



Comparative assessment of satellite images spectral characteristics in identifying the different levels of soil salinization using machine learning techniques in Google Earth Engine

S. Kabiraj¹ · M. Jayanthi¹ · S. Vijayakumar¹ · M. Duraisamy¹

Received: 10 June 2022 / Accepted: 10 September 2022 / Published online: 17 October 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Assessing salt-affected land (SAL) remains a major challenge worldwide, especially in developing countries due to limited data availability. The development of remote sensing (RS) digital satellite data and their spectral characteristics pave the way to assess soil salinity. A comparative assessment has been carried out between two different spatial resolution satellite images to find out the most suitable satellite image for assessing soil salinity. Sentinel-2 and Landsat-8 data of 2020 were imported from Google Earth Engine (GEE) data catalog. Spectral indices of vegetation (Normalized Difference Vegetation Indices, Enhanced Vegetation Indices, Soil Adjusted Vegetation Indices, Generalized Difference Vegetation Index) and soil salinity (Normalized Difference Salinity Index, Canopy Response Salinity Index, Salinity Index (SI), SI-I, SI-II, SI-III, Salinity index 1 (SI1), SI2, SI3, SI4, SI5) were calculated. The Random Forest model was used for the analysis. Collected soil samples' electrical conductivity values were used to train and validate the model. 70% of the soil samples were used for model training and 30% for validation. The result shows Sentinel-2 is more capable of detecting saline soil regions compared to Landsat-8. Based on 5-fold cross-validation, the overall accuracy of Sentinel-2 and Landsat-8 were assessed as 98.31% and 97.19%. The total SAL area using Landsat-8 and Sentinel-2 were 12,446 ha and 13,225 ha respectively. The study showed the effectiveness of RS techniques to detect the SALs in different spatial resolutions, which can help to evaluate unproductive land and its management at the state or regional level for the creation of alternative livelihood options.

Keywords Google Earth Engine · Random Forest · Soil salinity · Spectral indices

Introduction

Soil salinization is a major ecological and socioeconomic problem that adversely affects global food productivity and projected to be more intense due to the impacts of climate change in coastal countries (Mukhopadhyay et al. 2021). Apart from global warming and associated sea-level rise,

poor water management is an important factor in soil salinization in arid areas (Li et al. 2018). Researchers have indicated the area under salt-affected land (SAL) for different regions and also increasing globally, but yet there is a lack of information on the global SAL extent. As per FAO (2021) report on salt-affected soil, almost 73% of global land is affected due to soil salinity, that includes 424 mha of topsoil between 0 and 30 cm, and 833mha of subsoil between 30 and 100 cm. Of the different continents, Asia has the largest SAL followed by East and North Africa, and Europe has the least extent of SAL.

Soil salinization emerges as a key environmental constraint hampering soil productivity, agricultural sustainability, and food security, particularly in arid and semi-arid regions of the world (Cuevas et al. 2019). Soil fertility is defined by the ability to supply plant nutrients in sufficient levels for the agricultural crops, which in turn is affected

Communicated by H. Babaie

✉ M. Jayanthi
m.jayanthi@icar.gov.in
S. Kabiraj
sabyasachi.kabiraj2@gmail.com

¹ ICAR-Central Institute of Brackishwater Aquaculture, Chennai, India

by soil salinization. Annual losses of the agricultural sector as 27.3 million US dollars, either due to gradually reduced productivity of salinized lands or their disuse because of degraded soil quality (Qadir et al. 2006). The level of soil salinity will indicate the health of the soil, thus allow us to plan for the management measures, alternative crops, and systems. The conventional measure of soil salinity has been carried out by collecting the soil samples from the field and measuring through laboratory analysis in terms of electrical conductivity (Rhoades et al. 1999). It is not only time-consuming but does not permit us to assess the salinity at a larger scale at a periodical level. Large-scale studies on soil salinization are very much required to assess the soil quality at an administrative unit level for planning and devising policy framework.

Remote sensing data and methods are extensively being used to map soil salinity. Extensive research in soil salinity mapping has been carried out over the last three decades using satellite imagery. Advances in the availability of satellite data and analytical capabilities have paved the way for precise and timely assessment of soil salinity across different places and times. Numerous spatial models were tested to assess saline soils based on topographic information, climatic conditions, land use information, etc. (Aksoy et al. 2022; Wei et al. 2020). Several spectral indices of vegetation e.g. Normalized Differential Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), General Differential Vegetation Index (GDVI); salinity indices such as Normalized Difference Salinity Index (NDSI), Salinity Index (SI), SI 1, SI 2, SI 3, SI-I, Canopy Salinity Index (CRSI) (Aksoy et al. 2022; Taillie et al. 2019).

Remote sensing data with high and medium spatial resolution recently has become available. Various studies have found high potential of Sentinel-2 images for SAL modelling (Davis et al. 2019). The Sentinel-2 satellite provides high spatial, spectral, and temporal resolution images with multi-spectral bands. These images offer several aspects that can help to identify soil salinity more precisely (Banari et al. 2018). For SAL modelling and mapping, researchers have also commonly used Landsat satellite series images (Scudiero et al. 2014; Nguyen et al. 2020). In both methods (Sentinel-2 and Landsat satellite images), SAL was identified and mapped using soil salinity indices based on various band reflectance values (McBratney et al. 2003). As a result of its smoothness, soil that has a surface salt layer has high reflectance values in the visible and near-infrared bands (Elnaggar and Noller 2009; Meier et al. 2018). The spectral reflectance of the image bands can be influenced by several variables, including surface roughness, soil colour, wetness, and physical and chemical characteristics (Asfaw et al. 2018). The spectral reflectance of saline soil with a puffy black surface crust is lower (Davis et al. 2019).

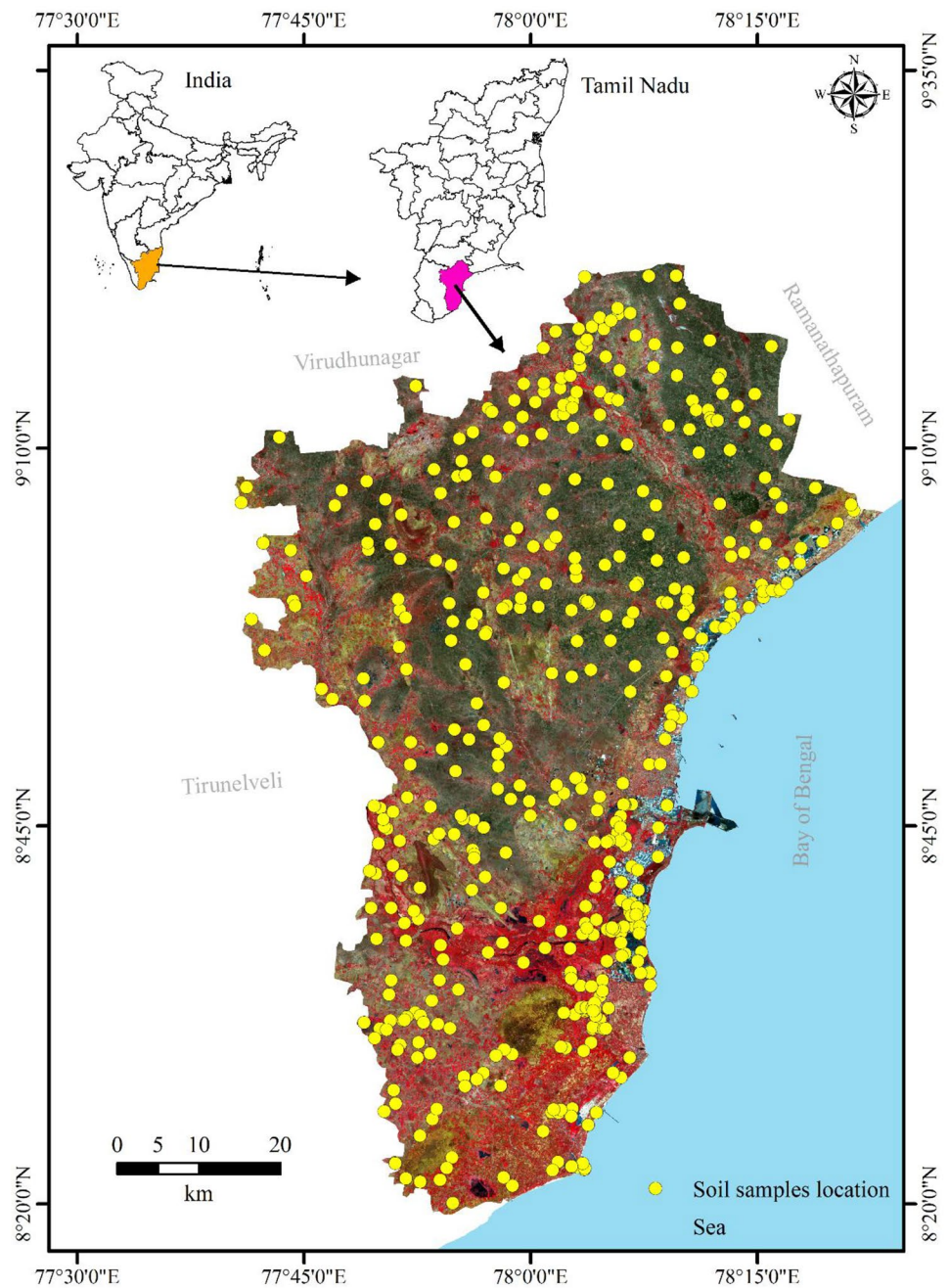
Different modeling techniques used in these studies include partial least squares regression (PLSR) (Fan et al. 2015; Bai et al. 2018), Multiple Linear Regression (MLR) (Li et al. 2019, 2021), Support Vector Regression (SVR) (Wu et al. 2018), Artificial Neural Networks (ANN) (Farifteh et al. 2007), Random Forest (RF) (Nabiollahi et al. 2021). A large number of studies have shown that machine learning methods (ANN, SVM, RF) achieve high predictive accuracy compared to other methods, especially RF (Wu et al. 2018; Wang et al. 2021; Aksoy et al. 2022). The advantages of RF application in remote sensing are it works effectively over a large number of datasets, it can control the large number of variables without deletion, it evaluates the importance of each variable in the classification, and it produces an accurate internal generalization error estimate (OOB error), calculates distances between pairs that can be used to find outliers, it is reasonably resistant to noise and outliers, and it requires less computing power than other tree ensemble approaches (e.g. Boosting).

India has 2,956 Mha of saline soil and 3.77 Mha of sodic soil (Arora and Sharma 2017). 16.84 million tonnes of agricultural production annually lose due to soil salinity in India (Mandal et al. 2018). The projected population growth shows India will need about 311 and 350 million tonnes of grains in 2030 and 2050, respectively, to feed about 1.43 and 1.8 billion people, respectively (Kumar et al. 2016). Estimations show that additionally each year 10% of the land becomes salinized and around 50% of the arable land will become saline by 2050 (Kumar and Sharma 2020). Climate change may enhance the level of soil salinity especially in densely populated nations due to sea-level rise and make it more vulnerable.

Soil salinity has been assessed by Indian researchers using both traditional and remote sensing data (Narjary et al. 2019; Bhardwaj et al. 2019; Kumar et al. 2015), but these studies were in bits and pieces, mainly focused on spectral indices, followed by regression analysis. Few studies has focused on machine learning application in detecting soil salinity using different spatial resolution satellite images. Several studies on the comparative assessment of different spatial resolution satellite images such as Landsat TM (30 m), SPOT (5 m), IKONOS (1 m) in assessing soil salinity showed that IKONOS data provide much better prediction accuracy than SPOT and Landsat TM (Dwivedi et al. 2008; Eldeiry and Garcia 2008).

Growing SAL is a major environmental issue in developing and densely populated regions like India, as land availability is the limiting factor. Though researchers have used different automated techniques, regional studies in relation to tropical and semi-arid regions are meagre. The global assessment of SAL provides an overall picture (Ivushkin et al. 2019), however, does not give the exact details required

Fig. 1 Study area map showing sample point location in Thoothukudi District, Tamil Nadu, India



for management at a regional scale, due to limited sample points. This study aims to assess the SAL changing pattern using different resolution images in the semi-arid tropical region of the densely populated coastal District.

Materials and methods

Study area

The study area of the present study is the Thoothukudi district of Tamil Nadu, India. Extended between 8° 19' N and 9° 20' N latitude, 77° 40' E, and 78° 10' E longitude, coastal length and spatial extent of 163.5 km and 4621sq.km respectively (Fig. 1). The Gulf of Mannar is an important ecological important site in the district. It is home to 3,600 marine species of flora and fauna. The district is the largest

Table 1 Spectral bands of Sentinel-2 used in the present study

Satellite Sensor	Band used	Spatial resolution	Date of acquisition	Tile Numbers
Sentinel 2 A MSI	B2 (Blue), B3 (Green), B4 (Red), B8 (Near infra-red)	10	03/05/2020	T43PGK, T43PGL, T43PHK, T43PHL, T44PKQ, T44PKR
			08/05/2020	T43PHL, T44PKR
			23/05/2020	T43PGK, T43PGL, T43PHK, T43PHL, T44PKQ, T44PKR
			28/05/2020	T43PGK, T43PHK, T44PKQ
			12/06/2020	T43PGK, T44PKR
			17/06/2020	T43PHL, T43PKR
			27/06/2020	T43PHK, T43PHL, T43PKQ, T44PKR
			02/07/2020	T44PKR
			07/07/2020	T43PGK, T43PGL, T43PHK, T43PHL, T44PKQ, T44PKR
			12/07/2020	T43PGL, T43PHL, T44PKR
			27/07/2020	T43PGK, T43PGL, T43PHL, T44PKR
			01/08/2020	T43PHL, T44PKR
			11/08/2020	T42PGK, T43PHK, T43PHL, T44PKQ, T44PKR
			16/08/2020	T44PKR
			Landsat-8 OLI	B2 (Blue), B3 (Green), B4 (Red), B5 (Near infra-red), B6 (Shortwave infra-red1), B7 (Shortwave infra-red 2)

salt producer in the state, producing 70% of the total salt of

the state and 30% of the country's salt requirements. The district experiencing the semi-arid tropical type of climate with many climate change impacts such as sea-level rise (Sheikh 2011), seawater intrusion, shoreline change (Satheskumar et al. 2021), increase in salt-affected land (Selvam et al. 2013) and meteorological drought (Sheik and Chandrasekar 2011). The annual mean rainfall of the district is 661.6 m with a maximum and minimum temperature ranging from 29.5 to 40.5 °C and 18.4 to 26.7 °C, respectively.

A total of 593 soil samples were used in the present study. Of 593 soil samples, 258 Soil samples from different locations were collected from 30 July to 5 August 2020 for EC analysis (Fig. 1). The remaining samples were taken from the Indian Council of Agricultural soil database. Each soil sample's geographic coordinate was measured and stored by GPS TDC 600 with less than 2 m positional accuracies. At each location of soil samples, four topsoils from the four corners of the quadrant were collected and mixed well. The samples were dried completely in the air and passed through a 2 mm sieve to remove non-soil material. Soil leachate was prepared at a 1:2.5 soil: water ratio, after that soil EC was determined with a digital multiparameter meter (Systronics EC - TDS meter 308) at room temperature 25 °C. EC values were classified as non-saline (<2ds/m), slightly saline (2–4 ds/m), moderately saline (>4, <8 ds/m), highly saline (8–16 ds/m) and extremely saline. (>16) (Brown et al. 1954; Aksoy et al. 2022). After the EC analysis of all soil samples in the laboratory, a shapefile was created using ArcGIS 10.5, the EC value, stored location detail, and corresponding salinity classes were assigned to each point in the attribute table of the shapefile.

Data used

This study used Sentinel-2 and Landsat-8 (Table 1) of 2020 satellite imagery from May to August month with less than 40% cloud cover available at the Google Earth Engine (GEE) platform (Table 1). GEE have an extensive geospatial dataset, including Sentinel, Moderate Resolution Imaging Spectroradiometer (MODIS), ASTER, Landsat, and SRTM imagery, as well as analytical abilities to handle a large number of dataset. Topographic maps (58 K/3,4,7,8; 58 H/9,13–15; 58 L/1-3.5; 58G/12,15,16) from the Survey of India (SOI) were utilized to demarcate the features like administrative boundaries, rivers, etc.

Data Processing

Selected all satellite images were imported from the GEE data catalog to the GEE code editor section. Selected image collections were then filtered between May to August month, 2020 using `.filterDate()` script in the GEE code

Table 2 Spectral indices of vegetation, soil salinity, and topographical attributes used in the model

Category	Sl.no	Spectral Indices	Formula	References
Vegetation Spectral Indices	1.	Normalized Difference Vegetation Index (NDVI)	$(\text{NIR}-\text{Red})/(\text{NIR}+\text{Red})$ (1)	Khan et al. 2005
	2.	Enhanced Vegetation Index (EVI)	$2.5 \times (\text{NIR}-\text{Red})/(\text{NIR}+6 \times \text{Red}-7.5 \times \text{Blue}+0.5)$ (2)	Huete et al. 2002
	3.	Soil Adjusted Vegetation Index (SAVI)	$[(\text{NIR}-\text{Red})/(\text{NIR}+\text{Red}+0.5)] \times (1+0.5)$ (3)	Huete et al. 1998
	4.	Generalized difference vegetation index (GDVI)	$(\text{NIR}^2-\text{Red}^2)/(\text{NIR}^2+\text{Red}^2)$ (4)	Wu et al. 2014
	Salinity spectral Indices	5.	Normalized difference salinity Index (NDSI)	$(\text{Red}-\text{NIR})/(\text{Red}+\text{NIR})$ (5)
6.		Canopy Response Salinity Index (CRSI)	$\sqrt{\frac{(\text{NIR}*\text{Red})-(\text{Green}*\text{Blue})}{(\text{NIR}*\text{Red})+(\text{Green}*\text{Blue})}}$ (6)	Scudiero et al. 2014
7.		Salinity index (SI)	$(\text{Blue}+\text{Red})^{0.5}$ (7)	Douaoui et al. 2006
8.		Salinity index I (SI1)	$(\text{Green} \times \text{Red})^{0.5}$ (8)	
9.		Salinity index 2 (SI2)	$[(\text{Green})^2+(\text{Red})^2+(\text{NIR})^2]^{0.5}$ (9)	
10.		Salinity index 3 (SI3)	$[(\text{Red})^2+(\text{Green})^2]^{0.5}$ (10)	
11.		Salinity index I (S-I)	Blue/Red (11)	Khan et al. 2005
12.		Salinity index II (S-II)	$(\text{Blue}-\text{Red})/(\text{Blue}+\text{Red})$ (12)	Khan et al. 2005
13.		Salinity index III (S-III)	$(\text{Green} \times \text{Red})/\text{Blue}$ (13)	Khan et al. 2005
14.		Salinity index-IV (S-IV)	$(\text{Blue} \times \text{Red})/\text{Green}$ (14)	Khan et al. 2005
15.		Salinity index V (S-V)	$(\text{Red} \times \text{NIR})/\text{Green}$ (15)	

editor. A total of 7 images of Landsat-8 and 47 images of Sentinel-2 were used. The study region shapefile and soil sample shapefile were then uploaded through ‘assets’ section and imported to the code editor in GEE and using the `.filterBounds()` script, clipped the satellite image collection to the study area extent. Finally, using `.median()` script, calculated the median of all image collections falling under the May to August month 2020 to reduce the data volume and for faster analysis (Carrasco et al. 2019).

Selection of predictors

The spectral indices are important to assess soil salinity in arid and semi-arid areas (Aksoy et al. 2022). By considering the various studies and nature of the study, commonly used soil salinity indicators such as vegetation indices, and salinity indices were selected for the present study to generate a powerful combination in assessing soil salinity. Various spectral bands of Sentinel-2 such as B2, B3, B4, and B8 of 10 m spatial resolution and spectral bands of Landsat-8 such as B2, B3, B4, B5, B6, and B7 of 30 m spatial resolution were selected for the study. Other spectral bands of Sentinel-2 and Landsat-8 were not considered in the analysis to make the model prediction output of 10 and 30 m. Various spectral bands, vegetation indices, and soil salinity indices (Tables 1 and 2) were combinedly named as predictors. `.select()` expression were used to select particular spectral band and `.expression()` script is used to calculate all predictors. A total of 19 predictors (fifteen indices and four spectral bands) using Sentinel-2 (Figs. 3) and 21 predictors (fifteen indices and 6 spectral bands) using Landsat-8 and collected soil sample’s salinity class were used to assess soil

salinity. `.addBands()` expression were used to add particular predictors into the predictor list to predict soil salinity.

Random forest modelling

The RF model was used to predict soil salinity in the present study. 70% of the collected soil sample’s salinity category was used to train the model and the remaining 30% were used for validation purposes using the `.filter()` expression. 593 soil samples were used in the model. 178 soil samples were used for training and the remaining samples were used for the model validation. All the predictors and collected soil sample’s salinity class have been integrated into the model and using the ‘`ee.Classifier.smileRandomForest()`’ expression the model was performed. The hyperparameter was used to find the optimum number of trees and bag fraction with the highest training accuracy. Under different settings ranging from 1 to 500 number of trees at an interval of 10 and bag-fraction ranging from 0.1 to 0.9 with an interval of 0.1 were calculated to get the optimum number of trees and bag-fraction with the highest training accuracy. Through the bag-fraction method, unused samples can participate in making the decision tree creation process to evaluate each tree’s accuracy to improve model performance by considering the mean accuracy value of all trees. The model was trained using `.sampleRegions()` expression. All predictors and their importance in the model prediction were generated using “`ee.Feature(null, ee.Dictionary().get(‘importance’))`” expression.

Using `.confusionMatrix()` expression confusion matrix was calculated and `.accuracy()` expression the overall training accuracy was calculated. Similarly, validation accuracy

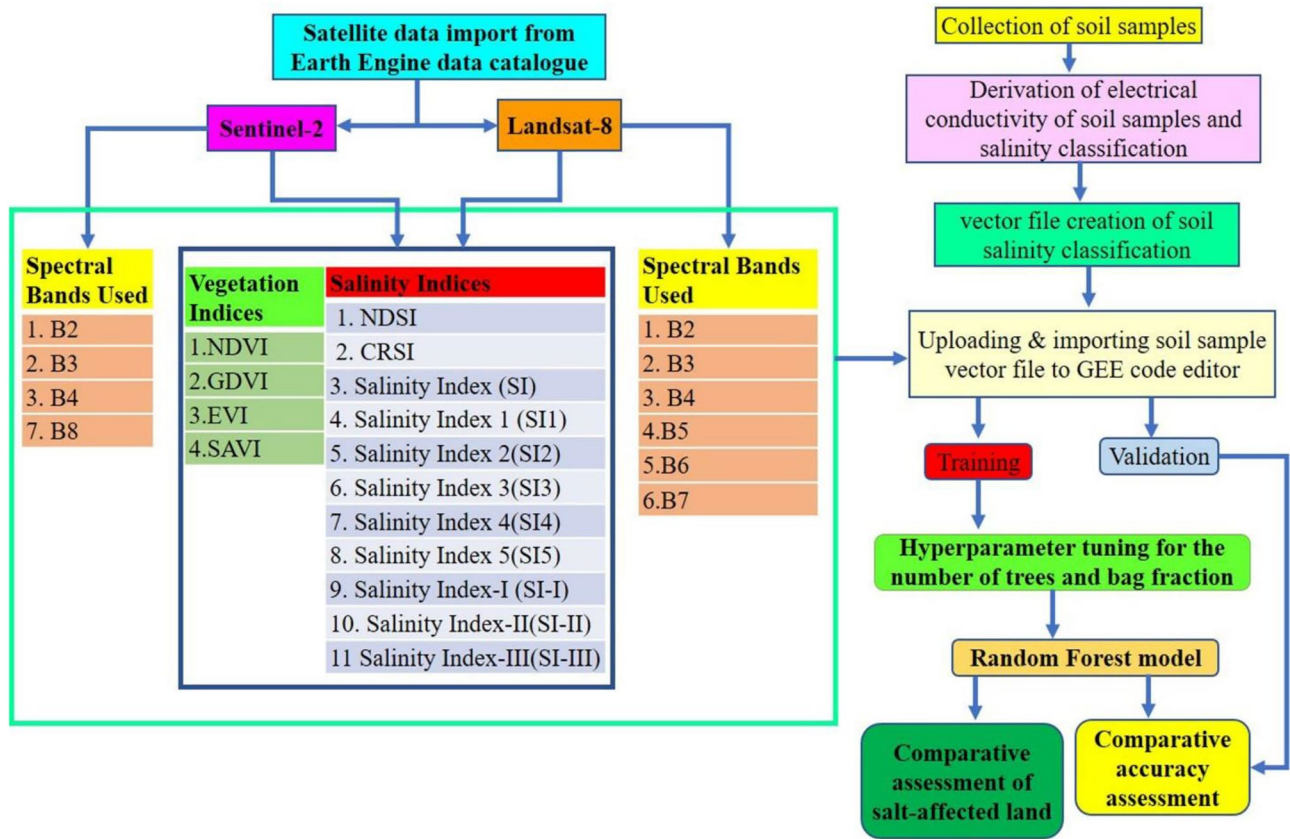


Fig. 2 Methodology flow chart of automated delineation of salt-affected land using Sentinel-2 and Landsat-8 satellite image

for classified raster were calculated using the **.errorMatrix()** and **.accuracy()** expression. 5-fold cross-validation method, training and validation accuracy were calculated to validate the model performance. Very limited amount of points in moderately, highly, and extremely saline regions were selected because of less soil samples in these regions. Area of different classes of soil salinity of the classified raster was calculated using **“ee.Image.pixelArea()”** expression. The detailed flowchart methodology was given in Fig. 2.

593 samples were collected and used for the user’s accuracy and producer’s accuracy. F1 score of each class and micro-averaged, macro-averaged F1 score were calculated using the formula given below:

$$MiP = \frac{\sum_{i=1}^r TP_i}{\sum_{i=1}^r (TP_i + FP_i)} \tag{1}$$

$$MiR = \frac{\sum_{i=1}^r TP_i}{\sum_{i=1}^r (TP_i + FN_i)} \tag{2}$$

Both Mip and MiR, the denominator is the sum of all the elements (diagonal and off-diagonal) of the confusion matrix. TP_i is the i -th diagonal element, FP_i is the total of

off-diagonal elements of the i -th row and FN_i is the total of off-diagonal elements of the i -th column. Finally, the micro-averaged F_1 (MiF_1) score is computed as the harmonic mean of these quantities:

$$MiF_1 = 2 \times \frac{MiP_i \times MiR_i}{MiP_i + MiR_i} \tag{3}$$

To calculate the macro-averaged (MaF1), first need to evaluate precision (P_i) and recall (R_i) using the formula given below within each class, $I = 1, \dots, r$:

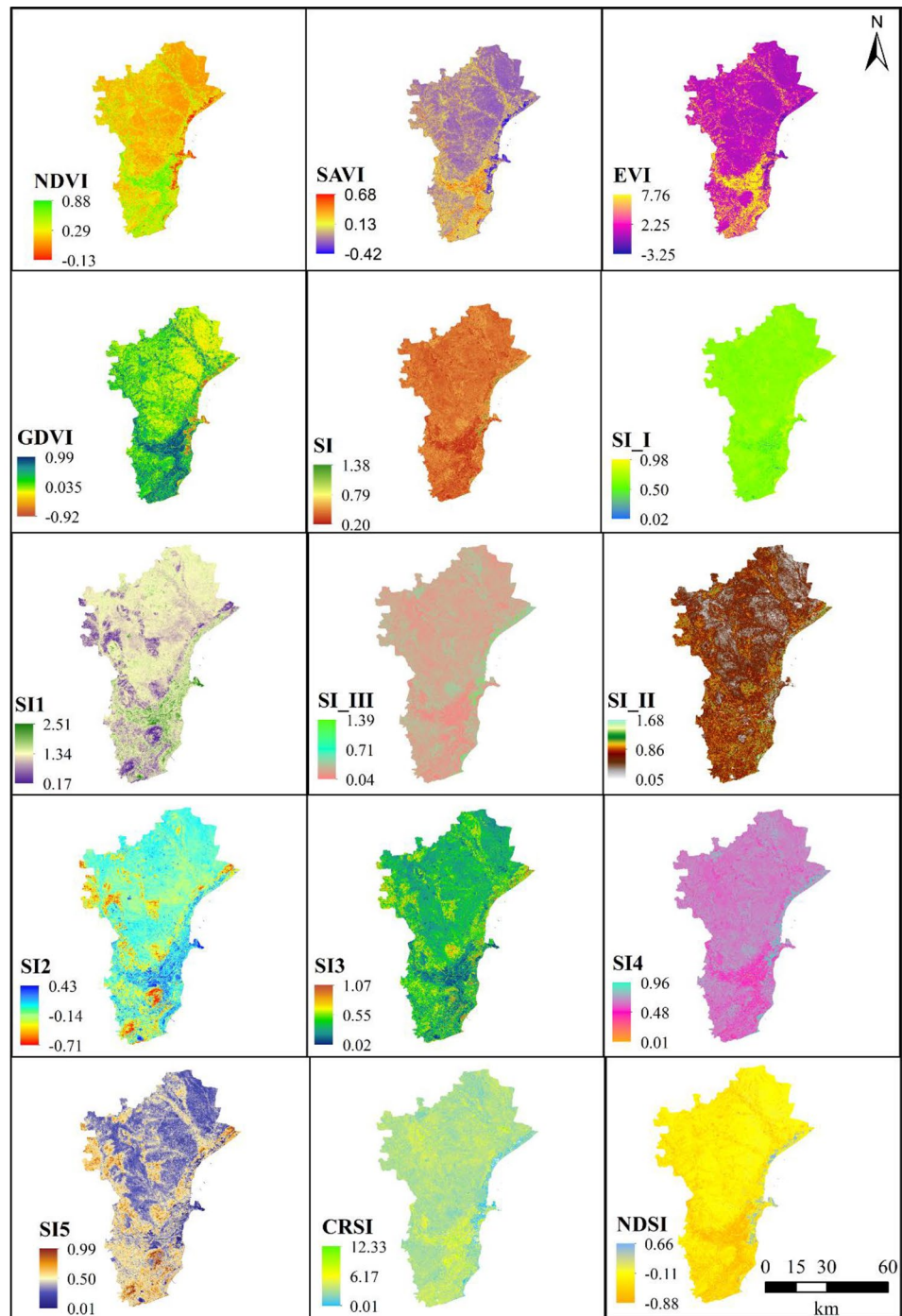
$$P_i = \frac{TP_i}{(TP_i + FP_i)} \tag{4}$$

$$R_i = \frac{TP_i}{(TP_i + FN_i)} \tag{5}$$

F1 score within each class ($F1_i$) is calculated by using the formula given below

$$F1_i = 2 \times \frac{P_i \times R_i}{P_i + R_i} \tag{6}$$

Fig. 3 Different soil salinity predictors used to derive salt-affected land for Thoothukudi District, India



The macro-averaged F1 (*MaF1*) score is defined as the simple arithmetic mean of *F1_i* :

$$MaF1 = \frac{1}{r} \sum_{i=1}^r F1_i \tag{7}$$

A popular technique for assessing the effectiveness of the RF models is the receiver operating characteristic (ROC)

curve. The true positive rate and the false positive rate for all potential cut-off value are plotted on the ROC curve.

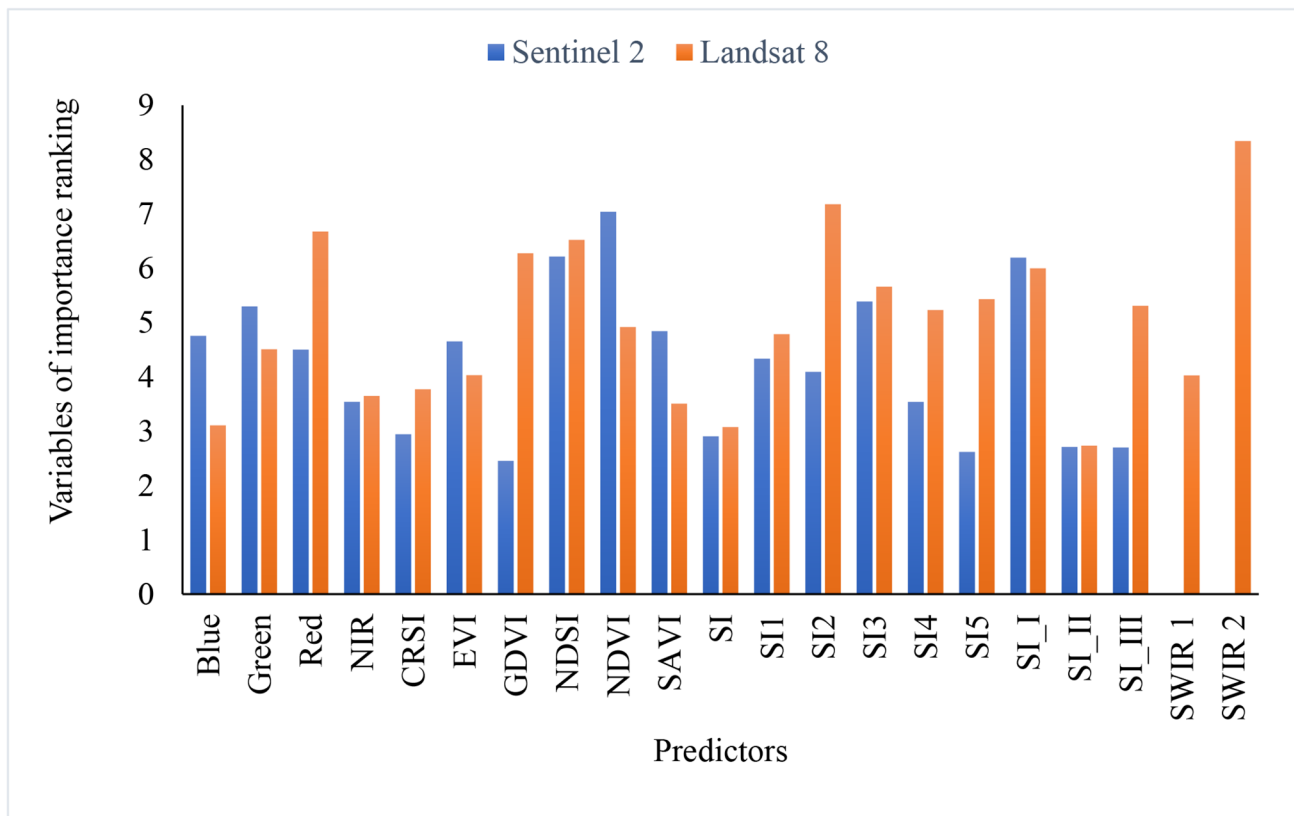


Fig. 4 Variables of importance ranking of model predictors using Sentinel-2 and Landsat-8

Results

Soil sample assessment for electrical conductivity

The EC values of the collected soil varied from 0.02 to 72 ds/m, with an average value of 2.87 ds/m at a depth ranging from 0 to 20 cm. The study area's major part belongs to agricultural land, also the dominance of saltpan close to the coastline. The maximum portion of the saline samples was located near the coastline of the study area.

Machine learning with Random Forest (RF) model

Two parameters, numbers of trees and bag-fraction were used in the hyperparameter tuning process. Result suggests that 20 number of trees with 0.6 bag-fraction has the highest training accuracy of 98.58% using Sentinel-2, whereas 30 number of trees with 0.5 bag-fraction has the highest training accuracy of 97.16% using Landsat-8. Variables of importance (VIMP) ranking were evaluated for the 19 predictors. Among the 19 predictors in the RF model using Sentinel-2, the top 10 important predictors were NDVI, SI5, SAVI, SI1, CRSI, SI2, SI4, GDVI, EVI and SI-I with VIMP scores of 8.87, 8.45, 8.36, 7.98, 7.66, 6.95, 6.55, 6.35, 6.26

and 5.7 respectively (Fig. 4). Among the 21 predictors in the RF model using Landsat-8 satellite image, the top 10 important predictors were CRSI, SI-III, SI-I, SI-II, NDVI, Blue, NDSI, GDVI, SI2 and Green with VIMP score of 8.7, 8.64, 8.11, 7.84, 6.94, 6.92, 6.87, 6.76, 6.75, and 6.36 (Fig. 4) respectively.

Spatial extent of salt-affected lands

Using the RF model of machine learning in the GEE platform SAL of Thoothukudi District was assessed (Fig. 5a and 5b). The estimated total SAL using Sentinel-2 in the Thoothukudi district in 2020 was 13,225 ha, including 7055 ha (Fig. 6) as moderately saline soil, 4070 ha as highly saline soil, and 2100 ha as extremely saline soil region. The overall training and validation accuracy was 98.97% and 98.31% respectively. The user's accuracy of slightly saline, moderately saline and extremely saline was 100% followed by non-saline (97.81%). The producer's accuracy of non-saline and extremely saline was 100%, followed by slightly saline (93.75%), moderately saline (92.85%) and highly saline (88.88%) (Table 3). Whereas using Landsat-8, the total assessed SAL was 12,446 ha, including 5249 ha as moderately saline, 3721 ha as highly saline and

Fig. 5a Salt-affected land of 2020 using Landsat-8 satellite images

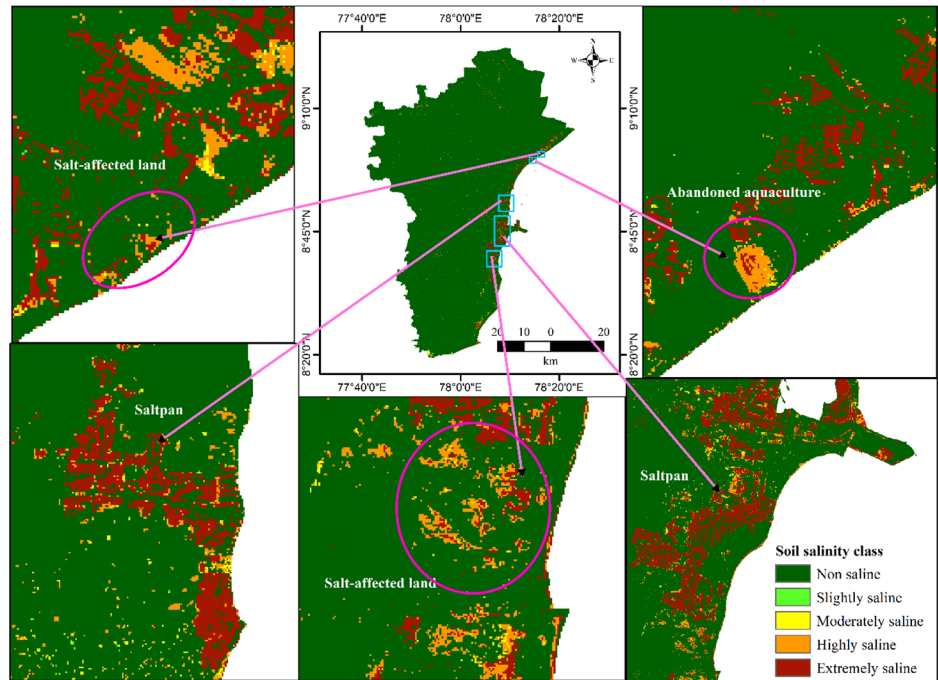
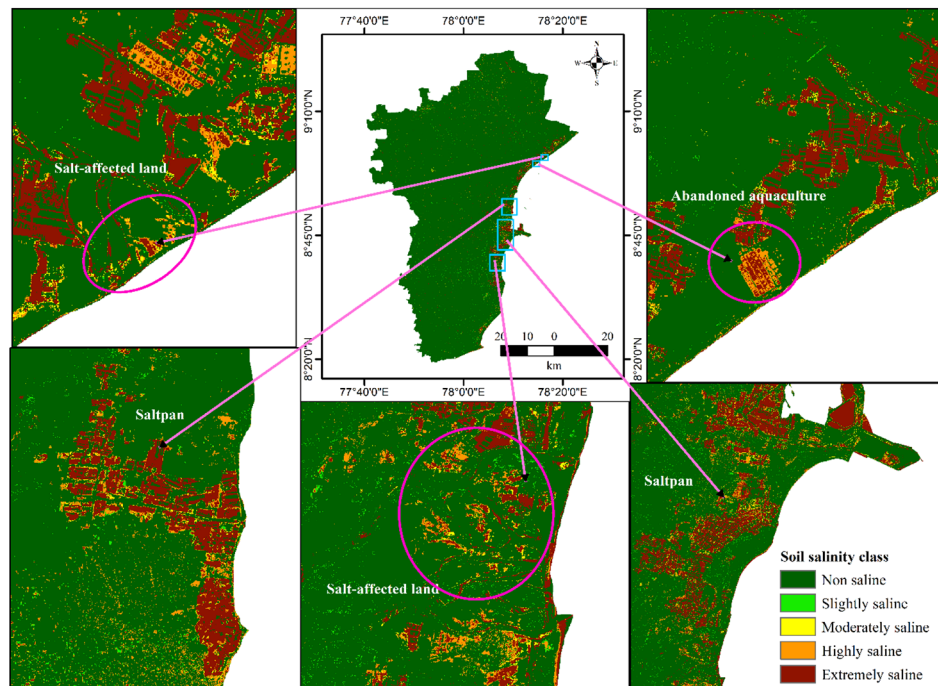


Fig. 5b Salt-affected land in 2020 using Sentinel-2 satellite data



3476 ha as extremely saline (Table 4) with a training and validation accuracy of 97.60% and 97.19% respectively. The user’s accuracy of slightly saline, moderately saline, highly saline and extremely saline was 100%, followed by non-saline soil (97.10) and the producer’s accuracy of non-saline and extremely saline was 100%, followed by highly

saline (88.88%), slightly saline (87.5%), moderately saline (85.71%) (Table 5).

F_1 scores of both Landsat-8 and Sentinel-2 were calculated for non-saline, slightly saline, moderately saline, highly saline and extremely saline. Using the Sentinel-2 image, the highest F_1 score was for extremely saline soil and slightly

Fig. 6 Class wise extension of salt-affected land using Sentinel-2 and Landsat-8 satellite images

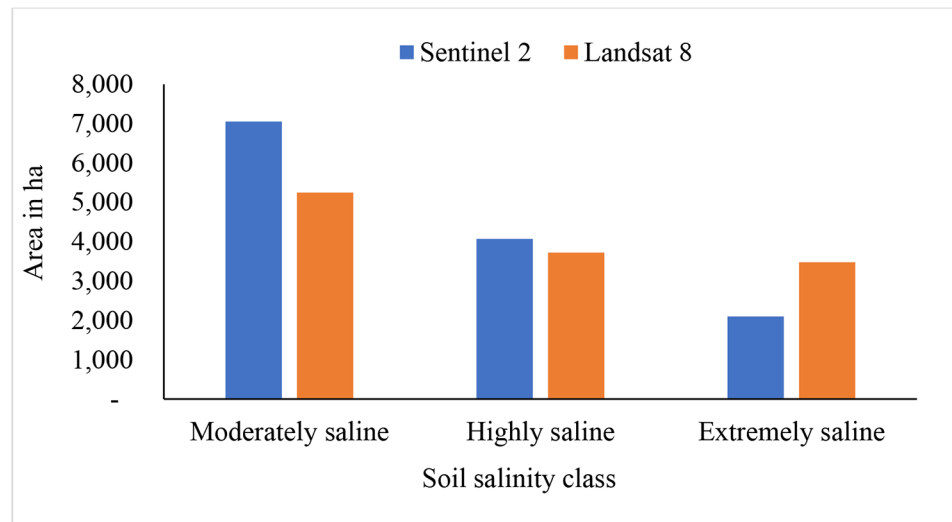


Table 3 Accuracy assessment of different class-wise soil salinity using Sentinel 2 of 2020

Salinity class		Predicted					Total sample	Producers' accuracy (%)
		Non-saline	Slightly saline	Moderately saline	Highly saline	Extremely saline		
Observed	Non-saline	134	0	0	0	0	134	100.00
	Slightly saline	1	15	0	0	0	16	93.75
	Moderately saline	1	0	13	0	0	14	92.85
	Highly saline	1	0	0	8	0	9	88.88
	Extremely saline	0	0	0	0	5	5	100
	Total sample	137	15	13	8	5	178	
	User's accuracy (%)	97.81	100	100	100	100		

Table 4 Spatial distribution of different soil salinity classes

Spatial extent of different class of soil salinity								
Sl.no	Year	Data used	Non-saline	Slightly saline	Moderately saline	Highly saline	Extremely saline	Total salt-affected soil area(ha)
1	2020	Sentinel-2	441,887	4744	7055	4070	2100	13,225
2	2020	Landsat-8 OLI	440,032	2831	5249	3721	3476	12,446

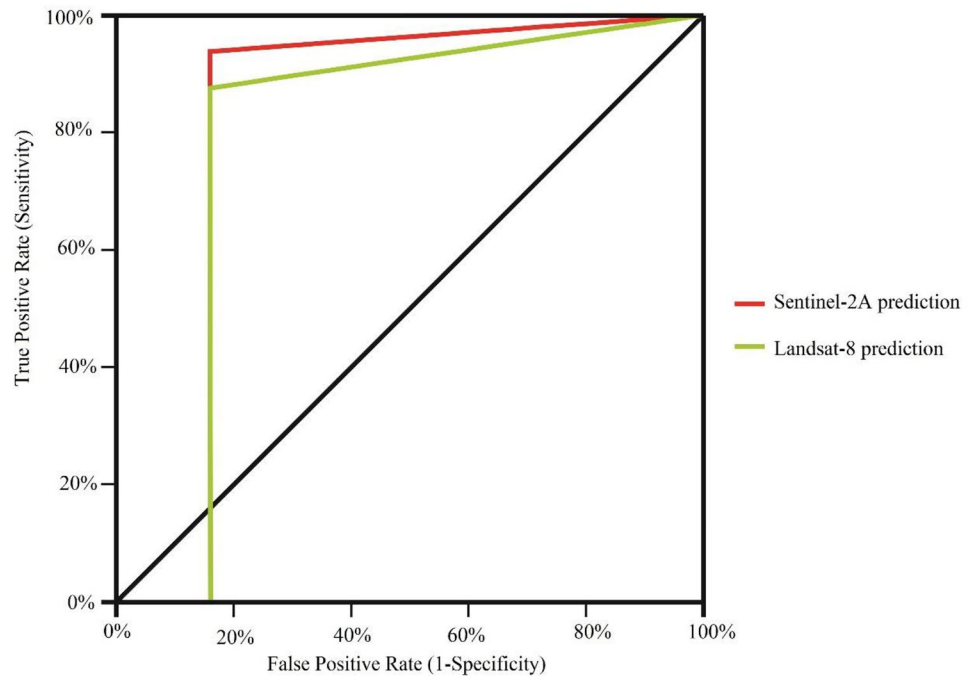
Table 5 Accuracy assessment of different class wise soil salinity using Landsat- 8 of 2020

Salinity class		Predicted					Total sample	Producers' accuracy (%)
		Non-saline	Slightly saline	Moderately saline	Highly saline	Extremely saline		
Observed	Non-saline	134	0	0	0	0	134	100.00
	Slightly saline	2	14	0	0	0	16	87.5
	Moderately saline	2	0	12	0	0	14	85.71
	Highly saline	1	0	0	8	0	9	88.88
	Extremely saline	0	0	0	0	5	5	100
	Total sample	138	14	13	8	5	178	
User's accuracy (%)	97.10	100.00	100	100	100			

saline (1), followed by non-saline (0.99), moderately saline (0.96) and highly saline (0.94). The overall micro-averaged and macro-averaged F_1 scores were 0.98 and 0.97 respectively. Whereas using Landsat-8 the highest F_1 score was for Extremely saline (1), followed by non-saline (0.98),

moderately saline (0.96), highly saline (0.94), and slightly saline (0.87). The overall micro-averaged and macro-averaged F_1 scores were 0.98 and 0.95 respectively.

Fig. 7 Model validation using ROC curve



ROC and AUC for model validation

The RF model performance using two different datasets of Sentinel-2 and Landsat-8 is illustrated in Fig. 7. The figure shows that Sentinel-2 data is more capable of detecting salt-affected regions accurately than Landsat-8. The area under curve (AUC) value using the Sentinel-2 dataset was 0.813, whereas the AUC value using Landsat-8 data was 0.792 (Fig. 7).

Discussion

Multispectral optical remote sensing satellite images and machine learning have been used to assess the SALs in the coastal Thoothukudi district of Tamil Nadu, India. Combination of spectral bands and indices were used as predictors in the model. Two different spatial resolution satellites of Sentinel-2 and Landsat-8 were used in the RF model and the predicted result shows the significance of the model with an overall validation accuracy of 98.31% and 97.19% respectively.

Satellite remote sensing data plays a vital role in assessing EC as saline soil reflects specific surface reflectance and helps to predict saline soil (Aksoy et al. 2022). White salt crust areas refer to high to extreme saline areas (Pessoa et al. 2016). However, high spectral surface reflectance in every band does not necessarily indicate high salinity in multispectral satellite data. This makes it difficult to use multispectral bands and their spectral indices to assess SAL

directly (Davis et al. 2019). However, the RF model evaluates the VIMP and represents the importance of each factor from highest to lowest importance in the model prediction. Several studies have suggested that the RF model is the best model with higher accuracy for monitoring soil salinity (Aksoy et al. 2022; Li et al. 2021).

Vegetation indices, salinity indices, and spectral bands of Sentinel-2 and Landsat-8 were considered as predictors in this study to assess SAL in 2020. Vegetation indices and salinity indices are the most widely used spectral indices to assess soil salinity (Aksoy et al. 2022; Li et al. 2021), but usually, spectral indices of vegetation and soil salinity response to EC are influenced by many factors which include resistance to salinity, percentage of vegetation cover, soil moisture and soil type (Peng et al. 2019). The results can vary significantly under different environmental settings. Since several other physical properties such as color, texture, and moisture affect the surface reflectance of saline soils, the unique soil salinity index may not provide an accurate estimate for all cases (Daliakopoulos et al. 2016).

Many strategies have been followed to get better accuracy results, such as (a) existing spectral indices selection based on the environmental condition of the study area, (b) new spectral indices creation based on local environmental conditions, (c) sensitive spectral indices selection based upon vegetation coverage, for example, salinity indices can be given more priority in the less vegetative cover region, whereas vegetation indices can be used in a high percentage of vegetation cover areas.

In the present study, the RF model was applied by considering all the predictors such as indices of vegetation, soil salinity, and spectral bands of Sentinel-2, Landsat-8 and collected soil samples EC. Vegetation indices, soil salinity indices and spectral bands have been used by many researchers worldwide for assessing SAL (Morgan et al. 2018; Aksoy et al. 2022; Allbed and Kumar 2013). Among the spectral band of Sentinel-2, green, blue, and red, among the Landsat-8, shortwave infra-red2, and red were found to be more important spectral bands in SAL prediction. Worldwide various studies on soil salinity assessment suggested that near-infrared, shortwave infrared, and blue bands were useful in identifying SAL (Nguyen et al. 2020; Khan and Sato 2001; Li et al. 2021).

Very less mix was observed between saline and non-saline regions while carrying out accuracy assessment by considering only saline and non-saline soil samples using both Sentinel-2 and Landsat images, also the overall accuracy was improved by 98.87 for both the Sentinel-2 and Landsat-8. Therefore, both Sentinel-2 and Landsat-8 were well capable in distinguishing saline regions from non-saline regions.

Prediction results obtained from sentinel-2, non-saline, and extremely saline classes had the highest producer's accuracy (100%) followed by slightly saline (93.75%), moderately saline (92.85%) and highly saline (88.88%), whereas prediction result from Landsat-8 shows non-saline soil has the highest producer accuracy, followed by moderately saline soil (92.85%), highly saline (88.88%), slightly saline (87.5%) and extremely saline (71.43%) class. This shows the Sentinel-2 image is more capable of predicting different soil salinity classes with more accuracy as compared to the Landsat-8 image. The F1-score result also shows that Sentinel-2 is more capable of distinguishing different levels of soil salinity as compared to Landsat-8.

In addition, Sentinel-2 data spatial resolution is 10 m, it is much more capable of detecting soil salinity with better accuracy as compared to Landsat-8 as increasing spatial resolution gives better accuracy (Li et al. 2021). Saline areas are largely found in nearby mining areas, near the coast, saltpan, and river mouths of both Thamirabarani and Vaippar river. This region experiences meteorological drought, shoreline change, tsunamis, and seawater intrusion. SAL results show high to extreme soil saline areas mostly concentrated around the saltpan regions and the Thamirabarani river mouth nearby. As seawater enters through both the river (Thamirabarani and Vaippar river) to the inland region, the surroundings become moderately to highly saline. In addition, due to saline groundwater and the high evaporation rate, the salt is deposited on the soil surface. Various studies reported seawater dominance in the Thamirabarani river region (Selvam et al. 2013; Satheskumar et al. 2021)

and also various landuse such as sandy beaches, dunes and mudflats conversion to saltpan may be the reason of SAL in the region. Soil salinity varies by location due to the influence of geological, hydrological, biological, and climatic factors that affects the influence of the soil and water balance (Wang et al. 2020). The spectral range of salt-affected land does not have a single spectral signature to detect, also surface cover creates mixed spectral responses which cause difficulties in identifying saline soil. Also, land surface cover, soil EC and sodicity prevent in the assessment of soil salinity using moderate resolution satellite images and spectral indices of vegetation and soil salinity (Kılıc et al. 2022). Yearly long-term rainfall trends can be incorporated and it can help in understanding the SAL over the study region in the future. The SAL detection and demarcation will help local stakeholders to manage and develop alternate livelihood options, also limiting the saltpan region to safeguard other coastal ecosystems. The current study model can be applied to other coastal regions in delineating the SAL at the state or national level which may help in making action plans to control and manage SALs.

Conclusion

Combination of satellite images and machine learning techniques coupled with field-level soil EC measurements has been used to detect the SALs of different saline classes from non-saline to extremely saline class. Spectral indices of vegetation, soil salinity, and spectral bands were incorporated into the model. The Hyperparameter tuning was used to evaluate the optimum number of decision trees and bag-fraction of highest training accuracy to improve the prediction accuracy. To delineate SAL, Sentinel-2, and Landsat-8 both can be utilized, both satellites were more or less equally sensitive to different classes of SAL. But Sentinel-2 satellite images is little better in detecting the different level of saline soil regions as compared to Landsat-8. The model can assist in making regional or national scale policies to control and manage SAL regions for different livelihood options. To demarcate precisely SALs, high-resolution satellite image is necessary. In addition, assessing soil salinity using spectral indices, land use patterns, long-term rainfall, and seawater rise in the region can provide a better framework to make policy planning to support alternative livelihood options for the local people.

Acknowledgements The funding support for the project on resource mapping for aquaculture in Tamil Nadu, by the Indian Council of Agricultural Research- Extramural project is gratefully acknowledged. Our sincere thanks to the Director, ICAR-Central Institute of Brackishwater Aquaculture for providing support and facilities.

Author contribution Methodology and GIS analysis: S. Kabiraj; Man-

uscript preparation and supervision: M. Jayanthi; Sample collection and data creation: S.Vijayakumar; GIS analysis: M. Duraisamy. All authors read and approved the final manuscript.

Funding The present research work was funded by the Indian Council of Agricultural Research.

Data Availability Sources of all the data have been described properly. Derived data supporting the findings of this study are available from the corresponding author on request.

Declarations

Conflict of interest The authors report there are no competing interests to declare.

References

- Aksoy S, Yildirim A, Gorji T, Hamzehpour N, Tanik A, Sertel E (2022) 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. <https://doi.org/10.1016/j.asr.2021.10.024>
- Allbed A, Kumar L (2013) Soil salinity mapping and monitoring in arid and semi-arid regions using remote sensing technology: a review. *Adv Remote Sens* 2:373–385. <https://doi.org/10.4236/ars.2013.24040>
- Arora S, Sharma V (2017) Reclamation and management of salt-affected soils for safeguarding agricultural productivity. *J Saf Agric* 1(1):1–10
- Asfaw E, Suryabhagavan KV, Argaw M (2018) Soil salinity modeling and mapping using remote sensing and GIS: The case of Wonji sugar cane irrigation farm. *Ethiopia J Saudi Soc Agric Sci* 17(3):250–258. <https://doi.org/10.1016/j.jssas.2016.05.003>
- Bai L, Wang C, Zang S, Wu C, Luo J, Wu Y (2018) Mapping soil alkalinity and salinity in Northern Songnen Plain, China with the HJ-1 hyperspectral imager data and partial least squares regression. *Sensors* 18(11):3855. <https://doi.org/10.3390/s18113855>
- Bannari A, El-Battay A, Bannari R, Rhinane H (2018) Sentinel-MSI VNIR and SWIR bands sensitivity analysis for soil salinity discrimination in an arid landscape. *Remote Sens* 10(6):855. <https://doi.org/10.3390/rs10060855>
- Bhardwaj AK, Mishra VK, Singh AK, Arora S, Srivastava S, Singh YP, Sharma DK (2019) Soil salinity and land use-land cover interactions with soil carbon in a salt-affected irrigation canal command of Indo-Gangetic plain. *CATENA* 180:392–400. <https://doi.org/10.1016/j.catena.2019.05.015>
- Brown JW, Hayward HE, Richards A, Bernstein L, Hatcher JT, Reeve RC, Richards LA (1954) *Diagnosis and Improvement of Saline and Alkali Soils*, 60. United States Department of Agriculture, Agriculture handbook
- Carrasco L, O'Neil AW, Morton RD, Rowland CS (2019) Evaluating combinations of temporally aggregated Sentinel-1, Sentinel-2 and Landsat 8 for land cover mapping with Google Earth Engine. *Remote Sens* 11(3):288. <https://doi.org/10.3390/rs11030288>
- Cuevas J, Daliakopoulos IN, del Moral F, Hueso JJ, Tsanis IK (2019) A review of soil-improving cropping systems for soil salinization. *Agron* 6295. <https://doi.org/10.3390/agronomy9060295>
- Daliakopoulos IN, Tsanis IK, Koutroulis A, Kourgialas NN, Varouchakis AE, Karatzas GP, Ritsema CJ (2016) The threat of soil salinity: a European scale review. *Sci Total Environ* 573:727–739. <https://doi.org/10.1016/j.scitotenv.2016.08.177>
- Davis E, Wang C, Dow K (2019) 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. <https://doi.org/10.1080/01431161.2019.1587205>
- Douaoui AEK, Nicolas H, Walte C (2006) Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. *Geoderma* 134(1–2):217–230. <https://doi.org/10.1016/j.geoderma.2005.10.009>
- Dwivedi RS, Kothapalli RV, Singh AN, Metternicht G, Zinck J (2008) Generation of farm level information on salt-affected soils using IKONOS-II multispectral data. *Remote Sensing of Soil Salinization: Impact on Land Management*. CRC press, Taylor & Francis, Boca Raton, pp 73–90
- Eldeiry AA, Garcia LA (2008) Detecting soil salinity in alfalfa fields using spatial modeling and remote sensing. *Soil Sci Soc Am J* 72(1):201–211. <https://doi.org/10.2136/sssaj2007.0013>
- Elnaggar AA, Noller JS (2009) Application of remote sensing data and decision tree analysis to mapping salt-affected soils over large areas. *Remote Sens* 151–165. <https://doi.org/10.3390/rs2010151>
- Fan X, Liu Y, Tao J, Weng Y (2015) Soil salinity retrieval from advanced multi-spectral sensor with partial least square regression. *Remote Sens* 7(1):488–511. <https://doi.org/10.3390/rs70100488>
- FAO (2021) The World Map of Salt Affected Soil [WWW Document]. FOOD Agric. Organ, United Nations. <https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/global-map-of-salt-affected-soils/en/>. Accessed 27 Apr 2022
- Farifteh J, Van der Meer F, Atzberger C, Carranza EJ (2007) Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN). *Remote Sens Environ* 110(1):59–78. <https://doi.org/10.1016/j.rse.2007.02.005>
- Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG (2002) Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens Environ* 83(1–2):195–213. [https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)
- Huete AR (1988) A soil-adjusted vegetation index (SAVI). *Remote Sens Environ* 25(3):295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- Ivushkin K, Bartholomeus H, Bregt AK, Pulatov A, Kempen B, De Sousa L (2019) Global mapping of soil salinity change. *Remote Sens Environ* 231:111260. <https://doi.org/10.1016/j.rse.2019.111260>
- Khan NM, Rastoskuev VV, Sato Y, Shiozawa S (2005) Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators. *Agric Water Manag* 77(1–3):96–109. <https://doi.org/10.1016/j.agwat.2004.09.038>
- Khan NM, Sato Y (2001) Monitoring hydro-salinity status and its impact in irrigated semi-arid areas using IRS-1B LISS-II data. *Asian J Geoinform* 1(3):63–73
- Kılıc OM, Budak M, Gunal E, Acır N, Halbac-Cotoara-Zamfir R, Alfarraj S, Ansari MJ (2022) Soil salinity assessment of a natural pasture using remote sensing techniques in central Anatolia, Turkey. *PLoS ONE* 17(4):e0266915. <https://doi.org/10.1371/journal.pone.0266915>
- Kumar P, Joshi PK, Mittal S (2016) Demand vs supply of food in India-futuristic projection. *Proc Indian Nat Sci Acad* 82 (5):1579–1586. <https://doi.org/10.16943/ptinsa/2016/48889>
- Kumar S, Gautam G, Saha SK (2015) Hyperspectral remote sensing data derived spectral indices in characterizing salt-affected soils: a case study of Indo-Gangetic plains of India. *Environ Earth Sci* 73(7):3299–3308. <https://doi.org/10.1007/s12665-014-3613-y>
- Kumar P and Sharma PK (2020) Soil salinity and food security in India. *Front Sustain Food Syst* 4:533781. <https://doi.org/10.3389/fsufs.2020.533781>
- Li Y, Wang C, Wright A, Liu H, Zhang H, Zong Y (2021) Combination of GF-2 high spatial resolution imagery and land surface

- factors for predicting soil salinity of muddy coasts. *CATENA* 202:105304. <https://doi.org/10.1016/j.catena.2021.105304>
- Li Z, Li Y, Xing A, Zhuo Z, Zhang S, Zhang Y, Huang Y (2019) Spatial prediction of soil salinity in a semiarid oasis: environmental sensitive variable selection and model comparison. *Chin Geogr Sci* 29(5):784–797. <https://doi.org/10.1007/s11769-019-1071-x>
- Li P, Qian H, Wu J (2018) Conjunctive use of groundwater and surface water to reduce soil salinization in the Yinchuan Plain, North-West China. *Int J Water Resour Dev* 34(3):337–353. <https://doi.org/10.1080/07900627.2018.1443059>
- Mandal S, Raju R, Kumar A, Kumar P, Sharma PC (2018) Current status of research, technology response and policy needs of salt-affected soils in India—A review. *J Indian Soc Coastal Agri Res* 36(2):40–53
- McBratney A, Santos M, Minasny B (2003) On digital soil mapping. *Geoderma* 117(1–2):3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- Meier M, Souza ED, Francelino MR, Fernandes Filho EI, Schaefer C E G, R (2018) Digital soil mapping using machine learning algorithms in a tropical mountainous area. *Revista Brasileira de Ciênciã do Solo* 42:e0170421. <https://doi.org/10.1590/18069657rbcs20170421>
- Morgan RS, El-Hady MA, Rahim IS (2018) Soil salinity mapping utilizing sentinel-2 and neural networks. *Indian J Agric Res* 52:524–529. <https://doi.org/10.18805/IJArE.A-316>
- Mukhopadhyay R, Sarkar B, Jat HS, Sharma PC, Bolan NS (2021) Soil salinity under climate change: Challenges for sustainable agriculture and food security. *J Environ Manage* 280:111736. <https://doi.org/10.1016/j.jenvman.2020.111736>
- Nabiollahi K, Taghizadeh-Mehrjardi R, Shahabi A, Heung B, Amirian-Chakan A, Davari M, Scholten T (2021) Assessing agricultural salt-affected land using digital soil mapping and hybridized random forests. *Geoderma* 1;385:114858. <https://doi.org/10.1016/j.geoderma.2020.114858>
- Narjary B, Meena MD, Kumar S, Kamra SK, Sharma DK, Triantafilis J (2019) Digital mapping of soil salinity at various depths using an EM38. *Soil Use and Management* 35(2):232–244. <https://doi.org/10.1111/sum.12468>
- Nguyen KA, Liou YA, Tran HP, Hoang PP, Nguyen TH (2020) Soil salinity assessment by using near-infrared channel and Vegetation Soil Salinity Index derived from Landsat 8 OLI data: a case study in the Tra Vinh Province, Mekong Delta, Vietnam. *Prog Earth Planet Sci* 7(1):1–16. <https://doi.org/10.1186/s40645-019-0311-0>
- Peng J, Biswas A, Jiang Q, Zhao R, Hu J, Hu B, Shi Z (2019) Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China. *Geoderma* 337:1309–1319. <https://doi.org/10.1016/j.geoderma.2018.08.006>
- Pessoa LG, Freire MB, Wilcox BP, Green CH, De Araújo RJ, De Araújo Filho JC (2016) Spectral reflectance characteristics of soils in northeastern Brazil as influenced by salinity levels. *Environ Monit Assess* 188(11):1–1. <https://doi.org/10.1007/s10661-016-5631-6>
- Qadir M, Noble AD, Schubert S, Thomas RJ, Arslan A (2006) Sodicty-induced land degradation and its sustainable management: problems and prospects. *Land Degrad Dev* 17:661–676. <https://doi.org/10.1002/ldr.751>
- Rhoades JD, Chanduvi F, Lesch SM (1999) Soil salinity assessment: Methods and interpretation of electrical conductivity measurements. Food & Agriculture Org
- Satheeskumar V, Subramani T, Lakshumanan C, Roy PD, Karunanidhi D (2021) Groundwater chemistry and demarcation of seawater intrusion zones in the Thamirabarani delta of south India based on geochemical signatures. *Environ Geochem Health* 43:757–770. <https://doi.org/10.1007/s10653-020-00536-z>
- Scudiero E, Skaggs TH, Corwin DL (2014) Regional scale soil salinity evaluation using Landsat 7, western San Joaquin Valley, California, USA. *Geoderma Reg* 2–3:82–90. <https://doi.org/10.1016/j.geoder.2014.10.004>
- Selvam S, Manimaran G, Sivasubramanian P (2013) Hydrochemical characteristics and GIS-based assessment of groundwater quality in the coastal aquifers of Tuticorin corporation, Tamilnadu, India. *Appl Water Sci* 3(1):145–159. <https://doi.org/10.1007/s13201-012-0068-8>
- Sheikh M (2011) A shoreline change analysis along the coast between Kanyakumari and Tuticorin, India, using digital shoreline analysis system. *Geo Spat Inf Sci* 14(4):282–293. <https://doi.org/10.1007/s11806-011-0551-7>
- Sheik M, Chandrasekar (2011) A shoreline change analysis along the coast between Kanyakumari and Tuticorin, India, using digital shoreline analysis system. *Geo Spat Inf Sci* 14(4):282–293. <https://doi.org/10.1007/s11806-011-0551-7>
- Taille PJ, Moorman CE, Poulter B, Ardón M, Emanuel RE (2019) Decadal-scale vegetation change driven by salinity at leading edge of rising sea level. *Ecosystems* 22(8):1918–1930. <https://doi.org/10.1007/s10021-019-00382-w>
- Wang J, Peng J, Li H, Yin C, Liu W, Wang T, Zhang H (2021) Soil Salinity Mapping Using Machine Learning Algorithms with the Sentinel-2 MSI in Arid Areas, China. *Remote Sens* 13(2):305. <https://doi.org/10.3390/rs13020305>
- Wang N, Xue J, Peng J, Biswas A, He Y, Shi Z (2020) 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. <https://doi.org/10.3390/rs12244118>
- Wei Y, Shi Z, Biswas A, Yang S, Ding J, Wang F (2020) Updated information on soil salinity in a typical oasis agroecosystem and desert-oasis ecotone: Case study conducted along the Tarim River, China. *Sci Total Environ* 716. <https://doi.org/10.1016/j.scitotenv.2019.135387135387>
- Wu W, Al-Shafie WM, Mhaimeed AS, Ziadat F, Nangia V, Payne WB (2014) Soil salinity mapping by multiscale remote sensing in Mesopotamia, Iraq. *IEEE J Sel Top Appl Earth Obs Remote Sens* 7(11):4442–4452. <https://doi.org/10.1109/JSTARS.2014.2360411>
- Wu W, Zucca C, Mhaimeed AS, Al-Shafie WM, Fadhil Al-Quraishi AM, Nangia V, Zhu M, Liu G (2018) Soil salinity prediction and mapping by machine learning regression in Central Mesopotamia, Iraq. *Land Degrad Dev* 29(11):4005–4014. <https://doi.org/10.1002/ldr.3148>

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

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.