

Article

Predicting Soil Fertility in Semi-Arid Agroecosystems Using Interpretable Machine Learning Models: A Sustainable Approach for Data-Sparse Regions

Nurullah Acir 

Department of Soil Science and Plant Nutrition, Faculty of Agriculture, Kırşehir Ahi Evran University, Kırşehir 40100, Türkiye; nurullah.acir@ahievran.edu.tr

Abstract

The accurate assessment of soil fertility is critical for guiding nutrient management and promoting sustainable agriculture in semi-arid agroecosystems. In this study, a machine learning-based Soil Fertility Index (SFI) model was developed using regularized regression techniques to evaluate fertility across a dryland maize-growing region in southeastern Türkiye. A total of 64 composite soil samples were collected from the Batman Plain, characterized by alkaline and salinity-prone conditions. Five soil chemical indicators, electrical conductivity (EC), pH, organic matter (OM), zinc (Zn), and iron (Fe), were selected for SFI estimation using a standardized rating approach. The dataset was randomly split into training (80%) and test (20%) subsets to calibrate and validate the models. Ridge, Lasso, and Elastic Net regression models were employed to predict SFI and assess variable importance. Among these, the Lasso model achieved the highest predictive accuracy on test data ($R^2 = 0.746$, RMSE = 0.060), retaining only EC and Zn as significant predictors. Ridge and Elastic Net captured OM and pH, though their contributions were minimal ($|\beta| < 0.01$). Spatial predictions showed moderate alignment with observed SFI values (range: 0.48–0.76), but all models underestimated high-fertility zones (>0.69), likely due to coefficient shrinkage. Despite its simplicity, the Lasso model offered superior interpretability and spatial resolution. The results reveal the potential of interpretable machine learning for supporting sustainable, site-specific fertility assessment and informed nutrient management in data-scarce and environmentally vulnerable regions.



Academic Editor: Primo Proietti

Received: 23 July 2025

Revised: 11 August 2025

Accepted: 18 August 2025

Published: 21 August 2025

Citation: Acir, N. Predicting Soil Fertility in Semi-Arid Agroecosystems Using Interpretable Machine Learning Models: A Sustainable Approach for Data-Sparse Regions. *Sustainability* **2025**, *17*, 7547. <https://doi.org/10.3390/su17167547>

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: soil quality indexing; regularized predictive models; Ridge regression; Lasso regression; Elastic Net; sustainable agriculture; soil fertility mapping

1. Introduction

Soil fertility is a fundamental component of sustainable agriculture, directly influencing crop yield, soil resilience, and long-term ecosystem stability. Global projections indicate that up to 89% of agricultural soils may experience fertility degradation by 2050 [1], underscoring the need for more adaptive and data-driven approaches to fertility assessment. Traditional methods, primarily based on laboratory testing and manual sampling, are resource-intensive and often insufficient to capture the spatial heterogeneity of modern agricultural landscapes [2,3].

In response, researchers have developed composite indicators like the Soil Fertility Index (SFI), which integrate multiple soil attributes into a unified metric. Among them, Tunçay et al. [4] proposed a comprehensive SFI model that integrates chemical (e.g., pH,

EC), physical (e.g., texture, bulk density), and biological (e.g., organic matter) soil indicators to classify fertility levels into five distinct classes: very low, low, moderate, high, and very high. This classification facilitates rapid, spatially explicit evaluation of fertility conditions, supporting localized management interventions, and land use planning. However, the methods used to construct SFI vary widely. Conventional statistical approaches, such as factor analysis, offer interpretability but may struggle with multicollinearity and nonlinear relationships [5]. Recently, machine learning (ML) techniques have gained prominence for their predictive power and ability to handle complex variable interactions [6]. The comparative evaluation by Jia et al. [6] across multiple agroecological zones revealed that machine learning models outperform conventional regression methods. Moreover, these models offer robust tools for variable importance analysis. This provides a semi-interpretable framework that helps bridge the gap between prediction and practical soil management. Complementary research has further shown the success of ML algorithms such as Support Vector Machines in modeling key fertility indicators like cation exchange capacity (CEC), pH, and soil organic carbon [7–9].

Despite their predictive accuracy, many ML models function as “black boxes,” offering limited interpretability, which constrains their application in agronomic decision-making. This limitation has prompted increased interest in interpretable machine learning (IML) techniques within soil science. Among them, regularized regression models, such as Ridge, Lasso, and ElasticNet, are particularly well-suited for soil fertility studies because they address multicollinearity while yielding interpretable coefficients. This allows researchers to assess the relative importance of highly correlated soil variables such as EC, pH, OM, Zn, and Fe [10–12].

In response to this need, the present study proposes a transparent and interpretable modeling framework for predicting the SFI using regularized regression methods. While most existing studies prioritize model accuracy, our focus is on understanding the underlying drivers of fertility by analyzing coefficient magnitudes and directions. The modeling pipeline outputs standardized coefficients and interceptions, allowing for meaningful integration of expert agronomic knowledge with data-driven insights. To operate this framework, the study pursues one general aim supported by three specific objectives. The general aim is to assess the predictive and interpretive capacity of regularized regression models for estimating a composite SFI based on key soil chemical properties in irrigated maize systems. Specifically, the study first constructs an SFI by standardizing and scoring five soil indicators, pH, EC, OM, Zn, and Fe, using agronomic thresholds and normalized weights. Second, it applies ridge, lasso, and elastic net regression models to predict the SFI from 64 composite soil samples collected from maize fields in the Batman Plain, using cross-validation and hyperparameter tuning for model optimization. Finally, model performance is evaluated using statistical indicators (R^2 , RMSE), and standardized coefficients are examined to interpret the relative contribution of each predictor to the overall fertility assessment.

2. Materials and Methods

2.1. Study Area and Soil Sampling

The study area is located between the city center of Batman and the Bismil district of Diyarbakır province, within the coordinates of $34^{\circ}48'$ – $37^{\circ}55'$ N latitude and $40^{\circ}55'$ – $41^{\circ}05'$ E longitude. This region is heavily used for maize cultivation. Situated north of the confluence of the Dicle River and the Batman Stream, the study area consists of young alluvial deposits formed by the accumulation of sediments from rivers and streams (Figure 1).

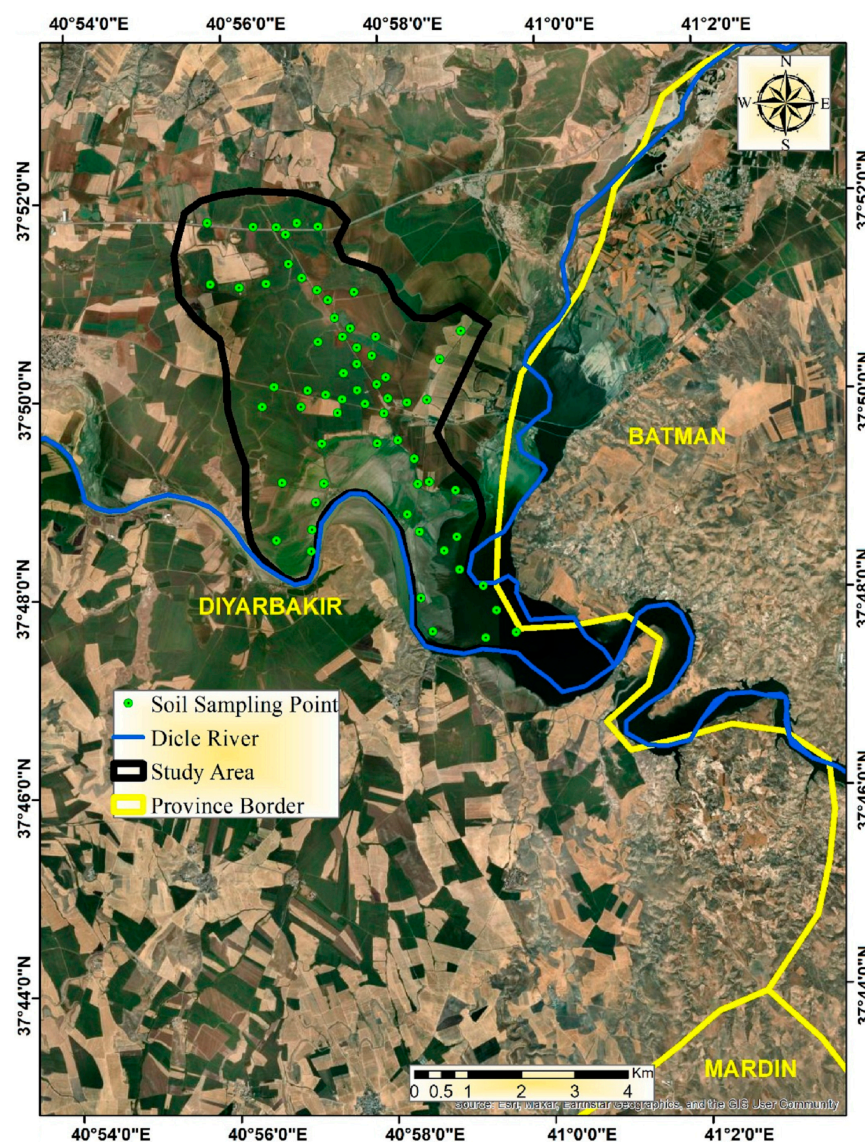


Figure 1. Location of study area and sampling points. Created in ArcGIS 10.8 (Esri, Redlands, CA, USA) using Google Satellite Hybrid Map.

The area lies within a semi-arid climate zone, with a long-term annual average precipitation of 494 mm and an average annual temperature of 15.9 °C [13]. The region is composed primarily of the Şelmo Formation [14], which consists of grayish-green, pink, and occasionally brownish-purple sandstones, shales, and sandy siltstones. These are interlayered with gypsum, carbonate-cemented, poorly sorted, coarse-textured, porous materials, mollusk-bearing layers, and include fine-bedded calcareous gravels and thick cross-bedded hard sandstones in the upper parts. The dominant soil orders in the area are Entisols and Inceptisols, and a significant portion of the land lies within the area submerged by the Ilisu Dam reservoir.

A total of 64 composite soil samples (0–20 cm) were collected from a 1957 ha area characterized by geologically uniform Şelmo Formation and topographical homogeneity. Samples were obtained through a randomized design focused on intensively cultivated maize fields. The resulting sampling density (~1 sample per 30.6 ha) was deemed sufficient based on the area's limited spatial variability and consistent land use. Similar findings in digital soil mapping studies (e.g., [15]) suggest that prediction performance improves with density but reaches an acceptable threshold near 1 profile per 2 km², especially when using models such as quantile random forest and kriging in more heterogeneous contexts.

The geographic coordinates of each sampling location were recorded using a GPS device. The collected soil samples were air-dried and passed through a 2 mm sieve prior to laboratory analysis. Soil texture was determined using the Bouyoucos hydrometer method following the procedure described by Gee and Bauder [16]. Soil reaction (pH) and electrical conductivity (EC) were measured in a 1:2 soil-to-water suspension in accordance with the method outlined by Rhoades et al. [17]. Organic matter content was assessed using the modified Walkley–Black wet oxidation method as proposed by Nelson and Sommers [18], which provides reliable estimates of organic carbon through dichromate oxidation. Extractable zinc (Zn) and iron (Fe) concentrations were determined using the DTPA (diethylenetriaminepentaacetic acid) extraction method developed by Lindsay and Norvell [19], with subsequent quantification performed via atomic absorption spectrophotometry. These soil parameters were selected due to their foundational roles in soil fertility assessment and nutrient availability.

2.2. Calculation of Soil Fertility Index

This study introduces an interpretable machine learning (IML) approach to soil fertility modeling by applying regularized linear regression techniques to develop a scalable SFI. The SFI serves as an integrative metric that encapsulates yield potential by evaluating multiple soil chemical properties in a composite manner. The primary goal of the SFI is to synthesize complex soil fertility data into a single, standardized index, enabling objective comparisons and spatial analysis across sites [20].

In this study, five key parameters, soil pH, EC, OM, and DTPA-extractable Zn and Fe, were used in the SFI computation. The indicators were chosen based on their established roles in governing nutrient availability and soil productivity in semi-arid agroecosystems. Soil pH is a key determinant of nutrient solubility, while EC reflects the extent of salinity stress that can inhibit plant uptake. Organic matter contributes to cation exchange capacity and nutrient retention. DTPA-extractable Zn and Fe were included due to their frequent deficiencies in alkaline soils, where their availability is pH-sensitive. This selection also aligns with prior SFI frameworks (e.g., [4,21]) and supports both agronomic interpretability and statistical robustness in regularized regression modeling.

Each parameter was assigned a score ranging from 0 to 1, reflecting its suitability for agricultural productivity. These scores were derived using a rating system (Table 1) based on known threshold and optimal values for each parameter. The SFI was calculated using two mathematical formulations. The basic formulation is expressed as (Equation (1)):

$$\text{SFI} = R_{\max} \times f(V_1, V_2, \dots, V_n). \quad (1)$$

Table 1. Soil Parameter Rating Scales and Assigned Values (adapted from Steven Merumba et al. [21]).

Parameter	1	0.8	0.5	0.2	0
pH	6.5–7.5	7.4–8.5	5.5–6.4	4.5–5.4	<4.5 or >8.5
EC (dS m ⁻¹)	0–2	2.1–4	4.1–6	6.1–8	>8
OM (g kg ⁻¹)	>30	20.1–30	10.1–20	5.1–10	0–5
Zn (mg kg ⁻¹)	0.71–2.41	2.4–8.0	0.2–0.7	>8	<0.2
Fe (mg kg ⁻¹)	2.1–4.5	1.1–2.0	0.2–1.0	>4.5	<0.2

Notes: EC, Electrical Conductivity; OM, Organic Matter; Zn, Available Zinc; Fe, Available Iron.

Here, V_1, V_2, \dots, V_n represent the individual rating scores of each parameter, and f is either an arithmetic or geometric mean function. R_{\max} denotes the maximum attainable index value under ideal soil conditions. To better reflect the relative importance of each parameter, a weighted geometric mean formulation was also used:

$$FI = R_{\max} \times \left(\prod_{i=1}^n \left(\frac{V_i}{100} \right)^{w_i} \right)^{\frac{1}{\sum_{i=1}^n w_i}}. \quad (2)$$

V_i is the rating value (between 0 and 1) for the i th parameter, w_i is the weight assigned to that parameter, n is the number of parameters, R_{\max} normalizes the final score based on maximum achievable productivity.

This dual-formula framework ensures interpretability and flexibility, allowing for domain expert judgment to be incorporated without introducing circular logic in the machine learning validation stage.

A comprehensive soil parameter rating system was used to objectively score each indicator, following the scale shown in Table 1, adapted from Steven Merumba et al. [21]. Each parameter was categorized into five levels (1, 0.8, 0.5, 0.2, 0) based on established agronomic thresholds. For instance, a pH between 6.5 and 7.5 received a full score of 1.0, while values below 4.5 or above 8.5 were rated 0.0, indicating critical fertility limitations. Similar thresholds were applied to EC, OM, Zn, and Fe.

2.3. Modeling Approach

2.3.1. Dataset Preparation

The independent variables include soil electrical EC, OM content, soil pH, and the concentrations of micronutrients, namely Zn and Fe, which are considered as pedo-variables. The dependent variable is the SFI, computed as described previously. To ensure the robust training of the regression models and to mitigate potential issues arising from the differing scales of the input variables, all variables were standardized according to Equation (3). This preprocessing step is critical for the accurate functioning and comparative evaluation of the Ridge, Lasso (Least Absolute Shrinkage and Selection Operator), and ElasticNet regression models. By standardizing the data, we can effectively harmonize the input variables, thereby reducing scale-induced biases and enhancing model performance.

Such an approach not only facilitates the proper training and convergence of these regression models, but it also allows for a direct comparison of their predictive capabilities when applied to complex, multi-dimensional soil datasets.

$$z = \frac{x - \mu}{\sigma} \quad (3)$$

In the equation, z is the standardized (or z -score) value, x is the original (raw) data value, μ is the mean of the data, σ is the standard deviation of the data. Thus, z represents the transformed value obtained by subtracting the mean from the original data value and then dividing by the standard deviation.

2.3.2. Establishment of Modeling Architecture

To address common challenges in regression analysis, such as multicollinearity, overfitting, and difficulties in model interpretation, modern studies frequently employ regularization techniques like Ridge, Lasso, and Elastic Net [22]. Among these, Ridge regression is particularly effective in high-dimensional contexts. It counteracts the impact of intercorrelated predictors by incorporating a penalty term into the loss function, which systematically reduces the size of the regression coefficients [23]. The objective function of Ridge regression is given in Equation (4), where a regularization parameter (λ) controls the extent of coefficient shrinkage and helps diminish the variance of the model's predictions [24]. By minimizing the magnitude of the coefficients, this approach curbs the risk of overfitting and enhances the stability and interpretability of the model. In practice, these regularization methods enable researchers to extract clearer insights from complex datasets, ensuring that each predictor's contribution is appropriately weighted.

$$\min_{\beta} \left(\sum_{i=1}^n (y_i - X_i \beta)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right) \quad (4)$$

In this equation, the parameter (λ) governs the strength of the penalty term. When (λ) is small, the results approximate those of ordinary least squares (OLS); conversely, a larger (λ) value induces a greater shrinkage of the regression coefficients, which is advantageous in handling multicollinearity [25]. Ridge regression, in particular, has proven to be highly effective in high-dimensional datasets where predictors exhibit strong correlations [26].

Lasso regression adopts an alternative regularization approach by incorporating the sum of the absolute values of the regression coefficients as a penalty term. This method not only reduces the magnitude of the coefficients but can also shrink some of them to exactly zero, thereby performing automatic variable selection. Consequently, the resulting model becomes simpler and more interpretable [27]. The objective function for Lasso regression is presented in Equation (3), where the penalty term imposes sparsity by favoring solutions with fewer non-zero coefficients [28,29].

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - X_i \beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (5)$$

In this equation, β is regression coefficient, y_i is dependent variable, X_i is vector of independent variables, n is number of observations, p is number of variables, λ is penalty coefficient, $y_i - X_i \beta$ squared errors, β_j is absolute value of the coefficients.

Elastic Net regression combines the strengths of both Ridge and Lasso regression, aiming to overcome the limitations encountered when each method is applied individually. By incorporating both λ_1 and λ_2 penalty terms, it encourages group selection and maintains stability even in the presence of high multicollinearity. The objective function to be minimized for Elastic Net is expressed as follows (Equation (6)).

$$\min_{\beta} \left\{ \sum_{i=1}^n (y_i - X_i \beta)^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right\} \quad (6)$$

Here, y_i represents the dependent variable, while $x_{[ij]}$ denotes the independent variables. The parameter λ provides a balance between the λ_1 and λ_2 penalty terms [30,31].

2.3.3. Hyperparameter Optimization (GridSearchCV)

To determine the optimal values of the hyperparameters (e.g., λ and ρ) for the Ridge, Lasso, and Elastic Net models, the GridSearchCV method was employed. Each model was systematically scanned across predefined hyperparameter ranges (λ : 0.001, 0.01, 0.1, 1, 10; ρ : 0.2, 0.5, 0.8). The performance of each hyperparameter combination was evaluated using five-fold cross-validation ($cv = 5$) based on the negative mean squared error metric. This approach aimed to reduce the risk of overfitting and enhance model stability.

2.3.4. Model Validation and Performance Metrics

To evaluate the model's generalizability, the dataset was divided into training (80%) and testing (20%) subsets. The training data facilitated the model's learning process, while the test data allowed for the evaluation of model performance on previously unseen examples. Each model was retrained using the best-performing hyperparameter values on the training data.

To compare model performance, the Mean Squared Error (MSE, Equation (7)), Root Mean Squared Error (RMSE, Equation (8)), and the Coefficient of Determination (R^2 , Equation (9)) were calculated for both training and testing datasets. Additionally, the Residual Predictive Deviation (RPD, Equation (10)) metric was employed to assess the predictive performance of the models in relation to the natural variability of the dataset.

$$\text{MSE} = (1/n) \times \sum (y_i - \hat{y}_i)^2 \quad (7)$$

$$\text{RMSE} = \sqrt{\left((1/n) \times \sum (y_i - \hat{y}_i)^2 \right)} \quad (8)$$

$$R^2 = 1 - \left[\frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \right] \quad (9)$$

$$\text{RPD} = \sigma(y_{\text{test}}) / \text{RMSE} \quad (10)$$

In the equations above, y_i denotes the actual values, \hat{y}_i the predicted values, and $\sigma(y_{\text{test}})$ represents the standard deviation of the test data.

2.3.5. Interpretation and Recording of Coefficients

Unlike conventional modeling efforts that prioritize only prediction accuracy, our methodology explicitly focuses on quantifying the agronomic significance of each soil parameter through model-derived coefficient analysis. The regression coefficients obtained from the trained models are crucial for quantitatively supporting expert opinions. In particular, the coefficients derived from Ridge, Lasso, and Elastic Net regressions objectively reveal the relative influence of each independent variable (EC, OM, pH, Zn, Fe) on the SFI score. Finally, the slope (coefficient) and intercept values for each model were recorded in a DataFrame for further analysis.

2.4. Statistical Analysis

Descriptive statistical measures, including mean, standard deviation, minimum, maximum, and interquartile range, were computed for all soil variables (EC, OM, pH, Zn, Fe) as well as the SFI, providing a comprehensive overview of central tendencies, variability, and distributional characteristics. To examine bivariate relationships, Pearson correlation coefficients were calculated, and the resulting correlation matrix enabled a quantitative evaluation of linear associations. All statistical computations and visualizations were performed using Python 3.9 (Python Software Foundation, Wilmington, DE, USA) with relevant packages from the Scikit-learn (version 1.2.2) library and Pandas library (version 2.0.3).

3. Results

3.1. Descriptive Statistics

The descriptive statistics (Table 2) reveal significant variability in soil parameters across sampling sites. Electrical conductivity exhibited moderate variability (mean = 5.39 dS/m; SD = 3.30), with extremes ranging from 1.33 dS/m (non-saline) to 15.71 dS/m (severely saline). Organic matter averaged 59.99 g/kg (SD = 15.13), indicating generally fertile soils, though localized hotspots (max = 97.13 g/kg) contrasted with lower values (min = 26.10 g/kg). Soil pH remained consistently alkaline (range = 7.54–8.19), with minimal spatial variability (SD = 0.13).

Micronutrient availability varied widely. Zinc concentrations were critically low (mean = 0.03 mg/kg), while Fe showed high variability (range = 0.35–248.97 mg/kg). The SFI scores indicated pH as the most stable fertility parameter (SFI pH = 0.76 ± 0.08), whereas Zn availability (SFI Zn = 0.22 ± 0.06) emerged as a critical limiting factor. OM consistently scored near-optimal (SFI OM = 0.99 ± 0.04), while EC-driven salinity stress (SFI EC = 0.60 ± 0.26) and Fe variability (SFI Fe = 0.49 ± 0.27) contributed to moderate overall soil fertility (SFI Score = 0.61 ± 0.07).

Table 2. Descriptive statistics of the data.

Variable	Unit	Mean	Std Dev	Minimum	1.Quartile	Median	3.Quartile	Maximum
EC	dS/m	5.39	3.30	1.33	3.22	4.47	6.20	15.71
Organic Matter	g/kg	59.99	15.13	26.10	49.53	61.47	68.98	97.13
pH		7.90	0.13	7.54	7.81	7.92	7.99	8.19
Zn		0.03	0.04	0.00	0.01	0.02	0.04	0.25
Fe		27.22	39.62	0.35	6.13	13.79	31.73	248.97
SFI pH		0.76	0.08	0.60	0.80	0.80	0.80	0.80
SFI EC		0.60	0.26	0.20	0.40	0.60	0.80	1.00
SFI OM	mg/kg	0.99	0.04	0.80	1.00	1.00	1.00	1.00
SFI Zn		0.22	0.06	0.20	0.20	0.20	0.20	0.60
SFI Fe		0.49	0.27	0.20	0.20	0.50	0.80	1.00
SFI Score		0.61	0.07	0.48	0.56	0.60	0.68	0.76

Notes: EC, Electrical Conductivity; Zn, Available Zinc; Fe, Available Iron; SFI pH, Soil Fertility Index pH; SFI EC, Soil Fertility Index Electrical Conductivity; SFI OM, Soil Fertility Index Organic Matter; SFI Zn, Soil Fertility Index Available Zinc; SFI Fe, Soil Fertility Index Available Iron; SFI Score, Soil Fertility Index Score.

The correlation matrix (Figure 2) reveals complex interdependencies among soil properties and fertility indices. The EC demonstrates a strong negative correlation with SFI_EC (-0.90) and soil pH (-0.59), indicating that elevated salinity (high EC) is associated with reduced pH-adjusted fertility scores and alkaline conditions. EC also shows a moderate negative relationship with SFI_Total_Score (-0.53), underscoring the detrimental impact of salinity stress on overall soil fertility. Soil pH further exhibits a significant negative correlation with SFI_PH (-0.66), reflecting diminished nutrient availability under alkaline regimes. Micronutrients Zn and Fe display an exceptionally strong positive correlation (0.97), likely due to shared solubility controls (e.g., pH, redox dynamics) or co-sourced inputs, with both elements also positively linked to SFI_Zn (0.86 and 0.88 , respectively), suggesting their bioavailability directly enhances Zn-specific fertility metrics. Soil OM content correlates moderately with SFI_OM (0.39), reinforcing its role in stabilizing nutrient retention. The SFI_Fe and SFI_Total_Score share a strong positive relationship (0.76), highlighting critical contribution of Fe to fertility assessments, while SFI_EC's positive association with SFI_Total_Score (0.68) emphasizes the importance of salinity management in improving fertility outcomes. Contrasting trends emerge in the negative correlations between SFI_Fe and Zn/Fe (-0.57 and -0.53), suggesting potential antagonistic interactions or competing retention mechanisms in the soil matrix.

The correlation matrix reveals a significant negative trend between soil pH and several plant nutrients (Figure 3). Additionally, low to moderate positive correlations were observed between OM content and micronutrient concentrations. These observations highlight the multifaceted nature of nutrient dynamics in soil systems: while higher pH can limit micronutrient solubility, sufficient levels of OM have the potential to counterbalance these effects by facilitating nutrient mobilization and buffering against losses.

Potential linear and nonlinear relationships between selected pairs of variables are shown in Figure 4. These visualizations reveal that parameters such as EC, OM, pH, Zn, and Fe do not consistently exhibit strong linear correlations. Instead, multiple nuanced and complex interaction patterns emerge, indicating that soil properties are governed by multivariate relationships rather than simple pairwise associations. The interaction between EC and OM shows that a significant concentration of OM values occurs under low to moderate EC conditions. As OM content increases, EC values become more broadly distributed, suggesting that OM has a modulating but not singularly controlling effect on salinity levels. Soil pH remains relatively stable across the study area, ranging narrowly from 7.5 to 8.2. In contrast, EC varies widely, spanning from 1 to 16 dS/m. Although both

parameters are known to interact in certain soil environments, the current data does not suggest a strong or consistent linear relationship between them.

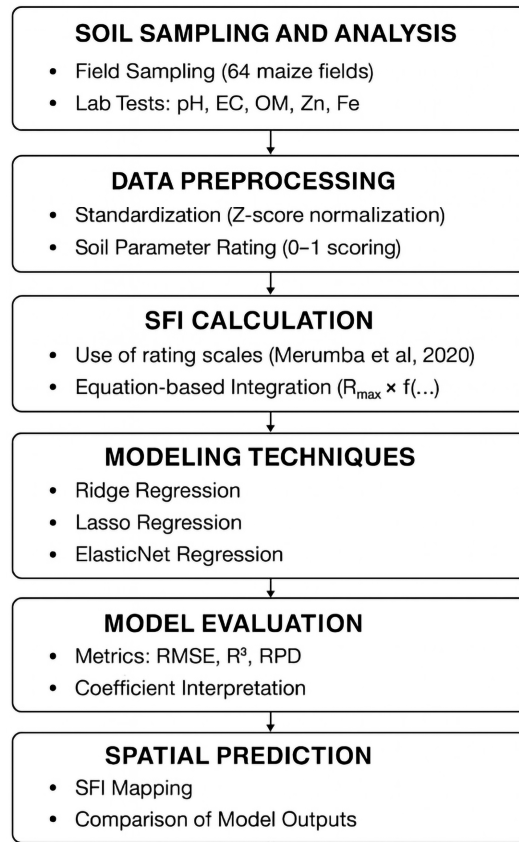


Figure 2. Flowchart of Soil Fertility Index (SFI) Modeling Framework.

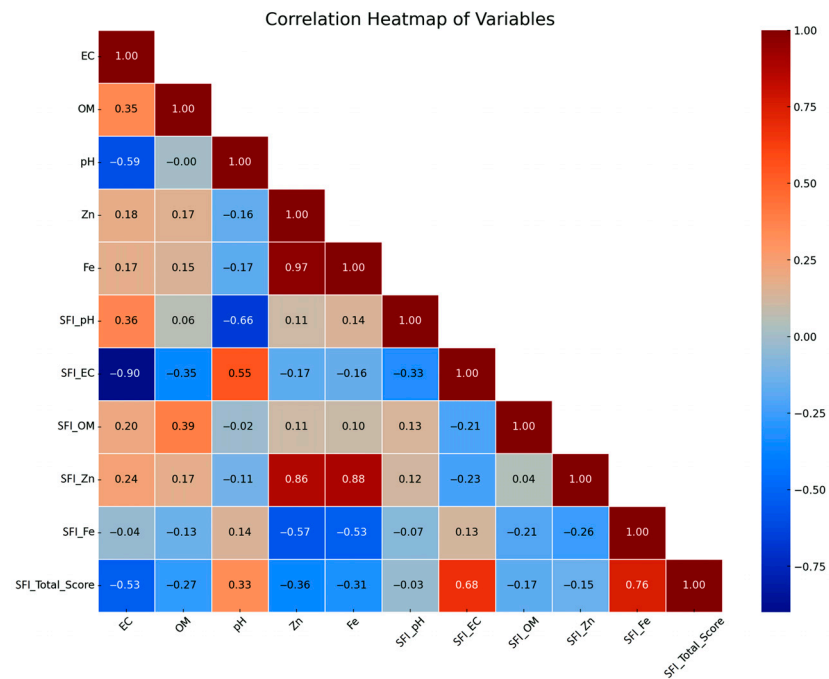


Figure 3. Correlation Patterns Among Soil Properties.

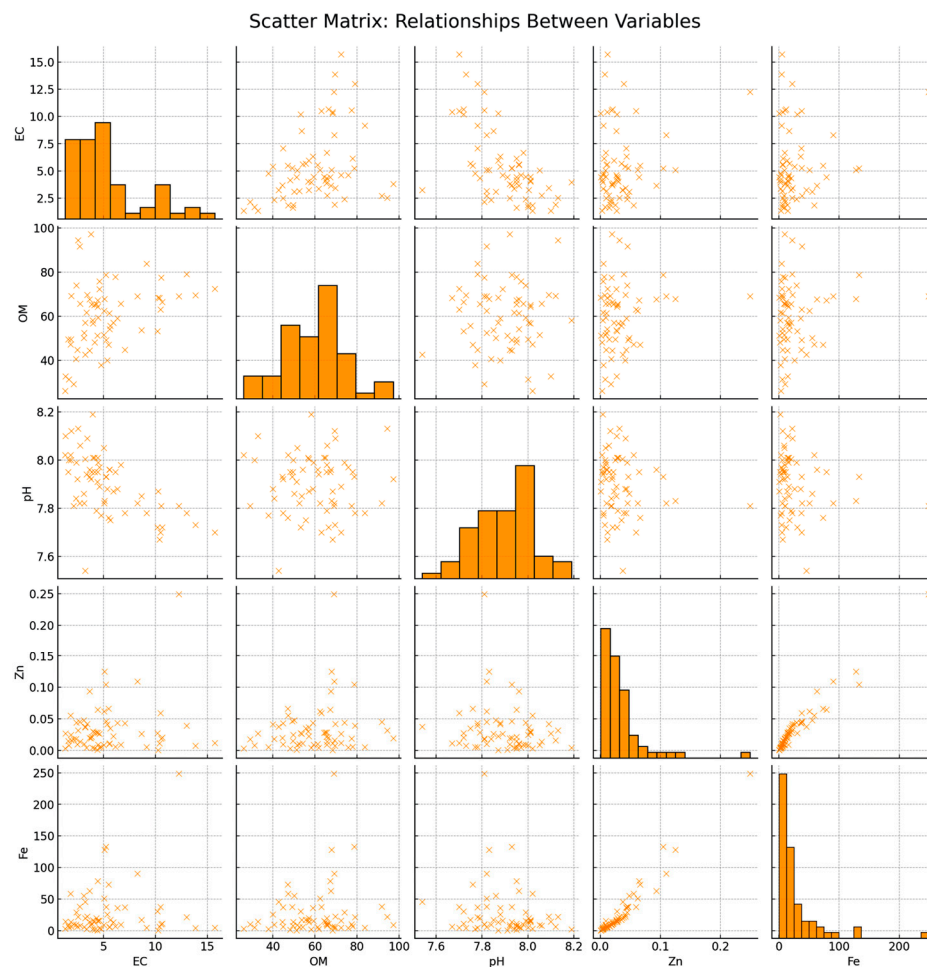


Figure 4. Bivariate Distribution of Soil Properties. Each × symbol marks one data point (individual soil sample) for the corresponding variable pair.

In examining the relationship between Zn and EC, it is evident that Zn concentrations exhibit limited variation at low EC levels, generally remaining within the 0–0.10 mg/kg range. Even as EC increases, Zn concentrations do not rise proportionally, implying a weak or non-existent linear correlation between Zn availability and salinity. A more consistent pattern is observed between OM and Zn, where increasing OM content is associated with a gradual rise in Zn concentration. This trend highlights the potential role of OM in enhancing Zn mobility and bioavailability in soil. Another important pattern is the clustering of low Zn (0.00–0.05 mg/kg) and moderate Fe (0–50 mg/kg) concentrations within the higher pH range of 7.8 to 8.2. This suggests that rising pH levels are associated with reduced solubility and availability of these micronutrients, particularly in alkaline soil conditions.

3.2. Model Evaluation

In this study, Ridge, Lasso, and ElasticNet regression techniques were employed to quantify the influence of soil properties (EC, OM, pH) and micronutrient concentrations (Zn, Fe) on SFI scores. These regularization methods were selected not only to mitigate multicollinearity risks but also to identify robust predictors of soil fertility. The magnitude and significance of the regression coefficients derived from these models provide empirical validation for expert-driven assessments of variable importance, thus bridging domain knowledge with data-driven insights. In addition, error metrics such as RMSE and MAE were computed to evaluate predictive accuracy and ensure the models' applicability for agronomic decision-making.

The results from the Ridge regression analysis indicate that all independent variables contributed to the prediction of the SFI, although each had relatively modest individual effects. The calculated β coefficients were as follows: EC = -0.009403 , OM = -0.003393 , pH = 0.005240 , Zn = -0.006554 , and Fe = -0.004987 , with an intercept term of 0.617255 (see Equation (8)). Owing to Ridge regression’s nature of penalizing coefficient magnitude rather than excluding variables, every variable was retained in the model despite the relatively small coefficient values, which nonetheless suggest limited independent predictive power. The Lasso regression results, on the other hand, revealed significant predictive influence for EC ($\beta = -0.021645$) and Zn ($\beta = -0.012754$), while OM, pH, and Fe were excluded from the model as their coefficients were shrunk to zero. This sparsity-inducing property of Lasso regression simplifies the model by retaining only the most informative predictors [32]. Furthermore, the ElasticNet approach, which integrates regularization penalties from both Ridge and Lasso regression, demonstrated that while EC ($\beta = -0.012621$) and Zn ($\beta = -0.004580$) retained significant contributions to the model, OM, pH, and Fe exhibited near-zero or negligible coefficients, indicating their limited influence on SFI predictions. The final predictive equations derived from the three approaches were as follows (Equations (11)–(13)):

$$SNI_{Ridge} = 0.617 - 0.010 \times OM + 0.005 \times pH - 0.018 \times EC - 0.011 \times Zn - 0.006 \times Fe, \quad (11)$$

$$SNI_{Lasso} = 0.617 - 0.022 \times EC - 0.013 \times Zn, \quad (12)$$

$$SNI_{ElasticNet} = 0.617 - 0.013 \times EC - 0.005 \times Zn. \quad (13)$$

3.3. Accuracy Assessment

Ridge, Lasso, and ElasticNet regression models were systematically evaluated to predict SFI scores using machine learning (Table 3). Hyperparameter tuning identified optimal configurations: Ridge regression achieved peak performance at a regularization strength (α) of 100, Lasso at $\alpha = 0.01$, and ElasticNet at $\alpha = 0.10$ with an l1 ratio of 0.2. Predictive performance on the test dataset revealed Lasso as the top performer, with an RMSE of 0.060, R^2 of 0.746, and RPD of 1.152. Ridge ranked second (RMSE = 0.064, $R^2 = 0.632$, RPD = 1.073), while ElasticNet yielded slightly lower accuracy (RMSE = 0.066, $R^2 = 0.684$, RPD = 1.045).

Table 3. Hyperparameters and Validation Metrics for Ridge, Lasso, and ElasticNet Regression Models.

Model	Ridge	Lasso	ElasticNet
Best Alpha	100	0.01	0.1
l1 ratio			0.2
MSE Train	0.00	0.00	0.00
RMSE Train	0.06	0.06	0.06
R^2 Train	0.83	0.89	0.78
MSE Test	0.00	0.00	0.00
RMSE Test	0.06	0.06	0.07
R^2 Test	0.63	0.75	0.68
RPD Test	1.07	1.15	1.04

Notes: MSE, Mean Square Error; RMSE, Root Mean Squared Error; R^2 , the coefficient of determination; RPD, Residual Predictive Deviation.

Training dataset evaluations showed Lasso with the strongest fit ($R^2 = 0.885$), followed by Ridge ($R^2 = 0.833$) and ElasticNet ($R^2 = 0.781$). The discrepancy between training and test R^2 values was minimal for Lasso (0.885 vs. 0.746), whereas larger drops were observed for Ridge (0.833 vs. 0.632) and ElasticNet (0.781 vs. 0.684), indicating varying degrees of overfitting.

3.4. Model-Based Spatial Predictions

Figures 5 and 6 presents the spatial distribution of SFI values based on observed data and predictions from three regularized linear regression models: Ridge, Lasso, and ElasticNet. The observed SFI scores exhibited a broad range (0.48–0.76), reflecting significant spatial heterogeneity in soil fertility across the study area. Higher values concentrated in the southern and southwestern zones, while lower values were prevalent in the northern and northeastern parts (Figure 5A). The Ridge model yielded predicted values ranging from 0.52 to 0.64, predominantly showing medium fertility scores throughout the region (Figure 5B).

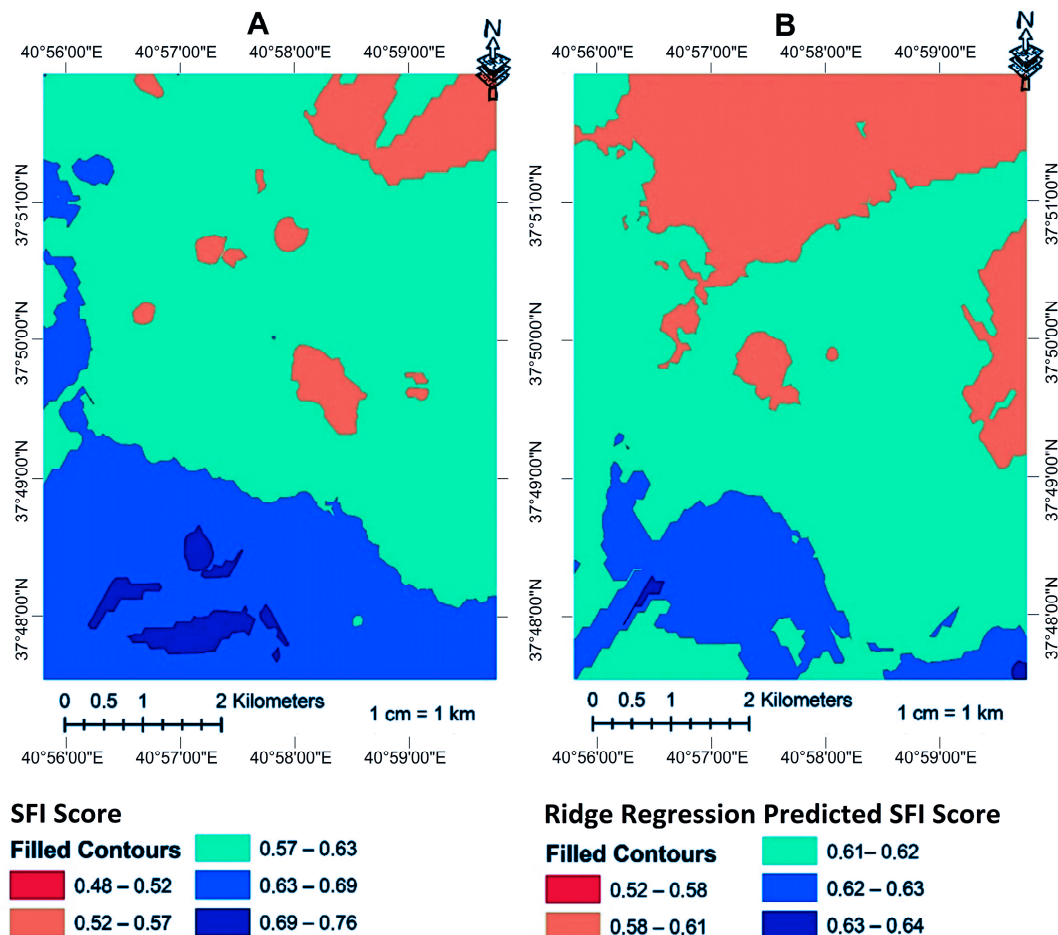


Figure 5. Comparison of observed and predicted SFI scores across regression models. The contour plots illustrate (A) actual SFI distribution and (B) Ridge predictions.

While it captured broad spatial patterns, its predictions were smoother, underestimating both the highest and lowest SFI values seen in the observed dataset. Lasso demonstrated a better capacity to replicate spatial variation in SFI, particularly in distinguishing low-fertility zones in the north and higher scores in the southern part of the study area (Figure 6A). Predicted values spanned 0.52 to 0.64, aligning closely with the observed SFI map. The ElasticNet model produced predictions highly consistent with Lasso but demonstrated slightly improved delineation in transitional areas (Figure 6B), maintaining spatial heterogeneity without over-smoothing. All models consistently underestimated high SFI values (>0.69), failing to resolve localized nutrient-rich zones.

The Lasso model, which demonstrated superior statistical performance (RMSE = 0.060, $R^2 = 0.746$), generated spatial predictions that closely mirrored the Ridge and ElasticNet outputs. This paradox highlights that while Lasso achieved higher numerical accuracy, its

regularization-driven sparsity (retaining only EC and Zn as predictors) did not translate to distinct spatial differentiation compared to other models. Ridge regression ranked second (RMSE = 0.064, $R^2 = 0.632$), and ElasticNet yielded slightly lower accuracy (RMSE = 0.066, $R^2 = 0.684$). Training dataset evaluations revealed Lasso's strong fit ($R^2 = 0.885$) with minimal overfitting (test $R^2 = 0.746$), whereas Ridge ($R^2 = 0.833$ vs. 0.632) and ElasticNet ($R^2 = 0.781$ vs. 0.684) exhibited larger training-test gaps.

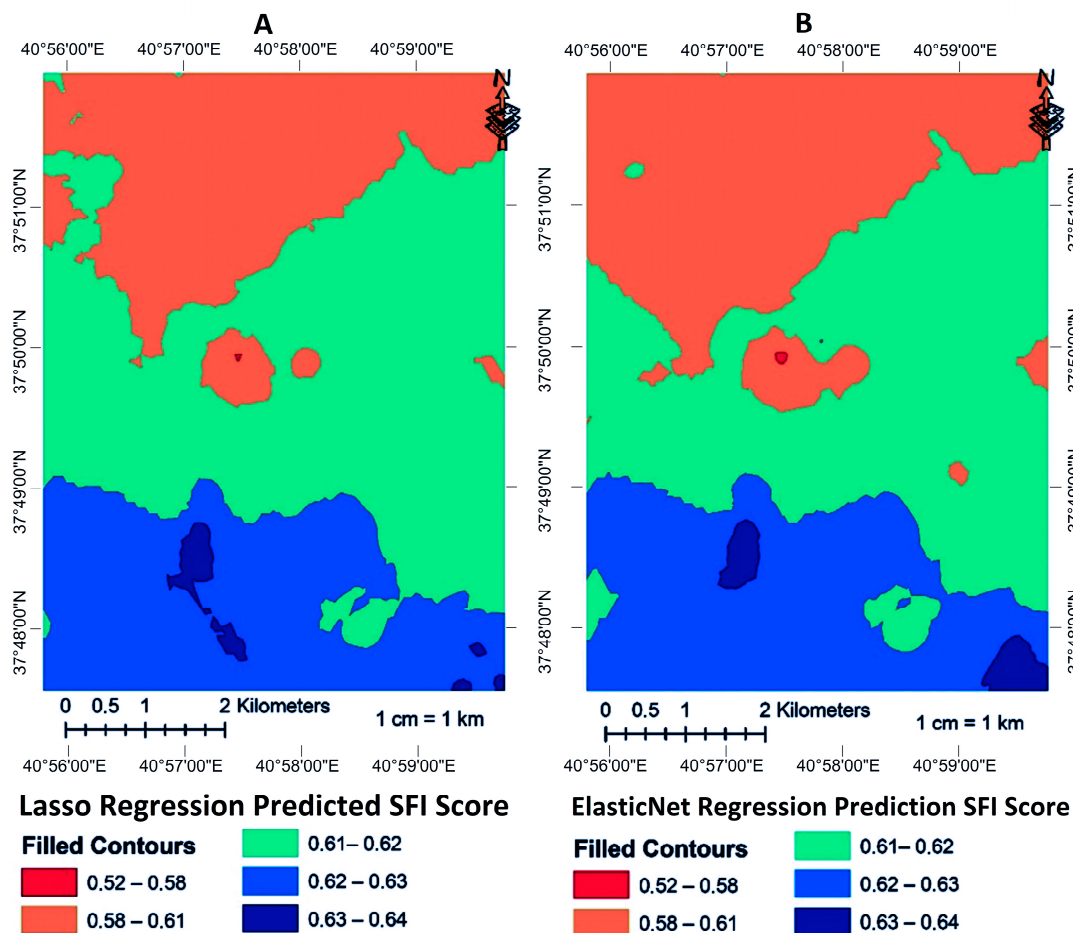


Figure 6. Comparison of observed and predicted SFI scores across regression models. The contour plots illustrate (A) Lasso and (B) ElasticNet predictions.

4. Discussion

Slightly alkaline soil pH (6.5–8.0) is a critical determinant of nutrient bioavailability, especially micronutrients such as Fe, Zn, Mn, and Cu. In this range, nutrient solubility is highly sensitive to minor pH shifts, with alkaline conditions promoting precipitation or conversion of these elements into insoluble forms, thereby limiting plant uptake [33–35]. Electrical conductivity (EC) is another key electrochemical indicator, reflecting both salinity status and ionic nutrient mobility. While moderate EC levels may facilitate nutrient diffusion in the rhizosphere, elevated EC often signifies salinity stress, potentially linked to excessive irrigation, poor drainage, or inherent parent material composition. High salinity impairs osmotic potential and restricts nutrient assimilation, ultimately reducing crop yield [36,37]. While moderate EC can enhance ionic nutrient mobility, excessive salinity impairs osmotic balance and inhibits nutrient uptake, ultimately threatening crop yield [38].

Organic matter, by contrast, showed a relatively symmetrical spatial distribution in our dataset, interrupted by localized organic-rich patches. Such patterns likely stem from vege-

tative cover, crop residue accumulation, or targeted anthropogenic amendments. Increased OM enhances nutrient retention capacity, microbial biomass activity, and cation exchange capacity, particularly under no-tillage or organic management regimes [39]. These synergistic interactions between pH, OM, and EC suggest a multifactorial regulatory network governing soil fertility status. Therefore, integrated interpretation of these parameters is essential for designing precision soil management strategies that optimize nutrient use efficiency while sustaining long-term agricultural productivity.

4.1. Exploring Zinc and Iron Mobility Under the Influence of Alkaline pH and Organic Matter in Cultivated Soils

The observed right-skew in Zn and Fe concentrations may reflect localized accumulation processes influenced by variable fertilizer application rates, differing irrigation water quality, or historical land use practices. Particularly in peri-urban and intensively cultivated areas, cumulative agrochemical loading has been shown to induce these spatial anomalies [40,41]. Although soil heterogeneity driven by parent material and pedogenic processes contributes to this distribution, anthropogenic impacts remain dominant in cultivated regions [42].

The solubility and mobility of Zn and Fe are largely governed by soil pH, OM content, and redox potential. For instance, Fe precipitates under alkaline conditions, forming immobile oxides, especially under fluctuating moisture regimes. Similarly, availability of Zn is reduced at higher pH due to precipitation as carbonates or hydroxides [34,35]. These effects become more pronounced when irregular fertilizer application interacts with soil buffering capacity and microbial dynamics, necessitating integrated nutrient management [43].

The heatmap patterns reveal the complex, nonlinear nature of soil interactions. Weak linear correlations among EC, OM, pH, Zn, and Fe indicate that single-variable regressions are insufficient to capture soil behavior. The broad EC range observed in high-OM areas, under relatively stable pH, suggests a multifactorial influence involving soil texture, irrigation quality, and drainage performance, previously identified as key EC modulators [44].

While alkaline pH is commonly linked with increased salinity and sodium-induced stress due to disrupted ionic balance, emerging evidence indicates that this relationship is not universally consistent. Recent studies have shown that the interplay between soil pH, salinity, and nutrient availability is modulated by complex biogeochemical interactions rather than direct pH-EC correlation [45]. Our dataset similarly revealed no significant correlation between pH and EC, highlighting the importance of assessing salinity within a broader matrix of soil physical and chemical properties. Additionally, Zn concentrations did not proportionally increase with higher EC values, consistent with findings that Zn availability in saline-alkaline soils is more strongly governed by pH-induced solubility and ion competition than by salinity alone [46].

Organic matter plays a context-dependent but substantial role in soil fertility. OM can lower pH through the release of organic acids and facilitate micronutrient solubility by forming stable chelates with Zn^{2+} and Fe^{3+} , enhancing their mobility and bioavailability [47,48]. This chelation mechanism becomes especially crucial in calcareous and alkaline soils, where Zn and Fe are prone to forming insoluble compounds. Recent studies have confirmed that humic acid-enriched organic amendments can improve micronutrient uptake and yield in crops grown in high-pH soils [49]. The positive association between OM and Zn in our data aligns with previous findings that OM additions reduce Zn sorption by up to 40% in alkaline soils [50]. Furthermore, field research continues to support the positive correlation between OM and Zn uptake in cereals like wheat and maize [51]. These insights affirm the importance of OM not only as a structural component but as a biochemical mediator of nutrient dynamics.

The co-occurrence of low Zn and moderate Fe levels at higher pH reinforces the susceptibility of micronutrient solubility to alkaline conditions. Classic studies [19,52] remain relevant in explaining how elevated pH favors the formation of insoluble Zn and Fe complexes. However, modern research emphasizes the interaction of pH with OM, redox state, and soil microbial activity as equally critical factors shaping micronutrient availability in semi-arid and calcareous systems [48,49].

The observed interactions between EC, pH, OM, and micronutrients align with established principles of soil fertility in agricultural systems. Elevated EC is widely documented to reduce crop yield by restricting water uptake and nutrient mobility, especially in semi-arid environments [53]. Our findings support this, showing strong negative correlations between EC and SFI_Total_Score. Similarly, alkaline pH impairs the availability of key micronutrients such as Zn and Fe [50], which explains the negative correlation trends between soil pH and SFI_Zn/Fe. In contrast, OM acts as a stabilizer of soil fertility by enhancing cation exchange, buffering pH, and forming complexes with metal ions [54]. The strong positive correlations observed between OM and SFI_OM, as well as between micronutrients and their SFI scores, support their use as diagnostic indicators of soil quality. These relationships are supported by findings from Bünemann et al. [55], who emphasize that nutrient availability and retention processes, modulated by pH, organic matter, and salinity, are fundamental to soil function and productivity assessments. Therefore, the inclusion of these indicators in our modeling framework is not only statistically relevant but also agronomically grounded, reflecting processes that govern soil productivity in intensively cultivated regions.

4.2. Data-Driven Identification of Critical Soil Fertility Drivers Using Regularized Regression Models

The negative coefficients for EC and Zn observed in the Ridge regression analysis are consistent with agronomic understanding of how elevated salinity and excessive micronutrient levels can inhibit plant nutrient uptake. High soil EC reflects salinity stress, which compromises osmotic balance and root membrane function, thereby reducing the efficiency of ion absorption and nutrient transport. Recent studies confirm that elevated salinity leads to reduced nutrient use efficiency due to disrupted soil–plant nutrient dynamics and increased physiological stress [56,57].

The positive coefficient associated with soil pH reaffirms its essential role in governing micronutrient availability, especially in alkaline soils where solubility dynamics are highly pH sensitive. Micronutrients such as Fe and Zn tend to form insoluble compounds at elevated pH, making them inaccessible to plants. Empirical relationships derived from regularized regression models reinforce the need for multivariate frameworks to capture these interactive and often nonlinear dynamics. Moreover, excess Zn may competitively inhibit the uptake of other micronutrients like Fe and Mn due to ionic competition at the root interface and altered cation exchange dynamics. These interactions are increasingly recognized as major contributors to nutrient imbalance and suboptimal crop growth, as corroborated by Kumar et al. [58] and supported by recent empirical evidence [59].

In the Lasso regression model, EC and Zn emerged as the only retained variables (EC, $\beta = -0.021645$; Zn, $\beta = -0.012754$), while OM, pH, and Fe were excluded due to their low explanatory contributions. This selective feature retention reflects the model's capacity to isolate high-influence predictors, suggesting that EC and Zn are dominant fertility drivers under the specific physiochemical context of the study. Feng et al. [60] similarly demonstrated the robustness of EC and Zn as explanatory variables in soil nutrient transport models across diverse edaphic conditions.

The consistent prominence of EC and Zn across Ridge, Lasso, and ElasticNet models suggests that these variables play central roles in nutrient mobility, root uptake efficiency,

and micronutrient homeostasis. EC's negative effect is particularly notable in saline-prone zones where osmotic and ionic stress constrains root functionality, reducing the uptake of essential elements. Simultaneously, Zn's dual role, as a nutrient and an antagonist, shows its delicate balance in soil systems. Excess Zn not only disrupts cationic balance but also exacerbates deficiencies in co-regulated nutrients like Fe and Mn, particularly in soils with low buffering capacity [57].

These results align with agronomic understanding that EC and Zn reflect critical constraints in irrigated alkaline soils. To further contextualize their inclusion in fertility modeling, it is essential to consider the broader roles of macro- and micronutrients in defining soil health. A meaningful SFI must capture the integrative effects of both macro- and micronutrients, as they jointly regulate plant health, productivity, and long-term soil function. Macronutrients such as nitrogen, phosphorus, and potassium govern vegetative and reproductive growth, while micronutrients, including An and Fe, play pivotal roles in enzymatic activity, chlorophyll formation, and hormonal balance. Importantly, deficiencies in micronutrients often occur despite adequate macronutrient supply and can lead to "hidden hunger" in plants [61]. The inclusion of DTPA-extractable Zn and Fe in our SFI framework is supported by findings from Kaur et al. [62], who reported strong correlations between micronutrient availability and soil pH, EC, and OM in Himalayan soils. Additionally, Athokpam et al. [63], demonstrated that micronutrient concentrations are often influenced by OM content and salinity, reinforcing the logic behind selecting pH, EC, and OM as explanatory variables. Recognizing these interactions enhances our index's ability to reflect latent fertility constraints, particularly in semi-arid alkaline soils where Zn and Fe deficiencies are prevalent. Thus, this multidimensional approach bridges chemical, biological, and management-driven processes, and supports site-specific fertility assessment beyond a purely yield-focused framework.

ElasticNet, combining the strengths of Ridge and Lasso, confirmed this trend by assigning substantial weights to EC and Zn while minimizing the influence of OM, Fe, and pH. The approach effectively balanced sparsity with collinearity management, but its marginally lower predictive performance suggests that Lasso's selective simplicity may offer greater utility for applied agronomy, particularly in digital soil fertility mapping contexts. This pattern mirrors findings from Swaminathan et al. [64], who noted Lasso's practical superiority in filtering relevant predictors from high-dimensional soil datasets.

Lasso's superior performance, reflected in a low training-test R^2 gap, shows its generalization strength, avoiding overfitting through its capacity to suppress noise variables. In contrast, Ridge regression, with a high regularization parameter ($\alpha = 100$), favored parameter shrinkage over feature selection, which may lead to underfitting in complex agronomic systems. The ElasticNet's intermediate L1 ratio (0.2) produced smoother parameter gradients but failed to surpass Lasso in predictive accuracy, likely due to the overwhelming influence of dominant features that benefited more from sparse selection than coefficient smoothing.

Together, these results suggest that Lasso is best suited for scenarios requiring both model interpretability and actionable agronomic insights. It enables precise identification of key fertility constraints, offering a pathway for targeted nutrient interventions. Future work should explore hybrid ensemble models or kernel-based methods (e.g., SVM, XGBoost 2.1.1) to capture potential nonlinearities and spatial heterogeneity without sacrificing generalizability.

4.3. Trade-Offs in Regularization, Spatial Realism, and Practical Implications for Soil Fertility Modeling

The spatial prediction maps and model diagnostics reveal key contrasts in how different regularization methods address the heterogeneity of soil fertility across landscapes.

The L2 distributed coefficient weights evenly, yielding smoothed, generalized predictions that favor model stability but suppress fine-scale variability, an issue particularly limiting for precision agriculture [65]. This trade-off reduces variance but often underrepresents nutrient hotspots critical for site-specific management.

Conversely, L1 regularization enhanced model interpretability and spatial differentiation by selecting a sparse set of predictors, primarily EC and Zn in this study. Its capacity to minimize overfitting, evidenced by a narrow training–test R^2 gap, is advantageous for generalization. However, this sparsity came at the cost of excluding variables like OM and pH, which are known to modulate micronutrient mobility and retention, especially under alkaline conditions [34]. The omission of such variables likely contributed to homogenized predictions in microsites enriched in organic matter.

ElasticNet, which combines L1 and L2 penalties, offered a more balanced performance. While it retained multiple predictors, including OM and pH (albeit with low-weight coefficients), it delivered smoother spatial transitions between fertility zones, beneficial for preventing over-segmentation in zonal management [66]. This reflects a nuanced balance between variable selection and coefficient shrinkage, supporting gradual nutrient interpolation and avoiding abrupt classification changes across field boundaries.

The observed convergence of predicted SFI values across models, particularly in moderate fertility classes, indicates the dominant roles of EC and Zn as fertility drivers in the region. However, regularization techniques systematically attenuated extreme SFI values (>0.69), aligning with the right-skewed distributions of Zn and Fe. This highlights a broader limitation: regularization penalizes outlier-driven patterns, potentially masking fertility hotspots that are agronomically significant [58].

In practical terms, Lasso stands out for its simplicity and robust predictive capacity, making it well-suited for operational digital soil mapping. Yet, in contexts requiring gradient-based recommendations, such as variable-rate fertilization or land use zoning, ElasticNet's continuous spatial rendering may offer superior agronomic value. Future research should explore hybrid frameworks integrating domain-specific rules (e.g., OM-pH interactions) or federated learning systems to maintain interpretability while capturing fine-grained spatial variability [67].

To complement these modeling insights, the spatial patterns illustrated in Figures 5 and 6 offer additional implications for interpreting local fertility constraints and targeting site-specific management. The spatial distribution maps presented in Figure 5 reveal key patterns in observed and predicted SFI values across the Batman Plain. The southern and southeastern zones consistently appear as lower fertility areas across both observed and modeled maps. These regions correspond with higher EC and reduced Zn concentrations, two factors known to limit nutrient uptake in semi-arid alkaline soils [68–70]. In contrast, central and northeastern subregions demonstrate moderately high fertility scores, likely supported by more favorable OM content and lower salinity stress.

The predictive maps generated by Ridge, Lasso, and Elastic Net regressions follow the general spatial trend of the observed SFI but differ in the intensity and sharpness of contour transitions. Elastic Net and Lasso, in particular, better capture spatial smoothing and localized variability, aligning with their superior test performance and lower RMSE. These models offer strong potential for spatial diagnostics, especially when used in conjunction with observed soil constraints.

From a management perspective, the discrete fertility zones revealed in these maps provide a practical foundation for site-specific nutrient interventions, enabling more efficient targeting of amendments and conservation practices [71]. The ability to spatially delineate

areas of suboptimal fertility supports the development of variable-rate fertilization plans, reducing input costs and environmental risks while improving productivity.

The use of regularized regression in this study represents an intermediate step toward more advanced artificial intelligence (AI)-based modeling in soil fertility diagnostics. These interpretable machine learning models (Ridge, Lasso, and Elastic Net) serve as robust tools capable of balancing model accuracy with agronomic interpretability. Their performance in capturing critical drivers of fertility (e.g., EC and Zn) underscores their value in early-stage digital agriculture applications. Looking ahead, the future of AI in soil fertility modeling lies in the integration of large-scale, multi-temporal data (e.g., remote sensing, crop yield sensors) with nonlinear modeling frameworks such as random forests, support vector machines, or deep neural networks [6,60]. These models could offer real-time diagnostics and adaptively update recommendations based on observed crop responses. While this study did not benchmark against traditional statistical models, the chosen regularized methods already extend classical regression by addressing multicollinearity and variable sparsity. Future research should incorporate side-by-side evaluations to quantify the added value of AI models over conventional approaches in real-world agronomic decision-making.

Comparative research in Mediterranean and similar semi-arid regions increasingly applies interpretable and high-performing machine learning to soil quality modeling. For instance, Mohammed et al. [72] used Recursive Feature Elimination and SHAP-enhanced Nu-Support Vector Regression and other ML models to pinpoint cation exchange capacity, Ca^{2+} , Mg^{2+} , and Na^+ as key predictors of sodium adsorption ratio in the eastern Mediterranean. Further, Mohamed et al. [73] combined embedded Lasso regression with multisource remote sensing to enhance salinity mapping in arid Egypt; Mohammadifar et al. [74] merged deep learning with SHAP to reveal climatic and textural controls on salinity in southern Iran. Nutrient-focused studies such as Bouslihim et al. [75] in Morocco and Khosravani et al. [76] in Iran used feature-selected Random Forests for transparent, accurate soil fertility and aggregate stability maps. Regionally, Ozturk et al. [77] demonstrated that a stacked Multi-Layer Perceptron (MLP) meta-model outperformed kriging in mapping soil pH and EC; Abakay et al. [78] achieved superior particle size predictions using Boruta-selected covariates with XGBoost in southeastern Turkey; and Gunal et al. [79] reported that MLP-ANN and Support Vector Regression pH maps were up to 55% more accurate than ordinary kriging under limited covariate conditions. Together, these studies underscore how feature selection, model stacking, and explainable algorithms can outperform traditional approaches while maintaining agronomic interpretability. Our study builds on this foundation by showing that, in semi-arid alkaline soils, EC and Zn consistently emerge as dominant predictors of a composite fertility index, reinforcing their relevance for targeted management in data-scarce environments.

5. Conclusions

This study presents an interpretable framework for evaluating soil fertility using Ridge, Lasso, and ElasticNet regression models. These techniques reliably predicted Soil Fertility Index (SFI) scores using key indicators such as EC, OM, pH, Zn, and Fe, while also enabled transparent evaluation of variable importance, reinforcing the link between empirical data and expert knowledge. Lasso regression identified EC and Zn as primary predictors, aligning with known salinity and micronutrient dynamics in semi-arid soils. Ridge and ElasticNet added interpretive nuance by retaining context specific variables like OM and pH.

Despite moderate R^2 values (0.632–0.746), the models performed well given the limited predictors and complexity of the SFI. Lasso achieved low RMSE (0.060) with minimal overfitting, supporting its use in data-scarce environments. Importantly, EC-related salinity

stress may involve nonlinear effects not fully captured here, and model outputs may smooth out subtle spatial features or outliers. These caveats suggest future studies should explore hybrid approaches to enhance resolution and robustness.

While spatial maps smooth extreme values, they still show meaningful fertility gradients that support more efficient nutrient management. The findings reaffirm established chemical relationships, such as the inverse link between pH and micronutrient solubility, and the positive role of organic matter in enhancing micronutrient retention. Although regularized regressions addressed multicollinearity and overfitting effectively, they may underrepresent localized soil anomalies or nonlinear salinity responses. Future studies could adopt hybrid models to improve prediction robustness. This research advances digital soil mapping in data-sparse environments by combining predictive modeling with agronomic interpretability.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: For access to the data used in this analysis, please contact the author.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Krasilnikov, P.; Taboada, M.A.; Amanullah. Fertilizer use, soil health and agricultural sustainability. *Agriculture* **2022**, *12*, 462. [CrossRef]
2. Minasny, B.; McBratney, A.B. Digital soil mapping: A brief history and some lessons. *Geoderma* **2016**, *264*, 301–311. [CrossRef]
3. Liptzin, D.; Norris, C.E.; Cappellazzi, S.B.; Mac Bean, G.; Cope, M.; Greub, K.L.; Rieke, E.L.; Tracy, P.W.; Aberle, E.; Ashworth, A.; et al. An evaluation of carbon indicators of soil health in long-term agricultural experiments. *Soil Biol. Biochem.* **2022**, *172*, 108708. [CrossRef]
4. Tunçay, T.; Kılıç, Ş.; Dedeoğlu, M.; Dengiz, O.; Başkan, O.; Bayramin, I. Assessing soil fertility index based on remote sensing and GIS techniques with field validation in a semiarid agricultural ecosystem. *J. Arid Environ.* **2021**, *190*, 104525. [CrossRef]
5. Adhikari, K.; Hartemink, A.E. Linking soils to ecosystem services—A global review. *Geoderma* **2016**, *262*, 101–111. [CrossRef]
6. Jia, X.; Fang, Y.; Hu, B.; Yu, B.; Zhou, Y. Development of Soil Fertility Index Using Machine Learning and Visible–Near Infrared Spectroscopy. *Land* **2023**, *12*, 2155. [CrossRef]
7. Adeniyi, O.D.; Brenning, A.; Maerker, M. Spatial prediction of soil organic carbon: Combining machine learning with residual kriging in an agricultural lowland area (Lombardy region, Italy). *Geoderma* **2024**, *448*, 116953. [CrossRef]
8. Matyunin, G.; Ogorodnikova, S.; Murmantseva, E.; Rozanov, V.; Palyga, R. Assessment of soil fertility indicators based on remote sensing data. In *BIO Web of Conferences, Proceedings of the XVII International Scientific and Practical Conference "State and Development Prospects of Agribusiness" (INTERAGROMASH 2024), Rostov-on-Don, Russia, 22–25 May 2024*; EDP Sciences: London, UK, 2024; Volume 113, p. 04013. [CrossRef]
9. Demir, S.; Sahin, E.K. The effectiveness of data pre-processing methods on the performance of machine learning techniques using RF, SVR, Cubist and SGB: A study on undrained shear strength prediction. *Stoch. Environ. Res. Risk Assess.* **2024**, *38*, 3273–3290. [CrossRef]
10. Kebonye, N.M.; John, K.; Chakraborty, S.; Agyeman, P.C.; Ahado, S.K.; Eze, P.N.; Němeček, K.; Drábek, O.; Borůvka, L. Comparison of multivariate methods for arsenic estimation and mapping in floodplain soil via portable X-ray fluorescence spectroscopy. *Geoderma* **2021**, *384*, 114792. [CrossRef]
11. Friedman, J.H.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22. [CrossRef]
12. Rahman Nur, A.; Jaya, A.K.; Siswanto, S. Comparative Analysis of Ridge, Lasso and Elastic Net Regularization Approaches in Handling Multicollinearity for Infant Mortality Data in South Sulawesi. *J. Mat. Stat. Dan Komputasi* **2024**, *20*, 311–319. [CrossRef]
13. General Directorate of Meteorology. 2025. Available online: <https://www.mgm.gov.tr/veridegerlendirme/il-ve-ilceler-istatistik.aspx?m=BATMAN> (accessed on 10 May 2025).
14. Eren, Y. General Assessment of the Geology, Tectonics, Seismicity, and New Settlement Area Selection of the Immediate Vicinity of Batman Province. *Tech. Rep.* **2011**. [CrossRef]

15. Loiseau, T.; Arrouays, D.; Richer-de-Forges, A.C.; Lagacherie, P.; Ducommun, C.; Minasny, B. Density of soil observations in digital soil mapping: A study in the Mayenne region, France. *Geoderma Reg.* **2021**, *24*, e00358. [CrossRef]
16. Gee, G.W.; Boudier, J.W. Particle size analysis. In *Methods of Soil Analysis. Part I, Agronomy No. 9*; Clute, A., Ed.; American Society of Agronomy: Madison, WI, USA, 1986.
17. Rhoades, J.; Chandavi, D.; Lesch, S.F. *Soil Salinity Assessment Methods and Interpretation of Electrical Conductivity Measurements*; FAO Irrigation and Drainage Paper 57; Food and Agriculture Organization of the United Nations: Rome, Italy, 1999. Available online: <https://openknowledge.fao.org/handle/20.500.14283/x2002e> (accessed on 20 August 2025).
18. Nelson, D.W.; Sommers, L.E. *Methods of Soil Analysis, Part 2. Chemical and Microbiological Properties*, 2nd ed.; Page, A.L., Miller, R.H., Keeney, D.R., Eds.; SSSA Inc.: Madison, WI, USA, 1982.
19. Lindsay, W.L.; Norvell, W. Development of a DTPA soil test for zinc, iron, manganese, and copper. *Soil Sci. Soc. Am. J.* **1978**, *42*, 421–428. [CrossRef]
20. Saraswat, A.; Ram, S.; AbdelRahman, M.A.E.; Raza, M.B.; Golui, D.; HC, H.; Lawate, P.; Sharma, S.; Dash, A.K.; Scopa, A.; et al. Combining Fuzzy, Multicriteria and Mapping Techniques to Assess Soil Fertility for Agricultural Development: A Case Study of Firozabad District, Uttar Pradesh, India. *Land* **2023**, *12*, 860. [CrossRef]
21. Merumba, M.S.; Semu, E.; Semoka, J.M.; Msanya, B.M. Soil Fertility Status in Bukoba, Missenyi and SharmaBiharamulo Districts in Kagera Region, Tanzania. *Int. J. Appl. Agric. Sci.* **2020**, *6*, 96. [CrossRef]
22. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*. In *Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009. [CrossRef]
23. Roozbeh, M. Shrinkage ridge estimators in semiparametric regression models. *J. Multivar. Anal.* **2015**, *136*, 56–74. [CrossRef]
24. Shen, X.; Alam, M.; Fikse, F.; Rönnegård, L. A novel generalized ridge regression method for quantitative genetics. *Genetics* **2013**, *193*, 1255–1268. [CrossRef]
25. Xing, L.; Pittman, J.J.; Inostroza, L.; Butler, T.J.; Munoz, P. Improving predictability of multisensor data with nonlinear statistical methodologies. *Crop Sci.* **2018**, *58*, 972–981. [CrossRef]
26. Huang, S.; Wu, Y.; Wang, Q.; Liu, J.; Han, Q.; Wang, J. Estimation of chlorophyll content in pepper leaves using spectral transmittance red-edge parameters. *Int. J. Agric. Biol. Eng.* **2022**, *15*, 85–90. [CrossRef]
27. Emmert-Streib, F.; Dehmer, M. High-dimensional LASSO-based computational regression models: Regularization, shrinkage, and selection. *Mach. Learn. Knowl. Extr.* **2019**, *1*, 359–383. [CrossRef]
28. Ruan, Y.; Huang, G.; Zhang, J.; Mai, S.; Gu, C.; Rong, X.; Huang, L.; Zeng, W.; Wang, Z. Risk analysis of noise-induced hearing loss of workers in the automobile manufacturing industries based on back-propagation neural network model: A cross-sectional study in Han Chinese population. *Br. Med. J. Open* **2024**, *14*, 079955. [CrossRef]
29. Ternès, N.; Rotolo, F.; Michiels, S. Empirical extensions of the lasso penalty to reduce the false discovery rate in high-dimensional Cox regression models. *Stat. Med.* **2016**, *35*, 2561–2573. [CrossRef]
30. Sancar, N.; Onakpojeruo, E.P.; Inan, D.; Uzun Ozsahin, D. Adaptive Elastic Net Based on Modified PSO for Variable Selection in Cox Model with High-Dimensional Data: A Comprehensive Simulation Study. *IEEE Access* **2023**, *11*, 127302–127316. [CrossRef]
31. Suzuki, T.; Tomioka, R.; Sugiyama, M. Sharp Convergence Rate and Support Consistency of Multiple Kernel Learning with Sparse and Dense Regularization. *arXiv* **2011**. [CrossRef]
32. Wei, B.; Liu, Y.; Liu, X.; Wei, C. Commentary: Analysis of Risk Factors for Painful Diabetic Peripheral Neuropathy and Construction of a Prediction Model Based on Lasso Regression. *Front. Endocrinol.* **2025**, *16*, 1519556. [CrossRef]
33. Hartemink, A.E.; Barrow, N.J. Soil pH-Nutrient Relationships: The Diagram. *Plant Soil* **2023**, *486*, 209–215. [CrossRef]
34. Malathi, P.; Babu, B.G.; Sellamuthu, K.M. Zinc Availability in Calcareous Soil as Influenced by Various Levels, Sources of Zn and Zn Solubilizing Bacteria. *Asian Res. J. Agric.* **2024**, *17*, 1123–1133. [CrossRef]
35. He, Y.; Zhang, J.; Li, C.; Zhang, L.; Fu, D. The Effect of Biochar Application Rates on Soil Fertility and Phyto-Availability of Heavy Metals is Dependent on Soil Type and pH. *Commun. Soil Sci. Plant Anal.* **2025**, *56*, 1291–1305. [CrossRef]
36. Sary, D.H.; Abd EL-Rahman, Z.M.; El-Sedfy, O.M.F. Impact of Irrigation Regimes and Organic Amendments on Soil Physical Properties, Nutrient Availability, and Productivity in Calcareous Soil. *Asian J. Soil Sci. Plant Nutr.* **2024**, *10*, 830–851. [CrossRef]
37. Tarolli, P.; Luo, J.; Park, E.; Barcaccia, G.; Masin, R. Soil Salinization in Agriculture: Mitigation and Adaptation Strategies Combining Nature-Based Solutions and Bioengineering. *iScience* **2024**, *27*, 108830. [CrossRef]
38. Safdar, H.; Amin, A.; Shafiq, Y.; Ali, A.; Yasin, R.; Shoukat, A.; Ul Hussan, M.; Sarwar, M.I. A Review: Impact of Salinity on Plant Growth. *Nat. Sci.* **2019**, *17*, 34–40. [CrossRef]
39. Carlos, F.S.; Marcolin, É.; Kunde, R.J.; Weinert, C.; Pasa, E.H.; Schäffer, N.; de Sousa, R.O.; dos Reis, R.B.; Andrezza, R.; de Oliveira, J.R.; et al. Long-Term No-Tillage Increases Soil Organic Matter and Cation Exchange Capacity, but Reduces P and K, in Irrigated Rice. *Soil Use Manag.* **2025**, *41*, e70016. [CrossRef]
40. Van Eynde, E.; Fendrich, A.N.; Ballabio, C.; Panagos, P. Spatial Assessment of Topsoil Zinc Concentrations in Europe. *Sci. Total Environ.* **2023**, *892*, 164512. [CrossRef]

41. Chen, J.; Liu, R.; Jian, Y.; Ma, T. Spatial Distribution and Factors Influencing the Various Forms of Iron in Alluvial–Lacustrine Clayey Aquitard. *Water* **2023**, *15*, 3934. [[CrossRef](#)]
42. Sun, G.; Liu, H.; Cui, D.; Chai, C. Spatial Heterogeneity of Soil Nutrients in Yili River Valley. *PeerJ* **2022**, *10*, e13311. [[CrossRef](#)] [[PubMed](#)]
43. Noulas, C.; Tziouvalekas, M.; Karyotis, T. Zinc in Soils, Water and Food Crops. *J. Trace Elem. Med. Biol.* **2018**, *49*, 252–260. [[CrossRef](#)] [[PubMed](#)]
44. Singh, A. Soil Salinity: A Global Threat to Sustainable Development. *Soil Use Manag.* **2022**, *38*, 39–67. [[CrossRef](#)]
45. Busoms, S.; Almira-Casellas, M.; Barceló, J.; Poschenrieder, C. Plant Tolerance to Alkaline Salinity: A Gordian Knot to Untie. In *Plant Stress Tolerance*; CRC Press: Boca Raton, FL, USA, 2025; pp. 37–69. [[CrossRef](#)]
46. Yang, S.; Xu, Y.; Tang, Z.; Jin, S.; Yang, S. The Impact of Alkaline Stress on Plant Growth and Its Alkaline Resistance Mechanisms. *Int. J. Mol. Sci.* **2024**, *25*, 13719. [[CrossRef](#)]
47. Hoffland, E.; Kuyper, T.W.; Comans, R.N.; Creamer, R.E. Eco-Functionality of Organic Matter in Soils. *Plant Soil* **2020**, *455*, 1–22. [[CrossRef](#)]
48. Kabelka, D.; Konvalina, P.; Kopecký, M.; Klenotová, E.; Šíma, J. Assessment of Soil Organic Matter and Its Microbial Role in Selected Locations in the South Bohemia Region (Czech Republic). *Land* **2025**, *14*, 183. [[CrossRef](#)]
49. Laik, R.; Awad Eltahira, E.B.; Pramanick, B.; Nidhi, N.; Singh, S.K.; van Es, H.M. Enhancing Soil Health in Rice Cultivation: Optimized Zn Application and Crop Residue Management in Calcareous Soils. *Sustainability* **2025**, *17*, 489. [[CrossRef](#)]
50. Alloway, B.J. Micronutrients and Crop Production: An Introduction. In *Micronutrient Deficiencies in Global Crop Production*; Springer: Dordrecht, The Netherlands, 2008; pp. 1–39. [[CrossRef](#)]
51. Singh, P.; Saini, S.P. Micronutrients Availability in Soil–Plant System as Influenced by Long-Term Integrated Nutrient Management under Rice–Wheat Cropping. *J. Plant Nutr.* **2022**, *45*, 457–470. [[CrossRef](#)]
52. Mossa, A.W.; Gashu, D.; Broadley, M.R.; Dunham, S.J.; McGrath, S.P.; Bailey, E.H.; Young, S.D. The effect of soil properties on zinc lability and solubility in soils of Ethiopia—an isotopic dilution study. *Soil* **2021**, *7*, 255–268. [[CrossRef](#)]
53. Munns, R.; Gilliham, M. Salinity tolerance of crops—What is the cost? *New Phytol.* **2015**, *208*, 668–673. [[CrossRef](#)] [[PubMed](#)]
54. Voltr, V.; Menšík, L.; Hlisnikovský, L.; Hruška, M.; Pokorný, E.; Pospíšilová, L. The soil organic matter in connection with soil properties and soil inputs. *Agronomy* **2021**, *11*, 779. [[CrossRef](#)]
55. Bünemann, E.K.; Bongiorno, G.; Bai, Z.; Creamer, R.E.; De Deyn, G.; de Goede, R.; Fleskens, L.; Geissen, V.; Kuyper, T.W.; Mäder, P.; et al. Soil quality—A critical review. *Soil Biol. Biochem.* **2018**, *120*, 105–125. [[CrossRef](#)]
56. Majumder, S.; Shankar, T.; Maitra, S.; Kumar, A.; Gudade, B.A.; Sagar, L.; Sairam, M.; Das, S.; Dash, S. Effect of Nutrient Omission Plot Technique Based Nutrient Management in Rabi Rice (*Oryza sativa*) on Crop Productivity, Nutrient Uptake and Soil Health. *Indian J. Agron.* **2024**, *69*, 357–363. [[CrossRef](#)]
57. Gawande, Y.D.; Hadole, S.S.; Mankar, E.P. Effect of Different Sources and Levels of Zinc on the Nutrient Content, Uptake and Fertility Status of Wheat. *Int. J. Plant Soil Sci.* **2025**, *37*, 288–309. [[CrossRef](#)]
58. Kumar, S.; Kumar, S.; Mohapatra, T. Interaction Between Macro- and Micro-Nutrients in Plants. *Front. Plant Sci.* **2021**, *12*, 665583. [[CrossRef](#)]
59. Thapa, S.; Bhandari, A.; Ghimire, R.; Xue, Q.; Kidwaro, F.; Ghatrehsamani, S.; Maharjan, B.; Goodwin, M. Managing Micronutrients for Improving Soil Fertility, Health, and Soybean Yield. *Sustainability* **2021**, *13*, 11766. [[CrossRef](#)]
60. Feng, B.; Ma, J.; Liu, Y.; Wang, L.; Zhang, X.; Zhang, Y.; Zhao, J.; He, W.; Chen, Y.; Weng, L. Application of Machine Learning Approaches to Predict Ammonium Nitrogen Transport in Different Soil Types and Evaluate the Contribution of Control Factors. *Ecotoxicol. Environ. Saf.* **2024**, *284*, 116867. [[CrossRef](#)] [[PubMed](#)]
61. Dhaliwal, S.S.; Naresh, R.K.; Mandal, A.; Walia, M.K.; Gupta, R.K.; Singh, R.; Dhaliwal, M.K. Effect of manures and fertilizers on soil physical properties, build-up of macro and micronutrients and uptake in soil under different cropping systems: A review. *J. Plant Nutr.* **2019**, *42*, 2873–2900. [[CrossRef](#)]
62. Kaur, T.; Sehgal, S.K.; Singh, S.; Sharma, S.; Dhaliwal, S.S.; Sharma, V. Assessment of seasonal variability in soil nutrients and its impact on soil quality under different land use systems of lower Shiwalik foothills of Himalaya, India. *Sustainability* **2021**, *13*, 1398. [[CrossRef](#)]
63. Athokpam, H.; Wani, S.H.; Kamei, D.; Athokpam, H.S.; Nongmaithem, J.; Kumar, D.; Singh, Y.K.; Naorem, B.S.; Devi, T.R.; Devi, L. Soil macro-and micro-nutrient status of Senapati district, Manipur (India). *Afr. J. Agric. Res.* **2013**, *8*, 4932–4936. [[CrossRef](#)]
64. Swaminathan, B.; Palani, S.; Vairavasundaram, S. Fertility Level Prediction in Precision Agriculture Based on an Ensemble Classifier Model. *Int. J. Sustain. Agric. Manag. Inform.* **2021**, *7*, 270. [[CrossRef](#)]
65. Birol, M.; Günal, H. Field Scale Variability in Soil Properties and Silage Corn Yield. *Soil Stud.* **2022**, *11*, 27–34. [[CrossRef](#)]
66. Kujawska, J.; Kulisz, M.; Cel, W.; Kwiatkowski, C.A.; Harasim, E.; Bandura, L. Evaluating and Predicting CO₂ Flux from Agricultural Soils Treated with Organic Amendments: A Comparative Study of ANN and ElasticNet Models. *J. Soils Sediments* **2025**, *25*, 864–882. [[CrossRef](#)]

67. Zhao, Z.; Luo, S.; Zhao, X.-X.; Zhang, J.; Li, S.; Luo, Y.; Dai, J. A Novel Interpolation Method for Soil Parameters Combining RBF Neural Network and IDW in the Pearl River Delta. *Agronomy* **2024**, *14*, 2469. [[CrossRef](#)]
68. Tekin, A.B.; Gunal, H.; Sindir, K.; Balci, Y. Spatial structure of available micronutrient contents and their relationships with other soil characteristics and corn yield. *Fresenius Environ. Bull.* **2011**, *20*, 783–792.
69. Vasu, D.; Sahu, N.; Tiwary, P.; Chandran, P. Modelling the spatial variability of soil micronutrients for site specific nutrient management in a semi-arid tropical environment. *Model. Earth Syst. Environ.* **2021**, *7*, 1797–1812. [[CrossRef](#)]
70. Dharumarajan, S.; Lalitha, M.; Niranjana, K.V.; Hegde, R. Evaluation of digital soil mapping approach for predicting soil fertility parameters—A case study from Karnataka Plateau, India. *Arab. J. Geosci.* **2022**, *15*, 386. [[CrossRef](#)]
71. Srinivasan, R.; Shashikumar, B.N.; Singh, S.K. Mapping of soil nutrient variability and delineating site-specific management zones using fuzzy clustering analysis in eastern coastal region, India. *J. Indian Soc. Remote Sens.* **2022**, *50*, 533–547. [[CrossRef](#)]
72. Mohammed, S.; Arshad, S.; Bashir, B.; Ata, B.; Al-Dalahmeh, M.; Alsalman, A.; Ali, H.; Alhennawi, S.; Kiwan, S.; Harsanyi, E. Evaluating machine learning performance in predicting sodium adsorption ratio for sustainable soil-water management in the eastern Mediterranean. *J. Environ. Manag.* **2024**, *370*, 122640. [[CrossRef](#)]
73. Mohamed, S.A.; Metwaly, M.M.; Metwalli, M.R.; AbdelRahman, M.A.; Badreldin, N. Integrating active and passive remote sensing data for mapping soil salinity using machine learning and feature selection approaches in arid regions. *Remote Sens.* **2023**, *15*, 1751. [[CrossRef](#)]
74. Mohammadifar, A.; Gholami, H.; Golzari, S. Assessment of the uncertainty and interpretability of deep learning models for mapping soil salinity using DeepQuantreg and game theory. *Sci. Rep.* **2022**, *12*, 15167. [[CrossRef](#)]
75. Bouslihim, Y.; Bouasria, A.; Jelloul, A.; Khiari, L.; Dahhani, S.; Mrabet, R.; Moussadek, R. Baseline high-resolution maps of soil nutrients in Morocco to support sustainable agriculture. *Sci. Data* **2025**, *12*, 1389. [[CrossRef](#)] [[PubMed](#)]
76. Khosravani, P.; Moosavi, A.A.; Baghernejad, M.; Kebonye, N.M.; Mousavi, S.R.; Scholten, T. Machine Learning Enhances Soil Aggregate Stability Mapping for Effective Land Management in a Semi-Arid Region. *Remote Sens.* **2024**, *16*, 4304. [[CrossRef](#)]
77. Ozturk, M.; Kilic, M.; Günal, H. Digital Mapping of Soil pH and Electrical Conductivity: A Comparative Analysis of Kriging and Machine Learning Approaches. *MAS J. Appl. Sci.* **2024**, *9*, 1168–1185. [[CrossRef](#)]
78. Abakay, O.; Kılıç, M.; Günal, H.; Kılıç, O.M. Tree-based algorithms for spatial modeling of soil particle distribution in arid and semi-arid region. *Environ. Monit. Assess.* **2024**, *196*, 264. [[CrossRef](#)]
79. Gunal, H.; Kılıç, M.; Altındal, M.; Gündoğan, R. Rapid spatial estimation of soil pH using machine learning under limited covariate conditions. *Levantine J. Appl. Sci.* **2021**, *1*, 30–37. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.