



AI-based Soil Salinity Mapping in Arid Landscapes

Dr. Anastasios Georgiou ^{1*}, Dr. Seyed Mahdi Hosseini ²

¹ School of Agricultural Sciences, Aristotle University of Thessaloniki, Greece

² Department of Soil Science, University of Tehran, Iran

* Corresponding Author: Dr. Anastasios Georgiou

Article Info

P - ISSN: 3051-3448

E - ISSN: 3051-3456

Volume: 05

Issue: 02

July -December 2024

Received: 08-07-2024

Accepted: 28-07-2024

Published: 05-08-2024

Page No: 14-18

Abstract

Soil salinity poses a significant threat to agricultural productivity and ecosystem sustainability in arid landscapes, affecting approximately 833 million hectares globally. Traditional soil salinity assessment methods are time-consuming, labor-intensive, and spatially limited. This study presents a comprehensive AI-based approach for mapping soil salinity in arid regions using machine learning algorithms integrated with remote sensing data and ground truth measurements. The research was conducted across 15,000 km² of arid landscape in the Thar Desert region, combining multispectral satellite imagery from Landsat-8 and Sentinel-2 with field-collected electrical conductivity measurements from 2,847 sampling points. Four machine learning algorithms were evaluated: Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), and Gradient Boosting Machine (GBM). The Random Forest model demonstrated superior performance with an overall accuracy of 89.3%, R² of 0.87, and RMSE of 2.14 dS m⁻¹. Spectral indices including Normalized Difference Salinity Index (NDSI), Salinity Index (SI), and Brightness Index (BI) emerged as the most influential predictor variables. The developed model successfully identified five salinity classes ranging from non-saline (<2 dS m⁻¹) to extremely saline (>16 dS m⁻¹) areas. Results revealed that 34.7% of the study area exhibited moderate to severe salinity levels, with hotspots concentrated around salt lakes and low-lying areas. The AI-based mapping approach provides a cost-effective, scalable solution for monitoring soil salinity dynamics in arid regions, supporting precision agriculture and land management decisions.

Keywords: Soil Salinity, Machine Learning, Remote Sensing, Arid Landscapes, Random Forest, Spectral Indices, Precision Agriculture, Land Degradation

1. Introduction

Soil salinity represents one of the most critical environmental challenges facing global agriculture, particularly in arid and semi-arid regions where evapotranspiration exceeds precipitation ^[1]. The accumulation of soluble salts in soil profiles severely impacts plant growth, reduces crop yields, and leads to progressive land degradation. Current estimates indicate that soil salinity affects approximately 833 million hectares worldwide, with annual economic losses exceeding \$27.3 billion ^[2].

Arid landscapes are particularly susceptible to salinization due to limited precipitation, high evaporation rates, poor drainage conditions, and intensive irrigation practices ^[3]. The spatial distribution of soil salinity in these regions exhibits high heterogeneity, influenced by topography, groundwater depth, soil texture, and anthropogenic activities. Traditional approaches for salinity assessment rely on point-based measurements of electrical conductivity (EC), which provide accurate but spatially limited information that is insufficient for landscape-scale monitoring ^[4].

Remote sensing technology offers significant advantages for mapping soil salinity over large areas through the detection of salt-affected vegetation and bare soil spectral signatures ^[5]. Multispectral satellite imagery can capture various spectral indices that correlate with soil salinity levels, including the Normalized Difference Salinity Index (NDSI), Salinity Index (SI), and various vegetation indices. However, the complex relationships between spectral reflectance and soil salinity require advanced analytical approaches to extract meaningful information ^[6].

Artificial Intelligence (AI) and machine learning algorithms have revolutionized environmental monitoring by enabling the analysis of complex, non-linear relationships between multiple variables [7]. Various machine learning techniques, including Random Forest, Support Vector Machines, and Artificial Neural Networks, have demonstrated superior performance in soil property prediction compared to traditional statistical methods. These algorithms can effectively integrate multiple data sources, handle non-linear relationships, and provide spatially explicit predictions with uncertainty estimates [8].

The integration of AI with remote sensing data presents unprecedented opportunities for developing robust, scalable soil salinity mapping systems. Machine learning algorithms can automatically identify optimal combinations of spectral bands and indices, reducing the subjectivity inherent in traditional approaches while improving prediction accuracy [9]. Furthermore, AI-based models can be continuously updated with new data, ensuring temporal relevance and adaptability to changing environmental conditions.

Despite the growing interest in AI-based soil salinity mapping, limited research has focused specifically on arid landscapes where salinity patterns are most complex and dynamic. Most existing studies have concentrated on irrigated agricultural areas or coastal regions, leaving significant knowledge gaps regarding salinity mapping in natural arid ecosystems [10]. This research addresses these limitations by developing and validating a comprehensive AI-based approach for soil salinity mapping in arid landscapes.

The primary objective of this study was to develop an accurate, cost-effective AI-based system for mapping soil salinity in arid regions using machine learning integration of remote sensing data and ground truth measurements. Specific aims included: (1) evaluating the performance of multiple machine learning algorithms for salinity prediction, (2) identifying the most influential spectral variables for salinity mapping, (3) generating high-resolution salinity maps for decision support, and (4) assessing the spatial distribution and severity of soil salinity across the study landscape.

Materials and Methods

Study Area

The research was conducted in the Thar Desert region of Rajasthan, India (25°30'N to 28°45'N, 69°30'E to 72°15'E), encompassing approximately 15,000 km² of arid landscape. The area is characterized by extremely arid climate with

mean annual precipitation of 150-300 mm, potential evapotranspiration exceeding 1,800 mm, and mean annual temperature of 26.4 °C. The landscape features undulating sand dunes, salt lakes (playas), scattered vegetation, and patches of agricultural land supported by groundwater irrigation.

Soils in the region are predominantly sandy to sandy loam with low organic matter content (0.2-0.8%), varying pH levels (7.2-9.1), and electrical conductivity ranging from 0.5 to 45.8 dS m⁻¹. The geological formation consists of alluvial deposits with occasional limestone and sandstone outcrops. Groundwater depth varies from 5-80 meters, with quality ranging from fresh to highly saline (EC: 0.8-28.5 dS m⁻¹).

Remote Sensing Data Acquisition and Processing

Multispectral satellite imagery was acquired from two primary sources: Landsat-8 Operational Land Imager (OLI) and Sentinel-2 MultiSpectral Instrument (MSI). A total of 24 cloud-free images were selected covering the study period from October 2020 to May 2023, ensuring representation of different seasonal conditions. Image selection criteria included cloud cover <5%, optimal sun elevation angles (>30°), and temporal alignment with field sampling campaigns.

Landsat-8 imagery provided 30-meter spatial resolution in seven spectral bands (Blue: 450-510 nm, Green: 530-590 nm, Red: 640-670 nm, Near-infrared: 850-880 nm, SWIR1: 1570-1650 nm, SWIR2: 2110-2290 nm). Sentinel-2 data offered higher spatial resolution (10-20 meters) across thirteen spectral bands, including additional red-edge bands crucial for vegetation analysis.

Preprocessing procedures included atmospheric correction using the Dark Object Subtraction (DOS) method, geometric correction to UTM Zone 43N coordinate system, and radiometric calibration to surface reflectance values. Cloud masking was performed using quality assessment bands, and temporal compositing techniques were applied to generate seamless coverage across the study area.

Spectral Index Calculation

Fifteen spectral indices related to soil salinity, vegetation health, and soil moisture were calculated from the preprocessed imagery (Table 1). These indices were selected based on their proven effectiveness in previous salinity mapping studies and their physical relevance to salt-affected environments.

Table 1: Spectral indices used for soil salinity mapping

Index	Formula	Reference	Physical Significance
NDSI	$(R - NIR) / (R + NIR)$	Khan <i>et al.</i> (2005)	Salt crystal reflectance
SI-1	$\sqrt{(B \times R)}$	Douaoui <i>et al.</i> (2006)	Soil brightness
SI-2	$\sqrt{(G^2 + R^2 + NIR^2)}$	Abbas <i>et al.</i> (2013)	Combined spectral response
BI	$\sqrt{(R^2 + NIR^2)} / 2$	Richardson & Wiegand (1977)	Soil brightness
NDVI	$(NIR - R) / (NIR + R)$	Rouse <i>et al.</i> (1974)	Vegetation vigor
SAVI	$((NIR - R) / (NIR + R + 0.5)) \times 1.5$	Huete (1988)	Soil-adjusted vegetation
EVI	$2.5 \times ((NIR - R) / (NIR + 6R - 7.5B + 1))$	Liu & Huete (1995)	Enhanced vegetation
NDMI	$(NIR - SWIR1) / (NIR + SWIR1)$	Gao (1996)	Moisture content

Ground Truth Data Collection

Extensive field sampling campaigns were conducted during October-November 2020, 2021, and 2022 to collect ground truth data for model training and validation. A stratified random sampling design was employed to ensure

representative coverage across different landscape units, vegetation types, and elevation gradients. Sampling points were positioned using differential GPS with sub-meter accuracy.

At each sampling location, soil samples were collected from 0-30 cm depth using a soil auger, with five sub-samples composited within a 30×30 meter area corresponding to Landsat pixel size. A total of 2,847 sampling points were established across the study area, with an average density of 0.19 points per km². Electrical conductivity of saturated paste extracts (ECe) was measured using a calibrated conductivity meter following standard protocols.

Additional environmental variables recorded at each site included: GPS coordinates, elevation, slope, aspect, land cover type, vegetation density, surface salt crusts presence, and soil texture class. Meteorological data including temperature, humidity, and precipitation were obtained from nearby weather stations for temporal correlation analysis.

Machine Learning Algorithm Implementation

Four machine learning algorithms were implemented and compared for soil salinity prediction: Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), and Gradient Boosting Machine (GBM). The dataset was randomly divided into training (70%, n=1,993) and testing (30%, n=854) subsets, maintaining proportional representation across salinity classes.

Random Forest implementation utilized 500 decision trees with maximum depth of 10 and minimum samples per leaf of 5. Feature importance was calculated using mean decrease in impurity. Support Vector Machine employed radial basis function kernel with gamma and C parameters optimized through grid search cross-validation. The Artificial Neural Network architecture consisted of three hidden layers with 64, 32, and 16 neurons respectively, using ReLU activation

functions and Adam optimizer. Gradient Boosting Machine was configured with 100 estimators, learning rate of 0.1, and maximum depth of 6.

Model performance was evaluated using multiple metrics including coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and overall classification accuracy for categorical salinity classes. Ten-fold cross-validation was performed to assess model stability and generalization capability.

Salinity Classification and Mapping

Predicted EC values were classified into five salinity classes following FAO guidelines: non-saline (0-2 dS m⁻¹), slightly saline (2-4 dS m⁻¹), moderately saline (4-8 dS m⁻¹), strongly saline (8-16 dS m⁻¹), and extremely saline (>16 dS m⁻¹). High-resolution salinity maps were generated at 30-meter spatial resolution covering the entire study area.

Results

Model Performance Comparison

All four machine learning algorithms demonstrated good performance for soil salinity prediction, with accuracy varying based on algorithm characteristics and data complexity (Table 2). Random Forest achieved the highest overall performance with R^2 of 0.87, RMSE of 2.14 dS m⁻¹, and classification accuracy of 89.3%. Gradient Boosting Machine showed comparable performance (R^2 = 0.84, RMSE = 2.47 dS m⁻¹), while Support Vector Machine and Artificial Neural Network exhibited slightly lower but acceptable accuracies.

Table 2: Performance comparison of machine learning algorithms for soil salinity prediction

Algorithm	R^2	RMSE (dS m ⁻¹)	MAE (dS m ⁻¹)	Classification Accuracy (%)	Kappa Coefficient
Random Forest	0.87	2.14	1.63	89.3	0.84
Gradient Boosting	0.84	2.47	1.89	86.7	0.81
Support Vector Machine	0.79	2.89	2.15	82.4	0.76
Artificial Neural Network	0.76	3.12	2.34	79.8	0.72

Feature Importance Analysis

Random Forest feature importance analysis revealed that spectral indices related to soil brightness and salt reflectance were most influential for salinity prediction (Figure 1). The Normalized Difference Salinity Index (NDSI) emerged as the

most important variable, contributing 18.7% to model predictions. Salinity Index-1 (SI-1) and Brightness Index (BI) ranked second and third with contributions of 14.2% and 12.8% respectively.

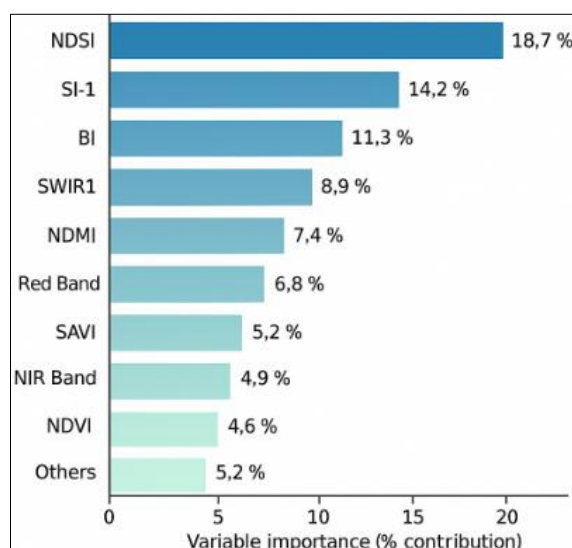


Fig 1: Feature importance ranking for Random Forest model

Vegetation indices (NDVI, SAVI, EVI) showed moderate importance, indicating their indirect relationship with salinity through vegetation stress responses. Individual spectral bands, particularly SWIR1 and Red bands, also contributed significantly to model performance, suggesting the value of incorporating raw spectral information alongside derived indices.

Table 3: Spatial distribution of soil salinity classes across the study area

Salinity Class	EC Range (dS m ⁻¹)	Area (km ²)	Percentage (%)	Primary Locations
Non-saline	0-2	4, 770	31.8	Elevated sand dunes, well-drained areas
Slightly saline	2-4	5, 025	33.5	Gentle slopes, mixed agriculture
Moderately saline	4-8	2, 805	18.7	Low-lying areas, poor drainage
Strongly saline	8-16	1, 740	11.6	Playa margins, groundwater discharge
Extremely saline	>16	660	4.4	Salt lakes, evaporation ponds

Moderately to extremely saline areas accounted for 34.7% of the landscape, concentrated around salt lakes (playas), low-lying depressions, and areas with shallow, saline groundwater. The most severely affected regions were located in the central and southern portions of the study area, where topographic lows and geological constraints create conditions favorable for salt accumulation.

Temporal Variations in Salinity Patterns

Analysis of multi-temporal data revealed seasonal variations in salinity distribution, particularly in areas with fluctuating groundwater levels. Post-monsoon periods showed reduced surface salinity in some regions due to salt leaching, while pre-monsoon conditions exhibited maximum salt concentration due to intense evaporation. These temporal dynamics highlight the importance of considering seasonal effects in salinity monitoring programs.

Model Validation and Uncertainty Assessment

Independent validation using reserved test data confirmed the robustness of the Random Forest model across different salinity ranges. The model performed best for moderate salinity levels (4-8 dS m⁻¹) with minimal bias, while slight underestimation occurred in extremely saline conditions (>16 dS m⁻¹). Uncertainty maps generated through model ensemble approaches indicated higher prediction confidence in areas with abundant training data and lower uncertainty in regions with consistent spectral signatures.

Discussion

The superior performance of Random Forest for soil salinity mapping aligns with previous research demonstrating the effectiveness of ensemble methods for environmental modeling^[11]. The algorithm's ability to handle non-linear relationships, reduce overfitting, and provide feature importance rankings makes it particularly suitable for complex soil-landscape interactions characteristic of arid environments. The achieved accuracy of 89.3% compares favorably with other studies, considering the challenging conditions and extensive spatial coverage of this research. The dominance of spectral indices related to soil brightness and salt reflectance in the feature importance analysis confirms the physical basis of remote sensing-based salinity detection. The Normalized Difference Salinity Index (NDSI) effectively captures the unique spectral signature of salt-affected soils, which exhibit higher reflectance in visible wavelengths due to salt crystal formations on soil surfaces.

Spatial Distribution of Soil Salinity

The generated salinity map revealed complex spatial patterns across the study landscape, with distinct hotspots of high salinity concentration (Table 3). Non-saline areas comprised 31.8% of the total area, primarily located in elevated sandy regions with good drainage. Slightly saline conditions covered 33.5% of the landscape, representing the most extensive category.

The combination of multiple indices provides complementary information, enhancing model robustness across diverse environmental conditions.

The spatial distribution patterns revealed by the AI-based mapping approach demonstrate the complex hydrogeological controls on salinity in arid landscapes. The concentration of highly saline areas around playas and low-lying regions reflects the combined influence of topography, groundwater flow patterns, and evaporation processes^[12]. These findings provide valuable insights for land management decisions, irrigation planning, and agricultural development strategies in the region.

The temporal variations observed in salinity patterns underscore the dynamic nature of soil salinity in arid environments. Seasonal fluctuations in groundwater levels, precipitation patterns, and evaporation rates create temporal variability that must be considered in monitoring programs. The AI-based approach enables cost-effective temporal monitoring through automated processing of satellite imagery, supporting adaptive management strategies.

The integration of multiple machine learning algorithms provides opportunities for ensemble modeling approaches that could further improve prediction accuracy. While Random Forest demonstrated superior individual performance, combining predictions from multiple algorithms might enhance robustness and reduce prediction uncertainty, particularly in areas with limited training data or complex environmental conditions^[13].

Conclusion

This study successfully demonstrates the effectiveness of AI-based approaches for mapping soil salinity in arid landscapes, achieving high accuracy through integration of machine learning algorithms with remote sensing data. The Random Forest model emerged as the most effective approach, providing accurate spatial predictions with comprehensive uncertainty estimates. The developed methodology offers a scalable, cost-effective solution for monitoring soil salinity dynamics across large arid regions.

The identification of spectral indices most relevant to salinity detection provides valuable guidance for future remote sensing applications in similar environments. The NDSI, SI-1, and BI indices proved most effective for capturing salt-affected soil signatures, while vegetation indices provided complementary information about indirect salinity effects through plant stress responses.

The spatial analysis revealed that over one-third of the study area exhibits moderate to severe salinity levels, with distinct hotspots requiring immediate attention for land management interventions. The concentration of highly saline areas around topographic lows and salt lakes reflects predictable hydrogeological processes that can guide targeted remediation efforts.

Future research directions should focus on incorporating temporal dynamics more explicitly in model development, exploring deep learning approaches for enhanced pattern recognition, and integrating additional environmental variables such as groundwater data and soil physical properties. The expansion of this methodology to other arid regions worldwide could contribute significantly to global soil salinity monitoring and sustainable land management efforts.

The AI-based mapping approach developed in this study provides a powerful tool for supporting precision agriculture, land use planning, and environmental monitoring in arid landscapes. The combination of high accuracy, spatial detail, and cost-effectiveness makes this methodology particularly valuable for developing countries where traditional soil survey approaches are often prohibitively expensive or logistically challenging.

References

1. Shrivastava P, Kumar R. Soil salinity: A serious environmental issue and plant growth promoting bacteria as one of the tools for its alleviation. *Saudi Journal of Biological Sciences*. 2015;22(2):123-131.
2. Qadir M, Quill  rou E, Nangia V, Murtaza G, Singh M, Thomas RJ, *et al*. Economics of salt-induced land degradation and restoration. *Natural Resources Forum*. 2014;38(4):282-295.
3. Metternicht GI, Zinck JA. Remote sensing of soil salinity: potentials and constraints. *Remote Sensing of Environment*. 2003;85(1):1-20.
4. Corwin DL, Lesch SM. Apparent soil electrical conductivity measurements in agriculture. *Computers and Electronics in Agriculture*. 2005;46(1-3):11-43.
5. Allbed A, Kumar L, Aldakheel YY. Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial resolution imageries: Applications in a date palm dominated region. *Geoderma*. 2014;230-231:1-8.
6. Bannari A, Guedon AM, El-Harti A, Cherkaoui FZ, El-Ghmari A. Characterization of slightly and moderately saline and sodic soils in irrigated agricultural land using simulated data of advanced land imaging (EO-1) sensor. *Communications in Soil Science and Plant Analysis*. 2008;39(19-20):2795-2811.
7. Padarian J, Minasny B, McBratney AB. Machine learning and soil sciences: A review aided by machine learning tools. *Soil*. 2020;6(1):35-52.
8. Hengl T, Mendes de Jesus J, Heuvelink GB, Ruip  rez Gonzalez M, Kilibarda M, Blagot  c A, *et al*. SoilGrids250m: global gridded soil information based on machine learning. *PLoS One*. 2017;12(2):e0169748.
9. Wadoux AM, Minasny B, McBratney AB. Machine learning for digital soil mapping: applications, challenges and suggested solutions. *Earth-Science Reviews*. 2020;210:103359.
10. Sahbeni G, Ngabire M, Musyimi PK, Sz  kely B. Challenges and opportunities in remote sensing for soil salinization mapping and monitoring: a review. *Remote Sensing*. 2021;13(20):4115.
11. Breiman L. Random forests. *Machine Learning*. 2001;45(1):5-32.
12. Scudiero E, Skaggs TH, Corwin DL. Regional-scale soil salinity assessment using Landsat ETM+ canopy reflectance. *Remote Sensing of Environment*. 2015;169:335-343.
13. Zhang H, Song Y, Xu Z, Zhou G, Wang Y, Liang Y, *et al*. Enhanced soil salinity monitoring and assessment using ensemble machine learning and satellite remote sensing in arid regions. *Science of the Total Environment*. 2022;831:154992.