



OPEN

A longitudinal analysis of soil salinity changes using remotely sensed imageries

Soraya Bandak^{1✉}, Seyed Alireza Movahedi-Naeini¹, Saeed Mehri^{2✉} & Aynaz Lotfata³

Soil salinization threatens agricultural productivity, leading to desertification and land degradation. Given the challenges of conducting labor-intensive and expensive field studies and laboratory analyses on a large scale, recent efforts have focused on leveraging remote sensing techniques to study soil salinity. This study assesses the importance of soil salinity indices derived from remotely sensed imagery. Indices derived from Landsat 8 (L8) and Sentinel 2 (S2) imagery are used in Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Decision Tree (DT), and Support Vector Machine (SVR) are associated with the electrical (EC) conductivity of 280 soil samples across 24,000 hectares in Northeast Iran. The results indicated that the DT is the best-performing method (RMSE = 12.25, MAE = 2.15, $R^2 = 0.85$ using L8 data and RMSE = 10.9, MAE = 2.12, and $R^2 = 0.86$ using S2 data). Also, the results showed that Multi-resolution Valley Bottom Flatness (MrVBF), moisture index, Topographic Wetness Index (TWI), and Topographic Position Indicator (TPI) are the most important salinity indices. Subsequently, a time series analysis indicated a reduction in salinity and sodium levels in regions with installed drainage networks, underscoring the effectiveness of the drainage system. These findings can assist decision-making about land use and conservation efforts, particularly in regions with high soil salinity.

Keywords Soil salinization, Remote sensing, Predictive modeling, Decision tree

Soil salinization profoundly affects soil productivity, nutrient availability, and plant physiology and biochemistry, especially in arid and semi-arid regions. Intensive irrigation in these areas brings saline groundwater to the surface, leading to overflow zones where evaporite minerals precipitate^{1–3}. The salinity adversely affects crop water and fertilizer uptake and soil fertility enhancement, impacting more than 20% of the world's irrigated land⁴.

Soil Electrical Conductivity (EC) is a key indicator of soil salinity measurements, strongly correlating with salinity levels⁵. Mapping EC can enhance understanding of soil genesis processes in arid and semi-arid soils and aid agricultural management. However, due to its complex spatial variation, soil EC estimation is more challenging than other soil properties (e.g., soil organic carbon), necessitating the development of more reliable methods.

Remotely sensed images, including Landsat 8 (L8) and Sentinel 2 (S2), are extensively used for salinity analysis, offering spectral diversity and cost-effectiveness in soil salinity mapping^{6–10}. Within this context, Erkin et al.¹¹ used L8 for temporal soil salinity analysis in Kashgar. They found that due to the continuous increase in inland reclamation and insufficient drainage, salinized arable land has steadily increased, and the average salinity of the cropland reached higher than 5.1 g per kilogram of soil. Taghizadeh-Mehrjardi et al.¹² identified that the S2 satellite image is a suitable and cost-effective data source for soil salinity assessment due to its short revisit interval, multiple spectral bands, and high spatial resolution. Wu et al.¹⁶ used Landsat 5 TM and ALOS L-band radar to create a soil salinity map in the Mussaib region of Central Mesopotamia, finding that the Random Forest (RF) algorithm outperformed Support Vector Regression (SVR). Wang et al.¹³ compared L8 and S2 images in the Ebinur Lake wetland using the cubist model, finding that the S2 image is superior for salinity estimation. Wang et al.¹⁴ have shown that the Cubist model with L8 image is superior for salinity estimation compared to using S2 with that model in Ebinur Lake Wetland National Nature Reserve, China. Wang et al.¹⁵ demonstrate the effectiveness of S2 images in distinguishing between saline and non-saline areas using the RF algorithm. Their

¹Department of Soil Sciences, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran. ²Department of Geospatial Information Systems, Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran. ³Department of Pathology, Microbiology, and Immunology, School Of Veterinary Medicine, University of California, Davis, USA. ✉email: soraya.bandak@gmail.com; sa.mehri20@gmail.com

study also highlights the capability to monitor changes in soil salinity levels between dry and wet seasons using remotely sensed images by generating region-specific maps.

Ma et al.¹⁶ digitally mapped salinity distribution in the Werigan-Kuqa oasis, analyzing the evolution characteristics and driving factors using a machine learning approach and field data. The eXtreme Gradient Boosting (XGBoost) model significantly enhances prediction accuracy and salinity mapping, illustrating spatial and temporal changes over 25 years. Ge et al.¹⁷ introduced a hybrid machine learning framework utilizing S2 image and environmental determinants, achieving a notable improvement in soil salinity mapping accuracy by 5–8%. These studies show that the L8 satellite image is more effective for monitoring soil salinity than the S2 image^{13,18–20}. Also, researchers have employed RF²¹, XGBoost²², Decision Tree (DT)²³, and SVR^{24,25} as predictive tools in their investigations, and these models consistently demonstrated robust and reliable performance in the context of soil salinity prediction.

This study's main aim is to analyze soil salinity variations in a drainage area using L8 and S2 imagery data alongside machine learning methods. Additionally, it compares salinity levels before and after installing a regional drainage network.

Materials and methods

Figure 1 provides a summary of the workflow in this study. The initial step involves preprocessing the imagery data for image segmentation delineating areas for collecting soil sampling field data. The second step entails collecting field soil sample data to measure EC as an indicator of soil salinity. Following this, preprocessed S2 and L8 imagery data is utilized to extract indices, which are then used for feature selection using machine learning algorithms.

The measured EC of soil samples is used as a dependent variable for predicting soil salinity, while the independent variables are derived from S2 and L8 imagery data. Considering that indices are extracted at the object-level segments rather than the pixel level. An object-based feature extraction strategy simplifies analysis, reduces noise, and enhances accuracy^{26–28}. Also, this study uses a hyperparameter tuning evaluation in regression models to optimize model performance and identify the most suitable parameters. Subsequently, sample migration is performed based on the selected most informative features to conduct a time-series analysis using the Continuous Change Detection and Classification (CCDC) algorithm²⁹. Finally, the paper creates a time series of EC maps to analyze temporal variations of soil salinity over time. These maps give a better understanding of the EC patterns and trends over time.

Study area

Within the geographical coordinates ranging from 55° 10' to 55° 22' East longitude and 37° 15' to 37° 25' North latitude, the study area was conducted in the Gonbad region of Golestan Province, situated in the northern part of Iran (Fig. 2a). This region experiences a temperate climate and features predominantly flat terrain. Over the past two decades, it has received an average annual rainfall of 455 mm, with a mean annual temperature of 17 °C³⁰. Notably, the yearly minimum temperature recorded in this period was 15.6 °C, while the maximum temperature reached 37.5 °C. Furthermore, based on the soil taxonomy system established by the United States, the soil type in this area is classified as a typical Haploxerept³¹.

The study area, located on the right flank of the Gorganroud River, spans approximately 24,000 hectares of farmlands in the Golestan and Boostan drainage networks (Fig. 2b), which are irrigated from the Golestan and Boostan dams, respectively. Soil salinity is a prevalent issue in the Gonbad Kavus area, attributed to high groundwater levels and subsequent water evaporation, leaving behind salts in the soil. Therefore, this study aimed to assess the potential of optical Earth observation imagery for predicting soil surface salinity in the eastern region of Gonbad City.

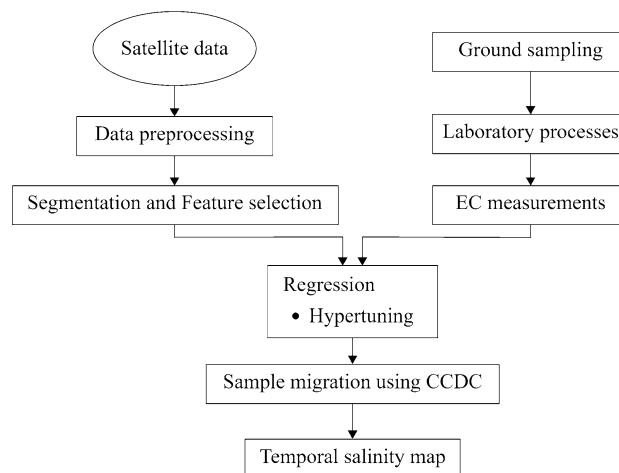
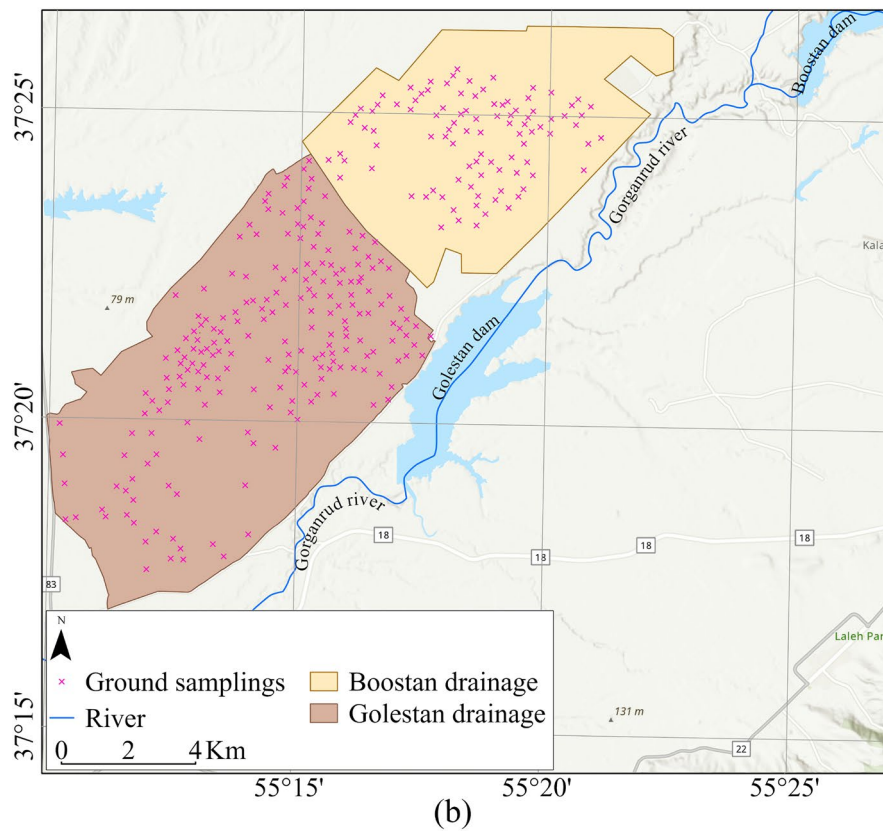
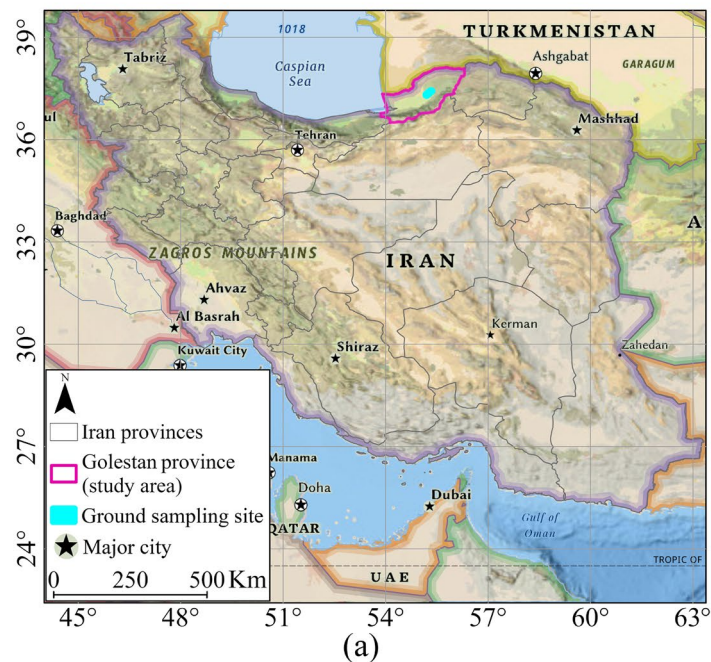


Figure 1. Workflow of the study.



(b)



(a)

Figure 2. Location of Golestan province and ground sampling site (a), overlaying on the National Geographic Style Map in Esri ArcGIS (version 10.8)³², Boostan and Golestan drainage area, and ground sampling locations (b), overlaid on the World Hillshade base map in Esri ArcGIS (version 10.8)³².

Data

Image segmentation

We utilized S2 imagery for land segmentation, which was employed for field sampling and extraction of indices

within each segment. The S2 multispectral imagery comprises 13 bands and has a spatial resolution of 10 m in visible bands and 20 m in the remaining. The S2 dataset, processed by ESA³³, is also accessed through the Google Earth Engine (GEE)³⁴. Utilizing S2 imagery data, we grouped nearby pixels based on similar characteristics such as color and texture³⁹. We employed the K-Nearest Neighbors (KNN) algorithm to group nearby pixels or image segmentation. We tried different grouping levels to find the best one, considering things like how clear the images were. We used various maps and data layers, like contour maps and soil maps, to help with this process. Then, we looked for soil samples in each segmented region (see Supplementary A for detailed information on the segmentation process).

Sampled soil data as the dependent variable

On July 1, 2020, during the summer season, a field survey and soil sample collection were carried out to ensure compatibility with imagery from the S2 and L8 satellites. Global Positioning System (GPS) receivers were employed to pinpoint the sampling locations. Each sampling point was established as a circle with a 5-m radius, from which eight individual soil samples were extracted from the 0–10 cm depth and subsequently combined into a single composite sample. All soil samples were transported to the laboratory for further analysis to determine their moisture content and conductivity. First, 200 g of fresh soil samples were weighed and placed in a drying apparatus. Subsequently, 20.00 g from each naturally air-dried soil sample were precisely measured to prepare a soil extract with saturated paste extracts. The EC was then calculated using the saturation extract method.

S2 and L8 imagery data were used to extract indices as independent variables

The independent variables used in this study are sourced from L8 and S2 imagery data. The L8 is a multispectral satellite with a spatial resolution of 30 m. It is provided by the United States Geological Survey (USGS)³⁵ and accessed through GEE³⁵. To select the most suitable images, we filtered image collections available between April and July 2020 to exclude images having more than 20 percent cloud cover. The filtering process allows the creation of a composite image with no cloud cover.

Twenty-six environmental variables (including auxiliary variables) are used for soil salinity estimation. These variables are used in feature selection to identify the most important features in soil salinity estimation (see Supplementary A for detailed information on feature selection). Table 1 provides a detailed overview of these variables (indices), including their formulas.

Machine learning regression analysis

Regression analysis is a commonly used statistical tool to study relationships between factors, making it straightforward to analyze multifactor data⁵⁴. This study employs SVR^{44,55–59}, RF^{22,58,60–62}, DT^{63,64}, and the XGBoost^{16,65–68} methods to investigate the association between remotely sensed imagery data and field soil salinity. The Scikit-learn for Python (version 1.3.0) is used to implement these algorithms⁶⁹.

Regression and hyperparameter tuning for feature selection

Hyperparameter tuning is essential for maximizing the performance of machine learning algorithms^{70,71}. Algorithms, such as DT, RF, XGBoost, and SVR, involve various types of hyperparameters, and the fine-tuning of these parameters directly influences the algorithm's effectiveness⁷². Several methods can be utilized to optimize parameters and enhance the performance of models, such as the local search method⁷³. This paper uses the BayesSearchCV and GridSearchCV methods to fine-tune the hyperparameters of RF, XGBoost, DT, and SVR models using the scikit-optimize library (version 0.8.1) in Python⁷⁴. Bayesian search CV is grounded in Bayes's rule of conditional probability, using prior knowledge to calculate posterior probabilities. Detailed information on the hyperparameters is provided in supplementary B.

DT

A DT, one of the simplest yet most successful machine learning methods, uses a divide-and-conquer approach to classify and regress large databases^{75,76}. This makes DT one of the common machine-learning methods⁷⁵. A DT is a non-parametric supervised learning method with a tree-like structure⁷⁷. It acts as a function taking input from a vector of attribute values and returning a “decision.” The decision is reached through a series of tests. Each internal node in the tree corresponds to an examination of the value of one of the input attributes, and the branches from that node are labeled with the attribute's possible values. Each leaf node in the tree specifies a value to be returned by the function⁷⁶. Both input and output values are discrete (classification) or continuous (regression). A regression tree has a linear function of some subset of numerical attributes at each leaf rather than a single value. The learning algorithm must decide when to stop splitting and begin applying linear regression over the attributes^{54,76}. DTs are prone to overfitting, meaning they follow the peculiarities of the training dataset too closely and may not perform well on a new dataset, i.e., the test dataset. In such cases, the general predictive accuracy of overfitting DTs will be low, i.e., generalization accuracy⁷⁸. One approach to improve the generalization accuracy is to construct multiple individual trees using a subset of the observations^{78,79}, which is the main idea of the RF algorithm⁸⁰. Detailed information on the DT hyperparameters are provided in Supplementary Table S1.

RF

The RF is an ensemble supervised learning algorithm using a collection of DTs for prediction. Individual trees are constructed by bootstrapping the dataset and averaging the results of all the trees to make the final prediction. Bootstrap aggregating or bagging helps reduce overfitting⁷⁸. The RF can be used to classify categorical target variables and the regression of continuous variables⁸¹. For regression purposes, at each branching of the

| Covariates | Definition | Reference |
|--|--|-----------|
| Elevation ³⁶ | Height above sea level (m) | DEM |
| Diffuse Insolation ³⁷ | Calculate the diffuse incoming solar radiation | DEM |
| Topographic Position Index (TPI) | Calculate with SAGA | DEM |
| Mean Annual Precipitation (MAP) ³⁸ | Derived from the monthly rainfall values | WordClim |
| Mean Annual Temperature (MAT) ³⁸ | Derived from the monthly temperature values | WordClim |
| Normalized Difference Salinity Index (NDSI) ³⁹ | $\frac{(Red-NIR)}{(NIR+Red)}$ | L8/S2 |
| Vegetation soil salinity index (VSSI) ⁴⁰ | $2 \times GREEN - 5 \times (RED + NIR)$ | L8/S2 |
| Soil Adjusted Vegetation Index (SAVI) ⁴¹ | $\frac{(1+L)(NIR-Red)}{(NIR+Red+L)}$ | L8/S2 |
| Normalized Difference Vegetation Index (NDVI) ⁴² | $\frac{(NIR-Red)}{(NIR+Red)}$ | L8/S2 |
| Enhanced Vegetation Index (EVI) ⁴³ | $\frac{2.5(NIR-Red)}{(NIR+6R-7.5B)+1}$ | L8/S2 |
| Generalized Difference Vegetation Index (GDVI) ⁴⁴ | $\frac{(NIR^2-RED^2)}{(NIR^2+RED^2)}$ | L8/S2 |
| Normalized Difference Water Index (NDWI) ⁴⁵ | $\frac{(NIR-SWIR)}{(NIR+SWIR)}$ | L8/S2 |
| Modified Normalized Difference Water Index (MNDWI) ⁴⁶ | $\frac{(GREEN-MIR)}{(GREEN+MIR)}$ | L8/S2 |
| Normalized Difference Salinity Index (NDSI) ³⁹ | $\frac{(RED-NIR)}{(RED+NIR)}$ | L8/S2 |
| Multi-resolution Valley Bottom Flatness (MrVBF) ⁴⁷ | DEM | DEM |
| Topographic Wetness Index (TWI) ⁴⁸ | $\ln\left(\frac{local\ slope}{slope}\right)$ | DEM |
| Brightness Index (BI) ³⁹ | $\sqrt{(Green)^2 + (RED)^2}$ | L8/S2 |
| Salinity 1 (S1) ³⁹ | $\sqrt{(Green \times Red)}$ | L8/S2 |
| Salinity 2 (S2) ⁴⁹ | $\sqrt{(Blue \times Red)}$ | L8/S2 |
| Salinity 3 (S3) ⁵⁰ | $\frac{(Red \times Green)}{Blue}$ | L8/S2 |
| Salinity 4 (S4) ⁵⁰ | $\frac{Blue}{Red}$ | L8/S2 |
| Extended NDVI ¹³ | $\frac{(NIR + SWIR2 - Red)}{(NIR + SWIR2 + Red)}$ | L8/S2 |
| Extended EVI ¹³ | $\frac{2.5 \times (NIR + SWIR1 - Red)}{NIR + SWIR1 + 6 \times Red - 7.5 \times Blue + 1}$ | L8/S2 |
| Canopy Response Salinity Index ⁵¹ | $\sqrt{\frac{((NIR \times Red) - (Green \times Blue))}{((NIR \times RED) + (Green \times Blue))}}$ | L8/S2 |
| Two-band EVI ⁵² | $2.5 \times \frac{(NIR - Red)}{(NIR + 2.4 \times RED + 1)}$ | L8/S2 |
| Difference Infrared Index (NDII) ⁵³ | $\frac{(NIR - SWIR1)}{(NIR + SWIR1)}$ | L8/S2 |

Table 1. Variables used in predicting soil salinity.

regression tree, the mean of the samples on the leaf nodes and the Root Mean Square Error (RMSE) formed between each sample are calculated. Following the minimum RMSE of leaf nodes as a branching condition, the regression tree stops when no more features are available, or the overall RMSE is optimal⁷⁷. The key to creating an accurate model is ensuring that the base learner, typically a regression tree, is as uncorrelated as possible to produce a robust generalization ability⁸⁰. Detailed RF hyperparameter information is provided in Supplementary Fig. S2 and Supplementary Table S2.

SVR

The Support Vector (SV) algorithm, a set of related supervised learning methods, was first introduced for pattern recognition⁸² and then generalized to solve regression problems⁸³. The SVR is a tool for overall and short-term forecasting or when real-time analysis is required⁸⁴. Also, it works in an infinite-dimensional space, giving it an edge over similar networks⁸⁴. The SVR investigates the relationship between one or more predictor variables and a real-valued (continuous) dependent variable⁸⁵. SVR finds a best-fitting hyperplane to data points in a continuous space, while in linear regression, a line is fitted to the data points. In SVR, the best-fitting hyperplane passes through as many sample points as possible within a certain distance, called a margin, which is defined to be the smallest distance between the decision boundary and any of the samples⁸⁶. SVR is very sensitive to the input data type, as it can produce incorrect results if the data spans a wide range. Therefore, data normalization is an essential step in using SVR⁸⁴. Detailed information on the SVR hyperparameters are provided in Supplementary Fig. S3.

XGBoost

The XGBoost algorithm was proposed by Chen and Guestrin⁸⁷. A scalable implementation of XGBoost is robust and highly efficient^{88,89}. It uses the Classification and Regression Trees (CART) and is jointly decided by multiple related DTs⁹⁰. In this structure, the input sample of the next DT is related to the training and prediction results of the previous DT. Like most machine learning algorithms, in XGBoost, the objective is to minimize the sum of the loss function to control the accuracy and complexity of the model⁹⁰. Detailed information on the XGBoost hyperparameters are provided in Supplementary Table S3.

Migration sample with algorithm CCDC

The fundamental concept behind CCDC involves fitting a simple harmonic model to a cloud-free time series and detecting changes when the difference between observed and predicted pixel values surpasses a predefined threshold for consecutive periods. Notably, Chen et al.⁸⁷ enhanced the CCDC algorithm by introducing a multi-harmonic model capable of fitting intricate phenological profiles in cultivated land. The paper used the CCDC method to select unchanged samples and generate additional field samples for each period during which field data were not collected in those years. These unchanged field samples serve as training and test data for all years. The process commenced with cloud masking of L8 images and the computation of two essential spectral indices: the NDWI and the NDSI (Table 1). Subsequently, these indices were leveraged with the CCDC model, as outlined in Eq. (1), to identify and isolate the unchanged field samples⁹¹.

$$F(i, x) = a_{0,i} + a_{1,i}\cos(2\pi/t_x) + b_{1,i}\sin(2\pi/t_x) + c_{1,i} \quad (1)$$

where i , x , and t represent the spectral index, Julian date, and the number of days in a year (i.e., 365.25 days); $a_{0,i}$ stands for the overall value of spectral index i of an L8 image; $a_{1,i}$ and $b_{1,i}$ specify the intra-year change. Furthermore, $c_{1,i}$ pertains to interannual values, and these new samples estimate model values and residuals by comparing observed and modeled sample values, as outlined in Eq. (1). In this context, a threshold value of 20% was employed. If the residual exceeded this threshold, it was assumed that an interannual variation had occurred. Samples failing to meet this threshold were excluded from other time interval classifications. The remaining residual samples were considered to exhibit stable spectral responses and were assumed to remain unchanged throughout the study period.

Before classification, these unaltered samples were randomly divided into a training group (70%) and a test group (30%). This division allowed for the assessment of the accuracy of the generated salinity maps using both the classification algorithm's training and test groups. The process CCDC is as follows: Identification of essential features for salinity determination based on field data; Time series analysis based on the above feature to identify field points whose value does not change over time (280 samples); Time series analysis based on the above feature to generate artificial points whose value does not change over time (576 sample artifacts); Producing a map with points 1 and 2 with different regression methods and comparing their accuracy and choosing the most accurate map from the previous step.

Accuracy assessment

Three distinct criteria, RMSE, R^2 , and Mean Absolute Error (MAE), were used to estimate the accuracy of the prepared prediction models (see Supplementary D).

Results

Soil salinity estimation using L8 imagery data

Feature extraction and segmentation of L8 data

According to Table 2, the results revealed that out of 26 environmental variables (including several auxiliary variables), the "Random Forest-Backward Feature Elimination" (RFE-RF) method⁹². The relative importance of independent variables for estimating soil salinity using L8 imagery is shown in Supplementary Fig. S4. It can be concluded that the parameters obtained from remote sensing are more important than other factors of soil formation, i.e., geomorphology, topography, and EC, in the spatial estimation of soil salinity in the surface horizon.

The MNDWI (0.35%) and NDSI (0.33%) were determined to be important predictors of soil salinity in the RF method (Supplementary Fig. S4a). The NDWI is the third important auxiliary variable, with a relative importance of 0.23%. Moreover, the XGBoost has shown better results in salinity estimation with L8 data. Therefore, the critical factors using L8 data, as shown in Supplementary Fig. S4b, are B3, MNDWI, NDSI, NDWI, NDVI, EVI, MrBVF, TWI, TPI, index S1, and S4.

Furthermore, based on the results presented in Supplementary Fig. S4c, the DT algorithm ranks the relative importance of auxiliary variables as follows: MNDWI, NDSI, NDWI, TWI, SAVI, MrBVF, TPI, NDVI, EVI, S3 (Table 1), in an ascending to descending trend.

Accuracy assessments of salinity estimation using L8 data

The scatter diagram of soil EC is shown in Fig. 3. The highest R^2 value is related to the DT model. Therefore, DT can predict the salinity more accurately than others with RMSE = 12.25, MAE = 2.15, and R^2 = 0.85. Furthermore, RF has almost the same performance as DT, while the XGBoost algorithm has the lowest coefficient of explanation for soil salinity estimation with RMSE = 18.62, MAE = 2.87, and R^2 = 0.58).

Spatial estimation of Soil salinity with L8 data

Suitable bands were determined to train the regression algorithms and create salinity level maps, and pixel data corresponding to sample points within each class were employed as training data. The regression was performed based on five salinity classes, and Fig. 4 displays the results obtained from various supervised regression algorithms.

Soil salinity estimation using S2 imagery data

Feature extraction and segmentation S2 data

As indicated in Table 2, the results revealed that out of a total of 26 environmental variables (including several auxiliary variables), the RFE-RF method⁹² identified the following eight variables as having the highest feature importance: Clay, Carbonate, NDVI, S3, S2, S1, Greens, and Brightness. These variables were derived from band

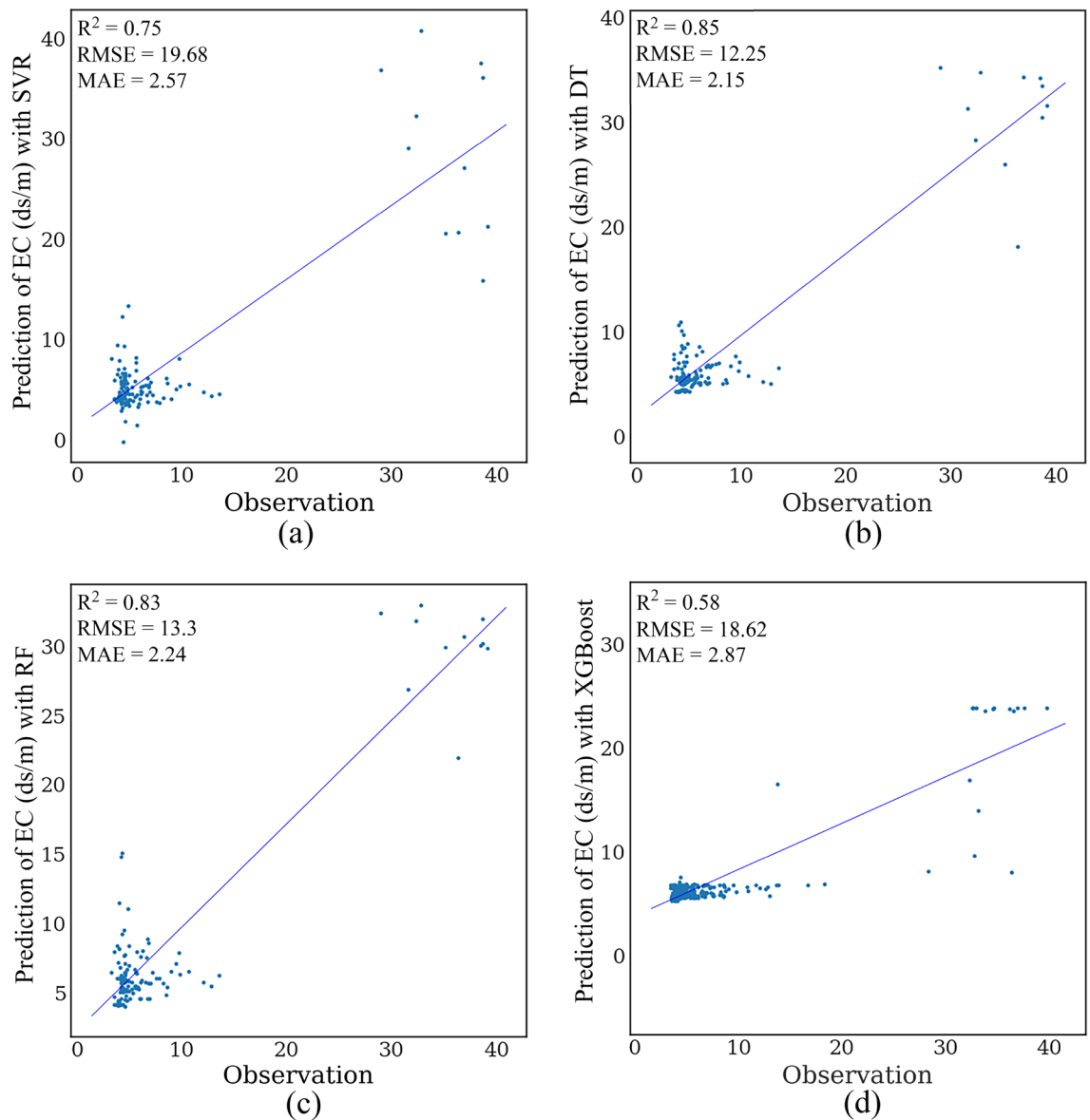


Figure 3. Scatter plots of SVR (a), RF (b), DT (c), and XGBoost (d) methods estimating the soil salinity value using the L8 image. The Matplotlib for Python (version 3.7.2) draws graphs⁹³.

assignments of the L8. Additionally, six parameters, Diffuse, MrVBF, Midslope, Standard, and SAGA, extracted from the DEM⁹⁴, and two climatic parameters, MAP and MAT, were selected for analysis.

Supplementary Figure S5 shows that the importance of environmental auxiliary variables differed in each algorithm using S2 data. The results showed that remote sensing data were the most important predictors of soil salinity. In Supplementary Fig. S5, the essential controllers are auxiliary variables of the humidity index, salinity, modified soil index, MrVBF, and climate variables of MAT, NDVI, S3, and TWI (Table 1). Soil salinity models with advanced DT algorithms have the highest effect on the prediction of soil salinity in the study area. This shows that humidity, altitude, elevation, and vegetation are the most important factors of soil formation in the study area. It is effective in the spatial distribution of soil and its characteristics because the topography and vegetation of each region are some of the essential and influential characteristics of the soil characteristics of that region, including soil salinity.

Accuracy assessments of salinity estimation using S2 data

As seen in Fig. 5, the DT had the highest value of R^2 and was the best model for estimating EC. The best-fitted model with the highest explanation coefficient and the lowest error was selected. Based on the results obtained from the estimator's machine learning, the SVR has the lowest accuracy, and the DT method has the most accuracy.

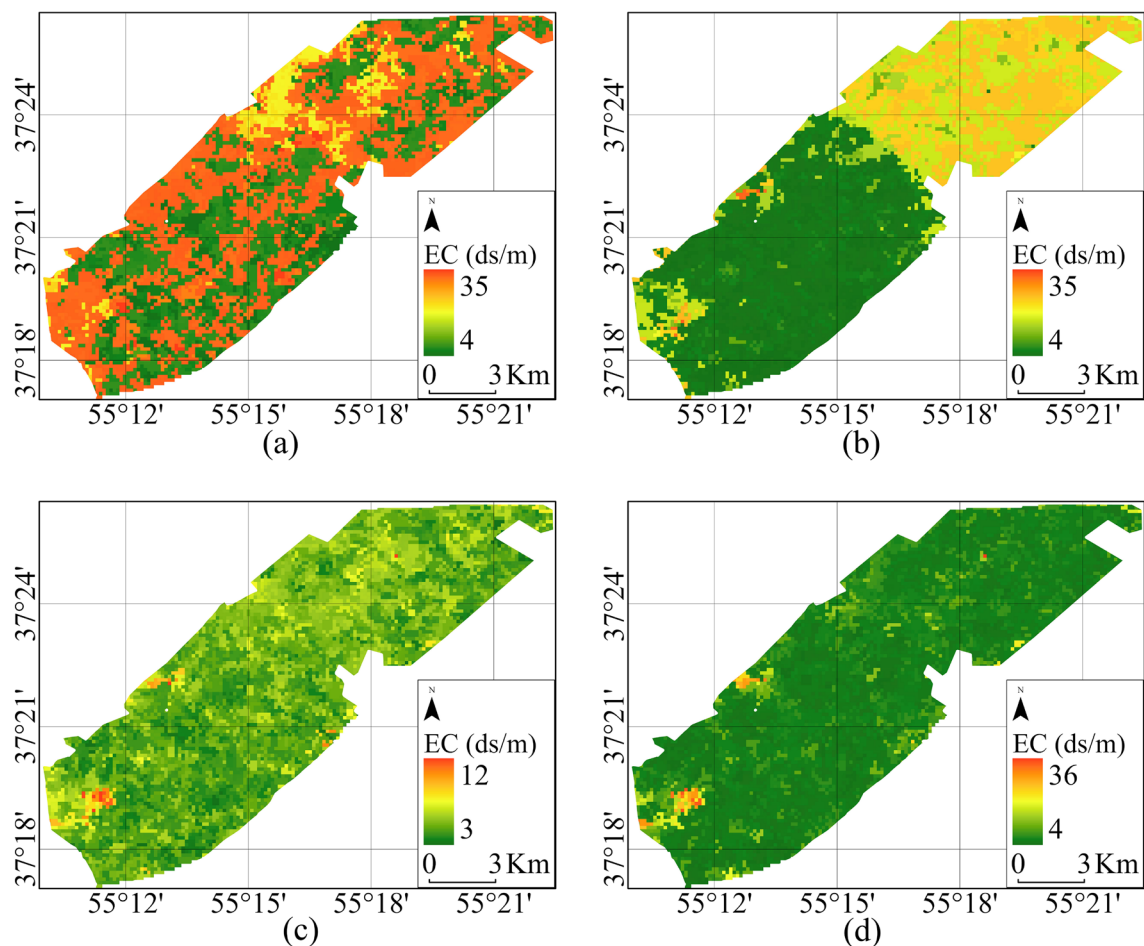


Figure 4. Soil salinity map created by L8 data in 2020 (a) XGBoost, (b) DT, (c) SVR, and (d) RF methods.

Spatial estimation of Soil salinity with S2 data

Figure 6 illustrates the classified salinity map of the study area, created using S2 data. It highlights the effectiveness of drainage in mitigating soil salinity. The RF algorithm exhibited higher accuracy than others due to its utilization of more trees and optimal bootstrap sampling techniques for auxiliary variables and observation points. The DT algorithm also performed well, while the SVR demonstrated improved accuracy over the XGBoost model.

The map indicates a range of soil salinity values in the study area, with the highest recorded at 24 ds/m and the lowest at 4 ds/m. These areas primarily consist of agricultural fields with low slopes and are equipped with pipe drainage systems.

Temporal analysis of soil salinity using L8 imagery data

We conducted temporal analysis on soil salinity exclusively using L8 imagery data. Supplementary Figures S6, S7, and S8 show the soil salinity map generated using RF and L8 data between 2013 and 2020. They revealed a decreasing trend in salinity levels in the study area. The Golestan drainage area started in 2013 and ended in 2019, and the Bostan drainage area began in 2019 and was completed in 2023. This can be attributed to the extensive use of pipe drainage in agricultural regions and potential climate changes. Also, the average soil salinity time series ten years after installing the drainage network is shown in Supplementary Figs. S9 and S10. In 2012, the predominant and least prevalent land classifications were those subject to high salinity restrictions (54%) and those free of salinity restrictions (2%), respectively. However, by 2023, the most prevalent land classification had shifted to areas without salinity limitations (40%), while those with very high salinity restrictions constituted 7% of the region's landscape.

Over the period from 2012 to 2023, the extent of land subject to high salinity restrictions experienced a significant decrease in drainage areas, nearing parity with the extent of land free of salinity restrictions. This transformation suggests a notable reduction in salinity levels within the drainage area and an overall increase in salinity across other non-drained regions.

Upon analyzing the histograms for 2012, it's apparent that salinity in agricultural regions primarily falls within the range of 8–18, with a peak around 13 (Supplementary Fig. S11). In the 2023 histogram shown in Supplementary Fig. S12, the minimum and maximum values have decreased, now confined to a range of 4–15, indicating an overall reduction in salinity in agricultural areas. Notably, the shift in the peak of the histogram to the left suggests that regions with higher initial salinity values within this range have experienced more pronounced reductions, indicating a significant improvement in their salinity conditions.

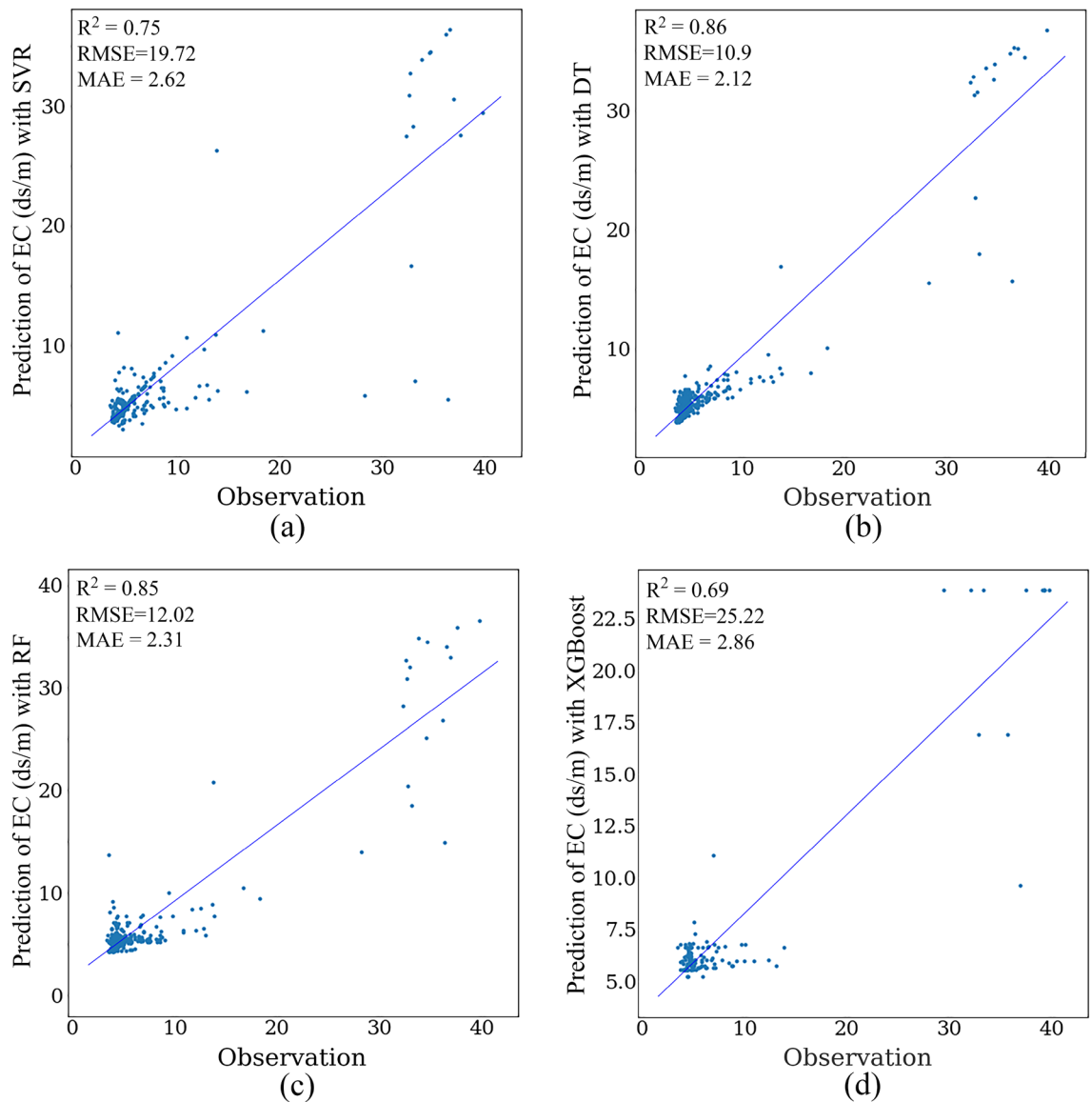


Figure 5. Scatter plots of SVR (a), RF (b), DT (c), and XGBoost (d) methods estimating the soil salinity value using the S2 data. The Matplotlib for Python (version 3.7.2) is used to draw graphs⁹³.

A 7-year analysis of soil salinity spanning from 2013 to 2020 was conducted using Landsat 8 (L8) imagery, focusing on the differentiation of pixels between bare soil and plant cover. This analysis aimed to identify optimal indices and suitable timing for studying changes. A comparison of the soil salinity maps for 2013 and 2020, illustrated in Supplementary Fig. S13, indicates a decreasing trend in salinity, particularly noticeable in agricultural areas with tile drainage systems. However, rangeland areas display a notable trend potentially influenced by climate change.

A slight increase is observed within the range of high salinity levels when comparing data from 2013 to 2020, consistent with global trends. Consequently, salinity has risen in non-agricultural regions but declined in agricultural areas. A distribution chart comparing values from 2013 has been generated for further investigation. The histograms and this distribution chart confirm that changes diminish within the 8–18 salinity range, displaying a negative regression slope. This indicates that areas with initially higher salinity levels experienced more significant reductions. However, in cases exceeding a salinity value of 18, a relative increase in salinity is observed. On average, this increase amounts to approximately five units of EC for values greater than 18 (Supplementary Fig. S14).

Figure 7 compares soil salinity and provides the slope (β) and significance level (P) values related to the salinity trend in drained areas. Based on these values, it can be concluded that soil salinity has decreased with drainage. Drained regions exhibit a significant decreasing trend at a 5% level, with the most substantial decrease observed in 2013. Based on Kendall's statistics, the analysis of soil salinity changes in the study area over the 11 years (2012–2023) indicates a consistent downward trend in salinity changes following drainage implementation. However, the situation is different for non-drained areas.

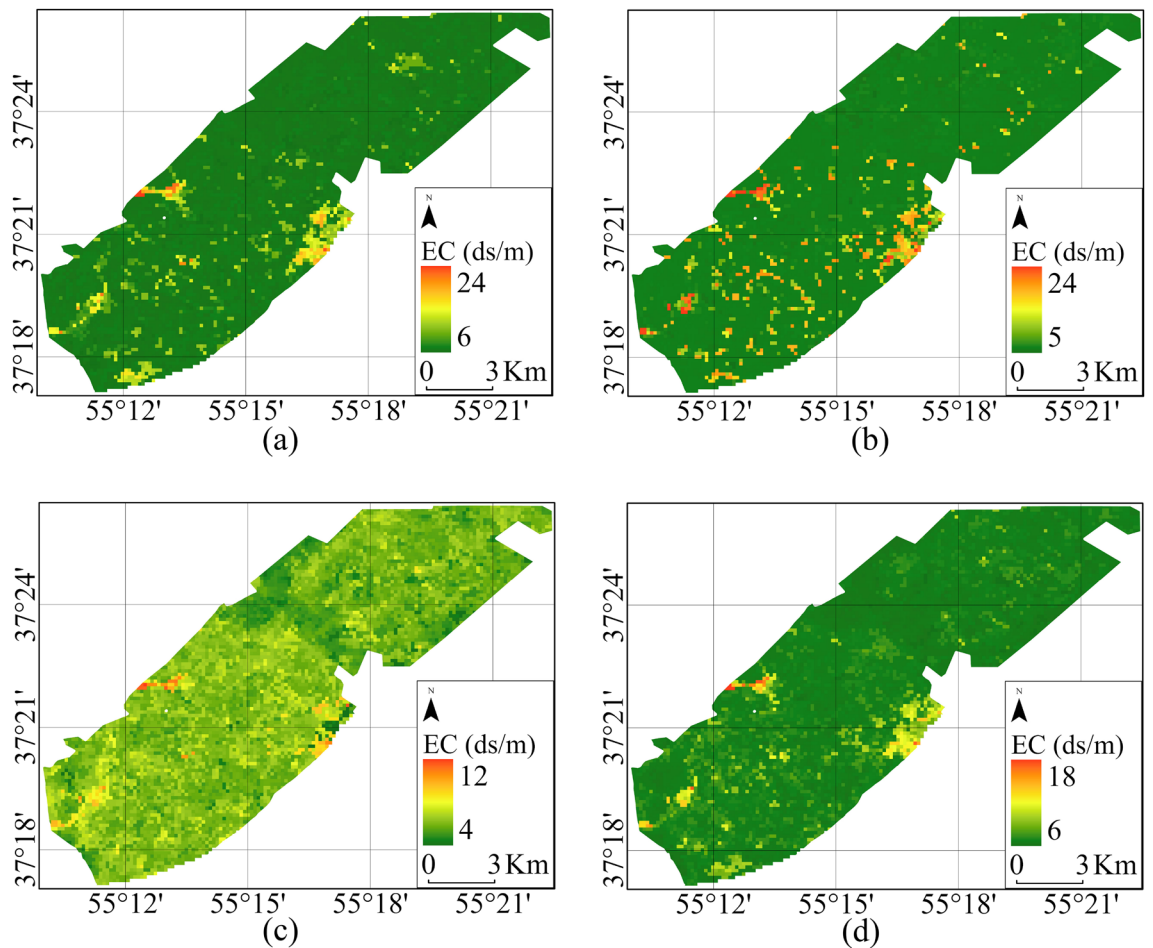


Figure 6. Soil Salinity map created with S2 data in 2020 using (a) RF, (b) XGBoost, (c) SVR, and (d) DT.

Discussion

Estimation of soil salinity using S2 and L8 data

The study compares four machine learning methods, RF, DT, XGBoost, and SVR, to estimate soil salinity levels using remotely sensed L8 and S2 data. We have found that the accuracy of DT and RF in salinity estimation at a depth of 10 cm is higher than XGBoost and SVR. This shows that RF and DT are more powerful in modeling highly nonlinear dimensional relationships than XGBoost and SVR. Our findings concur with those found in^{95,96}. Merembayev et al.⁹⁷ indicated that DT and RF have almost the same performance for soil salinity estimation, which is aligned with our findings. Ding et al.⁹⁸ used the Landsat Enhanced Thematic Mapper Plus (ETM+) image, stating that DT is suitable for extracting saline soil information. Qi-sheng et al.⁹⁹ found the effectiveness of DT for soil salinity extraction from ETM data, which is aligned with our findings. Fu et al.¹⁰⁰ reported that a combination of DT using Landsat TM images had superior performance in land salinization classification. In addition, Breiman⁸⁰ stated that by aggregating multiple models, the instability of a single-tree model is minimized, which leads to an improvement in consistency¹⁰¹. While our results demonstrate the superiority of DT over RF for indices developed by L8 imagery data, Haq et al.¹⁰² have argued that RF outperforms DT using L8 imagery data. It is worth noting that their soil samples were collected at a depth of 15 cm, whereas our study samples were collected at a depth of 10 cm. Nevertheless, our findings indicate a close alignment between the predicted and actual EC values when using the RF and DT models. This contrasts several studies showing RF superior to DT for estimating soil salinity^{103–105}.

Feature importance for estimation of soil salinity

Our study shows the advantage of ranking the relative importance of predictor variables (indices) using machine learning methods. Salcedo et al.¹⁰⁶ have stated that applying different spectral indices helps characterize soil salinity. Accordingly, our results suggest that MNDWI is an optimal index for assessing soil salinity using the DT and RF methods with both L8 and S2 imagery data. This is aligned with Xu⁴⁶, Qi-sheng et al.⁹⁹, and Fu et al.¹⁰⁰, stating that MNDWI is suitable for salinization mapping. Also, Ding et al.⁹⁸ reported the usefulness of MNDWI combined with NDVI in soil salinity classification. Additionally, our results indicated that the NDSI is the optimal index for salinity estimation using S2 data. Aligning with our findings, Shrestha et al.¹⁰⁷ have found that the only significant predictor of observed soil salinity was NDSI.

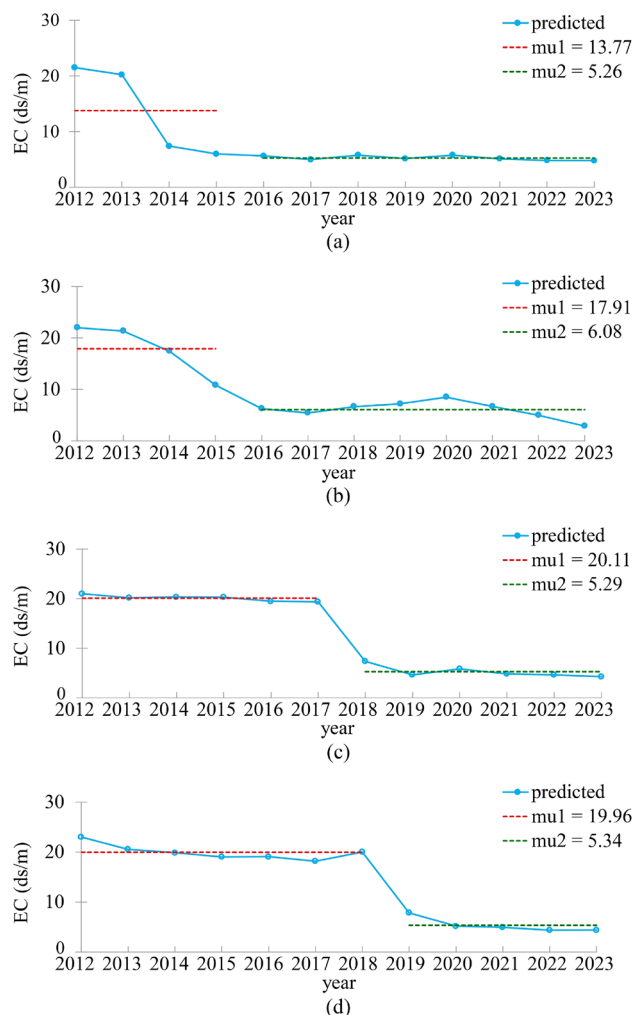


Figure 7. Trend diagram of Kendall and Pettit Mutation test for phases 1, 2, 3, and 4 of drained areas.

According to our study, topography and vegetation indices are important indicators of soil salinity. This is aligned with the findings of Wang et al.¹⁰⁸ They indicated that environmental factors contributed substantially to soil salinity estimation, including the digital elevation model (DEM) and Green Atmospherically Resistant Vegetation Index (GARI).

Additionally, our findings show that the salinity index reaches its maximum value, particularly near the soil surface in specific sections. Factors such as capillary currents, soil texture, groundwater, and moisture levels are pivotal in transporting solutes to surface layers, especially during dry seasons. Consistent with our findings, Taghizadeh-Mehrjardi et al.⁵⁵ identified climate parameters, particularly prolonged droughts, as the primary factors influencing soil salinity in lowland areas. They also highlighted saline parent materials, soil texture, and the lack of surface irrigation and drainage as contributing factors.

Moreover, our findings showed that in areas with saline soil, an increase in soil moisture leads to a decrease in reflectance within the visible and near-infrared regions, resulting in enhanced soil salinity estimation, as also reported by Cao et al.¹⁰⁹.

Temporal and spatial changes in soil salinity

Analyzing soil salinity changes from 2012 to 2023 in the study area indicates that most of these lands did not transition from saline to non-saline over 11 years in the undrained areas. The extent of moderate and high salinity classes gradually decreased, while the area of lands with very high and severe salinity increased. Despite the similar influence of climate and weather changes on salinity increase, agricultural areas with drainage systems have experienced reduced salinity. Additionally, areas with improved soil amendments and established drainage systems exhibit more favorable salinity classifications than non-drained areas¹¹⁰. Moreover, regions with drainage systems in place report higher crop yields than non-drained areas; this is corroborated by the soil salinity chart in this range, which demonstrates a declining trend, affirming the effective functioning of drainage systems. Additionally, the high EC of drainage water, signifying a very high salinity class, indicates that drainage systems have efficiently removed salinity from the soil, and these changes have stabilized over time¹¹⁰. This is aligned with Gopalakrishnan et al.¹¹¹, as they reported that poor drainage intensifies soil salinity, and salinization has severe

implications for food production and security. Therefore, this indicates the importance of installing a proper drainage system aligned with Singh's findings¹¹².

Since the soil salinity of the grazing lands around the outlet drainage evaporation pond has gradually increased over the 11 years following the implementation of the pipe drainage project, it is recommended to consider a comprehensive Integrated Water Resources Management (IWRM) plan. This plan would utilize serial biological drainage for agricultural drainage waters multiple times, each cycle devoted to cultivating salt-tolerant crops. The highly saline drainage water produced at the end of this process can be repurposed for non-agricultural activities such as aquaculture or salt production. Eventually, a small volume of water should be directed into the outlet drawing, or ideally, no water should enter the outlet and instead be discharged directly into the sea. Mardanifar et al.¹¹³ investigated the declining salinity trend in the Golestan and Boustan dam areas. Their findings regarding the trend of electrical conductivity (EC) changes in the sampled wells within the project area generally indicate a reduction during the operation of the drainage system. Factors such as the terrain slope, positioning of drainage systems, and the known distribution of soil salinity are considered significant in this region.

Overall, proper drainage is an essential aspect of managing soil salinity. It helps to remove excess salts from the soil and prevent their accumulation. Drainage systems such as tile drains or subsurface drainage can be implemented to improve water movement and reduce salt build-up in the root zone. Effective drainage can help to lower the water table, preventing the capillary rise of saline groundwater. It also promotes leaching, which involves flushing salts from the soil profile through water movement. This helps to maintain a more favorable salt balance in the soil, reducing the risk of salinity problems. Farmers can effectively manage soil salinity by implementing proper drainage measures and creating a better plant-growing environment. This can improve yields, improve water-use efficiency, and enhance soil fertility. In addition, proper drainage can help minimize the environmental impact of soil salinity. It reduces the risk of saltwater intrusion into freshwater sources, protecting water quality and preserving aquatic ecosystems. Overall, drainage plays a crucial role in soil salinity by preventing salt accumulation, improving water movement, and promoting optimal plant growth and agricultural productivity.

Limitations

The assessment of soil salinity change trends through the utilization of remote sensing data has a few limitations. The principal constraints in this study are as follows: (1) Remote sensing data is typically collected based on a specific scale, with each data pixel potentially representing several meters. This spatial scale may result in losing fine details related to soil salinity changes at smaller scales. (2) Acquiring remote sensing data over time and for specific areas can be challenging. Some regions may have limited access to satellite imagery, which can restrict the comprehensive evaluation of soil salinity changes. (3) Cloud cover in satellite imagery can introduce interferences in the remote sensing process and degrade image quality. Accurate soil salinity monitoring requires high-quality, cloud-free images.

Conclusion

The study utilized remote sensing imagery data and soil salinity samples in machine learning algorithms to explore factors affecting salinity levels in salt-affected crop fields. The DT machine learning model effectively predicted total soil salinity in heavily vegetated croplands. The study highlighted that undrained areas exhibit greater sensitivity to salinity, likely influenced by climate patterns. Furthermore, a significant disparity in average salinity levels between undrained and drained land was observed, with lower EC in agricultural land due to salt leaching. To mitigate salinity, the study recommends the installation of drainage pipes at a depth of 2 m to reduce and stabilize soil electrical conductivity. The results suggest incorporating environmental variables in time series modeling to predict soil salinity, considering that climate and weather variations contribute to increased salinity.

Data availability

According to the first author's university regulations, the datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 23 November 2023; Accepted: 18 April 2024

Published online: 06 May 2024

References

1. AbdelRahman, M. A. An overview of land degradation, desertification and sustainable land management using GIS and remote sensing applications. *Rendiconti Lincei. Scienze Fisiche e Naturali* **34**(3), 767–808 (2023).
2. Metwaly, M. M., AbdelRahman, M. A. & Abdellatif, B. Heavy metals and micronutrients assessment in soil and groundwater using geospatial analyses under agricultural exploitation in dry areas. *Acta Geophys.* **71**(4), 1937–1965 (2023).
3. AbdelRahman, M. A. et al. Determining the extent of soil degradation processes using trend analyses at a regional multispectral scale. *Land* **12**(4), 855 (2023).
4. Hafez, E. M. et al. Minimizing hazard impacts of soil salinity and water stress on wheat plants by soil application of vermicompost and biochar. *Physiol. Plant.* **172**(2), 587–602 (2021).
5. Vermeulen, D. & Van Niekerk, A. Machine learning performance for predicting soil salinity using different combinations of geomorphometric covariates. *Geoderma* **299**, 1–12 (2017).
6. Gao, L. et al. Road extraction from high-resolution remote sensing imagery using refined deep residual convolutional neural network. *Remote Sens.* **11**(5), 552 (2019).
7. Yu, H. et al. Mapping soil salinity/sodicity by using Landsat OLI imagery and PLSR algorithm over semiarid West Jilin Province, China. *Sensors* **18**(4), 1048 (2018).
8. Wu, Z. et al. Remote sensing monitoring and driving force analysis of salinized soil in grassland mining area. *Sustainability* **14**(2), 741 (2022).

9. AbdelRahman, M. A. *et al.* Detecting and mapping salt-affected soil with arid integrated indices in feature space using multi-temporal Landsat imagery. *Remote Sens.* **14**(11), 2599 (2022).
10. Aboelsoud, H. M. *et al.* Quantitative estimation of saline-soil amelioration using remote-sensing indices in arid land for better management. *Land* **11**(7), 1041 (2022).
11. Erkin, N. *et al.* Method for predicting soil salinity concentrations in croplands based on machine learning and remote sensing techniques. *J. Appl. Remote Sens.* **13**(3), 034520–034520 (2019).
12. Taghizadeh-Mehrjardi, R., Nabiollahi, K. & Kerry, R. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma* **266**, 98–110 (2016).
13. Wang, J. *et al.* Machine learning-based detection of soil salinity in an arid desert region, Northwest China: A comparison between Landsat-8 OLI and Sentinel-2 MSI. *Sci. Total Environ.* **707**, 136092 (2020).
14. Wang, F. *et al.* Multi-algorithm comparison for predicting soil salinity. *Geoderma* **365**, 114211 (2020).
15. Wang, J. *et al.* Capability of Sentinel-2 MSI data for monitoring and mapping of soil salinity in dry and wet seasons in the Ebinur Lake region, Xinjiang, China. *Geoderma* **353**, 172–187 (2019).
16. Ma, S. *et al.* Investigation of the spatial and temporal variation of soil salinity using Google Earth Engine: A case study at Werigan-Kuqa Oasis, West China. *Sci. Rep.* **13**(1), 2754 (2023).
17. Ge, X. *et al.* Updated soil salinity with fine spatial resolution and high accuracy: The synergy of Sentinel-2 MSI, environmental covariates and hybrid machine learning approaches. *Catena* **212**, 106054 (2022).
18. Gorji, T. *et al.* Soil salinity analysis of Urmia Lake Basin using Landsat-8 OLI and Sentinel-2A based spectral indices and electrical conductivity measurements. *Ecol. Indic.* **112**, 106173 (2020).
19. Bannari, A. *et al.* Sentinel-msi and landsat-oli data quality characterization for high temporal frequency monitoring of soil salinity dynamic in an arid landscape. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **13**, 2434–2450 (2020).
20. Davis, E., Wang, C. & Dow, K. Comparing Sentinel-2 MSI and Landsat 8 OLI in soil salinity detection: A case study of agricultural lands in coastal North Carolina. *Int. J. Remote Sens.* **40**(16), 6134–6153 (2019).
21. Nabiollahi, K. *et al.* Assessing agricultural salt-affected land using digital soil mapping and hybridized random forests. *Geoderma* **385**, 114858 (2021).
22. Xiao, C. *et al.* Prediction of soil salinity parameters using machine learning models in an arid region of northwest China. *Comput. Electron. Agric.* **204**, 107512 (2023).
23. He, B. *et al.* Spatiotemporal variation and future predictions of soil salinization in the Werigan-Kuqa River Delta Oasis of China. *Sustainability* **15**(18), 13996 (2023).
24. Haq, Y. U. *et al.* Spatial mapping of soil salinity using machine learning and remote sensing in Kot Addu, Pakistan. *Sustainability* **15**(17), 12943 (2023).
25. Rajath, E. *et al.* Soil Salinity Mapping Using Multisensor Data Employing Machine-Learning Technique: A Case Study from Indo-gangetic Plain. In *Remote Sensing of Soils* 439–453 (Elsevier, 2024).
26. Mehri, S., Hooshangi, N. & Ghaffari Razin, M. R. Providing a knowledge-based method for distinguishing crops and estimating a cultivation area (Case study: The Moghan Plain). *Geograph. Eng. Territory.* **7**(1), 191–208 (2023).
27. Awad, M. An unsupervised artificial neural network method for satellite image segmentation. *Int. Arab J. Inf. Technol.* **7**(2), 199–205 (2010).
28. Blaschke, T., Burnett, C. & Pekkarinen, A. Image segmentation methods for object-based analysis and classification. In *Remote Sensing Image Analysis: Including the Spatial Domain* 211–236 (Springer, 2004).
29. Zhu, Z. & Woodcock, C. E. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens. Environ.* **144**, 152–171 (2014).
30. Ghorbani, K., Zakerinia, M. & Hezarjaribi, A. The effect of climate change on water requirement of soybean in Gorgan. *J. Agric. Meteorol.* **1**(2), 60–72 (2014).
31. Roozitalab, M. H. *et al.* Major soils, properties, and classification. In *The Soils of Iran* (eds Roozitalab, M. H. *et al.*) 93–147 (Springer International Publishing, 2018).
32. Esri. *Working with basemap layers*. (2018). <https://desktop.arcgis.com/en/arcmap/latest/map/working-with-layers/working-with-basemap-layers.htm>.
33. *Copernicus Sentinel-2 (processed by ESA)*. MSI Level-2H/F Harmonized/Fused Reflectance Product. Collection 1. (European Space Agency, 2021).
34. Catalog, E. E. D. *Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A*. (2020).
35. Catalog, E.E.D. *USGS Landsat 8 Level 2, Collection 2, Tier 1*. (2022). https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C02_T1_L2#terms-of-use.
36. Olaya, V. & Conrad, O. Geomorphometry in SAGA. *Dev. Soil Sci.* **33**, 293–308 (2009).
37. Wilson, J. P., Gallant, J. C. *Terrain Analysis: Principles and Applications* (Wiley, 2000).
38. Fick, S. E., Hijmans, R. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **1**–14 (2017).
39. Khan, N. M. *et al.* Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators. *Agric. Water Manag.* **77**(1–3), 96–109 (2005).
40. Dehni, A. & Lounis, M. Remote sensing techniques for salt affected soil mapping: Application to the Oran region of Algeria. *Proc. Eng.* **33**, 188–198 (2012).
41. Huete, A. R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **25**(3), 295–309 (1988).
42. Rouse Jr, J. W. *et al.* Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. (1974).
43. Priya, M. *et al.* Monitoring vegetation dynamics using multi-temporal Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) images of Tamil Nadu. *J. Appl. Nat. Sci.* **15**(3), 1170–1177 (2023).
44. Wu, W. *et al.* Mapping soil salinity changes using remote sensing in Central Iraq. *Geoderma Reg.* **2**, 21–31 (2014).
45. Gao, B.-C. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* **58**(3), 257–266 (1996).
46. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **27**(14), 3025–3033 (2006).
47. Gallant, J. C., Dowling, T. I. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resour. Res.* **39**(12) (2003).
48. Böhner, J., Selige, T. Spatial prediction of soil attributes using terrain analysis and climate regionalization. In *SAGA-Analyses and Modelling Applications* (Goltze, 2006).
49. Allbed, A. & Kumar, L. Soil salinity mapping and monitoring in arid and semi-arid regions using remote sensing technology: A review. *Adv. Remote Sens.* **02**(04), 13 (2013).
50. Abbas, A. *et al.* Characterizing soil salinity in irrigated agriculture using a remote sensing approach. *Phys. Chem. Earth Parts A/B/C* **55**, 43–52 (2013).
51. Scudiero, E., Skaggs, T. H. & Corwin, D. L. Regional-scale soil salinity assessment using Landsat ETM+ canopy reflectance. *Remote Sens. Environ.* **169**, 335–343 (2015).
52. Jiang, Z. *et al.* Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.* **112**(10), 3833–3845 (2008).

53. Sriwongsitanon, N. *et al.* Comparing the Normalized Difference Infrared Index (NDII) with root zone storage in a lumped conceptual model. *Hydrol. Earth Syst. Sci.* **20**(8), 3361–3377 (2016).
54. Chatterjee, S., Hadi, A. S. *Regression Analysis by Example* (Wiley, 2013).
55. Taghizadeh-Mehrjardi, R. *et al.* Improving the spatial prediction of soil salinity in arid regions using wavelet transformation and support vector regression models. *Geoderma* **383**, 114793 (2021).
56. Tang, Y., Wang, Z. & Zhang, T. Soil salinity estimation in Shule River Basin using support vector regression model. *Land Degrad. Dev.* **34**(13), 4094–4108 (2023).
57. Wu, W. *et al.* Soil salinity prediction and mapping by machine learning regression in Central Mesopotamia, Iraq. *Land Degrad. Dev.* **29**, 4005–4014 (2018).
58. Aksoy, S. *et al.* Assessing the performance of machine learning algorithms for soil salinity mapping in Google Earth Engine platform using Sentinel-2A and Landsat-8 OLI data. *Adv. Space Res.* **69**(2), 1072–1086 (2022).
59. Li, J. *et al.* Comparing machine learning algorithms for soil salinity mapping using topographic factors and sentinel-1/2 data: A case study in the yellow river delta of China. *Remote Sens.* **15**(9), 2332 (2023).
60. Yin, H. *et al.* Synergistic estimation of soil salinity based on Sentinel-1 image texture and Sentinel-2 salinity spectral indices. *J. Appl. Rem. Sens.* **17**(1), 018502 (2023).
61. Andrade Foronda, D. & Colinet, G. Prediction of soil salinity/sodicity and salt-affected soil classes from soluble salt ions using machine learning algorithms. *Soil Syst.* **7**(2), 47 (2023).
62. Chakraborty, S., Elahi, F. Comparison of Soil salinity prediction by Machine Learning algorithms in coastal areas of Bangladesh. *Authorea Preprints* (2023).
63. Shahare, Y. *et al.* A comprehensive analysis of machine learning-based assessment and prediction of soil enzyme activity. *Agriculture* **13**(7), 1323 (2023).
64. Liu, Q. *et al.* Estimation of soil moisture using multi-source remote sensing and machine learning algorithms in farming land of Northern China. *Remote Sens.* **15**(17), 4214 (2023).
65. Jiang, Z. *et al.* Simulating soil salinity dynamics, cotton yield and evapotranspiration under drip irrigation by ensemble machine learning. *Front. Plant Sci.* **14** (2023).
66. Abedi, F. *et al.* Salt dome related soil salinity in southern Iran: Prediction and mapping with averaging machine learning models. *Land Degrad. Dev.* **32**(3), 1540–1554 (2021).
67. Zhou, Y. *et al.* Global soil salinity prediction by open soil vis-NIR spectral library. *Remote Sens.* **14**(21), 5627 (2022).
68. Ma, G. *et al.* Digital mapping of soil salinization based on Sentinel-1 and Sentinel-2 data combined with machine learning algorithms. *Reg. Sustain.* **2**(2), 177–188 (2021).
69. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
70. Probst, P., Wright, M. N. & Boulesteix, A.-L. Hyperparameters and tuning strategies for random forest. *WIREs Data Min. Knowl. Discov.* **9**(3), e1301 (2019).
71. Bardenet, R. *et al.* Collaborative hyperparameter tuning. In *International Conference on Machine Learning* (PMLR, 2013).
72. Wistuba, M., Schilling, N. & Schmidt-Thieme, L. Two-stage transfer surrogate model for automatic hyperparameter optimization. In *European Conference on Machine Learning and Knowledge Discovery in Databases* Vol. 9851 199–214 (Springer, 2016).
73. Hutter, F. *Automated Configuration of Algorithms for Solving Hard Computational Problems* (University of British Columbia, 2009).
74. Head, T. *et al.* Scikit-optimize/scikit-optimize: v0. 8.1. Zenodo (2020).
75. Myles, A. J. *et al.* An introduction to decision tree modeling. *J. Chemometr.* **18**(6), 275–285 (2004).
76. Russell, S. J. *Artificial Intelligence a Modern Approach* (Pearson Education, Inc, 2010).
77. Zhang, W. *et al.* Assessment of pile drivability using random forest regression and multivariate adaptive regression splines. *Georisk. Assess. Manag. Risk Eng. Syst. Geohazards.* **15**(1), 27–40 (2021).
78. Schonlau, M. & Zou, R. Y. The random forest algorithm for statistical learning. *Stata J.* **20**(1), 3–29 (2020).
79. Ho, T. K. Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition* (IEEE, 1995).
80. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
81. Jaiswal, J. K., Samikannu, R. Application of Random forest algorithm on feature subset selection and classification and regression. In *2017 World Congress on Computing and Communication Technologies (WCCCT)*. (2017).
82. Boser, B. E., Guyon, I. M., Vapnik, V. N. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory* (1992).
83. Vapnik, V. *The Nature of Statistical Learning Theory*. (Springer Science & Business Media, 1999).
84. Kurani, A. *et al.* A comprehensive comparative study of Artificial Neural Network (ANN) and Support Vector Machines (SVM) on stock forecasting. *Ann. Data Sci.* **10**(1), 183–208 (2023).
85. Zhang, F., O'Donnell, L. J. Chapter 7—Support vector regression. In *Machine Learning* (eds. Mechelli, A., Vieira, S.) 123–140 (Academic Press, 2020).
86. Bishop, C. M., Nasrabadi, N. M. *Pattern Recognition and Machine Learning*, vol. 4. (Springer, 2006).
87. Chen, T., Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016).
88. Friedman, J., Hastie, T. & Tibshirani, R. Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors). *Ann. Stat.* **28**(2), 337–407 (2000).
89. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
90. Ma, M. *et al.* XGBoost-based method for flash flood risk assessment. *J. Hydrol.* **598**, 126382 (2021).
91. Peng, Y. *et al.* Automated glacier extraction using a Transformer based deep learning approach from multi-sensor remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **202**, 303–313 (2023).
92. Zhou, J. Y. *et al.* Prediction of hepatic inflammation in chronic hepatitis B patients with a random forest-backward feature elimination algorithm. *World J. Gastroenterol.* **27**(21), 2910–2920 (2021).
93. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**(03), 90–95 (2007).
94. Khosravani, P. *et al.* Digital mapping to extrapolate the selected soil fertility attributes in calcareous soils of a semiarid region in Iran. *J. Soils Sediment.* **23**, 4032–4054 (2023).
95. Mzid, N. *et al.* Salinity properties retrieval from Sentinel-2 satellite data and machine learning algorithms. *Agronomy* **13**(3), 716 (2023).
96. Wu, W. *et al.* Soil salinity prediction and mapping by machine learning regression in Central Mesopotamia, Iraq. *Land Degrad. Dev.* **29**(11), 4005–4014 (2018).
97. Merembayev, T. *et al.* Soil salinity classification using machine learning algorithms and radar data in the case from the South of Kazakhstan. *J. Ecol. Eng.* **23**(10), 61–67 (2022).
98. Ding, J.-L., Wu, M.-C. & Tiyyip, T. Study on Soil salinization information in arid region using remote sensing technique. *Agric. Sci. China* **10**(3), 404–411 (2011).
99. Qi-sheng, H., Chun-xiang, C. & Tiyyip, T. Study on the extraction of saline soil information in arid area based on multiple source data. *Remote Sens. Technol. Appl.* **25**(2), 209–215 (2010).

100. Fu, H. *et al.* Land salinization classification method using Landsat TM images in Western Jilin Province of China. In *SPIE Optical Engineering + Applications*, vol. 9220 (SPIE, 2014).
101. Breiman, L. Bagging predictors. *Mach. Learn.* **24**, 123–140 (1996).
102. Haq, Y. U. *et al.* Identification of soil type in Pakistan using remote sensing and machine learning. *PeerJ Comput. Sci.* **8**, e1109 (2022).
103. Merembayev, T., Amirgaliyev, Y., Saurov, S. & Wójcik, W. Soil Salinity Classification Using Machine Learning Algorithms and Radar Data in the Case from the South of Kazakhstan. *J. Ecol. Eng.* **23**(10), 61–67 (2022).
104. Yahiaoui, I., Bradaï, A., Douaoui, A. & Abdennour, M. A. Performance of random forest and buffer analysis of Sentinel-2 data for modelling soil salinity in the Lower-Chelif plain (Algeria). *Int. J. Remot. Sens.* **42**(1), 148–171 (2021).
105. Schulz, K., Hänsch, R. & Sörgel, U. Machine learning methods for remote sensing applications: an overview. In *Proc. SPIE 10790, Earth Resources and Environmental Remote Sensing/GIS Applications IX, 1079002*. <https://doi.org/10.1117/12.2503653>
106. Salcedo, F. P. *et al.* Use of remote sensing to evaluate the effects of environmental factors on soil salinity in a semi-arid area. *Sci. Total Environ.* **815**, 152524 (2022).
107. Shrestha, R. P., Qasim, S. & Bachri, S. Investigating remote sensing properties for soil salinity mapping: A case study in Korat province of Thailand. *Environ. Chall.* **5**, 100290 (2021).
108. Wang, N. *et al.* Integrating remote sensing and landscape characteristics to estimate soil salinity using machine learning methods: A case study from Southern Xinjiang, China. *Remote Sens.* **12**(24), 4118 (2020).
109. Cao, X. *et al.* Multidimensional soil salinity data mining and evaluation from different satellites. *Sci. Total Environ.* **846**, 157416 (2022).
110. Ostad-Ali-Askari, K. & Shayan, M. Subsurface drain spacing in the unsteady conditions by HYDRUS-3D and artificial neural networks. *Arab. J. Geosci.* **14**, 1–14 (2021).
111. Gopalakrishnan, T. & Kumar, L. Linking long-term changes in soil salinity to paddy land abandonment in Jaffna Peninsula, Sri Lanka. *Agriculture* **11**(3), 211 (2021).
112. Singh, A. Soil salinization management for sustainable development: A review. *J. Environ. Manag.* **277**, 111383 (2021).
113. Mardanifar, M. *et al.* Evaluating the drainage process of agricultural lands in Golestan province based on agricultural drainage reuse. *Nat. Ecosyst. Iran* **13**(3), 1–13 (2022).

Author contributions

S.B.: Writing the original draft, sampling and investigation, methodology, analysis, and software. S.A.M.-N.: Writing-review & editing, project administration. S.M.: Visualizing, writing-review & editing. A.L.: Conceptualization, methodology, writing-review & editing, supervision, project administration.

Funding

The authors receive no funding.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-60033-6>.

Correspondence and requests for materials should be addressed to S.B. or S.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024