BAREKENG: Journal of Mathematics and Its Applications
December 2025 Volume 19 Issue 4 Page 2815–2830
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss4pp2815-2830

PREDICTION OF SOIL PARTICLES USING A SPATIALLY ADAPTIVE GEOGRAPHICALLY WEIGHTED K-NEAREST NEIGHBORS ORDINARY LOGISTIC REGRESSION APPROACH

Henny Pramoedyo [1], **Wigbertus Ngabu** [2*], **Atiek Iriany** [3], **Sativandi Riza** [4]

[1,3]*Statistic Department, Faculty of Mathematics and Science, University of Brawijaya*
Jln. Veteran, Ketawanggede, Lowokwaru, Malang, 65141, Indonesia

[2]*Mathematics Education Study Program, Faculty of Teacher Training and Education, Universitas Riau*
Jln HR. Soebrantas, Simpang Baru, Pekanbaru, Riau, 28293, Indonesia

[4]*Soil Science Department, Faculty of Agriculture, University of Brawijaya*
Jln. Veteran, Ketawanggede, Lowokwaru, Malang, 65141, Indonesia

*Corresponding author's e-mail: * **wigbertus.ngabu@lecturer.unri.ac.id***

Article History:

Received: 3rd March 2025
Revised: 2nd May 2025
Accepted: 16th June 2025
Available online: 1st September 2025

Keywords:

Geographically Weighted;
K-Nearest Neighbors;
Logistic Regression;
Spatial modeling;
Soil texture prediction.

ABSTRACT

Soil particle prediction is crucial in various fields, including agriculture, environmental management, and geotechnical applications. The spatial variation of soil texture significantly affects land fertility, erosion risk, and construction feasibility. However, conventional statistical methods and machine learning techniques often fail to capture the complex spatial heterogeneity in soil distribution. This study proposes the Geographically Weighted K Nearest Neighbors Ordinary Logistic Regression (GWKNNOLR) method to improve the accuracy of soil particle classification by integrating geographically weighted regression with an adaptive spatial weighting mechanism using the K Nearest Neighbors (KNN) algorithm. The objective of this research is to develop and evaluate a spatially adaptive classification model that more accurately predicts soil particle categories, namely sand, silt, and clay, by incorporating local spatial dependencies using GWKNNOLR in the Kalikonto watershed (DAS Kalikonto) in Batu. This study utilizes field measurement data combined with digital terrain modeling to analyze the relationship between local morphological variables and soil texture classification (sand, silt, and clay). The study area includes 50 observation points and 8 test variables. The model's performance is compared to the Ordinary Logistic Regression (OLR) method. The results indicate that GWKNNOLR achieves a classification accuracy of 88 percent, outperforming OLR, which only reaches 80 percent. Integrating KNN as a spatial weighting mechanism enhances adaptability to variations in sample distribution, leading to more accurate predictions. These findings emphasize the importance of considering spatial dependencies in soil texture modeling. The proposed method can support sustainable land resource management, erosion risk mitigation, and precision agriculture by providing more reliable soil classification. Future research may explore further optimization of spatial weighting mechanisms and the application of this method in different geographical regions.

How to cite this article:

H. Pramoedyo, W. Ngabu, A. Iriany and S. Riza., "PREDICTION OF SOIL PARTICLES USING A SPATIALLY ADAPTIVE GEOGRAPHICALLY WEIGHTED K-NEAREST NEIGHBORS ORDINARY LOGISTIC REGRESSION APPROACH", *BAREKENG: J. Math. & App.*, vol. 19, iss. 4, pp. 2815-2830, December, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*
Research Article · Open Access

1. INTRODUCTION

Predicting soil particles is crucial in various disciplines, including agriculture, environmental science, and geotechnical engineering [1]. Information regarding soil particle distribution plays a significant role in determining land fertility levels, identifying erosion risks, and assessing soil suitability for construction purposes. Various analytical methods have been employed to model soil characteristics, ranging from conventional statistical approaches to machine learning techniques [2]. However, these methods often struggle to effectively capture complex spatial variations due to the heterogeneity of soil characteristics across different locations [3].

Logistic regression is a widely used method for analyzing the relationship between dependent and independent variables in soil particle prediction [4]. However, this method assumes that the relationship between variables remains spatially constant, which is a major limitation when applied to data with significant geographical distribution patterns [5]. To address this limitation, the Geographically Weighted Ordinary Logistic Regression (GWOLR) method has been developed to account for geographical variations in logistic regression analysis [6].

Although GWOLR offers improvements in spatial analysis, its effectiveness highly depends on the weighting mechanism used to determine spatial influence. Conventional approaches in GWOLR often utilize kernel functions based on Euclidean distance, which do not always provide the best representation of spatial relationships among analyzed samples. Therefore, using the K-Nearest Neighbors (KNN) algorithm as a weighting mechanism in GWOLR presents a promising approach to improving model accuracy [7].

The KNN algorithm allows for selecting geographical neighbors based on more adaptive spatial proximity compared to conventional kernel-based approaches [8]. By incorporating KNN as the weighting method in GWOLR, the model becomes more responsive to changes in geographical distribution patterns, yielding more accurate estimations of soil characteristics across various locations. Consequently, the Geographically Weighted K-Nearest Neighbors Ordinary Logistic Regression (GW-KNN-OLR) approach is proposed to enhance soil particle prediction performance by dynamically considering spatial influences.

The state-of-the-art in this research is the GWOLR method, which has been widely applied in various fields, including spatial analysis, ecology, and geotechnical studies. Numerous studies have shown that GWOLR enhances prediction accuracy by capturing finer spatial variations. However, the method still faces limitations in its weighting mechanism. Typically, GWOLR utilizes kernel functions based on Euclidean distance to assign spatial weights, which may not be sufficiently flexible to model complex spatial relationships, especially in areas with uneven sample distributions or significant topographic variability [9].

Moreover, in previous studies, parameter selection in kernel functions has often been conducted manually or through cross-validation-based optimization methods, which require high computational costs. The K-Nearest Neighbors (KNN)-based approach is gaining attention as a more adaptive alternative for determining spatial weights, as it can adjust the number of neighbors based on sample density in a given area [10]. This method enables a more flexible approach to capturing dynamic spatial patterns than fixed-distance kernel-based functions [11].

The novelty of this research lies in the proposed Geographically Weighted K-Nearest Neighbors Ordinary Logistic Regression (GW-KNN-OLR) approach, which serves as a new method for predicting soil particle size [6]. The primary novelty of this research lies in implementing the KNN algorithm as the weighting mechanism in the GWOLR model. This approach differs from conventional methods as it does not rely on fixed-distance kernel functions. Instead, it adaptively selects geographical neighbors based on the most relevant samples in a given area [12]. This allows the model to respond more to complex spatial variations, especially in areas with uneven sample density.

Additionally, this study focuses on improving prediction accuracy and enhancing the interpretability of results through a more flexible spatial weighting approach. By combining geographically weighted logistic regression and KNN-based neighbor selection, this approach is expected to improve soil particle distribution modeling by incorporating more accurate spatial factors.

In the context of this research, GW-KNN-OLR is expected to overcome the limitations of conventional methods in capturing the complexity of spatial soil particle distribution. This approach enhances prediction accuracy and better interprets geographical influences on soil particle distribution [13]. Therefore, this study aims to evaluate the effectiveness of the GW-KNN-OLR method in soil particle prediction compared to conventional approaches and to assess its potential applications in various related fields.

This research's findings are expected to significantly contribute to developing more adaptive spatial analysis methods, particularly in soil characteristic modeling. Furthermore, this approach can support more effective land resource management, particularly in sustainable agriculture, soil conservation, and erosion risk mitigation. Thus, this study has the potential to provide broad benefits for spatial data-driven decision-making across various sectors.

2. RESEARCH METHODS

2.1 Research Data

This study utilizes two main types of data: primary data and secondary data. Primary data is obtained through direct field measurements of soil texture, while secondary data is derived from digital terrain modeling analysis [14]. Combining these two data types enables the examination of the relationship between local morphological parameters and soil texture characteristics [15].

This study uses field data for model development **(**training data**)** and model validation (testing data). Specifically, 50 observation points are used as training data and 8 test variables. The data is derived from soil texture analysis conducted in the Kalikonto Watershed (DAS Kalikonto)**,** with the location map presented in **Figure 1** below.

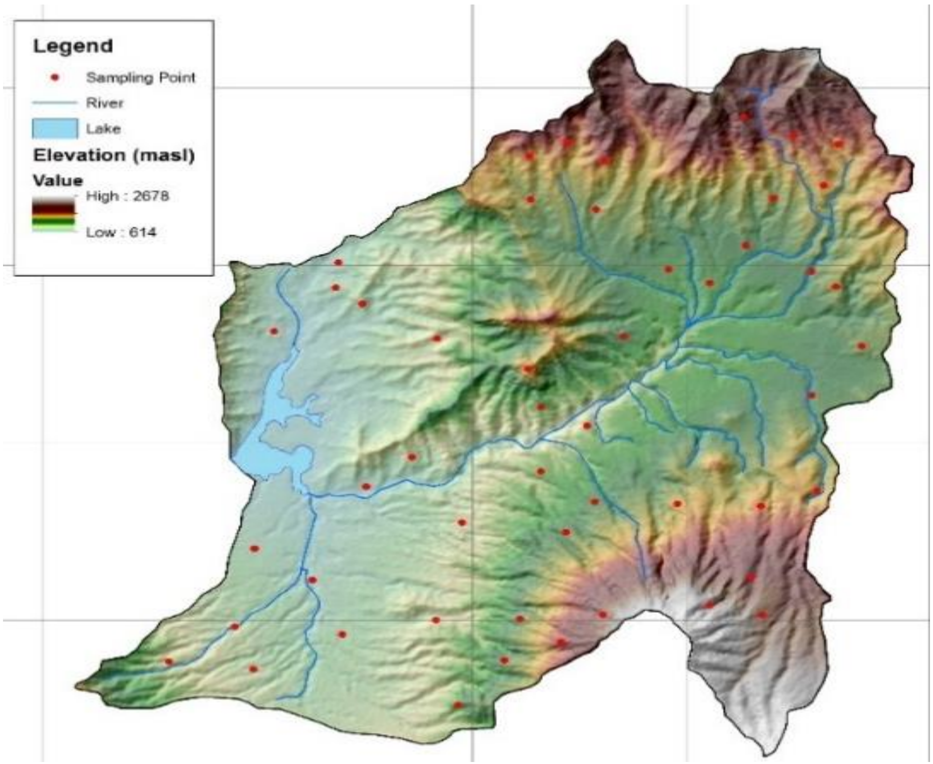

Figure 1. Research Location Map

The variables used in this study consist of eight Local Morphologic Variables (LMV), which reflect curvature variations in a given topography [16]. The LMV concept is based on previous studies, where these variables represent different morphological characteristics that may influence the distribution and physical properties of soil [17].

The variables in this study are categorized into two main groups**:** response variables and predictor variables. The response variable in this study is soil texture**,** classified into three categories: sand, silt, and clay, measured on an ordinal scale. Meanwhile, the eight other variables are predictor variables**,** measured on a ratio scale. The complete list of variables used in this study is presented in **Table 1** below:

Table 1. Research Variables

Variable		Measurement Scale	Description
Soil Particle Size	1 (Sand), 2(Silt), 3 (clay)	Ordinal	Response
Horizontal Curvature	Kh	ratio	Predictor
Vertical Curvature	Kv	ratio	Predictor
Slope	S	ratio	Predictor
Ring Curvature	Kr	ratio	Predictor
Elevation	Elev	ratio	Predictor
Northness Aspects	An	ratio	Predictor
Eastness Aspects	Ae	ratio	Predictor
Accumulation Curvature	Ka	ratio	Predictor

2.2 Independence Testing

The Chi-Square test of independence is used to examine whether there is a statistically significant association between two categorical variables. The null hypothesis ($\boldsymbol{H_0}$) states that the two variables are independent. The test statistic is calculated using the following formula [18]:

$$\boldsymbol{X^2} = \sum \frac{(\boldsymbol{O} - \boldsymbol{E})^2}{\boldsymbol{E}} \qquad (1)$$

where $\boldsymbol{O}$ represents the observed frequencies and $\boldsymbol{E}$ the expected frequencies, computed as:

$$\boldsymbol{E} = \frac{(\boldsymbol{row\ total})\mathbf{x}(\boldsymbol{column\ total})}{\boldsymbol{grand\ total}} \qquad (2)$$

The test is conducted at a 5% significance level with degrees of freedom $(\boldsymbol{r}-\mathbf{1})(\boldsymbol{c}-\mathbf{1})$. The result is evaluated by comparing the calculated value with the critical value or using the p-value for decision-making.

2.3 Ordinary Logistic Regression

Ordinal logistic regression is an extension of binary logistic regression, where ordinal logistic regression is a statistical method used to analyze data in which the response variable is measured on an ordinal scale consisting of three or more categories. Meanwhile, the predictor variables can be covariates (if measured on an interval or ratio scale) or factors (if measured on a nominal or ordinal scale) [19].

The logit model is commonly used for ordinal logistic regression**,** specifically the cumulative logit model [20]. In this logit model, the ordinal nature of the response variable (Y) is expressed through cumulative probabilities**.** The cumulative logit model is obtained by comparing the cumulative probability of the response variable being less than or equal to a given category G, given p predictor variables represented as vector **x,** $P(Y = g|\mathbf{x})$, with the probability of the response variable being greater than category g, $P(Y > g|\mathbf{x})$ [21].

Ordinal logistic regression is a statistical method used to analyze a dependent variable with an ordinal data scale that consists of three or more categories [22]. The independent (predictor) variables included in the model can be categorical or continuous and must consist of at least two variables [23].

The ordinal logistic regression model, which follows the cumulative logit form, expresses the ordinal nature of Y through cumulative probabilities [24]. Logistic regression is formulated by defining $P(Y = 1|\mathbf{x})$ as $\pi(\mathbf{x})$, which is denoted as follows [25].

$$\pi_g(\mathbf{x}_i) = \frac{\exp(\alpha_g + \boldsymbol{X}_i^T\ \boldsymbol{\beta})}{1 - exp(\alpha_g + \boldsymbol{X}_i^T\ \boldsymbol{\beta})} \qquad (3)$$

The logistic regression model is part of the Generalized Linear Models (GLM) framework. The model commonly used for ordinal logistic regression is the cumulative logit model (Cumulative Logit Models) [26]. Let the response variable Y have G ordinal categories, and let i represent the observation index while **x** denotes the vector of predictor variables for the i-th observation, $\mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix}^T$ with i =1,2,3,..,n. Thus, the ordinal logistic regression model can be expressed as follows [27]:

$$Logit[P(Y_i \leq g|\mathbf{x}_i)] = ln\left[\frac{P\ (Y_i\ \leq\ \mathrm{g}\ |\ \mathbf{x}_i)}{1-\ P\ (Y_i\ \leq\ \mathrm{g}\ |\ \mathbf{x}_i)}\right] = \propto_g + \ \mathbf{x}_i^T\beta \tag{4}$$

Let $\pi_g(\mathbf{x}_i) = [P(Y_i = g|\mathbf{x}_i)]$ represent the probability that the response variable for the i-th observation falls into category g given $\mathbf{x}_i$. Then, the ordinal logistic regression model can be formulated as follows:

$$\begin{aligned} P(Y_i \leq g\ |\ \mathbf{x}_i) &= (P\ (Y_i = 1|\mathbf{x}_i) + P(Y_i\ = 2\ |\ \mathbf{x}_i) + \ldots + P(Y_i = g|\mathbf{x}_i)) \\ &= \pi_1(\mathbf{x}_i) + \pi_2(\mathbf{x}_i) + \ldots + \pi_g(\mathbf{x}_i) \end{aligned}$$

2.4 Geographically Weighted Ordinary Logistic Regression

The Geographically Weighted Ordinal Logistic Regression (GWOLR) model is a combination of the Geographically Weighted Regression (GWR) model and the ordinal logistic regression model**.** GWOLR is used to model the relationship between an ordinal-scaled response variable and predictor variables, where each regression coefficient depends on the location where the data is observed [20].

GWOLR can also be regarded as a localized form of ordinal logistic regression, where the location effect is explicitly considered. The GWOLR technique incorporates geographical location into the model through a weighting function. Weights are assigned to each observation [28]. Let the response variable consist of G ordinal categories. Then, the GWOLR model for the i-th location can be expressed as follows:

$$logit[P(Y_i \leq |\boldsymbol{x}_i)] = ln\left[\frac{P(Y_i \leq g|x_i)}{1 - P(Y_i \leq g|x_i)}\right] = \alpha_g(u_i, v_i) + \boldsymbol{X}_i^T\boldsymbol{\beta}(u_i, v_i) \tag{5}$$

where g represents the category $(1,2, \ldots., G-1)$, With $\alpha_g(u_i, v_i)$ as the intercept parameter and the regression coefficient vector for the $i-th$ location, while (u_i, v_i) denotes the coordinate point (longitude, latitude) of the i-th location. The cumulative probability of the g-th response category can be expressed as [6]:

$$P(Y_i \leq g|\mathbf{x}_i) = \frac{\exp\left(\alpha_g(u_i, v_i) + \boldsymbol{X}_i^T\ \boldsymbol{\beta}(u_i, v_i)\right)}{1 + \exp\left(\alpha_g(u_i, v_i) + \boldsymbol{X}_i^T\ \boldsymbol{\beta}(u_i, v_i)\right)}, g = 1,2, \ldots, G-1 \tag{6}$$

If $\pi_g^*(\mathbf{x}_i) = P(Y_i \leq g|\mathbf{x}_i)$ represents the probability that the response variable at the i-th location falls into category g given $\mathbf{x}_i$, then [29] [30]:

$$\pi_g^*(\mathbf{x}_i) = \frac{\exp\left(\alpha_g(u_i, v_i) + \boldsymbol{X}_i^T\boldsymbol{\beta}(u_i, v_i)\right)}{1 + \exp\left(\alpha_g(u_i, v_i) + \boldsymbol{X}_i^T\boldsymbol{\beta}(u_i, v_i)\right)} - \frac{\exp\left(\alpha_{g-1}(u_i, v_i) + \boldsymbol{X}_i^T\boldsymbol{\beta}(u_i, v_i)\right)}{1 + \exp\left(\alpha_{g-1}(u_i, v_i) + \boldsymbol{X}_i^T\boldsymbol{\beta}(u_i, v_i)\right)} \tag{7}$$

with : $$\frac{\exp\left(\propto_g\ (u_i, v_i) + \boldsymbol{X}_i^T\boldsymbol{\beta}(u_i, v_i)\right)}{1 + \exp\left(\propto_g\ (u_i, v_i) + \boldsymbol{X}_i^T\boldsymbol{\beta}(u_i, v_i)\right)} = 0$$

$$\frac{\exp\left(\alpha_{g-1}(u_i, v_i) + \boldsymbol{X}_i^T\boldsymbol{\beta}(u_i, v_i)\right)}{1 + \exp\left(\propto_{g-1}\ (u_i, v_i) + \boldsymbol{X}_i^T\boldsymbol{\beta}(u_i, v_i)\right)} = 1$$

If the response variable in this study is assumed to be divided into three distinct categories (G = 3), then the Geographically Weighted Ordinary Logistic Regression (GWOLR) model can be constructed to account for spatial variations in the relationship between the dependent and independent variables. This model allows each location to have its own set of regression coefficients, reflecting local influences on the classification process. By incorporating geographically weighted parameters, the model enhances predictive accuracy by adapting to spatial heterogeneity in the data. The formulation of the GWOLR model for the i-th location is expressed as follows [30].

$$logit[P(Y_i \leq 1|\boldsymbol{x}_i)] = ln\left[\frac{P(Y_i \leq 1|x_i)}{1 - P(Y_i \leq 1|x_i)}\right] = \alpha_1(u_i, v_i) + \boldsymbol{X}_i^T\boldsymbol{\beta}(u_i, v_i) \tag{8}$$

$$logit[P(Y_i \leq 2|\boldsymbol{x}_i)] = ln\left[\frac{P(Y_i \leq 2|x_i)}{1 - P(Y_i \leq 2|x_i)}\right] = \alpha_2(u_i, v_i) + \boldsymbol{X}_i^T\boldsymbol{\beta}(u_i, v_i) \tag{9}$$

Thus, the probability for each response category at the i-th location is given by:

$$\pi_1^*(\mathbf{x}_i) = P(Y_i = 1|\boldsymbol{x}_i) = P(Y_i \le 1|\boldsymbol{x}_i) = \frac{\exp\left(\alpha_1(u_i, v_i) + \boldsymbol{X}_i^T\ \boldsymbol{\beta}(u_i, v_i)\right)}{1 + \exp\left(\alpha_1(u_i, v_i) + \boldsymbol{X}_i^T\ \boldsymbol{\beta}(u_i, v_i)\right)} \quad (10)$$

$$\pi_2^*(\mathbf{x}_i) = P(Y_i = 2|\boldsymbol{x}_i) = P(Y_i \le 2|\boldsymbol{x}_i) - P(Y_i \le 1|\boldsymbol{x}_i)$$
$$= \frac{\exp\left(\alpha_2(u_i, v_i) + \boldsymbol{X}_i^T\ \boldsymbol{\beta}(u_i, v_i)\right)}{1 + \exp\left(\alpha_2(u_i, v_i) + \boldsymbol{X}_i^T\ \boldsymbol{\beta}(u_i, v_i)\right)} - \frac{\exp\left(\alpha_1(u_i, v_i) + \boldsymbol{X}_i^T\ \boldsymbol{\beta}(u_i, v_i)\right)}{1 + \exp\left(\alpha_1(u_i, v_i) + \boldsymbol{X}_i^T\ \boldsymbol{\beta}(u_i, v_i)\right)} \quad (11)$$

$$\pi_3^*(\mathbf{x}_i) = P(Y_i = 3|\boldsymbol{x}_i) = P(Y_i \le 3|\boldsymbol{x}_i) - P(Y_i \le 2|\boldsymbol{x}_i) = 1 - \frac{\exp\left(\alpha_2(u_i, v_i) + \boldsymbol{X}_i^T\ \boldsymbol{\beta}(u_i, v_i)\right)}{1 + \exp\left(\alpha_2(u_i, v_i) + \boldsymbol{X}_i^T\ \boldsymbol{\beta}(u_i, v_i)\right)} \quad (12)$$

2.5 K-Nearest Neighbors Weighting

Weighting in the K-Nearest Neighbors (KNN) algorithm functions to assign different influences to each neighbor based on its distance from the test point [31]. In the nearest neighbor-based approach, weights are assigned according to spatial proximity, where closer data points to the test point receive higher weights. KNN works by identifying a set number of nearest neighbors to predict the value or label of a given test point [32]. However, not all neighbors have the same influence. Therefore, distance-based weighting is necessary to: (1) improve prediction accuracy, especially when data is not evenly distributed; (2) give greater influence to closer neighbors, ensuring results better reflect local patterns; and (3) reduce the effect of distant points, which may be less relevant for classification or regression.

Distance-based weighting in KNN is typically implemented using a mathematical function that assigns weights based on the proximity between the test point and its neighbors [33]. This method gives higher weights to closer neighbors, using the formula: $w_i = \frac{1}{d_{ij}}$, where d_{ij} represents the distance between the test point and its *i*-th neighbor. This weighting ensures that smaller distances correspond to greater influence in prediction.

Distance weighting in KNN allows the model to assign greater importance to closer neighbors, enhancing accuracy in classification or regression tasks [34]. The weighting method should be adapted to the dataset characteristics and analytical objectives to achieve optimal results.

2.6 Classification Accuracy

Classification accuracy is a crucial measure in evaluating the performance of a classification model. This accuracy is calculated by comparing the model's predicted results with the actual observed data. One commonly used classification evaluation method is the confusion matrix, which represents the distribution between the model's predictions and the actual values of the classified categories [35].

The **confusion matrix** provides information on the number of correctly classified observations and those the model misclassified. The confusion matrix is presented in Table 2 below:

Table 2. Confusion Matrix

Observation	Prediction		
	y_1	y_2	y_3
y_1	n_{11}	n_{12}	n_{13}
y_2	n_{21}	n_{22}	n_{23}
y_3	n_{31}	n_{32}	n_{33}

Thus, the classification accuracy percentage is calculated as follows:

$$\text{Classification Accuracy} = \frac{n_{11} + n_{22} + n_{33}}{n_{11} + n_{21} + n_{31} + n_{12} + n_{22} + n_{32} + n_{13} + n_{23} + n_{33}} \text{x } 100\%$$

3. RESULTS AND DISCUSSION

3.1 Independence Testing

The first step in logistic regression is to conduct an independence test. This test is performed to determine whether there is a significant relationship between the predictor variables and the response variable. The hypothesis testing for the independence test is conducted using the Chi-square test, with the results presented in **Table 3** below:

Table 3. Independence Test

Variable	Pearson Chi Square	P-value	Decision	Description
Kh	56.980	0.0041	Reject H_0	Dependent
Kv	45.876	0.0062	Reject H_0	Dependent
S	51.286	0.0051	Reject H_0	Dependent
Kr	42.872	0.0067	Reject H_0	Dependent
Elev	28.093	0.0362	Reject H_0	Dependent
An	41.876	0.0071	Reject H_0	Dependent
Ae	31.076	0.0276	Reject H_0	Dependent
Ka	26.987	0.0392	Reject H_0	Dependent

Based on the results presented in **Table 3**, we obtain the Pearson Chi-Square value and P-value for the eight tested variables. All variables show a P-value < 0.05 (α = 5%), indicating that the null hypothesis (H_0) is rejected. Consequently, it can be concluded that there is a significant relationship between the predictor variables and the response variable, suggesting that these variables are dependent on the response variable.

Furthermore, the high Pearson Chi-Square values indicate that the dependency between the predictor variables and the response variable has a significant impact. These results suggest that the tested variables explain the variation in the response variable. Therefore, these variables can be used in further analytical models, such as logistic regression or other predictive methods, to understand causal relationships better and improve model accuracy in data modeling.

3.2 Ordinal Logistic Regression

Ordinal Logistic Regression Analysis was conducted to examine the factors influencing Particle-Size Fraction, which is categorized into three groups: Sand, Silt, and Clay. This method was selected because the dependent variable has more than two ordinal-scale categories, allowing for a more accurate modeling of the relationship between the independent variables and Particle-Size Fraction categories.

In the Ordinal Logistic Regression model, the Estimate values (regression coefficients) indicate the influence of each predictor variable on the probability of a sample belonging to a specific Particle-Size Fraction category (Sand, Silt, or Clay). A positive coefficient in the model suggests that an increase in the predictor variable raises the probability of a sample transitioning to a higher ordinal category (e.g., Sand to Silt or Silt to Clay). Conversely, a negative coefficient indicates that an increase in the predictor variable reduces the likelihood of the sample shifting to a higher category.

In addition, the model includes intercepts for category $[Y = 1]$ (Sand) and $[Y = 2]$ (Clay), which represent the threshold points between categories in ordinal logistic regression. These intercepts indicate the cut-off points where the probability of category transition occurs. Significant predictor variables in the model contribute to shifting the probability distribution of Particle-Size Fraction, whereas non-significant variables have a weak or negligible effect on determining the final category.

The following **Table 4** presents the estimated regression parameters for the Ordinal Logistic Regression model, based on the regression coefficient values for each variable.

Table 4. Ordinal Logistic Regression Results

Variable	Estimate	Std. Error	Wald	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
[Y = 1]	-3.923	1.219	10.362	0.001	-6.312	-1.535
[Y = 2]	-2.942	1.133	6.741	0.009	-5.163	-0.721
Kh	0.133	0.511	0.068	0.794	-0.868	1.135
Kv	0.709	0.415	2.919	0.048	-0.104	1.522
S	34.636	31.312	1.224	0.269	-26.733	96.006

Variable	Estimate	Std. Error	Wald	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Kr	-1.222	0.955	1.638	0.201	-3.093	0.649
Elev	-1.524	2.365	0.415	0.520	-6.160	3.113
An	7.339	5.004	2.151	0.042	-2.469	17.146
Ae	68.010	60.973	1.244	0.265	-51.495	187.516
Ka	-0.935	1.815	0.266	0.606	-4.494	2.623

Based on **Table 4**, the logit function model is constructed. This model estimates the probability of a sample belonging to a specific category based on the values of the predictor variables included in the logit equation. The displayed equation represents two cumulative logit models, which estimate the probability of a sample falling into a category lower than or equal to Y = 1 (Sand) and Y = 2 (Silt). Each predictor variable has a regression coefficient, indicating the direction and magnitude of its influence on the Particle-Size Fraction categories. The logit function is formulated as follows:

$$\text{Logit}\left[\hat{P}(Y \leq 1|x)\right] = -3.923 + 0.133\ \text{Kh} + 0.709\ \text{Kv} + 34.636\ \text{S} - 1.222\ \text{Kr} - 1.524\ \text{Elev} + 7.339\ \text{An} + 68.010\ \text{Ae} - 0.935\ \text{Ka}$$

$$\text{Logit}\left[\hat{P}(Y \leq 2|x)\right] = -2.942 + 0.133\ \text{Kh} + 0.709\ \text{Kv} + 34.636\ \text{S} - 1.222\ \text{Kr} - 1.524\ \text{Elev} + 7.339\ \text{An} + 68.010\ \text{Ae} - 0.935\ \text{Ka}$$

The ordinal logistic regression results indicate that each predictor variable plays a distinct role in determining the probability of a soil sample falling into a specific Particle-Size Fraction category **(**Sand, Silt, or Clay**)**. A positive coefficient suggests that as the value of the predictor variable increases, the likelihood of the soil containing coarser particles or predominantly belonging to the Sand category increases. Conversely, a negative coefficient indicates that as the predictor variable decreases, the probability of the sample falling into finer particle categories, such as Clay**,** increases.

For example, the Kr (Ring Curvature) variable has a negative coefficient (-1.222)**,** indicating that as Kr (Ring Curvature) decreases at a given location, the probability of the soil containing a finer Particle-Size Fraction, such as Clay, increases. This means that a one-unit decrease in Kr (Ring Curvature) increases the likelihood of the soil being classified as Clay. This finding suggests a negative relationship between Kr (Ring Curvature) and soil particle coarseness, where lower values lead to a finer soil fraction**.**

On the other hand, the An (Northness Aspects) variable has a positive coefficient (7.339), meaning that as the An (Northness Aspects) value increases at a given location, the probability of the soil containing coarser particles, such as Sand, also increases. In other words, an increase in An (Northness Aspects) corresponds to a higher likelihood of the soil belonging to the Sand category rather than Silt or Clay. This suggests that An (Northness Aspects) is a contributing factor to the dominance of coarse soil particles, where higher values correspond to a larger or coarser soil fraction.

Additionally, the Kv (Vertical Curvature) variable (0.709) also has a significant positive influence on Particle-Size Fraction, although to a lesser extent than An (Northness Aspects). This implies that an increase in Kv (Vertical Curvature) tends to increase the probability of a soil sample being classified as Sand, though its effect is not as strong as that of An (Northness Aspects). Conversely, the Ka (Accumulation Curvature) variable (-0.935) shows a negative relationship with Particle-Size Fraction**,** indicating that as Ka (Accumulation Curvature) decreases, the probability of the soil being classified as Clay increases.

3.3 Classification Accuracy of Ordinal Logistic Regression

Ordinal logistic regression is a statistical method used to model the relationship between predictor variables and an ordinal response variable. In this study, the ordinal logistic regression model was applied to classify the particle-size fraction in the DAS Kalikonto area, Batu City, in 2024. The model evaluation was conducted by measuring classification accuracy to assess its performance in categorizing data accurately.

Classification accuracy was calculated by comparing the number of correct predictions to the total sample, as presented in **Table 5**.

Table 5. Accuracy of Logistic Regression Model Classification

Category	Number of Predictions per Category			Accuracy Percentage
	1	2	3	
1	32	3	1	88.88%
2	0	5	3	62.50%
3	2	1	3	50%

Based on Table 5, the analysis results indicate that the ordinal logistic regression model exhibits varying accuracy across different categories. Category 1 has the highest accuracy rate at 88.88%, followed by Category 2 with 62.50%, and Category 3 with 50%. The overall classification accuracy is calculated as follows:

$$Classification\ Accuracy = \left(\frac{Number\ of\ correct\ predictions}{N}\right) x\ 100\% = \left(\frac{32+5+3}{50}\right)\ x\ 100\ \% = 80\%$$

Based on the calculations, the overall classification accuracy reaches 80%, indicating that the model employed demonstrates a relatively high level of accuracy in predicting the particle-size fraction categories. This result suggests that the applied ordinal logistic regression model is reasonably effective in classifying the data, with an accuracy rate exceeding the commonly used 50% threshold for classification models. However, variations in accuracy across categories indicate that the model performs better in classifying certain particle-size fractions.

Therefore, although the ordinal logistic regression model demonstrates satisfactory performance, further analysis is still required. This includes incorporating spatial influences and exploring potential techniques to enhance accuracy, such as adjusting predictor variables or applying spatial weighting to improve the balance of data distribution across categories.

3.4 Geographically Weighted K-Nearest Neighbors Ordinary Logistic Regression

Geographically Weighted K-Nearest Neighbors Ordinary Logistic Regression (GW-KNN OLR) is an analytical method that combines logistic regression with spatial weighting using K-Nearest Neighbors (KNN). This model incorporates geographical factors in determining the relationship between dependent and independent variables, making it more effective in capturing local variations than conventional logistic regression. By utilizing KNN, the model adjusts estimations based on the influence of the nearest neighbors, enhancing its ability to account for spatial heterogeneity in the data.

3.4.1 Formation of K-Nearest Neighbors Weighting

K-Nearest Neighbors (KNN) weighting in the context of geographical modeling aims to capture spatial relationships between data points based on their proximity. In this model, each point is connected to its two nearest neighbors ($K = 2$), represented by red lines in the visualization. The value of $K = 2$ was determined through trial evaluations of several K values and was selected because it provided the best balance between model complexity and prediction accuracy, given the distribution of the analyzed data. This approach enables location-based analysis by incorporating spatial influence into logistic regression estimation. By considering the interconnection between points in geographic space, the model can capture local patterns that non-spatial approaches might overlook.

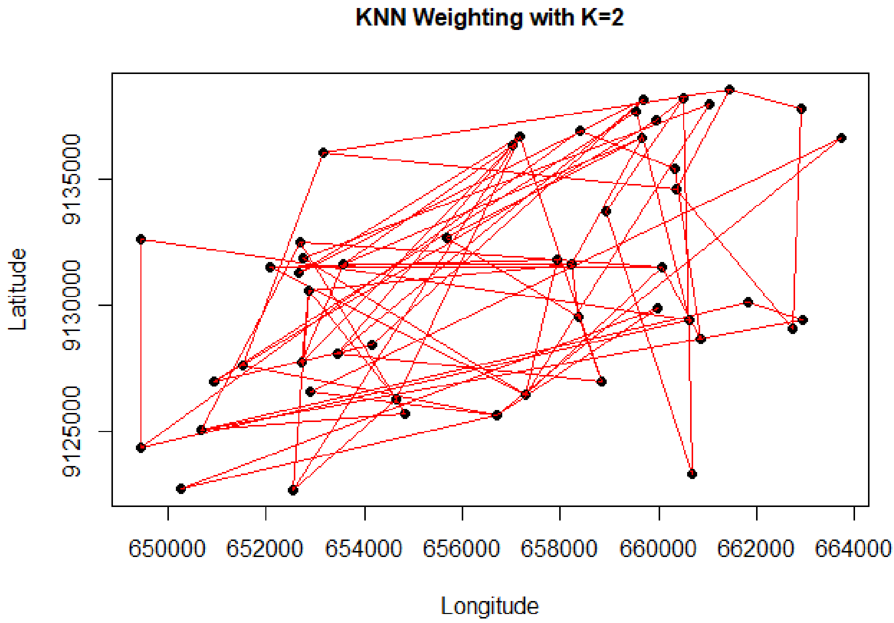

Figure 2. KNN weighting results
source:*R-Studio application*

The visualization above shows the spatial weighting relationships between points based on the K-Nearest Neighbors (K=2) weighting algorithm. Black dots represent data locations within a coordinate system (Longitude, Latitude), while red lines connect each point to its two nearest neighbors. The resulting network pattern depicts how spatial relationships are estimated within the Geographically Weighted KNN Ordinary Logistic Regression (GW-KNN OLR) model. The closer the connections between points, the greater the spatial influence in logistic regression weighting. By incorporating this approach, the model can more accurately capture local variations, making it particularly useful in fields such as epidemiological analysis, regional economics, and geographic risk mapping. Each of the 50 points is connected to its two nearest neighbors, as represented by the red lines. This reflects the weighting structure embedded within a 50×50 matrix, where each row and column correspond to a specific location point. In this matrix, non-zero elements indicate geographical proximity relationships based on the KNN algorithm, highlighting the spatial dependencies considered in the regression model.

3.4.2 GWKNNOLR Results

The geographically weighted k-nearest neighbors ordinary logistic regression (GW-KNN OLR) model integrates k-nearest neighbors (KNN) with ordinary logistic regression (OLR), where KNN serves as a weighting mechanism to account for the influence of geographic location on the probability estimation of particle-size fraction classification. This approach allows the model to capture spatial variations and provide more accurate estimations of soil categories based on specific locations.

The resulting logit equation expresses the likelihood that a given location belongs to each particle-size fraction category: sand, silt, and clay. Probability values are computed based on a combination of predictor variables, including Kh, Kv, S, Kr, Elev, An, Ae, and Ka, with each coefficient indicating the magnitude and direction of influence on the predicted category.

1. The equation $P_1(x_1)$ represents the probability that a given location is classified into the sand category.
2. The equation $P_2(x_1)$ indicates the likelihood that the location falls into the silt category, which is calculated by considering the cumulative probability between sand and silt.
3. The equation $P_3(x_1)$ represents the probability of the location being classified as clay, which complements the probabilities of the other categories, formulated as $(1 - P_1 - P_2)$..

In this model, coefficients with positive values indicate that an increase in the corresponding variable enhances the probability of classification into a higher ordinal category, from sand to silt or from silt to clay. Conversely, negative coefficients suggest that an increase in the variable decreases the likelihood of classification into a higher category.

The parameter estimation in the GW-KNN OLR model results in a unique local model for each location. Thus, each of the 50 locations has a distinct model tailored to its specific geographical characteristics. The generated model enables probability predictions for each particle-size fraction category. For instance, the model derived for the first location, positioned at coordinates (u_i, v_i), can be represented as follows:

$$P_1(x_1) = \frac{exp(5{,}6551 + 3{,}765\,Kh + 1{,}875\,Kv - 6{,}875\,S\; - \;6.8642\,Kr\; - \;2{,}543\,Elev + 0{,}9876\,An - 4{,}987\,Ae\; - 0{,}7623\,Ka\,)}{1 + exp(5{,}6551 + 3{,}765\,Kh + 1{,}875\,Kv - 6{,}875\,S\; - \;6.8642\,Kr\; - \;2{,}543\,Elev + 0{,}9876\,An - 4{,}987\,Ae\; - 0{,}7623\,Ka)}$$

$$P_2(x_1) = \frac{exp(-0{,}897 + 3{,}765\,Kh + \;1{,}875\,Kv - 6{,}875\,S\; - \;6.8642\,Kr\; - \;2{,}543\,Elev + 0{,}9876\,An - 4{,}987\,Ae\; - 0{,}7623\,Ka\,)}{1 + exp(-0{,}897 + 3{,}765\,Kh + 1{,}875\,Kv - 6{,}875\,S\; - 6.8642\,Kr\; - 2{,}543\,Elev + 0{,}9876\,An - 4{,}987\,Ae\; - 0{,}7623\,Ka\,)}$$

$$- \frac{exp(5{,}6551 + 3{,}765\,Kh + \;1{,}875\,Kv - 6{,}875\,S\; - \;6.8642\,Kr\; - \;2{,}543\,Elev + 0{,}9876\,An\; - 4{,}987\,Ae\; - 0{,}7623\,Ka\,)}{1 + exp(5{,}6551 + 3{,}765\,Kh + 1{,}875\,Kv - 6{,}875\,S\; - \;6.8642\,Kr\; - \;2{,}543\,Elev + 0{,}9876\,An - 4{,}987\,Ae\; - 0{,}7623\,Ka)}$$

$$P_3(x_1) = 1 - \frac{exp(-0{,}897 + 3{,}765\,Kh + 1{,}875\,Kv - 6{,}875\,S\; - \;6.8642\,Kr\; - \;2{,}543\,Elev + 0{,}9876\,An - 4{,}987\,Ae\; - 0{,}7623\,Ka\,)}{1 + exp(-0{,}897 + 3{,}765\,Kh + 1{,}875\,Kv - 6{,}875\,S\; - 6.8642\,Kr\; - \;2{,}543\,Elev + 0{,}9876\,An - 4{,}987\,Ae\; - 0{,}7623\,Ka\,)}$$

The probability of a given location being classified into a specific particle-size fraction category is influenced by the values of independent variables within the model. For instance, the parameter estimate β_1 has a positive value, indicating that as the value of the variable Kh (horizontal curvature) increases at a particular location, the likelihood of the soil in that location containing a higher proportion of sand also increases. This suggests that an increase in Kh contributes to a higher presence of coarser soil particles.

Conversely, the parameter estimate of β_5, which has a negative value, implies that as the value of the variable Kr (ring curvature) decreases, the probability of the soil containing a higher proportion of silt and clay increases. In other words, a reduction in Kr indicates a tendency for the soil to have finer particle fractions, shifting towards the silt and clay categories.

Additionally, the variable Elev (elevation), with a negative coefficient, also plays a role in determining soil fractions. Lower elevation is associated with a higher probability of increased silt and clay content, whereas higher elevation tends to increase the likelihood of sand dominance. The same applies to the Ka (accumulation curvature) variable, where a negative coefficient suggests that lower soil aeration increases the likelihood of finer soil particle content.

Thus, the GW-KNN OLR model demonstrates how environmental and topographical characteristics influence the distribution of particle-size fractions at a given location. This model can be used as a predictive tool to understand how spatial factors affect soil particle composition locally.

Based on the GW-KNN OLR model results, each location has a predicted probability of belonging to each particle-size fraction category (sand, silt, and clay). These probabilities are calculated based on the influence of predictor variables at each location while accounting for spatial factors through K-nearest neighbors (KNN) weighting. The following presents the predicted probability results:

Table 6. Results of Predicted Opportunities

Location	$\boldsymbol{\pi_1}(\mathbf{x}_i)$	$\boldsymbol{\pi_2}(\mathbf{x}_i)$	$\boldsymbol{\pi_3}(\mathbf{x}_i)$
1	0,7699	0,1294	0,1008
2	0,8536	0,086	0,0604
3	0,8127	0,1078	0,0795
4	0,9444	0,034	0,0216
5	0,3872	0,2404	0,3724
6	0,6062	0,1979	0,1958
7	0,6174	0,1941	0,1885
8	0,9985	0,001	0,0006
9	0,8868	0,0675	0,0457
10	0,2229	0,2105	0,5666
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
25	0,8588	0,0831	0,0581
30	0,5251	0,2217	0,2532
35	0,9414	0,0358	0,0228

Location	$\pi_1(\mathbf{x}_i)$	$\pi_2(\mathbf{x}_i)$	$\pi_3(\mathbf{x}_i)$
40	0,8323	0,0975	0,0702
45	0,8257	0,1009	0,0733
50	0,9929	0,0044	0,0027

The predicted probability results for particle-size fraction categories (sand, silt, and clay) obtained from the GW-KNN OLR model are subsequently visualized in the form of maps. These maps are designed to provide a spatial representation of the probability distribution for each category at the analyzed locations. These maps facilitate a more intuitive and interpretable understanding of soil distribution patterns using interpolation methods or symbolization based on the computed probabilities. The following map visualization presents the predicted probability results derived from the GW-KNN OLR model, as illustrated in **Figure 3**:

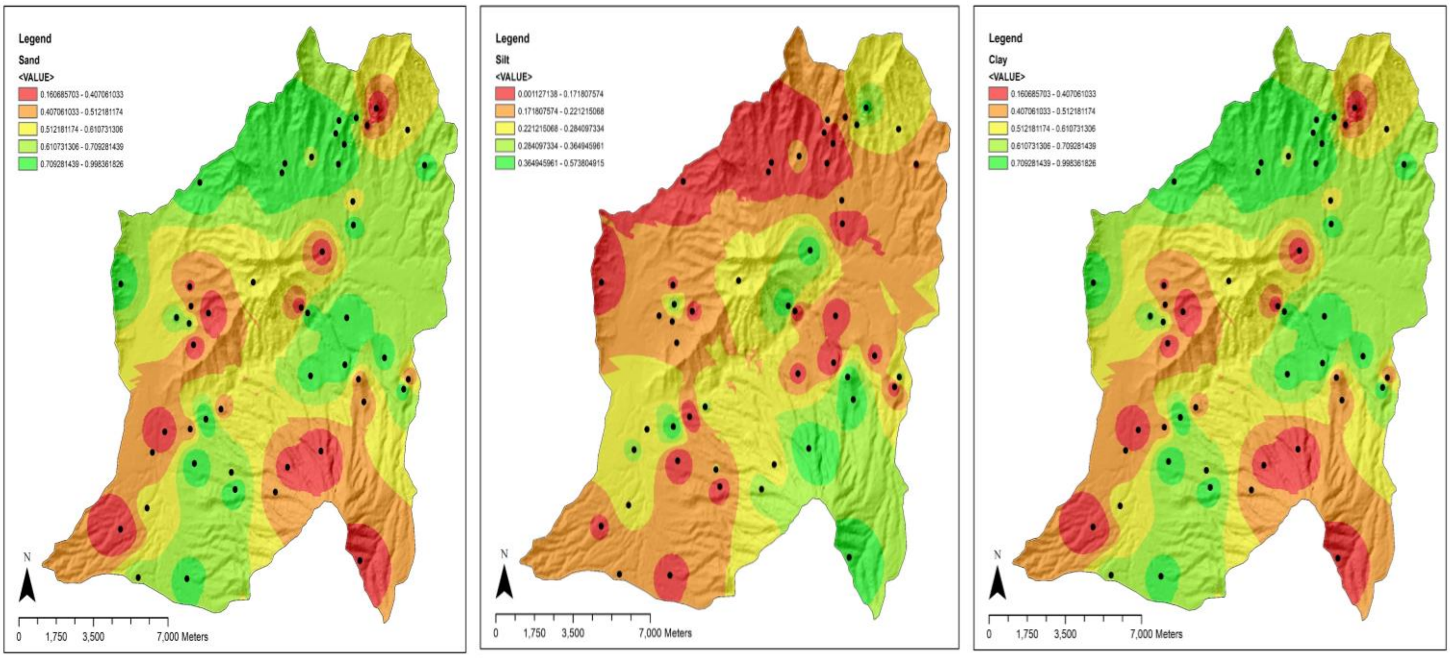

Figure 3. Probability Prediction Map for 3 categories
***source**: ArcGIS 10.8 application*

Figure 3 presents the predicted probability maps for the three particle-size fraction categories: sand, silt, and clay, based on the estimation results of the GW-KNN OLR model. Each map illustrates the spatial distribution of the probability for each category across the study area, with color variations representing the likelihood of dominance for each soil fraction.

1. The first map (left) represents the probability of the sand category. Green areas indicate regions with a higher probability of sand dominance, while red and yellow areas correspond to locations with lower probabilities for this category.
2. The second map (center) depicts the probability of the silt category. The distribution of this category is scattered across different locations, with some areas displaying moderate to high probabilities (indicated by green), while others exhibit lower probabilities (marked by red and yellow).
3. The third map (right) shows the probability of the clay category. Green areas represent regions with a higher probability of clay dominance, whereas yellow and red areas indicate lower probabilities for this category.

Based on these visualizations, it can be concluded that the distribution patterns of particle-size fractions vary significantly across locations. Certain areas tend to be dominated by sand, as indicated by the extensive green regions in the first map. In contrast, others exhibit a higher probability of clay dominance, as seen in the green areas of the third map. These patterns suggest that different spatial and environmental factors influence soil characteristics in the study area.

3.5 Accuracy of GWKNNOLR Classification

Classification accuracy is a key parameter in evaluating the performance of the geographically weighted k-nearest neighbors ordinary logistic regression (GW-KNN OLR) model in classifying particle-size fractions (sand, silt, and clay). This model aims to predict soil particle categories based on environmental variables while incorporating spatial factors through KNN weighting.

The predicted classification results were compared with actual observational data to assess the model's performance. Classification accuracy was calculated as the proportion of correctly predicted classifications relative to the total tested data. This evaluation is presented as a confusion matrix, which illustrates the distribution of the model's predictions against the actual soil category values.

A higher classification accuracy percentage indicates better model performance in accurately identifying particle-size fraction categories. Thus, this analysis not only aids in understanding the model's reliability but also provides insights into the extent to which the GW-KNN OLR model can be effectively applied. The probability estimates for each soil particle-size fraction category are presented in **Table 7**, along with the classification accuracy as follows:

Table 7. Accuracy of GWKNNOLR Classification

Observation	Number of Predictions per Category			Accuracy Percentage
	1	2	3	
1	34	1	1	94.44%
2	1	6	1	75 %
3	0	1	4	80%

The overall classification accuracy of the predictions is obtained from:

$$classification\ accuracy = \left(\frac{Number\ of\ correct\ predictions}{N}\right) x\ 100\% = \left(\frac{34+6+4}{50}\right) x\ 100\ \% \ = \ 88\%$$

These results indicate that the GW-KNN OLR model achieves an accuracy of 88%, demonstrating its ability to classify most samples correctly. This level of accuracy suggests that the applied method effectively captures the distribution patterns of particle-size fractions.

3.6 Best Model Selection

The selection of the best classification model is a crucial step in evaluating the predictive performance of various statistical approaches. This study compares the accuracy of Ordinary Logistic Regression (OLR) and Geographically Weighted K-Nearest Neighbors Ordinary Logistic Regression (GW-KNN OLR) in classifying Particle-Size Fraction. The accuracy of each model is assessed based on its ability to correctly classify soil categories while considering spatial influences.

The classification accuracy results for both models are presented in **Table 8** below.

Table 8. Model Comparison

Model	Accuracy
OLR	80 %
GWKNNOLR	88 %

Based on **Table 8**, the results indicate that the Ordinary Logistic Regression model achieves an accuracy of 80%, whereas the GW-KNN OLR model demonstrates a higher accuracy of 88%. This increase in accuracy suggests that incorporating spatial weighting and local geographical influences in GW-KNN OLR enhances the model's ability to capture variations in soil characteristics.

These findings highlight the advantage of spatial-based models in classification tasks, particularly in geographic and environmental studies, where location-based variations significantly impact prediction accuracy. The GW-KNN OLR model has proven more reliable, demonstrating better classification performance than standard logistic regression. This study supports the application of spatial analysis techniques in predictive modeling, leading to more accurate and contextually relevant results.

3.7 Research Implications and Applications

The findings of this study have significant implications across various fields, particularly in soil management and agriculture. Land-use strategies can be optimized to enhance agricultural productivity while mitigating soil degradation risks by understanding the relationship between environmental variables and particle-size fractions. Areas dominated by clay and silt are more suitable for water conservation, soil moisture management, and crops with high water requirements. In contrast, regions with a higher proportion of sand are better suited for deep-rooted plants and well-drained agricultural lands. This understanding facilitates more precise and sustainable agricultural planning, supporting agroecological practices and long-term food security.

Beyond agriculture, this study is crucial in land-use planning and environmental policy. More accurate soil texture distribution data enable the identification of erosion-prone areas, sedimentation zones, and soil degradation risks, which are essential for designing effective soil and water conservation systems. Regions with high clay content require careful consideration in drainage system design and infrastructure development to mitigate risks associated with waterlogging and soil instability. By leveraging more precise spatial data, land-use policies can be adapted to enhance ecosystem resilience and promote sustainable natural resource management.

From a methodological perspective, this study demonstrates that the GW-KNN OLR approach outperforms conventional logistic regression in capturing spatial heterogeneity in soil texture distribution. The model improves classification accuracy to 88%, compared to standard regression models that do not adequately account for spatial factors. By applying this method, geotechnics, hydrology, and ecological mapping research can achieve a more accurate understanding of soil distribution and its impact on surrounding ecosystems. Furthermore, this model can be further developed to support more complex spatial analyses in soil mapping and environmental landscape assessments.

Overall, this research contributes significantly to the spatial understanding of soil distribution, which serves as a foundation for decision-making in soil resource management, environmental risk mitigation, and climate adaptation strategies. With a more accurate predictive model, land-use and soil conservation policies can be better directed, supporting more sustainable and data-driven land management practices. Additionally, the findings of this study can be utilized in disaster-resilient infrastructure planning, further strengthening environmental resilience in the long term.

4. CONCLUSION

Based on the analysis results, the GW-KNN-OLR method demonstrates an 88% classification accuracy, higher than the 80% accuracy achieved by the Ordinary Logistic Regression (OLR) method. This indicates that the spatial weighting approach using KNN can capture variations in soil distribution more accurately than the conventional kernel-based approach.

The key advantage of this method lies in its ability to adjust spatial weighting based on the geographic distribution of samples, resulting in more precise estimations of soil texture categories (sand, silt, and clay). By incorporating location-based influences in the analysis, this method provides a better understanding of soil distribution patterns and the factors affecting them.

The findings of this study have significant implications across various fields, including land resource management, erosion risk mitigation, and precision agriculture applications. The GW-KNN-OLR model can aid in making more accurate spatial data-driven decisions, particularly in soil conservation planning and ecosystem management.

As a further step, this research can be expanded by exploring more efficient optimization mechanisms for spatial weighting and applying this method to regions with more diverse soil characteristics. Thus, the proposed method has the potential to become a more effective approach for spatial analysis based on geographic data.

AUTHOR CONTRIBUTIONS

Henny Pramoedyo: Writing - Original Draft, Writing - Review and Editing. Wigbertus Ngabu: Software, Formal Analysis, Visualization. Atiek Iriany: Methodology. Sativandi Riza: Project Administration, Data Curation. All authors discussed the results and contributed to the final manuscript.

FUNDING STATEMENT

This study was funded by the Professor Research Grant Program, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya, contract number 02172.15/UN10.F0901/B/KS/2024.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Faculty of Mathematics and Natural Sciences, Brawijaya University, for supporting this research through the Professor Research Grant Program. This support has been instrumental in the successful completion of this study.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest to report in this study, and all authors have approved the final version of the manuscript.

REFERENCES

[1] D. Kim, B. H. Nam, and H. Youn, "EFFECT OF CLAY CONTENT ON THE SHEAR STRENGTH OF CLAY–SAND MIXTURE," *International Journal of Geo-Engineering*, vol. 9, pp. 1–12, 2018.doi: https://doi.org/10.1186/s40703-018-0087-x

[2] E. J. Parent, S.-É. Parent, and L. E. Parent, "DETERMINING SOIL PARTICLE-SIZE DISTRIBUTION FROM INFRARED SPECTRA USING MACHINE LEARNING PREDICTIONS: METHODOLOGY AND MODELING," *PLoS One*, vol. 16, no. 7, p. e0233242, 2021.doi: https://doi.org/10.1371/journal.pone.0233242

[3] J. M. Moreno-Maroto and J. Alonso-Azcarate, "EVALUATION OF THE USDA SOIL TEXTURE TRIANGLE THROUGH ATTERBERG LIMITS AND AN ALTERNATIVE CLASSIFICATION SYSTEM," *Appl Clay Sci*, vol. 229, p. 106689, 2022.doi: https://doi.org/10.1016/j.clay.2022.106689

[4] M. Heidari, S. Zalaghaie, M. R. Nikudel, and A. Seif, "PREDICTION OF THE LOESS SOILS COLLAPSE SENSITIVITY CLASSES BY APPLICATION OF ORDINAL LOGISTIC REGRESSION, NORTHEASTERN IRAN," *Iranian Journal of Engineering Geology*, vol. 16, no. 2, 2023.

[5] S. Zalaghaie, M. Heidari, M. R. Nikudel, S. H. Jalali, and A. Seif, "APPLICATION OF ORDINAL LOGISTIC AND GENE EXPRESSION PROGRAMMING METHODS TO PREDICT THE COLLAPSE SENSITIVITY CLASSES OF LOESS SOILS, A CASE STUDY: GOLESTAN PROVINCE, NORTHEASTERN IRAN," *Environ Monit Assess*, vol. 195, no. 2, p. 305, 2023.doi: https://doi.org/10.1007/s10661-023-10915-5

[6] H. Pramoedyo, W. Ngabu, S. Riza, and A. Iriany, "SPATIAL ANALYSIS USING GEOGRAPHICALLY WEIGHTED ORDINARY LOGISTIC REGRESSION (GWOLR) METHOD FOR PREDICTION OF PARTICLE-SIZE FRACTION IN SOIL SURFACE," in *IOP Conference Series: Earth and Environmental Science*, IOP Publishing, 2024, p. 012005.doi: https://doi.org/10.1088/1755-1315/1299/1/012005

[7] M. Kubara and K. Kopczewska, "AKAIKE INFORMATION CRITERION IN CHOOSING THE OPTIMAL K-NEAREST NEIGHBOURS OF THE SPATIAL WEIGHT MATRIX," *Spat Econ Anal*, vol. 19, no. 1, pp. 73–91, 2024.doi: https://doi.org/10.1080/17421772.2023.2176539

[8] X. Peng, R. Chen, K. Yu, F. Ye, and W. Xue, "AN IMPROVED WEIGHTED K-NEAREST NEIGHBOR ALGORITHM FOR INDOOR LOCALIZATION," *Electronics (Basel)*, vol. 9, no. 12, p. 2117, 2020.doi: https://doi.org/10.3390/electronics9122117

[9] H. Pramoedyo, A. Iriany, W. Ngabu, and S. Riza, "SOIL PARTICLE PREDICTION USING SPATIAL ORDINARY LOGISTIC REGRESSION," *Advances in Decision Sciences*, vol. 28, no. 1, pp. 66–92, 2024.doi: https://doi.org/10.47654/v28y2024i2p66-92

[10] X. Yang, Y. Yang, S. Xu, J. Han, Z. Chai, and G. Yang, "A NEW ALGORITHM FOR LARGE-SCALE GEOGRAPHICALLY WEIGHTED REGRESSION WITH K-NEAREST NEIGHBORS," *ISPRS Int J Geoinf*, vol. 12, no. 7, p. 295, 2023.doi: https://doi.org/10.3390/ijgi12070295

[11] P. Jafary, D. Shojaei, A. Rajabifard, and T. Ngo, "AUTOMATING PROPERTY VALUATION AT THE MACRO SCALE OF SUBURBAN LEVEL: A MULTI-STEP METHOD BASED ON SPATIAL IMPUTATION TECHNIQUES, MACHINE LEARNING AND DEEP LEARNING," *Habitat Int*, vol. 148, p. 103075, 2024.doi: https://doi.org/10.1016/j.habitatint.2024.103075

[12] W. Ngabu, R. Fitriani, H. Pramoedyo, and A. B. Astuti, "CLUSTER FAST DOUBLE BOOTSTRAP APPROACH WITH RANDOM EFFECT SPATIAL MODELING," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 17, no. 2, pp. 945–954, 2023.doi: https://doi.org/10.30598/barekengvol17iss2pp0945-0954

[13] F. M. Seaton, P. B. L. George, I. Lebron, D. L. Jones, S. Creer, and D. A. Robinson, "SOIL TEXTURAL HETEROGENEITY IMPACTS BACTERIAL BUT NOT FUNGAL DIVERSITY," *Soil Biol Biochem*, vol. 144, p. 107766, 2020.doi: https://doi.org/10.1016/j.soilbio.2020.107766

[14] L. Xiong, G. Tang, X. Yang, and F. Li, "GEOMORPHOLOGY-ORIENTED DIGITAL TERRAIN ANALYSIS: PROGRESS AND PERSPECTIVES," *Journal of geographical sciences*, vol. 31, pp. 456–476, 2021.doi: https://doi.org/10.1007/s11442-021-1853-9

[15] A. Iriany, W. Ngabu, D. Ariyanto, and H. Pramoedyo, "KRIGING PREDICTION AND SIMULATION MODEL: ANALYSIS OF SURFACE SOIL PARTICLE SIZE DISTRIBUTION.," *Mathematical Modelling of Engineering Problems*, vol. 12, no. 4, 2025, doi: https://doi.org/10.18280/mmep.120408 .

[16] Z. Guo *et al.*, "SOIL TEXTURE IS AN IMPORTANT FACTOR DETERMINING HOW MICROPLASTICS AFFECT SOIL HYDRAULIC CHARACTERISTICS," *Environ Int*, vol. 165, p. 107293, 2022.doi: https://doi.org/10.1016/j.envint.2022.107293

[17] H. Pramoedyo, W. Ngabu, S. Riza, and L. Amaliana, "SINGLE BOOTSTRAP APPROACH WITH GEOGRAPHICALLY WEIGHT REGRESSION MODELING USING PARTICLE-SIZE FRACTION," *J Theor Appl Inf Technol*, vol. 31, no. 10, 2023, [Online]. Available: www.jatit.org

[18] R. J. Tallarida, R. B. Murray, R. J. Tallarida, and R. B. Murray, "CHI-SQUARE TEST," *Manual of pharmacologic calculations: with computer programs*, pp. 140–142, 1987.doi: https://doi.org/10.1007/978-1-4612-4974-0_43

[19] M. Alrumaidhi and H. A. Rakha, "FACTORS AFFECTING CRASH SEVERITY AMONG ELDERLY DRIVERS: A MULTILEVEL ORDINAL LOGISTIC REGRESSION APPROACH," *Sustainability*, vol. 14, no. 18, p. 11543, 2022.doi: https://doi.org/10.3390/su141811543

[20] H. Khaulasari and F. Suhaeri, "MAPPING OF GENDER INEQUALITY IN INDONESIA BASED ON INFLUENCING FACTORS USING GEOGRAPHICALLY WEIGHTED ORDINAL LOGISTIC REGRESSION," *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, vol. 18, no. 1, pp. 233–244, 2024.doi: https://doi.org/10.30598/barekengvol18iss1pp0233-0244

[21] J. Liang, G. Bi, and C. Zhan, "MULTINOMIAL AND ORDINAL LOGISTIC REGRESSION ANALYSES WITH MULTI-CATEGORICAL VARIABLES USING R," *Ann Transl Med*, vol. 8, no. 16, p. 982, 2020.doi: https://doi.org/10.21037/atm-2020-57

[22] B. French and M. S. Shotwell, "REGRESSION MODELS FOR ORDINAL OUTCOMES," *JAMA*, vol. 328, no. 8, pp. 772–773, 2022.doi: https://doi.org/10.1001/jama.2022.12104

[23] B. S. Tusa, S. A. Kebede, and A. B. Weldesenbet, "SPATIAL DISTRIBUTION AND DETERMINANT FACTORS OF ANEMIA AMONG ADULTS AGED 15–59 IN ETHIOPIA; USING MIXED-EFFECTS ORDINAL LOGISTIC REGRESSION MODEL," *BMC Nutr*, vol. 7, pp. 1–12, 2021.doi: https://doi.org/10.1186/s40795-021-00424-4

[24] I. O. Asare and A. C. Mensah, "CRASH SEVERITY MODELLING USING ORDINAL LOGISTIC REGRESSION APPROACH," *Int J Inj Contr Saf Promot*, vol. 27, no. 4, pp. 412–419, 2020.doi: https://doi.org/10.1080/17457300.2020.1790615

[25] W. Ngabu, H. Pramoedyo, A. Iriany, and R. Fitriani, "HYBRID MODELING OF DEEP NEURAL NETWORKS AND ORDINARY LOGISTIC REGRESSION FOR PREDICTING GROUNDWATER QUALITY," *Mathematical Modelling of Engineering Problems*, vol. 12, no. 3, pp. 774–786, 2025, doi: https://doi.org/10.18280/mmep.120304.

[26] G. Tutz, "ORDINAL REGRESSION: A REVIEW AND A TAXONOMY OF MODELS," *Wiley Interdiscip Rev Comput Stat*, vol. 14, no. 2, p. e1545, 2022.doi: https://doi.org/10.1002/wics.1545

[27] S. Jayawardena, J. Epps, and E. Ambikairajah, "ORDINAL LOGISTIC REGRESSION WITH PARTIAL PROPORTIONAL ODDS FOR DEPRESSION PREDICTION," *IEEE Trans Affect Comput*, vol. 14, no. 1, pp. 563–577, 2020.doi: https://doi.org/10.1109/TAFFC.2020.3031300

[28] A. Iriany, W. Ngabu, and D. Ariyanto, "RAINFALL MODELING USING THE GEOGRAPHICALLY WEIGHTED POISSON REGRESSION METHOD," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 18, no. 1, pp. 627–636, 2024.doi: https://doi.org/10.30598/barekengvol18iss1pp0627-0636

[29] A. Comber *et al.*, "THE GWR ROUTE MAP: A GUIDE TO THE INFORMED APPLICATION OF GEOGRAPHICALLY WEIGHTED REGRESSION," *arXiv preprint arXiv:2004.06070*, 2020.

[30] V. N. Mishra, V. Kumar, R. Prasad, and M. Punia, "GEOGRAPHICALLY WEIGHTED METHOD INTEGRATED WITH LOGISTIC REGRESSION FOR ANALYZING SPATIALLY VARYING ACCURACY MEASURES OF REMOTE SENSING IMAGE CLASSIFICATION," *Journal of the Indian Society of Remote Sensing*, vol. 49, pp. 1189–1199, 2021.doi: https://doi.org/10.1007/s12524-020-01286-2

[31] S. Zhang, S. Ray, R. Lu, and Y. Zheng, "SPRIG: A LEARNED SPATIAL INDEX FOR RANGE AND KNN QUERIES," in *Proceedings of the 17th International Symposium on Spatial and Temporal Databases*, 2021, pp. 96–105.doi: https://doi.org/10.1145/3469830.3470892

[32] S. Liu, R. De Lacerda, and J. Fiorina, "PERFORMANCE ANALYSIS OF ADAPTIVE K FOR WEIGHTED K-NEAREST NEIGHBOR BASED INDOOR POSITIONING," in *2022 IEEE 95th Vehicular Technology Conference:(VTC2022-Spring)*, IEEE, 2022, pp. 1–5.doi: https://doi.org/10.1109/VTC2022-Spring54318.2022.9860699

[33] X. Chen and Y. Chen, "QUANTIFYING THE RELATIONSHIPS BETWEEN NETWORK DISTANCE AND STRAIGHT-LINE DISTANCE: APPLICATIONS IN SPATIAL BIAS CORRECTION," *Ann GIS*, vol. 27, no. 4, pp. 351–369, 2021.doi: https://doi.org/10.1080/19475683.2021.1966503

[34] X. Yang, Y. Yang, S. Xu, J. Han, Z. Chai, and G. Yang, "A NEW ALGORITHM FOR LARGE-SCALE GEOGRAPHICALLY WEIGHTED REGRESSION WITH K-NEAREST NEIGHBORS," *ISPRS Int J Geoinf*, vol. 12, no. 7, p. 295, 2023.doi: https://doi.org/10.3390/ijgi12070295

[35] M. K. Campbell, A. Donner, and K. M. Webster, "ARE ORDINAL MODELS USEFUL FOR CLASSIFICATION?," *Stat Med*, vol. 10, no. 3, pp. 383–394, 1991.doi: https://doi.org/10.1002/sim.4780100310