



ARTICLE

A robust geographically weighted LASSO model with Huber Loss for outlier-resilient spatial analysis: application to COVID-19 dynamics in Indian districts

 Megha Sharma,^{*}¹  Shalini Chandra,¹ and  Gargi Tyagi²

¹Department of Mathematics and Statistics, Banasthali Vidyapith, Rajasthan, India.

²S.N.Medical College, Agra, India.

*Corresponding author. Email: meghasharma15aug@gmail.com

(Received: September 7, 2025; Revised: May 19, 2026; Accepted: June 17, 2026; Published: September 30, 2026)

Abstract

Geographically Weighted Regression (GWR) is a widely used technique for modeling spatially varying relationships in epidemiological data. However, its sensitivity to multicollinearity and outliers can limit its reliability in complex datasets, such as those involving COVID-19. To address these limitations, this study first employed the RGWR model, which uses the Huber loss to enhance robustness against outliers. Simultaneously, to handle multicollinearity and enable local variable selection, the Geographically Weighted LASSO (GWL) model is employed. Despite their individual strengths, both RGWR and GWL face challenges when outliers and multicollinearity coexist. To overcome this, we propose a Robust Geographically Weighted LASSO (RGWL) model that integrates the Huber loss function with a spatially adaptive LASSO penalty. This combined formulation enhances robustness to outliers while effectively performing localized regularization and variable selection. We apply these models to district-level COVID-19 data from India, incorporating a comprehensive set of socio-demographic, environmental, and health-related variables. Comparative analysis reveals that the proposed RGWL model outperforms both RGWR and GWL in terms of predictive accuracy and goodness of fit, particularly in the presence of extreme observations and collinear predictors. The spatially varying coefficients further highlight region-specific patterns, demonstrating the utility of the RGWL model for robust, localized analysis in epidemiological research.

Keywords: Geographically Weighted Regression, Robustness, LASSO, Heteroscedasticity, Outliers

1. Introduction

While the immediate threat of COVID-19 has diminished, its widespread effects continue to offer valuable insights for research. Understanding the factors that contributed to its spread and

severity remains crucial, especially in countries like India, which faced significant economic disruption, unemployment, and pressure on its healthcare system (Sardar *et al.*, 2020). The pandemic highlighted social inequalities that increased vulnerability to disease transmission (Ahmed *et al.*, 2020), including poor living conditions (Pereira & Oliveira, 2020), high population density (Rocklöv & Sjödin, 2020), inadequate healthcare access, and a high proportion of vulnerable individuals (Dutta *et al.*, 2021). Environmental factors such as temperature were also associated with disease severity in various countries (Chen *et al.*, 2020; Tosepu *et al.*, 2020; Chung *et al.*, 2021; Bashir *et al.*, 2020). Additionally, urban slums (Sridhar, 2023), smoking habits, and other socio-behavioral factors contributed to disparities in infection risk and healthcare access. Beyond these socio-environmental elements, individual baseline demographics such as age, gender, and education, alongside underlying comorbidities like neurological disorders, have been identified as highly significant predictors of severe clinical degradation in hospitalized patients (CARMO *et al.*, 2021).

Such pandemic outbreaks are not only driven by individual and societal factors but are also spatially dependent in nature, as shown in numerous studies (Sarkar *et al.*, 2021; Ramírez-Aldana *et al.*, 2020; Appiah-Otoo & Kursah, 2022; Adekunle *et al.*, 2020). Spatial dependency implies that disease outcomes in one region are influenced by those in nearby regions due to geographic, social, or environmental connections. To capture this, spatial modeling approaches have gained importance in epidemiological research. Among these, the Geographically Weighted Regression (GWR) model has emerged as a widely used method to account for spatial heterogeneity by estimating location-specific relationships between risk factors and disease outcomes (Sarkar *et al.*, 2021; Ramírez-Aldana *et al.*, 2020; Appiah-Otoo & Kursah, 2022; Adekunle *et al.*, 2020). While GWR has been widely applied in health geography and epidemiology, its performance can be compromised by two critical challenges: local multicollinearity and outlier sensitivity. Multicollinearity inflates variance in coefficient estimates, while outliers often stemming from data irregularities or localized outbreaks can unduly influence model results.

Multicollinearity inflates variance in coefficient estimates, while outliers often stemming from data irregularities or localized outbreaks can unduly influence model results. In classical regression frameworks, researchers have addressed these issues through robust and penalized estimation methods. The utility of penalized frameworks like LASSO and Elastic-Net for managing multicollinearity and isolating critical features has been well-documented across diverse datasets, including clinical anthropometry (NEISSE *et al.*, 2021), high-dimensional biological data (Carrilho & Lopes, 2022), and remote sensing (Barros *et al.*, 2024). Furthermore, combining these regularization techniques with robust loss functions such as Huber or Tukey's biweight has been shown to simultaneously reduce variance and limit the influence of extreme observations (Wang *et al.*, 2007; Fan *et al.*, 2012; Alfons *et al.*, 2013; Khan *et al.*, 2022; Zhang & Mei, 2011; Wulandari *et al.*, 2019). These approaches integrate the strengths of regularization for handling multicollinearity with robust estimation techniques for mitigating the effects of outliers.

In classical regression frameworks, researchers have addressed these issues through robust and penalized estimation methods. For instance, combining ridge or LASSO regularization with robust loss functions such as Huber or Tukey's biweight has been shown to simultaneously reduce variance and limit the influence of extreme observations (Wang *et al.*, 2007; Fan *et al.*, 2012; Alfons *et al.*, 2013; Khan *et al.*, 2022; Zhang & Mei, 2011; Wulandari *et al.*, 2019). These approaches integrate the strengths of regularization for handling multicollinearity with robust estimation techniques for mitigating the effects of outliers.

However, in the context of spatial regression frameworks, particularly GWR, the challenges of multicollinearity and outliers have generally been addressed independently. To mitigate multicollinearity, various penalized GWR extensions have been developed, such as Geographically Weighted Ridge Regression (GWRR), Geographically Weighted LASSO (GWL), and Geographically Weighted Principal Component Regression (GWPCR), which stabilize coefficient estimates

through localized regularization techniques (Wheeler, 2009). In contrast, efforts to address outliers in the GWR framework have led to the development of robust variants that modify the loss function. Methods incorporating Huber loss, Tukey's biweight, or adaptive weighting schemes aim to limit the influence of extreme observations on local coefficient estimates (Foody, 2003). Robust GWR models improve estimation stability under outlier contamination but overlook multicollinearity concerns. Consequently, while both types of methodological adaptations exist, a unified approach that simultaneously handles multicollinearity and outliers within a spatially varying context remains underexplored.

To address this methodological gap, we propose a Robust Geographically Weighted LASSO (RGWL) model that incorporates the Huber loss function into the GWL framework. The Huber loss blends the sensitivity of least squares with the resilience of absolute deviations, providing a balanced approach to estimation in the presence of outliers. This integration allows the RGWL model to simultaneously account for spatial nonstationarity, local variable selection, and robustness to data anomalies a critical advancement for real-world epidemiological applications.

In this study, we develop and apply the RGWL model to district-level COVID-19 data from India. The proposed approach enhances classical and spatial regression techniques by unifying penalization and robustness within a geographically adaptive framework. Through this, we aim to offer a more reliable tool for spatial analysis in public health, particularly in settings where data quality and heterogeneity pose significant challenges.

2. Models and Estimators

This section outlines the spatial modeling framework employed in the study. We begin with the Geographically Weighted Regression (GWR) model, which captures spatial heterogeneity by allowing regression coefficients to vary across locations. We then describe the Robust Geographically Weighted Regression (RGWR) model, which modifies GWR by incorporating a robust loss function to reduce sensitivity to outliers. Building further, the Geographically Weighted LASSO (GWL) model extends GWR by introducing localized L1-penalization, enabling both spatial adaptivity and variable selection.

Finally, we present the proposed Robust Geographically Weighted LASSO (RGWL) model, which integrates the robustness of RGWR with the sparsity-inducing properties of GWL. This unified approach enhances estimation accuracy in the presence of outliers while capturing spatial non-stationarity and selecting relevant predictors. Together, these models form a comprehensive and flexible framework for spatial regression analysis.

Geographically Weighted Regression (GWR)

GWR model captures local variations in relationships between the dependent and independent variables by fitting a separate regression model for each spatial location (Brunsdon *et al.*, 1998). The GWR model is specified as:

$$\gamma_i = \beta_{i0} + \sum_{j=1}^m X_{ij}\beta_{ij} + \epsilon_i, \quad i = 1, 2, \dots, n. \quad (1)$$

Where γ_i denotes the dependent variable at location i , β_{i0} is the intercept at that location, X_{ij} is the value of the j^{th} predictor, β_{ij} is the local coefficient for the j^{th} variable at location i , and ϵ_i is the error term.

Parameter estimates at each location i are derived as:

$$\hat{\beta}(i)_{GWR} = (X^T W(i) X)^{-1} X^T W(i) \gamma, \quad (2)$$

where X denotes the matrix of predictors, γ is the matrix of dependent variable and $W(i)$ is a diagonal spatial weight matrix computed using a kernel function. Specifically, the adaptive bisquare kernel function is:

$$w_k(i) = \begin{cases} \left[1 - \left(\frac{d_{ik}}{bw} \right)^2 \right]^2, & \text{if } k \in \{N_i\}, \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

where d_{ik} is the distance between locations i and k , and bw is the bandwidth controlling the influence zone. The set $\{N_i\}$ includes the N nearest neighbors around location i . Points beyond this range receive zero weight.

Bandwidth selection is carried out using cross-validation (CV) to minimize prediction error:

$$CV(bw) = \sum_{i=1}^n [\gamma_i - \hat{\gamma}_{\neq i}(bw)]^2, \quad (4)$$

where $\hat{\gamma}_{\neq i}(bw)$ is the leave-one-out prediction of γ_i with the i^{th} observation excluded. The bandwidth bw is optimized by minimizing the CV score over iterations.

Robust Geographically Weighted Regression (RGWR)

RGWR extends the standard GWR model by incorporating robustness to outliers through the use of M-estimators, particularly the Huber loss function (Zhang & Mei, 2011; Wulandari *et al.*, 2019). While GWR assumes Gaussian errors and is sensitive to extreme observations, RGWR mitigates this issue by replacing the traditional squared loss with a robust loss function.

The model estimates spatially varying coefficients by minimizing a locally weighted objective function based on the Huber loss:

$$\hat{\beta}(i) = \arg \min_{\beta} \sum_{k=1}^n w_k(i) \cdot \rho_{\delta} \left(\gamma_k - \mathbf{x}_k^{\top} \beta \right), \quad (5)$$

where $w_k(i)$ denotes the spatial kernel weight for observation k relative to location i , $\mathbf{x}_k$ is the vector of predictors at location k and $\rho_{\delta}(\cdot)$ is the Huber loss function:

$$\rho_{\delta}(r) = \begin{cases} \frac{1}{2} r^2, & \text{if } |r| \leq \delta, \\ \delta(|r| - \frac{1}{2}\delta), & \text{if } |r| > \delta \end{cases} \quad (6)$$

where r denoted as the residuals and δ is the parametr controls the threshold between quadratic and linear loss behavior, allowing the model to down-weight the influence of outliers while maintaining efficiency for normally distributed errors.

Geographically Weighted Ridge Regression (GWRR)

GWRR is a modified technique within the domain of spatial regression, in which GWR model combine with the ridge regression Wheeler, 2009. Estimator of parametres of GWRR model at i^{th} location is obtained by

$$\hat{\beta}(i)_{GWRR} = (X^T W(i)X + CI)^{-1} X^T W(i)y. \quad (7)$$

The process involves predicting bandwidth values to form a weighted matrix, minimizing bias using CV, and iteratively determining the coefficient value C for each bandwidth. These results are then applied to estimate spatial model with ridge regression coefficients.

Geographically Weighted LASSO (GWL)

GWL extends the LASSO framework to account for spatial heterogeneity in the relationships between predictors and response variables (Wheeler, 2009). GWL combines the local estimation of GWR with the variable selection capability of LASSO, thereby addressing both spatial nonstationarity and multicollinearity.

In GWL, LASSO-based parameter estimation is performed locally using a kernel function to define spatial weights. The model assigns a unique shrinkage parameter to each location, allowing for spatially adaptive regularization. The general formulation of the GWL estimator incorporates both spatial weights and LASSO penalties:

$$\hat{\beta}(i)_{GWL} = \arg \min_{\beta} \left\{ (y - X\beta)^T W(i)(y - X\beta) + \lambda(i) \sum_{j=1}^m |\beta_j| \right\}, \quad (8)$$

where $W(i)$ is the spatial weight matrix at location i , and $\lambda(i)$ is the local shrinkage parameter determined through cross-validation. This adaptive penalty allows the GWL model to retain locally significant predictors while shrinking or eliminating others, leading to improved prediction accuracy and model parsimony in spatial contexts.

Proposed Model- Robust Geographically Weighted LASSO (RGWL)

This study proposed the RGWL model, as RGWL integrates the robust loss formulation of RGWR with the sparsity-inducing LASSO penalty of GWL.

The RGWL estimator is defined as:

$$\hat{\beta}(i)_{RGWL} = \arg \min_{\beta} \sum_{k=1}^n w_k(i) \cdot \rho_{\delta} \left(y_k - \mathbf{x}_k^T \beta \right) + \lambda(i) \sum_{j=1}^m |\beta_j|, \quad (9)$$

where $w_k(i)$ denotes the spatial kernel weight for observation k relative to location i , and $\rho_{\delta}(\cdot)$ is the Huber loss function as defined in Equation 6, and $\lambda(i)$ is the location-specific tuning parameter that controls the degree of L1-penalization (Wang *et al.*, 2007; Wulandari *et al.*, 2019).

The model is estimated through an iterative reweighting scheme that updates residual-based weights according to the Huber function and solves the penalized local regression problem. This formulation enhances robustness against outliers while enabling automatic local variable selection in spatially varying contexts.

Estimation Procedure for Robust Geographically Weighted LASSO (RGWL):

1. **Initial Estimation Using GWL:**

For each location i , estimate the Geographically Weighted LASSO (GWL) model using spatial kernel weights. This provides initial local coefficient estimates that account for spatial heterogeneity and perform variable selection via L1 regularization.

2. **Residual Calculation:**

Calculate the residuals for each observation based on the fitted GWL model. These residuals reflect the local model's accuracy and serve to identify potential outliers or influential data points.

3. **Robustification Using M-estimation:**

Evaluate the Huber loss function on the residuals obtained in Step 2. This robust loss function down-weights the influence of observations with large residuals, thereby mitigating the effect of outliers.

4. **Iterative Reweighting:**

Update the local estimation procedure by adjusting the influence of observations according to their residual-based robustness weights. The procedure is iterated until the coefficient estimates converge, ensuring stable and robust solutions.

5. **Final Coefficient Estimation:**

At each location i , solve the robustified penalized regression problem that integrates spatial weighting, the Huber loss, and L1 regularization.

This yields the final RGWL estimates, which are simultaneously robust to outliers and capable of localized variable selection.

3. Data Analysis and Key Findings

3.1 Data

This study covers districts from all states and union territories of India, with the exception of Assam, Delhi, Goa, Manipur, Telangana, and Sikkim, due to the lack of district-level COVID-19 data in their official state bulletins. The analysis is based on district boundaries as defined in 2019. The two main outcome variables used in this study are the cumulative COVID-19 cases (CCC) and cumulative COVID-19 deaths (CCD), which capture the overall burden and severity of the pandemic across districts.

Data for CCC and CCD were obtained from (<http://www.covidindia.org>), a publicly available platform that compiled information from official government sources and state bulletins. a publicly accessible data repository that aggregated and disseminated COVID-19 statistics from official government reports, state health department bulletins, and other authorized public health sources. The platform provided regularly updated district-level information on COVID-19 infections, recoveries, and mortality, facilitating detailed spatial and temporal analyses of the pandemic across India. Due to its comprehensive geographic coverage and systematic compilation of data from multiple official sources, the database has been widely utilized in COVID-19-related epidemiological and spatial studies.

Alongside these outcomes, a set of independent variables given in Table 1 was selected based on prior evidence and their potential influence on the transmission and impact of COVID-19.

Table 1 presents the selected variables, which were compiled from multiple reliable secondary data sources, including the National Family Health Survey-5 (NFHS-5), Global Data Lab (GDL), Rural Health Statistics (RHS), and the Registrar General of India (RGI). These sources collectively provide district-level information on demographic structure, socioeconomic development, health status, and healthcare infrastructure. NFHS-5 contributes detailed household and health-related indicators, GDL supplies standardized measures of regional development, RHS provides information

on healthcare facilities and workforce availability, and RGI serves as the primary source of official demographic statistics. The use of these complementary datasets ensured broad coverage of the factors potentially influencing the spatial distribution of COVID-19 prevalence across India.

Table 1. List of independent variables with their corresponding abbreviations for analysis of COVID-19 outcomes in Indian districts

Abbreviation	Description
Pop_U15	Population below age 15 years
Elec_HH	Population living in households with electricity
Sanit_HH	Population living in households using improved sanitation facilities
CleanFuel	Households using clean fuel for cooking
HealthIns	Households with at least one member covered under health insurance
Women_Lit	Women who are literate
Women_School	Women with 10+ years of schooling
Undernourished_W	Proportion of undernourished women
Obese_W	Proportion of obese women
Tobacco_15+	Tobacco use among population aged 15+
Alcohol_15+	Alcohol use among population aged 15+
Pop_Density	Population density
Urban_Pop	Proportion of urban population
HealthCenters	Number of health centers (Sub-centers + PHCs + CHCs)
HTN_Prev	Hypertension prevalence among adults (15+)
BloodSugar	Blood sugar levels among adults (15+)
Avg_Temp	Average temperature
Rel_Humidity	Relative humidity
Poverty_Pop	Proportion of population below poverty line

To ensure the quality and consistency of the dataset, preliminary data preparation involved cleaning procedures to address issues such as missing values and duplicate records, achieved through cross-verification from multiple data sources. Due to the heterogeneous nature of the data, a critical step in the process was data integration, wherein information from various origins was merged to form a coherent and unified dataset. Since data on COVID-19 outcomes and associated risk factors were derived from distinct sources, standardization was essential to harmonize variable scales and facilitate meaningful comparisons across different data types.

These preparatory steps were completed prior to aggregating the data at the district level and merging it with the district boundary shapefile for India, obtained from the DIV-GIS database, using ArcGIS Desktop 10.7. All data analyses were conducted using R software with packages such as *sp*, *spgwr*, *spdep*, and *spatstat*.

3.2 Preliminary Data Insights

Before proceeding with the modeling framework, it is essential to understand the underlying structure and characteristics of the dataset. This section aims to explore the data patterns and evaluate key assumptions that inform the selection of appropriate models. During this exploratory phase, a range of statistical and spatial diagnostics were performed. Notably, three critical aspects emerged: spatial dependency, multicollinearity, and the presence of outliers.

Spatial patterns of COVID-19 cases and deaths across Indian districts are illustrated in the choropleth maps presented in Figure 1.

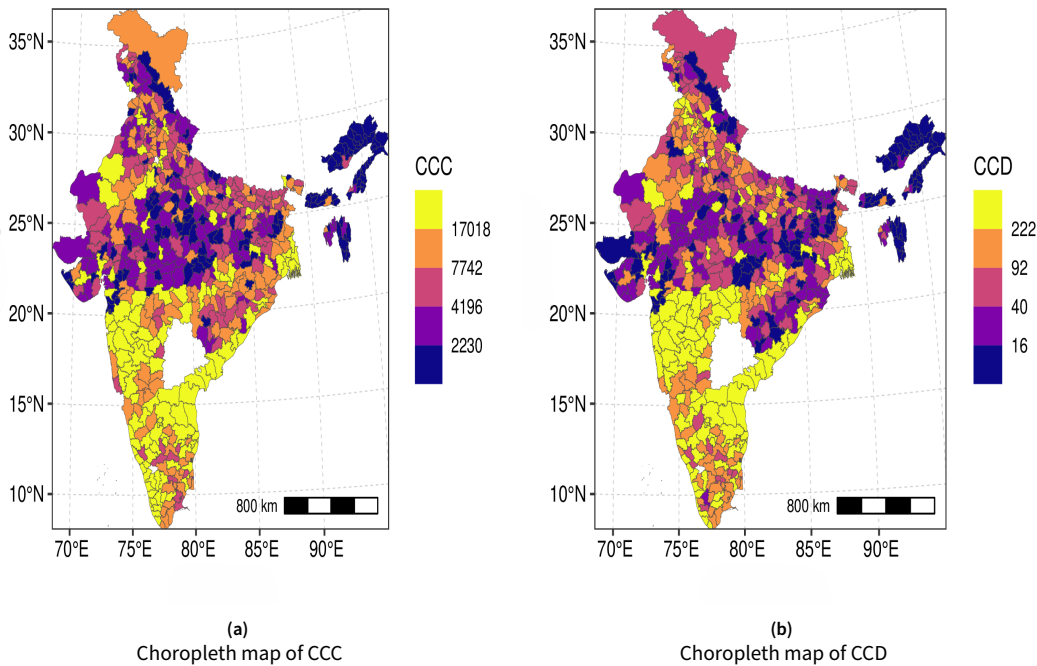


Figure 1. Choropleth maps of CCC and CCD showing spatial distribution and regional variation across Indian districts.

These maps reveal clear regional variations and clustering of both CCC and CCD, indicating that the outcomes are not uniformly distributed across districts. The observed spatial patterns were further assessed using the Global Moran’s I statistic, which quantifies the degree of spatial autocorrelation for each dependent variable (Cliff & Ord, 1981). The results indicated significant spatial autocorrelation for both CCC (0.43) and CCD (0.47) with $p\text{-value} = 0.0001$, confirming that COVID-19 outcomes exhibit strong spatial dependence across Indian districts. This pronounced geographic clustering strongly echoes the macro-level findings of Dutta *et al.* (2021), whose analysis confirmed a significant spatial dependency in COVID-19 transmission trends across Indian states.

Additionally, multicollinearity among explanatory variables was examined through the Variance Inflation Factor (VIF), where several variables exceeded the conventional threshold of 10 (See Table 2), indicating that the variable is highly linearly correlated with one or more of the other predictors (Gujarati, 2004).

Table 2. Multicollinearity Assessment Using VIF Values of Independent Variables in India

Variable	VIF	Variable	VIF
Pop_U15	3.20	Alcohol_15+	2.41
Elec_HH	1.59	Pop_Density	1.32
Sanit_HH	2.63	Urban_Pop	2.53
CleanFuel	3.54	HealthCenters	1.06
HealthIns	1.53	HTN_Prev	38.13
Women_Lit	6.01	BloodSugar	24.47
Women_School	5.48	Avg_Temp	1.99
Undernourished_W	24.67	Rel_Humidity	2.84
Obese_W	22.43	Poverty_Pop	2.89
Tobacco_15+	3.44		

Furthermore, Outlier detection using boxplots in Figure 2 revealed extreme values in dependent variables, suggesting the need for robust estimation techniques to prevent distortion of model estimates (Gujarati, 2004).

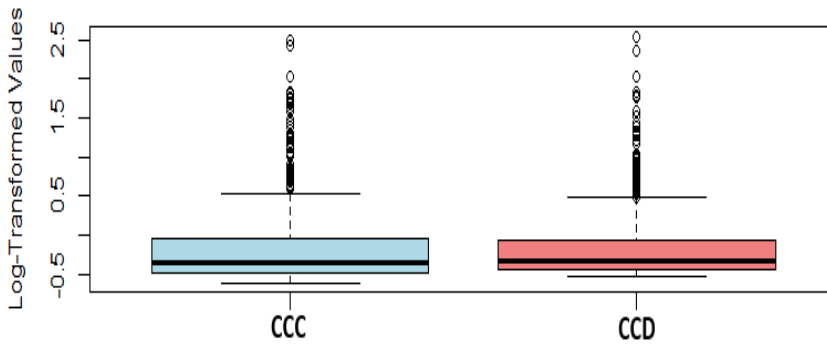


Figure 2. Boxplots of independent variables showing their distribution, variability, and potential outliers across Indian districts.

Together, these findings establish a strong empirical foundation for adopting spatial, regularized, and robust regression models in the subsequent analysis.

3.3 Models Implementation

As highlighted in the data insights section, the dependent variables exhibit strong spatial dependency. In such situations, spatial analysis is commonly approached using the GWR model. The GWR model is widely used because it produces location-specific parameter estimates and effectively captures local spatial heterogeneity.

However, as will be shown in the residuals and multicollinearity diagnostics sections, the GWR model is highly sensitive to both multicollinearity and outliers. To handle these issues separately, RGWR has been used to deal with outliers, while ridge- and LASSO-based extensions such as GWRR and GWL have been applied to mitigate multicollinearity. Despite their usefulness, these models cannot address both challenges simultaneously, a limitation that will also be evident in the diagnostic results.

To overcome this gap, we propose the RGWL model, which integrates Huber weights with LASSO regularization. By jointly handling outliers and multicollinearity, the RGWL model ensures more stable and reliable local parameter estimates. The methodology of RGWL has already been detailed in the Models and Estimators section, and in the following subsections, we compare its performance with existing approaches.

3.3.1 Residuals Diagnostics

To evaluate model adequacy and reliability, residual heteroscedasticity was assessed using the Breusch-Pagan (BP) test (Gujarati, 2004). Table 3 outlines these test statistics and p-values, while Figures 3 and 4 display the corresponding residual boxplots. Together, these statistical and visual diagnostics provide a comprehensive view of how each model handles variance instability.

Table 3. Breusch-Pagan (BP) test results across spatial regression models for CCC and CCD

Metric	CCC					CCD				
	GWR	RGWR	GWRR	GWL	RGWL	GWR	RGWR	GWRR	GWL	RGWL
Test statistic	20.67	9.12	11.89	18.65	5.21	17.89	9.35	10.97	9.12	5.78
p-value	0.001	0.061	0.007	0.033	0.062	0.002	0.002	0.011	0.028	0.051

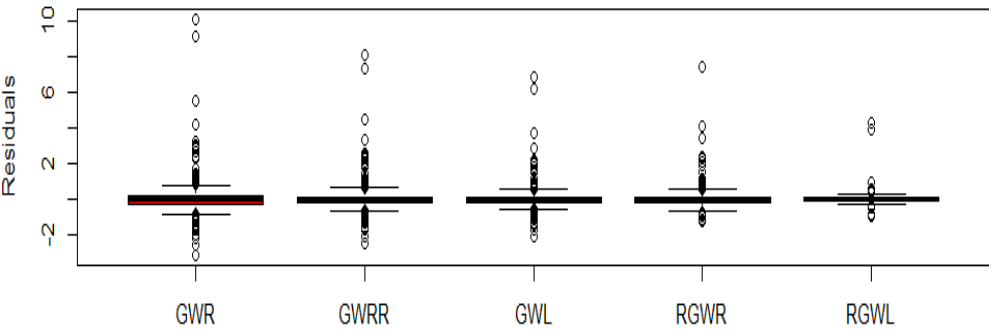


Figure 3. Boxplot visualization of residuals from spatial regression models for CCC, highlighting distribution, variability, and potential outliers.

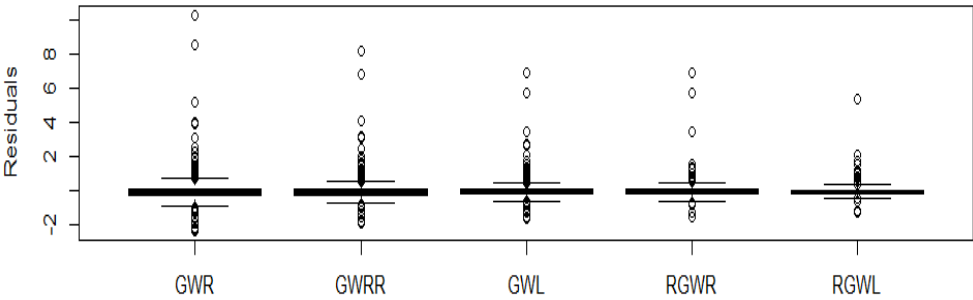


Figure 4. Boxplot visualization of residuals from spatial regression models for CCD, highlighting distribution, variability, and potential outliers.

As seen in Table 3, Breusch–Pagan test results reveal clear differences in how spatial regression models handle heteroscedasticity. The GWR model shows the highest test statistics (20.67 for CCC and 17.89 for CCD) with highly significant p-values (0.001 for CCC and 0.002 for CCD), indicating strong evidence of heteroscedasticity. The GWRR and GWL models also exhibit heteroscedasticity, though with slightly lower test statistics than GWR, suggesting partial improvement but not complete resolution. The RGWR model reduces heteroscedasticity further, as reflected in higher p-values (0.061 for CCC and 0.062 for CCD), yet some concerns remain. Importantly, the RGWL model demonstrates the best performance, yielding the lowest test statistic (5.21 for CCC and 5.78 for CCD) and non-significant p-values, confirming its robustness against heteroscedasticity.

Overall, these findings emphasize that while GWR is most vulnerable to variance instability, the proposed model RGWL provides the most reliable estimates. This robustness is further evident in

the residual distribution visualized through the boxplots presented in Figures 3 and 4.

The residual boxplots in Figures 3 and 4 visually confirm the superior variance stabilization achieved by the proposed framework. While classical GWR yields an unstable error profile with extreme localized outliers, the proposed RGWL model successfully compresses this error distribution. RGWL restricts the maximum outlier and stabilizes the overall residual variance. This symmetric, attenuated outlier profile demonstrates that the integration of robust loss functions with regularization effectively insulates the model against spatial anomalies

3.3.2 Multicollinearity Diagnostics

To evaluate the presence of multicollinearity within local spatial regression models such as GWR, RGWR, GWRR, GWL, and RGWL it is essential to examine it at each geographical location. In this context, the local VIF serves as a key diagnostic tool, as it measures the degree of multicollinearity among predictors in a spatially varying framework (Wheeler, 2007). These diagnostics provide insight into the extent of linear dependence among the independent variables at the local level. Table 4 presents a summary of the local VIF values across locations for all models.

Table 4 reports the minimum, mean, and maximum VIF values for the spatial regression models: GWR, RGWR, GWRR, GWL, and RGWL, providing an assessment of local multicollinearity. In the GWR model, several variables (Undernourished_W, Obese_WObese_W, HTN_Prev, and BloodSugar) show high multicollinearity, with maximum VIFs exceeding 10. The RGWR model exhibits a similar pattern but with slightly lower VIFs due to robust estimation. The GWRR model, which applies ridge penalization, shows marginal reductions in VIFs compared to GWR. The most notable improvements are seen in the GWL and RGWL models, where LASSO regularization substantially reduces VIF values, particularly for variables that previously exhibited high multicollinearity.

Table 4. Summary of local VIF values (minimum, mean, and maximum) across spatial regression models

Variable	GWR	RGWR	GWRR	GWL	RGWL
Pop_U15	3.0 / 3.2 / 3.4	1.3 / 2.3 / 3.0	1.5 / 2.8 / 3.5	2.4 / 2.6 / 2.9	2.2 / 2.4 / 2.7
Elec_HH	1.5 / 1.5 / 1.6	1.1 / 1.3 / 1.5	1.3 / 1.5 / 1.7	1.5 / 1.5 / 1.5	1.4 / 1.5 / 1.5
Sanit_HH	2.5 / 2.7 / 2.9	1.4 / 2.0 / 2.6	1.7 / 2.3 / 2.9	2.0 / 2.2 / 2.4	1.8 / 2.1 / 2.3
CleanFuel	3.3 / 3.4 / 3.5	2.1 / 2.7 / 3.2	2.6 / 3.1 / 3.8	2.8 / 3.0 / 3.1	2.6 / 2.8 / 3.0
HealthIns	1.3 / 1.4 / 1.5	1.0 / 1.2 / 1.4	1.3 / 1.5 / 1.7	1.3 / 1.4 / 1.4	1.2 / 1.3 / 1.4
Women_Lit	6.2 / 6.3 / 6.5	3.5 / 4.3 / 5.1	4.9 / 5.4 / 5.9	4.8 / 4.9 / 5.2	4.5 / 4.7 / 5.0
Women_School	5.0 / 5.2 / 5.5	2.8 / 3.6 / 4.5	4.0 / 4.6 / 5.0	3.9 / 4.0 / 4.3	3.6 / 3.8 / 4.1
Undernourished_W	26.1 / 28.5 / 33.6	24.2 / 27.1 / 33.2	8.7 / 9.5 / 9.9	8.2 / 9.7 / 10.9	7.8 / 9.2 / 9.5
Obese_W	19.4 / 20.1 / 22.4	18.6 / 19.9 / 21.6	8.1 / 9.0 / 9.8	7.4 / 8.0 / 9.6	7.0 / 7.7 / 8.9
Tobacco_15+	1.8 / 1.8 / 1.9	1.3 / 1.6 / 1.9	1.6 / 1.8 / 2.2	1.7 / 1.8 / 1.8	1.6 / 1.7 / 1.8
Alcohol_15+	1.7 / 1.8 / 2.2	1.2 / 1.4 / 1.6	1.5 / 1.6 / 2.1	1.5 / 1.7 / 1.9	1.4 / 1.6 / 1.8
Pop_Density	1.3 / 1.3 / 1.3	1.0 / 1.1 / 1.2	1.2 / 1.3 / 1.4	1.3 / 1.3 / 1.3	1.2 / 1.3 / 1.3
Urban_Pop	2.5 / 2.6 / 2.6	1.6 / 1.9 / 2.1	2.0 / 2.2 / 2.6	2.0 / 2.1 / 2.2	1.9 / 2.0 / 2.1
HealthCenters	1.1 / 1.1 / 1.1	0.9 / 1.0 / 1.1	1.1 / 1.2 / 1.3	1.1 / 1.1 / 1.1	1.0 / 1.1 / 1.1
HTN_Prev	25.9 / 28.1 / 32.7	24.3 / 27.9 / 32.0	9.1 / 9.9 / 9.8	7.9 / 8.5 / 10.1	7.5 / 9.0 / 9.7
BloodSugar	21.8 / 23.4 / 25.0	20.2 / 23.6 / 24.8	8.0 / 8.8 / 9.9	6.8 / 7.9 / 8.9	6.4 / 7.5 / 8.6
Avg_Temp	1.8 / 2.0 / 2.3	1.2 / 1.6 / 1.9	1.6 / 2.0 / 2.5	1.6 / 1.8 / 2.0	1.5 / 1.7 / 1.9
Rel_Humidity	2.4 / 2.6 / 2.8	1.5 / 1.9 / 2.4	2.1 / 2.5 / 3.1	1.9 / 2.1 / 2.3	1.7 / 2.0 / 2.2
Poverty_Pop	2.5 / 2.7 / 2.9	1.6 / 1.9 / 2.3	2.2 / 2.6 / 3.3	2.0 / 2.2 / 2.4	1.8 / 2.1 / 2.3

The residual and multicollinearity diagnostics demonstrate the robustness of the proposed model in handling both issues simultaneously. The performance of the proposed RGWL model is then compared with existing models using performance metrics in the following section.

3.3.3 Models Comparison

The performance of GWR, RGWR, GWRR, GWL, and RGWL using adjusted (adj.) R^2 , Root mean square error (RMSE), AIC, and BIC. adj. R^2 accounts for model explanatory power with predictor adjustment, RMSE quantifies predictive accuracy, and AIC/BIC provide information-based criteria for balancing fit and complexity. Higher adj. R^2 and lower RMSE, AIC, and BIC indicate superior performance. The results are summarized in Table 5, highlighting performance variations across models.

Table 5. Performance evaluation of spatial regression models for CCC and CCD

	GWR	RGWR	GWRR	GWL	RGWL
CCC					
adj R^2	0.45	0.71	0.78	0.83	0.87
RMSE	0.7921	0.5205	0.5322	0.4130	0.4017
AIC	398.25	360.93	354.10	341.56	336.41
BIC	415.16	378.12	371.44	359.27	354.02
CCD					
adj R^2	0.53	0.70	0.82	0.86	0.89
RMSE	0.6978	0.4896	0.4597	0.4398	0.4132
AIC	377.82	342.17	330.75	325.93	318.12
BIC	393.74	359.09	347.62	343.71	336.10

As shown in Table 5, the proposed RGWL model demonstrates the strongest performance among all models considered. It achieves the highest adj. R^2 values for both CCC (0.87) and CCD (0.89), indicating superior explanatory power with lowest RMSE (0.4017 for CCC and 0.4132 for CCD). The performance of the RGWL model reflects its ability to balance prediction accuracy with resistance to outliers and local heterogeneity, which are particularly relevant in spatial epidemiological modeling.

3.4 Findings from the proposed model RGWL

In the RGWL modeling framework, three key parameters bandwidth (bw), regularization parameter (λ), and the Huber threshold (δ) were selected using a CV approach to ensure optimal model performance for both CCC and CCD. For the CCC model, the CV procedure identified a bandwidth of 0.15, which enables the model to capture finer spatial variations. The regularization parameter was selected as $\lambda = 0.12$, promoting shrinkage of less influential predictors towards zero, thus enhancing model parsimony and interpretability. The Huber threshold was set at $\delta = 1.345$ to provide a balanced trade-off between robustness to outliers and estimation efficiency.

For the CCD model, the optimal bandwidth determined through CV was 0.32, which captures broader spatial trends. A slightly smaller regularization strength of $\lambda = 0.08$ was chosen to retain more relevant predictors in the model. The Huber threshold was increased to $\delta = 1.75$ to accommodate potentially extreme residuals in the mortality data.

These CV-optimized parameters allowed the RGWL model to effectively address spatial heterogeneity, local multicollinearity, and sensitivity to outliers, leading to improved predictive accuracy and robustness in both CCC and CCD modeling contexts.

Similar to GWR, the RGWL model also produces location-specific coefficient estimates, reflecting the spatial heterogeneity inherent in the data. Each spatial unit is assigned its own set of regression coefficients, influenced by the local context defined through the calibrated bandwidth and regularization parameters. Given this spatially varying nature, it becomes essential to summarize the distribution of these local coefficients to gain insights into their overall behavior across the study area.

The table 6 presents the summary statistics of the RGWL parameter estimates, including the minimum, first quartile (Q1), median, mean, third quartile (Q3), and maximum values across districts. These values help in understanding both the magnitude and the spatial distribution of coefficients. A coefficient equal to zero in Q1, the median, or Q3 indicates that the corresponding variable was excluded in at least 25%, 50%, or 75% of the districts, respectively. Similarly, if the minimum and maximum values are also zero, it implies that the variable was entirely excluded across all districts. Conversely, when coefficients are nonzero, their sign and magnitude reflect the nature (positive or negative) and strength of the variable’s contribution in the districts where it was selected.

Table 6. Summary statistics of RGWL parameter estimates for COVID-19 outcomes (CCC and CCD) across Indian districts, including measures of central tendency and variability

Variable	CCC						CCD					
	Min.	1st	Median	Mean	3rd	Max.	Min.	1st	Median	Mean	3rd	Max.
Intercept	-2.015	-0.120	0	0.031	0	3.762	-2.137	-0.115	0	0.021	0	4.102
Pop_U15	-0.014	0	0	0	0	0.101	-0.035	0	0	-0.002	0	0.096
Elec_HH	-0.011	0	0	0	0	0	-0.021	0	0	0	0	0
Sanit_HH	-0.020	0	0	0.001	0	0.114	-0.109	0	0	-0.008	0	0
CleanFuel	0	0.009	0.031	0.030	0.047	0.087	0	0.095	0.122	0.106	0.139	0.178
Women_Lit	-0.069	0	0	-0.001	0	0.024	0	0	0.030	0.041	0.063	0.285
Women_School	-0.042	0	0	-0.001	0	0.030	-0.012	0	0.014	0.017	0.029	0.120
Undernourished_W	0	0.061	0.089	0.099	0.147	0.296	-0.541	0	0	-0.012	0	0
Obese_W	-0.472	0	0	0.002	0	0.056	-0.339	0	0	-0.004	0	0
HealthIns	-0.396	0	0	-0.011	0	0	-0.116	0	0	-0.001	0	0
Tobacco_15+	0	0	0	0.001	0	0.035	0	0	0	0	0	0
Alcohol_15+	-0.001	0	0	0	0	0	0	0	0	-0.001	0	0
Pop_Density	-0.030	0	0	-0.002	0	0	0	0	0	0.001	0	0.029
Urban_Pop	0	0	0	0.001	0	0.019	0.018	0.033	0.048	0.060	0.072	0.097
HealthCenters	0	0	0	0.001	0	0	0	0	0	0.001	0	0.044
BloodSugar	-0.003	0	0	-0.001	0	0	0	0	0	0.001	0	0.027
Avg_Temp	0	0	0	0.001	0	0.023	-0.017	0	0	0	0	0.012
Rel_Humidity	-0.019	0	0	0	0	0.017	-0.006	0	0	0	0	0.006
Poverty_Pop	-0.007	0	0	0	0	0.009	-0.025	0	0	-0.002	0	0
HTN_Prev	-0.027	0	0	-0.001	0	0	-0.006	0	0	-0.001	0	0

For instance,Pop_U15 has a minimum value that is negative and a maximum value that is positive, indicating that in some districts it exerts a negative impact, while in others it has a positive impact. Its first quartile, median, and third quartile are all equal to zero, suggesting that at least half of the districts did not retain this variable in the model. Elec_HH shows zero values for Q1, median, Q3, and maximum, suggesting that it was excluded from at least 75% of the districts, and in the districts where it was retained, the coefficients carried a negative sign, reflecting an adverse impact. Similarly, Urban_Pop has zero values for the minimum, Q1, median, and Q3, implying that at least 75% of the districts excluded it. However, in those districts where it was included, the coefficients were positive, pointing to a favorable association with CCC.

Building on this interpretation, household conditions such as CleanFuel exhibit small but consistently positive coefficients, reflecting modest yet favorable local effects on COVID-19 cases and

deaths. HealthIns generally shows negative coefficients, especially in CCD, suggesting a protective effect, though the estimates vary by region. Education-related variables (Women_Lit and Women_School) display slightly positive coefficients in CCD, indicating that districts with higher female education may experience better awareness, testing, or reporting of COVID-19 mortality. Nutritional indicators (Undernourished_W and Obese_W) contribute minimally overall, though Obese_W exhibits negative coefficients in many areas, potentially reflecting confounding factors such as under reporting or healthcare disparities.

Substance use variables (Tobacco_15+ and Alcohol_15+) show regionally varying effects, with some districts experiencing positive associations between tobacco or alcohol use and COVID-19 outcomes. In contrast, Pop_Density and Urban_Pop show consistently strong positive effects, particularly on CCD, reinforcing that urban and densely populated areas were more severely impacted. HealthCenters, measured by the number of health centers, has negligible coefficients, suggesting that facility count alone does not adequately capture healthcare effectiveness or accessibility.

Chronic conditions like HTN_Prev and BloodSugar are positively associated with both CCC and CCD, supporting their established role as risk factors for severe COVID-19 outcomes. Environmental variables (Avg_Temp and Rel_Humidity) show small or near-zero coefficients, indicating limited spatial influence in this context. Finally, socioeconomic status, captured by Poverty_Pop, displays mostly negative associations, particularly with CCC, possibly due to underreporting or reduced exposure in rural poor populations.

The variability observed in the RGWL coefficient estimates highlights substantial spatial heterogeneity, indicating that the effects of predictors differ markedly across districts. The range between minimum and maximum values captures this regional variation, while the median values provide a measure of central tendency. These findings demonstrate the advantage of robust and spatially adaptive models like RGWL, which not only capture non-stationary relationships but also reduce the influence of outliers and local noise. Additionally, the RGWL model addresses multicollinearity through LASSO regularization, allowing for automatic variable selection. This facilitates the identification of the most relevant predictors in each district, thereby enhancing model interpretability and stability. As a result, district-specific sets of significant covariates are identified that explain the variation in COVID-19 cases and deaths. The model shows strong explanatory performance, with high local adjusted R^2 values across most districts, reflecting its ability to capture localized pandemic dynamics effectively.

On average, the RGWL model explains approximately 87% and 89% of the variation in CCC and CCD (see Figure 5), respectively, although this explanatory power varies across regions.

Figures 5 and 6 illustrate the local adjusted R^2 values, indicating the proportion of variance accounted for by the selected variables in each spatial unit. Figures 5 and 6 show the local adj. R^2 values for CCC and CCD, ranging from 0.7012–0.9025 for CCC and 0.7210–0.9250 for CCD, with mean values of 0.8240 and 0.8912, respectively. These results indicate strong model performance across districts, with higher explanatory power for deaths, suggesting the RGWL model captures spatial variation in mortality more effectively. Minimum values confirm robustness even in less favorable regions, while the interquartile ranges demonstrate that most districts exhibit consistently high model fit.

Overall, these findings underscore the advantage of the RGWL model in capturing spatially varying effects while mitigating noise and outliers. The coefficient distributions reflect real regional disparities in COVID-19 dynamics and validate the use of a robust, geographically weighted framework for modelling health outcome in diverse spatial settings.

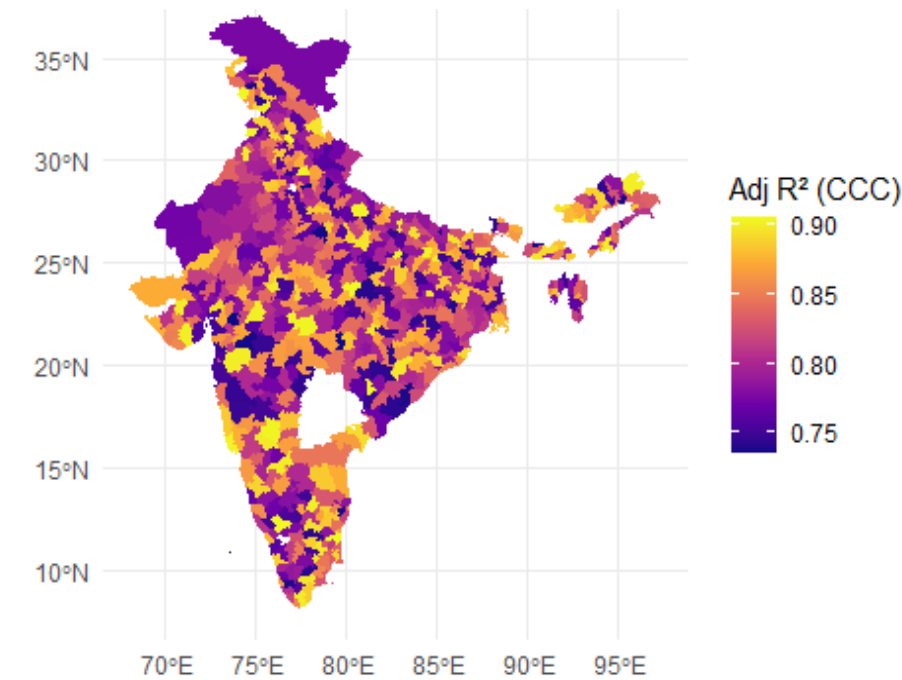


Figure 5. Choropleth maps showing the spatial distribution of local adjusted R^2 values for the RGWL model (CCC).

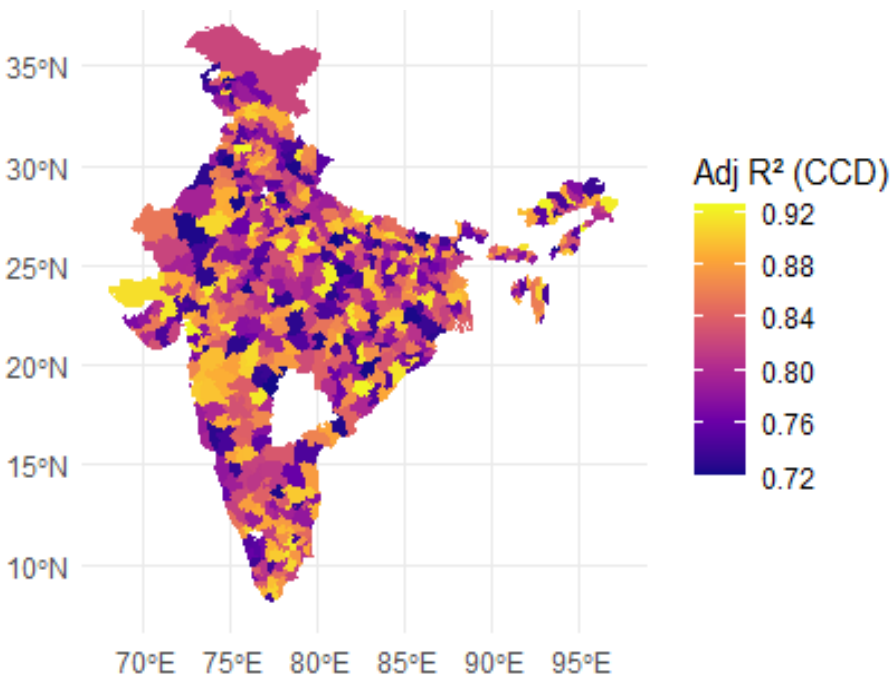


Figure 6. Choropleth maps showing the spatial distribution of local adjusted R^2 values for the RGWL model (CCD).

4. Discussion

The empirical results from the proposed RGWL model provide critical insights into the socio-demographic, clinical, and environmental determinants of COVID-19 across Indian districts by effectively capturing severe spatial non-stationarity. The substantial variation between the minimum, median, and maximum coefficient values highlights how localized relationships change across regions, which is a phenomenon that traditional global models fail to capture due to local multicollinearity and outlier vulnerability (Wheeler, 2007; Wheeler, 2009; Wulandari *et al.*, 2019). RGWL addresses these vulnerabilities through its sparse variable selection mechanism, where variables like Pop_U15, Elec_HH, and Urban_Pop were shrunk to exactly zero across specific percentiles. This automatic exclusion ensures that a covariate is only retained if its localized explanatory power outweighs the penalty, a strategy heavily supported by recent biometric and spatial modeling literature for managing highly collinear predictors without overfitting (NEISSE *et al.*, 2021; Carrilho & Lopes, 2022; Barros *et al.*, 2024).

Additionally, risk factors such as Pop_Density and Urban_Pop and chronic clinical comorbidities (HTN_Prev and BloodSugar) emerged as highly significant predictors, exerting consistent and powerful positive impacts on COVID-19 outcomes. These significant findings align with numerous empirical investigations establishing high-density environments and metabolic risk profiles as the primary drivers of severe pandemic waves (Rocklöv & Sjödin, 2020; Sridhar, 2023; Dutta *et al.*, 2021). Conversely, Poverty_Pop displayed unexpected negative associations, reflecting structural disparities in healthcare access and diagnostic under-reporting among rural populations rather than a protective biological effect (Dutta *et al.*, 2021). Structural metrics such as HealthCenters yielded negligible coefficients, proving that raw facility counts do not equate to operational capacity. Meanwhile, education-related variables (Women_Lit and Women_School) showed slight positive trends with mortality, pointing to a reporting bias where highly literate districts engaged in more proactive data tracking, a phenomenon similarly documented by (CARMO *et al.*, 2021). Behavioral factors displayed high spatial instability across regions, while environmental variables remained near zero, proving that human mobility ultimately overrode localized climatic effects. Overall, high local adjusted R^2 values confirm that RGWL successfully isolates true localized pandemic trends from spatial noise.

5. Conclusions

In this study, we proposed and implemented the RGWL model to analyze the spatial determinants of COVID-19 cumulative cases and deaths across districts in India. By incorporating Huber loss into the GWL framework, the RGWL model combines local adaptiveness with robustness against outliers and the ability to manage multicollinearity through regularization.

Empirical comparisons revealed that RGWL consistently outperforms the conventional GWR, RGWR and GWL models. Specifically, it achieved lower RMSE and higher adjusted R-squared, indicating improved predictive accuracy. The model also exhibited greater resilience to influential observations and reduced local multicollinearity, as confirmed by lower local VIFs and condition indices. Residual analysis, including boxplot visualization and BP test, further reinforced the robustness of RGWL in handling extreme deviations.

The spatially varying relationships captured by the RGWL model revealed distinct regional patterns in the influence of demographic, health, socioeconomic, and environmental variables. These findings underscore the importance of using robust spatial regression techniques for analyzing geographically heterogeneous epidemiological data.

Overall, the RGWL model offers a reliable and interpretable approach for spatial data modeling in the presence of outliers and multicollinearity. It provides a valuable extension of existing spatial

regression methods and holds potential for broader application in public health and environmental research contexts.

Acknowledgments

The authors sincerely acknowledge the Department of Mathematics and Statistics, Banasthali Vidyapith, for providing the necessary facilities, academic environment, and continuous support that made this research work possible.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Megha Sharma: Concept Development, Methodology, Writing – original draft; **Shalini Chandra:** Concept Development, Supervision; **Gargi Tyagi:** Validation.

References

1. Adekunle, I. A., Onanuga, A. T., Akinola, O. O. & Ogunbanjo, O. W. Modelling spatial variations of coronavirus disease (COVID-19) in Africa. *Science of the Total Environment* **729**, 138998. doi:10.1016/j.scitotenv.2020.138998 (2020).
2. Ahmed, F., Ahmed, N., Pissarides, C. & Stiglitz, J. Why inequality could spread COVID-19. *The Lancet Public Health* **5**, e240. doi:10.1016/S2468-2667(20)30085-2 (2020).
3. Alfons, A., Croux, C. & Gelper, S. Sparse least trimmed squares regression for analyzing high-dimensional large data sets. *Annals of Applied Statistics* **7**, 226–248. doi:10.1214/12-AOAS575 (2013).
4. Appiah-Otoo, I. & Kursah, M. B. Modelling spatial variations of novel coronavirus disease (COVID-19): evidence from a global perspective. *GeoJournal* **87**, 3203–3217. doi:10.1007/s10708-021-10427-0 (2022).
5. Barros, A. C. A. V. B. d., Silva, M. A., Marques, R. D., Lobos, C. M. V. & Luciano, A. C. d. S. Classification of sugarcane areas in Landsat images using machine learning algorithms. *Brazilian Journal of Biometrics* **42**, 147–157. doi:10.28951/bjb.v42i2.693 (2024).
6. Bashir, M. F., Ma, B., Komal, B., Bashir, M. A., Tan, D., Bashir, M., *et al.* Correlation between climate indicators and COVID-19 pandemic in New York, USA. *Science of the Total Environment* **728**, 138835. doi:10.1016/j.scitotenv.2020.138835 (2020).
7. Brunsdon, C., Fotheringham, S. & Charlton, M. Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)* **47**, 431–443. doi:10.1111/j.1538-4632.1996.tb00936.x (1998).
8. CARMO, C. R. S., de Lima CANEPPELE, F. & NUNES, F. C. Analysis of covid-19 contamination and deaths cases in brazil according to the newcomb-benford. *Brazilian Journal of Biometrics* **39**, 522–535. doi:10.28951/rbb.v39i4.535 (2021).
9. Carrilho, J. & Lopes, M. B. Classification and biomarker selection in lower-grade glioma using robust sparse logistic regression applied to RNA-seq data. *Brazilian Journal of Biometrics* **40**, 371–381. doi:10.28951/bjb.v40i4.634 (2022).

10. Chen, B., Liang, H., Yuan, X., Hu, Y., Xu, M., Zhao, Y., Zhang, B., Tian, F. & Zhu, X. Roles of meteorological conditions in COVID-19 transmission on a worldwide scale. *MedRxiv*, 2020–03. doi:10.1101/2020.03.16.20037168 (2020).
11. Chung, H. W. *et al.* Effects of government policies on the spread of COVID-19 worldwide. *Scientific Reports* **11**, 20495. doi:10.1038/s41598-021-99368-9 (2021).
12. Cliff, A. D. & Ord, J. K. *Spatial Processes: Models & Applications* (Pion, London, 1981).
13. Dutta, I., Basu, T. & Das, A. Spatial analysis of COVID-19 incidence and its determinants using spatial modeling: A study on India. *Environmental Challenges* **4**, 100096. doi:10.1016/j.envc.2021.100096 (2021).
14. Fan, J., Fan, Y. & Barut, E. Adaptive robust variable selection. *Annals of Statistics* **40**, 1748–1772. doi:10.1214/13-AOS1191 (2012).
15. Foody, G. M. Local characterization of thematic classification accuracy through spatially constrained confusion matrices. *International Journal of Remote Sensing* **24**, 1217–1225. doi:10.1080/01431160512331326521 (2003).
16. Gujarati, D. N. *Basic Econometrics* 4th (McGraw-Hill, New York, 2004).
17. Khan, M., Javed, M. & Ahmed, S. A robust penalized regression estimator for simultaneous variable selection and outlier detection. *Statistical Papers* **63**, 1151–1173. doi:10.1007/s00362-021-01304-7 (2022).
18. NEISSE, A. C., de OLIVEIRA, F. L. P., de OLIVEIRA, A. C. S., NETO, R. M. N., *et al.* Chronic fatigue syndrome and its relation with absenteeism: elastic-net and stepwise applied to biochemical and anthropometric clinical measurements. *Brazilian Journal of Biometrics* **39**, 221–239. doi:10.28951/rbb.v39i1.533 (2021).
19. Pereira, M. & Oliveira, A. M. Poverty and food insecurity may increase as the threat of COVID-19 spreads. *Public health nutrition* **23**, 3236–3240. doi:10.1017/S1368980020003493 (2020).
20. Ramírez-Aldana, R., Gomez-Verjan, J. C. & Bello-Chavolla, O. Y. Spatial analysis of COVID-19 spread in Iran: Insights into geographical and structural transmission determinants at a province level. *PLoS neglected tropical diseases* **14**, e0008875. doi:10.1371/journal.pntd.0008875 (2020).
21. Rocklöv, J. & Sjödin, H. High population densities catalyse the spread of COVID-19. *Journal of travel medicine* **27**, taaa038. doi:10.1093/jtm/taaa038 (2020).
22. Sardar, S., Abdul-Khaliq, I., Ingar, A., Amaidia, H. & Mansour, N. COVID-19 lockdown: A protective measure or exacerbator of health inequalities? A comparison between the United Kingdom and India. a commentary on “the socio-economic implications of the coronavirus and COVID-19 pandemic: A review. *International Journal of Surgery (London, England)* **83**, 189. doi:10.1016/j.ijssu.2020.09.044 (2020).
23. Sarkar, S. K., Ekram, K. M. M. & Das, P. C. Spatial modeling of COVID-19 transmission in Bangladesh. *Spatial Information Research*, 1–12. doi:10.1007/s41324-021-00387-5 (2021).
24. Sridhar, K. S. Urbanization and COVID-19 Prevalence in India. *Regional Science Policy & Practice* **15**, 493–505. doi:10.1111/rsp3.12503 (2023).
25. Tosepu, R., Gunawan, J., Effendy, D. S., Lestari, H., Bahar, H., Asfian, P., *et al.* Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. *Science of the total environment* **725**, 138436. doi:10.1016/j.scitotenv.2020.138436 (2020).

26. Wang, H., Li, G. & Jiang, G. Robust regression shrinkage and consistent variable selection through the LAD-Lasso. *Journal of Business & Economic Statistics* **25**, 347–355. doi:10.1198/073500106000000251 (2007).
27. Wheeler, D. C. Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environment and planning A* **39**, 2464–2481. doi:10.1068/a38325 (2007).
28. Wheeler, D. C. Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted lasso. *Environment and planning A* **41**, 722–742. doi:10.1068/a40256 (2009).
29. Wulandari, P. P., Djuraidah, A. & Wigena, A. H. Robust Geographically Weighted Regression Modeling using Least Absolute Deviation and M-Estimator. *International Journal of Scientific Research in Science, Engineering and Technology* **6**, 238–245. doi:10.32628/IJSRSET196123 (2019).
30. Zhang, H. & Mei, C. Local least absolute deviation estimation of spatially varying coefficient models: robust geographically weighted regression approaches. *International Journal of Geographical Information Science* **25**, 1467–1489. doi:10.1080/13658816.2010.528420 (2011).