International Journal of Design & Nature and Ecodynamics
Vol. 21, No. 10, October, 2026, pp. 3055-3064
Journal homepage: http://iieta.org/journals/ijdne

Spatial Prediction of a Texture-Derived Physical Soil Quality Index in Tal Kaif, Iraq, Using Sentinel-1/2 and Terrain Data

Riyad H. Zubair[1*], Muntadhar Aidi Sharif[1], Abdelmalek Tuomi[2]

[1] Department of Surveying Engineering Techniques, Kirkuk Technical Engineering College, Northern Technical University, Kirkuk 36001, Iraq
[2] Lab-STICC, ENSTA Bretagne, Brest 29806, France

Corresponding Author Email: riyad.hazem@ntu.edu.iq

Copyright: ©2026 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/ijdne.211016

Received: 17 August 2026
Revised: 16 October 2026
Accepted: 23 October 2026
Available online: 31 October 2026

Keywords:
desertification vulnerability, machine learning, Pedotransfer Functions, Physical Soil Quality Index, Random Forest, Sentinel-1 synthetic aperture radar

ABSTRACT

Determining subsurface soil physical characteristics is vital for assessing moisture retention, yet conventional sampling is resource-demanding and optical sensors lack penetration. This study leverages remote sensing and machine learning (ML) to explore the spatial distribution and environmental drivers of a subsurface (30–40 cm) Physical Soil Quality Index (PSQI) in the semi-arid Tal Kaif district, Iraq. Using 70 field samples, textural fractions were determined to compute hydrological parameters via Pedotransfer Functions (PTFs). Principal Component Analysis (PCA) synthesized these into a unified PSQI. An ML framework—including Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), and Artificial Neural Network (ANN)—was established using Sentinel-1 synthetic aperture radar (SAR) backscatter, Sentinel-2 Soil Adjusted Vegetation Index (SAVI), and topographic variables. Five-fold out-of-fold (OOF) validation revealed that absolute point-based prediction remains challenging, with models exhibiting negative R^2. Although SVR provided the lowest validation error (Root Mean Square Error; RMSE = 21.48%) compared to RF (RMSE = 23.96%), the latter was strategically selected as an explanatory modeling framework. Rather than optimizing solely for absolute predictive accuracy, RF was utilized to extract variable importance and understand the non-linear spatial environmental drivers of soil quality. Sentinel-1 cross-polarization (VH) was the most influential driver (21% normalized importance), reflecting an indirect spatial covariation with subsurface soil texture. The resulting map demarcated potential zones of physical soil limitations, demonstrating strong spatial consistency with regional desertification models. This framework offers a scalable, technology-driven tool for screening relative soil vulnerability in data-scarce zones.

1. INTRODUCTION

Soil physical condition is one of the essential pillars of agricultural sustainability and ecosystem resilience in semi-arid regions suffering from water crises and desertification [1]. The infiltration rate, Field Capacity (FC), and Permanent Wilting Point (PWP) are important hydrological properties of soils. These properties are controlled by physical and structural properties of the soil matrix such as soil texture and bulk density. When these elements are combined, they are evaluated to give the Available Water Capacity (AWC), i.e., the maximum water that is available to the roots of plants. In semi-arid zones (i.e., Tal Kaif district in Northern Iraq), assessing these subsurface physical properties inside the active root zone (30–40 cm) helps with the most effective water management and alleviating the effects of prolonged periods of drought [2]. Assessment of soil physical quality at broad spatial scales has, traditionally, relied on extensive sampling in the field and laboratory and is known to be costly in time, labour, and money. Digital Soil Mapping (DSM) depends on remote sensing as a low-cost alternative to handle such spatial limitations. Despite this capability, there remains a major limitation: optical multispectral indices (e.g., Normalized Difference Vegetation Index (NDVI) or Soil Adjusted Vegetation Index (SAVI)) strictly measure the surface reflectance. They cannot drill through the soil surface to offer a reliable indication of what is happening in the subsurface root zone, often the most important layer for plant survival and structural degradation [3]. Synthetic aperture radar (SAR) microwave remote sensing may help mitigate the limitations of optical imagery. As per the study by Nxumalo et al. [4], SAR signals such as Sentinel-1, which operate at C-band or longer wavelengths, are able to penetrate sparse semi-arid vegetation canopies and are indirectly sensitive to surface moisture and roughness, acting as an environmental proxy for the shallow subsurface. The sensitivity of SAR backscatter intensity to the soil dielectric constant makes the latter a significant parameter. Furthermore, SAR backscatter is well-documented for estimating surface roughness and volumetric soil moisture [4, 5], and the soil's ability to retain moisture essentially depends on the physical texture and structural arrangement. Thus, temporal SAR backscatter may be a

powerful indirect proxy of the subsurface physical quality for its mapping [6]. The dielectric properties represented by SAR backscatter are not just descriptors; they are physical proxies for soil porosity and water-holding architecture, providing a much-needed link between surface reflectance and subsurface physical function [4, 6]. Advancements in machine learning (ML) algorithms, particularly ensemble methods like Random Forest (RF), have further revolutionized spatial modeling when they effectively accommodate complex non-linear relationships between environmental covariates and soil properties [7, 8]. As such, the study aims to develop a soil Physical Quality Index (PSQI) with established Pedotransfer Functions (PTFs) and Principal Component Analysis (PCA) for the extraction of the most influential physical indicators [9]. Furthermore, it will explore the relative spatial distribution of this soil index over the Tal Kaif district using multi-temporal Sentinel-1 SAR, Sentinel-2 optical data, and topographic indices under an ML framework [10]. In essence, this paper puts forward a spatially explicit, powerful tool for mapping relative patterns of physical soil degradation and its hydrological consequences that could assist in regional desertification monitoring and sustainable land management [2]. The overall methodological framework adopted in this study is illustrated in Figure 1.

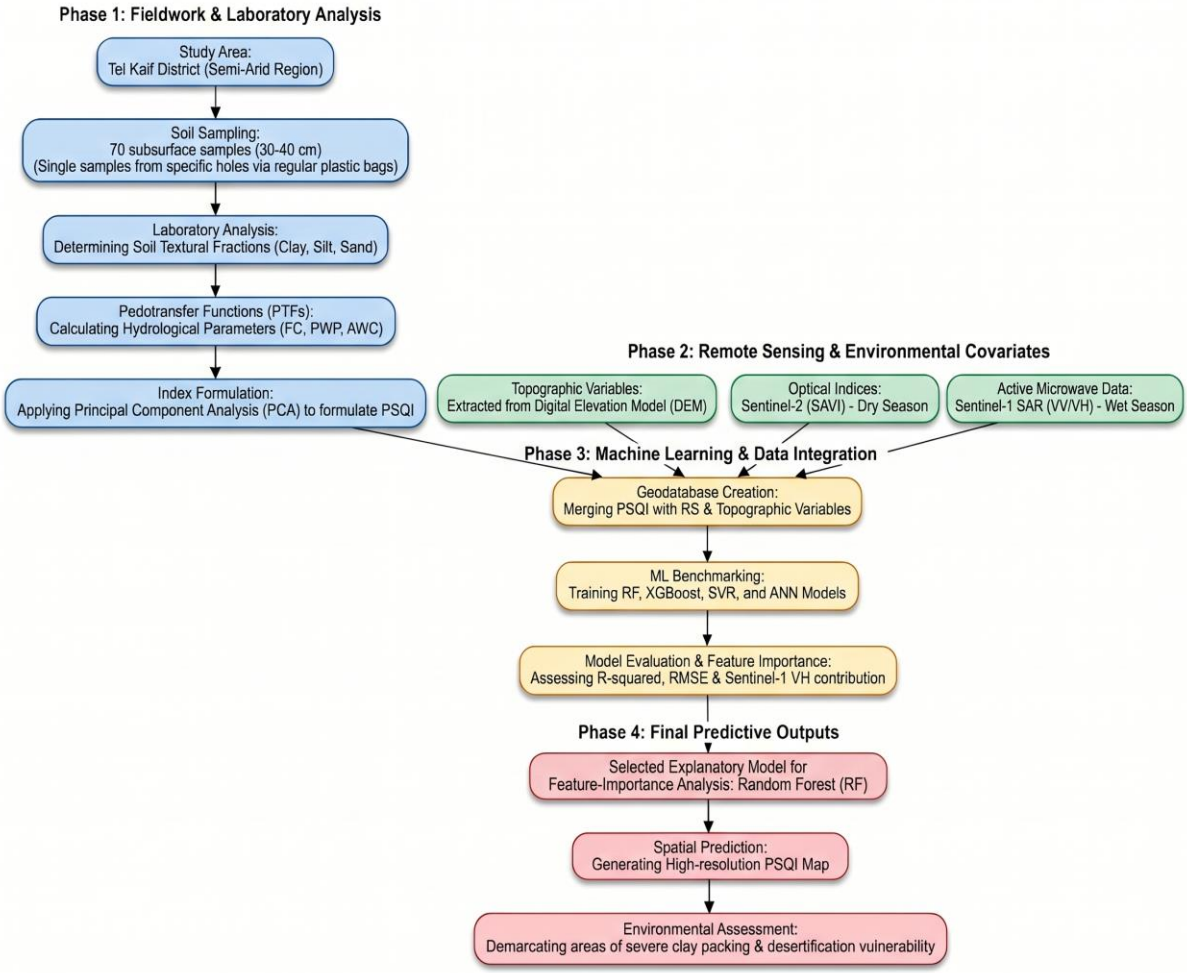

Figure 1. Flowchart of the methodological framework for exploratory spatial mapping of the Physical Soil Quality Index (PSQI)

2. MATERIALS AND METHODS

2.1 Field sampling and baseline data

The Tal Kaif district, located to the north of Mosul in Nineveh Governorate, Northern Iraq, was the research site of this study (Figure 2).

The Mediterranean climate of the research area (semi-arid), which is used in this study, exhibits hot, dry summers with a pedoclimatic xeric soil moisture regime and Thermic soil temperature regime, with an average annual rainfall of about 345 mm. In this district, the soils of cultivated lands are often affected by moisture deficit [2]. As a result, understanding how soil types vary in terms of physical properties in hilly terrain with diverse climatic conditions is important to ensure sustainable cultivation in the region. For the purpose of this study, a stratified random sampling design was adopted to reflect the spatial heterogeneity of soil properties. It allows for a fair representation of the different topography and environmental gradients present in the area [11]. The study established a total of 70 georeferenced soil sampling sites. Unlike conventional DSM and remote sensing studies that strictly target surface soil properties (usually 0–10 cm), mapping soil properties of deeper layers (10–100 cm) is also important but rarely done. The subsurface layer, which is 30–40 cm deep, was selected as the subject of the present study. Predicting and modelling physical attributes at this root-zone depth is very critical as this governs crop root establishment, retention of moisture within the soil profile, and agricultural response to droughts [6, 12]. The estimation of root-zone soil

properties represents an important modelling problem, requiring the use of ML frameworks integrated with active microwave (Sentinel-1 SAR) and multispectral (Sentinel-2) satellite observations, because standard optical sensors cannot physically penetrate the soil surface to these depths [3]. Fieldwork was conducted during the wet season, spanning a single week from February 19 to February 25, 2026. A total of 70 discrete soil samples were collected utilizing a systematic random sampling design, stratified across the main topographic gradients (e.g., elevation and slope classes) to ensure an unbiased yet representative spatial coverage of the landscape (Table 1).

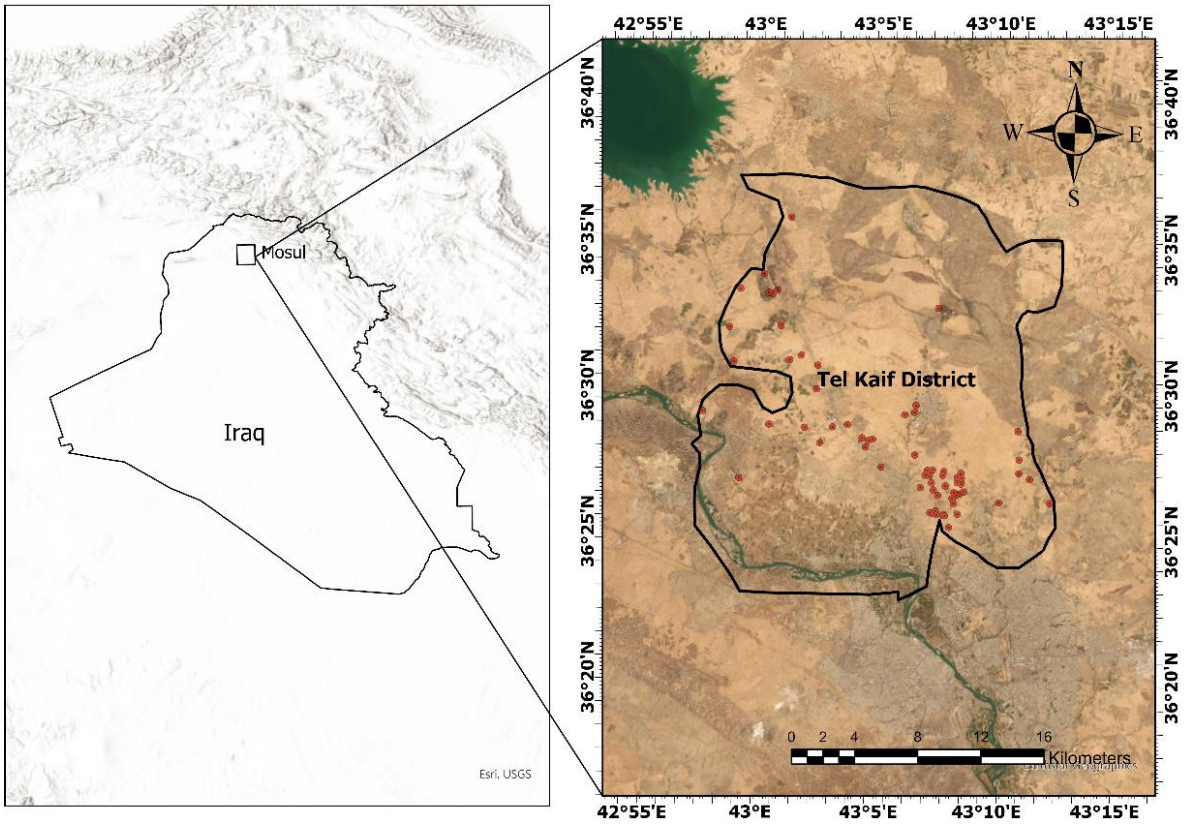

Figure 2. Map of the study area showing the location of the Tal Kaif district in Iraq and the spatial distribution of the 70 soil sampling points

Table 1. Stratified systematic random sampling distribution across the topographic gradients (elevation and slope) in Tal Kaif district (n = 70)

Elevation Class	Gentle Slope (0-2%)	Moderate Slope (2-5%)	Steep Slope (>5%)	Total
Low (<250 m)	0	0	1	1
Medium (250–280 m)	15	20	3	38
High (>280 m)	13	17	1	31
Total	28	37	5	70

At each predefined site, navigated via a handheld GPS receiver (±3 m accuracy), a single soil sample was extracted from the root-zone subsurface layer (30–40 cm depth) using standard soil augers and stored in plastic bags. Given the inherent spatial heterogeneity of soil, a point-to-pixel matching rule was established during data extraction using a 3 × 3 averaging window (approximately 30 × 30 m for Sentinel-2, aligned to the Sentinel-1 grid). This spatial aggregation mitigates GPS positioning errors and reduces SAR speckle noise. However, it should be acknowledged that a sample size of (n = 70) is relatively modest for complex ML models, which inherently limits the absolute spatial generalization across the broader landscape. When the soil samples reached the laboratory, they were immediately subjected to a gravimetric soil moisture content determination taking place under standard conditions. The standard oven-drying method was used for analysis at a temperature of 105 ℃ for 24 hours. Soil samples were air-dried at room temperature, mechanically crushed, and passed through a 2 mm sieve to remove coarse fragments and organic debris. Quantitative particle size distribution analysis was done using the standard hydrometer method to establish the sand, silt, and clay fractions of the soil matrix [13, 14]. These baseline laboratory measurements were employed to compute the Physical Soil Quality Index (PSQI) and fine-tune the predictive ML models.

2.2 Estimation of hydrological properties via Pedotransfer Functions

Accurate quantification of soil hydrological parameters, specifically FC, PWP, and AWC, is fundamental for assessing subsurface moisture retention and overall soil physical quality in semi-arid environments. However, direct in-situ or laboratory measurements of these hydraulic properties across expansive and spatially heterogeneous networks are notoriously resource-intensive, costly, and time-consuming [1].

To solve logistical limitations issues and also maintaining

scientific rigor, PTFs were used. PTFs are widely validated empirical models that translate readily available basic soil data, such as particle size distribution (texture fractions), into difficult-to-measure hydraulic characteristics, a method increasingly utilized in DSM frameworks [15]. The hydrological parameters, namely FC and PWP, were estimated using the globally established empirical PTFs developed by Cosby et al. [16]. These PTFs were originally calibrated using a large dataset of soils across the United States. This decision is methodologically justified for two primary reasons. First, these empirical equations rely on fundamental physical relationships between primary soil textural fractions (sand, clay) and moisture retention at specific matric potentials (33 kPa for FC, 1500 kPa for PWP), which operate independently of geographic boundaries. The textural range of the Tal Kaif samples falls within the broad operational capacity of these generalized models. Second, in the absence of high-resolution, locally calibrated PTFs for the deep subsurface layer (30–40 cm) in this specific semi-arid region, the Cosby models provide a standardized and robust baseline. Most importantly, the primary objective of this spatial modeling framework is to assess the relative spatial variation and vulnerability of the physical soil matrix across the landscape, rather than attempting to predict absolute volumetric water content at point precision. This relative spatial assessment significantly minimizes the impact of potential regional calibration biases. In the applied PTF equations, Sand and Clay are expressed as percentages (0–100), rather than mass fractions (0–1). The estimated FC and PWP represent volumetric water content, expressed in m^3/m^3. Accordingly, AWC, calculated as the difference between FC and PWP, is also expressed in m^3/m^3.

The formulas used are as follows:

$$\text{Field Capacity (FC)} = 0.3456 - (0.0028 \times \text{Sand \%}) + (0.0031 \times \text{Clay \%})$$

The PWP in terms of the sand percent and clay percent is given as:

$$\text{Permanent Wilting Point (PWP)} = 0.1080 - (0.0016 \times \text{Sand\%}) + (0.0044 \times \text{Clay\%})$$

Then, the AWC was calculated as the arithmetic difference between FC and PWP:

$$\text{Available Water Capacity (AWC)} = (\text{FC} - \text{PWP})$$

AWC is the amount of water available that the soil or soil subsurface matrix can hold against drainage by gravity reaching the wilting point. As the water within the root zones is the one that will dictate the resilience of the crop in semi-arid regions, AWC, together with FC and PWP, will be one of the key bases for the construction of the PSQI in later modeling stages [1, 17].

2.3 Derivation of the Physical Soil Quality Index

The PSQI was created to combine individual physical and hydrologic variables into a single measure of soil health. Because the soil properties we measured (textural fractions and derived hydrological parameters) are potentially redundant and quite high dimensional, we applied PCA to objectively weight each indicator [9, 17]. PCA efficiently reduces the data dimension and removes multicollinearity through the transformation of the initial correlated variables into a number of Principal Components (PCs) which are linearly uncorrelated [9, 18]. The PCA input matrix used in the study consisted of three textural fractions (sand, silt and clay) and three hydrological parameters (FC, PWP and AWC). The PSQI was calculated using only the first principal component (PC1), which explained the highest variance. Although FC, PWP, and AWC are mathematically derived from texture fractions, all six variables were intentionally retained in the PCA to holistically weight both the physical skeleton and the hydrological capacity of the soil. PC1 captured the dominant variance. Note that the PCA loadings presented represent the original eigenvector weights before the sign reversal was applied. Because the initial PCA loadings resulted in negative eigenvector weights for favorable hydrological parameters (such as AWC), the PC1 scores were systematically inverted. This mathematical inversion was necessary to ensure logical physical interpretability, where higher final scores consistently represent better soil structural matrix and moisture-retention capacity. The inverted scores were then normalized using min-max scaling to a standardized 0–100 scale. The underlying approach is consistent with various soil quality assessment frameworks [19, 20], which assign weight to each indicator based on eigenvector loading to the principal component. The PC1 scores extracted were standardized to linear scoring functions for spatial interpretation and mapping and normalized on a 0–100 basis [17]. The normalized PSQI formula is given by:

$$\text{PSQI} = \left(\frac{\text{PSQI}_{\text{observed}} - \text{PSQI}_{\text{min}}}{\text{PSQI}_{\text{max}} - \text{PSQI}_{\text{min}}}\right) \times 100 \quad (1)$$

Higher PSQI values denote better physical matrix with higher moisture retention capacity and aeration essential to alleviate drought stress in semi-arid Tal Kaif district environment [2, 19].

2.4 Spatial modeling framework: Machine learning and remote sensing

The advanced spatial modelling framework was implemented to predict the continuous spatial distribution of the PSQI of Tal Kaif using multi-source environmental and satellite data. The soil-landscape relationships in semi-arid environments are complex and non-linear by nature. Thus, ML algorithms can capture these relationships better than traditional geostatistical methods [12, 21]. The integration of bi-seasonal (two scenes) satellite data led to the commencement of predictive modeling. To represent the specific environmental conditions, a Sentinel-1 (SAR C-band) image was acquired on February 26, 2026, immediately following the field campaign, to capture the wet-season soil dielectric properties (Table 2). Conversely, a Sentinel-2 multispectral image acquired on August 30, 2025, was utilized to extract the SAVI during the dry season (bare soil conditions) to eliminate vegetation interference [3, 16]. The Sentinel-1 GRD product was preprocessed using SNAP software, which included precise orbit application, thermal noise removal, radiometric calibration to Sigma0, Lee Sigma speckle filtering, and Range-Doppler terrain correction using the SRTM 30 m Digital Elevation Model (DEM). These signify sub-surface dielectric moisture signatures, indicative of the soil's physical structure and porosity [22, 23]. The main algorithm used for prediction is the RF algorithm. RF is increasingly gaining popularity in DSM due to its efficiency

with high-dimensional, nonlinear environmental data and natural resistance to overfitting [7, 21]. The RF model was fitted to the correlation between the target (PCA-derived PSQI) and all the previously mentioned satellite indices and static topographic variables such as elevation, slope, and Topographic Wetness Index (TWI) for better prediction. DEM topography is essential in understanding the geomorphological controls on soil development and moisture redistribution, according to studies [23, 24].

Figure 3 illustrates the distribution of multi-source predictors across the Tal Kaif district, which is capable of high-resolution mapping of PSQI.

Table 2. Characteristics of the bi-seasonal satellite data utilized

Satellite / Sensor	Acquisition Date	Product Level	Spatial Resolution	Cloud Cover / Masking
Sentinel-2 (MSI)	August 30, 2025	L2A (BOA Reflectance)	10 m	0% (Clear sky, no masking required)
Sentinel-1 (SAR C-band)	February 26, 2026	GRD (Interferometric Wide)	10 m	N/A (Microwave)

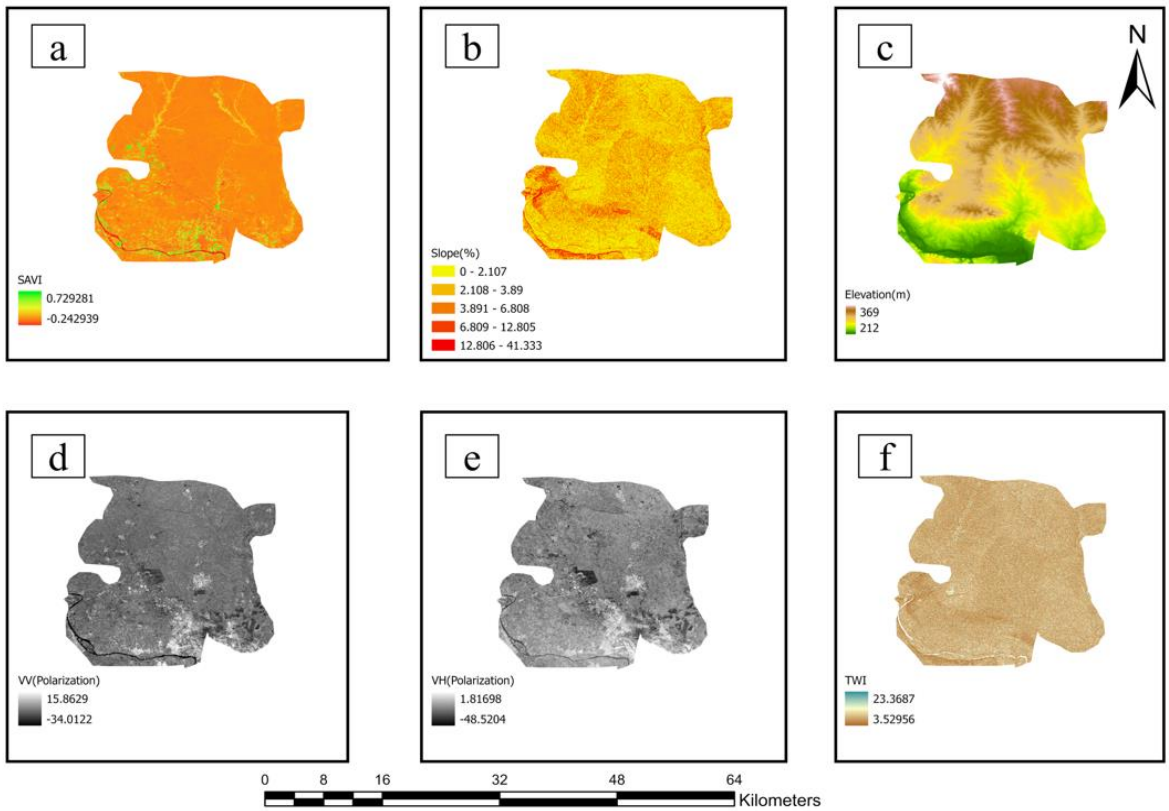

Figure 3. Spatial distribution of the environmental and remote sensing predictor variables used in the spatial modeling: (a) Soil Adjusted Vegetation Index (SAVI), (b) slope (%), (c) elevation (m), (d) co-polarization (VV) backscatter σ^0 (dB), (e) cross-polarization (VH) backscatter σ^0 (dB), and (f) Topographic Wetness Index (TWI)

2.5 Model evaluation

To rigorously evaluate the predictive performance of the ML algorithms, the Coefficient of Determination (R^2) and Root Mean Square Error (RMSE) were calculated using the following equations:

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

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (3)$$

In these equations, y_i represents the observed PSQI value for sample i, while $\hat{y}_i$ represents the corresponding predicted PSQI value obtained from the five-fold out-of-fold (OOF) validation. The symbol n denotes the total number of soil samples (n = 70), and $\bar{y}$ represents the mean of all 70 observed PSQI values. All evaluation metrics were calculated using the aggregated OOF predictions.

3. RESULTS AND DISCUSSION

3.1 Statistical correlations and model validation

The statistical testing of the formulated PSQI confirmed its validity before exploring the spatial correlations. As the hydrological parameters (FC, PWP, AWC) were extracted from the textural fractions, including all six variables in the PCA ensured that the resultant PSQI inherently weighted the hydrologically active fine fractions (clay and silt) besides soil moisture retention capacity. As indicated in Table 3, PC1 explained 69.37% of total variance (with an eigenvalue of 4.1624), confirming its strong representation of overall physical quality of soil.

Worked Calculation Example for PSQI (Sample 9 and Sample 61):

To clarify the derivation of the PSQI, the initial variables (sand, silt, clay, FC, PWP, and AWC) were scaled (Z-scores) and multiplied by the eigenvectors (loadings) of PC1, which captured exactly 69.37% of the total variance. Because the initial orientation of PC1 assigned negative weights to favorable attributes (e.g., AWC), and thus higher PC1 scores implied poorer soil quality, we multiplied PC1 scores by -1 so that higher values indicate better soil physical quality, and subsequently min-max normalized to a 0–100 scale. Below is the step-by-step calculation for two extreme samples: Sample 9 (heavy clay, very low PSQI):

Raw fractions: Sand = 0.00%, Silt = 31.79%, Clay = 68.21%

PTF derived: FC = 0.557, PWP = 0.408, AWC = 0.149

Standardized (Z-scores): Z_Sand = -1.080, Z_Silt = -2.423, Z_Clay = 3.083, Z_FC = 2.714, Z_PWP = 2.973, Z_AWC = -2.525

PC1 calculation: Sum of (Z-score × inverted loading) = (-1.080 × 0.197) + (-2.423 × 0.370) + (3.083 × -0.494) + (2.714 × -0.446) + (2.973 × -0.482) + (-2.525 × 0.387)

Raw PC1 Score = -6.251

Normalized PSQI (0–100 scale) = 0.00

Sample 61 (silty texture, very high PSQI):

Raw fractions: Sand = 14.70%, Silt = 62.37%, Clay = 22.93%

PTF derived: FC = 0.376, PWP = 0.185, AWC = 0.190

Standardized (Z-scores): Z_Sand = 1.122, Z_Silt = 0.956, Z_Clay = -1.708, Z_FC = -1.743, Z_PWP = -1.766, Z_AWC = 1.036

Raw PC1 Score = 3.448

Normalized PSQI (0–100 scale) = 100.00

An analysis of correlation indicated complex non-linear relations between PSQI and the environmental variables, as their correlation coefficient ranged from -0.16 to 0.13 (Figure 4).

Table 3. Principal Component Analysis (PCA) loadings and variance explained for the Physical Soil Quality Index (PSQI)

Variables / Metrics	Original Loading (PC1)	Loading Used after Inversion
Sand (%)	-0.1974	0.1974
Silt (%)	-0.3699	0.3699
Clay (%)	0.4936	-0.4936
Field Capacity (FC)	0.4457	-0.4457
Permanent Wilting Point (PWP)	0.4816	-0.4816
Available Water Capacity (AWC)	-0.3873	0.3873
Eigenvalue	4.1624	
Variance Explained (%)	69.37%	

In the absence of a strong linear relationship, we utilized the RF algorithm, which is a non-linear ensemble learning method, to enable discovery of complex spatial patterns that the linear models cannot capture. It was aimed to rigorously validate the predictive robustness capability of RF model the validation plot was constructed to compare the PSQI values of the 70 sampling points which were observed versus the predicted values of RF model.

The validation plot (Figure 5) illustrates the relationship between the observed and predicted PSQI values based on the strict OOF predictions. Rather than clustering tightly along the 1:1 ideal line, the predictions exhibit a narrower variance compared to the observed field data. This dispersion aligns with the negative R^2 values, confirming that while the model captures broad spatial trends, exact point-by-point prediction of subsurface properties remains highly challenging.

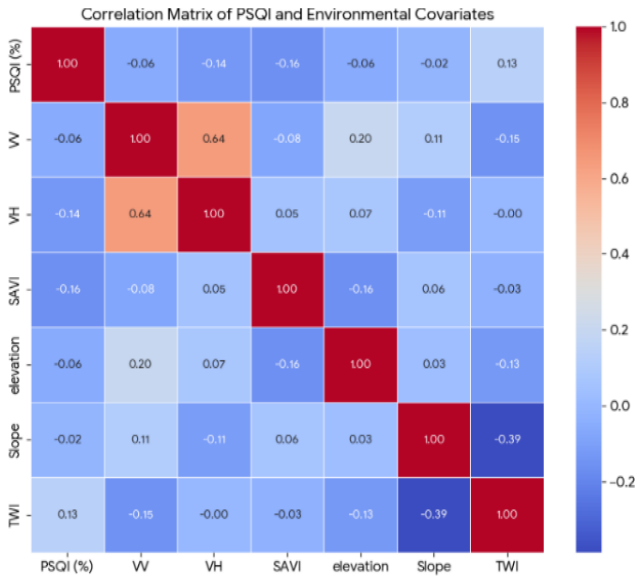

Figure 4. Correlation matrix demonstrating the non-linear relationships between the Physical Quality Index (PSQI) and the environmental covariates used in the Random Forest (RF) modeling framework

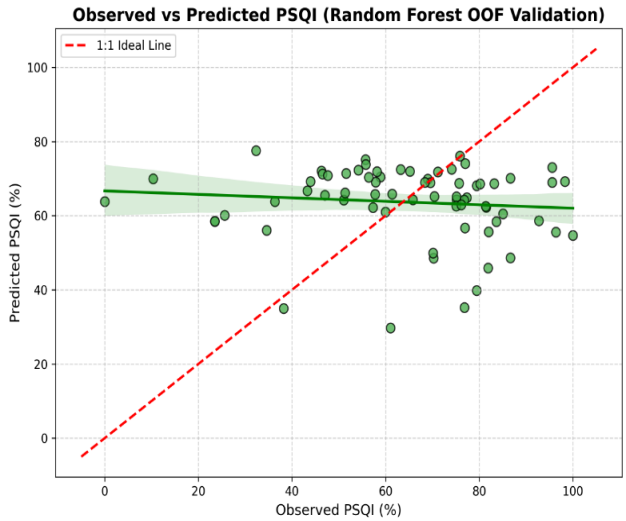

Figure 5. Observed vs. predicted Physical Quality Index (PSQI) values showing the predictive performance of the Random Forest (RF) model (based on strict out-of-fold (OOF) validation) relative to the 1:1 ideal line

3.2 Predictive performance and benchmarking

To fully assess the predictive limits of the spatial framework applied, a strict five-fold OOF benchmarking was performed for RF, Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), and Artificial Neural Network (ANN). As indicated in Table 4 and Figure 6, all pure ML models exhibited negative R^2 values, highlighting the inherent "spatial blindness" and the extreme complexity of predicting subsurface (30–40 cm) properties solely from surface-level remote sensing covariates [10].

Table 4. Predictive performance benchmarking of the evaluated machine learning (ML) algorithms for predicting the Physical Soil Quality Index (PSQI)

Algorithm	R^2	RMSE (%)
Support Vector Regression (SVR)	-0.06	21.48
Random Forest (RF)	-0.32	23.96
Artificial Neural Network (ANN)	-0.59	26.35
Extreme Gradient Boosting (XGBoost)	-0.82	28.21

Note: R^2: Coefficient of Determination; RMSE: Root Mean Square Error.

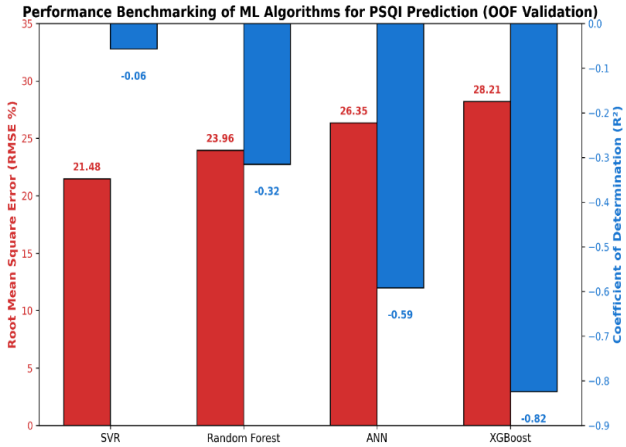

Figure 6. Performance benchmarking of the evaluated machine learning (ML) algorithms based on strict out-of-fold (OOF) validation, illustrating the comparative Coefficient of Determination (R^2) and Root Mean Square Error (RMSE) metrics

Although SVR achieved the lowest OOF error (RMSE = 21.48%) and slightly better predictive accuracy, the RF algorithm (RMSE = 23.96%) was strategically selected to generate the final spatial maps. This decision was driven by the study's primary objective: to interpret environmental drivers rather than solely maximizing point-based predictive accuracy. Unlike SVR, which acts as a "black box," RF inherently calculates variable importance, allowing for an explanatory interpretation of the non-linear spatial trends. The analysis of variable importance is carried out after selection of the model. It indicates the most important environmental drivers of physical soil quality in the Tal Kaif district (Table 5).

Table 5. Relative importance of multi-source environmental covariates in predicting Physical Soil Quality Index (PSQI)

Rank	Importance Weight (%)	Source / Type	Variable
1	21.0	Sentinel-1 (Cross-polarization)	VH
2	17.0	Sentinel-1 (Co-polarization)	VV
2	17.0	DEM (Topography)	Elevation
2	17.0	Topographic Wetness Index	TWI
5	14.0	Sentinel-2 (Optical)	SAVI
5	14.0	DEM (Topography)	Slope

Note: VH: Cross-polarization; VV: Co-polarization; DEM: Digital Elevation Model; SAVI: Soil Adjusted Vegetation Index.

The backscatter of Sentinel-1 SAR cross-polarization (VH) was found to be the most significant predictor, accounting for 21% of the normalized impurity-based RF feature importance. It is crucial to note that C-band SAR (such as Sentinel-1) penetrates only the top few centimeters of the soil surface. Therefore, the high predictive importance of the VH backscatter does not imply direct measurement of the 30–40 cm depth. Instead, it reflects a strong indirect spatial covariation; surface roughness, micro-topography, and topsoil moisture patterns captured by the radar act as highly effective environmental proxies for the deeper subsurface soil texture and PSQI distribution. It should be acknowledged that standard impurity-based feature importance in RF can be biased when predictor variables are highly correlated. Therefore, the ranking of VH and topography reflects an indirect environmental proxy rather than an isolated physical mechanism. Elevation, TWI, and SAR co-polarization (VV) followed equally at 17% each. The foremost contribution of topography factors reflects their significant role in controlling moisture redistribution, lateral runoff, and soil formation in this area [24, 25]. In addition, the significant sensitivity of the VH channel to volume scattering provides a strong verification of our main hypothesis that active microwave backscatter can penetrate the sparse semi-arid vegetation cover and interact with the dielectric properties of the soil; these are influenced by the soil's moisture retention capacity (AWC) and subsurface physical matrix [3, 22].

3.3 Spatial distribution of Physical Quality Index and hydrological implications

The spatial prediction surface produced (Figure 7) reveals significant variation in physical soil quality across the Tal Kaif district. The final point-based prediction map shows values ranging from 26.79 to 82.11. This narrow range, compared to the theoretical limits (0–100) of the calculated samples, is a known result of the RF algorithm's tendency to approximate extreme predictions toward the mean. The contiguous map was classified into five PSQI quality classes (very low, low, medium, high, very high) using the natural Jenks interval classification method (Table 6).

Table 6. Areal extent and percentage distribution of the predicted Physical Soil Quality Index (PSQI) classes in Tal Kaif district

PSQI Quality Class	Area (km^2)	Percentage (%)
Very Low	45.56	9.76%
Low	106.69	22.85%
Moderate	118.29	25.33%
High	99.04	21.21%
Very High	97.41	20.85%

It is important to note the inherent uncertainty in the boundaries between these discrete categories due to the interpolation limits of the model. Based on the multivariate connections established by the PCA framework, the highest PSQI values, approaching the 100% threshold (e.g., Sample No. 61), are largely concentrated in zones that feature silty textures, especially silty clay loams. Soils that fall under this textural class have an ideal water regime, having enough capillary porosity that maximises their AWC and can also ensure aeration of the root zone. Consequently, agricultural soils situated within these high-PSQI zones demonstrate superior structural resilience against drought stress, which is

of paramount importance for sustainable dryland farming in semi-arid environments [1, 9].

Conversely, areas exhibiting the lowest PSQI values (approaching 0%, e.g., sampling site No. 9) strongly correlate with heavy clay extremes where the clay content exceeds 68%. In accordance with established pedological principles, although fine-textured clayey matrices possess a high FC, their exceptionally high specific surface area generates intense matric suction that consequently leads to a highly elevated PWP [1]. Consequently, the moisture held within these soil systems is locked in tightly in micropores and is unavailable for crop root uptake, leading to very low AWC [17]. These areas, which are predominantly composed of heavy clay, are highly susceptible to physical degradation. The map highlights these areas with low PSQI values as areas with potential structural limitations and insufficient moisture in the root zone, making it a rough regional map of spatial susceptibility rather than a direct measure of mechanical compaction.

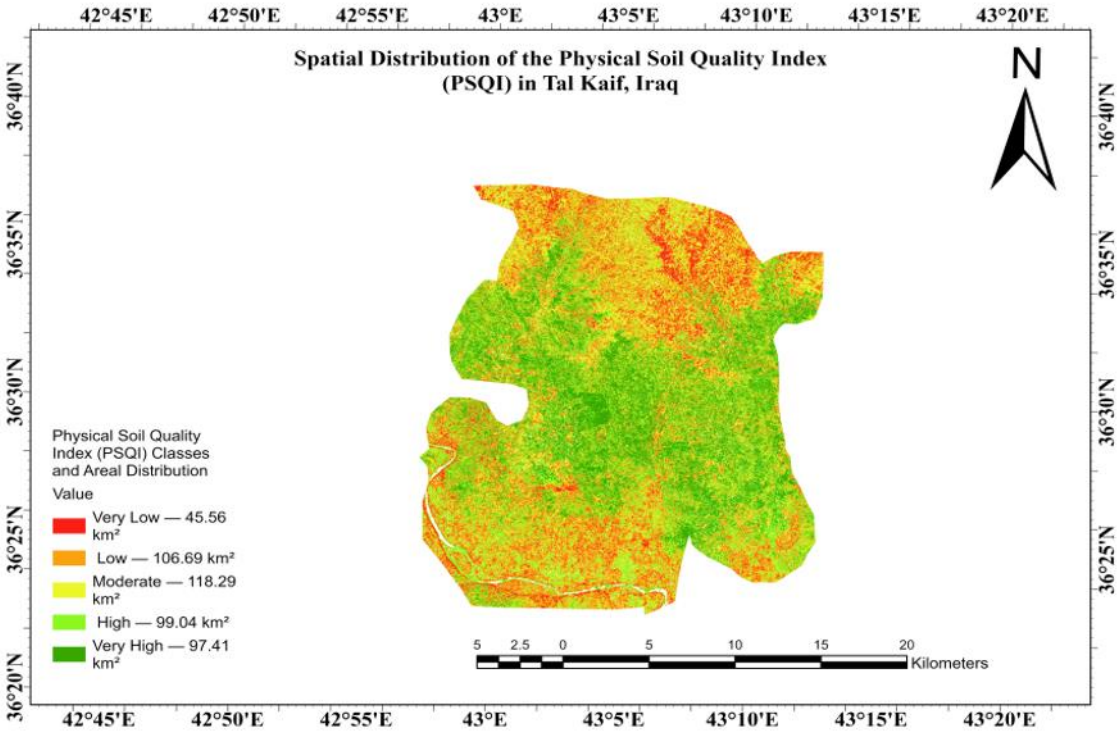

Figure 7. Spatial distribution map of the predicted Physical Soil Quality Index (PSQI) in the Tal Kaif district using the Random Forest (RF) algorithm

3.4 Spatial comparison with regional desertification indicators (MEDALUS)

The generated PSQI map was cross-validated using spatial patterns of physical degradation derived from independent desertification assessments within the region. According to the predictive ML framework, the spatial zones identified as having a low PSQI—characterized by substandard root-zone moisture retention and deteriorated physical structure—demonstrate a remarkable spatial congruence with the areas classified as "Highly Environmentally Sensitive to Desertification" (Critical C3 type) under recent applications of the MEDALUS model in the Tal Kaif district [2].

Physically, such structural and textural degradation restricts root proliferation, exacerbates surface runoff, and severely depletes plant-available water, thereby accelerating the positive feedback loops of land degradation [26]. Integrating multi-temporal, SAR-derived covariates with ML algorithms to predict subsurface PSQI provides a robust, physically based proxy for mapping future desertification vulnerability. By adopting this approach, the logistical bottlenecks and prohibitive costs associated with extensive field campaigns and lab-based soil characterization can be effectively bypassed, offering a highly scalable, technology-driven early warning system for regional land degradation monitoring [18].

4. CONCLUSIONS

The integration of multi-sensor remote sensing with ML algorithms has provided valuable insights into mapping subsurface PSQI in semi-arid areas. Although strict OOF validation revealed that absolute point-based prediction of subsurface properties solely from surface observations remains highly challenging, the RF algorithm was used to interpret general spatial trends and environmental factors leading to physical soil degradation. As the models yielded negative R^2 values, indicating poorer performance than the simple average baseline, the RF model in this study is used exclusively to determine the significance of explanatory variables and to explore relative spatial patterns; it is not considered reliable for absolute point-based spatial prediction. Among the limitations of this study is the lack of spatially split cross-validation, which would have provided a more accurate estimate of spatial autocorrelation and mapping accuracy.

The results reveal that active microwave remote sensing, especially Sentinel-1 VH, is a critical indirect proxy, showing greater normalized feature importance than optical vegetation indices for assessing subsurface structural and hydrological properties. It is imperative to note that VH does not directly penetrate to a depth of 30–40 cm; rather, its importance stems from strong spatial covariance with surface roughness and topsoil moisture, which act as proxies for deeper textural

variations. Consequently, the high predictive significance of the "VH" backscatter serves strictly as an exploratory environmental indicator, enabling the identification of spatial areas that may be subject to structural constraints or contain heavy clay soils. The generated PSQI spatial map highlights areas of critical vulnerability in the Tal Kaif district, which spatially aligns with regional desertification patterns. The spatial patterns predicted by the ML framework highlight potential zones of physical degradation based on textural and pedotransfer estimates, rather than direct in-situ measurements of soil compaction or structural damage. Ultimately, this approach provides decision-makers and agricultural planners with a scalable, cost-effective tool to monitor general trends of soil physical degradation, significantly reducing the reliance on labor-intensive traditional field sampling. Future research should focus on integrating geostatistical techniques (such as Regression Kriging) or deeper profile data to improve the absolute predictive accuracy of these models across varying climatic conditions.

REFERENCES

[1] Di Prima, S., Fernandes, G., Burguet, M., et al. (2024). Assessing soil physical quality in a layered agricultural soil: A comprehensive approach using infiltration experiments and time-lapse ground-penetrating radar surveys. Applied Sciences, 14(20): 9268. https://doi.org/10.3390/app14209268

[2] Alhamdany, S.A., Alnaser, Y.H. (2024). Assessing environmental sensitivity areas (ESA) using MEDALUS model in Telkaif district of northern Iraq. Asian Journal of Soil Science and Plant Nutrition, 10(3): 384-393. https://doi.org/10.9734/ajsspn/2024/v10i3349

[3] Bousbih, S., Zribi, M., Pelletier, C., et al. (2019). Soil texture estimation using radar and optical data from Sentinel-1 and Sentinel-2. Remote Sensing, 11(13): 1520. https://doi.org/10.3390/rs11131520

[4] Nxumalo, G.S., Ramabulana, T.S., Dlamini, Z., János, T., Kiss, N.É., Nagy, A. (2026). AI-driven integration of Sentinel-1 SAR for high-resolution soil water content estimation to enhance precision irrigation in smallholder maize systems, Vhembe District. Water, 18(4): 499. https://doi.org/10.3390/w18040499

[5] Baghdadi, N., El Hajj, M., Choker, M., et al. (2018). Potential of Sentinel-1 images for estimating the soil roughness over bare agricultural soils. Water, 10(2): 131. https://doi.org/10.3390/w10020131

[6] Ali, C.A., Shareef, M.A., Al Nuaimy, Q.A.M. (2024). Significance and distribution of soil parameters in the Laylan area, Kirkuk city, Northern Iraq using interpolation methods based on GIS technique. Scientific Research Journal of Multidisciplinary, 4(2): 1-13. https://doi.org/10.47310/srjmd.2024.v04i02.01

[7] Breiman, L. (2001). Random forests. Machine Learning, 45(1): 5-32. https://doi.org/10.1023/A:1010933404324

[8] Louzada, R.O., Bergier, I., Bolfe, É.L., Barbedo, J.G.A. (2025). Integrating GIS and remote sensing for soil attributes mapping in degraded pastures of the Brazilian Cerrado. Soil Advances, 3: 100044. https://doi.org/10.1016/j.soilad.2025.100044

[9] Shokr, M.S., Abdellatif, M.A., El Baroudy, A.A., et al. (2021). Development of a spatial model for soil quality assessment under arid and semi-arid conditions. Sustainability, 13(5): 2893. https://doi.org/10.3390/su13052893

[10] Mirzaeitalarposhti, R., Shafizadeh-Moghadam, H., Taghizadeh-Mehrjardi, R., Demyan, M.S. (2022). Digital soil texture mapping and spatial transferability of machine learning models using sentinel-1, sentinel-2, and terrain-derived covariates. Remote Sensing, 14(23): 5909. https://doi.org/10.3390/rs14235909

[11] Win, K., Sato, T., Hiroshima, T. (2025). Applicability of remote sensing and machine learning for predicting bulk soil electrical conductivity under different forest types in central Japan. Soil Advances, 3: 100045. https://doi.org/10.1016/j.soilad.2025.100045

[12] Arshad, S., Kazmi, J.H., Harsányi, E., et al. (2025). Predictive modeling of soil salinity integrating remote sensing and soil variables: An ensembled deep learning approach. Energy Nexus, 17: 100374. https://doi.org/10.1016/j.nexus.2025.100374

[13] Gee, G.W., Bauder, J.W. (1986). Particle-size analysis. In Methods of Soil Analysis: Part 1 Physical and Mineralogical Methods, 2nd ed., pp. 383-411. https://doi.org/10.2136/sssabookser5.1.2ed.c15

[14] Khudhair, D.S. (2024). Evaluating soil texture and elemental composition in Wasit governorate, Iraq: A remote sensing approach. Journal of Al-Muthanna for Agricultural Sciences, 11(1). https://doi.org/10.52113/mjas04/11.1/12

[15] Odebiri, O., Mutanga, O., Odindi, J., Peerbhay, K., Dovey, S., Ismail, R. (2020). Estimating soil organic carbon stocks under commercial forestry using topo-climate variables in KwaZulu-Natal, South Africa. South African Journal of Science, 116(3-4): 6339. https://doi.org/10.17159/sajs.2020/6339

[16] Cosby, B.J., Hornberger, G.M., Clapp, R.B., Ginn, T. (1984). A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. Water Resources Research, 20(6): 682-690. https://doi.org/10.1029/WR020i006p00682

[17] Reis, D.A., Lima, C.L.R.D., Bamberg, A.L. (2019). Developing a soil physical quality index (SPQI) for lowlands under different deployment times of no-tillage. Scientia Agricola, 76(2): 157-164. https://doi.org/10.1590/1678-992X-2017-0196

[18] Fathizad, H., Ardakani, M.A.H., Heung, B., et al. (2020). Spatio-temporal dynamic of soil quality in the central Iranian desert modeled with machine learning and digital soil assessment techniques. Ecological Indicators, 118: 106736. https://doi.org/10.1016/j.ecolind.2020.106736

[19] Damiba, W.A.F., Gathenya, J.M., Raude, J.M., Home, P.G. (2024). Soil quality index (SQI) for evaluating the sustainability status of Kakia-Esamburmbur catchment under three different land use types in Narok County, Kenya. Heliyon, 10(5): e25611. https://doi.org/10.1016/j.heliyon.2024.e25611

[20] Vasu, D., Singh, S.K., Ray, S.K., et al. (2016). Soil quality index (SQI) as a tool to evaluate crop productivity in semi-arid deccan plateau, India. Geoderma, 282: 70-79. https://doi.org/10.1016/j.geoderma.2016.07.010

[21] Shafizadeh-Moghadam, H., Minaei, F., Talebi-Khiavi, H., Xu, T., Homaee, M. (2022). Synergetic use of multi-temporal sentinel-1, sentinel-2, NDVI, and topographic factors for estimating soil organic carbon. Catena, 212: 106077. https://doi.org/10.1016/j.catena.2022.106077

[22] Mohammed, P.A., Al Nuaimy, Q.A.M., Shareef, M.A.

(2024). Analyzing soil physical properties in Kirkuk City, Iraq, utilizing a geographic information system–based inverse distance-weighted technique. Ecological Engineering & Environmental Technology, 25(11): 118-133. https://doi.org/10.12912/27197050/192346

[23] Odebiri, O., Mutanga, O., Odindi, J., et al. (2024). Mapping sub-surface distribution of soil organic carbon stocks in South Africa's arid and semi-arid landscapes: Implications for land management and climate change mitigation. Geoderma Regional, 37: e00817. https://doi.org/10.1016/j.geodrs.2024.e00817

[24] Sulaiman, S.M., Mezaal, M.R., Al-Attar, A.A.A. (2024). Evaluation of machine learning models for predicting soil texture using Sentinel-1A SAR and topographic information. Ecological Engineering & Environmental Technology, 25(11): 200-217. https://doi.org/10.12912/27197050/192397

[25] Azizi, K., Garosi, Y., Ayoubi, S., Tajik, S. (2023). Integration of sentinel-1/2 and topographic attributes to predict the spatial distribution of soil texture fractions in some agricultural soils of western Iran. Soil and Tillage Research, 229: 105681. https://doi.org/10.1016/j.still.2023.105681

[26] D'Odorico, P., Bhattachan, A., Davis, K.F., Ravi, S., Runyan, C.W. (2013). Global desertification: Drivers and feedbacks. Advances in Water Resources, 51: 326-344. https://doi.org/10.1016/j.advwatres.2012.01.013