

Article

Investigating Spatial Effects through Machine Learning and Leveraging Explainable AI for Child Malnutrition in Pakistan

Xiaoyi Zhang ¹, Muhammad Usman ², Ateeq ur Rehman Irshad ³, Mudassar Rashid ^{4,*} and Amira Khattak ⁵

¹ College of Liberal Arts and Science, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA; xiaoyi6@illinois.edu

² Faculty of Economic Sciences, University of Warsaw, 00-927 Warszawa, Poland; m.usman@uw.edu.pl

³ Deanship of Educational Services, Prince Sultan University, Riyadh 11586, Saudi Arabia; airshad@psu.edu.sa

⁴ Department of Economics, COMSATS University Islamabad, Islamabad 45550, Pakistan

⁵ Department of Marketing, College of Business Administration, Prince Sultan University, Riyadh 11586, Saudi Arabia; akhattak@psu.edu.sa

* Correspondence: mudassar.rashid@comsats.edu.pk

Abstract: While socioeconomic gradients in regional health inequalities are firmly established, the synergistic interactions between socioeconomic deprivation and climate vulnerability within convenient proximity and neighbourhood locations with health disparities remain poorly explored and thus require deep understanding within a regional context. Furthermore, disregarding the importance of spatial spillover effects and nonlinear effects of covariates on childhood stunting are inevitable in dealing with an enduring issue of regional health inequalities. The present study aims to investigate the spatial inequalities in childhood stunting at the district level in Pakistan and validate the importance of spatial lag in predicting childhood stunting. Furthermore, it examines the presence of any nonlinear relationships among the selected independent features with childhood stunting. The study utilized data related to socioeconomic features from MICS 2017–2018 and climatic data from Integrated Contextual Analysis. A multi-model approach was employed to address the research questions, which included Ordinary Least Squares Regression (OLS), various Spatial Models, Machine Learning Algorithms and Explainable Artificial Intelligence methods. Firstly, OLS was used to analyse and test the linear relationships among selected variables. Secondly, Spatial Durbin Error Model (SDEM) was used to detect and capture the impact of spatial spillover on childhood stunting. Third, XGBoost and Random Forest machine learning algorithms were employed to examine and validate the importance of the spatial lag component. Finally, EXAI methods such as SHapley were utilized to identify potential nonlinear relationships. The study found a clear pattern of spatial clustering and geographical disparities in childhood stunting, with multidimensional poverty, high climate vulnerability and early marriage worsening childhood stunting. In contrast, low climate vulnerability, high exposure to mass media and high women's literacy were found to reduce childhood stunting. The use of machine learning algorithms, specifically XGBoost and Random Forest, highlighted the significant role played by the average value in the neighbourhood in predicting childhood stunting in nearby districts, confirming that the spatial spillover effect is not bounded by geographical boundaries. Furthermore, EXAI methods such as partial dependency plot reveal the existence of a nonlinear relationship between multidimensional poverty and childhood stunting. The study's findings provide valuable insights into the spatial distribution of childhood stunting in Pakistan, emphasizing the importance of considering spatial effects in predicting childhood stunting. Individual and household-level factors such as exposure to mass media and women's literacy have shown positive implications for childhood stunting. It further provides a justification for the usage of EXAI methods to draw better insights and propose customised intervention policies accordingly.

Keywords: regional health inequalities; childhood stunting; climate vulnerability; spatial hotspots; machine learning; explainable artificial intelligence



Citation: Zhang, X.; Usman, M.; Irshad, A.u.R.; Rashid, M.; Khattak, A. Investigating Spatial Effects through Machine Learning and Leveraging Explainable AI for Child Malnutrition in Pakistan. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 330. <https://doi.org/10.3390/ijgi13090330>

Academic Editors: Lan Mu, Jue Yang and Wolfgang Kainz

Received: 2 June 2024

Revised: 6 September 2024

Accepted: 13 September 2024

Published: 16 September 2024



Copyright: © 2024 by the authors. Published by MDPI on behalf of the International Society for Photogrammetry and Remote Sensing. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In 2015, the United Nations declared the Sustainable Development Goals (Target 2.2), with an objective of “ending all forms of malnutrition by 2030”. The first 1000 days of a child’s life are considered crucial for mental and physical development. Good early days are proven to be key catalysts for the overall growth of a child [1]. Therefore, a gritty and realistic focus on reducing childhood stunting is required to comprehend the most critical period of children’s development.

Globally, in 2019, 144 million children under the age of five years had stunted growth. Although there has been a consistent decline in the growth of stunting since 2000 (199.5 million were stunted), the progress seems insufficient to meet the 40% reduction target by 2025 set by World Health Assembly [2].

Africa and Asia are burdened with the highest number of stunted children with 54% and 40%, respectively. Specifically, in South Asia, Pakistan bears the third-highest childhood stunting rate followed by India. Childhood stunting has been a persistent threat to child growth in Pakistan. Historic data on child malnutrition reveals that childhood stunting has remained alarmingly high with minimal improvement over time. In the year 2011, the stunting rate was 44% and, despite some efforts, it only decreased slightly to 38% by 2018. Recent estimates on childhood stunting present a rather dismal picture as it increased to 40%. These historical data suggest that the country has been grappling with this enduring issue for an extended period of time [3].

Socioeconomic and environmental challenges such as high poverty, widespread food insecurity, poor quality of water and inadequate sanitation facilities and stagnating agricultural productivity are the major contributors to poor health outcomes and deteriorating human capital [4]. In Pakistan, 30% of households lack adequate sanitation facilities and more than approximately 2.1 million individuals are deprived of safe drinking water. It has been firmly established that if the current conditions such as high poverty are not improved, the lack of basic sanitation and safe drinking water and prevalence of diseases would further worsen in the coming years [5].

Several studies have elucidated that illiteracy, poor sanitation facilities, lack of improved drinking water, poverty and bigger household size may in part contribute to childhood stunting. Existing literature on Lower Middle Income Countries (LMICs) has empathized household and individual factors as the significant contributors to childhood stunting. In [6] it was concluded that breast feeding practices, household wealth status, maternal nutrition and child food intake are crucial factors in determining childhood stunting. Similarly, using a mutually adjusted logistic regression, [7] determined that in children who were aged between 6 and 8 months, limited intake of complementary foods and short maternal stature were significantly linked with higher risk of childhood stunting. For children who were aged between 6 and 23 months, household wealth, maternal height, maternal education and age at marriage were found to be significant correlates of childhood stunting. The authors of [8] used the two recent most Indonesian Family Life Survey (IFLS) to examine the variations in prevalence of childhood stunting between 2007 and 2014. They revealed that household wealth status, mothers’ education, place of delivery and the availability of sanitation facilities contribute most to childhood stunting. Furthermore, they found that improvement in health care access significantly contributed to reducing the stunting gap between richer and poorer children. Additionally, [9] asserted that children from poor households with unavailability of sanitation facilities and lack of safe drinking water were at more risk of having stunted growth.

Maternal stature and education play a crucial role in child’s development. The authors of [10] used the fourth Indian National Family Health Survey to identify risk factors for chronic undernutrition in India. They calculated odds ratios using logistic regression. Based on their findings, they determined the five most vital factors for childhood stunting. They argued that illiterate mothers, poor households, short maternal stature, poor dietary intake and low maternal weight are the main contributing factors. These factors had a population-attributable risk of 67.2% for stunting. Similarly, using the latest available

Indian National Family Health Survey (2015–2016), [11] analysed the relative importance of 23 correlates of stunting. They estimated logistic regression models to evaluate the importance of selected predictors. Based on their results, they identified short maternal stature as the most important contributing factor to stunting. Other key factors were poor household status, poor air quality of households, lack of maternal education and low maternal BMI. Additionally, they found weaker associations for dietary intake, breastfeeding and parental factors.

Maternal caretaker behaviour and early female child marriage have been identified as risk factors for child development. The authors of [12] inspected the intergenerational effects of early child marriage on stunting and child development in Sub Saharan Africa. They concluded that the odds of being stunted are 29% higher for children born to women who married before the age 18 in comparison to those women who married later in their lives. They further examined the mechanisms through which risk of stunting increases. They added that early child marriage is not the sole path to stunting. However, contextual factors such as maternal education and household wealth matter for child development and health.

In recent years, researchers have assessed the role of environmental covariates such as rainfall, temperature, droughts and floods, etc., on stunting. A recent study by [13] revealed that along with socioeconomic features, high climate vulnerability exacerbates childhood stunting in comparison to low climate vulnerability. In sub-Saharan Africa the odds ratio for child health outcomes were higher in the presence of droughts and political instability [14]. Similarly, using multivariate regression models, authors assessed the relationship between exposure to climate variations (rainfall and temperature) in utero and early child development. They suggested that higher rainfall during early life bears a negative relationship with stunting. However, higher prenatal temperatures have negative repercussions during the early life of children and consequently lead to severe stunting. Additionally, they identified potential pathways such as heat stress, infectious diseases caused by climatic variations and more time devoted to work during pregnancies in dry seasons through which the odds of stunting increase [15]. In Indonesia, a similar study was conducted to detect the association between climate exposures and child undernutrition. They argued that delays in monsoon rainfall during the prenatal period had an adverse effect on child height during the early 2–4 years [16].

Lately, there has been growing recognition of integrating spatial data and geographic factors to uncover the multifaceted nature of childhood stunting. Numerous studies have illustrated that childhood stunting varies across regions, a term often used to describe this phenomenon being geographical heterogeneity. For instance, spatial hotspots were identified with high stunting rates amongst districts in the Upper West, Upper East and Western North regions in Ghana. These districts were characterised by low access to basic health services, high poverty levels and low educational attainment. On the other hand, clusters with low childhood stunting were observed in the Greater Accra, Volta and Bono regions [17]. Similarly, a study conducted in India identified 146 hot spots as high prevalence and 130 cold spots as low prevalence districts. Districts with high stunting rates were mostly concentrated in the states of Bihar, Jharkhand and Uttar Pradesh, while low prevalence districts were largely observed in the states of Himachal Pradesh, Kerala, Tamil Nadu and West Bengal. Factors such as large household size, open defecation, female illiteracy, household poverty and unavailability of improved drinking water were associated with higher stunting rates [18]. In Ethiopia, a similar study was conducted which revealed a spatial dependency in childhood stunting. Districts with the highest prevalence of childhood stunting were found in Northern Ethiopia. By leveraging logistic regression, the authors concluded that factors such as low level of female education, poor household wealth and older age were exacerbating stunting in the children under 5 years of age [19]. Similarly, in another study, spatial heterogeneity was observed in Ethiopia across locations suggesting a variation in childhood stunting. The estimates of logistic regression indicated that children who had anaemic history, malnourished mothers, low educational

attainment and mothers with multiple births were most likely to be severely stunted [20]. A secondary spatial analysis of the RDHS 2014–15 cross sectional data revealed that low maternal education, non-optimal feeding practices, lack of improved water and belonging to poor households were significant contributors to childhood stunting [21].

It is evident from the existing literature that researchers have focused mainly either on household- or on individual-level factors. Furthermore, from the methodological perspective, there exists a vast amount of literature which has predominantly employed ordinary least squares or logistic regression. These methods are very well known for testing linear relationships [22]. Spatial studies are limited to the spatial distribution of child malnutrition rates. Rare usages of spatial models, the possible significant impact of spatial lag component on nearby districts, and integrated spatial machine learning approaches in the area of regional health inequalities have not yet been thoroughly explored. Additionally, as childhood stunting is influenced by a complex interaction of multiple factors, traditional linear models such as OLS may not adequately capture potential nonlinear relationships. In this regard, employing Explainable Artificial Intelligence (EXAI) methods such as LIME and SHAP could help us identify such intricate relationships, which are not identifiable with conventional approaches. Furthermore, to best of our knowledge, no study has been conducted to examine spatial correlates of childhood stunting for all the districts of Pakistan. Therefore, this study attempts to quantify the possible spatial correlates of childhood stunting for all the regions of Pakistan. This study is more comprehensive and inclusive. Moreover, the primary focus of this research is to provide a novel methodological perspective and implementation of spatial machine learning and EXAI methods for regional health economics in a spatial context.

Specifically, this research aims to address these aforementioned research gaps through (1) investigating the spatial distribution of childhood stunting in Pakistan, (2) spatial examination of the relationships between socioeconomic, climatic features and childhood stunting, (3) identification and validation of spatial lag using spatial machine learning algorithms and (4) examination of nonlinear relationships using EXAI techniques.

The present study brings novelty to the recognition and validation of the significance of spatial lag components (i.e., average value in the neighbourhood) and proposes a methodological approach using spatial machine learning. The integrated methodology provides additional insights to researchers and policymakers compared to conventional statistical methods, which typically assume that effects remain stationary. Additionally, the inclusion of climatic covariates along with other socioeconomic features could help address the issue of sparse determinants of stunting. Understanding the relationship between climate covariates and stunting would allow for the understanding of how climate change might shape childhood stunting in the future. Moreover, examination of nonlinear relationships will aid in developing targeted health intervention programs and policies for populations with significant needs.

2. Materials and Methods

This study covered all 145 districts from all provinces (Punjab, Sindh, Kpk, Gilgit-Baltistan), excluding ICT (Islamabad Capital Territory represented as white region), made available in the form of shapefiles. All districts were included in the analysis. Districts were mapped using shapefiles of Pakistan from DIV-GIS. All computations and estimations were performed in R Studio. The following Figure 1 is the official map of Pakistan portraying all the districts covered.

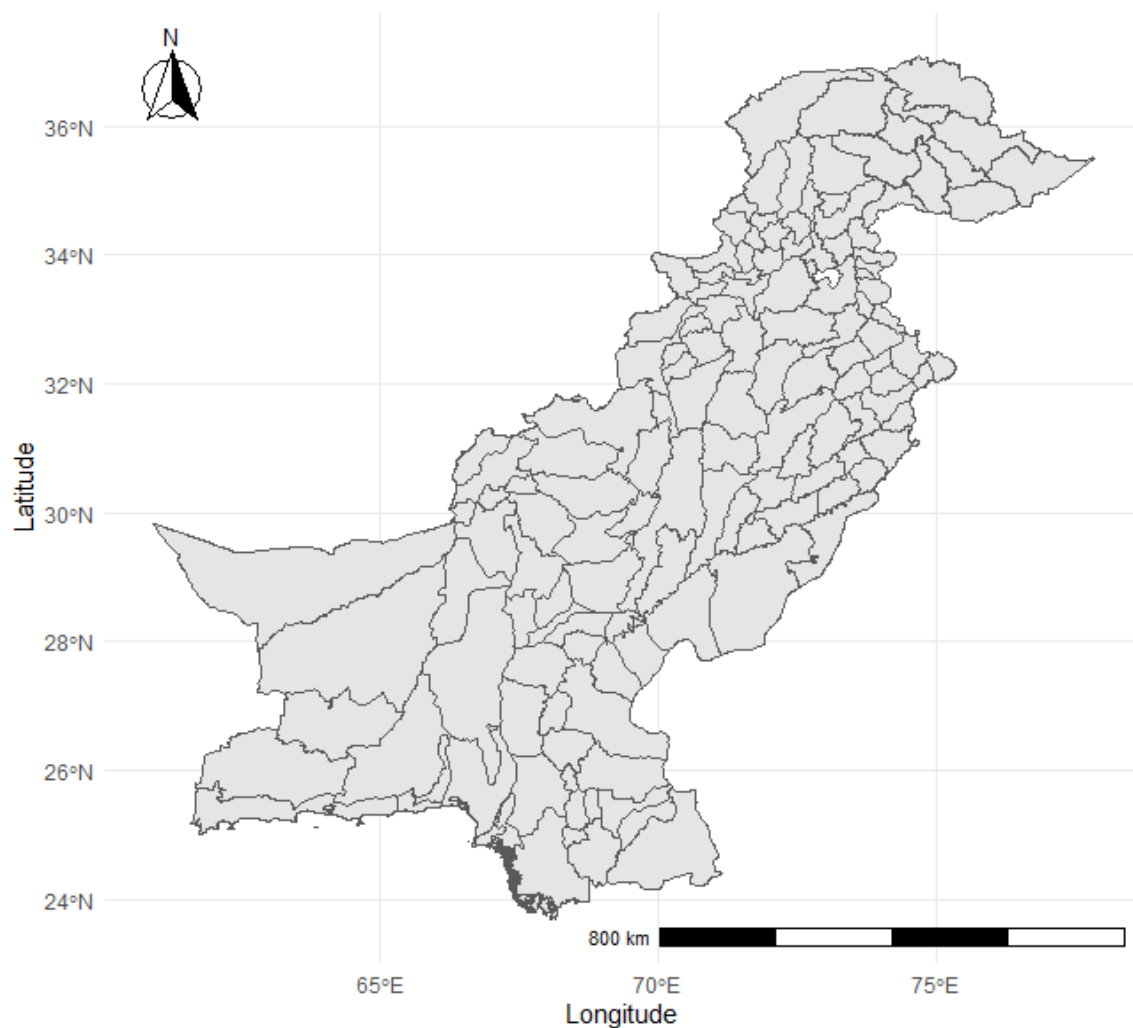


Figure 1. Regional map of Pakistan.

2.1. Sampling and Data Collection

The study uses a cross-sectional dataset from multiple sources, the Multiple Indicator Cluster Survey (MICS) and the Integrated Household Survey (ICA), which were matched accordingly at the district level. MICS (accessed on 13 December 2023) (<https://mics.unicef.org/surveys>) is an internationally recognized household survey that provides district level estimates of more than 125 key indicators related to demographic, socioeconomic, health, to assess the overall situation of women aged 14–49 years and children under five years of age. The latest available MICS survey from year 2017–2018 was used in this study. A two-stage standardized sampling approach was followed to select the survey samples from all regions. In the first step, enumeration areas were selected from the primary sampling units. In the second step, a household was selected from the chosen enumeration areas. The enumeration areas were selected with unequal probability means with probability proportional to size. The total sample was comprised of 53,480 households in 2692 cluster samples, with a response rate of approximately 98% across all regions [23].

2.2. Variables

In this study, the outcome variable was the percentage of stunted children under the age of five, defined by MICS as children whose height-for-age z-scores were below minus two standard deviations. Independent features were obtained from three different sources. Socioeconomic features, including female literacy, meaning the percentage of women who can read a short simple statement and also attended secondary school

education; premature birth rate, representing the percentage of all children who were born before 37 weeks of pregnancy; marriage before age 15, showing the percentage of women who got married before reaching the age of 15; multidimensional poverty (MDP), entailing people of all age groups living in poverty; women's exposure to mass media, meaning the percentage of women who read a newspaper or watch any sort of electronic media at least once per week; and full immunization, meaning the percentage of vaccinated children, were retrieved from MICS 2018. All the abovementioned features were aggregated as percentages. Additionally, district-level climate vulnerability was retrieved from Integrated Contextual Analysis [24]. Climate vulnerability includes two primary hazards: (1) floods and (2) droughts. Flood hazard data were retrieved from the National Disaster Management Authority (NDMA). It reflects the total number of recorded flood events from 1950–2015. Drought data, obtained from the Pakistan Meteorological Department, are based on multiple factors such as soil moisture, precipitation, dependency on seasonal rainfall and drought frequency (based on Standardized Precipitation Index). All districts were categorized into a three-point scale—low, medium and high—according to their vulnerability to floods and droughts (<https://reliefweb.int/report/pakistan/integrated-context-analysis-ica-vulnerability-food-insecurity-and-natural-hazards>, (accessed on 10 November 2023)). Climate vulnerability was classified into three levels: low, medium and high. Each district was assigned a vulnerability classification based on its exposure to climate change.

2.3. Statistical Analysis

We started with an OLS estimation. In the second step, we calculated the variance inflation factor (VIF) for all our explanatory features to assess potential multicollinearity. The VIF value for all the independent features was below 7. In the next step, we performed spatial distribution and visualization of childhood stunting. In order to test spatial autocorrelation in childhood stunting, we used Moran's statistic as given by (1):

$$I = \frac{N}{\sum_i \sum_j \omega_{ij}} \frac{\sum_i \sum_j \omega_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2} \quad (1)$$

The computation of Moran's statistic is based on the famous spatial weights matrix. ω_{ij} in the above equation signifies the spatial weights matrix. Moran's value is in the range $[-1, 1]$. A positive value reveals a positive spatial autocorrelation and vice versa. A positive value suggests that districts exhibit similar values to their neighbours. Furthermore, Moran's scatter plot was generated, which is composed of four quadrants: Hotspots (regions with high values next to each other), Cold spots (regions with low values next to each other) and Spatial outliers (districts with high and low values next to each other and vice versa) [25,26].

In the presence of spatial autocorrelation in the error term, estimates from OLS are biased and inconsistent. This allows one to employ spatial models. Therefore, after confirming that errors were not random but spatially dependent, we employed the Spatial Durbin Error Model (SDEM). SDEM allows us to control for spatial dependence in error terms and examine the relationships between dependent and independent features with a reference to the average value of independent variables in the neighbourhood. Furthermore, it allows us to examine and evaluate whether the value from the neighbourhood matters. The model equations are given as follows:

$$\text{OLS Regression Model : } y = \alpha + \beta X + \varepsilon \quad (2)$$

$$\text{SDEM Model : } y = \beta_0 + \beta X + WX\theta + u, \text{ where } u = \lambda Wu + \varepsilon \quad (3)$$

where y depicts the dependent variable, i.e., childhood stunting, α and β_0 are both intercepts, β is the regression coefficient, X entails independent variables and ε represents the error term. In Equation (3), W is the spatial weights row standardized matrix, θ is known

as the Durbin component, which consists of the spatial lag of explanatory features, and u is the spatial error term that is used to capture spatial autocorrelation in the error term.

The core of this study was to examine if the average value (spatial lag) of our selected independent features impacted childhood stunting in the nearby districts. To fulfill this objective, we implemented an integrated spatial machine learning approach, which represents a move from explainability to forecasting [27]. We incorporated the significant spatial lag component of Flood and Drought Vulnerability (average value from the neighbourhood) from the SDEM estimates and trained that feature value as an input variable into XGBoost and Random Forest machine learning algorithms. After estimating models, we extracted the most important features for childhood stunting. Both models have built in feature selection methods for identifying the most important features [28]. This step was performed to validate and identify the importance of spatial lag with the help of machine learning algorithms.

In the last step we leveraged EXAI methods to identify the existence of nonlinear relationships with our outcome variable. We employed these methods on the trained machine learning algorithms to extract the desired results. In general, the main purpose of EXAI is to offer more than the machine learning black box and explain why the algorithm is acting the way it does. There are two main categories of EXAI techniques: SHapley Additive exPlanations (SHAP) and Local Interpretable Model Agnostics Explanations (LIME). Compared to LIME, SHAP gives a more holistic explanation when it comes to the outcomes of the machine learning algorithm [29,30]. In this way, it employs a game theory approach to determine the extent of each feature in arriving at the final prediction. In addition, it reveals how much each feature contributes to the model and the influence of the features on the prediction. This is carried out by assessing the relevance of the feature in the range of all the features as opposed to the relevance of the feature in isolation. Therefore, SHAP can help to explain how the model leverages individual features to make predictions and consequently provide more insights into the functioning of the model [31–33].

3. Results

The spatial distribution clearly reveals patterns of local densification in childhood stunting in Pakistan. Districts belonging to Punjab province (light shaded regions) possess lower to moderate stunting rates. As one moves into the northeast (Kpk) and southwest (Sindh) regions, we find a clear and visible increase in childhood stunting ranging between 33.5% and 52.3%. Furthermore, the largest province by area, i.e., Baluchistan, on the western side of Pakistan exhibits severe childhood stunting. Figure 2 below clearly shows the district and regional spatial diversity in childhood stunting in Pakistan.

Descriptive statistics are summarized in Table 1, which presents the severity of the discussed issue. The average rate of childhood stunting in Pakistan (excluding ICT) was approximately 40% with an alarming maximum value of 78.60%. Furthermore, the mean values for other chosen measures such as multidimensional poverty stood at 40.98%, female literacy 38.48%, premature birth rate 11.37%, full immunization 42.56%, etc. Table 1 also depicts the corresponding maximum and minimum values for selected independent features.

Table 1. Descriptive statistics.

Variables	Mean	Std. Dev.	Min	Max
Stunting (%)	40.33	14.95	10.10	78.60
Mean household size	7.57	2.40	3.90	19.0
Premature birth rate (%)	11.37	11.48	0.10	75.30
Full immunization (%)	42.56	24.93	13.30	85.40
Multidimensional poverty (%)	40.98	21.53	2.0	87.80
Female literacy (%)	38.48	22.86	3.50	84.70
Women's exposure to mass media (%)	31.12	30.62	0.00	86.60
Marriage before 15 (%)	6.89	4.49	0.80	29.40

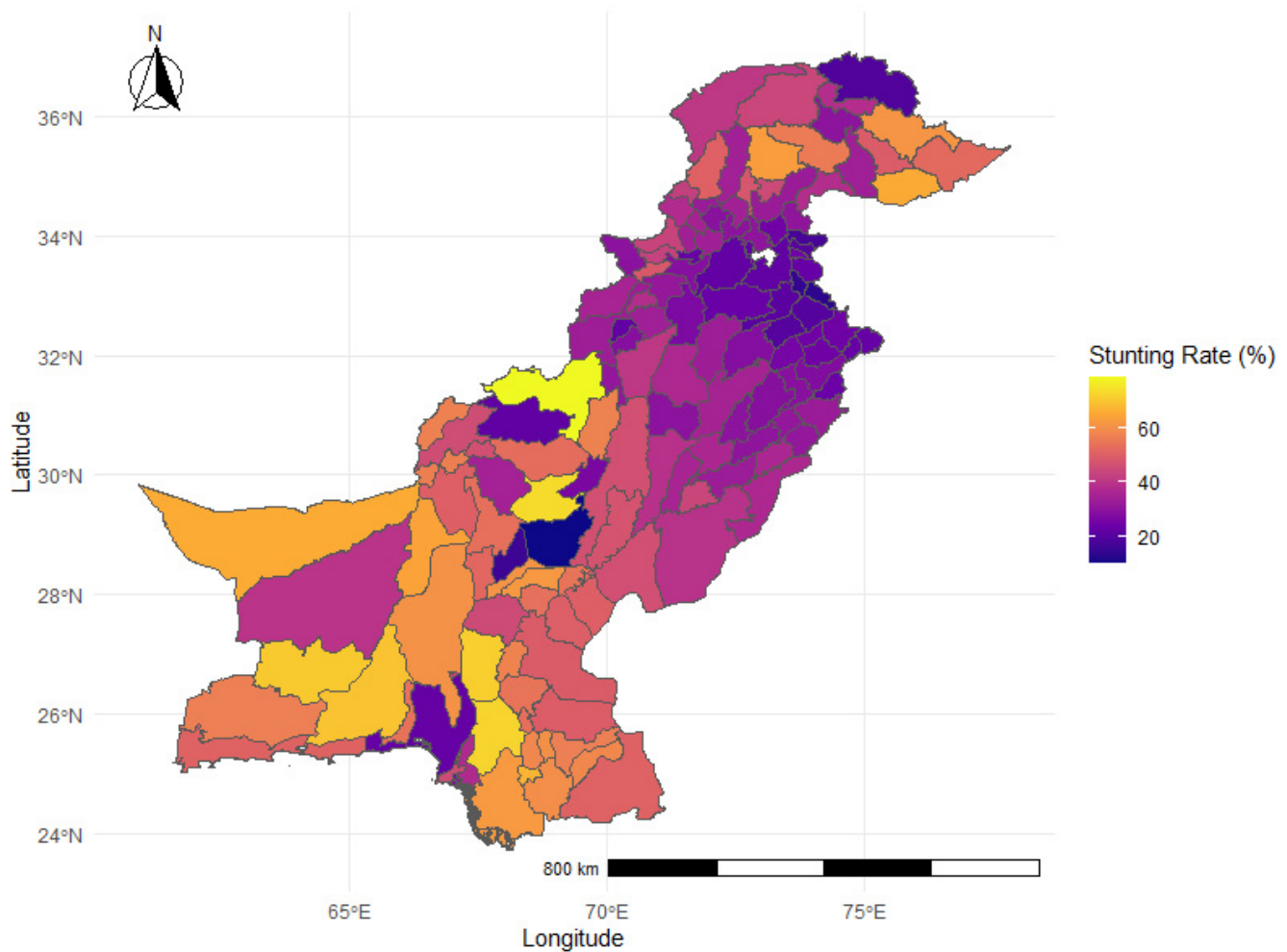


Figure 2. Spatial distribution of childhood stunting at the regional level in Pakistan.

In the next step we tested childhood stunting for spatial autocorrelation using Moran's statistic. The significant Moran's value of 0.56 ($p < 0.000001$) confirmed a positive spatial autocorrelation in childhood stunting, which negates spatial independence and suggests that similar problems exist in neighbouring regions. Figure 3 shows Moran's scatter plot and is composed of four quadrants. Districts with low similar values are grouped together in quadrant III. Districts with high positive values are grouped together in quadrant I. Quadrant II and quadrant IV contain spatial outliers where high and low values are grouped together.

Figure 4 illustrates the same thing as a scatter plot. It shows districts in green colour with low values surrounded by similar districts (low–low regions), light shaded districts possessing higher stunting rates and surrounded by similar districts (high–high regions) and in between a few outliers (low–high regions).

Table 2 below summarizes the estimation results. OLS results demonstrates that multidimensional poverty, women's exposure to mass media, marriage before age 15, and low and high climate vulnerability possess significant relationships with childhood stunting. Specifically, keeping all other factors constant, on average if we observe a percentage point (p.p) increase in multidimensional poverty, it may increase childhood stunting by 0.34 p.p. Additionally, on average high climate vulnerability increases childhood stunting by 0.5 p.p in comparison to medium climate vulnerability (base category). By contrast, on average, low climate vulnerability is negatively associated with childhood stunting. Furthermore, keeping other factors constant, if on average women's exposure to mass media increases by one p.p, we might observe a 0.13 p.p reduction in childhood stunting.

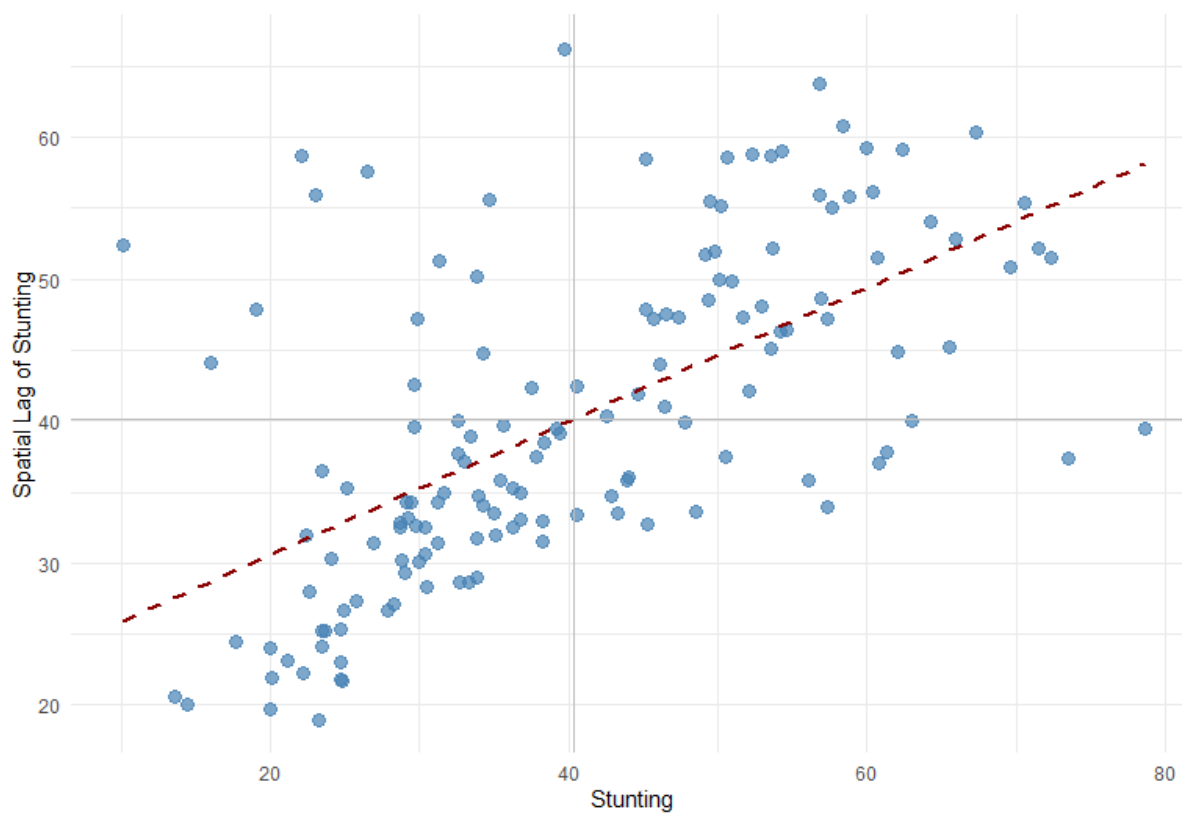


Figure 3. Moran's scatter plot.

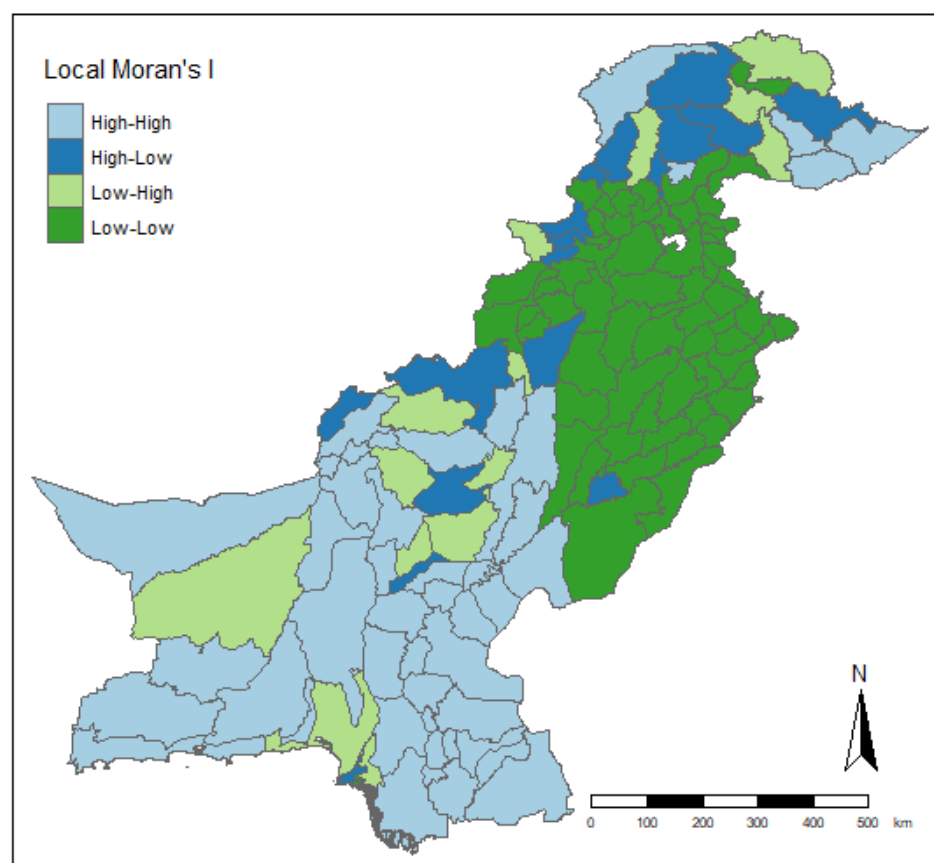


Figure 4. Moran's map.

Table 2. Estimation results.

Variables	OLS Model	SDEM Model
Female literacy	−0.07 (0.07)	−0.004 (0.08)
Multidimensional poverty	0.34 (0.08 ***)	0.24 (0.08 *)
Full immunization	−0.08 (0.05)	−0.01 (0.05)
Marriage before age 15	0.72 (0.227 **)	0.81 (0.21 **)
Women’s exposure to mass media	−0.13 (0.04 **)	−0.01 (0.07)
Premature birth rate	0.07 (0.07)	0.06 (0.06)
Mean household size	0.15 (0.35)	−0.07 (0.32)
Flood and drought vulnerability Low	−0.41 (2.20 *)	−4.67 (2.32 *)
Flood and drought vulnerability High	0.51 (2.10 *)	2.08 (1.98 *)
Spatial lag		
Mass media exposure		−0.33 (0.09 **)
Flood and drought vulnerability High		7.06 (3.79 *)
Spatial lag- λ		0.09 (0.13)
Model diagnostics		
Adjusted R-square	0.58	0.71
AIC	1065.2	1063.1

Note: Standard errors in parentheses; *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; $\dot{p} < 0.1$.

SDEM estimates show that marriage before age 15 and multidimensional poverty have significant relationships with childhood stunting. In contrast to OLS results, SDEM results demonstrate a severe impact of high climate vulnerability on childhood stunting. Furthermore, the significant λ value demonstrates the existence of spatial dependency in the error term. Additionally, significant values of spatial lag (women’s exposure to mass media and high climate vulnerability) suggest a possible impact of these features on childhood stunting in nearby and neighbouring districts. These results confirm that the average value from the neighbourhood of these features (spatial lag) does impact childhood stunting not only in the given district but also in the neighbouring districts as well. In other words, the impact of spatial lag is not geographically bounded.

Model diagnostics show that SDEM outperformed OLS estimates, as the AIC value for SDEM is $1063 < 1065.2$ (OLS AIC). The small difference in AIC values between OLS and SDEM demonstrates that both models fit the data well, with a slight advantage for the SDEM estimates. Additionally, it implies that spatial dependence may not improve the model performance. However, SDEM estimates are still preferable as they account for spatial autocorrelation, which OLS fails to capture, resulting in biased and inconsistent estimates. Furthermore, the value-adjusted R-square is also higher in the case of SDEM, which indicates the total variation in childhood stunting as explained by our independent variables.

Validation of Spatial Lag Component through XGBoost and Random Forest

This section involves the validation of significant spatial lag features as evident from SDEM results. We formulated an integrated approach in which we incorporated the spatial lag value of high climate vulnerability along with other independent features into XGBoost and Random Forest machine learning algorithms. These steps were performed to double-check and validate the significance of spatial lag. Both these machine learning algorithms follow a two-step methodology in extracting the most important features. Both use recursive feature elimination by measuring relative importance and ranking features in several ways such as the split method and average gain method. The most important features are used for building the final model and the remaining ones are used in minimizing residuals [34].

Figure 5 above shows the most important features as extracted by XGBoost and Random Forest. Both extracted the same features, and one noticeable thing is that both extracted the spatial lag of high climate vulnerability as one of the important features.

These results further confirm that the average value from the neighbourhood indeed plays an important role and its impacts are not geographically bounded.

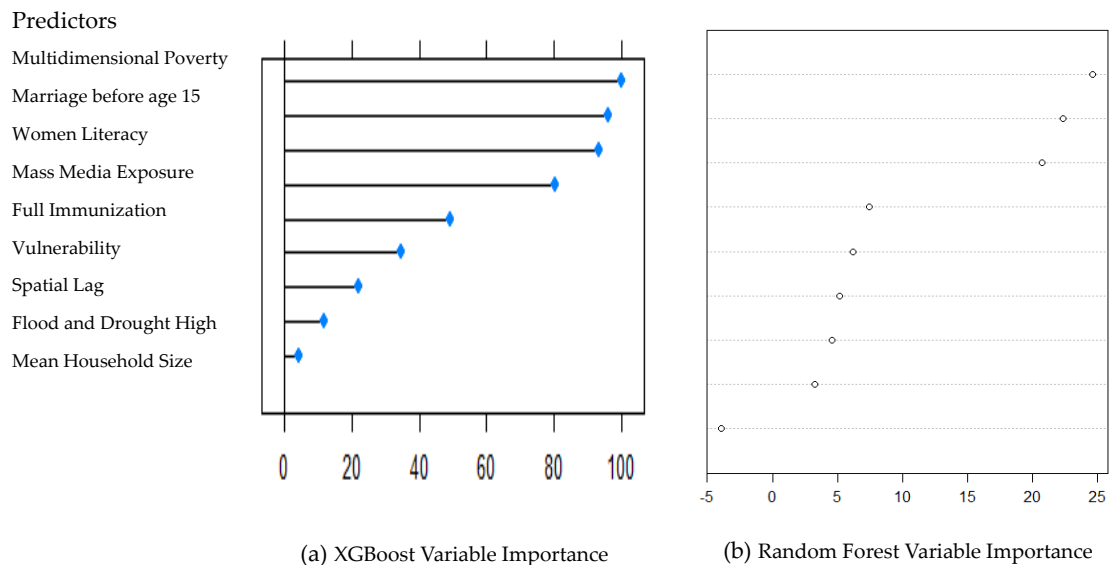


Figure 5. Validation of spatial lag and variable importance via (a) XGBoost, (b) Random Forest.

Figure 6 below exhibits a partial dependency plot that indicates the existence of a nonlinear relationship between multidimensional poverty and childhood stunting, as identified by EXAI Random Forest. However, the ordinary least square regression shows a perfect linear relationship to represent the predicted value of childhood stunting for a random observation. It clearly shows that OLS predicts an approximate value of 54, while EXAI suggests the predicted value to be 57.23 (the real value was 60). This further shows the predictive performance of ML algorithms in comparison to traditional OLS regression.

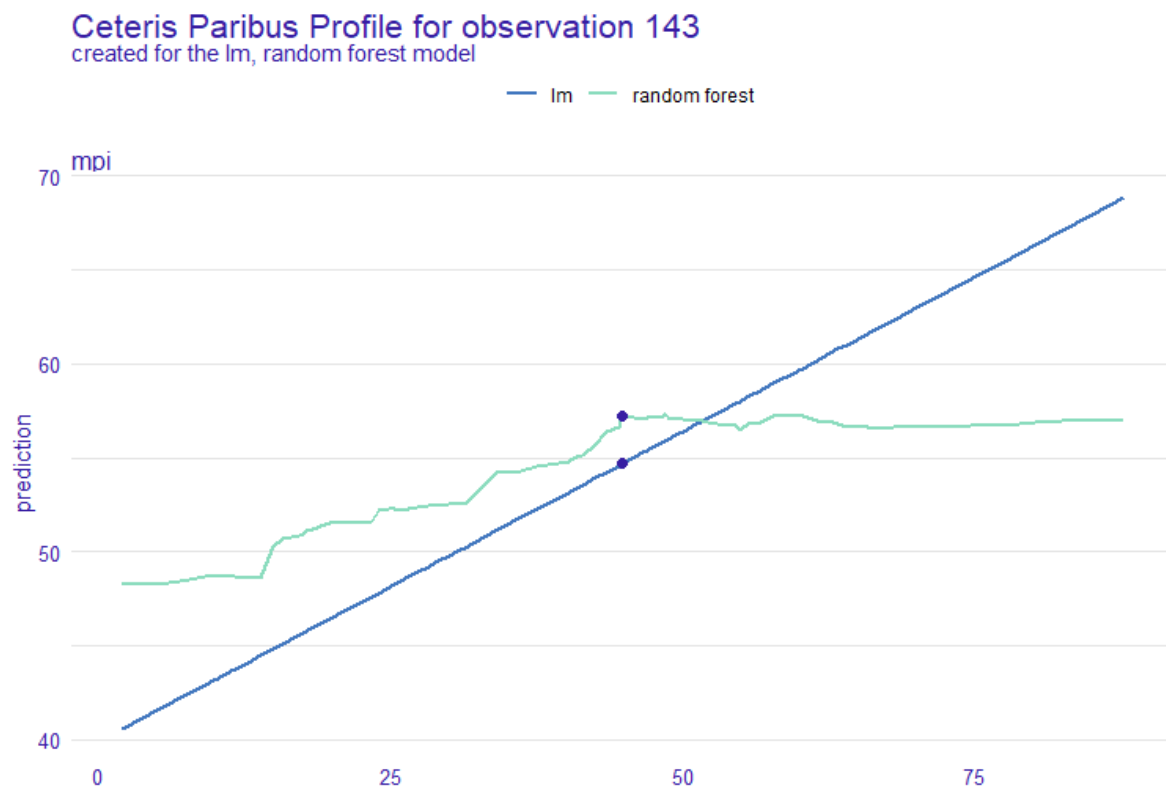


Figure 6. Partial dependency plot.

Figure 7 provides further explanation of local predictions using the SHapley method. It explains the contribution from each feature with positive large values, suggesting a large contribution from each specific feature. For this specific observation, it shows that marriage before age 15 contributes most to this predicted value.

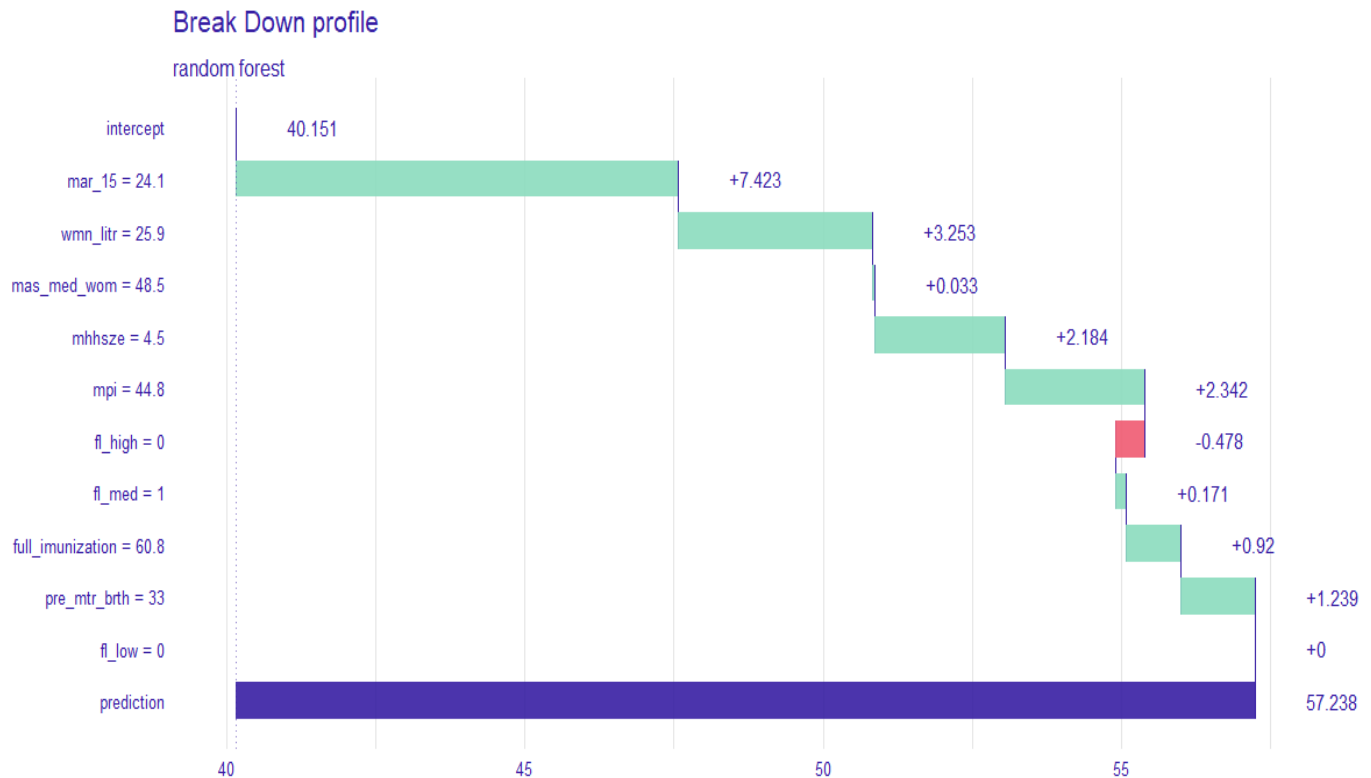


Figure 7. Breakdown profile for a local prediction.

4. Discussion

Child malnutrition remains a significant threat to the development prospects of low- and middle-income countries, including Pakistan. Addressing child malnutrition in all its forms is a key primary objective within the framework of Sustainable Development Goals (SDGs). The present study illustrated two crucial outcomes pertaining to the long-lasting issue of child malnutrition. The initial findings revealed significant spatial inequalities (clusters and spatial outliers) concerning childhood stunting across Pakistan. Moreover, the conducted study analyzed the impact of socioeconomic and climate factors on childhood stunting across all districts in Pakistan, specifying the importance of spatial spillover effects and the existence of nonlinear relationships between the studied features and childhood stunting.

The current study highlights the spatial dependence in childhood stunting across districts in Pakistan. Out of 144 districts examined, 46 exhibited childhood stunting rates exceeding 50%. Notably, the majority of these districts are clustered in southern Punjab Province, northern Kashmir and northwest Khyber Pakhtunkhwa. In contrast, regions in central Pakistan and Gilgit-Baltistan Province have low rates of childhood stunting.

The study investigated the impact of climate vulnerability on childhood stunting revealing significant spatial spillover effects. The findings reveal that high climate vulnerability not only influences childhood stunting within a district but also affects childhood stunting in nearby regions. This suggests the existence of a spatial spillover effect that is not bounded by geographical boundaries [35,36]. The study also found that multidimensional poverty exacerbates childhood stunting. Poor households lack access to fundamental health and other essential services, which consequently impacts children's growth and

development. These results align with existing research on the nexus between poverty and child malnutrition [37,38].

The results show the presence of a nonlinear relationship between multidimensional poverty and childhood stunting, suggesting that childhood stunting does not increase proportionally with poverty. Poor households are financially constrained, limiting their access to essential services like education, health care and nutritious food, etc. Low- and middle-income countries bear the substantial burden of child malnutrition. Such existing nutritional disparities hamper productivity and human capital, making countries prone to vicious cycles of poverty where child malnutrition and poverty reinforce each other.

5. Conclusions

The study explored the spatial correlates of child malnutrition at a district level. The findings highlight the geographical variation in childhood stunting in Pakistan, with 46 districts facing an alarming rate of childhood stunting above 50%, concentrated in the west, northeast and south regions of Pakistan. Socioeconomic features and climate vulnerability both play an important role in determining childhood stunting. Additionally, the study found evidence of spatial spillover effects throughout the neighbourhood of high climate vulnerability and women's exposure to mass media, suggesting the potential significant impact of spatial lag in nearby locations.

The findings from this study clearly demonstrate the importance of considering climate vulnerability as a key determinant of childhood stunting. Pakistan has experienced devastating floods. According to the reports from World Health Organization, more than 116 districts were affected and declared as calamity-hit last year [39]. Not only that, more than 2000 health facilities were completely destroyed, leaving millions without access to health and medical treatment [40]. These events further exacerbate food insecurity and elevate waterborne diseases, thereby worsening child malnutrition and consequently childhood stunting. Integrated and context-based policies are necessary to tackle the alarming issue of childhood stunting. It is pertinent to consider the underlying contextual features and design and implement policies accordingly.

Author Contributions: Conceptualization, Muhammad Usman and Mudassar Rashid; methodology, Muhammad Usman, Ateeq ur Rehman Irshad and Amira Khattak; software, Xiaoyi Zhang and Muhammad Usman; validation, Muhammad Usman, Mudassar Rashid and Ateeq ur Rehman Irshad; formal analysis, Muhammad Usman and Xiaoyi Zhang; investigation, Ateeq ur Rehman Irshad; resources, Muhammad Usman and Amira Khattak; data curation, Muhammad Usman; writing—original draft, Xiaoyi Zhang and Muhammad Usman; writing—review and editing, Muhammad Usman, Ateeq ur Rehman Irshad and Amira Khattak; visualization, Muhammad Usman, Ateeq ur Rehman Irshad and Amira Khattak; supervision, Mudassar Rashid and Ateeq ur Rehman Irshad; project administration, Mudassar Rashid and Ateeq ur Rehman Irshad. All authors have read and agreed to the published version of the manuscript.

Funding: The authors Ateeq Ur Rehman Irshad and Aamira Khattak would like to thank Prince Sultan University for paying the APC and the support through TAS research laboratory.

Data Availability Statement: All data were collected from publicly open repositories and are available upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. United Nations System Standing Committee on Nutrition (UNSCN). *Nutrition and the Post-2015 Sustainable Development Goals*; UNSCN: Geneva, Switzerland, 2014.
2. UNICEF. *Global Nutrition Report*; United Nations Global Marketplace: New York, NY, USA, 2018.
3. UNICEF. *National Nutrition Survey Pakistan*; United Nations Global Marketplace: New York, NY, USA, 2017.
4. Khan, A.Y.; Fatima, K.; Ali, M. Sanitation ladder and undernutrition among under-five children in Pakistan. *Environ. Sci. Pollut. Res.* **2021**, *28*, 38749–38763. [[CrossRef](#)] [[PubMed](#)]

5. Qamar, K.; Nchasi, G.; Mirha, H.T.; Siddiqui, J.A.; Jahangir, K.; Shaeen, S.K.; Islam, Z.; Essar, M.Y. Water sanitation problem in Pakistan: A review on disease prevalence, strategies for treatment and prevention. *Ann. Med. Surg.* **2022**, *82*, 104709. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Smith, L.C.; Haddad, L. Reducing Child Undernutrition: Past Drivers and Priorities for the Post-MDG Era. *World Dev.* **2015**, *68*, 180–204. [\[CrossRef\]](#)
7. Kim, R.; Mejía-Guevara, I.; Corsi, D.J.; Aguayo, V.M.; Subramanian, S. Relative importance of 13 correlates of child stunting in South Asia: Insights from nationally representative data from Afghanistan, Bangladesh, India, Nepal, and Pakistan. *Soc. Sci. Med.* **2017**, *187*, 144–154. [\[CrossRef\]](#)
8. Rizal, M.F.; van Doorslaer, E. Explaining the fall of socioeconomic inequality in childhood stunting in Indonesia. *SSM Popul. Health* **2019**, *9*, 100469. [\[CrossRef\]](#)
9. Saaka, M.; Saapiire, F.N.; Dogoli, R.N. Independent and joint contribution of inappropriate complementary feeding and poor water, sanitation and hygiene (WASH) practices to stunted child growth. *J. Nutr. Sci.* **2021**, *10*, e109. [\[CrossRef\]](#)
10. Beal, T.; Le, D.T.; Trinh, T.H.; Burra, D.D.; Huynh, T.; Duong, T.T.; Truong, T.M.; Nguyen, D.S.; Nguyen, K.T.; de Haan, S.; et al. Child stunting is associated with child, maternal, and environmental factors in Vietnam. *Matern. Child Nutr.* **2019**, *15*, e12826. [\[CrossRef\]](#)
11. Kumar, P.; Chauhan, S.; Patel, R.; Srivastava, S.; Bansod, D.W. Prevalence and factors associated with triple burden of malnutrition among mother-child pairs in India: A study based on National Family Health Survey 2015–16. *BMC Public Health* **2021**, *21*, 391. [\[CrossRef\]](#)
12. Kim, R.; Rajpal, S.; Joe, W.; Corsi, D.J.; Sankar, R.; Kumar, A.; Subramanian, S. Assessing associational strength of 23 correlates of child anthropometric failure: An econometric analysis of the 2015–2016 National Family Health Survey, India. *Soc. Sci. Med.* **2019**, *238*, 112374. [\[CrossRef\]](#)
13. Efevbera, Y.; Bhabha, J.; Farmer, P.E.; Fink, G. Girl child marriage as a risk factor for early childhood development and stunting. *Soc. Sci. Med.* **2017**, *185*, 91–101. [\[CrossRef\]](#)
14. Usman, M.; Kopczewska, K. Spatial and Machine Learning Approach to Model Childhood Stunting in Pakistan: Role of Socio-Economic and Environmental Factors. *Int. J. Environ. Res. Public Health* **2022**, *19*, 10967. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Kinyoki, D.K.; Berkley, J.A.; Moloney, G.M.; Odundo, E.O.; Kandala, N.-B.; Noor, A.M. Environmental predictors of stunting among children under-five in Somalia: Cross-sectional studies from 2007 to 2010. *BMC Public Health* **2016**, *16*, 654. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Thiede, B.C.; Gray, C. Climate exposures and child undernutrition: Evidence from Indonesia. *Soc. Sci. Med.* **2020**, *265*, 113298. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Johnson, F.A. Spatiotemporal clustering and correlates of childhood stunting in Ghana: Analysis of the fixed and nonlinear associative effects of socio-demographic and socio-ecological factors. *PLoS ONE* **2022**, *17*, e0263726.
18. Gupta, A.K.; Santhya, K.G. Proximal and contextual correlates of childhood stunting in India: A geo-spatial analysis. *PLoS ONE* **2020**, *15*, e0237661. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Tamir, T.T.; Techane, M.A.; Dessie, M.T.; Atalell, K.A. Applied nutritional investigation spatial variation and determinants of stunting among children aged less than 5 y in Ethiopia: A spatial and multilevel analysis of Ethiopian Demographic and Health Survey 2019. *Nutrition* **2022**, *103–104*, 111786. [\[CrossRef\]](#)
20. Hailu, B.A.; Bogale, G.G.; Beyene, J. Spatial heterogeneity and factors influencing stunting and severe stunting among under-5 children in Ethiopia: Spatial and multilevel analysis. *Sci. Rep.* **2020**, *10*, 16427. [\[CrossRef\]](#)
21. Ndagijimana, A.; Nduwayezu, G.; Kagoyire, C.; Elfving, K.; Umubyeyi, A.; Mansourian, A.; Lind, T. Childhood stunting is highly clustered in Northern Province of Rwanda: A spatial analysis of a population-based study. *Heliyon* **2024**, *10*, e24922. [\[CrossRef\]](#)
22. Ahmad, D.; Afzal, M.; Imtiaz, A. Effect of socioeconomic factors on malnutrition among children in Pakistan. *Future Bus. J.* **2020**, *6*, 30. [\[CrossRef\]](#)
23. Multiple Indicator Cluster Survey 2017–2018—Punjab. (2018, November 1). UNICEF Pakistan. Available online: <https://www.unicef.org/pakistan/reports/multiple-indicator-cluster-survey-2017-18-punjab> (accessed on 13 December 2023).
24. Integrated Context Analysis (ICA): On Vulnerability to Food Insecurity and Natural Hazards—Pakistan, 2017. Available online: <https://reliefweb.int/report/pakistan/integrated-context-analysis-ica-vulnerability-food-insecurity-and-natural-hazards> (accessed on 10 November 2023).
25. Anselin, L. The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In *Spatial Analytical Perspectives on GIS*; Routledge: Abingdon-on-Thames, UK, 2019; pp. 111–126.
26. Getis, A.; Ord, J.K. The Analysis of Spatial Association by Use of Distance Statistics. *Geogr. Anal.* **1992**, *24*, 189–206. [\[CrossRef\]](#)
27. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.
28. Li, X.; Du, M.; Chen, J.; Chai, Y.; Lakkaraju, H.; Xiong, H. M⁴: A Unified XAI Benchmark for Faithfulness Evaluation of Feature Attribution Methods across Metrics, Modalities and Models. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 1630–1643.
29. Rothman, D. *Hands-On Explainable AI (XAI) with Python: Interpret, Visualize, Explain, and Integrate Reliable AI for Fair, Secure, and Trustworthy AI Apps*; Packt Publishing Ltd.: Birmingham, UK, 2020.
30. Naz, Z.; Khan MU, G.; Saba, T.; Rehman, A.; Nobanee, H.; Bahaj, S.A. An explainable AI-enabled framework for interpreting pulmonary diseases from chest radiographs. *Cancers* **2023**, *15*, 314. [\[CrossRef\]](#) [\[PubMed\]](#)

31. Shi, X.; Wong, Y.D.; Li, M.Z.-F.; Palanisamy, C.; Chai, C. A feature learning approach based on XGBoost for driving assessment and risk prediction. *Accid. Anal. Prev.* **2019**, *129*, 170–179. [[CrossRef](#)]
32. Akhtar, K.; Yaseen, M.U.; Imran, M.; Khattak SB, A.; Nasralla, M.M. Predicting inmate suicidal behavior with an interpretable ensemble machine learning approach in smart prisons. *PeerJ Comput. Sci.* **2024**, *10*, e2051. [[CrossRef](#)]
33. Rehman, A.; Haseeb, K.; Saba, T.; Lloret, J.; Sendra, S. An optimization model with network edges for multimedia sensors using artificial intelligence of things. *Sensors* **2021**, *21*, 7103. [[CrossRef](#)]
34. Mahapatra, B.; Walia, M.; Rao, C.A.R.; Raju, B.M.K.; Saggurti, N. Vulnerability of agriculture to climate change increases the risk of child malnutrition: Evidence from a large-scale observational study in India. *PLoS ONE* **2021**, *16*, e0253637. [[CrossRef](#)]
35. Okunlola, O.A.; Kassouri, Y. Empirical investigation of the agriculture–malnutrition nexus in Africa: Spatial clustering and spillover effects. *Rev. Dev. Econ.* **2023**, *27*, 685–709. [[CrossRef](#)]
36. Rahman, M.A.; Halder, H.R.; Rahman, M.S.; Parvez, M. Poverty and childhood malnutrition: Evidence-based on a nationally representative survey of Bangladesh. *PLoS ONE* **2021**, *16*, e0256235.
37. Siddiqui, F.; Salam, R.A.; Lassi, Z.S.; Das, J.K. The Intertwined Relationship Between Malnutrition and Poverty. *Front. Public Health* **2020**, *8*, 453. [[CrossRef](#)]
38. Adeyeye, S.A.O.; Ashaolu, T.J.; Bolaji, O.T.; Abegunde, T.A.; Omoyajowo, A.O. Africa and the Nexus of poverty, malnutrition and diseases. *Crit. Rev. Food Sci. Nutr.* **2023**, *63*, 641–656. [[CrossRef](#)] [[PubMed](#)]
39. United Nations. *Global Nutrition Overview*; United Nations: New York, NY, USA, 2022.
40. National Disaster Management Authority. *Pakistan (2022) Floods Response Plan: 1—Pakistan*; United Nations Office for the Coordination of Humanitarian Affairs: New York, NY, USA, 2022.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.