Integrating Satellite Imageries and Multiple Geospatial Big Data for Granular Mapping of Spatial Distribution of the Human Development Index in East Java, Indonesia

R Ramadhan [1], **A W Wijayanto**[1,2,*]

[1] Department of Statistical Computing, Politeknik Statistika STIS, Jakarta, Indonesia
[2] BPS-Statistics Indonesia, Jakarta, Indonesia

*Corresponding author's e-mail: ariewahyu@stis.ac.id

Abstract. The availability of data on the Human Development Index (HDI) is crucial as a gauge of regional performance, particularly in terms of assessing the development of human resources. In Indonesia, the collecting of HDI data uses the conventional method, such as undirect estimation, National Socio-Economic Survey (SUSENAS), The Ministry of Religion, or inventory of sectoral data that used the large cost, time, and effort. Additional data are required to provide more detailed poverty data at a lower cost and with more recent information to overcome these limitations. According to recent studies, the quality of life for measuring HDI can be identified down to the granular level using geospatial big data. Therefore, the contribution of this research is to implement the use of geospatial big data, such as integrated multi-source satellite imagery data and Point of Interest (POI). Besides that, this study develops the relative spatial-human development index in 11 km x 11 km resolution for the granular mapping of the quality of life to measure the HDI in East Java, Indonesia. The kinds of weighted sum models used in this study such as equal weight (EWS), Pearson (PCCWS), Spearman (SCCWS) correlation-based weight, and Principal Component Analysis (PCA)-based weight (PCAWS). The best RSHDI PCCWS for representing the human development index in East Java in 2022, which was determined using a weight-sum model based on Pearson correlation, has a correlation coefficient of 0.7858 (p-value = 5.078 x 10-9) and is highly correlated with official HDI data. The use of this RSHDI as a predictor variable in the estimation of HDI data shows the ideal model had an RMSE of 3.098% and an R^2 of up to 61.75% using RSHDI PCCWS. According to the findings of the descriptive analysis of this map, areas with low RSHDI scores typically in some regencies areas in Madura Island and the east area of East Java with geographically depressed, while areas with high RSHDI scores typically have dense populations and have better accessibility such as urban area in Surabaya and Kota Malang. As a result, the official human development index data can be supported by the RSHDI's ability to map spatially deprive areas.

1. Introduction

As one measure of regional performance, especially in terms of evaluating the process of human resource development, the provision of data on the Human Development Index (HDI) is very important [1]. This data provides a general indication of the development needs and priorities of the population. According to the United Nations Development Program (UNDP), the process of increasing people's options for their income, health, education, and physical environment is known as "human development," and the four main factors that must be taken into account are productivity, equity, sustainability, and empowerment [2]. While the HDI is defined by BPS-Statistics Indonesia as a composite index created by combining the life expectancy index, education index, and decent standard of living index [3]. The HDI is pertinent to be used as a benchmark to determine human development success because the calculation of the life expectancy index, education index, and a decent standard of living index includes economic and non-economic components such as quality education, health, and population.

The calculation of the HDI in Indonesia is updated once a year to the level of the district/city only due to the presence of data. For example UNDP uses the dimensions of longevity and healthy life which are represented by indicators of life expectancy at birth but this calculation in BPS-Statistics Indonesia is done through an indirect approach (indirect estimation), the knowledge dimension is represented by the Expected Years of School and the Average Years Schooled which are the results of conventional survey namely National Socio-Economic Survey (SUSENAS) and the number of students data who are studying with a residence from the Ministry of Religion is the results of an inventory of sectoral data in the regions in Indonesia [4]. Additionally, the UNDP uses information on Gross National Income (GNI) per capita as a measure of a respectable standard of living. [1]. However, since this information is not available at the regional level, an adjusted real per capita expenditure indicator—which is adjusted and calculated using the SUSENAS results as well is used as a substitute. [4]. Thus, the data collection to measure HDI is generally still using conventional methods such as SUSENAS.

SUSENAS typically gathers data through a ground-based house-hold survey that is conducted every six months. Unfortunately, this traditional approach to data collecting is expensive, time-consuming, labor-intensive, and has a limited reach [5]. Due to a lack of coverage and duration across several dimensions, Indonesia's human development data representation is constrained. Therefore, to overcome the shortcomings of the current home survey-based data collection, a human development index estimation with greater scope granularity and needing less expense and time to update is required. Compared to the collection of existing household survey data, the use of satellite image data from remote sensing and other geospatial big data such as point of interest (POI) from OpenStreetMap (OSM) has the potential to complement the limitations of development data and the quality of human life. Remote sensing is a technique for obtaining information about the Earth's surface through electromagnetic radiation generated by a device that is not in direct contact with the object [6, 26, 27]. The POI data contains locations of interest based on specific categories and is personalized according to context [7]. Thus, data collection using such methods can provide more accurate data [8, 28], because its uniqueness and objectivity in observing socio-economic and physical phenomena from different perspectives effectively [9, 29], and the use of lower costs, up to date, as well as the representation of the area coverage is granular [10,30,31]. However, data derived from such sources cannot be used directly as a substitute for conventionally obtained data such as SUSENAS, but can be used as alternative supporting data [11].

Recent research has demonstrated that geographic big data, such as nighttime light (NTL), may reveal population density [12], gross domestic product (GDP)[13], and electricity usage [14] which might identify the socioeconomic situations. Besides that, using the normalized difference vegetation index (NDVI) was significantly high positive correlated with HDI [15], and land surface temperature (LST) could show the high and low income with urban thermal which could identify the expenditure indicator [16]. The normalized difference built-up index (NDBI) could have the potential to identify the urban areas with the normalized difference water index (NDWI) which provide the accurate urban land to show the good quality of life area [17]. Moreover, the air pollution such as the carbon monoxide (CO) and the nitrogen dioxide (NO_2) is related with the regional economic growth and city level characteristic that could identify the quality of life indicator [18][19], and the sulfur dioxide (SO_2) could be used to identify the energy consumption [20]. The other geospatial big data such as POI density and POI cost distance could show the regional economic development to identify the poverty for show the quality of life [9]. Therefore, the difference of geographical characteristic could show the quality of life for show the human development. In Indonesia, the use of geospatial big data such as remote sensing and POI for HDI mapping is still limited in granular level. To support official HDI data, this research is focused on developing a relative spatial human development index (RSHDI) based on multisource remote sensing and other geospatial big data, such as POI. This method uses an aggregation at 11 x 11 km level with granular mapping estimation that requires less time to update. As a result, it is anticipated that policy decisions will be more successful in achieving their objectives.

2. Method

2.1. Study Area

The East Java is the province in Indonesia which has 38 regencies/municipalities with 666 sub-districts and has been selected in this research due to the lowest number of HDI in Java in 2022 in 72.75 [4]. This research is expected to make development planning in East Java will be good and accurate if available data and information to the granular level (in this study using 11 km) so that it can monitor and evaluate towards human resource development comprehensively to know faster progress in quality of life up from the side of the coverage of its territory. The distribution of official HDI data at the regency/municipality level is shown in Figure 1 along with a map of East Java, the case study, for the year 2022. The high HDI values tend in urban areas such as Kota Surabaya and Kota Malang, which have a HDI of 82.74 and 82.71 respectively. While areas with low HDI tend to be in central areas such as Lumajang District (66.95) and Madura Island such as Sampang District (63.69) and Pamekasan (66.99). Then in the western area has an average HDI above 70 or has a fairly good quality of human life such as Ngawi District (71.75), Nganjuk (72.93), and Jombang (74.05).

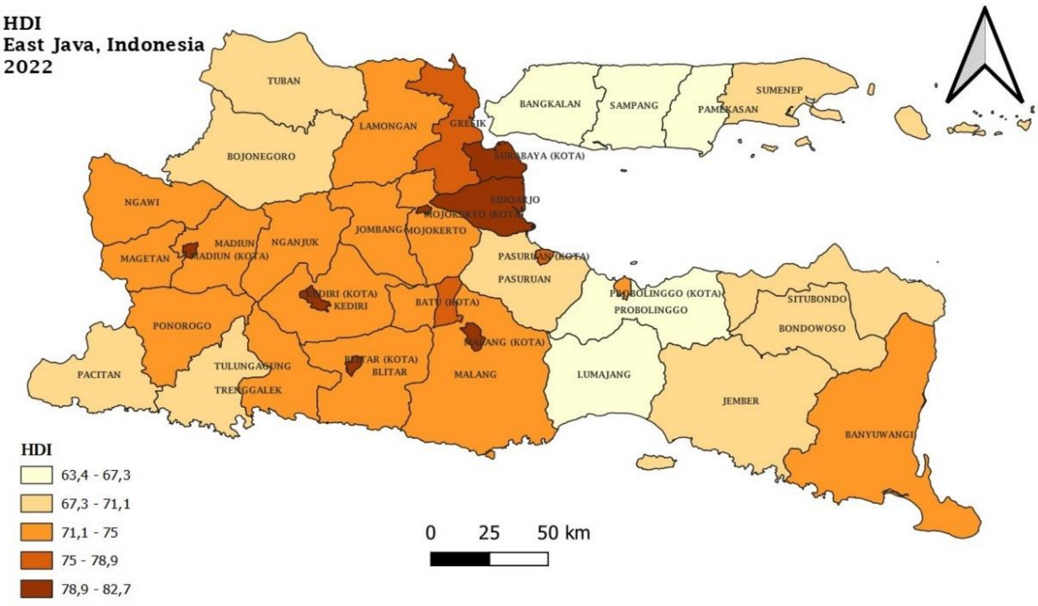

Figure 1. East Java Province as the Study Area with its Official Human Development Index (HDI)

2.2. Data Source

In this research, point-of-interest (POI and multi-source satellite images were used to construct a model for estimating poverty. Table 1 contains comprehensive details about the datasets.

Table 1. Summary of Data Source and Variables.

Data Source	Variable	Band	Unit	Spatial Resolution
Visible Infrared Imaging Radiometer Suite (VIIRS)	Nighttime Light Intensity (NTL)	avg_radian	nanoWatts/ cm2/sr	750 m
Sentinel Multispectral Level 2A	Normalized Difference Vegetation Index (NDVI)	B4 (Red) dan B8 (NIR)	index	10 m
	Normalized Difference Water Index (NDWI)	B3 (Green) dan B8 (NIR)		
	Normalized Difference Built-Up Index (NDBI)	B8 (NIR) dan B11 (SWIR 1)		
	Soil Adjusted Vegetation Index (SAVI)	B4 (Red) dan B8 (NIR)		
Moderate-resolution Imaging Spectroradiometer (MODIS)	Day Time Land Surface Temperature (LST)	LST Day 1 km	Kelvin	1000 m
	Night Time Land Surface Temperature (LST)	LST Night 1 km		
Sentinel-5P	Carbon Monoxide (CO)	CO Column Number Density	mol/m2	1113.2 m
	Nitrogen Dioxide (NO_2)	NO2 Column Number Density		
	Sulfur Dioxide (SO_2)	SO2 Column Number Density		
OpenStreetMap (OSM)	POI Density	-	point	-
	POI Distance	-	meter	-

In this study, data from remote sensing—including multi-source satellite images—as well as other geographic big data—including Point of Interest (POI) data from OpenStreetMap (OSM)—are used. The multi-source satellite images used in this work include Nighttime Light (NTL) intensity from NOAA-VIIRS, Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI), Normalized Difference Water Index (NDWI), and Normalized Difference Built-Up Index (NDBI). NDVI and SAVI are a vegetation index that is analyzed through reflection brightness and absorption of Near-Infrared (NIR) and red band and can use to identify the quality of life [21]. Then, NDBI could have the potential to identify the urban areas [16], and NDWI which provide the accurate urban land and water area [17], that both can identify the poverty areas. These are the composing indices that were utilized for this study. The formula used to calculate NDVI, NDBI, and NDWI is as follows:

$$NDVI = \frac{NIR_{band\ 8} - RED_{band\ 4}}{NIR_{band\ 8} + RED_{band\ 4}} \tag{1}$$

$$NDWI = \frac{Green_{band\ 3} - NIR_{band\ 8}}{Green_{band\ 3} + NIR_{band\ 8}} \tag{2}$$

$$NDBI = \frac{SWIR_{band\ 1} - NIR_{band\ 8}}{SWIR_{band\ 1} + NIR_{band\ 8}} \quad (3)$$

$$SAVI = \frac{NIR_{band\ 8} - RED_{band\ 4}}{(NIR_{band\ 8} + RED_{band\ 4} + 0.5)}\ x\ (1 + 0.5) \quad (4)$$

Other data from Sentinel-5P and MODIS were also used in this analysis, including daytime land surface temperature (LST), nitrogen dioxide (NO2), carbon monoxide (CO), and sulfur dioxide (SO2). Remote sensing satellite imaging data is received and analyzed using Google Earth Engine (GEE), a cloud-based platform designed to store and analyze geographic data for the Earth. POI data is a different type of geographic big data collection that was gathered using OSM and used in this investigation. Up until December 31, 2022, there were 17.542 points in East Java according to this study. These points were evaluated based on several criteria, including those related to economics, tourism, health, and education. Figure 2 through 13 in this study's data visualization display the study's data.

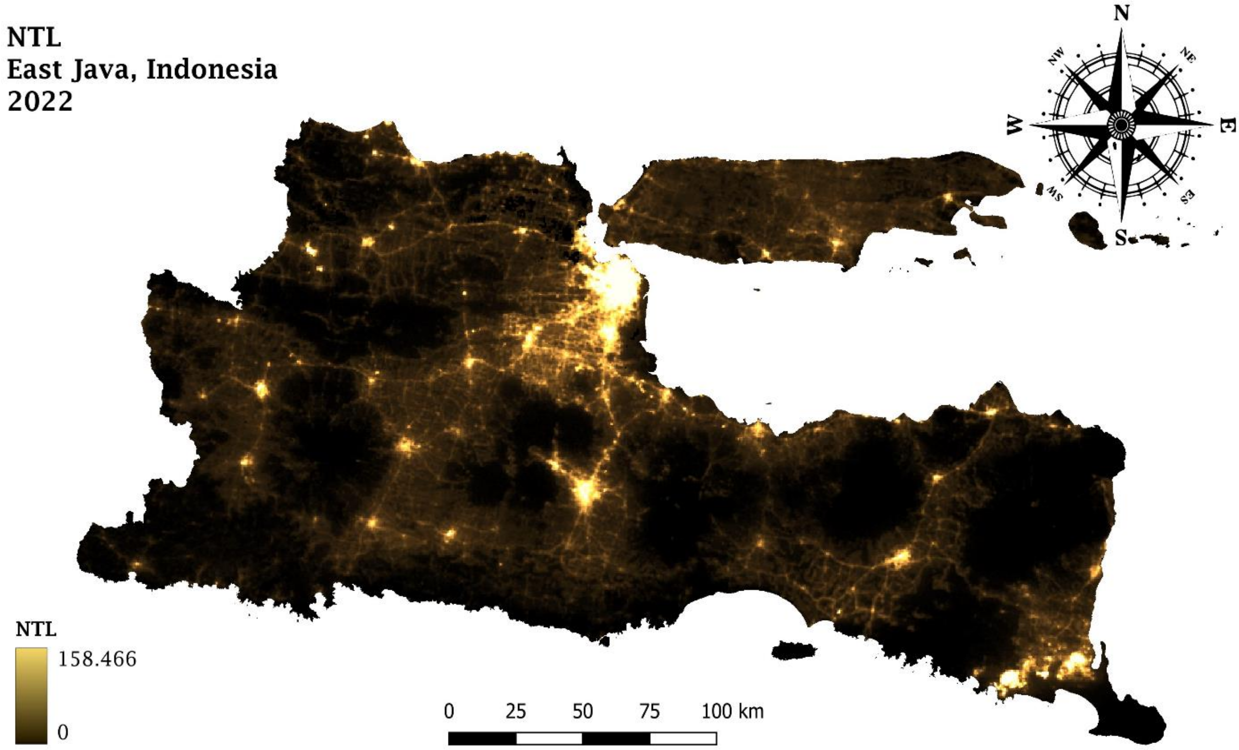

Figure 2. Data visualization of NTL (nanoWatts/cm^2/sr) in East Java, Indonesia 2022 from VIIRS

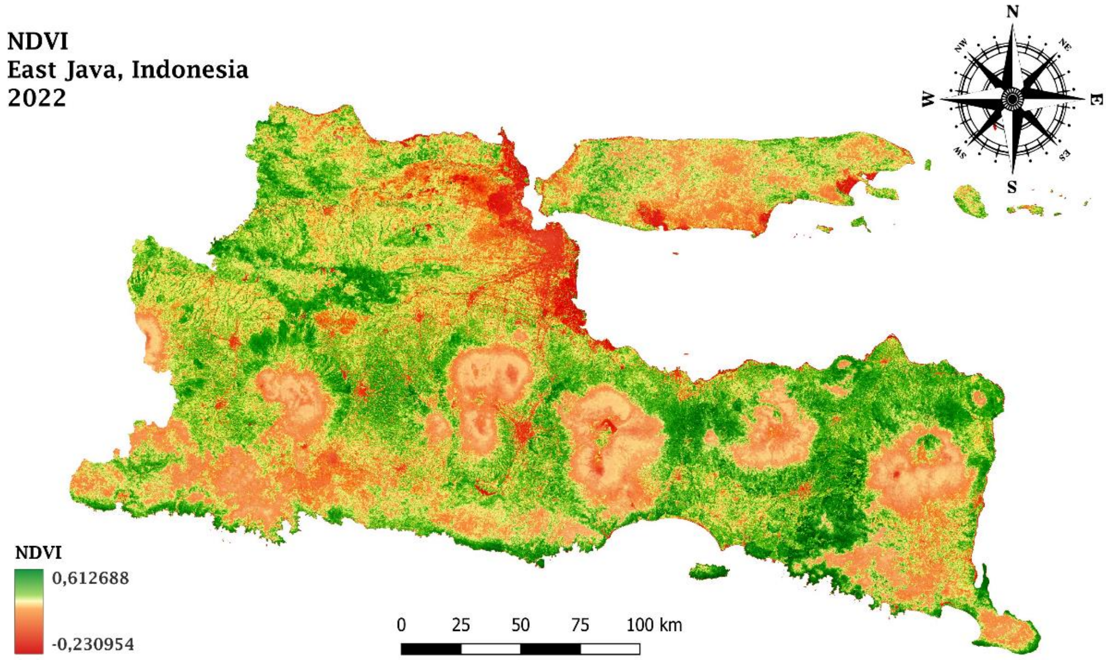

Figure 3. Data visualization of SAVI (index) in East Java, Indonesia 2022 from Sentinel MSI-Level 2A

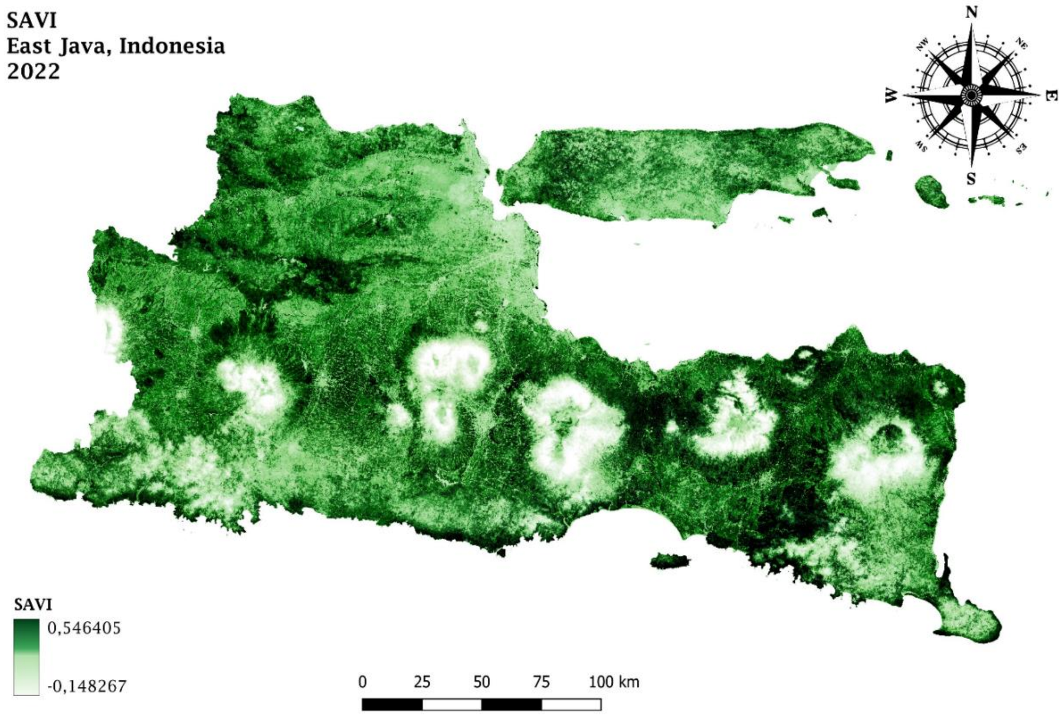

Figure 4. Data visualization of SAVI (index) in East Java, Indonesia 2022 from Sentinel MSI-Level 2A

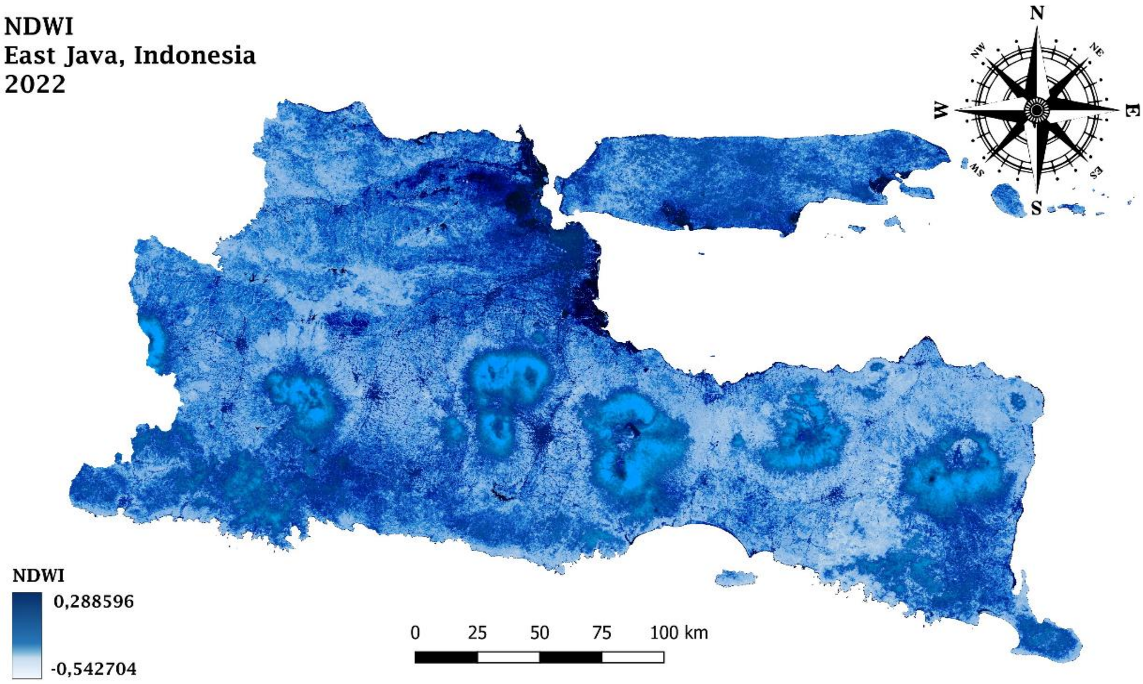

Figure 5. Data visualization of NDWI (index) in East Java, Indonesia 2022 from Sentinel MSI-Level 2A

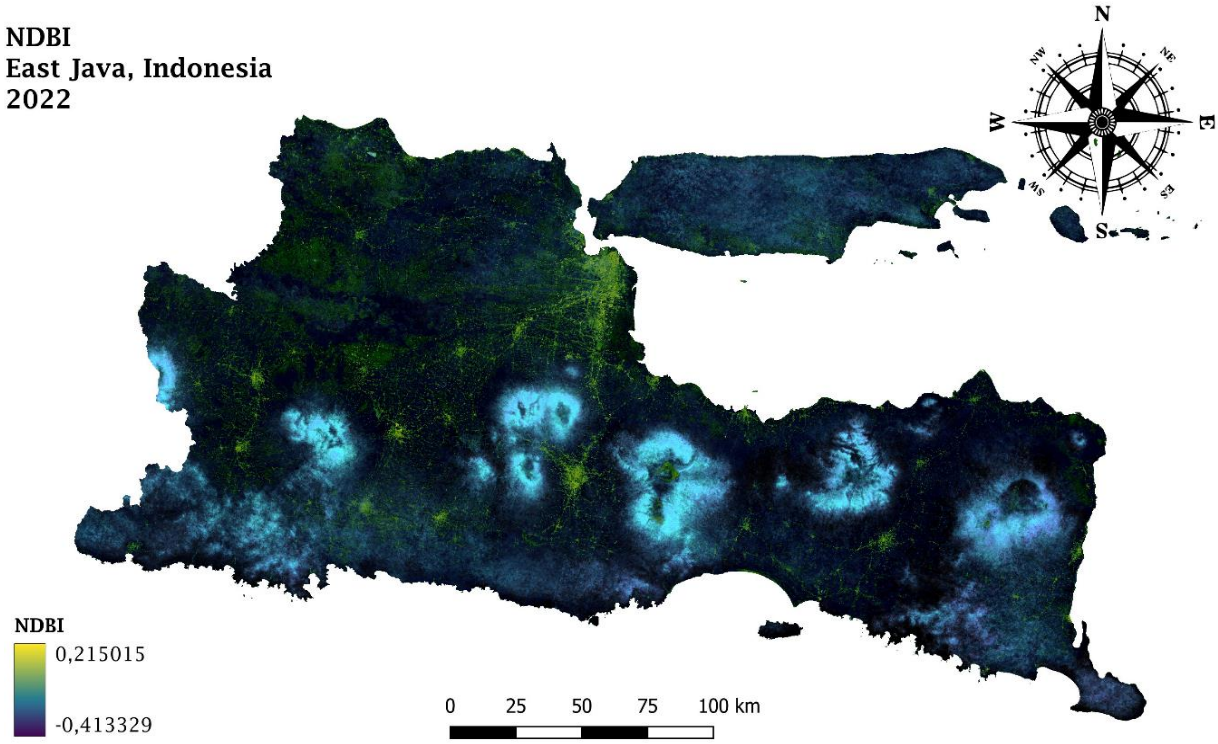

Figure 6. Data visualization of NDBI (index) in East Java, Indonesia 2022 from Sentinel MSI-Level 2A

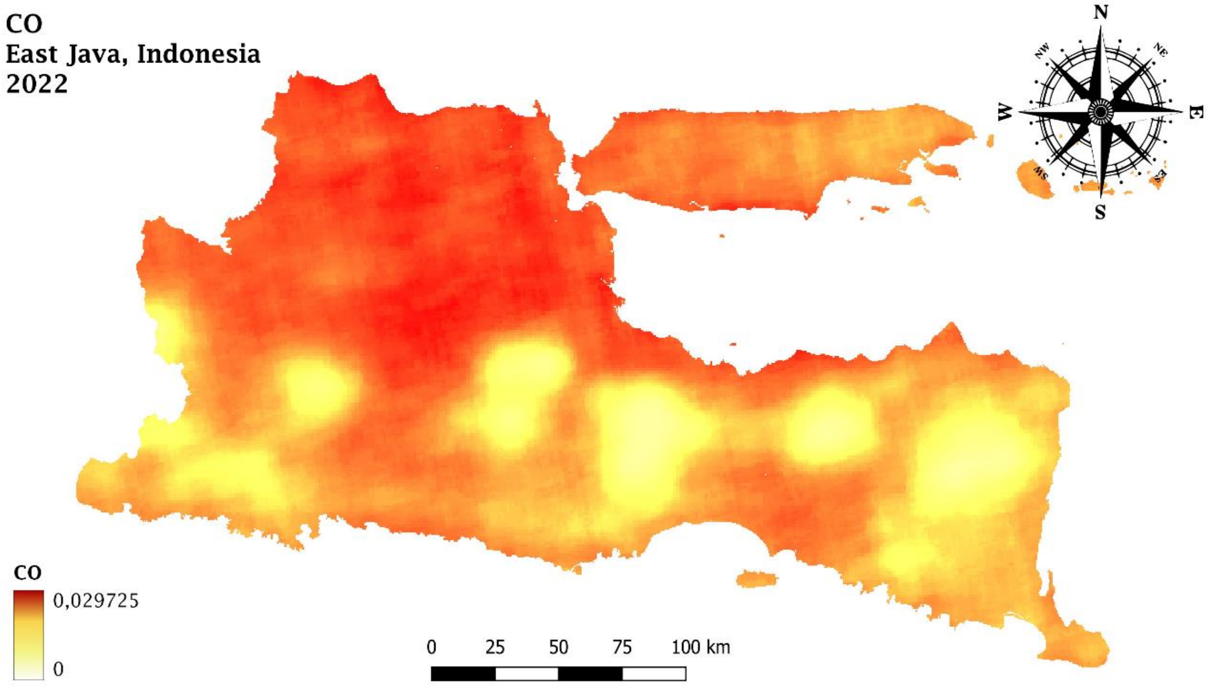

Figure 7. Data visualization of CO (mol/m^2) in East Java, Indonesia 2022 from Sentinel-5P

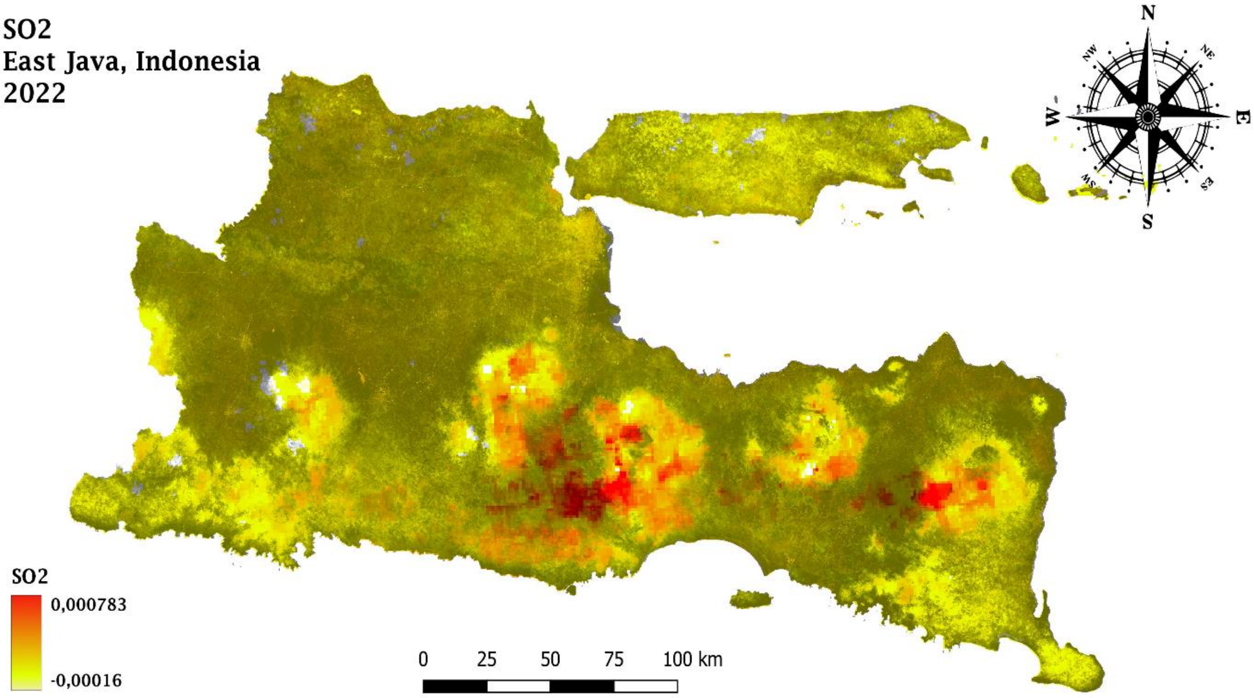

Figure 8. Data visualization of SO_2 (mol/m^2) in East Java, Indonesia 2022 from Sentinel-5P

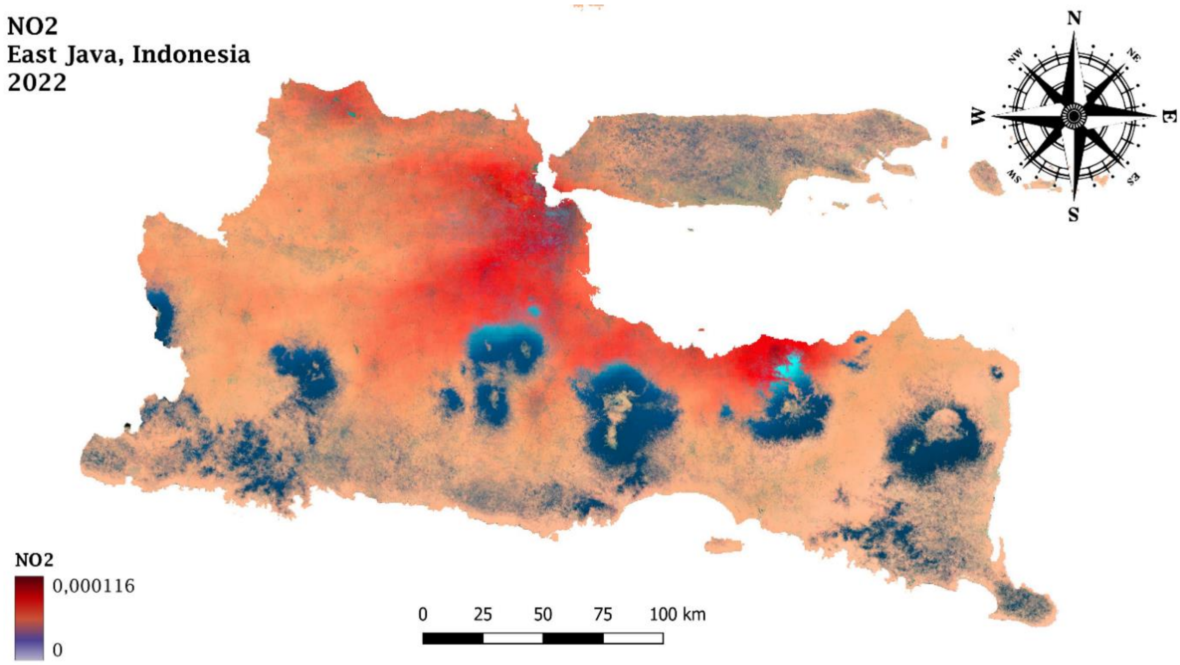

Figure 9. Data visualization of NO_2 (mol/m^2) in East Java, Indonesia 2022 from Sentinel-5P

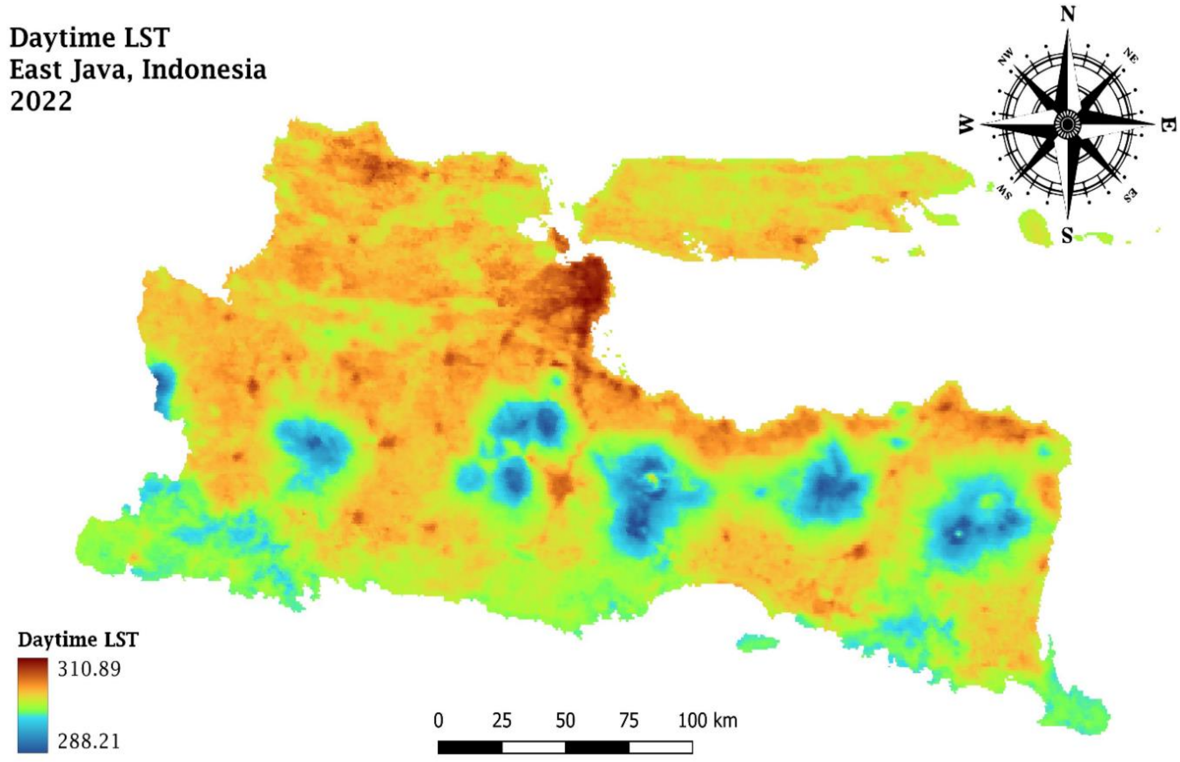

Figure 10. Data visualization of Daytime LST (Kelvin) in East Java, Indonesia 2022 from MODIS

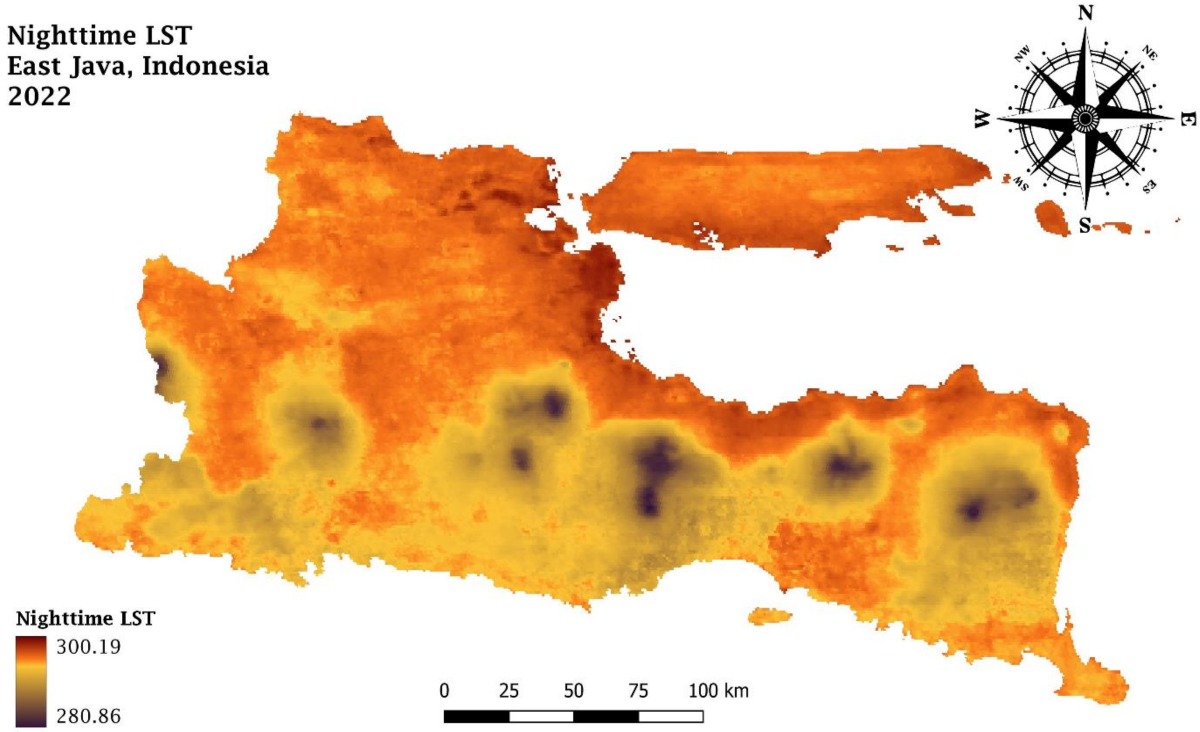

Figure 11. Data visualization of Nighttime LST (Kelvin) in East Java, Indonesia 2022 from MODIS

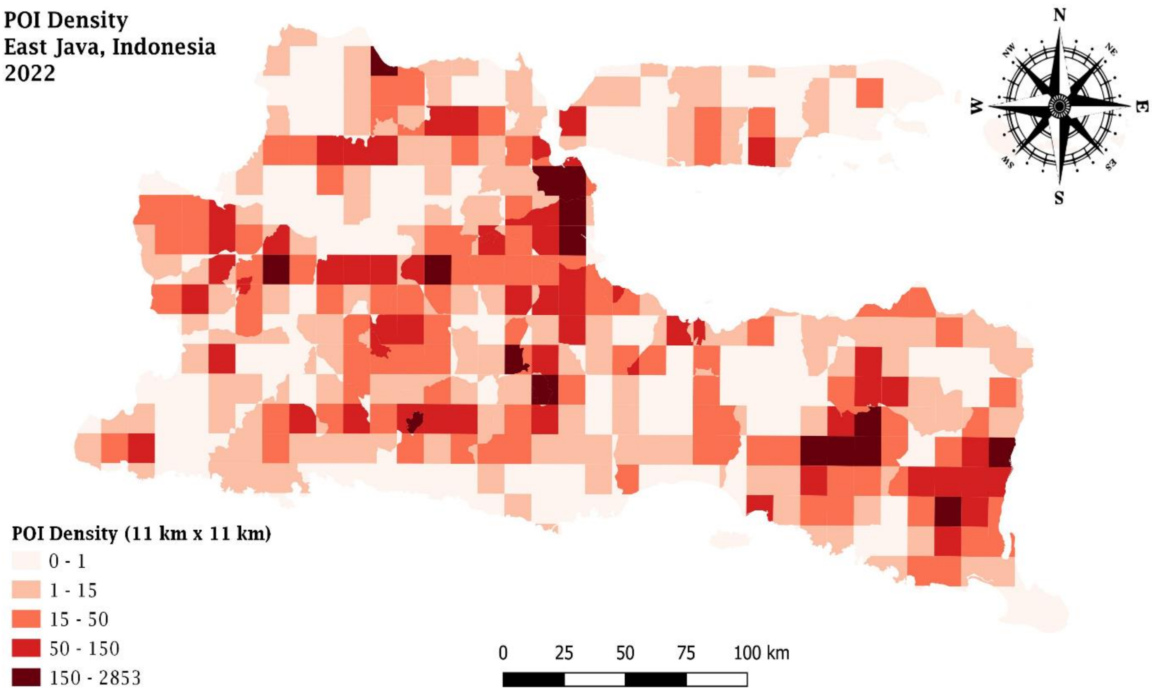

Figure 12. Data visualization of POI Density (points) in East Java, Indonesia 2022 from OSM

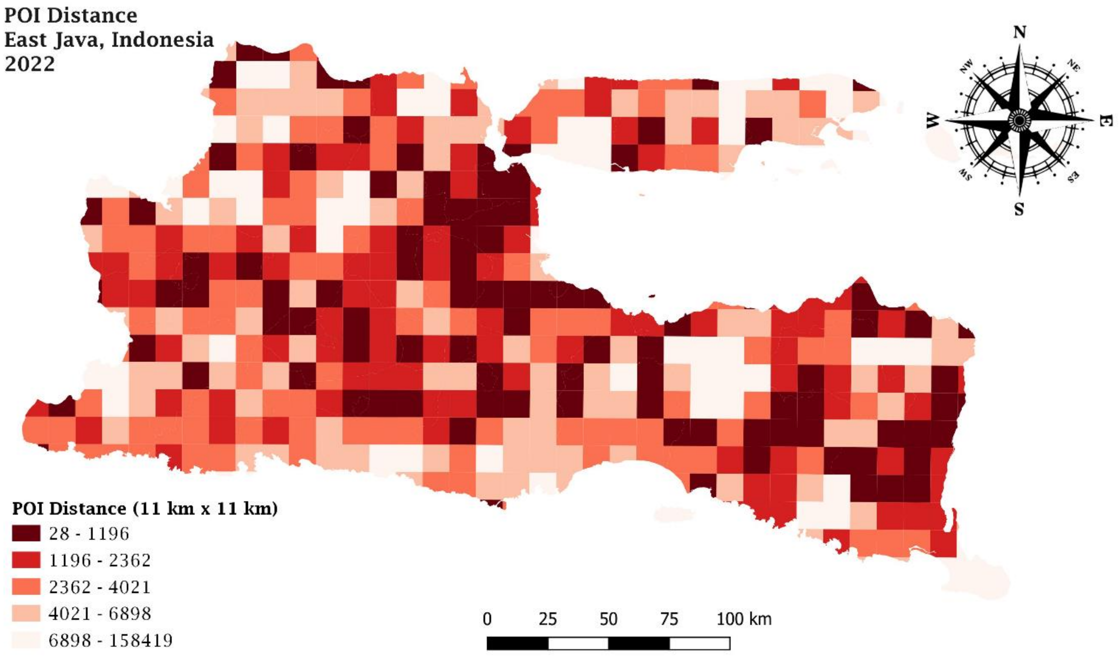

Figure 13. Data visualization of POI Distance (meter) in East Java, Indonesia 2022 from OSM

2.3. Research Framework

In this study, a relative spatial human development index (RSHDI), which describes human development index with better scope granularity, at a lower cost, and with shorter update times, is being developed. The purpose of relative spatial human development index (RSHDI) is to improve upon the shortcomings of the current household survey-based approaches to poverty data collection. Beginning with data collection, pre-processing, integration, transformation, correlation analysis, variable selection, and relative spatial human development index (RSHDI) calculation, the study then validated and interpreted the findings. We used R Studio, Python 3.6.9, and QGIS 3.28 to carry out our analysis and visualization. The 11 km x 11 km resolution relative spatial human development index (RSHDI) spatial human development index map and its validation were the expected results. Figure 14 of the explanations below provide more thorough explanations.

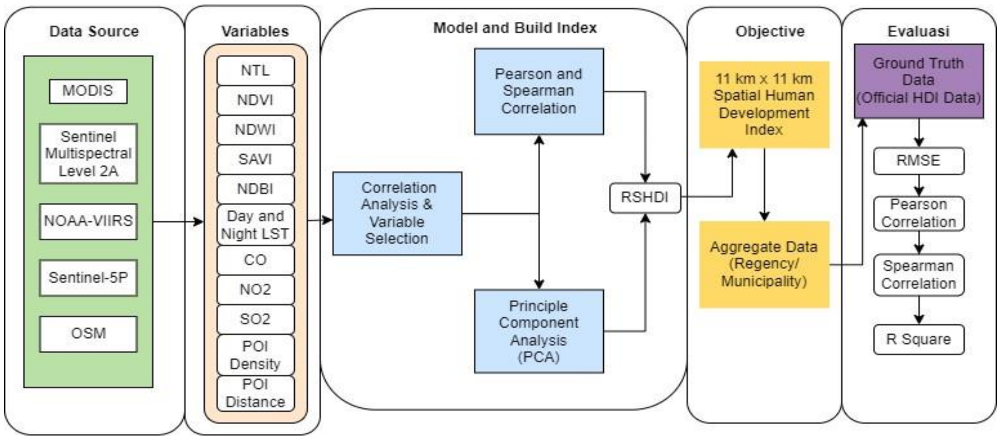

Figure 14. The Research Framework

2.4. Data Transformation

In order to handle data values that can be both positive and negative, the Yeo-Johnson power transformation, a variant of the Box-Cox transformation, is used [22]. By restructuring the variables to fit a Gaussian or more normal distribution, this transformation strategy is excellent for dealing with variables with different units throughout all ranges [23]. The following describes the formula data transformation strategy applied in this study:

$$y_\lambda(x) \begin{cases} \dfrac{(1+x)^\lambda - 1}{\lambda}, \lambda \neq 0 \; dan \; x \geq 0 \\ log(1+x), \lambda = 0 \; dan \; x \geq 0 \\ -\dfrac{(1-x)^{2-\lambda}}{2-\lambda}, \lambda \neq 2 \; dan \; x < 0 \\ -log(1-x), \lambda = 2 \; dan \; x < 0 \end{cases} \tag{5}$$

If the variables follow a normal distribution, a parameter, which is calculated using the Maximum Likelihood approach, is used to apply the Yeo-Johnson transformation to each individual variable or input data value, indicated as *x*. A linear connection is created when the family of transformations has a parameter value of λ=1. The distribution is then modified by the transformation, which condenses the right tail at $\lambda < 1$ to make the distribution right-skewed and more symmetric. In contrast, when $\lambda > 1$, the left tail becomes more symmetrical, especially for left-skewed distributions.

2.5. Relative Spatial Human Development Index (RSHDI) Calculation

Selected geographical variables that indicate human development in East Java are superimposed to construct the relative spatial human development index (RSHDI). The variables were overlaid using a weighted sum model that we created. The weighted sum model is applied for relative spatial human development index (RSHDI) construction in the formula that follows:

$$RSHDI = \sum_{i=1}^{p} w_i x_i \tag{6}$$

where *w* is the allocated weight, *x* is the observed value, and *p* is the total number of overlay variables. The weight computation was done using two different methods. First, we developed the correlation-based weights with Equal Weighted Sum (EWS) that based on the Pearson Correlation Coefficient Weighted Sum (PCCWS) and Spearman Correlation Coefficient Weighted Sum (SCCWS). It is considered that variables with larger correlations more accurately reflect the quality of life. Second, we used the weight based on Principle Component Analysis Weighted Sum (PCAWS). Some previous researches use the correlation and PCA weighted based for granular mapping to estimate such as poverty and quality of life [24][21][25]. This study uses those methods for granular mapping of spatial distribution for human development in East Java, Indonesia.

2.6. Model Evaluation

To ascertain the association between each geographical variable described in this study and the official human development index (HDI) data for East Java, correlation analysis was carried out. Although we wanted to measure the link at the 11 km x 11 km level, we were only able to do so due to the official human development index data's limitations. To ascertain this, both Pearson and Spearman correlation studies were used. The formula to calculate the Pearson and Spearman correlation coefficient (*r*) is shown in the equation below:

$$r_{xy} = \frac{n \sum_{i=1}^{n} x_i \, y_i \; - \; (\sum_{i=1}^{n} x_i) \; (\sum_{i=1}^{n} y_i)}{\sqrt{(n \sum_{i=1}^{n} x_i^2 - \; (\sum_{i=1}^{n} x_i)^2} \sqrt{(n \sum_{i=1}^{n} y_i^2 \; - \; (\sum_{i=1}^{n} y_i)^2}} \tag{7}$$

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n^3 - n} \tag{8}$$

where n is the number of observations and $\boldsymbol{r_{xy}}$ is the correlation between the features and the second feature. By converting the observation value to its difference ranking value ($\boldsymbol{d_i^2}$), the Spearman correlation is determined in the same manner. Between 0 and 1 is the correlation coefficient (r). A positive or negative sign denotes the direction of the connection. After that, the correlation significance test is run to see if the correlation coefficient was statistically significant at the level of 0.05 (significance level). The alternative hypothesis is characterized as supposing that there is a correlation between the two variables, as opposed to the null hypothesis, which states that there is no association between the two variables. The table 2 below instructions for analyzing the correlation coefficient (r) data based on the correlation coefficient [23].

Table 2. Correlation coefficient interpretation

Correlation Coefficient	Interpretation
$0.00 \leq \lvert r \rvert \leq 0.199$	Very weak
$0.20 \leq \lvert r \rvert \leq 0.399$	Weak
$0.40 \leq \lvert r \rvert \leq 0.599$	Moderate
$0.60 \leq \lvert r \rvert \leq 0.799$	Strong
$0.80 \leq \lvert r \rvert \leq 1.000$	Very strong

The variables we selected from this correlation study are statistically significant and are correlated with the East Java official human development index data, as determined by hypothesis testing. To develope the relative spatial human development index (RSHDI), variables that linearly reflect the human development index of East Java were used. The validation evaluation is a critical step in determining how well the relative spatial human development index (RSHDI) can represent human development in East Java. We used numerical evaluation by root mean square error (RMSE) and R^2 to measure the numerical similarity between the generated result and the supplied ground truth data. The calculation is shown in the following formulas:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \underline{y_i})^2} \tag{10}$$

where n is the number of observations, $\boldsymbol{y_i}$ is the real value, $\boldsymbol{\bar{y}_i}$ is the predicted value. Due to the restrictions of the availability of ground-truth data, each pixel cannot be evaluated. Therefore, to provide administrative-based statistics for comparison, we averaged the pixel values for each regency/municipality by computing the mean.

3. Result

3.1. Correlation Model Development

Correlation analysis was used in this study to determine the strength and direction of the relationship between each geographic variable and the official human development statistics. Correlation analysis is carried out after the data have been pre-processed and transformed. Due to the limitations of official human development index statistics, which are only available at the regency/municipality level, we combined the pixel-sized geographic variable data by taking the median value for each regency/community. 38 observations were thus collected for each geographic variable. To better understand the relationship between geographic factors and the official human development index at the regency/municipality level, we used correlation analysis. We obtained correlation coefficients from this study, which show how closely related the variables are, and p-values, which demonstrate whether correlations are statistically significant. The correlation coefficient was calculated using the Pearson and Spearman rank correlations. The results of the correlation analysis are shown in Table 3, which demonstrates that (except for SO_2) the variables NTL, NDWI, NDBI, Daytime LST, Nighttime LST, CO, NO2, SO_2, and POI Density have a positive direction connection and are statistically significant. This illustrates how the increment values of these variables will rise in lockstep with the official HDI value. With HDI and significant, NDVI, SAVI, and POI Distance all move in the opposite direction.

Table 3. The correlation analysis between geospatial variables and the official human development index at the regency/municipality level.

Variable	Pearson Correlation Test		Spearman Rank Test		Direction & Closeness	Statistically Significant
	Correlation Coefficient	p-value	Correlation Coefficient	p-value		
NTL	$\mathbf{0.7244}$	$\mathbf{2.728 \, x \, 10^{-}}$	$\mathbf{0.6744}$	$\mathbf{6.358 \, x \, 10^{-}}$	Positive Strong	Yes
NDVI	$\mathbf{-0.5039}$	$\mathbf{1.258 \, x \, 10^{-}}$	$\mathbf{-0.4796}$	$\mathbf{2.581 \, x \, 10^{-}}$	Negative Moderate	Yes
SAVI	$\mathbf{-0.6059}$	$\mathbf{5.539 \, x \, 10^{-}}$	$\mathbf{-0.5735}$	$\mathbf{2.173 \, x \, 10^{-}}$	Negative Strong/Moderate	Yes
NDWI	$\mathbf{0.4403}$	$\mathbf{5.673 \, x \, 10^{-}}$	$\mathbf{0.4072}$	$\mathbf{1.171 \, x \, 10^{-}}$	Positive Moderate	Yes
NDBI	$\mathbf{0.6565}$	$\mathbf{7.615 \, x \, 10^{-}}$	$\mathbf{0.6164}$	$\mathbf{5.469 \, x \, 10^{-}}$	Positive Strong	Yes
LST Day	$\mathbf{0.6035}$	$\mathbf{6.054 \, x \, 10^{-}}$	$\mathbf{0.6505}$	$\mathbf{1.599 \, x \, 10^{-}}$	Positive Strong	Yes
LST Night	$\mathbf{0.4867}$	$\mathbf{1.944 \, x \, 10^{-}}$	$\mathbf{0.5098}$	$\mathbf{1.250 \, x \, 10^{-}}$	Positive Moderate	Yes
CO	$\mathbf{0.3975}$	$\mathbf{1.346 \, x \, 10^{-}}$	$\mathbf{0.4100}$	$\mathbf{1.110 \, x \, 10^{-}}$	Positive Weak/Moderate	Yes
NO_2	$\mathbf{0.5052}$	$\mathbf{1.216 \, x \, 10^{-}}$	$\mathbf{0.5186}$	$\mathbf{1.001 \, x \, 10^{-}}$	Positive Moderate	Yes
SO_2	$\mathbf{0.2363}$	$\mathbf{1.533 \, x \, 10^{-}}$	$\mathbf{0.1688}$	$\mathbf{3.097 \, x \, 10^{-}}$	Positive Weak/Very Weak	No
POI Density	$\mathbf{0.7819}$	$\mathbf{6.807 \, x \, 10^{-}}$	$\mathbf{0.7313}$	$\mathbf{1.840 \, x \, 10^{-}}$	Positive Strong	Yes
POI Distance	$\mathbf{-0.6453}$	$\mathbf{1.218 \, x \, 10^{-}}$	$\mathbf{-0.6619}$	$\mathbf{1.035 \, x \, 10^{-}}$	Negative Strong	Yes

3.2. PCA Model Development

PCA is used to reduce dimensions so that the main components have their respective variance proportions to all the variables in the data. Based on table 4, the components used are the first three components, namely component 1, component 2, and component 3 with a cumulative variance of 84.282 % of all variables in the data.

Table 4. The PCA and each component variance performance analysis.

Variable	Eigen Value	Proportion of Variance	Variance Cumulative Percentage
Component 1	5.5767	0.4647	46.471%
Component 2	2.7466	0.2289	69.360%
Component 3	1.8146	0.1512	84.482%
Component 4	0.7834	0.0653	91.010%
Component 5	0.4561	0.0380	94.811%
Component 6	0.2107	0.0176	96.567%
Component 7	0.1695	0.0142	97.980%
Component 8	0.1234	0.0103	99.009%
Component 9	0.0926	0.0077	99.781%
Component 10	0.0168	0.0014	99.920%
Component 11	0.0074	0.0006	99.982%
Component 12	0.0022	0.0002	100.000%

The following table 5 shows the loadings or components in the eigenvectors for each component to all variables. In this study, components 1, 2 and 3 will be weighted to form a relative spatial human development index (RSHDI) where a combination will be carried out, namely the sum of component 1 alone, the sum of component 1 and component 2, as well as the sum between components 1,2, and 3.

Table 5. The eigen vector of PCA component (loadings) for each variable.

Variable	NTL	NDVI	SAVI	NDWI	NDBI	LST Day	LST Night	CO	NO_2	SO_2	POI Density	POI Distance
Comp. 1	0.381	-0.242	-0.279	0.203	0.339	0.299	0.260	0.259	0.327	0.094	0.333	-0.335
Comp. 2	0.046	-0.481	-0.438	0.499	0.008	-0.352	-0.364	-0.107	-0.072	-0.002	-0.149	0.176
Comp. 3	0.055	0.064	-0.007	-0.119	0.354	-0.105	-0.072	0.516	0.202	-0.604	-0.296	0.280
Comp. 4	0.289	0.196	0.169	-0.234	0.192	-0.399	-0.511	0.113	0.304	0.470	0.106	-0.032
Comp. 5	0.219	0.041	0.033	-0.115	0.289	-0.082	-0.283	-0.151	-0.596	-0.389	0.311	-0.374
Comp. 6	0.198	0.002	0.043	0.076	0.023	0.075	0.136	0.532	-0.617	0.400	-0.028	0.324
Comp. 7	0.038	0.035	0.041	0.083	-0.375	-0.136	0.021	0.119	0.116	-0.243	0.758	0.409
Comp. 8	0.803	-0.135	0.155	0.004	-0.408	-0.014	0.121	-0.184	0.038	-0.168	-0.293	0.0236
Comp. 9	0.168	0.045	-0.159	-0.201	0.464	0.219	0.086	-0.516	-0.005	0.077	0.076	0.596
Comp. 10	0.023	-0.006	-0.703	-0.615	-0.309	0.048	-0.082	0.129	-0.049	0.014	-0.0216	-0.054
Comp. 11	0.012	0.066	0.060	0.152	-0.146	0.728	-0.635	0.070	0.049	-0.028	-0.046	0.067
Comp. 12	0.057	0.812	-0.386	0.418	0.009	-0.076	0.054	-0.032	-0.012	-0.014	-0.045	-0.047

3.3. Relative Spatial Human Development Index (RSHDI) Weighted Calculation

We determined the relative spatial human development index (RSHDI) using a weighted sum overlay and variables that had a strong correlation to the official East Java HDI data. In this study, four methods for calculating weights were used: EWS (W_1), PCCWS (W_2), SSCWS (W_3), and PCAWS (W_4). The derived weight calculations are shown in Table 6

Table 6. The Weight values for RSHDI approaches

Variable	W_1 (EWS)	W_2 (PCCWS)	W_3 (SCCWS)	W_4 (PCAWS)
NTL	+1	$\mathbf{0.7244}$	$\mathbf{0.6744}$	0.482
NDVI	-1	$\mathbf{-0.5039}$	$\mathbf{-0.4796}$	-0.687
SAVI	-1	$\mathbf{-0.6059}$	$\mathbf{-0.5735}$	-0.724
NDWI	+1	$\mathbf{0.4403}$	$\mathbf{0.4072}$	0.583
NDBI	+1	$\mathbf{0.6565}$	$\mathbf{0.6164}$	0.701
LST Day	+1	$\mathbf{0.6035}$	$\mathbf{0.6505}$	-0.158
LST Night	+1	$\mathbf{0.4867}$	$\mathbf{0.5098}$	-0.176
CO	+1	$\mathbf{0.3975}$	$\mathbf{0.4100}$	0.668
NO_2	+1	$\mathbf{0.5052}$	$\mathbf{0.5186}$	0.457
SO_2	Not Significant	Not Significant	Not Significant	-0.512
POI Density	+1	$\mathbf{0.7819}$	$\mathbf{0.7313}$	-0.112
POI Distance	-1	$\mathbf{-0.6453}$	$\mathbf{-0.6619}$	0.121

In this study, the W_1 through W_5 are calculated on a grid with a spatial resolution of 11 km x 11 km to produce a map of human development, as shown in Figures 12 through 16. We present the min-max scaled relative spatial poverty map, with values displayed that range from 0 to 1, making interpretation easier. From Figures 15 to 18, the human development maps produced by the RSHDI$_1$ through the RSHDI$_4$ have produced results that are remarkably similar. High values are primarily found in the urban areas of Kota Surabaya, Kota Malang, Kota Batu, Kota Kediri, Kota Madiun, and other parts of East Java.

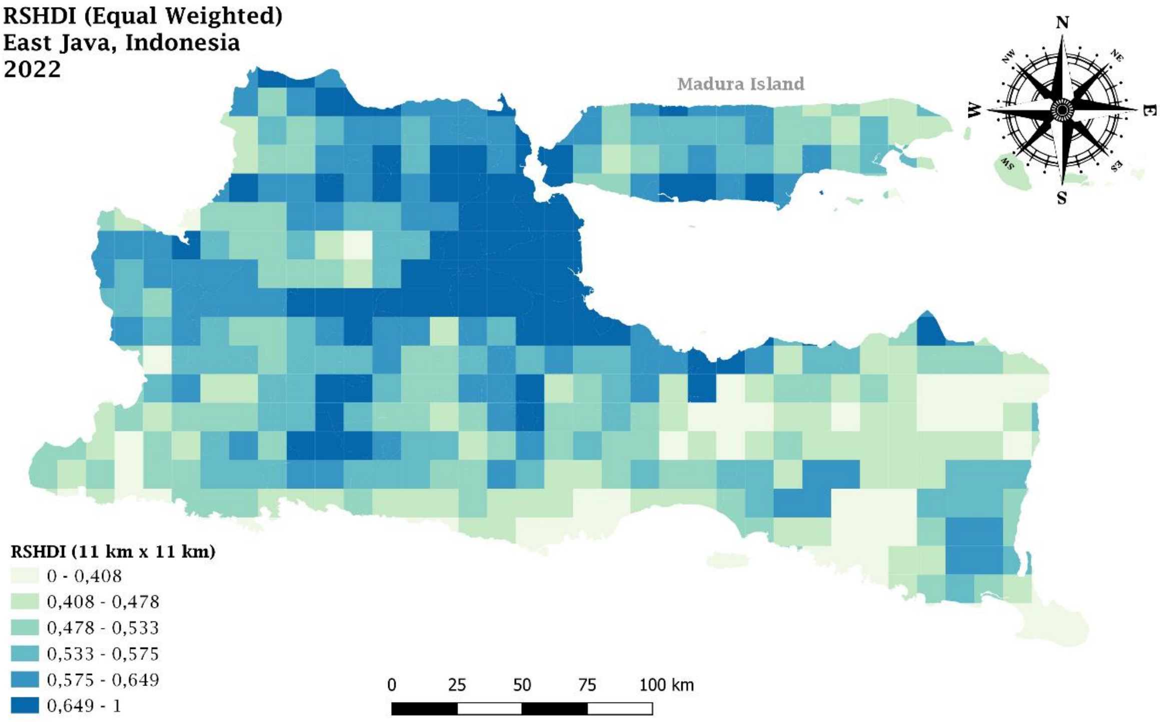

Figure 15. The obtained scaled of RSHDI EWS

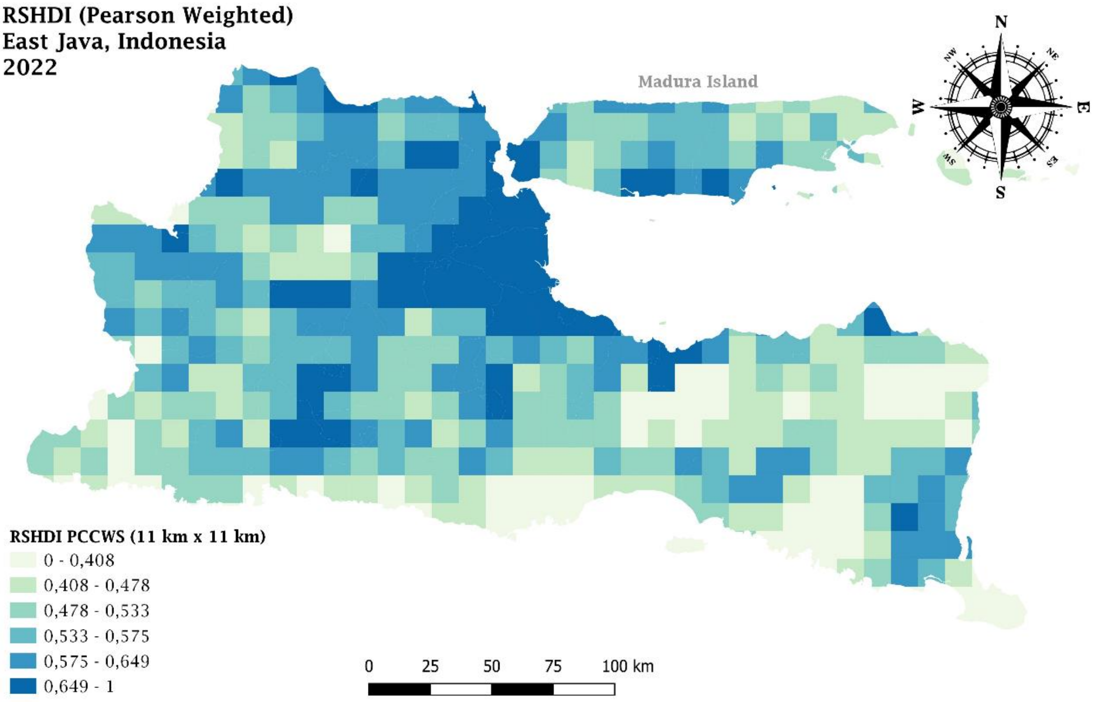

Figure 16. The obtained scaled of RSHDI PCCWS

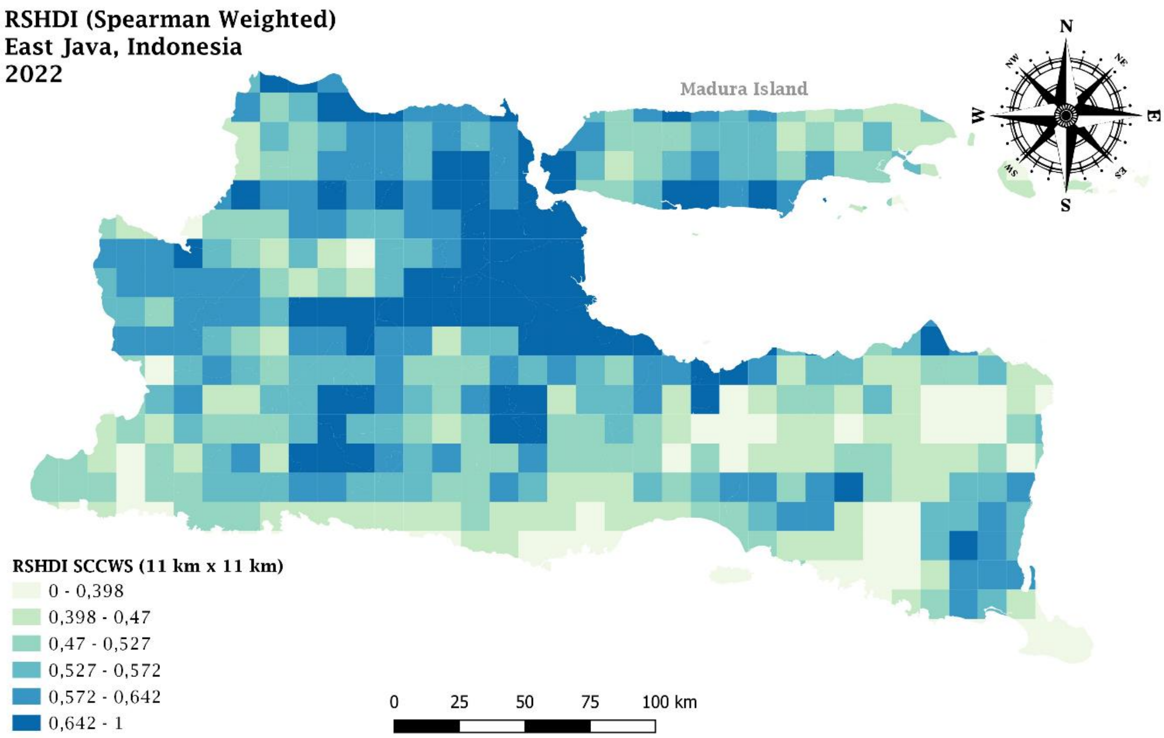

Figure 17. The obtained scaled of RSHDI SCCWS

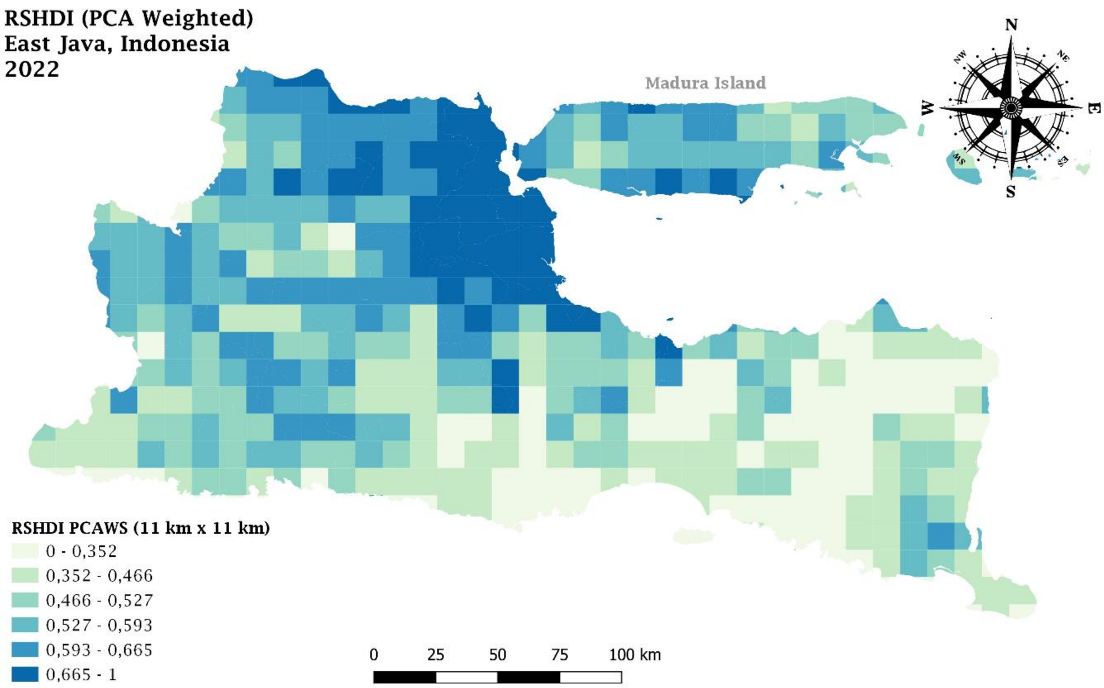

Figure 18. The obtained scaled of RSHDI PCAWS

4. Discussion

The obtained relative spatial human development index (RSHDI) was then validated through numerical and descriptive analysis. During the numerical analysis, we focused on figuring out how closely the relative spatial human development index (RSHDI) results matched the official HDI data numerically. Due to the limitations of the official poverty statistics, which are only available up to the regency/municipality level, it is not possible to examine every pixel. To combine the relative spatial human development index (RSHDI) pixel values, we took the mean of the 38 relative spatial human development index (RSHDI) values that were obtained for each regency/municipality. Correlation analysis (Pearson and Spearman Rank) and RMSE computation from linear regression are the two techniques used for numerical evaluation. Table 7 presents the correlation analysis that was obtained.

Table 7. Relationship between the official regency/municipality level HDI and the RSHDI

RSHDI	Pearson Correlation Test		Spearman Rank Test		Closeness	Direction	Statistically Significant
	Correlation Coefficient	p-value	Correlation Coefficient	p-value			
RSHDI EWS	0.7762	$\mathbf{1.031\ x\ 10^{-}}$	0.7523	3.921×10^{-7}	Strong	Positive	Yes
RSHDI PCCWS	0.7858	5.078×10^{-9}	0.7665	2.671×10^{-7}	Strong	Positive	Yes
RSHDI SCCWS	0.7837	5.970×10^{-9}	0.7779	1.976×10^{-7}	Strong	Positive	Yes
RSHDI PCAWS	0.6583	7.056×10^{-6}	0.5968	1.046×10^{-4}	Strong/Moderate	Positive	Yes

Table 7 shows that each relative spatial human development index (RSHDI) has a statistically significant correlation to the official HDI data (p-value 0.05). Except for the RSHDI PCAWS, which has a moderately positive Spearman rank correlation, the Pearson and Spearman rank correlation coefficient reveals that the RSHDI EWS, RSHDI PCCWS, and RSHDI SCCWS are strongly positively correlated to the official human development index data. The RSHDI PCCWS, which uses correlation-based weight to calculate correlation coefficients, had the highest correlation coefficient (Pearson correlation coefficient = 0.7858 (p-value = 5.078×10^{-9}) and Spearman rank correlation coefficient = 0.7837 (p-value = 5.970×10^{-9}). The direction is positive, indicating that the RSHDI variables' increment values tend to match the official HDI data's increment percentage. Additionally, using the relative spatial human development index (RSHDI) as the independent variable and the HDI data as the dependent variable, we constructed a simple linear regression model. The obtained model, along with its RMSE and R^2 value for each relative spatial human development index (RSHDI), are displayed in Table 8.

Table 8. Simple linear regression result

	RSHDI EWS	RSHDI PCCWS	RSHDI SCCWS	RSHDI PCAWS
Model	$71.0510 + 0.7662RSHDI$	$70.9925 + 1.3029RSHDI$	$70.9947 + 1.3359RSHDI$	$71.7070 + 1.3836RSHDI$
RMSE	3.159%	3.098%	3.112%	3.771%
R^2	0.6024	0.6175	0.6141	0.4333

The model created by RSHDI PCCWS, as can be seen, has the lowest RMSE value (3.098%) and the highest R^2 value (61.75%). The R^2 value obtained is 0.6175 which means 61.75% of the official HDI variance data can be explained through estimates on linear regression equations using RSHDI PCCWS as an independent variable. Then the value of 1.3029 as the slope on the equation indicates that each increase in the PCCWS RSHDI value will increase the HDI value by 1.3029. Then, the value 70.9925 is the intercept value, which shows that if the PCCWS RSHDI value is 0 then the estimated HDI value is 70.9925.

4.1. Comparison Between PCCWS RSHDI with Official HDI Data

As a result, we can say that RSHDI PCCWS is the best index to quantitatively predict the data from the official human development index (HDI). So, we decide that the RSHDI PCCWS is the index that best captures the official HDI data for granular mapping estimation.

Figure 19 compares the official HDI data with the aggregated relative spatial human development index (RSHDI) with Pearson Correlation Coefficient Weighted Sum (PCCWS) at the regency/municipality level with scatter plot. Due to the official HDI data's limitations, which limit its availability to the regency/municipality level, aggregation is done. Figure 20 illustrates the relative spatial human development index (RSHDI) mapping at the municipal/regent level for simple interpretation.

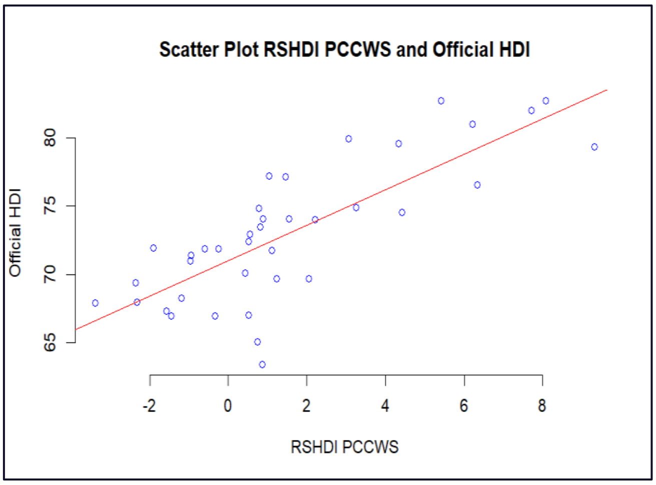

Figure 19. The Scatter Plot of RSHDI PCCWS and Official HDI data at Regency/Municipality

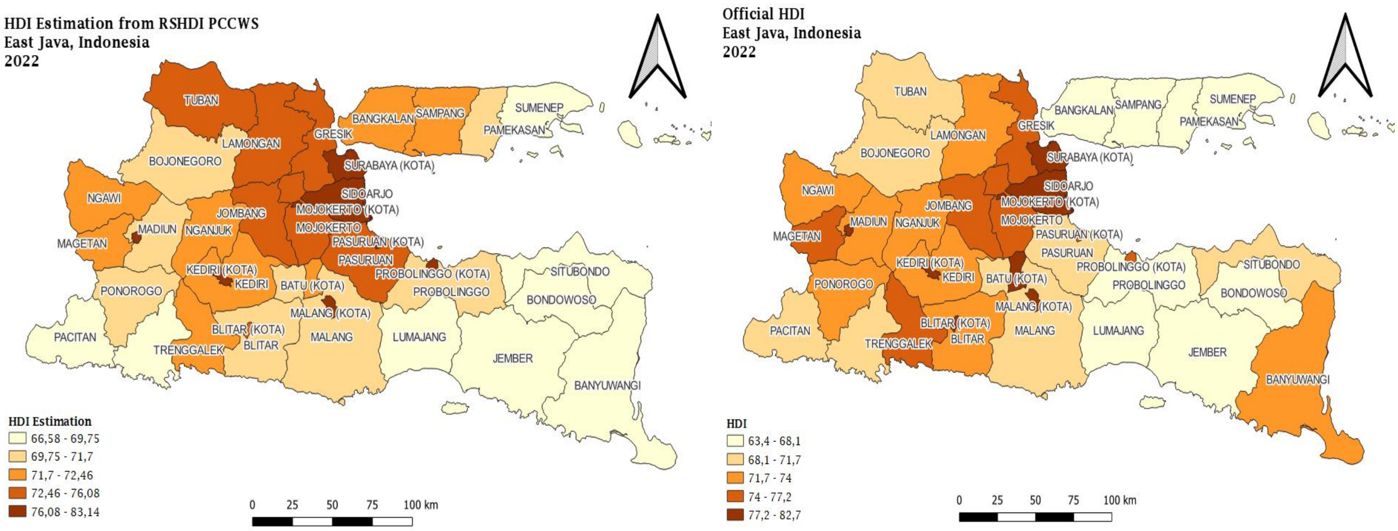

Figure 20. Comparison of HDI Estimation from RHSDI PCCWS (left) and Official HDI Data (right)

Based on figure 20, the results of both mapping show that high estimates of HDI are concentrated in urban areas such as Kota Surabaya, Kota Malang, Kota Batu, Kota Probolinggo and so on. This is in line with the Official HDI data that shows that in urban areas with densely populated, good accessibility, as well as good economics show higher HDI values than other areas. Then, the results of the HDI estimates also showed that in the eastern region of East Java such as Jember, Banyuwangi, Lumajang, Bondowoso, and so on has a lower result of the HDI compared to other areas which are also shown by official data as well. By visually mapping, the estimate is not much different from the official data so the PCCWS RSHDI is well used in the linear regression model with RMSE of 3.098%.

RSHDI PCCWS is the development of a granular spatial-based human development index with a correlation weight approach and its composer variables are big geospatial data variables such as NTL, NDVI, NDWI, LST, Air pollution, LST, and POI data. This developed index can be used to monitor and see the distribution of quality of life and human development in a more granular manner at a cheaper cost, faster time, and up to date. This data can be used as an accompanying or supporting the official data source to look at areas with poor or good spatially granular quality of life and human development. However, the use of such data cannot be used to replace conventional data sources such as SUSENAS as this data does not contain data in detail for individuals such as name by address to have a direct impact on individuals in the area. Better use is to combine both such data as a source of material for the evaluation of quality of life and human development as at the granular level such as a region has low PCCWS RSHDI values (assuming low HDI values as well) then can be done development planning in the area for example construction of schools, hospitals, the provision of social assistance based on detailed data from conventional data, and so on to improve the standard of living and quality of human life in the region. Thus, geospatial big data sources and conventional data sources are jointly beneficial in policy making and so on for quality of life and human development.

5. Conclusion

To address the limitations of the current household survey-based HDI data collection, this study offers a relative spatial human development index (RSHDI) that has better scope granularity, is less expensive, and requires less time to update. The relative spatial human development index (RSHDI) calculations use multisource remote sensing satellite imagery. Geospatial variables used in the calculation of the relative spatial human development index (RSHDI) in the case study area of East Java, Indonesia include the normalized difference built-up index (NDBI), the normalized difference vegetation index (NDVI), the soil adjusted vegetation index (SAVI), the normalized difference water index (NDWI), the day and nighttime land surface temperature (LST), carbon monoxide (CO), nitrogen dioxide (NO2), sulfur dioxide (SO2), and the nighttime light intensity (NTL). After that, a weighted sum model is applied to these variables, using equal weight (EWS), Pearson (PCCWS) and spearman (SCCWS) correlation-based weight, and PCA-based weight (PCAWS). It was discovered that Indonesia's human development index in 2022 has a good chance of succeeding using multisource remote sensing to represent East Java. This is demonstrated by the strong correlation between the official poverty statistics and the relative spatial human development index (RSHDI) at the regency/municipality level. According to a weight-sum model based on Pearson correlation, the best RSHDI PCCWS for representing human development index in East Java in 2022 has a Pearson correlation coefficient of 0.7858 (p-value = 5.078×10^{-9}) and is strongly correlated with official HDI data. It is also very promising to use this relative spatial human development index (RSHDI) as a predictor variable when estimating HDI data. We created a straightforward linear regression model, using relative spatial human development index (RSHDI) as the sole predictor variable, to estimate the official HDI for East Java, Indonesia, in 2022. Using RSHDI PCCWS, the ideal model had an RMSE of 3.098% and an R^2 of up to 61.75%. Then, RSHDI is presented using a non-technical, user-friendly HDI map with a spatial resolution of 11 km x 11 km. According to the findings of the descriptive analysis of this map, regions with low relative spatial human development index (RSHDI) scores tend to be sparsely populated, geographically depressed regions with inadequate accessibility, such as some of the regencies in Madura Island and the eastern region of East Java, while regions with high relative spatial human development index (RSHDI) scores tend to be densely populated regions with adequate accessibility, such as Surabaya and Kota Malang. As a result, the official human development index data can be supported by the RSHDI's ability to map spatially deprive areas.

References

[1] Badan Pusat Statistik Indonesia, *Indeks Pembangunan Manusia 2020*. 2020.
[2] United Nations Development Programme, *Human Development Report 1995*, vol. 54, no. 1. 1995.
[3] Badan Pusat Statistik, *Indeks Pembangunan Manusia 2021*. Jakarta, Indonesia, 2022.
[4] Badan Pusat Statistik, *Indeks Pembangunan Manusia 2022*. Jakarta, Indonesia, 2023.
[5] M. Jerven, “Benefits and costs of the data for development targets for the Post-2015 Development Agenda,” *Data Dev. Assess. Pap. Work. Pap.*, vol. Copenhagen, no. September 2014, p. 41, 2014, [Online]. Available: http://www.copenhagenconsensus.com/sites/default/files/data_assessment_-_jerven.pdf.
[6] Badan Pusat Statistik RI, *Teknik Pengumpulan Data dan Preprocessing Citra Satelit*. Jakarta, Indonesia: Jakarta: Badan Pusat Statistik, 2022.
[7] B. Liu and H. Xiong, “Point-of-interest recommendation in location based social networks with topic and location awareness,” *Proc. 2013 SIAM Int. Conf. Data Mining, SDM 2013*, pp. 396–404, 2013, doi: 10.1137/1.9781611972832.44.
[8] A. W. Wijayanto, D. W. Triscowati, and A. H. Marsuhandi, “Maize field area detection in East Java, Indonesia: An integrated multispectral remote sensing and machine learning approach,” *ICITEE 2020 - Proc. 12th Int. Conf. Inf. Technol. Electr. Eng.*, pp. 168–173, 2020, doi: 10.1109/ICITEE49829.2020.9271683.
[9] K. Shi, Z. Chang, Z. Chen, J. Wu, and B. Yu, “Identifying and evaluating poverty using multisource remote sensing and point of interest (POI) data: A case study of Chongqing, China,” *J. Clean. Prod.*, vol. 255, p. 120245, 2020, doi: 10.1016/j.jclepro.2020.120245.
[10] A. Irwansyah Fauzi *et al.*, “Evaluating mangrove forest deforestation causes in Southeast Asia by analyzing recent environment and socio-economic data products,” *Proc. - 39th Asian Conf. Remote Sens. Remote Sens. Enabling Prosper. ACRS 2018*, vol. 2, no. August 2019, pp. 880–889, 2018.
[11] N. Pokhriyal, O. Zambrano, J. Linares, and H. Hernández, “Estimating and Forecasting Income Poverty and Inequality in Haiti Using Satellite Imagery and Mobile Phone Data,” *Estim. Forecast. Income Poverty Inequal. Haiti Using Satell. Imag. Mob. Phone Data*, 2020, doi: 10.18235/0002466.
[12] K. Shi *et al.*, “Modeling and mapping total freight traffic in China using NPP-VIIRS nighttime light composite data,” *GIScience Remote Sens.*, vol. 52, no. 3, pp. 274–289, 2015, doi: 10.1080/15481603.2015.1022420.
[13] Z. Zhao *et al.*, “Analysis of the Spatial and Temporal Evolution of the GDP in Henan Province Based on Nighttime Light Data,” *Remote Sens.*, vol. 15, no. 3, 2023, doi: 10.3390/rs15030716.
[14] Y. Gu, Z. Shao, X. Huang, and B. Cai, “GDP Forecasting Model for China’s Provinces Using Nighttime Light Remote Sensing Data,” *Remote Sens.*, vol. 14, no. 15, 2022, doi: 10.3390/rs14153671.
[15] O. J. R. Pereira, L. G. Ferreira, F. Pinto, and L. Baumgarten, “Assessing pasture degradation in the Brazilian Cerrado based on the analysis of MODIS NDVI time-series,” *Remote Sens.*, vol. 10, no. 11, 2018, doi: 10.3390/rs10111761.
[16] S. Ahmed, “Assessment of urban heat islands and impact of climate change on socioeconomic over Suez Governorate using remote sensing and GIS techniques,” *Egypt. J. Remote Sens. Sp. Sci.*, vol. 21, no. 1, pp. 15–25, 2018, doi: 10.1016/j.ejrs.2017.08.001.
[17] Y. Zheng, Q. Zhou, Y. He, C. Wang, X. Wang, and H. Wang, “An optimized approach for extracting urban land based on log-transformed dmsp-ols nighttime light, ndvi, and ndwi,” *Remote Sens.*, vol. 13, no. 4, pp. 1–22, 2021, doi: 10.3390/rs13040766.
[18] Y. Wang *et al.*, “The impact of carbon monoxide on years of life lost and modified effect by individual- and city-level characteristics: Evidence from a nationwide time-series study in China,” *Ecotoxicol. Environ. Saf.*, vol. 210, p. 111884, 2021, doi: 10.1016/j.ecoenv.2020.111884.
[19] C. Han, Z. Gu, and H. Yang, “Ekc test of the relationship between nitrogen dioxide pollution and economic growth—a spatial econometric analysis based on chinese city data,” *Int. J. Environ.*

Res. Public Health, vol. 18, no. 18, 2021, doi: 10.3390/ijerph18189697.

[20] K. Bakhsh, T. Akmal, T. Ahmad, and Q. Abbas, “Investigating the nexus among sulfur dioxide emission, energy consumption, and economic growth: empirical evidence from Pakistan,” *Environ. Sci. Pollut. Res.*, vol. 29, no. 5, pp. 7214–7224, 2022, doi: 10.1007/s11356-021-15898-9.

[21] M. M. Yagoub, Y. T. Tesfaldet, M. G. Elmubarak, and N. Al Hosani, “Extraction of Urban Quality of Life Indicators Using Remote Sensing and Machine Learning: The Case of Al Ain City, United Arab Emirates (UAE),” *ISPRS Int. J. Geo-Information*, vol. 11, no. 9, 2022, doi: 10.3390/ijgi11090458.

[22] Y. I.-K. and R. A. Johnson, “A new family of power transformations to improve normality or symmetry,” *Biometrika*, vol. 156, no. I, pp. 87–90, 1989.

[23] J. Raymaekers and P. J. Rousseeuw, “Transforming variables to central normality,” *Mach. Learn.*, no. May 2020, 2021, doi: 10.1007/s10994-021-05960-5.

[24] S. R. Putri, A. W. Wijayanto, and A. D. Sakti, “Developing Relative Spatial Poverty Index Using Integrated Remote Sensing and Geospatial Big Data Approach: A Case Study of East Java, Indonesia,” *ISPRS Int. J. Geo-Information*, vol. 11, no. 5, 2022, doi: 10.3390/ijgi11050275.

[25] N. A. Utami, A. W. Wijayanto, S. Pramana, and E. T. Astuti, “Spatially granular poverty index (SGPI) for urban poverty mapping in Jakarta metropolitan area (JMA): a remote sensing satellite imageries and geospatial big data approach,” *Earth Sci. Informatics*, no. 0123456789, 2023, doi: 10.1007/s12145-023-01084-7.

[26] Afira N, Wijayanto AW 2022 Mono-temporal and multi-temporal approaches for burnt area detection using Sentinel-2 satellite imagery (a case study of Rokan Hilir Regency, Indonesia), Ecological Informatics, 69, 101677, Elsevier

[27] Wijayanto AW, Afira N, Nurkarim W 2022 Machine Learning Approaches using Satellite Data for Oil Palm Area Detection in Pekanbaru City, Riau, Proceedings of the 2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom).

[28] Putri SR, Wijayanto AW 2022 Learning Bayesian Network for Rainfall Prediction Modeling in Urban Area using Remote Sensing Satellite Data (Case Study: Jakarta, Indonesia), Proceedings of The International Conference on Data Science and Official Statistics, 2021, 1, 77-90

[29] Saadi T D T and Wijayanto A W 2021 Machine learning applied to Sentinel-2 and Landsat-8 multispectral and medium-resolution satellite imagery for the detection of rice production area in Nganjuk, East Java, Indonesia International Journal of Remote Sensing and Earth Sciences 18 19-32

[30] Nurmasari Y, Wijayanto AW 2021 Oil Palm Plantation Detection in Indonesia using Sentinel-2 and Landsat-8 Optical Satellite Imagery (Case Study: Rokan Hulu Regency, Riau Province), International Journal of Remote Sensing and Earth Sciences (IJReSES), 18, 1, 1-18, LAPAN