Spatial Regression Analysis of Leprosy in East Java 2025 Using Spatial Autoregressive and Spatial Error Models with Queen Contiguity

Nilna Meis Shafira[(1)], Danang Ariyanto[(2)]

Program Studi S1 Matematika, FMIPA, Universitas Negeri Surabaya

Jl. Ketintang, Ketintang, Kec. Gayungan, Surabaya City, East Java 60231
e-mail: nilna.22089@mhs.unesa.ac.id[(1)], danangariyanto@unesa.ac.id[(2)]

ABSTRAK

Kusta masih menjadi salah satu permasalahan kesehatan masyarakat di Indonesia, termasuk di Provinsi Jawa Timur yang memiliki jumlah kasus relatif tinggi. Perbedaan karakteristik sosial, ekonomi, dan kesehatan antarwilayah dapat menyebabkan variasi kasus kusta serta membentuk keterkaitan spasial. Berbeda dengan penelitian sebelumnya yang umumnya menggunakan jumlah kasus kusta absolut dan hanya menerapkan satu model regresi spasial, penelitian ini membandingkan model *Spatial Autoregressive* (SAR) dan *Spatial Error Model* (SEM) menggunakan *New Case Detection Rate* (NCDR) sebagai variabel respons. Penelitian ini bertujuan menganalisis faktor-faktor yang memengaruhi NCDR kusta di Jawa Timur tahun 2025 menggunakan pembobot *Queen Contiguity*. Variabel prediktor meliputi kepadatan penduduk, persentase rumah tangga tanpa fasilitas buang air besar, persentase penduduk miskin, dan jumlah puskesmas. Data yang digunakan merupakan data sekunder dari 38 kabupaten/kota di Jawa Timur. Hasil analisis menunjukkan adanya autokorelasi spasial positif yang signifikan dengan nilai Moran's I sebesar 0,6059. Persentase penduduk miskin merupakan satu-satunya variabel yang berpengaruh signifikan terhadap NCDR kusta. Berdasarkan nilai *Akaike Information Criterion* (AIC) dan *log-likelihood*, model SAR dipilih sebagai model terbaik. Temuan ini menunjukkan bahwa pengendalian kusta perlu didukung oleh intervensi sosial-ekonomi, khususnya melalui upaya pengurangan kemiskinan.
Kata kunci: Kusta; Regresi Spasial; *Spatial Autoregressive Model*; *Spatial Error Model*; *Queen Contiguity*

ABSTRACT

Leprosy remains a public health issue in Indonesia, including in East Java, which has a relatively high number of cases. Differences in social, economic, and health characteristics between regions can cause variations in leprosy cases and spatial clustering. Unlike previous studies, which generally used absolute case numbers and applied one spatial regression model, this study compares spatial autoregressive (SAR) and spatial error (SEM) models using the new case detection rate (NCDR) as the response variable. The study uses Queen Contiguity weighting to analyze the factors that influence the NCDR of leprosy in East Java in 2025. The predictor variables include population density, the percentage of households without toilet facilities, the percentage of impoverished residents, and the number of health centers. Secondary data from 38 districts/cities in East Java were used for the analysis. The analysis revealed significant positive spatial autocorrelation, with a Moran's I value of 0.6059. Of the predictor variables, the percentage of impoverished residents was the only one that significantly affected the NCDR of leprosy. Based on Akaike's information criterion (AIC) and log-likelihood values, the SAR model was selected as the best model. These findings suggest that controlling leprosy requires social and economic interventions, particularly poverty reduction efforts.
Keywords: Leprosy; Spatial Regression*; Spatial Autoregressive Model; Spatial Error Model; Queen Contiguity*

INTRODUCTION

Global health issues still face high rates of various infectious diseases, one of which is leprosy. Leprosy is a chronic infectious disease caused by Mycobacterium leprae that affects the skin and peripheral nerves [1]. It is a serious public health concern because new cases are reported every year in large numbers. According to the World Health Organization, most of the new leprosy cases detected each year originate from the Southeast Asia region, making it a primary focus in efforts to control the disease [2].

Globally, Indonesia, India, and Brazil are the countries with the highest number of new leprosy cases. Each of these countries reports more than 10,000 new cases per year, while some other countries still record more than 1,000 new cases per year [2]. Among these countries, Indonesia still faces a high burden of leprosy, with around 17,000-20,000 new cases reported annually. At the regional level, East Java is one of the Indonesian provinces with the highest number of new leprosy cases, accounting for around 15.9% of the national total, according to the East Java Provincial Health Office's 2024 data [3]. The high incidence rate indicates that leprosy is still spreading and poses a challenge to controlling the disease. If not treated properly, leprosy can cause permanent disability [4].

The distribution and occurrence of leprosy are influenced by social, economic, environmental, and health service factors. Higher population density increases contact between individuals, raising the risk of disease transmission [5]. Poor sanitation and environmental conditions can also increase the risk of leprosy in a community [6]. Economic factors, such as poverty, play a role because they are related to limited access to health services and a lower quality of life [7]. On the other hand, the availability of health facilities, such as health centers, plays a crucial role in detecting and addressing leprosy cases early on. Differences in social, economic, and health conditions between regions can cause variations in the number of leprosy cases in each district and create spatial links between regions. A study by Bueno et al. (2023) shows that leprosy cases are not distributed evenly and tend to form spatial clusters in certain areas [8]. These findings suggest that a spatial analysis approach is necessary to more accurately identify the factors influencing the occurrence of leprosy in each region.

Previous studies have used spatial regression approaches to model health data based on regions. Akolo (2022) showed that the Queen Contiguity weight and the SEM model are the best combination for modeling stunting cases [9]. Meanwhile, Pratiwi et al. (2023) demonstrated that the SEM model outperformed the SAR model in modeling diarrhea cases in Bali Province [10]. In the context of leprosy, Rahmawati and Bimanto (2021) used an SAR model to analyze factors influencing leprosy cases in East Java, demonstrating spatial dependence between regions [11]. Their results showed that average years of schooling and the number of male residents significantly influenced the incidence of leprosy in East Java. The SAR model with Queen contiguity weights has also been used to analyze population dependency ratios between districts, showing significant spatial influence between regions[12].

However, previous research has shown that the best spatial regression model can differ for each case and disease indicator used, as demonstrated by the differing results between Akolo [9], who found that SEM was more effective for stunting cases, and Pratiwi et al [10]. who also found that SEM was more effective than SAR for diarrhoea cases. Meanwhile, the leprosy study by Rahmawati and Bimanto [11] only applied the SAR model without comparing it with the SEM, so

a comparative evaluation of the SAR and SEM models in modelling leprosy cases in East Java has not yet been conducted. Furthermore, the response variable used in the study was the absolute number of leprosy cases, which is influenced by differences in population size between regions, thus failing to reflect the relative rate of new case detection. These differences show that the choice of spatial model cannot be generalised for leprosy cases, so a study is needed that specifically compares the performance of SAR and SEM in this context.

Therefore, this study has two novelties: (1) explicitly comparing SAR and SEM models for leprosy cases in East Java, which has not been applied in previous leprosy case studies [11], and (2) using the New Case Detection Rate (NCDR) as the response variable, which illustrates the rate of new leprosy case detection per population unit, making it more representative of early detection performance than the use of absolute case numbers in previous studies. In addition to comparing the performance of the two models, this study also provides empirical evidence of the performance of the SAR and SEM models in modelling the NCDR of leprosy in East Java, thus enriching spatial epidemiological studies and serving as a reference for selecting spatial regression models for the study of infectious diseases in similar geographical areas. Furthermore, the results of this study are expected to inform the development of more effective and targeted policies for the control of leprosy.

METHOD

Data and Variables

This study employed a quantitative approach using spatial statistical analysis to examine leprosy cases in East Java. The data consisted of the New Case Detection Rate (NCDR) of leprosy and several socio-economic factors across 38 districts and municipalities in East Java. Secondary data for the year 2025 were obtained from the official publication of Statistics Indonesia (BPS) of East Java Province, namely *East Java Province in Figures 2026* [13]. The analysis was conducted using the Spatial Autoregressive Model (SAR) and Spatial Error Model (SEM) with Queen Contiguity spatial weights. The variables used in this study consisted of one response variable (Y) and four predictor variables (X). The entire data processing and analysis process was carried out using the R software, utilising the sf and tmap packages for spatial data processing and visualisation, the spdep package for forming spatial weight matrices and testing spatial autocorrelation (Moran's I and Lagrange Multiplier), the spatialreg package for estimating SAR and SEM models, and the lmtest and car packages for testing the assumptions of classical regression (heteroskedasticity and multicollinearity). The description of the study variables is presented in Table 1.

Table 1. Study Variables

Variable	Description	Unit/Scale
Y	New Case Detection Rate (NCDR) of Leprosy	Per 100,000 population
X_1	Population Density	Persons/km²
X_2	Households Without Toilet Facilities	Percent (%)
X_3	Poverty Rate	Percent (%)
X_4	Number of Community Health Centers (*Puskesmas*)	Unit

In this study, the response variable is the new case detection rate (NCDR) of leprosy. The predictor variables are population density, percentage of households without toilet facilities, percentage of impoverished residents, and number of health centers. The analysis stages were performed in the following order:

1. Conduct a descriptive statistical analysis to provide an overview of the characteristics of the research data.
2. Form a global linear regression model using the ordinary least squares (OLS) method as the initial model.
3. Check the OLS model for:
   a. Perform an F-test to determine the significance of the model as a whole.
   b. A t-test to determine the significance of each independent variable on the dependent variable.
   c. Test for normality of residuals using the Shapiro-Wilk method.
   d. Multicollinearity testing using the variance inflation factor (VIF) value.
   e. A heteroskedasticity test using the Breusch-Pagan test.
4. Create a spatial weight matrix using the Queen contiguity approach to illustrate the relationship between regions.
5. Identify the presence of spatial autocorrelation using the Moran's I index.
6. Conduct a Lagrange multiplier (LM) test to determine the most suitable spatial regression model.
7. Estimate the parameters of the spatial regression model using:
   a. Spatial Autoregressive Model (SAR)
   b. Spatial Error Model (SEM).
8. Autocorrelation spatial testing of the residuals of the SAR and SEM models was performed using Moran's I as a diagnostic evaluation to ensure that spatial autocorrelation had been adequately accommodated in each model.
9. Perform a significance test on the parameters using the maximum likelihood ratio test (MLRT) and the Wald test.
10. The best model is determined based on the Akaike information criterion (AIC).

Spatial Weight Matrix

A spatial weight matrix is used to represent neighborhood relationships among regions in spatial analysis. This study employed the Queen Contiguity spatial weight matrix, which defines two regions as neighbors if they share either a common boundary or a common vertex [14]. This approach was chosen because Queen Contiguity considers not only the intersection of sides between regions, as in Rook Contiguity, but also the intersection of corners. Therefore, the relationship between districts and cities can be depicted more accurately. Meanwhile, distance-based weights are not used because this study focuses on administrative relationships between regions rather than geographical proximity. The elements of the spatial weight matrix are defined as follows:

$$w_{ij} \begin{cases} 1, & \text{if region i and region j are neighbors} \\ 0, & \text{if region i and region j are not neighbors} \end{cases}$$

where w_{ij} denotes the element of the spatial weight matrix representing the neighborhood relationship between region i and region j. The spatial weight matrix is then row-standardised by normalising the weights in each row such that the sum of the weights for each region across all neighbouring regions equals one. This standardisation produces proportional spatial weights, enabling spatial influences to be compared consistently across regions despite differences in the number of neighbouring districts and municipalities under the Queen Contiguity approach.

Spatial Autocorrelation

Spatial autocorrelation shows the relationship between the values of a variable in areas that are geographically close to each other. The presence of spatial autocorrelation can cause the assumption of independence between observations to be violated, so it needs to be considered in spatial data analysis. In this study, spatial autocorrelation was tested using Moran's I and the Lagrange Multiplier (LM) test [15].

a. Moran's I

Moran's I is used to identify the presence of global spatial autocorrelation in the data. The Moran's I value can be expressed as follows:

$$\boldsymbol{I} = \frac{n}{S_0}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} wij(y_i-\bar{y})(y_j-\bar{y})}{\sum_{j=1}^{n}(y_i-\bar{y})^2}$$

where $n$denotes the number of observed regions, y_i represents the observation value in region i, $\bar{y}$ is the sample mean, w_{ij} denotes the element of the spatial weight matrix, and S_0 represents the sum of all spatial weights, where $S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} wij$

Moran's I ranges from −1 to 1. A positive Morans'I value $(0 < I \leq 1)$indicates positive spatial autocorrelation, whereas a negative Moran's Index value $(-1 \leq I < 0)$ indicates negative spatial autocorrelation. If Moran's Index equals zero $(I = 0)$, it can be concluded that no spatial autocorrelation exists in the data.

The hypothesis used is as follows:

$H_0: I = 0$, This means that there is no spatial autocorrelation in the data

$H_1: I \neq 0$, This means that there is spatial autocorrelation in the data

The Z test statistic is defined as follows:

$$Z(I) = \frac{I - E(I)}{\sqrt{Var(I)}}$$

$$E(I) = -\frac{1}{n-1}\,;\; Var(I) = \frac{n^2S_1 - nS_2 + 3S_0^2}{(n^2-1)S_0^2} - [E(I)]^2$$

$$S_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(wij + wji)^2;\; S_2 = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} wij + \sum_{j=1}^{n} wij\right)^2$$

where I is Moran's I, $E(I)$ is the expected value of Moran's I, and $Var(I)$ is the variance of Moran's I. The null hypothesis H_0 is rejected if $|\, Z(I)\,| > Z_{\alpha/2}$ or if the *p-value* < α. Rejection of H_0 indicates the presence of spatial autocorrelation among regions.

b. Lagrange Multiplier

Spatial effect testing was performed using the Lagrange Multiplier (LM) test to identify spatial dependence in the data and determine the appropriate spatial regression model [16]. The test was performed using LM Lag and LM Error. LM Lag is used to detect spatial lag dependence, which leads to the use of a Spatial Autoregressive Model (SAR), while LM Error is used to detect spatial error dependence, which leads to the use of a Spatial Error Model (SEM).

The hypothesis used in LM Lag is:

$H_0 : \rho = 0$ (there is no spatial lag dependency)

$H_1 : \rho \neq 0$ (there is a spatial lag dependency)

The LM lag test statistic is defined as follows:

$$LM_{lag} = \frac{(e^T W y)/s^2}{T}$$

H_0 is rejected if the *p-value* < α. Rejection of H_0 indicates the presence of spatial lag dependence, suggesting that the SAR model may be considered for subsequent analysis.

The hypothesis used in LM Error is:

$H_0 : \lambda = 0$ (there is no spatial error dependency)

$H_1 : \lambda \neq 0$ (There is spatial error dependency)

The LM lag test statistic is defined as follows:

$$LM_{error} = \frac{(e^T W e)/s^2}{T}$$

H_0 is rejected if the *p-value* < α. Rejection of H_0 indicates the presence of spatial error dependence, suggesting that the SEM may be considered for further analysis.

Spatial Autoregressive Model (SAR)

A Spatial Autoregressive Model (SAR) is a spatial regression model used when spatial dependence occurs directly in the response variable. In this model, the value of the response variable in one region is influenced by the values of neighboring regions. Within a general spatial regression model framework, the SAR is a special case when the spatial lag parameter satisfies the condition of not equal to zero ($\rho \neq 0$) and the spatial error parameter satisfies the condition of equal to zero ($\lambda = 0$) [17]. The SAR model can be expressed as follows:

$$y_i = \rho \sum_{j=1}^{n} w_{ij}\, y_j + \beta_0 + \sum_{k=1}^{p} \beta_k\, x_{ki} + \varepsilon_i$$

The model can be represented in matrix form as follows:

$$\boldsymbol{Y} = \rho \boldsymbol{W y} + \boldsymbol{X}\beta + \boldsymbol{\varepsilon}$$

where $\boldsymbol{Y}$ is the response vector, ρ is the spatial lag parameter, $\boldsymbol{W}$ is the spatial weight matrix, $\boldsymbol{X}$ is the predictor variable matrix, β is the regression parameter vector, and $\boldsymbol{\varepsilon}$ is the random error.

Spatial Error Model (SEM)

The spatial error model (SEM) is one of the spatial regression models used when spatial dependence does not occur directly on the response variable, but rather on the error component. In the framework of a general spatial regression model, the SEM is a special case when the spatial lag

parameter satisfies the condition ($\rho = 0$) and the spatial error parameter satisfies the condition ($\lambda \neq 0$) [17]. The SEM model can be expressed as follows:

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ki} + u_i$$

where

$$u_i = \lambda \sum_{j=1}^{n} w_{ij} u_j + \varepsilon_i$$

Substituting the spatial error equation into the model yields:

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ki} + \lambda \sum_{j=1}^{n} w_{ij} u_j + \varepsilon_i$$

In matrix form, the SEM model can be written as follows:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{u}$$
$$\boldsymbol{u} = \lambda \boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{Y}$ is the response vector, $\boldsymbol{X}$ is the predictor matrix, β is the regression parameter vector, λ is the spatial autoregressive parameter in the error, $\boldsymbol{W}$ is the spatial weight matrix, $\boldsymbol{u}$ is the spatial error component, and ε is the random error.

Parameter Estimation and Significance Testing

Parameter estimation in Spatial Autoregressive (SAR) and Spatial Error (SEM) models is performed using the Maximum Likelihood Estimation (MLE) method . This method is used to obtain the model parameter values that best fit the observed data.

a. Overall Significance Test

An overall significance test is performed using the maximum likelihood ratio test (MLRT) to determine the significance of all predictor variables together against the response variable. The hypothesis used is:

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$

H_1 : at least one $\beta_k \neq 0,\ k = 1, 2, \ldots, p$

The Maximum Likelihood Ratio Test (MLRT) statistic is defined as follows:

$$LR = 2\left[\ln L\left(\hat{\beta}\right) - \ln L\left(\hat{\beta}_0\right)\right]$$

where LR denotes the Maximum Likelihood Ratio Test statistic, ln $L(\hat{\beta})$ is the log-likelihood value of the full model, and ln $L(\hat{\beta}_0)$ is the log-likelihood value of the null model (*restricted model*). The decision rule is to reject H_0 if $LR > \chi^2_{(\alpha,p)}$ or if the p-value $< \alpha$

b. Partial Test

The Wald Test is used to determine the significance of each parameter in the model. The hypothesis used is:

$H_0 : \beta_k = 0,\ k = 1, 2, \ldots, p$

$H_1 : \beta_k \neq 0,\ k = 1, 2, \ldots, p$

The Wald test statistic is defined as follows:

$$W = \left(\frac{\hat{\beta}_k}{se(\hat{\beta}_k)}\right)^2$$

where W denotes the Wald test statistic, $\hat{\beta}_k$ represents the estimated parameter value, and $se(\hat{\beta}_k)$ denotes the standard error of the estimated parameter. The decision rule is to reject $W > \chi^2_{(1,\alpha)}$ or if the p-value $< \alpha$

RESULT AND DISCUSSION

Descriptive Analysis

The descriptive statistics, including the mean, standard deviation, minimum, and maximum values, are presented in Table 2.

Table 2. Descriptive Statistics

Variable	Min	Max	Mean	Std. Dev
Y	0,00	27,65	4,58	5,40
X_1	411,00	8.727,00	1.969,97	2.299,08
X_2	0,00	20,44	2,97	4,98
X_3	2,86	20,61	9,48	4,21
X_4	3,00	63,00	25,68	12,88

Based on Table 2, the average New Case Detection Rate (NCDR) of leprosy in East Java in 2025 was 4.58 cases per 100,000 population, with a minimum value of 0.00 and a maximum value of 27.65. Furthermore, all predictor variables exhibited variation across regions, indicating differences in social, economic, and healthcare characteristics among districts and municipalities.

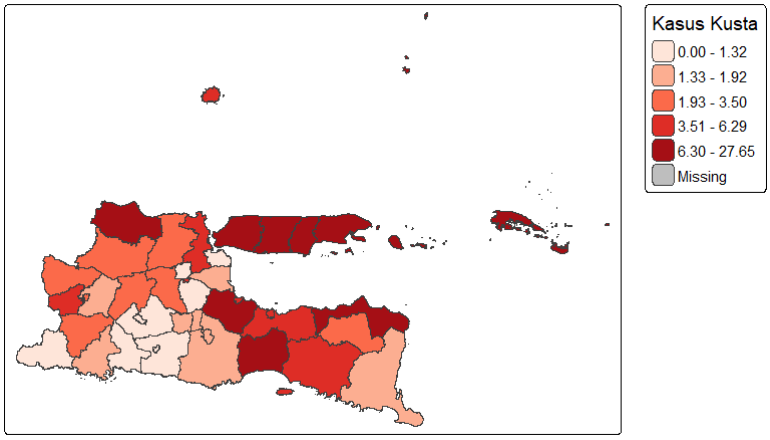

Figure 1. Spatial Distribution of Leprosy Cases in East Java 2025

Based on Figure 1, leprosy cases were not evenly distributed across districts and municipalities. Several regions exhibited relatively high case rates, particularly those within the NCDR range of 6.30–27.65 cases per 100,000 population. In contrast, some regions had relatively low NCDR values, ranging from 0.00 to 1.32 cases per 100,000 population. These differences indicate

variations in regional characteristics that may influence the occurrence of leprosy. In addition, the clustering pattern observed in the map suggests the presence of spatial dependence among regions, highlighting the need for further spatial analysis.

Before performing spatial regression modeling, the global regression model was tested to ensure the validity of the initial model. The assumptions tested include normality of residuals, multicollinearity, and heteroskedasticity.

Table 3. Normality and Heteroskedasticity Tests Result

Assumption	Statistic Value	p-value
Normality	0,9597	0,1855
Heteroskedasticity	15,405	0,0039

According to the results of the normality test in Table 3, which used the Shapiro-Wilk method, the statistical value was 0.9597 and the p-value was 0.1855. This indicates that the p-value is greater than 0.05, meaning the residuals in the global model are normally distributed. Meanwhile, the heteroskedasticity test using the Breusch-Pagan method yielded a statistical value of 15.405 with a p-value of 0.0039 (less than 0.05). This suggests that residual variance is inconsistent across regions and that the homoskedasticity assumption in the global model is not met.

Table 4. Multicollinearity Test (VIF) Result

Variable	X_1	X_2	X_3	X_4
VIF	1,71	1,13	1,57	1,10

The results of the multicollinearity test presented in Table 4 show that all independent variables have a Variance Inflation Factor (VIF) value of less than 10. Therefore, it can be concluded that the global regression model is not affected by multicollinearity, and the relationship between the independent variables is still within acceptable limits.

Based on the results of the test, the global model meets the normality assumption and does not experience multicollinearity. However, the presence of heteroskedasticity indicates that the global model cannot fully represent the spatial variation in the NCDR of leprosy data in East Java. To further identify spatial dependence between regions, spatial autocorrelation testing was performed using the Moran's I index.

Moran's I

The Moran's I test was used to identify global spatial autocorrelation in the number of new leprosy cases (NCDR) in East Java in 2025. The results of the Moran's I test are presented in Table 5.

Table 5. Moran's I test result

Test Statistic	I
5,5083	0,6059

According to Table 5, the Moran's I value obtained is 0.6059 with a test statistic value of 5.5083. At a 5% significance level, the p-value is less than 0.001, so H_0 is rejected. Therefore, it can be concluded that there is significant positive spatial autocorrelation in leprosy cases in East Java in 2025.

Lagrange Multiplier

A Lagrange Multiplier (LM) test was performed to determine the appropriate spatial regression model for modeling the number of new leprosy cases (NCDR) in East Java in 2025. The results of the Lagrange Multiplier test are presented in Table 6.

Table 6. Lagrange Multiplier Test Result

Lagrange Multiplier	p-value	Conclusion
LM Lag	0,0332	Significant
Robust LM Lag	0,0171	Significant
LM Error	0,3885	Not significant
Robust LM Error	0,1695	Not significant

Based on Table 6, the p-values of the LM Lag and Robust LM Lag tests are less than 0.05, indicating the presence of significant spatial lag dependence. In contrast, the LM Error and Robust LM Error tests are not significant because their p-values are greater than 0.05. Therefore, the SAR model is more appropriate for modeling leprosy cases in East Java in 2025.

***Spatial Autoregressive Model* (SAR)**

Next, a spatial autoregressive (SAR) model was used to estimate the parameters and test their significance in order to identify the factors influencing the NCDR of leprosy.

Table 7. Estimated parameters of the spatial autoregressive model (SAR).

Variables	Estimate	p-value	Conclusion
Intercept	-4,8735	0,0611	Not significant
X_1	0,000525	0,1194	Not significant
X_2	0,055319	0,6615	Not significant
X_3	0,852143	0,00001	Significant
X_4	-0,033635	0,4871	Not significant
ρ	0,2285	0,1201	Not significant

Based on the results of estimating parameters using the Spatial Autoregressive (SAR) model, the following model was obtained:

$$\hat{y}_i = -4{,}8735 + 0{,}2285\boldsymbol{WY} + 0{,}000525X_1 + 0{,}055319X_2 + 0{,}852143X_3 - 0{,}033635X_4$$

According on Table 7, the poverty variable (X3) has a positive and significant effect on the New Case Detection Rate (NCDR) of leprosy in East Java in 2025, with a coefficient of 0.852143 and a p-value of 0.00001. This indicates that an increase in the percentage of impoverished people is

likely to be followed by an increase in the NCDR of leprosy. This indicates that communities with higher levels of poverty tend to have limited access to healthcare services and lower quality of life, thereby increasing the risk of transmission and delayed detection of leprosy cases. This finding is consistent with previous research which states that poverty is a socio-economic factor closely related to limited access to health services and increased vulnerability of the population to various health issues [12].

Meanwhile, population density (X1), the percentage of the households without access to toilet facilities (X2) and the number of health centers (X4) had no significant influence on the NCDR of leprosy in East Java in 2025 ($p > 0.05$). The lack of significance of population density is possibly due to the use of NCDR as a normalised response variable against population size, which reduces the direct impact of density. Leprosy transmission generally requires prolonged and repeated contact with an infected person, so population density alone is not sufficient to describe the risk of transmission at the district/city level. Similarly, the percentage of households without access to toilet facilities and the number of health centers did not have a significant impact on the NCDR of leprosy. These results suggest that these two variables are not direct determinants of leprosy transmission. The success of identifying new cases is likely to be more influenced by the quality of surveillance and active case finding programmes than by the number of healthcare facilities available. Furthermore, the spatial parameter ρ was positive at 0.2285, but not significant at the 5% confidence level ($p = 0.120$). This shows that, once the predictor variables had been entered into the model, the direct spatial influence between regions had become relatively small, meaning that the parameter ρ was not statistically significant.

Spatial Error Model (SEM)

Next, a spatial error model (SEM) was used to estimate the parameters and test their significance to identify the factors that influence the NCDR of leprosy.

Table 8. Estimated parameters of the Spatial Error Model (SEM)

Variable	Estimate	p-value	Conclusion
Intercept	-4,1812	0,1339	Not significant
X_1	0,000508	0,1449	Not significant
X_2	0,034331	0,8139	Not significant
X_3	0,922462	0,000003	Significant
X_4	-0,040333	0,4201	Not significant
λ	0,1875	0,3041	Not significant

Based on the results of the parameter estimation using the Spatial Error Model (SEM) model, the following model was obtained:

$$\hat{Y}_i = -4{,}1812 + 0{,}000508X_1 + 0{,}034331X_2 + 0{,}922462X_3 - 0{,}040333X_4 + u_i$$

where $u_i = 0{,}1875Wu_i + \varepsilon_i$

According to Table 8, the poverty variable (X3) also has a positive and significant effect on the NCDR of leprosy in East Java in 2025, with a coefficient of 0.922462 and a p-value of 0.000003.

This result is consistent with the SAR model and further strengthens the idea that poverty is an important factor in the increase in new cases of leprosy in East Java, consistent with previous studies showing that poverty is closely linked to limited access to health services and increased susceptibility to leprosy [18].

Conversely, population density (X1), the percentage of households without access to toilet facilities (X2), and the number of health centers (X4) do not have a significant impact on the NCDR of leprosy, consistent with the results obtained in the SAR model. The estimated coefficients of the three variables are relatively similar in size and direction, indicating that the NCDR of leprosy is more strongly explained by the poverty factor than by the other variables.

Additionally, the spatial parameter λ is positive at 0.1875 but not significant at the 5% confidence level (p = 0.304). Although the Moran's I and Lagrange Multiplier (LM) tests in the initial stage indicated spatial dependence in the data, the parameters ρ in the SAR model and λ in the SEM model were not significant after the predictor variables were included in the model. This indicates that most of the spatial dependence detected in the initial stage can be explained by the predictor variables. After accommodating the influence of these variables, the remaining spatial dependence that needed to be captured by the parameters ρ and λ became relatively small, so neither was significant.

Evaluation of the SAR and SEM models

As part of the evaluation process, spatial autocorrelation testing was performed on the residuals of the SAR and SEM models using Moran's I index (two-sided) to ensure that the spatial dependence detected in the global model had been successfully explained by both models. The results of the test showed that the residuals of the SAR model had a Moran's I value of −0.2121 (p-value = 0.1405), while the residuals of the SEM model had a Moran's I value of −0.0496 (p-value = 0.8580). As both p-values are greater than 0.05, there is no significant spatial autocorrelation in the residuals of the two models. These results show that both the SAR and SEM models have successfully reduced the spatial dependence previously identified in the global model (OLS), making them both worthy candidates for further consideration before the best model is selected based on AIC and log-likelihood values.

The Best-Fitting Model

The best model was selected based on the Akaike Information Criterion (AIC) and log-likelihood values. The model with the lowest AIC and the highest log-likelihood was considered the most appropriate model for representing the spatial pattern of leprosy NCDR.

Table 9. AIC and Log-Likelihood Values for the SAR and SEM Models

Model	AIC	Log-likelihood
SAR	219,73	-102,8651
SEM	222,18	-104,0909

Based on Table 9, the SAR model has an AIC value of 219.73, which is lower than the SEM model's value of 222.18. Additionally, the SAR model produced a higher log-likelihood value (-102.8651)

than the SEM model (-104.0909), indicating that the SAR model is a relatively better fit for the data. Both goodness-of-fit criteria consistently show that the Spatial Autoregressive (SAR) model is the most suitable for modelling the New Case Detection Rate (NCDR) of leprosy in East Java Province in 2025. Based on the goodness-of-fit criteria used, the SAR model provides a better representation of the spatial pattern of the NCDR of leprosy than the SEM model.

CONCLUSION

The results of this study indicate that the percentage of the population living in poverty significantly affects the *New Case Detection Rate* (NCDR) of leprosy in East Java in 2025. Based on a lower *Akaike Information Criterion* (AIC) value and a higher log-likelihood value, the *Spatial Autoregressive* (SAR) model was selected as the best-performing model compared with the *Spatial Error* (SEM) model. Therefore, the SAR model is more appropriate for modeling the spatial distribution of leprosy NCDR in East Java. These findings can support government decision-making in developing more effective leprosy control strategies, particularly in areas with high poverty rates. Leprosy control should not rely solely on medical interventions but should also be supported by social and economic measures, including improved access to health services, health education, and poverty alleviation programs. More broadly, this study demonstrates that spatial regression can support regional health planning and strengthen spatial disease surveillance systems, enabling more targeted public health interventions according to regional characteristics.

This study has several limitations. First, the use of cross-sectional data from 2025 does not allow for temporal analysis of leprosy NCDR trends. Second, the number of observational units was relatively limited to 38 districts/cities in East Java. Finally, the analysis was restricted to a limited number of social, economic, and health service indicators. Therefore, future studies are recommended to use spatial-temporal data over a longer observation period and incorporate additional relevant variables to provide a more comprehensive understanding of the factors influencing leprosy occurrence.

REFERENCE

[1] Kementerian Kesehatan Republik Indonesia, “Profil Kesehatan Indonesia 2024,” Kementerian Kesehatan Republik Indonesia, 2024.

[2] World Health Organization, “Leprosy.” Accessed: May 20, 2026. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/leprosy

[3] Jatim Newsroom, “Yankes Bergerak Kangean Perkuat Langkah Jatim Menuju Eliminasi Kusta 2030,” Dinas Komunikasi dan Informatika Provinsi Jawa Timur. Accessed: May 20, 2026. [Online]. Available: https://kominfo.jatimprov.go.id/berita/yankes-bergerak-kangean-perkuat-langkah-jatim-menuju-eliminasi-kusta-2030

[4] World Health Organization, “Mewujudkan Komitmen Nasional Menjadi Kemajuan Nyata dalam Penanggulangan Kusta.” Accessed: Jun. 06, 2026. [Online]. Available: https://www.who.int/indonesia/id/news/detail/20-04-2026-turning-national-commitment-into-real-progress-against-leprosy

[5] R. A. D. Bestari, J. Maulana, N. L. Fitriyani, and H. Akbar, “Faktor Risiko Kejadian Kusta di Indonesia: Literature Review,” *Promot. J. Kesehat. Masy.*, vol. 13, no. 2, pp. 57–65, Dec. 2023, doi: 10.56338/promotif.v13i2.4492.

[6] G. F. A. Edi and R. Azizah, "Analysis of Environmental Risk Factors for Leprosy in Indonesian Society: Meta-Analysis," *Poltekita J. Ilmu Kesehat.*, vol. 17, no. 1, pp. 105–113, May 2023, doi: 10.33860/jik.v17i1.2119.

[7] A. I. Putri *et al.*, "A socio-ecological model of the management of leprosy reactions in Indonesia and India using the experiences of affected individuals, family members and healthcare providers," *BMC Health Serv. Res.*, vol. 25, no. 1, Dec. 2025, doi: 10.1186/s12913-025-12340-5.

[8] I. de C. Bueno, D. D. S. Lages, and F. C. F. Lana, "Spatial analysis of the epidemiological risk of leprosy in the municipalities of Minas Gerais," *PLoS Negl. Trop. Dis.*, vol. 17, no. 6, Jun. 2023, doi: 10.1371/journal.pntd.0011381.

[9] I. R. Akolo, "Perbandingan Matriks Pembobot Rook dan Queen Contiguity dalam Analisis Spatial Autoregressive Model (SAR) dan Spatial Error Model (SEM)," *Jambura J. Probab. Stat.*, vol. 3, no. 1, pp. 11–18, May 2022, doi: 10.34312/jjps.v3i1.13582.

[10] L. P. S. Pratiwi, "Perbandingan Metode Regresi Spatial Autoregressive Model dan Spatial Error Model untuk Kejadian Diare di Provinsi Bali," *J. Math. Educ. Sci.*, vol. 9, no. 1, pp. 67–72, Oct. 2023, doi: 10.30743/mes.v9i1.8099.

[11] D. Rahmawati and H. Bimanto, "Analysis of Factors Affecting Leprosy Cases in East Java Province with Spatial Autoregressive Model (SAR)," *J. Berk. Kesehat.*, vol. 7, no. 1, p. 59, 2021, doi: 10.20527/jbk.v7i1.10285.

[12] S. Wahyuni, N. Amalita, and F. Fitri, "Implementation of Spatial Autoregressive Analysis to Determine Factors Affecting the Population Dependency Ratio in West Sumatera," *J Stat. J. Ilm. Teor. Dan Apl. Stat.*, vol. 18, no. 2, pp. 957–967, 2025, doi: 10.36456/jstat.vol18.no2.a10786.

[13] Badan Pusat Statistik Jawa Timur, "Provinsi Jawa Timur dalam Angka 2026," Badan Pusat Statistik Provinsi Jawa Timur, 2026.

[14] L. Anselin, *Spatial Econometrics: Methods and Models*, 1st ed. Springer Dordrecht, 1988. [Online]. Available: https://link.springer.com/book/10.1007/978-94-015-7799-1

[15] J. Lee and D. W. S. Wong, *Statistical Analysis With ArcView GIS*. New York: John Wiley & Sons, 2001.

[16] I. P. G. I. B. Prasetya, B. Baharuddin, and G. N. A. Wibawa, "Pemodelan Regresi Spasial untuk Menentukan Faktor-Faktor yang Berpengaruh terhadap Tingkat Kriminalitas di Provinsi Bali dan Jawa Timur," *J. Syntax Admiration*, vol. 5, no. 6, pp. 2033–2046, 2024, doi: 10.46799/jsa.v5i6.1207.

[17] J. P. LeSage, "The Theory and Practice of Spatial Econometrics," Toledo, Feb. 1999.

[18] F. Sholihah, D. A. Wulandari, and N. Qisthi, "A Dynamic Spatial Durbin Panel Model for Analyzing Poverty Rates in West Java Province," *J Stat. J. Ilm. Teor. Dan Apl. Stat.*, vol. 18, no. 2, pp. 968–980, 2025, doi: 10.36456/jstat.vol18.no2.a10826.