporal analysis, 1993–2014. *BMJ Global Health*, *6*(4), e004544. https://doi.org/10.1136/bmjgh-2020-004544

Masquelier, B., Ashira Menashe-Oren, Georges Reniers, & Timæus, I. M. (2024). A new method for estimating recent adult mortality from summary sibling histories. *Population Health Metrics*, *22*(1). https://doi.org/10.1186/s12963-024-00350-0

Mazzucato, M., Dalla, V., Minichiello, C., Toto, E., Vianello, A., & Facchin, P. (2023). Estimating mortality in rare diseases using a population-based registry, 2002 through 2019. *Orphanet Journal of Rare Diseases*, *18*(1). https://doi.org/10.1186/s13023-023-02944-7

Mfateneza, E., Rutayisire, P. C., Biracyaza, E., Musafiri, S., & Mpabuka, W. G. (2022). Application of machine learning methods for predicting infant mortality in Rwanda: Analysis of Rwanda Demographic Health Survey 2014–15 Dataset. *BMC Pregnancy and Childbirth*, *22*(1), 1–13. https://doi.org/10.1186/s12884-022-04699-8

Mokhayeri, Y., Mohammad Riahi, S., & Rafiei, E. (2020). Indirect estimation of child mortality using 2011 census data in the Islamic Republic of Iran. *Eastern Mediterranean Health Journal*, *26*(2), 161–169. https://doi.org/10.26719/2020.26.2.161

Mosley, W. H., & Chen, L. C. (1984). An analytical framework for the study of child survival in developing countries. *Bulletin of the World Health Organization*, *81*(2), 140–145.

Murphy, G. A., Waters, D., Ouma, P. O., Gathara, D., Shepperd, S., Snow, R. W., & English, M. (2017). Estimating the need for inpatient neonatal services: An iterative approach employing evidence and expert consensus to guide local policy in Kenya. *BMJ Global Health*, *2*(4), 1–10. https://doi.org/10.1136/bmjgh-2017-000472

Neupert, R., Fernandez Menjivar, R. E., & Fernandez Castilla, R. E. (2019). Indirect estimation of infant mortality in small areas. *Revista Brasileira de Estudos de População*, *36*, 1–37. https://doi.org/10.20947/s0102-3098a0100

P, Mrs. B. K., Rohini, M., & M, P. G. (2024). Captionify: Bridging the gap Between Vision and Language with Neural Networks. *International Journal for Research in Applied Science and Engineering Technology*, *12*(5), 2155–2160. https://doi.org/10.22214/ijraset.2024.61513

Patel, A., Khatib, M. N., Kurhe, K., Bhargava, S., & Bang, A. (2017). Impact of neonatal resuscitation trainings on neonatal and perinatal mortality: a systematic review and meta-analysis. *BMJ Paediatrics Open*, *1*(1), e000183. https://doi.org/10.1136/bmjpo-2017-000183

Paul, M. (2023, January 19). *Neonatal mortality rate in Kenya shows no significant change in 8 years: Survey*. Down To Earth. https://www.downtoearth.org.in/news/africa/neonatal-mortality-rate-in-kenya-shows-no-significant-change-in-8-years-survey-87230

Reisman, J., Arlington, L., Jensen, L., Louis, H., Suarez-Rebling, D., & Nelson, B. D. (2016). Newborn resuscitation training in resource-limited settings: A systematic literature review. *Pediatrics*, *138*(2). https://doi.org/10.1542/peds.2015-4490

Santosa, A., Wall, S., Fottrell, E., Högberg, U., & Byass, P. (2014). The Development and Experience of Epidemiological Transition Theory over Four decades: a Systematic Review. *Global Health Action*, *7*(1). https://doi.org/10.3402/gha.v7.23574

Sartorius, B. K., & Sartorius, K. (2014). Global infant mortality trends and attributable determinants – an ecological study using data from 192 countries for the period 1990–2011. *Population Health Metrics*, *12*(1). https://doi.org/10.1186/s12963-014-0029-6

Shikuku, D. N., Milimo, B., Ayebare, E., Gisore, P., & Nalwadda, G. (2018). Practice and outcomes of neonatal resuscitation for newborns with birth asphyxia at kakamega county general hospital, kenya: A direct observation study. *BMC Pediatrics*, *18*(1). https://doi.org/10.1186/s12887-018-1127-6

Smith-Greenaway, E., & Thomas, K. J. (2014). Exploring child mortality risks associated with diverse patterns of maternal migration in Haiti. *Population Research and Policy Review*, *33*(6), 873–895. https://doi.org/10.1007/s11113-014-9326-9

Stoner, O., Economou, T., & Drummond Marques da Silva, G. (2019). A hierarchical framework for correcting under-reporting in count data. *Journal of the American Statistical Association*, *114*(528), 1481–1492. https://doi.org/10.1080/01621459.2019.1573732

Swanson, D. A. (2019). Estimating the underlying infant mortality rates for small populations, even those reporting zero infant deaths: A case study of 66 local health areas in British Columbia. *Canadian Studies in Population*, *46*(2), 173–187. https://doi.org/10.1007/s42650-019-00014-7

Tesema, G. A., Teshale, A. B., & Tessema, Z. T. (2021). Incidence and predictors of under-five mortality in East Africa using multilevel Weibull regression modeling. *Archives of Public Health*, *79*(1), 1–13. https://doi.org/10.1186/s13690-021-00727-9

Tin Oung, M., Richter, K., Prasartkul, P., & Tangcharoensathien, V. (2017). Myanmar mortality registration: An assessment for system improvement. *Population Health Metrics*, *15*(1). https://doi.org/10.1186/s12963-017-0153-1

UNICEF. (2023, May 22). *Kenya (KEN) - demographics, Health & Infant Mortality*. UNICEF DATA. https://data.unicef.org/country/ken/

Wairoto, K. G., Joseph, N. K., Macharia, P. M., & Okiro, E. A. (2020). Determinants of subnational disparities in antenatal care utilisation: A spatial analysis of demographic and Health Survey Data in Kenya. *BMC Health Services Research*, *20*(1), 1–12. https://doi.org/10.1186/s12913-020-05531-9

Wakibi, S., & Ngure, E. (2021). An assessment of knowledge, attitude, and practices of birth and death registration in Kilifi County in the coastal region in Kenya. *BioMed Research International*, *2021*, 1–8. https://doi.org/10.1155/2021/9963703

Wang, S., Ren, Z., & Liu, X. (2023). Spatiotemporal trends in neonatal, infant, and child mortality (1990–2019) based on Bayesian spatiotemporal modelling. *Frontiers in Public Health*, *11*, 1–11. https://doi.org/10.3389/fpubh.2023.996694

WHO. (2023). *Kenya Data*. World Health Organization. https://data.who.int/countries/404

Wilson, K., & Wakefield, J. (2020). Child mortality estimation incorporating summary birth history data. *Biometrics*, *77*(4), 1456–1466. https://doi.org/10.1111/biom.13383

Wilson, K., & Wakefield, J. (2022). A probabilistic model for analysing summary birth history data. *Demographic Research*, *47*, 291–344. https://doi.org/10.4054/demres.2022.47.11

Wong, J. S., Forster, J. J., & Smith, P. W. (2023). Bayesian model comparison for mortality forecasting. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *72*(3), 566–586. https://doi.org/10.1093/jrsssc/qlad021

Yadav, D., Sahoo, L., Sanjeev Kumar Mandal, G. Ravivarman, P. Vijayaraghavan, & Prasad B. (2023). Using Long Short-Term Memory Units for Time Series Forecasting. *Semantic Scholar*, *2023*(2). https://doi.org/10.1109/incoft60753.2023.10425756

Yang, H., Pan, Z., & Tao, Q. (2017). Robust and adaptive online time series prediction with long short-term memory. *Computational Intelligence and Neuroscience*, *2017*, 1–9. https://doi.org/10.1155/2017/9478952

Yego, F., Stewart Williams, J., Byles, J., Nyongesa, P., Aruasa, W., & D'Este, C. (2013). A retrospective analysis of maternal and neonatal mortality at a teaching and referral hospital in Kenya. *Reproductive Health*, *10*(1). https://doi.org/10.1186/1742-4755-10-13

Zhou, L., Zhang, D., Yang, C. C., & Wang, Y. (2018). Harnessing social media for Health Information Management. *Electronic Commerce Research and Applications*, *27*, 139–151. https://doi.org/10.1016/j.elerap.2017.12.003

APPENDICES

Appendix I: Migori Data

Year	Neonatal Mortality/1000	Infant mortality/1000	U5 Mortality/1000	Women Formal Education per Household (%)	Basic Sanitation (%)	Access to Clean Drinking Water (%)	Access to Clean Fuels %	Teenage Pregnancy %	Ceaserian Section	Births by Skilled Provider %	ANC visits (4+) %	Post-natal Visits %
1989	43	77.3	194.7	55	8.6	12.5	1	66	2.2	52	22	24.00022
1990	42	77.1	191.5	56.6	8.925464	12.6956	0.99	65.32	2.254676	54.60984	25	24.24867
1991	41	76.735	169	56.9	11.03128	13.66818	1	65	2.48323	55.37748	25.00001	24.38036
1992	41	75.935	148	57.1	11.32259	14.68415	1.28	64.21	2.748804	56.62491	25.00013	25.11724
1993	40	75.003	189	57.7	11.77707	18.87585	1.23	63.58	2.85026	58.17353	25.00099	25.28427
1994	39	74.871	149.5	57.94	12.25137	19.0612	1.48	55.26	2.865041	58.46477	25.00123	25.72876
1995	38	73.651	141	58.2	12.28726	19.34714	1.4	51.74	3.121076	60.01519	25.00478	27.1189
1996	36	73.195	135	58.41484	12.3833	19.66846	1.42	49.5	3.300983	60.64167	25.06411	27.57207
1997	36	73.585	141	60.68038	12.4333	19.81143	1.55	41.5	3.463767	62.19957	25.11993	27.97443
1998	38	72.97	130.5	61.14124	12.52525	19.83779	1.8	37.58	3.645406	65.72903	25.31116	28.29099
1999	35	143.4	240	61.85941	12.92645	20.07391	2	33.55	3.725769	66.69275	26.0657	29.09301
2000	34	71.658	126	61.95977	13.78136	20.28851	2.14	31.22	3.842955	67.16988	26.13981	31.5213
2001	32	70.043	125	62.18358	15.05716	20.55072	2.26	30.85	3.844557	69.00933	27.55847	35.25105
2002	32	69.4	125	63.18698	15.39485	20.81567	2.45	29.66	3.883158	69.71076	28.15494	35.75306
2003	32	66.758	124.6	64.93346	15.77508	21.16329	2.49	29.01	3.906238	71.63805	29.3051	36.08408
2004	31	64.758	91	65.31393	15.79737	21.34394	2.5	28.46	4.030386	72.09489	29.57547	42.92166
2005	30	63.429	119	66.4159	15.90029	21.39633	2.52	28.02	4.207661	74.95548	30.03802	43.65
2006	31	143.4	240	67.75333	15.99044	21.48704	3.5	27.78	4.215482	75.19162	33.84001	45
2007	30.5	59.656	100	68.52785	16.74708	21.6137	2.65	27.44	4.83	75.528	33.86303	46
2008	32	58.555	93.6	69.26646	16.75444	21.81258	2.74	27	5.5	76.21307	35.56558	47.66
2009	27	76	123	69.83106	22.2	28.81	2.9	27	6.2	76.29488	37.36234	48
2010	29	53.242	84	70.42129	22.67	30.87	3	27.52	6.52	76.55979	43.00453	48.85

2011	27	112	123	70.69454	23	32	3.4	26.33	6.8	76.66119	45.08457	49
2012	26.5	50.361	81	70.98152	25	32.55	3.6	26.61	7.11	78.00067	46.8393	53.37007
2013	28	49.556	77	80	26.3	33	3.9	27	7.84	78.13931	47	50.0472
2014	26.62	48.513	123	81.6	26.9	35.65	4.1	33	8.7	79.01506	47.45	51.52904
2015	25	46.47	84	82	29.87	37.19	4.2	33	9.23	79.24643	47.62	51.53699
2016	24.5	44.426	83	82.6	34.36	37.98	4.4	28	10.56	79.24684	48.65	51.72436
2017	24.02	42.383	73	83.2	44.56	42.36	4.8	26.6	12.32	79.3061	53.3	52.29004
2018	24	39.34	79	84	48.74	44.5	4.9	26	14.44	95.2	54.86	52.41526
2019	23.69	67.2	107	84.6	50	46.1	4.9	25	16.4	95.7	56.33	53.04364
2020	23	35.046	79	85.4	51.52	49.82	4.9	40	21.65	95	57	55
2021	22.96	32.271	76	86.62	52	52.63	5.2	24	28.39	95	58.3	55.64
2022	22.5	53	73	87.12	52.4	53	5.4	23	33	92.6	58.5	56

Appendix II: Nyamira Data

Year	Neonatal Mortality/1000	Infant mortality/1000	U5 Mortality/1000	Women Formal Education per Household (%)	Basic Sanitation (%)	Access to Clean Drinking Water (%)	Clean Fuels %	Teenage Pregnancy %	Ceaserian Section %	Births by Skilled Provider %	ANC visits (4+) %	Post-natal Visits %
1989	32	64	140	49	9.5	22.68	1	56	2.2	52	24.00022	24.00022
1990	31	67.06	138	49.6	14.05192	23.85786	0.99	55.22	2.254676	54.60984	24.24867	24.24867
1991	30.6	66.89	135	50.8	14.85425	24.00224	1	49.3	2.48323	55.37748	24.38036	24.38036
1992	33.4	66.63	128	58.65367	15.68202	25.12862	1.28	44.27239	2.748804	56.62491	25.11724	25.11724
1993	26	66.24	120	59.49664	17.69544	25.33052	1.23	43.86965	2.85026	58.17353	25.28427	25.28427
1994	29	65.71	118	60.04482	18.06089	26.4717	1.48	43.75474	2.865041	58.46477	25.72876	25.72876
1995	29	65.25	115	62.42426	18.33155	30.01234	1.4	43.35151	3.121076	60.01519	25	27.1189
1996	30	64.44	113	64.57925	19.14701	30.18002	1.42	43.22082	3.300983	60.64167	25.06411	27.57207
1997	30	64.03	112	64.6771	20.27513	30.74873	1.55	42.82696	3.463767	62.19957	25.11993	27.97443
1998	28	62.38	111	65.20543	21.12688	31.18315	1.8	42.65238	3.645406	65.72903	25.31116	28.29099
1999	31	60.48	115	65.23427	22.13404	31.52231	2	42.25902	3.725769	66.69275	26.0657	29.09301
2000	31	58.19	110	65.44715	22.20508	33.11744	2.14	40.99361	3.842955	67.16988	26.13981	31.5213
2001	32	55.59	116	67.1077	22.71478	33.84363	2.26	39.16405	3.844557	69.00933	27.55847	35.25105
2002	29	52.91	114	67.23276	23.20526	34.15316	2.45	38.63739	3.883158	69.71076	28.15494	35.75306
2003	33	50.21	115	68.92747	23.24027	34.49077	2.49	38.42482	3.906238	71.63805	29.3051	36.08408
2004	28	47.47	96	69.6459	23.58578	34.765	2.5	37.91573	4.030386	72.09489	29.57547	42.92166
2005	32	44.5	94	69.95685	23.64505	35.29475	2.52	35.04792	4.207661	74.95548	30.03802	43.65
2006	29	42.16	89	69.97686	23.9241	36.30223	3.5	34.69005	4.215482	75.19162	33.84001	45
2007	29	40.3	81	71.08841	24.13609	37.10548	2.65	32.74765	4.83	75.528	33.86303	46
2008	30	38.06	74	72.87521	24.35188	38.501	2.74	30.69083	5.5	76.21307	35.56558	31
2009	27	37.4	76	74.89173	33.2	48.9	2.9	29.89783	6.2	76.29488	37.36234	31
2010	29	36.93	75	76.55085	28	48	3	29.7	6.52	76.55979	38.66	46.3
2011	25	43	52	82	30	48	3.4	28.64	6.8	76.66119	39.7	52

2012	27	36.44	51	82.5	38.62	48.66	3.6	28	7.11	78.00067	40.56	52.96
2013	26	36.01	52	83	44.55	49.04	3.9	27.94757	7.84	78.13931	44.69	53.53
2014	28	35.19	52	84	50	49.52	4.1	46.01	8.7	79.01506	49.6	53.87
2015	27	34.09	50	85.6	53	49.98	4.2	25.45	9.23	79.24643	47.2	54.22
2016	24.2	33.61	48	86	54.63	50.01	4.4	25.01	10.56	79.24684	48.1	54.69
2017	24	33.41	45	86.2	56.88	50.56	4.8	24.66	12.32	79.3061	33.6	55.11
2018	23	32.76	44	86.4	62.47	51.02	4.9	24.12	14.44	95.2	38.8	57.65
2019	24	32.23	43.8	86.9	64.29	51.5	4.9	23.88	16.4	95.7	46.8	60.66
2020	23	31.77	46.3	87.4	66.11	52	4.9	23.52	21.65	95	57.8	69.55
2021	21	31.01	45	87.8	67	54	5.2	22.57	28.39	95	65.6	72.04
2022	27	31	42	88	68.8	56	5.4	22	33	95.2	65.8	73

Appendix III: Python Code

jupyter STAR Model Last Checkpoint: 14 days ago

File Edit View Run Kernel Settings Help

Code | JupyterLab | Python 3 (ipykernel) | Trusted

```
[47]: # Load necessary Libraries
import pandas as pd
import geopandas as gpd
import numpy as np
from sklearn.preprocessing import StandardScaler
from libpysal.weights import KNN
from spreg import ML_Lag
import matplotlib.pyplot as plt
import seaborn as sns
from docx import Document
from IPython.display import FileLink

# Suppress warnings for cleaner output
import warnings
warnings.filterwarnings('ignore')
```

•••

```
[49]: # Load data from the Excel file
file_path = r"C:\Users\ADMIN\OneDrive\Documents\Child mortality data.xlsx"
migori_data = pd.read_excel(file_path, sheet_name='Migori')
nyamira_data = pd.read_excel(file_path, sheet_name='Nyamira')

# Combine the data from both sheets
data = pd.concat([migori_data, nyamira_data], axis=0, ignore_index=True)
```

•••

```
[51]: # Treating as a categorical variable for temporal trends
data['Year'] = data['Year'].astype(str)

# List of independent variables
independent_vars = [
    'Women Formal Education per Household (%)', 'Basic Sanitation (%)',
    'Access to Clean Drinking Water (%)', 'Access to Clean Fuels %',
    'Teenage Pregnancy %', 'Caesarian Section', 'Births by Skilled Provider %',
    'ANC visits (4+) %', 'Post-natal Visits %', 'Malnutrition %',
    'HIV Infection Rate', 'Respiratory Complications/1000',
    'Congenital Malformation', 'Vaccination Coverage (%)',
    'Non Exclusive Breastfeeding %', 'Malaria Prevalence %',
    'Overall Poverty %', 'Solid Fuel Use'
]

# Normalize/Standardize the independent variables
scaler = StandardScaler()
data[independent_vars] = scaler.fit_transform(data[independent_vars])
```

jupyter STAR Model Last Checkpoint: 14 days ago

Edit View Run Kernel Settings Help

Code | JupyterLab | Python 3 (ipykernel) | Trusted

```
Model for Neonatal Mortality/1000 fitted successfully. •••
```

```
[63]: # Provide a download link (useful in Jupyter Notebooks)
FileLink(output_path)
```

```
<a href='C:\Users\ADMIN\OneDrive\Documents\STAR_model_summary.docx' target='_blank'>C:\Users\ADMIN\OneDrive\Documents\STAR_model_summary.docx</a><br> •••
```

```
[65]: print(model.summary)
```

```
REGRESSION RESULTS •••
```

```
[67]: import matplotlib.pyplot as plt
import seaborn as sns

# Visualization function
def plot_coefficients(dependent_var, coefficients, independent_vars):
    plt.figure(figsize=(12, 8))
    # Ensure the number of coefficients matches the number of independent variables
    if len(coefficients) == len(independent_vars) + 1:  # +1 for intercept
        sns.barplot(x=independent_vars, y=coefficients[1:])  # Skip intercept
        plt.axhline(y=0, color='black', linestyle='--')
        plt.title(f'Coefficients of the Model for {dependent_var}')
        plt.xticks(rotation=45, ha='right')
        plt.xlabel('Independent Variables')
        plt.ylabel('Coefficient Value')
        plt.show()
    else:
        print(f"Error: Number of coefficients ({len(coefficients)}) does not match number of variables ({len(independent_vars)})")

# Visualize coefficients for each model
for dependent_var in dependent_vars:
    y = data_cleaned[dependent_var].values.reshape(-1, 1)
    X = data_cleaned[independent_vars].values

    try:
        model = ML_Lag(y, X, w=weights, name_y=dependent_var, name_x=independent_vars)
        coefficients = model.betas.flatten()
        print(f"{dependent_var} - Number of coefficients: {len(coefficients)}, Number of variables: {len(independent_vars) + 1}")

        # Plot the coefficients
        plot_coefficients(dependent_var, coefficients, independent_vars)

    except Exception as e:
        print(f"Error in visualizing model for {dependent_var}: {e}")
```

Appendix IV: Similarity Report

University of Eldoret

University of Eldoret

Certificate of Plagiarism Check for Thesis

Author Name	Brian Omare SSCI/MAT/M/001/17
Course of Study	Type here...
Name of Guide	Type here...
Department	Type here...
Acceptable Maximum Limit	Type here...
Submitted By	similarity@uoeld.ac.ke
Paper Title	PREDICTIVE MODELING OF CHILD MORTALITY IN MIGORI AND NYAMIRA COUNTIES USING INDIRECT METHODS
Similarity	6%
Paper ID	4612493
Total Pages	135
Submission Date	2025-11-03 08:12:44

Signature of Student

Signature of Guide

Head of the Department

University Librarian

Director of Post Graduate Studies

* This report has been generated by DrillBit Anti-Plagiarism Software