

Examining the Correlation between COVID-19 Prevalence and Patient Behaviors, Healthcare, and Socioeconomic Determinants: A Geospatial Analysis of ASEAN Countries

Thammaboribal, P.,^{1*} Nitin, K. T.,¹ Junpha, J.,² Lipiloet, S.,³ and Wongpitak, K.,^{2*}

¹Department of Remote Sensing and Geographic Information System, Asian Institute of Technology, Thailand. E-mail: prapasgnss@gmail.com

²Faculty of Science and Technology, Rajamangala University of Technology Suvarnabhumi, Thailand

³Department of Civil Engineering, Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Thailand. E-mail: sukhom.l@en.rmutt.ac.th

⁴Department of Public Health, Collage of Medicine and Public Health, Ubon Ratchathani University, Thailand. E-mail: klarnarong.w@ubu.ac.th

*Corresponding Author

DOI: <https://doi.org/10.52939/ijg.v20i3.3159>

Abstract

This study aimed to identify COVID-19 clusters in ASEAN countries using Global Moran's I. The results revealed an absence of spatial autocorrelation in the study area, indicating the non-existence of clusters in ASEAN countries during the 2021 pandemic. However, high-low (HL) and low-high (LH) outliers were identified in the Philippines, Brunei Darussalam, and Timor-Leste, respectively. Bivariate local Moran's I was employed to analyze cluster areas between the dependent variable (COVID-19 confirmed cases) and independent variables categorized as 1.) behavioral (Smoking rate (SMK) and Number of older people (65+) (OLD)), 2.) Healthcare (Number of cardiovascular patients (CAR) and Number of lung cancer patients (LNG)), and 3.) Socioeconomic factors: Total population number (TPN), Urban population (UBP), Population density (PDT), Human development index (HDI), GDP per capita (GDP), Life expectancy index (LEI), and International tourist arrival (ITA). The results showed that Indonesia, with high COVID-19 confirmed cases, was surrounded by high rates of independent variables. Bivariate local Moran's I can be instrumental in formulating disease control plans for individual countries. Correlations between COVID-19 and independent variables were modeled using Ordinary Least Squares (OLS), Spatial Lag Model (SLM), and Spatial Error Model (SEM). OLS was deemed less effective compared to the other models. Additionally, p-values from SLM and SEM indicated that only specific socioeconomic factors were associated with COVID-19 confirmed cases. The correlation of Smoking Rate (SMK) to the number of COVID-19 cases was questioned due to non-identical p-values from SLM and SEM. After excluding unassociated factors, SEM emerged as the best-fit model, boasting the highest R² of 0.961 and the lowest AIC and BIC of 323.663 and 325.665, respectively, compared to other models. The correlation between COVID-19 confirmed cases and associated variables derived from SEM is $COV = 204,763 + 0.017TPN - 710.617PDT + 54,967GDP + 340.898ITA$. It is crucial to note that the correlations identified in this study do not imply a causal relationship between socioeconomics and COVID-19 infection in ASEAN countries.

Keywords: ASEAN, COVID-19, Geospatial Analysis, HealthGIS, OLS, SEM, SLM

1. Introduction

The year 2020 emerged as a pivotal period globally, not only due to its centennial occurrence but primarily owing to the onset of a widespread pandemic caused by the SARS-CoV-2 virus, commonly known as COVID-19. The novel coronavirus (nCoV) was initially identified in China

in December 2019, and it swiftly disseminated to various countries globally. Consequently, the World Health Organization (WHO) declared a Public Health Emergency of International Concern (PHEIC) on January 30, 2020, and officially classified the situation as a pandemic on March 11, 2020.



Believed to have originated in Wuhan District, China [1] and [2] towards the end of 2019, the virus prompted the Chinese Government to implement stringent measures, including lockdowns and healthcare provisions, to curb its spread. Despite these efforts, by January 2020, the World Health Organization (WHO) recognized the outbreak as a "Public Health Emergency of International Concern," subsequently declaring it a pandemic in March 2020 [1]. The ensuing pandemic induced unprecedented changes in human history, such as the closure of establishments, factories, and airlines, the postponement or cancellation of public events, the closure or transition to online classes for schools and universities, supply shortages resulting from closed international borders, leading to panic buying behaviors, and more. However, amid the challenges, some positive outcomes emerged, including a reduction in air pollution indices [3] and [4], greenhouse gases [5] and [6], and water pollution [7] and [8].

The global pandemic, with profound implications for healthcare, the economy, and various societal facets, has prompted researchers across diverse disciplines to conduct extensive studies. These investigations aim to comprehend the factors contributing to the elevated prevalence of COVID-19 and explore potential solutions to mitigate its spread and minimize associated damages. Conventional statistical analyses, coupled with spatial analysis methods, have been employed to unravel the correlations between various factors and the dynamics of COVID-19 [9][10][11][12][13][14] and [15] enabling the estimation of infectious rates and identification of key factors influencing transmission and mortality rates.

The global COVID-19 pandemic has underscored the intricate interplay between disease prevalence and a myriad of factors encompassing behavioral patterns, healthcare systems, and socioeconomic dynamics. Understanding the nuanced relationships between these determinants and the regional prevalence of COVID-19 is imperative for effective public health strategies. This article delves into a comprehensive examination of the intricate connections among COVID-19 prevalence, behavioral factors, healthcare infrastructure, and socioeconomic conditions within the context of the Association of Southeast Asian Nations (ASEAN) countries. Employing geospatial analysis as a methodological lens, our study aims to unravel spatial patterns, identify significant correlations, and elucidate the underlying dynamics that contribute to the varying impacts of the pandemic across the ASEAN region. Through this exploration, we seek to provide valuable insights that can inform targeted

interventions, policy formulations, and proactive measures to mitigate the spread and impact of COVID-19 in these diverse and interconnected nations.

Geospatial analysis is a process that involves examining the attributes and relationships of features within an aerial context, utilizing traditional analytical techniques to extract or generate new information from spatial data. Spatial information encompasses details pertaining to the position, area, shape, and size of regions of interest [16]. A key distinction between spatial analysis and statistical analysis lies in the consideration of feature or independent variable locations; traditional statistical methods typically neglect this spatial dimension. In contemporary times, Geographic Information System (GIS) technologies are adept at managing spatial data in conjunction with various factors, whether spatial or non-spatial in nature. Numerous studies have been conducted to analyze the impact of different factors on the transmission rate, mortality rate, and outcomes of the COVID-19 pandemic within specific regions of interest [17]. Utilizing geospatial analysis as a powerful tool, our research endeavors to provide a nuanced examination of the spatial distribution of COVID-19 prevalence across ASEAN countries. By scrutinizing the geospatial patterns, we aim to discern correlations between disease prevalence and various behavioral, healthcare, and socioeconomic factors. This analytical approach allows us to go beyond traditional statistical methods and explore the geographic dimension of the pandemic, shedding light on localized influences and disparities that may otherwise remain obscured.

This study specifically concentrates on conducting spatial analysis at the country level for ASEAN countries, with a focus on identifying significant factors influencing the COVID-19 pandemic in the ASEAN regions (Cambodia, Brunei, Philippine, Thailand, Myanmar, Laos, Indonesia, Malaysia, and Singapore). The main objective of this study is to understand Spatial Analysis and its application areas in air-borne diseases with the proposed factors and to analyse the outcome of different spatial regression models. Geostatistical analysis on the selected regions was performed using an opensource software so-called "GeoDa" and the analysis focusing on the total number of COVID-19 confirmed cases, total death cases of COVID-19. The spatial regression analysis was performed with the following regression models: Ordinary least square (OLS), spatial lag model (SLM) and spatial error model SEM. This study incorporates socioeconomic factors, behavioral factors, and healthcare factors as the variables factors.

The significance of this investigation lies not only in its potential to enhance our understanding of the multifaceted determinants influencing COVID-19 prevalence but also in its practical implications. The insights gained from this geospatial analysis can inform targeted interventions, guide healthcare resource allocation, and aid policymakers in crafting region-specific strategies to curb the spread of the virus and mitigate its impact on diverse communities within the ASEAN region.

2. Study Area

ASEAN is a regional intergovernmental organization comprising ten member states in Southeast Asia. Established on August 8, 1967, the founding members of ASEAN are Indonesia, Malaysia, the Philippines, Singapore, and Thailand. Over the years, Brunei Darussalam, Vietnam, Laos, Myanmar, and Cambodia have also joined the organization, contributing to its current composition [18]. Geographically, ASEAN is situated in the southeastern part of Asia as illustrates in Figure 1, and is characterized by its diverse landscapes, encompassing tropical rainforests, mountains, plains, and coastlines. The member states collectively cover a substantial portion of the region, sharing land and maritime borders. The organization's geographic location places it at the crossroads of major global trade routes, fostering economic interactions and cultural exchanges among member countries.

In terms of population, ASEAN is home to over 690 million people [19], making it one of the most populous regions in the world. The population is ethnically and culturally diverse, reflecting the rich tapestry of Southeast Asian societies. The member states exhibit a wide range of languages, religions, and traditions, contributing to the vibrant and dynamic character of the region. The geostrategic location of ASEAN has played a crucial role in shaping its economic and political significance. The member countries, collectively known as the ASEAN Community, work towards promoting regional cooperation, economic integration, and social progress. ASEAN serves as a platform for addressing shared challenges, fostering diplomatic ties, and enhancing the overall well-being of its diverse population. The population distribution among countries is illustrated in Figure 2.

3. Methodology

3.1 Data Collection

The necessary data, including shape files of ASEAN countries, essential statistical indices, and COVID-19 data, were gathered. The ASEAN country boundary

shapefile was acquired from DivaGIS available at <https://www.diva-gis.org/>. The COVID-19 infection data from December 2019 to January 2021 were obtained from the World Health Organization (WHO) at <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>. The dependent parameters as behavioral factors, healthcare, and socioeconomic conditions were mostly collected from Key Indicator Database (KIDB) provided by Asian Development Bank (ADB) available at <https://kidb.adb.org/>. The Gross Domestic Products (GDP) per capita was collected from International Monetary Fund at <https://www.imf.org/external/datamapper/NGDPDPC@WEO/THA/IDN/PHL/VNM/MYS>. To facilitate further processing and spatiotemporal analysis, geospatial shapefiles were generated using ArcMap Software (Version 10.8.1) and GeoDa. In the modeling process, the independent variables were categorized into three classes, with each class containing the variables as outlined in Table 1.

3.2 Study Workflow

The procedure for conducting this study depicts in Figure 3. To model the COVID-19 infection rate based on various influencing factors, an initial investigation into spatial autocorrelation is crucial. Global Moran's I was employed to determine whether spatial autocorrelation exists within the dataset encompassing all ASEAN countries. Global Moran's I is widely used in fields such as geography, ecology, and epidemiology to understand the spatial patterns of phenomena and identify areas where similar values are clustered or dispersed. If spatial autocorrelation is identified, bivariate Local Moran's I is then utilized to pinpoint local cluster areas and explore spatial correlations between COVID-19 confirmed cases and independent variables, namely behavioral factors, healthcare infrastructure, and socioeconomic factors. Subsequently, the spatial modeling of the COVID-19 infection rate is performed using both non-spatial regression models, such as Ordinary Least Squares (OLS), and spatial regression models, such as Spatial Lag Model (SLM) and Spatial Error Model (SEM). The distinction between non-spatial and spatial regression lies in the consideration of spatial dependency. The spatial lag model incorporates spatial dependency between dependent variables, while the spatial error model considers spatial error dependency.

Finally, the selection of the best-fitting model is based on criteria such as R-squared (R^2), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC).

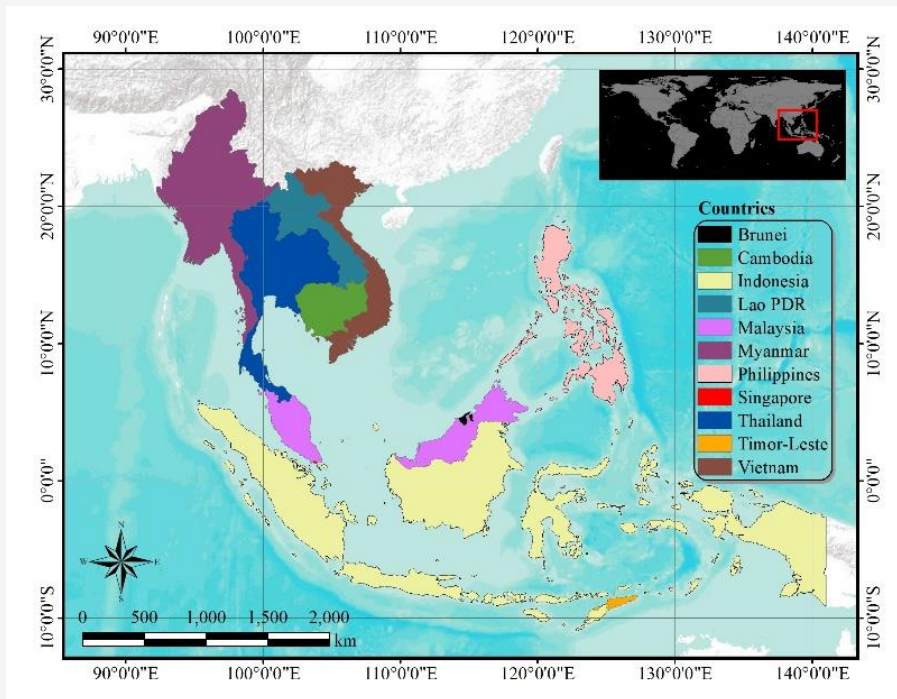


Figure 1: ASEAN community

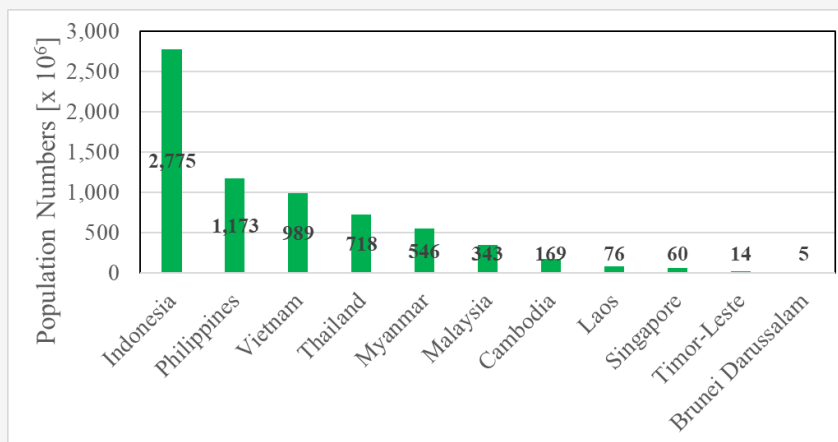


Figure 2: Total populations of the countries within ASEAN [19]

Table 1: Variables used in modeling

Dependent	COVID-19 confirmed cases in ASEAN countries (COV)
Independent	1. Behavior: Smoking rate (SMK), and Number of older people (65+) (OLD)
	2. Healthcare: Number of cardiovascular patients (CAR), and Number of lung cancer patients (LNG)
	3. Socioeconomic: Total population number (TPN), Urban population (UBP), Population density (PDT), Human development index (HDI), GDP per capita (GDP), Life expectancy index (LEI), International tourist arrival (ITA)

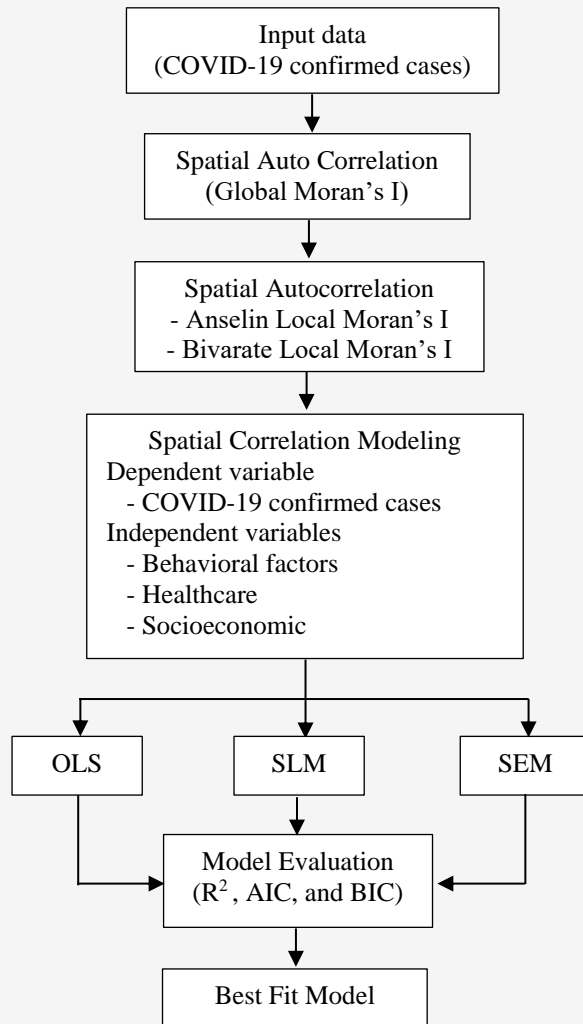


Figure 3: Study work flow

3.3 Global Moran's I

Global Moran's I is a statistical measure used in spatial analysis to assess the spatial autocorrelation of a variable across a geographic area. Spatial autocorrelation refers to the degree to which the values of a variable are correlated with the values of the same variable in neighboring locations. Global Moran's I provides a single summary statistic to indicate whether there is spatial clustering, dispersion, or randomness in the distribution of the variable. The values of Moran's I range from -1 (indicating perfect dispersion) to 1 (indicating perfect clustering), with 0 suggesting a random spatial pattern [20]. The Global Moran's I (I) is generally computed from Equation 1 [21].

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{X})(x_j - \bar{X})}{W \sum_{i=1}^n (x_i - \bar{X})^2}$$

Equation 1

where:

- n is the total number of considered polygons
- \bar{X} is the average of an attribute
- x_i is the value of attribute for polygon i
- x_j is the value of attribute for polygon j
- w is the spatial weight for a pair of polygons
- W is the sum of spatial weights

The formula for Global Moran's I involves comparing the observed spatial autocorrelation to what would be expected under spatial randomness. It considers both the values of the variable at each location and the values in neighboring locations. The calculated expected value of Moran's I criterion represents the level indicative of a random

distribution of values. The expected value of Global Moran's I ($E(I)$) is determined from Equation 2.

$$E(I) = -\frac{1}{n-1}$$

Equation 2

Equation 2 shows that the higher the number of objects, the expected Global Moran's I is getting close to zero.

3.4 Local Moran's I

Local Moran's I is a spatial autocorrelation statistic that extends the analysis provided by Global Moran's I by examining local patterns of spatial clustering, dispersion, or randomness for individual locations within a geographic dataset. While Global Moran's I provides a single measure for the entire dataset, Local Moran's I calculates spatial autocorrelation for each specific location, allowing for the identification of local clusters or outliers. Local Moran's I helps to uncover spatial heterogeneity by indicating whether specific areas exhibit statistically significant clustering or dispersion of values. This analysis is particularly useful for identifying local spatial patterns that might not be apparent when considering the entire dataset as a whole. The definition of Local Moran's I is given in Equation 3 [22].

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^n w_{i,j} (x_j - \bar{X})$$

Equation 3

$$S = \frac{\sum_{j=1, j \neq i}^n (x_j - \bar{X})^2}{n-1}$$

Equation 4

Where:

$w_{i,j}$ is spatial weight between feature i and j

The calculation of Local Moran's I involves comparing the value of a variable at a specific location with the average value of that variable in neighboring locations. The results classify each location into one of four categories [9]:

- High-High (HH): Locations with high values surrounded by other locations with high values. Indicates a cluster of high values in a specific area.
- Low-Low (LL): Locations with low values surrounded by other locations with low values.

Indicates a cluster of low values in a specific area.

- High-Low (HL): Locations with high values surrounded by locations with low values. Indicates a spatial outlier with high values in a region of low values.
- Low-High (LH): Locations with low values surrounded by locations with high values. Indicates a spatial outlier with low values in a region of high values.

The results of Local Moran's I are typically represented in a spatial map, known as a Moran scatterplot or Moran map, where each location is categorized based on its level of spatial autocorrelation [23] and [24]. The significance of these patterns is determined through statistical tests. High-High and Low-Low patterns represent areas of positive local spatial autocorrelation, indicating clustering of similar values, while High-Low and Low-High patterns suggest negative local spatial autocorrelation, indicating dissimilar values. Local Moran's I is valuable for identifying spatial clusters and outliers, helping researchers and analysts understand the local patterns and associations present in their data, and it is widely used in fields such as geography, ecology, and economics for spatial data analysis.

3.5 Bivariate Local Moran's I

In GeoDa software, Bivariate Local Moran's I is an extension of the Local Moran's I statistic that allows for the analysis of spatial autocorrelation between two different variables simultaneously [25]. The traditional Local Moran's I is applied to a single variable, indicating local patterns of clustering or dispersion for that variable at each location. In contrast, Bivariate Local Moran's I assess the spatial relationship between two variables, revealing local patterns of joint clustering, joint dispersion, or cross-correlation. The Bivariate Local Moran's I statistic provides insights into whether there are specific locations where both variables exhibit similar values (joint clustering) or dissimilar values (joint dispersion). This can be valuable for understanding spatial relationships and dependencies between different factors within a geographic dataset. Users can generate Bivariate Local Moran's I results in GeoDa by selecting two variables of interest and examining the local spatial patterns of association between them. The results are often visualized using Moran scatterplots or Moran maps, where each location is classified based on its joint spatial autocorrelation with the two variables.

Bivariate local Moran's I is defined in Equation 5 [26].

$$I_B = \frac{\sum_i (\sum_j w_{ij}(d)x_i y_i)}{\sum_i x_i^2}$$

Equation 5

where:

- $x_i y_i$ is cross product of first variable at location i and second variable at each neighboring location j
- $w_{ij}(d)$ is spatial weight matrix presenting neighborhood structure within a specified distance d .

3.6. Spatial Modeling

3.6.1 Ordinary Least Square (OLS)

Ordinary Least Squares (OLS) in spatial analysis is a standard statistical method used to estimate the parameters of a linear regression model when analyzing spatial data. The goal of OLS is to minimize the sum of squared differences between observed and predicted values, considering the relationships and the independent variables. OLS is commonly used in modeling relationships between variables, it stands out as one of the most well-known regression techniques and serves as the foundational approach for all regression analyses. Acting as a starting point, OLS constructs a global model for the variable or process under investigation, formulating a single regression equation as presented in equation 6 to encapsulate and represent the overarching patterns of that process [27].

$$y = \beta_0 + \sum_{i=1}^n \beta_i x_i + \varepsilon$$

Equation 6

where:

- y is predicted value (dependent variable)
- β_0 is model constant
- β_i is coefficient of parameter i
- x_i is independent parameter i
- ε_i is random error of parameter i
- n is total number of independent variables

OLS may exhibit inefficiency when dealing with spatial autocorrelation since it inherently overlooks spatial structures within the data [27]. The basic concept of OLS depicts in Figure 4.

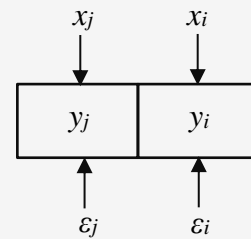


Figure 4: OLS basic concept

3.7.2 Spatial Lag Model (SLM)

The spatial lag model in spatial analysis is a regression framework that acknowledges and incorporates the influence of spatial dependencies among neighboring observations. Unlike traditional regression models, the spatial lag model accounts for the fact that the values of a variable in one location may be correlated with the values of the same variable in nearby locations. This model includes a spatial lag term, representing the weighted average of the variable in neighboring locations, and aims to capture the spatial autocorrelation inherent in the data. By considering spatial interactions, the spatial lag model provides a more nuanced understanding of how spatial relationships affect the dependent variable, making it particularly valuable in scenarios where spatial dependencies play a crucial role, such as in urban planning, economics, and geography. The SLM is expressed in Equation 4.

$$y = \beta_0 + \lambda WY + x\beta + \varepsilon$$

Equation 4

where:

- β_0 is model constant
- λ is spatial error coefficient
- W is spatial weight matrix
- x is independent variable
- β is coefficient of independent parameter
- ε is spatial error

Spatial lag occurs when the dependent variable (y) in location i is influenced by the independent variables present in both location i and location j as shown in Figure 5. The assumption of uncorrelated error terms is violated in OLS regression with spatial lag, and concurrently, the assumption of independent observations is also violated. This violation leads to biased and inefficient estimates.

Furthermore, spatial lag suggests a potential diffusion process, wherein events in one place predict an increased likelihood of similar events in neighboring places [29].

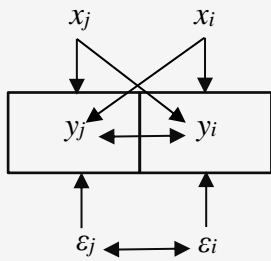


Figure 5: SLM basic concept [28]

3.7.3 Spatial Error Model (SEM)

The spatial error model in spatial analysis is a regression technique that accounts for spatial autocorrelation by incorporating a spatially correlated error term. In contrast to the spatial lag model, the spatial error model assumes that the spatial dependencies are captured through the error term rather than the dependent variable. This model recognizes that there may be unobserved or omitted spatially correlated factors influencing the dependent variable. By allowing for spatially correlated errors, the spatial error model captures the unaccounted spatial variation in the model, making it a valuable tool in situations where spatial autocorrelation may be present in the residuals of the regression model. The spatial error model is commonly used in spatial econometrics and geographic studies to improve the accuracy of parameter estimates and account for spatial patterns in the data. SEM is modelled by Equation 7.

$$y = \beta_0 + \lambda W\varepsilon + x\beta + \xi \quad \text{Equation 7}$$

where:

- ξ is white noise or vector of uncorrelated error terms
- λ is spatial error coefficient ($\lambda=0$ when no spatial correlation between the errors)

In SEM, the concept revolves around the correlation of error terms across spatial units. Consequently, the predicted y is influenced by errors (ε) from another predicted y , as illustrated in Figure 6. In OLS regression with spatial error, the assumption of uncorrelated error terms is violated, leading to inefficient estimates.

The presence of spatial error signals the possible omission of spatially correlated covariates, which, if overlooked, could impact the validity of inference [29].

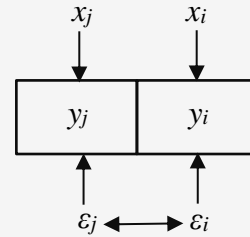


Figure 6: SEM basic concept [28]

3.8 Evaluation of Models

3.8.1 R-squared (R^2)

R-squared, also known as the coefficient of determination or the coefficient of multiple determination in the case of multiple regression, is a statistical metric that gauges the proximity of the data points to the fitted regression line. Simply put, it represents the percentage of variation in the response variable explained by a linear model. R^2 values range from 0 to 100%, where 0% indicates that the model does not account for any variability around the mean of the response data, and 100% suggests that the model comprehensively explains all variability. In general, a higher R^2 -value indicates a better fit of the model to the data. The strength of correlation is determined from Table 2.

Table 2: Strength of correlation and R^2 [30]

Range of R^2	Strength of correlation
1.0 – 0.7	Very strong
0.7 – 0.5	Strong
0.5 – 0.4	Moderate
0.4 – 0.3	Limited
0.3 – 0.2	Weak
0.2 – 0.0	Negligible

3.8.2 Akaike Information Criterion (AIC)

The Akaike Information Criterion (AIC) is a statistical measure used for model selection when comparing different models. It was developed by the Japanese statistician Hirotugu Akaike in 1973 [31]. The AIC takes into account both the goodness of fit of the model and the complexity of the model (the number of parameters). The AIC is calculated using Equation 8:

$$AIC = 2k - 2\ln(L)$$

Equation 8

where:

- L is the likelihood of the model given the data,
- k is the number of parameters in the model.

The AIC penalizes models for being too complex, favoring models that achieve a good fit with fewer parameters. When comparing models, the one with the lowest AIC is considered the best, indicating a good balance between goodness of fit and model simplicity. The AIC is widely used in various fields, including statistics, machine learning, and econometrics, for model selection and comparison.

3.8.3 Bayesian Information Criterion (BIC)

The Bayesian Information Criterion (BIC), also known as the Schwarz criterion, is a statistical measure used for model selection among a set of candidate models. It was developed by the statistician Gideon Schwarz. Similar to the Akaike Information Criterion (AIC) [32], the BIC considers both the goodness of fit and the complexity of the model. The BIC is defined as in Equation 9.

$$BIC = k \cdot \ln(n) - 2 \ln(L) \quad \text{Equation 9}$$

where:

- L is the likelihood of the model given the data
- k is the number of parameters in the model
- n is the sample size.

The BIC penalizes models for being too complex, and the penalty term is more severe than in the AIC, particularly for models with a larger number of parameters. Like the AIC, when comparing models, the one with the lowest BIC is considered the best, indicating a good balance between goodness of fit and model simplicity.

The BIC is commonly used in statistics, machine learning, and other fields for model selection as AIC.

4. Results

4.1 COVID-19 Situation

ASEAN Biodiaspora Virtual Center (ABVC) reports that as of 11 October 2021 at 2 PM (GMT+8), there were 239,004,429 confirmed COVID-19 cases worldwide, resulting in 4,872,719 deaths. The global Case Fatality Rate (CFR) stood at 2.0%. The COVID-19 confirmed cases as well as mortality rate for the ASEAN countries indicted in Table 3 and Figure 7. According to the data presented in the Table and Figure, the COVID-19 statistics for ASEAN countries, spanning from the first confirmed case report to the latest update, include 10,552,272 confirmed cases, 268,788 deaths, and a mortality rate of 2.14%.

4.2 Independent Variable

The cartographic visualization method was employed to depict data corresponding to the independent variables. as depicts in Figure 8 to Figure 10.

4.3 Special autocorrelation using Global Moran's I

In Figure 11, the Z-score of -0.217236 and a p-value of 0.828 collectively indicate that the observed pattern is not significantly different from randomness. As a result, there is no discernible evidence of spatial autocorrelation within the dataset. Put differently, the prevalence of COVID-19 in each ASEAN country does not exhibit correlation with that of another. The Global Moran's I results further reinforce this, suggesting that the prevalence of COVID-19 in each ASEAN country was not influenced by the COVID-19 situations in the surrounding countries.

Table 3: COVID 19 situations in ASEAN countries [33]

Country	1 st confirmed case	Last report	Total confirmed cases	Total deaths	Mortality rate [%]
Indonesia	2-Mar-20	10-Oct-21	4,227,932	142,651	3.4
Philippines	30-Jan-20	10-Oct-21	2,666,562	39,624	1.5
Vietnam	23-Jan-20	10-Oct-21	839,662	20,555	2.4
Thailand	13-Jan-20	10-Oct-21	1,710,884	17,691	1.0
Myanmar	23-Mar-20	10-Oct-21	478,651	18,134	3.8
Malaysia	25-Jan-20	09-Oct-21	2,332,221	27,265	1.2
Cambodia	27-Jan-20	10-Oct-21	114,810	2,506	2.2
Laos	24-Mar-20	10-Oct-21	28,540	26	0.1
Singapore	23-Jan-20	10-Oct-21	124,157	153	0.1
Timor-Leste	21-Mar-20	10-Oct-21	19,673	119	0.6
Brunei Darussalam	10-Mar-20	10-Oct-21	8,980	64	0.7

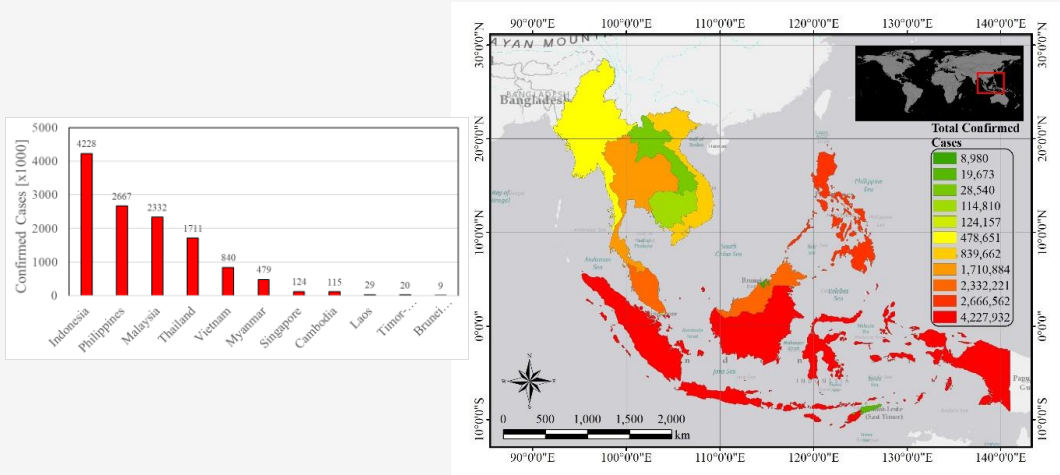


Figure 7: COVID 19 confirmed cases in ASEAN countries as of 10 October 2021

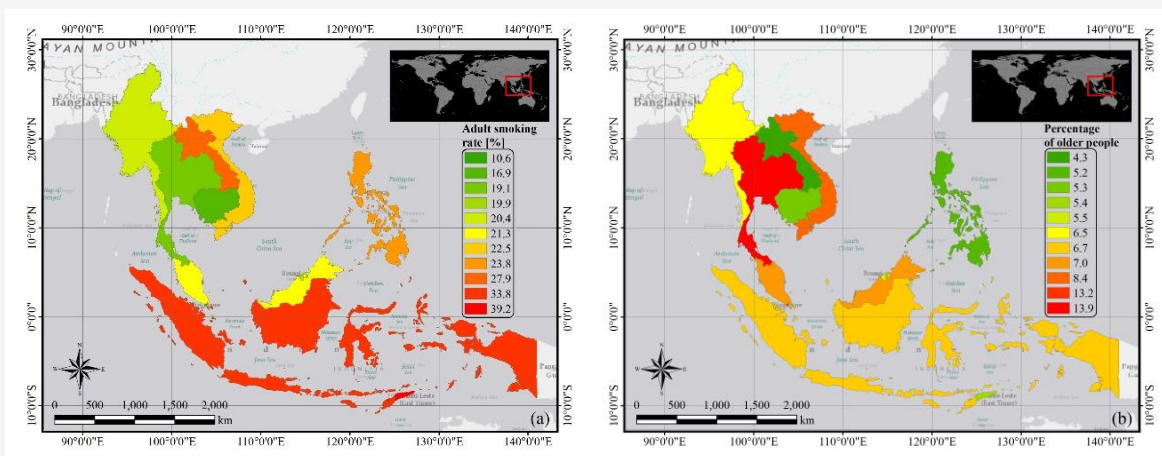


Figure 8: Behavioral variables (a) adult smoking rate (b) number of older people (65+)

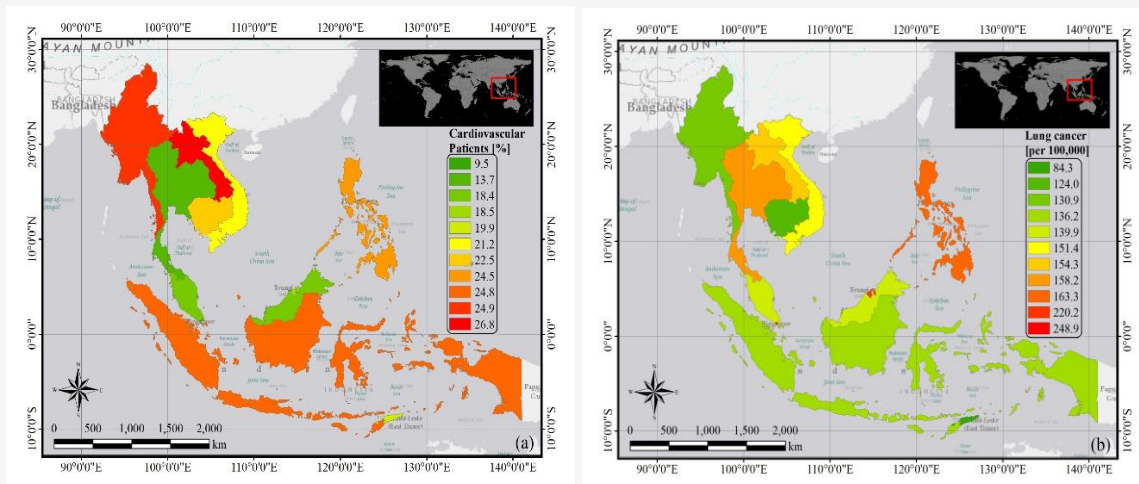


Figure 9: Healthcare variables (a) number of cardiovascular patients (b) number of lung cancer patients

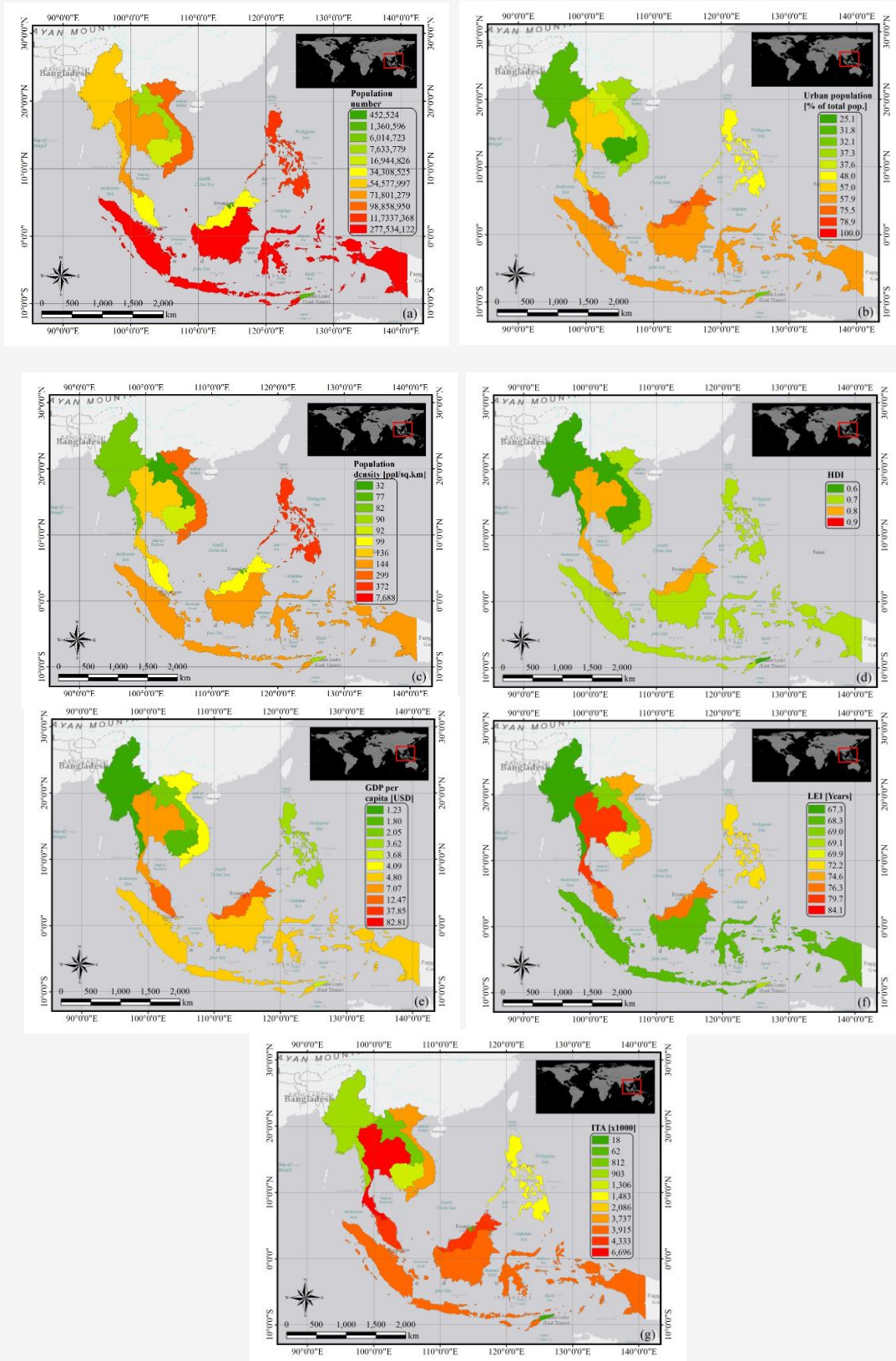


Figure 10: Socioeconomic (a) total population number (b) urban population (c) population density (d) human development index (e) GDP per capita (f) Life expectancy index (g) International tourist arrival

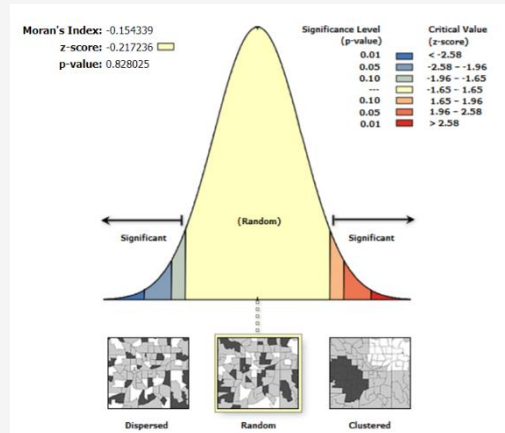


Figure 11: Spatial autocorrelation derived from Global Moran's I

4.4 Anselin Local Moran's I

Based on the findings presented in the preceding section, spatial autocorrelation was not detected within the dataset. Consequently, there were no discernible hot spots or cold spots within the study area. While theoretically, identifying the locations of these spots may not be necessary, in practical terms, outliers can be extracted using Anselin's Local Moran's I, as illustrated in Figure 12. Figure 11 illustrates a High-Low outlier pattern in the Philippines, signaling a high COVID-19 infection rate in the country compared to the lower rates in surrounding nations. Consequently, the Philippines should prioritize implementing robust measures to prevent the spread of the disease to other ASEAN countries. Conversely, Brunei Darussalam and Timor-Leste exhibit a Low-High outlier pattern, indicating lower infection rates compared to their neighbors. Therefore, these two countries should enhance disease screening for inbound travelers to prevent the spread of the disease from surrounding countries.

4.5 Bivariate Local Moran's I

The bivariate local Moran's I statistic characterizes the statistical relationship between a variable at a given location and a spatially lagged second variable at neighboring locations. The results of bivariate local Moran's I depict in Figure 12. Figure 12 illustrates the cluster areas of two variables: the confirmed cases of COVID-19 (COV) as of October 10, 2021, and the independent variables. The results indicate that cluster areas for COV and adult smoking rate (SMK), number of older people (OLD), and life expectancy index (LEI) were not observed in the ASEAN countries, as depicted in Figures 13(a), 13(b), and 13(j). Therefore, it can be inferred that COVID-19 was not associated with the

clustering of SMK, OLD, and LEI. A heightened incidence of disease infection in Indonesia and the Philippines was evident, with these countries surrounded by nations characterized by increased urban population, HDI, and GDP per capita, as depicted in Figures 13(f), 13(h), and 13(i), respectively. Malaysia exhibited a high number of COVID-19 confirmed cases, and it was encircled by countries with elevated HDI and GDP per capita, as illustrated in Figures 13(h) and 13(i), respectively.

In contrast, Myanmar and Vietnam showed low COVID-19 confirmed cases, with both countries surrounded by regions with low urban populations, as portrayed in Figure 12(f). Additionally, Cambodia displayed a lower incidence of COVID-19 cases compared to its neighboring countries, despite having a high number of international tourist arrivals, as depicted in Figure 12(k). Thailand, characterized by a high number of confirmed COVID-19 cases, found itself surrounded by countries exhibiting low urban population, HDI, and GDP per capita, as depicted in Figures 12(f), 12(h), and 12(i). The outcomes of bivariate local Moran's I are pivotal in assessing the likelihood of disease spread, potentially influenced by factors from surrounding countries. Additionally, these results can inform and guide the formulation of effective disease control strategies at the national level.

4.6 Spatial Regression Models

The association between confirmed COVID-19 cases and the independent variables, as illustrated in Table 2, was established through the application of Ordinary Least Squares (OLS), Spatial Lag Model (SLM), and Spatial Error Model (SEM) techniques. GeoDa served as the tool for conducting these modeling procedures, and the outcomes of the three models are presented in Table 4.

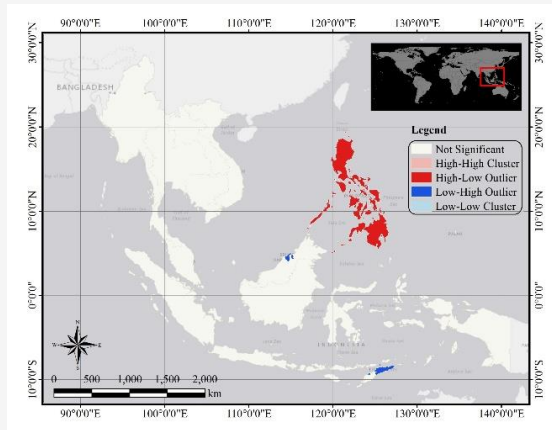


Figure 12: Cluster analysis utilizing Anselin Local Moran's I

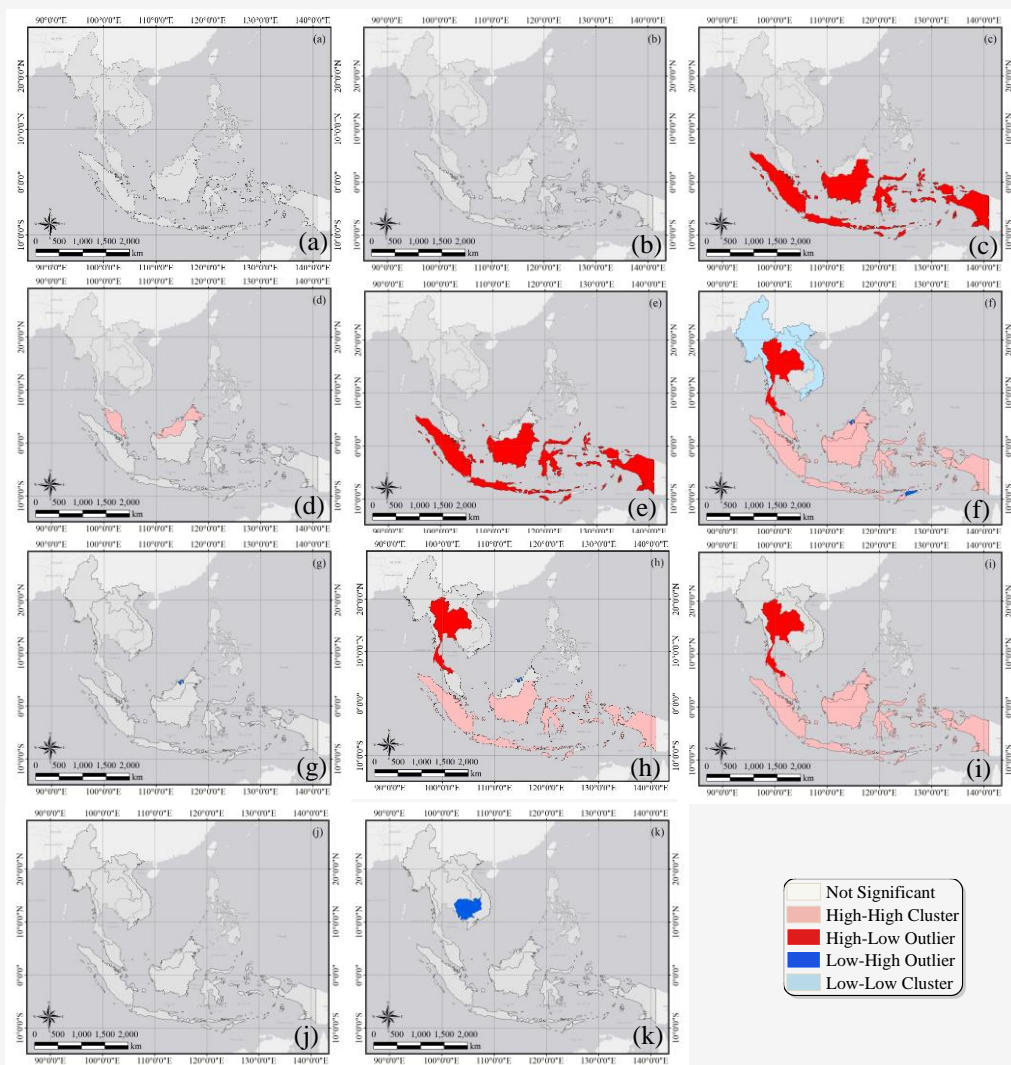


Figure 13: Bivariate Local Moran's I (a) adult smoking rate (b) number of older people (65+) (c) number of cardiovascular patients (d) number of lung cancer patients (e) total population number (f) urban population (g) population density (h) human development index (i) GDP per capita (j) Life expectancy index (k) International tourist arrival

Table 4: COVID-19 confirmed cases spatial regression models

Variables	OLS		SLM		SEM	
	Coef.	p-value	Coef.	p-value	Coef.	p-value
Behavioral factors						
Constant	0	0.60	-784,304	0.64	0	0.22
SMK	73,038.8	0.32	93,585.9	0.04*	97,636.7	0.02*
OLD	112,728	0.52	76,995.1	0.49	107,773	0.25
R ²	0.125		0.471		0.545	
AIC	364.263		344.303		341.643	
BIC	347.456		345.894		342.837	
Health condition factors						
Constant	568,175	0.89	0	0.68	0	0.68
CAR	44,196.6	0.71	5,6095.3	0.49	7,5771.1	0.39
LNG	-2,120.13	0.88	-3,149.7	0.74	6,054.86	0.50
R ²	0.04		0.32		0.00	
AIC	347.237		346.509		344.666	
BIC	348.431		348.100		345.849	
Socioeconomics factors						
Constant	0	0.91	0	0.49	0	0.79
TPN	0	0.22	0.0106	0.00*	0	0.05*
UBP	42,397.1	0.42	59,324.9	0.03*	36852.2	0.14
PDT	608.034	0.16	271.532	0.46	670.126	0.00*
HDI	0	0.52	0	0.90	0	0.16
GDP	-120,278	0.03*	-87,180.7	0.02*	-125,077	0.00*
LEI	-126,730	0.68	17734	0.93	-208,768	0.28
ITA	-172.595	0.40	-140.289	0.14	-185.886	0.03*
R ²	0.977		0.979		0.982	
AIC	316.194		317.364		315.45	
BIC	319.377		320.945		318.633	

*Significant at $p < 0.05$

Table 4 highlights that the SEM stands out as the most suitable among various models, primarily due to its possession of the highest R² and the lowest AIC and BIC values. Conversely, OLS exhibits the lowest R². However, all three models exhibit robust correlation, with R² values nearing 1.00, indicating their collective ability to explain approximately 97% of the variations in the target variable (number of COVID-19 confirmed cases). Notably, certain independent variables in the models have p-values exceeding the significance level of 0.05, signaling their lack of statistical significance. Consequently, these variables should be excluded from the model, as changes in them do not correlate with changes in the dependent variable. Specifically, Table 4 illustrates that only GDP is associated with COVID-19 confirmed cases in the OLS model. Additionally, across all three models, health conditions are not found to be significantly associated with COVID-19 confirmed cases.

Examining the p-values in Table 4 reveals that health condition variables, such as the number of patients with cardiovascular disease and lung

cancer, as well as behavioral variables like age greater than 65 years, and certain socioeconomic factors like the percentage of urban population, HDI, and LEI should be eliminated from the models due to p-values exceeding 0.05.

After excluding unassociated variables from the model, the development of all three models is detailed in Tables 5 and 6. It is evidenced that all of the R² in Table 5 were significantly improved compared to Table 4. However, the strength of correlation is moderately strong. Notably, in the realm of behavioral factors, smoking behavior exhibits a p-value below the designated threshold in the SEM, while it does not demonstrate statistical significance in the OLS and SLM. Consequently, the correlation between smoking behavior and COVID-19 appears somewhat questionable. This finding aligns with [34] who assert that assessing the relationship between smoking and COVID-19 is challenging. It is important to note that smoking may elevate the risk of severe COVID-19 symptoms and exacerbate the condition in patients with COVID-19 [34].

Table 5: Adjusted COVID-19 confirmed cases spatial regression models (behavioral factor)

Variables	OLS		SLM		SEM	
	Coef.	<i>p</i> -value	Coef.	<i>p</i> -value	Coef.	<i>p</i> -value
Behavioral factors						
Constant	-1,379.45	1.00	167,415	0.87	-431,593	0.66
SMK	49,206.1	0.41	77,991.1	0.05*	74,315.9	0.08
R ²	0.076		0.452		0.471	
AIC	344.859		342.762		340.865	
BIC	345.655		343.956		341.661	

Table 6: Adjusted COVID-19 confirmed cases spatial regression models (socioeconomics factors)

Variables	OLS		SLM		SEM	
	Coef.	<i>p</i> -value	Coef.	<i>p</i> -value	Coef.	<i>p</i> -value
Socioeconomics factors						
Constant	-11,44,212	0.81	0.815	0.01*	204,763	0.00*
TPN	0.013	0.01*	0.01	0.00*	0.017	0.00*
PDT	-132.975	0.62	-423.84	0.01*	-710.617	0.00*
GDP	10,264.2	0.68	27,242.3	0.04*	54,967	0.00*
ITA	162.922	0.26	111.049	0.01*	340.898	0.00*
R ²	0.821		0.911		0.961	
AIC	332.830		328.276		323.663	
BIC	334.819		330.664		325.665	

Table 6 illustrates that after the exclusion of variables with a *p*-value exceeding 0.05, there was a reduction in the R² values across the three models. Nevertheless, the correlation strength remains robust. In the OLS model, only TPN demonstrates an association with COVID-19 confirmed cases. Conversely, both the SLM and SEM reveal that TPN, PDT, GDP, and ITA are statistically linked to the disease. Employing the best-fit model evaluation criteria, SEM emerges as the most suitable model due to its superior R² and lower AIC and BIC in comparison to alternative models. The relationship between COVID-19 confirmed cases and socioeconomic factors is delineated in equation 10.

$$COV = 204,763 + 0.017TPN - 710.617PDT + 54,967GDP + 340.898ITA$$

Equation 10

where:

<i>COV</i>	is number of COVID-19 confirmed cases
<i>TPN</i>	is total population numbers
<i>PDT</i>	is population density [ppl/sq.km]
<i>GDP</i>	is gross domestic product per capita [USD]
<i>ITA</i>	is international tourist arrival

While the SEM model is deemed the most appropriate for describing the correlation between the number of confirmed COVID-19 cases and socioeconomic factors, as evidenced by a high R², it

is crucial to note that the variables written in Equation 10 do not imply causation of the disease but only present the geographically statistical correlation.

5. Conclusion

COVID-19 clusters, including both hot spots and cold spots, were not observed in ASEAN countries; however, outliers were identified in certain nations. The Philippines exhibited a high-low outlier, indicating a high number of COVID-19 patients compared to its neighboring countries. Conversely, Brunei Darussalam and Timor-Leste had low-high outliers, suggesting a low number of patients relative to their surrounding nations, during the 2021 pandemic. This highlights varying infection rates and patterns across the region.

Bivariate local Moran's I serves as a robust tool for examining spatial autocorrelation, identifying clusters, and pinpointing outliers between two variables. The outcomes of bivariate local Moran's I analysis indicate prominent hot spots (HL and HH) of COVID-19 and independent variables in Thailand and Indonesia, while cold spots (LH and LL) are predominantly observed in Brunei Darussalam and Timor-Leste. These findings are crucial for formulating targeted disease control strategies at the national level.

The correlation between confirmed COVID-19 cases and associated factors such as behavioral, healthcare, and socioeconomic parameters was investigated using three modeling techniques (OLS,

SLM, and SEM). The p-values of independent variables underscore that the COVID-19 infection rate is predominantly linked with socioeconomic variables, particularly total population numbers, population density, GDP per capita, and international tourist arrivals. The association between smoking behavior and the COVID-19 infection rate remains questionable, as the p-values derived from SEM and SLM differ. Consequently, only socioeconomic factors were considered in the modeling process.

In this study, SEM was chosen as the most appropriate model, as it accounts for 96.1% of the variability in COVID-19 confirmed cases and demonstrates the lowest AIC and BIC values among the considered models. However, it is important to note that the correlations identified in this study do not imply a causal relationship between socioeconomics and COVID-19 infection in ASEAN countries.

6. Limitations and Suggestion

This study concentrated on the regional scale, specifically the ASEAN region. Consequently, the findings presented herein provide an overarching perspective on the disease within the studied area. However, owing to travel restrictions and lockdowns in various countries during the COVID-19 pandemic from 2020 to 2022, the spatial relationships among ASEAN countries were disrupted. Consequently, there is a compelling rationale to downscale the study to the national level for a more nuanced and insightful analysis of spatial dynamics within individual nations.

References

- [1] World Health Organization., (2024). *Coronavirus Disease (COVID-19) Pandemic*. Available: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. [Accessed Dec. 25, 2023].
- [2] Zhang, H., Suepa, T., Hong, L., Navelin, P., Mot, L. and Chakpor, A., (2021). Geospatial Analysis of Covid-19 to Respond to Pandemic Outbreaks: A Case Study in Bangkok Metropolitan Region, Thailand. *International Journal of Geoinformatics*, Vol. 17(5), 68–80, <https://doi.org/10.52939/ijg.v17i5.2013>
- [3] Venter, Z. S., Aunan, K., Chowdhury, S. and Lelieveld, J. (2021). Air Pollution Declines During COVID-19 Lockdowns Mitigate the Global Health Burden. *Environmental Research*, Vol. 192. <https://doi.org/10.1016/j.envres.2020.110403>.
- [4] Venter, Z. S., Aunan, K., Chowdhury, S. and Lelieveld, J., (2020). COVID-19 Lockdowns Cause Global Air Pollution Declines. *Proc Natl Acad Sci U S A*. Vol. 117(32). 18984–18990. <https://doi.org/10.1073/pnas.2006853117>.
- [5] Saha, L., Kumar, A., Kumar, S., Korstad, J., Srivastava, S. and Baudhdh, K., (2022). The Impact of the COVID-19 Lockdown on Global Air Quality: A review. *Environmental sustainability (Singapore)*, Vol., 5(1). 5-23. <https://doi.org/10.1007/s42398-021-00213-6>.
- [6] Barua, S. and Nath, S. D., (2021). The Impact of COVID-19 on Air Pollution: Evidence from Global Data. *Journal of Cleaner Production*. Vol., 298. <https://doi.org/10.1016/j.jclepro.2021.126755>.
- [7] Raza, T., Shehzad, M., Farhan, Q. M., Kareem, H. A., Eash, N. S., Sillanpaa, M. and Hakeem, K. R. (2022). Indirect Effects of Covid-19 on Water Quality. *Water-Energy Nexus*. Vol., 5. 29-38. <https://doi.org/10.1016/j.wen.2022.10.001>.
- [8] Chakraborty, B., Bera, B., Adhikary, P. P., Bhattacharjee, S., Roy, S., Saha, S., Ghosh, A., Senquota, D., and Shit, P. K., (2021). Positive Effects of COVID-19 Lockdown on River Water Quality: Evidence from River Damodar, India. *Scientific Report*. Vol., 11(1). <https://doi.org/10.1038/s41598-021-99689-9>.
- [9] Yotha, N., Phimha, S., Prasit, N., Senahad, N., Sirikarn, P. and Nonthamat, A., (2023). Spatial Association Patterns with Cultural and Behaviour with the Situations of COVID-19. *International Journal of Geoinformatics*. Vol., 19(4). 51-63. <https://doi.org/10.52939/ijg.v19i4.2637>.
- [10] Pholputta, L., Glubvong, M., Amornmahaphun, S., Jiranukul, J., Thongkrajai, P. and Nithikathkul, C., (2021). COVID-19 Lock-Down Affecting Mental Health in Thailand; Review and Situation. *International Journal of Geoinformatics*. Vol. 17(5). 122-129. <https://doi.org/10.52939/ijg.v17i5.2023>.
- [11] Farhan, U., Moazzam, M., Paracha, T., Rahman, G., Lee, B. and Farid, N., (2021). Spatial and Temporal Mapping of COVID-19 Pandemic Using GIS Technique: A Case Study of Italy. *International Journal of Geoinformatics*. Vol., 17(5). 100-108. <https://doi.org/10.52939/ijg.v17i5.2019>.

- [12] Taher, M. T., (2022). Mapping the COVID-19 Spatial Behaviors and Narratives of Women in an Architecture School in the Midwest USA. *Geographies*. Springer International Publishing. 889-905. https://doi.org/10.1007/978-3-030-94350-9_48.
- [13] Xie, Z., Zhao, R., Ding, M. and Zhang, Z., (2021). A Review of Influencing Factors on Spatial Spread of COVID-19 Based on Geographical Perspective. *Int J Environ Res Public Health*. Vol., 18(22). <https://doi.org/10.3390/ijerph182212182>.
- [14] Dutta, I., Basu, T. and Das, A., (2021). Spatial Analysis of COVID-19 Incidence and Its Determinants using Spatial Modeling: A Study on India. *Environmental Challenges*. Vol. 4. <https://doi.org/10.1016/j.envc.2021.100096>.
- [15] Kauh, B., König, J. and Wolf, S., (2023). Spatial Distribution of COVID-19 Hospitalizations and Associated Risk Factors in Health Insurance Data Using Bayesian Spatial Modelling. *Int J Environ Res Public Health*. Vol. 20(5). <https://doi.org/10.3390/ijerph20054375>.
- [16] Paramasivam, C. R. and Venkatramanan, S., (2019). An Introduction to Various Spatial Analysis Techniques. *GIS and Geostatistical Techniques for Groundwater Science*. 23-30. <https://doi.org/10.1016/B978-0-12-815413-7.00003-1>.
- [17] Lin, S., Fu Y., Jia, X., Ding, S., Wu, Y. and Huang, Z., (2020). Discovering Correlations between the COVID-19 Epidemic Spread and Climate. *Int J Environ Res Public Health*, Vol., 17(21). <https://doi.org/10.3390/ijerph17217958>.
- [18] Associate of Southeast Asian Nations., (2024). About ASEAN. Available: <https://asean.org/about-asean/>. [Accessed Dec. 25, 2023].
- [19] Worldmeter., (2024). South-Eastern Asia Population. Available: <https://www.worldometers.info/world-population/south-eastern-asia-population/>. [Accessed Dec. 25, 2023].
- [20] Kanav, A., Yadav, B., Sharma, R. and Kumar J., (2024). Spatio-temporal Analysis of COVID-19 Hotspots in India Using Geographic Information Systems. *International Journal of Geoinformatics*. Vol., 20(1). 72-87. <https://doi.org/10.52939/ijg.v20i1.3027>.
- [21] Ukrainski, P., (2018). How Are Objects' Quantitative Characteristics Spatially Distributed? Finding an Answer with Global Moran's I. Available: <http://www.50northspatial.org/global-morans-i-spatial-autocorrelation/>. [Accessed Nov. 19, 2023].
- [22] Anselin, L., (1995). Local Indicators of Spatial Association—LISA. *Geogr Anal.*, Vol., 27(2). 93-115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>.
- [23] Rattanahon, P., Laohasiriwong, W., Nilnate, N., Yotha, N., Phanyotha, S., Thammasarn, K., Senahad, N., and Prasit, N. (2022). The Influence of the Density of Buddhist Temples on the Spatial Distribution of Happiness of the Elderly in Thailand. *International Journal of Geoinformatics*, Vol. 18(3), 123–130. <https://doi.org/10.52939/ijg.v18i3.2213>.
- [24] Nilnate, N., Jirapornkul, C. and Limmongkon, Y., (2022). Spatial Factors Associated with fall among the Elderly in Thailand. *International Journal of Geoinformatics*, Vol. 18(5), 105–113. <https://doi.org/10.52939/ijg.v18i5.2391>
- [25] GeoDa. Local Spatial Autocorrelation (3). (n.d.). Available: https://geodacenter.github.io/workbook/6c_local_multi/lab6c.html. [Accessed Sept. 9, 2023].
- [26] University Consortium for Geographic Information Science. AM-23 - Local Measures of Spatial Association. Available: <https://gistbok.ucgis.org/bok-topics/local-measures-spatial-association>. [Accessed Dec. 25, 2023].
- [27] Suerungruang, S., Laohasiriwong, W. and Sornlorm, K., (2023). Spatial Association and Modeling of Infant Mortality in Thailand, 2020. *International Journal of Geoinformatics*, Vol. 19(8), 28–41. <https://doi.org/10.52939/ijg.v19i8.2779>.
- [28] Catma, S., (2021). The Price of Coastal Erosion and Flood Risk: A Hedonic Pricing Approach. *Oceans*. Vol., 2(1), 149-161. <https://doi.org/10.3390/oceans2010009>.
- [29] GeoDa. (n.d.). Spatial Regression with GeoDa. Available at: <https://s4.ad.brown.edu/resources/tutorial/modul2/geoda3final.pdf>. [Accessed Dec. 25, 2023].
- [30] Kawasaki, A., Kawamura, G. and Zin, W., (2019). A Local Level Relationship between Floods and Poverty: A Case in Myanmar. *International Journal of Disaster Risk Reduction*. Vol., 42. <https://doi.org/10.1016/j.ijdrr.2019.101348>.

- [31] Wagenmakers, E. J. and Farrell, S., (2004). AIC Model Selection using Akaike Weights. *Psychometc Bullentin & Review*, Vol. 11(1), 192-196. <https://doi.org/10.3758/BF03206482>
- [32] Analyttica Datalab., (2019). What is Bayesian Information Criterion (BIC)?. Available at: <https://medium.com/@analyttica/what-is-bayesian-information-criterion-bic-b3396a894be6>. [Accessed Dec. 25, 2023].
- [33] ASEAN Biodiaspora Virtual Center., (2023). COVID-19 Situational Report in the ASEAN Region. Available at: https://asean.org/wp-content/uploads/2023/05/COVID-19-and-Mpox_Situational-Report_ASEAN-BioDia-spora-Regional-Virtual-Center_05May2023.pdf. [Accessed Dec. 25, 2023].
- [34] Xie, J., Zhong, R., Wang, W., Chen, O. and Zou, Y., (2021). COVID-19 and Smoking: What Evidence Needs Our Attention?. *Front Physiol.* Vol., 12. <https://doi.org/10.3389/fphys.2021.603850>.