

# Risk Factor and Spatial Pattern Analysis of Neonatal Mortality in Kenya

**Getrude Moraa Nyabuto\*, Bonface Malenje, Anthony Wanjoya**

Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya

## Email address:

gettymoh@gmail.com (Getrude Moraa Nyabuto), bmalenje@jkuat.ac.ke (Boniface Malenje), awanjoya@gmail.com (Anthony Wanjoya)

\*Corresponding author

## To cite this article:

Getrude Moraa Nyabuto, Boniface Malenje, Anthony Wanjoya. (2025). Risk Factor and Spatial Pattern Analysis of Neonatal Mortality in Kenya. *American Journal of Mathematical and Computer Modelling*, 10(2), 54-65. <https://doi.org/10.11648/j.ajmcm.20251002.12>

**Received:** 18 April 2025; **Accepted:** 29 April 2025; **Published:** 3 June 2025

---

**Abstract:** Neonatal health is a critical component of overall public health, providing the groundwork for a healthy life and making a substantial contribution to the social and economic advancement of any nation. Despite the progress that has been made in reducing the global neonatal mortality rate, substantial regional disparities persist, particularly in Sub-Saharan Africa. In Kenya, the NMR stands at 21 deaths per 1,000 live births (as of 2022) which is higher than the global average. The main objective for this study was to perform risk factor and spatial pattern analysis of neonatal mortality in Kenya. A multivariate logistic regression model was fitted that identified urban residence, underweight birth weight status, unimproved water sources, and non-hospital deliveries (especially in non standard locations) as the significant contributors of neonatal mortality in Kenya. Getis-Ord Gi statistics identified Wajir, Garissa, and Lamu counties as major hotspots in Kenya after showing a strong spatial clustering of high neonatal mortality rates. GWLR, utilized in this study, revealed that climatic factors, such as temperature and aridity, impact neonatal mortality differently across regions in Kenya. Generally, higher temperatures are a significant risk factor for neonatal mortality, particularly in arid counties like Mandera, Wajir, Garissa, Tana River, and Lamu.

**Keywords:** KDHS, WHO, GWLR, NMR, ANC, UNICEF

---

## 1. Introduction

In modern policymaking, it is widely recognized that successful national growth policies must prioritize public health as a fundamental component to enhance the well-being of all citizens. Neonatal mortality refers to a newborn's passing during the first 28 days of life. There are several reasons why these deaths may occur including preterm birth, suffocation at birth, neonatal infections (such as sepsis, pneumonia, and meningitis), congenital abnormalities, and complications during childbirth. Inadequate healthcare infrastructure, socioeconomic disparities, restricted access to trained birth attendants, and poor maternal health are the main causes of neonatal fatalities. Unsafe delivery methods, inadequate prenatal care, and limited access to specialized newborn care are other major contributing factors. With prompt measures like appropriate prenatal care and professional attendance, the majority of newborn deaths can be prevented.

Neonatal Mortality Rate is a measure of the number of neonate deaths per 1,000 live births. Over the past few decades, there has been a tremendous improvement in the global NMR. The global newborn mortality rate decreased by more than half, from 37 deaths per 1,000 live births in 1990 to 17 deaths per 1,000 in 2022 [5]. This indicates that 2.3 million babies worldwide perished in their first month of life. Neonatal deaths continue to constitute a considerable component of under-five mortality, accounting for roughly half of all under-five deaths as of 2021, despite the notable progress made in lowering newborn mortality. The first month of life is incredibly fragile because this period is the newborn's biggest and fastest transition to life outside the womb. Fostering healthy development and reducing the hazards that lead to newborn death require an understanding of the scope of this transition.

Neonatal mortality exhibits significant regional variation, with Sub-Saharan Africa and Asia bearing the heaviest

burden. Having 27 neonatal deaths for every 1,000 live births, Sub-Saharan Africa has the highest neonatal mortality rate. In comparison with newborns born in wealthy economies, newborns in this region have a roughly ten-fold higher chance of dying within the first 28 days of their lives. In reality, South Asia and Sub-Saharan Africa account for more than 75% of newborn fatalities worldwide. Premature birth defects, birth asphyxia, infections like sepsis, and congenital anomalies are the primary causes of these deaths. In these areas, the risk is further increased by limited access to high-quality treatment for mothers and newborns.

In Kenya, like many developing countries, the neonatal mortality rate remains higher than the global average, standing at approximately 20 deaths per 1,000 live births in recent years. While this shows some improvement, Kenya experiences varying neonatal mortality rates across different regions, influenced by factors such as access to skilled birth attendants, healthcare quality, socioeconomic disparities, and maternal health conditions. Implementing focused treatments to lower infant deaths and enhance newborn survival outcomes requires an indepth analysis and comprehension of these regional inequalities.

Clusters or enumeration areas are small area aggregates in which KDHS data on newborn mortality is distributed. Distinct units such as villages or neighborhoods are represented by these clusters, which are defined by administrative divisions or geographic borders. Establishing spatial patterns and relationships with sociodemographic factors is partly made possible by the examination of newborn mortality within these clusters. Few studies thoroughly examine population levels determined by DHS statistics, making it difficult to allocate resources and make evidence-based decisions.

While univariate logistic regression can analyze associations between a single predictor variable and an outcome, multivariate modeling accommodates the risk factor analysis of multiple correlated outcomes. Multivariate logistic modeling does not assume independence between predictor variables but rather accomodates the complex interrelationships and correlations among multiple predictors and response variables simultaneously. By taking into account both direct and indirect effects, multivariate logistic modeling can capture how several predictors interact to influence outcomes, in contrast to ordinary logistic regression, which may miss significant variable interactions. This approach has been applied to DHS data to investigate multiple factors that influence breastfeeding, the usage of contraceptives, the nutritional health of children under five, and many other issues [13, 14].

This study will leverage spatial and hotspot analysis to explore the geographic pattern of neonatal mortality in the country for informed public health interventions. Utilizing a Geographically Weighted Logistic Regression model will allow for the exploration of spatially varying relationships between risk factors and neonatal mortality, generating separate regression equations for different geographic locations and revealing location-specific risk factors. This

will yield a more accurate and spatially detailed understanding of neonatal mortality. For the hotspot identification, we will use the Getis-Ord  $G_i^*$  Statistic which gives a measure of the degree of spatial autocorrelation. The regions with the highest values will be considered the hotspots for neonatal mortality.

Multivariate risk factor analysis, spatial analysis, hotspot identification, and mapping are key elements in uncovering the complex interplay of the risk factors of neonatal mortality. Consequently, this study aims to add to the body of knowledge by applying these methods to conduct a thorough analysis of neonatal fatalities in order to pinpoint risk factors that are specific to a given region and guide focused measures that would lower neonatal mortality in Kenya. By fostering multidisciplinary collaboration and interacting with local communities, we can create strong structures that protect the health of newborns and contribute to the continued well-being of future generations.

## 2. Methodology

### 2.1. Data Source and Description

The 2022 Kenya Demographic and Health Survey, which is based on a carefully designed sample taken from the Kenya Household Master Sample Frame (K-HMSF), provides the data used in this study. The KDHS provides nationally representative estimates that are grouped into Kenya's 47 counties and stratified by urban/rural location. Individual neonates found within the primary sampling units are nested within the primary sample units of the survey, and these are further clustered within the counties [3], indicating a hierarchical structure in the data. This comprehensive multistage dataset includes individual-level infant mortality statistics together with associated demographic, socioeconomic, and climatic data.

### 2.2. Multivariate Logistic Regression

As an initial step in our analysis, we used bidirectional stepwise regression. Variables whose elimination led to a drop of the AIC below the predetermined threshold were removed from the model. On the other hand, factors that considerably reduced the AIC upon inclusion were considered for inclusion. Until the model with the lowest AIC -a representation of a balance between model fit and complexity was achieved, this iterative process of adding and eliminating variables persisted. The factors chosen from this first screening were then taken into account for the multivariate logistic regression analysis that followed where they statistically significant factors were determined further using adjusted odd ratio(aOR).

Logistic Regression is a type of GLM specifically designed for binary classification tasks. During the training phase, the algorithm uses labeled data to learn the intricate relationships between input characteristics and the corresponding binary outcome (neonatal survival or death). When given new unseen data, this learning process allows the model to precisely forecast the likelihood that an instance will belong to each

class. Multivariate logistic regression is particularly useful when the target variable is categorical, meaning it has only two possible outcomes and several predictor variables. In this study, the target variable is neonatal mortality, which can either result in "neonatal death" (1) or "survival" (0). The logistic regression model for neonatal mortality is given by:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (1)$$

where:

1.  $p_i$  is the probability of neonatal mortality for individual  $i$ ,
2.  $\beta_0$  is the intercept,
3.  $\beta_1, \beta_2, \dots, \beta_k$  are the regression coefficients for the predictor variables  $X_1, X_2, \dots, X_k$ . Solving for  $p_i$  gives:

$$p_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \quad (2)$$

where:

$$\eta_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k.$$

Substituting  $\eta_i$ :

$$p_i = \frac{\exp(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)}{1 + \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)}. \quad (3)$$

$$\text{Odds Ratio} = \exp(\beta_i).$$

The odds ratio represents how the odds of neonatal mortality change for a one-unit increase in  $X_i$ , holding all other factors constant.

### 2.3. Getis-Ord $G_i^*$ Statistic

This spatial analysis technique is used to identify regions with high and low outcome cases. In the context of this study, the statistic was used to measure the degree of spatial clustering [4] of high(hotspots) or low(coldspots) of neonatal mortality. In essence, it gives a measure of the degree to which values in one location are similar to values at nearby locations,

i.e, the spatial autocorrelation. It's formula is given as:

$$G_i^* = \frac{\sum_j w_{ij} x_j - \bar{X} \sum_j w_{ij}}{S \sqrt{\frac{n \sum_j w_{ij}^2 - (\sum_j w_{ij})^2}{n-1}}} \quad (4)$$

where:

1.  $G_i^*$  : Hotspot/coldspot indicator for location  $i$ .
2.  $x_j$  : Neonatal mortality rate at  $j$ .
3.  $w_{ij}$  : Spatial weight between  $i$  and  $j$ .
4.  $\bar{X}, S$  : Mean and standard deviation of mortality rates.
5.  $n$  : Total locations.

Spatial weights, based on survey cluster proximity, were used in the analysis, which was performed using spatial packages in R. High  $G_i^*$  values indicate hotspots, while low values indicate coldspots. Our research explicitly addressed the interconnection of regions by building a spatial weights matrix based on queen contiguity. In this approach, two regions are considered neighbors if they both share a vertex or an edge. This approach is a more accurate assessment of spatial dependencies and potential spillover effects across the entire study area.

### 2.4. Geographically Weighted Logistic Regression

GWLR is an extension of logistic regression designed to account for spatial variation in regression parameters. Unlike other global spatial models like Spatial Autoregressive Models(SAR) which assume that the relationships between variables are constant across the study area, GWLR acknowledges that these relationships can vary geographically. In a standard logistic regression model, the coefficients are assumed to be constant across the entire study area as well. However, GWLR allows these coefficients to vary spatially, meaning that the relationship between the dependent variable and the explanatory variables may differ across geographic locations [7]. This is especially useful in situations where spatial dependence or heterogeneity exists in the data, such as regional differences in economic, environmental, or demographic factors.

#### 2.4.1. Model Equations

##### Log-Likelihood Function

The log-likelihood function is used to estimate the parameters of the logistic model. In GWLR, it is modified to incorporate spatial variation in the coefficients. The log likelihood function for the data set is expressed as

$$L^*(\beta(\text{loni}, \text{lati})) = \sum_{i=1}^n [y_i \log(\pi_i) + (1 - y_i) \log(1 - \pi_i)] \quad (5)$$

Where:

1.  $y_i$  is the observed binary outcome(neonatal mortality  $i$  (1 for success, 0 for failure)),
2.  $\pi_i = \text{logit}^{-1}(X_i \beta(\text{loni}, \text{lati}))$  The probability that neonatal mortality occurs for observation  $i$ , given the

location and other risk factors.

##### Partial Derivatives (Gradient)

The gradient of the log-likelihood function with respect to the parameters is calculated to find the optimal coefficients. The partial derivatives represent the rate of change of the

log-likelihood with respect to each parameter  $\beta_k$  at a specific location:

$$\frac{\partial L^*(\beta(\text{loni}, \text{lati}))}{\partial \beta_k(\text{loni}, \text{lati})} = X^T W(\text{loni}, \text{lati}) y^* = 0 \quad (6)$$

Setting these derivatives to zero provides the necessary conditions to maximize the log-likelihood, which leads to the

optimal estimates for the coefficients. However, solving these equations analytically is complex due to their implicit nature.

#### *Fisher's Scoring Algorithm*

Fisher's Scoring algorithm is used to iteratively solve for the optimal parameter estimates. This is a numerical optimization technique that updates the parameter estimates based on the gradient and the information matrix. The iterative update rule is:

$$\hat{\beta}^{(t+1)}(\text{loni}, \text{lati}) = \hat{\beta}^{(t)}(\text{loni}, \text{lati}) - I^{-1}(\hat{\beta}^{(t)}(\text{loni}, \text{lati})) \cdot g(\hat{\beta}^{(t)}(\text{loni}, \text{lati})) \quad (7)$$

Where:

1.  $\hat{\beta}^{(t)}(\text{loni}, \text{lati})$  is the estimate of the parameters at the  $t$ -th iteration,
2.  $g(\hat{\beta}^{(t)}(\text{loni}, \text{lati}))$  is the gradient vector at the  $t$ -th iteration (the first derivative of the log-likelihood),
3.  $I(\hat{\beta}^{(t)}(\text{loni}, \text{lati}))$  is the information matrix (the inverse of the second derivative of the log-likelihood), which reflects the precision of the estimates.

#### **2.4.2. Key Components in the Methodology**

##### *Gradient Vector ( $g(\beta)$ )*

The gradient vector is a key component in the iterative process of parameter estimation. It consists of the first partial derivatives of the log-likelihood function with respect to each parameter:

$$g(\beta) = \left[ \frac{\partial L^*(\beta)}{\partial \beta_1}, \frac{\partial L^*(\beta)}{\partial \beta_2}, \dots, \frac{\partial L^*(\beta)}{\partial \beta_k} \right] \quad (8)$$

These partial derivatives tell us the direction in which the log-likelihood function increases most rapidly, allowing the algorithm to adjust the parameter estimates in that direction.

In GWLR, the gradient vector varies across locations ( $\text{loni}, \text{lati}$ ), reflecting how the influence of the explanatory variables changes spatially [11].

##### *Information Matrix ( $I(\beta)$ )*

The information matrix is the second derivative of the log-likelihood function, representing the curvature of the log-likelihood surface. In Fisher's Scoring algorithm, the information matrix is used to adjust the update step size, ensuring convergence to the optimal solution.

$$I(\beta) = - \frac{\partial^2 L^*(\beta)}{\partial \beta \partial \beta^T} \quad (9)$$

It is crucial because it informs the algorithm about the precision of the parameter estimates. A larger value of the information matrix indicates greater precision and faster convergence.

In GWLR, the information matrix also varies spatially, reflecting the different degrees of variability and precision across geographic locations.

##### *Iterative Process (Fisher's Scoring)*

Fisher's Scoring is an iterative optimization algorithm used to update the parameter estimates at each step:

1. Start with an initial guess for the parameter values  $\hat{\beta}^{(0)}(\text{loni}, \text{lati})$ .
2. Compute the gradient vector  $g(\hat{\beta}^{(t)})$  and the information matrix  $I(\hat{\beta}^{(t)})$ .
3. Update the parameters using the rule:

$$\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} - I^{-1} \cdot g(\hat{\beta}^{(t)})$$

4. Repeat steps 2 and 3 until the updates become small, indicating convergence.

##### *Convergence Criteria*

The algorithm continues to iterate until the updates to the parameters become sufficiently small, indicating that the log-likelihood function has been maximized. Convergence is typically determined based on a tolerance threshold (e.g., when the change in log-likelihood or parameter estimates between iterations is below a specified value).

## **3. Results and Discussion**

### **3.1. Descriptive Statistics**

This study utilized data from women aged 15-49 years who were interviewed in the Kenya Demographic and Health Survey (KDHS) regarding their children born within the five years preceding the survey. The statistical analyses were restricted to children under one month of age born to mothers aged 15-49 years during the five-year period prior to the survey. A total of 23,343 children were included in the study. The primary outcome of interest was neonatal survival, categorized into two groups: neonates who survived and those who did not.

Figure 1 presents a geographical visualization of neonatal mortality prevalence across counties in Kenya. Darker shades indicate higher prevalence, while lighter shades represent lower mortality rates.

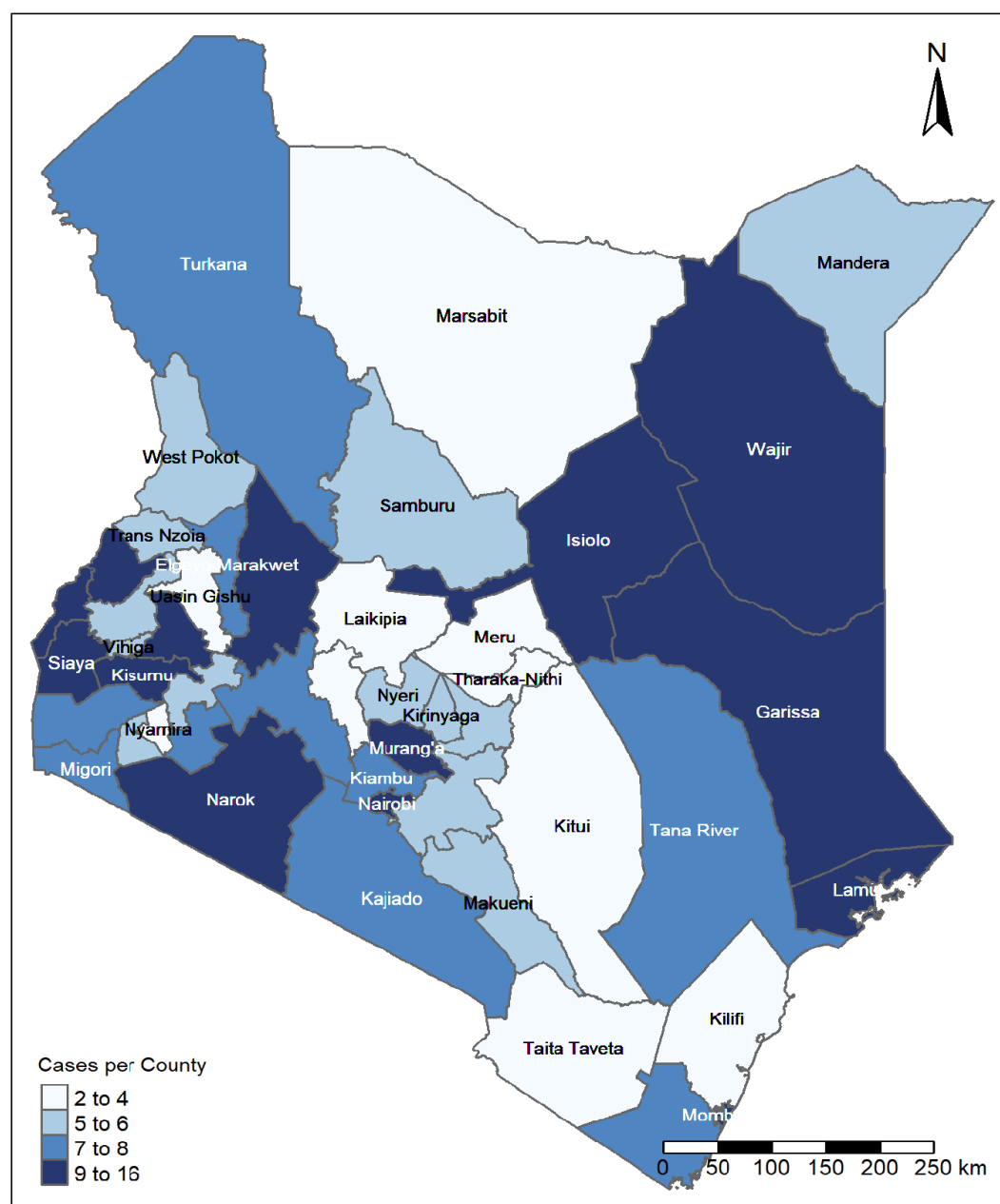


Figure 1. Cases of neonatal mortality across counties in Kenya.

The spatial distribution revealed that counties in the northeastern and coastal regions, including Wajir, Garissa, Mandera, and Lamu, exhibit the highest burden of neonatal mortality. Conversely, counties in the central and western regions, such as Nyeri, Kirinyaga, and some parts of Rift Valley, display relatively lower neonatal mortality. In addition, some counties in the western and lake regions, including Kisumu and Siaya, as well as parts of the Rift Valley, such as Turkana and West Pokot, fall within the moderate mortality range. Overall, the map highlights significant regional disparities in neonatal mortality.

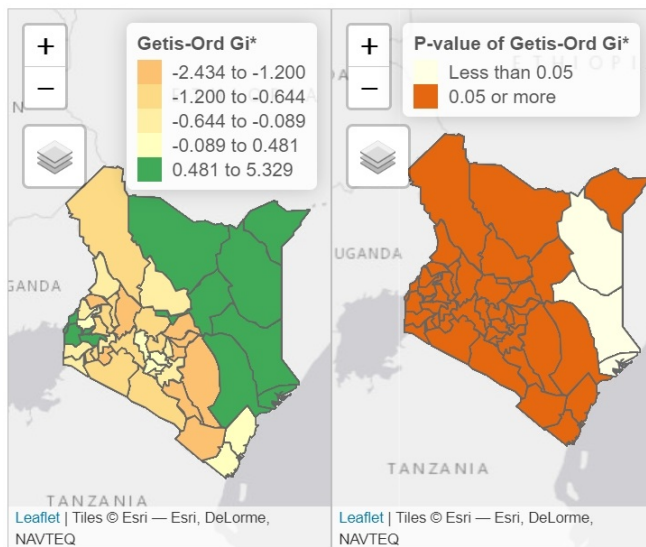
### 3.2. Hot Spot Analysis

Significant spatial autocorrelation in neonatal mortality rates was observed when Moran's I statistic was used to test the null hypothesis of spatial randomness. The computed Moran's I value was 0.012, with a variance of 4.969e-06, as indicated in Table 1. The null hypothesis is strongly refuted by the resulting Z-score of 5.4581 and p-value of 2.406e-08, which show that mortality rates are not distributed randomly. This low p-value ( $p < 0.05$ ) demonstrates a statistically significant positive spatial autocorrelation, indicating that both low and high mortality rates have a tendency to cluster together.

**Table 1.** Moran's I test for spatial Autocorrelation.

Statistic	Value
Moran's I	0.4012
Variance	4.969e-06
Z-score	5.4581
p-value	2.406e-08

The Getis-Ord  $G_i^*$  statistic was then utilized to unveil the clusters with the highest neonatal mortality rates (hotspots) and those with the lowest neonatal mortality rates (coldspots). Figure 2 presents a spatial visualization of the hotspots and coldspots of neonatal mortality in Kenya. The areas with the green shades on the left map indicate the hotspots ( $G_i^* > 0$ ), whereas the yellow shades indicate the coldspots ( $G_i^* < 0$ ). The regions with  $G_i^* \approx 0$  represent the regions where there was no significant spatial pattern of high or low values (no spatial clustering detected). The light-colored areas ( $p$ -value  $< 0.05$ ) on the right map of Figure 2 represent the statistically significant hotspots for neonatal mortality in Kenya. From this analysis, we can conclude that Wajir, Garissa, and Lamu counties have been identified as statistically significant hotspots for neonatal mortality in Kenya.

**Figure 2.** Spatial distribution of neonatal mortality hotspots and coldspots in Kenya.

### 3.3. Fitted Models

A multivariate logistic regression model was employed to identify individual- and household-level determinants of neonatal mortality, while a geographically weighted regression (GWR) was used to examine spatial heterogeneity in the relationship between neonatal mortality and climatic as well as socioeconomic factors.

#### 3.3.1. Logistic Regression

Table 2 presents the results of the multivariate logistic regression analysis examining the association between various maternal, socio-demographic, and neonatal factors with

neonatal mortality. The adjusted odds ratios (aORs), along with their 95% confidence intervals (CIs) and corresponding p-values, are reported. The analysis revealed several significant risk factors associated with neonatal mortality.

Birth weight was a statistically significant predictor of neonatal mortality. Compared to neonates with normal birth weight, those who were underweight had significantly higher odds of neonatal mortality (aOR = 4.48, 95% CI: 2.05 - 9.80,  $p < 0.001$ ). Moreover, neonates who were not weighed at birth faced an even higher risk of mortality (aOR = 17.10, 95% CI: 6.80 - 42.96,  $p < 0.001$ ), suggesting that lack of birth weight documentation may be associated with increased vulnerability.

Place of delivery was a key determinant of neonatal mortality. Compared to home deliveries, neonates born in hospitals had significantly increased odds of mortality (aOR = 9.25, 95% CI: 3.31 - 25.91,  $p < 0.001$ ). Despite being vital for complicated births, hospitals might paradoxically raise the risk of neonate infections due to pathogen exposure, invasive treatments, and extended stays in a crowded setting. The highest risk was observed among neonates delivered in "Other" locations (aOR = 24.74, 95% CI: 5.15 - 118.79,  $p < 0.001$ ), highlighting the potential dangers of unsafe birth environments.

Neonates born in urban areas had significantly higher odds of mortality compared to those in rural areas (aOR = 2.68, 95% CI: 1.32 - 5.46,  $p = 0.01$ ). This suggests that despite greater access to healthcare facilities, urban neonates may experience other risk factors such as preterm births, medical complications, or disparities in healthcare quality.

The type of household water source was also significantly associated with neonatal mortality. Neonates from households with unimproved water sources had higher odds of mortality compared to those with improved sources (aOR = 2.09, 95% CI: 1.11 - 3.91,  $p = 0.02$ ). This finding highlights the need for safe water access in preventing infections and ensuring better neonatal health outcomes.

ANC visits were categorized into 3: those who attended 4 and more visits (4 ANC visits), those who attended 1 to 3 visits (1-3 visits) and those who did not attend any (No ANC visits). Although increased ANC visits were associated with reduced odds of neonatal mortality, the results were not statistically significant. Mothers who had 1-3 ANC visits had an aOR of 0.21 (95% CI: 0.03 - 1.23,  $p = 0.08$ ), while those with more than four visits had an aOR of 0.25 (95% CI: 0.04 - 1.44,  $p = 0.12$ ). Despite the lack of significance, the trend suggests that better prenatal care may contribute to improved neonatal survival.

Birth order did not show a statistically significant effect on neonatal mortality. Compared to first-born children, the odds were slightly elevated for neonates who were the second or third child (aOR = 1.96, 95% CI: 0.89 - 4.31,  $p = 0.10$ ) and for those who were the fourth or later child (aOR = 1.88, 95% CI: 0.63 - 5.63,  $p = 0.26$ ).

There was no clear pattern between wealth and neonatal mortality. Compared to the poorest households, neonates from poorer (aOR = 0.94, 95% CI: 0.37 - 2.37,  $p = 0.90$ ), richer (aOR = 1.10, 95% CI: 0.47 - 2.59,  $p = 0.83$ ), and richest (aOR

= 1.46, 95% CI: 0.56 - 3.79,  $p = 0.44$ ) households did not have significantly different odds of mortality. The odds of neonatal mortality did not differ significantly between male and female neonates (aOR = 0.95, 95% CI: 0.57 - 1.56,  $p = 0.83$ ).

There was no significant relationship between maternal age and neonatal mortality. Compared to mothers aged below 20 years, older mothers had lower odds of neonatal mortality, but none of these associations reached statistical significance. For instance, mothers aged 30-34 years had an aOR of 0.31 (95% CI: 0.09 - 1.09,  $p = 0.07$ ), while those aged 20-24 years had an aOR of 0.44 (95% CI: 0.17 - 1.12,  $p = 0.09$ ).

Higher maternal education levels were not significantly associated with neonatal mortality. Compared to mothers with no formal education, those with primary education had an aOR of 1.76 (95% CI: 0.69 - 4.47,  $p = 0.24$ ), while those with secondary and higher education had aORs of 2.26 (95% CI: 0.78 - 6.59,  $p = 0.13$ ) and 1.47 (95% CI: 0.39 - 5.54,  $p = 0.57$ ), respectively.

There was no significant association between marital status and neonatal mortality. Widowed or divorced mothers had similar odds of neonatal mortality compared to those who were married and living together (aOR = 1.11, 95% CI: 0.25 - 4.90,  $p = 0.89$ ).

Religious affiliation did not significantly influence neonatal mortality risk. Compared to other religions, Protestants had an aOR of 0.74 (95% CI: 0.39 - 1.41,  $p = 0.35$ ), while Muslims had an aOR of 0.99 (95% CI: 0.41 - 2.39,  $p = 0.99$ ). Receiving three or more tetanus injections before birth was not significantly associated with neonatal mortality risk (aOR = 0.94, 95% CI: 0.36 - 2.47,  $p = 0.90$ ).

The findings highlight several key risk factors for neonatal mortality, with birth weight, place of delivery, residence, and water source showing statistically significant associations. Neonates who were underweight, not weighed at birth, or delivered in hospitals or unsafe locations faced significantly higher mortality risks. Additionally, urban residence and unimproved water sources were associated with increased odds of neonatal mortality.

While factors such as ANC visits, birth order, maternal age, and education showed some associations, they were not statistically significant. These results emphasize the need for targeted interventions to improve neonatal outcomes, particularly through enhanced prenatal care, safe delivery practices, and improved access to clean water. Nonetheless, the multivariate logistic regression approach resulted in wide confidence intervals, suggesting the need for further research to enhance the precision of the findings.

### 3.3.2. Geographically Weighted Logistic Regression

The study employed GWR to investigate how the effect of climatic and socioeconomic on neonatal mortality varies

across regions. Figure 3 presents the spatial heterogeneity in the effect of climatic variables on neonatal mortality.

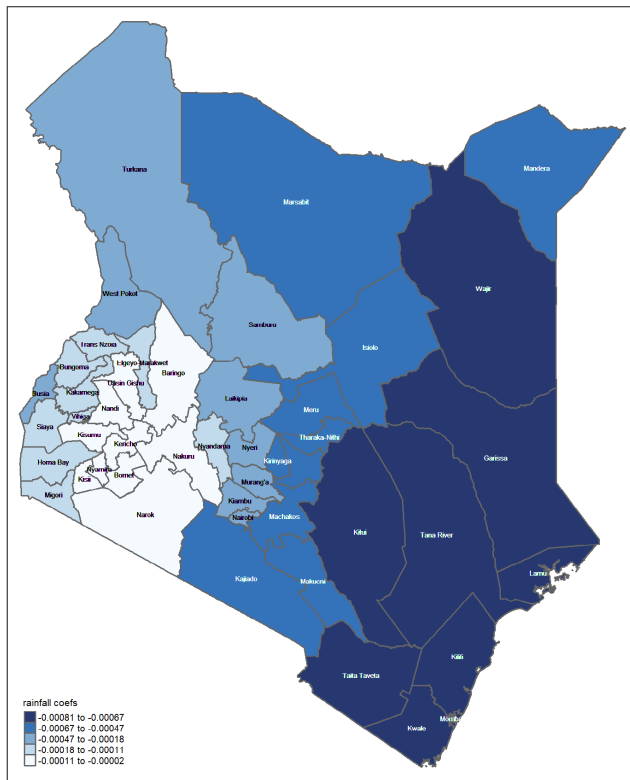
The results indicate a negative relationship between rainfall and neonatal mortality across most counties. This implies that higher rainfall levels are generally associated with lower neonatal mortality rates. However, the magnitude and significance of this effect vary across different regions, as shown in the spatial distribution of coefficients in figure 3 top left. This suggests that in these regions, increased rainfall is significantly linked to lower neonatal mortality rates. Counties such as Nairobi, Kisumu, and parts of Narok exhibit relatively weak or near-zero relationships (lighter shades on the map).

The results indicate a positive association between mean annual temperature and neonatal mortality across all counties. This suggests that higher temperatures are generally linked to increased neonatal mortality, although the strength of this relationship varies by region. Counties such as Mandera, Wajir, Garissa, Tana River, and Lamu exhibit the highest temperature coefficients (darker shades of blue). This indicates that in these regions, higher temperatures are significantly associated with increased neonatal mortality, likely due to heat stress, water scarcity, and increased prevalence of temperature-sensitive diseases such as dehydration and infections. Counties such as Nairobi, Kisumu, and parts of Narok exhibit the lowest temperature coefficients (lighter shades on the map).

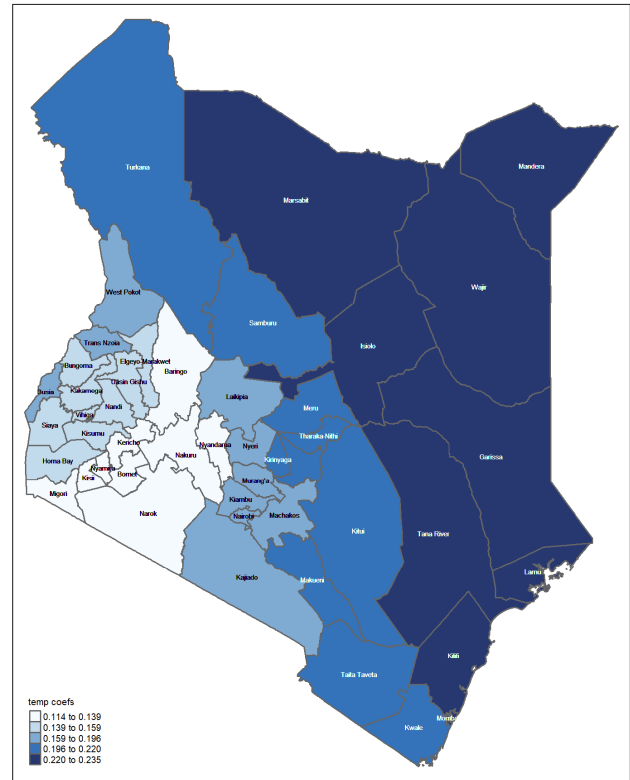
Diurnal temperature variation refers to the difference between daytime and nighttime temperatures, which can have significant effects on neonatal health, particularly in regions with extreme temperature fluctuations. The results indicate a positive association between diurnal temperature variation and neonatal mortality across all counties, meaning that larger differences between day and night temperatures are linked to higher neonatal mortality rates. However, the magnitude of this relationship varies geographically, as shown in figure 3 bottom left. Counties such as Mandera, Wajir, Garissa, Lamu, and Tana River exhibit the highest diurnal temperature coefficients (darker shades of blue). This suggests that in these regions, greater temperature fluctuations between day and night are strongly associated with higher neonatal mortality rates. The increased risk could be due to thermal stress, as newborns are highly sensitive to temperature changes and may experience hypothermia at night or heat stress during the day.

The findings indicate a positive association between mean annual aridity and neonatal mortality rate i.e higher levels of aridity are generally associated with increased neonatal mortality. The strongest positive associations are observed in the arid and semi-arid regions, particularly Mandera, Wajir, Garissa, and Tana River counties, where neonatal mortality rates tend to be higher. In contrast, regions with lower aridity, such as parts of the western highlands, including West Pokot and Trans Nzoia, exhibit weaker associations.

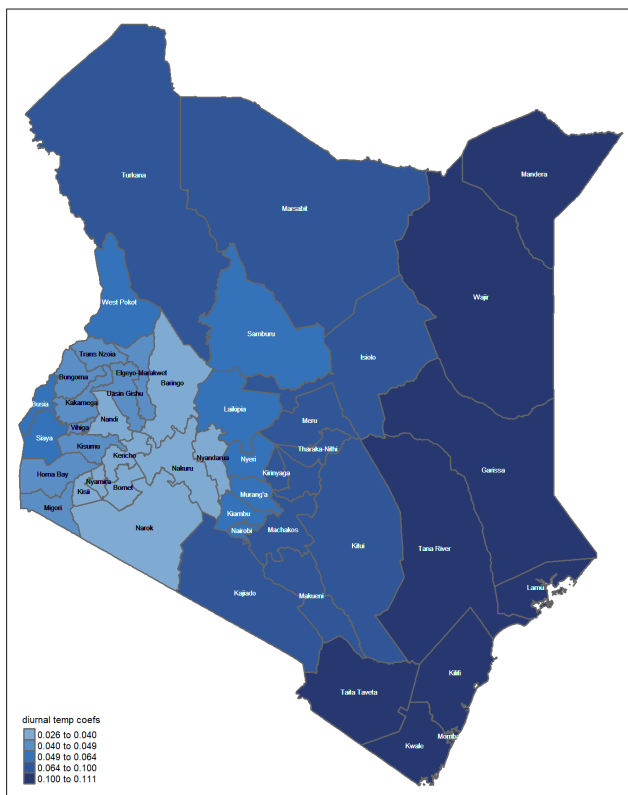




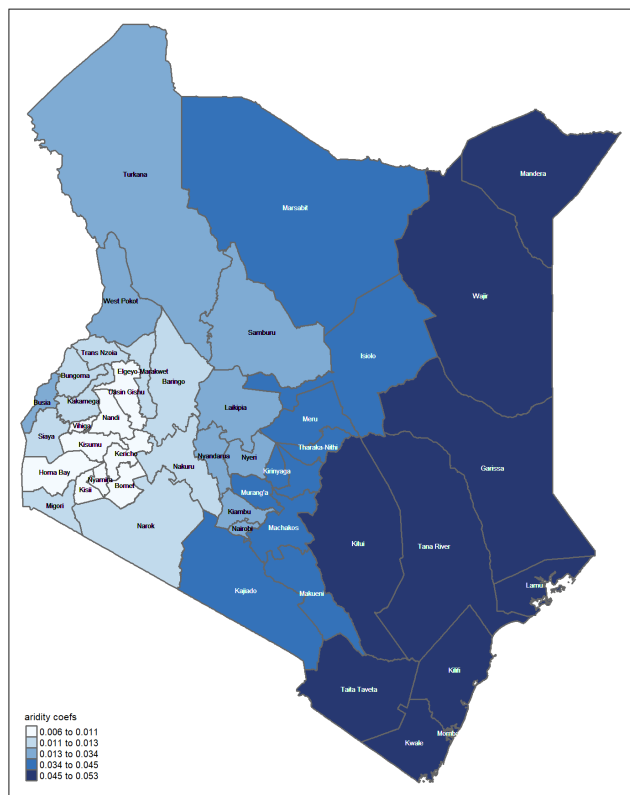
(a) Mean Annual Rainfall



(b) Mean Annual Temperature



(c) Mean Diurnal Temperature



(d) Aridity

Figure 3. Spatial distribution of GWR local estimates for climatic factors.



**Table 2.** Multivariate logistic regression results of neonatal mortality risk factors.

Variable	Adjusted Odds Ratio (aOR)	95% CI	P-value
(Intercept)	0.00	(0.00 - 0.04)	0.00
ANC Visits			
No ANC Visits (RC)	1.00	-	-
1-3 visits	0.21	(0.03 - 1.23)	0.08
4 visits	0.25	(0.04 - 1.44)	0.12
Residence			
Rural (RC)	1.00	-	-
Urban	2.68	(1.32 - 5.46)	0.01
Water Source			
Improved (RC)	1.00	-	-
Unimproved	2.09	(1.11 - 3.91)	0.02
Not a dejuire	2.60	(0.87 - 7.73)	0.09
Wealth Index of households			
Poorest (RC)	1.00	-	-
Poorer	0.94	(0.37 - 2.37)	0.90
Middle	-	-	-
Richer	1.10	(0.47 - 2.59)	0.83
Richest	1.46	(0.56 - 3.79)	0.44
Birth Order			
4th child (RC)	1.00	-	-
2nd to 3rd	1.96	(0.89 - 4.31)	0.10
3rd to 4th	1.87	(0.70 - 5.00)	0.21
Birth Weight Status			
Normal (RC)	1.00	-	-
Don't know	2.78	(0.55 - 13.96)	0.21
Not weighed	17.10	(6.80 - 42.96)	0.00
Underweight	4.48	(2.05 - 9.80)	0.00
Sex of Child			
Male (RC)	1.00	-	-
Female	0.95	(0.57 - 1.56)	0.83
Delivery Place			
Home (RC)	1.00	-	-
Hospital	9.25	(3.31 - 25.91)	0.00
Other	24.74	(5.15 - 118.79)	0.00
Marital Status			
Married/living together (RC)	1.00	-	-
Widowed/Divorced	1.11	(0.25 - 4.90)	0.89
Not living together	0.79	(0.23 - 2.73)	0.71
Religion			
Protestant (RC)	1.00	-	-
Other	1.00	(0.27 - 3.67)	1.00
Islam	0.99	(0.41 - 2.39)	0.99
Highest Education			
No Education (RC)	1.00	-	-
Primary	1.76	(0.69 - 4.47)	0.24
Secondary	2.26	(0.78 - 6.59)	0.13
Higher	1.47	(0.39 - 5.54)	0.57
Mother's Age			
< 20 (RC)	1.00	-	-
20-24	0.44	(0.17 - 1.12)	0.09
25-29	0.42	(0.14 - 1.23)	0.11
30-34	0.31	(0.09 - 1.09)	0.07
35-39	0.82	(0.23 - 2.96)	0.77
40-49	0.94	(0.20 - 4.42)	0.94

Variable	Adjusted Odds Ratio (aOR)	95% CI	P-value
Tetanus Injections Before Birth			
< 3 (RC)	1.00	-	-
3+	0.94	(0.36 - 2.47)	0.90

### 3.4. Model Diagnostics

The multivariate logistic regression model's goodness-of-fit was evaluated using the Hosmer-Lemeshow test. With eight degrees of freedom, the chi-squared statistic that resulted was 13.787, resulting in a p-value of 0.0875. Since the p-value is greater than 0.05, it suggests that the model fits the data adequately, with no significant difference between the observed and predicted outcomes. This indicates that the multivariate logistic regression model is a good fit for the data. The *pseudo* -  $R^2$  statistic was used to evaluate the Geographically Weighted Logistic Regression (GWLR) model's goodness-of-fit. The model produced a *pseudo* -  $R^2$  value of 0.65, meaning that the spatially varied coefficients accounted for around 65% of the variance in the outcome variable.

## 4. Conclusion and Recommendations

The fitted multivariate logistic regression revealed that urban residence, underweight status, unimproved water sources, and non-hospital deliveries (especially in non-standard locations) are the key significant risk factors of neonatal mortality. Counties in north eastern and coastal Kenya including Wajir, Garissa, and Lamu are the hotspots for neonatal mortality. GWLR revealed that climatic factors, such as temperature and aridity, impact neonatal mortality differently across regions in Kenya. Generally, higher temperatures are a significant risk factor for neonatal mortality, particularly in arid counties like Mandera, Wajir, Garissa, Tana River, and Lamu.

The government should enforce sanitation laws and increase access to clean drinking water in order to improve water safety. Enhancing antenatal care, enhancing hospital safety, and offering alternative birthing sites are more ways that healthcare stakeholders can improve maternal and neonatal care. Implementing digital tracking, educating healthcare professionals, and providing assistance to low birth weight babies are essential to ensuring accurate birth weight records. The government should prioritize integrated initiatives in hotspots, including as desert regions like Mandera, Wajir, Garissa, and Tana River, to target vulnerable areas and address climate and healthcare concerns. To assess the long-term impact of interventions, researchers should conduct longitudinal studies to track neonatal mortality trends over time. Finally, scientists should expand spatial analysis to explore the spatial variation of other relevant risk factors beyond climatic factors that influence neonatal deaths, potentially using GWLR.

## ORCID

0009-0003-7573-675X (Getrude Moraa Nyabuto)

## Abbreviations

GWLR	Geographically Weighted Logistic Regression
ANC	Antenatal Care
aOR	Adjusted Odds Ratio
OLS	Ordinary Least Squares
LR	Logistic Regression
UNICEF	United Nations International Children's Emergency Fund
OR	Odds Ratio
SSA	Sub-Saharan Africa
PDE	Partial Differential Equation
KNBS	Kenya National Bureau of Statistics
CI	Confidence Interval

## Acknowledgments

This work would not have been possible without the invaluable mentorship of my supervisors. Their expertise and insightful feedback were instrumental in sharpening the research question and refining the methodological approach. I am further grateful for the unwavering support network that sustained me throughout this endeavor. The encouragement of my family and the collaborative spirit of my friends and colleagues provided a constant source of motivation. Their belief in my potential has been a driving force in my academic journey, and I am deeply appreciative for the positive impact they have had on both my intellectual development and personal growth.

## Ethics Clearance

**Getrude Moraa Nyabuto:** Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing - original draft, Writing - review & editing

## Ethics Clearance

This study utilized publicly available secondary data from the Kenya Demographic and Health Survey (KDHS). The KDHS surveys are conducted with ethical approval obtained

by ICF International and the relevant ethical review boards in Kenya [e.g., the National Commission for Science, Technology and Innovation (NACOSTI)]. The DHS Program ensures the anonymity and confidentiality of respondents. Therefore, no further ethical approval was required for this secondary analysis. Permission to use the data was granted by the DHS Program upon approval of the data request.

## Data Availability

The data used in this study is from the 2022 Kenya Demographic and Health Survey (KDHS). It is publicly available upon request from the Demographic and Health Surveys (DHS) Program. Researchers can access the data after registering and submitting a data request through the DHS Program website (<https://dhsprogram.com/data/available-datasets.cfm>). Climate data for the study came from publicly accessible gridded datasets and was spatially connected to DHS survey clusters using geographic coordinates.

## Conflicts of Interest

The authors declare that there are no conflicts of interest related to this study.

## References

- [1] Imbo, A. E., Mbuthia, E. K., & Ngotho, D. N. (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*, 10(2), 287. <https://doi.org/10.21106/ijma.508>
- [2] Ogallo, W., Speakman, S., Akinwande, V., Varshney, K. R., Walcott-Bryant, A., Wayua, C., ... & Orobato, N. (2021, January). Identifying factors associated with neonatal mortality in Sub-Saharan Africa using machine learning. In *AMIA Annual Symposium Proceedings* (Vol. 2020, p. 963).
- [3] Elkasabi, M., Ren, R., & Pullum, T. W. (2020). Multilevel modeling using DHS surveys: a framework to approximate level-weights. *ICF*.
- [4] Abid, R. I. (2024). Identifying Spatial Patterns of Road Accidents in Madaba City by Applying Getis-Ord Gi\* Spatial Statistic. *International Journal of Engineering and Advanced Technology*, 13(4), 1-8. <https://doi.org/10.35940/ijeat.C4387.13040424>
- [5] UNICEF. (2017). Maternal and Newborn Health Disparities, Kenya. UNICEF Data.
- [6] Ahmed, K. Y., Thapa, S., Hassen, T. A., Tegegne, T. K., Dadi, A. F., Odo, D. B., ... & Ross, A. G. (2024). Population modifiable risk factors associated with neonatal mortality in 35 sub-Saharan Africa countries: analysis of data from demographic and health surveys. *EClinicalMedicine*, 73. <https://doi.org/10.1016/j.eclinm.2024.102682>
- [7] Appiah, E. K., Aidoo, E. N., Avuglah, R. K., & Appiah, S. K. (2024). Geographically weighted logistic regression model for identifying risk factors for malaria infection among under-5 children in Ghana. *Scientific African*, 26, e02398. <https://doi.org/10.1016/j.sciaf.2024.e02398>
- [8] Dare, S., Oduro, A. R., Owusu-Agyei, S., Mackay, D. F., Gruer, L., Manyeh, A. K., ... & Pell, J. P. (2021). Neonatal mortality rates, characteristics, and risk factors for neonatal deaths in Ghana: analyses of data from two health and demographic surveillance systems. *Global health action*, 14(1), 1938871. <https://doi.org/10.1080/16549716.2021.1938871>
- [9] Kibret, G. D., Demant, D., & Hayen, A. (2022). Bayesian spatial analysis of factors influencing neonatal mortality and its geographic variation in Ethiopia. *PloS one*, 17(7), e0270879. <https://doi.org/10.1371/journal.pone.0270879>
- [10] Mollalo, A., Vahedi, B., Bhattarai, S., Hopkins, L. C., Banik, S., & Vahedi, B. (2020). Predicting the hotspots of age-adjusted mortality rates of lower respiratory infection across the continental United States: Integration of GIS, spatial statistics and machine learning algorithms. *International Journal of Medical Informatics*, 142, 104248. <https://doi.org/10.1016/j.ijmedinf.2020.104248>
- [11] Mutwiri, R. M. (2010). Spatial analysis and mapping of infant mortality in Kenya on the basis of demographic and health survey data (Doctoral dissertation, University of NAIROBI).
- [12] Novianti, P., & Rosadi, D. (2023). Geographically Weighted Logistic Regression Model on Binomial Data to Explore Weather Spatial Non-Stationarity in Covid-19 Cases. *Engineering Letters*, 31(3).
- [13] Habtamu Kebebe Kasaye, H. K. K., Firew Tekle Bobo, F. T. B., Mekdes Tigistu Yilma, M. T. Y., & Mirkuzie Woldie, M. W. (2019). Poor nutrition for under-five children from poor households in Ethiopia: evidence from 2016 Demographic and Health Survey. <https://doi.org/10.1371/journal.pone.0225996>
- [14] Ogbo, F. A., Page, A., Idoko, J., & Agho, K. E. (2018). Population attributable risk of key modifiable risk factors associated with non-exclusive breastfeeding in Nigeria. *BMC public health*, 18, 1-9.
- [15] Ouma, P. O., Malla, L., Wachira, B. W., Kiarie, H., Mumo, J., Snow, R. W., ... & Okiro, E. A. (2022). Geospatial mapping of timely access to inpatient neonatal care and its relationship to neonatal mortality in Kenya. *PLOS global public health*, 2(6), e0000216. <https://doi.org/10.1371/journal.pgph.0000216>

- [16] Gruebner, O., Khan, M. M. H., Burkart, K., Lautenbach, S., Lakes, T., Krämer, A., ... & Galea, S. (2017). Spatial variations and determinants of infant and under-five mortality in Bangladesh. *Health & place*, 47, 156-164. <https://doi.org/10.1016/j.healthplace.2017.08.012>
- [17] 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. <https://doi.org/10.3390/ijerph19010399>