



# Automated delineation of salt-affected lands and their progress in coastal India using Google Earth Engine and machine learning techniques

Sabyasachi Kabiraj · Marappan Jayanthi ·  
Muthusamy Samynathan ·  
Selvasekar Thirumurthy

Received: 2 July 2022 / Accepted: 2 February 2023 / Published online: 21 February 2023  
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

**Abstract** Assessment of salt-affected land (SAL) is still a major challenging task worldwide, especially in developing nations. The advancement of remotely sensed digital satellite images of different spectral bands has enabled the assessment of soil salinity. Sentinel-2 and Landsat 8 and 5 images of 2020, 2015 and 2009 and Shuttle Radar Topographical Mission data of 2014 were obtained from the Google Earth Engine data catalogue. Twenty spectral indices have been used which include four vegetation indices, twelve soil salinity indices, four topographical characteristics and their spectral bands. The Random Forest model was used to detect SAL. A total of 593 soil samples were used in the model. Of the electrical conductivity values of samples collected in the field, 70% of the soil samples were used for the model training, and the remaining 30% were used for validation. Also, fivefold cross-validation was carried out to validate the model prediction. The predicted SAL extent identified during 2020 was 134.4 sq. km with an overall accuracy of 93% using fivefold cross-validation. In 2015 and 2009, the total SAL was 128.42 and 120.41 sq. km, respectively. The total SAL has increased by 11.6% during the study period. The present study demonstrated the strength of remote sensing techniques to assess the

SAL, which will help quantify the unproductive lands at the state or national level for reclamation or other productive use.

**Keywords** Salt-affected lands · Google earth engine · Machine learning · Spectral indices · Soil salinity

## Introduction

Soil salinization is a global environmental issue as it affects around 10% of global food production (Machado & Serralheiro, 2017), particularly in coastal countries, and is expected to be more intense in the future due to climate change scenarios (Das et al., 2020), viz. sea level rise impact on coastal areas, rise in temperature and thus increase in evapotranspiration. Precise statistics on the global salt-affected land (SAL) spatial database is not yet developed; various data sources provide different information. Globally, 424 million ha of topsoil (0–30 cm) and 833 million ha (30–100 cm) of subsoil are salt-affected, covering 73% of the global land area in 118 countries (FAO, 2021). Studies have shown that the SAL area has been increasing across the world: 932.2 million ha (Sparks, 2003), 952.2 m ha (Arora & Sharma, 2017) and 1128 m ha (Mandal et al., 2018). Of the salt-affected regions, Asia stands first (65%), followed by Africa (19%) and Europe (5%) (Siebert et al., 2013). The estimations show that globally, there is US\$27.3

S. Kabiraj · M. Jayanthi (✉) · M. Samynathan ·  
S. Thirumurthy  
ICAR-Central Institute of Brackishwater Aquaculture,  
Chennai, India  
e-mail: m.jayanthi@icar.gov.in; jayanthiciba@gmail.com

billion loss of crop production annually due to salt-induced land degradation in irrigated areas (Qadir et al., 2014).

Salinity and excessive alkalinity (Zhu et al., 2012) have a negative influence on soil fertility, thus damaging the land and creating difficulties for plant growth (Wang et al., 2014). In semi-arid regions, soil salinity has a major influence due to the difficult climatic conditions, especially as these areas are under pressure for food and fibre (Mushtak & Zhou, 2012). The traditional method of collecting soil samples with subsequent laboratory analysis (Allbed et al., 2014) has proven to be insufficient and unsuitable to achieve the speed of development of this phenomenon, especially as these methods are very time-consuming, costly and difficult to update.

Remote-sensing data and methods have been increasingly applied to map soil salinity. Extensive research on soil salinity mapping using satellite images has been carried out over the last three decades across the world. The advancement of satellite data availability and analytical capability has paved the way for accurate and timely assessment of soil salinity at different spaces and times. Different spatial models have been tested for saline soil assessment based on the topographic information, climate condition, land use information, etc. (Fatholoulumi et al., 2020; Peng et al., 2019). Several spectral indices are used in soil salinity mappings, such as normalized differential vegetation index (NDVI), enhanced vegetation index (EVI) and generalized difference vegetation index (GDVI), and soil salinity normalized difference salinity index (NDSI), salinity index (SI), SI 1, SI 2, SI 3, SI-I and canopy response salinity index (CRSI), etc. (Jiang et al., 2018; Peng et al., 2019; Wang et al., 2020). In addition, measured soil electrical conductivity (EC) values and various topographical attributes used in the soil salinity mapping include elevation, slope, aspect, hill shade and flow accumulation (Fatholoulumi et al., 2020; Peng et al., 2019); distance from the sea and distance from the tidal creek; land surface temperature (Ivushkin et al., 2019) and soil moisture indices.

Modelling techniques used in the study include partial least square regression (PLSR) (Jiang & Shu, 2018; Peng et al., 2019; Yahiaoui et al., 2021), cubist model (Peng et al., 2019; Wang et al., 2020), multiple linear regression (MLR) (Gorji et al., 2017; Yahiaoui et al., 2021), support vector machine (SVM) (Jiang et al., 2018), artificial neural network (ANN) (Habibi et al., 2021; Jiang

et al., 2018) and random forest (RF) (Fathizad et al., 2020; Wang et al., 2020; Yahiaoui et al., 2021). A large number of studies showed machine learning techniques (ANN, SVM, RF) achieved high prediction accuracy as compared to other methods, especially RF (Ivushkin et al., 2019; Lu et al., 2018; Yahiaoui et al., 2021). Researchers have used different machine learning methods such as classification and regression trees (CART), support vector machine (SVM) and random forest (RF), but among all these methods, they found the RF method as the most accurate (Wang et al., 2019; Wu et al., 2018) as compared to others, and output generated by the RF model was more reliable (Li et al., 2021) as it is matched with the visual interpretation data (Aksoy et al., 2022). So, we also have considered the RF model for the present study.

India has 6.727 m ha (2.1% of total geographical area) of salt-affected area, classified into 2.956 m ha of saline soil and 3.77 m ha of sodic soil (Arora & Sharma, 2017). The country loses 16.84 million tons of farm production annually due to soil salinization, costing Rs 230.20 billion (Mandal et al., 2018). Considering the national population projection growth, India would need around 311 and 350 million tons of grain (cereals and legumes) to feed around 1.43 and 1.8 billion people in 2030 and 2050, respectively (Kumar et al., 2016). It is estimated that nearly 10% of the additional area is being salinized each year, and by 2050, about 50% of the arable land will be affected by salinity (Kumar & Sharma, 2020). Changing climate may enhance the speed of soil salinization due to sea-level rise, which will make densely populated developing nations more vulnerable than other regions.

Soil salinity assessment has been carried out by researchers in India using traditional and remote sensing data (Sahana et al., 2020; Periasamy & Ravi, 2020; Paliwal et al., 2019; Kumar et al., 2015), but these studies were in bits and pieces, not representative to make an informed decision to policymakers. However, there is a need to find ways and means to monitor soil salinity in a multi-temporal mode through automation to get faster, cheaper and more reliable results. Hence, the present study aimed to identify and quantify the SALs using various freely available satellite data, spectral indices and topographical characteristics through machine learning techniques.

**Material and methods**

**Study area**

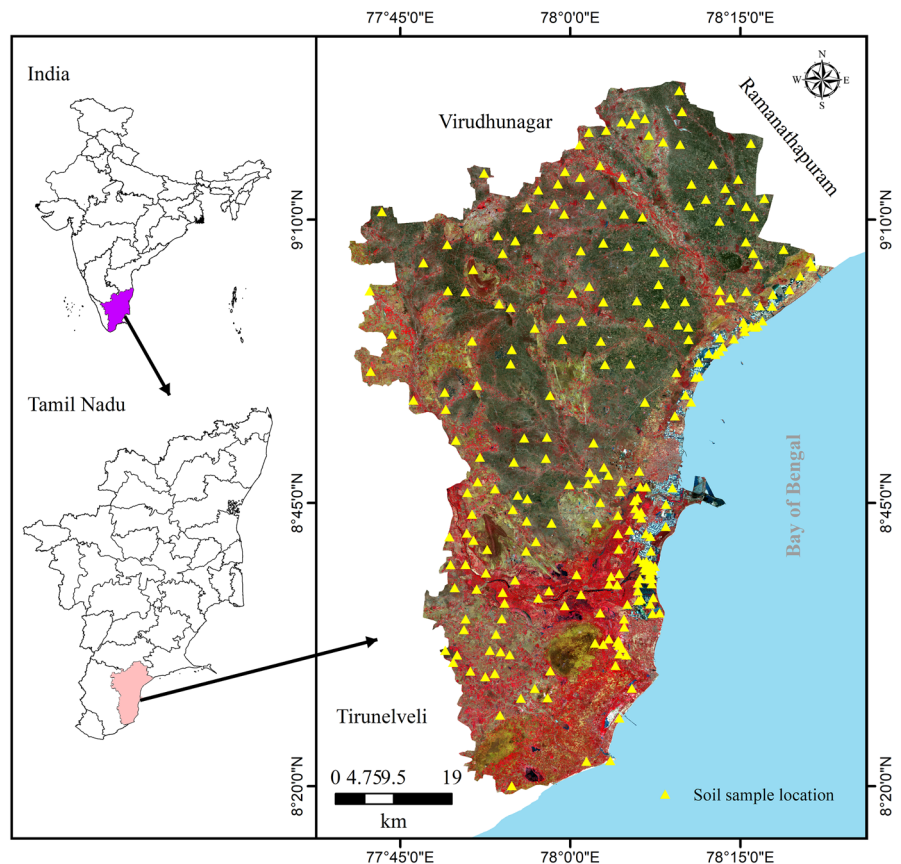
The study area, Thoothukudi District, lies between 8° 19' N and 9° 20' N latitude, 77° 40' E, and 78° 10' E longitude, covers an area of 4621 sq. km, with a coastal length of 163.5 km (Fig. 1). The region gains enormous ecological importance due to The Gulf of Mannar bioserve, home to marine biodiversity with 3600 species of flora and fauna. The district produces 70% of the total domestic salt production and meets 30% of our country’s needs. The study region is facing many climate change impacts such as meteorological drought (Sheik & Chandrasekar, 2011), sea-level rise (Sheikh, 2011), shoreline change and seawater intrusion (Satheeskumar et al., 2021) and increase in SAL (Selvam et al., 2013). Maximum and minimum temperatures in the study area ranged from 29.5 to 40.5 °C and from 18.4 to 26.7 °C, respectively. The

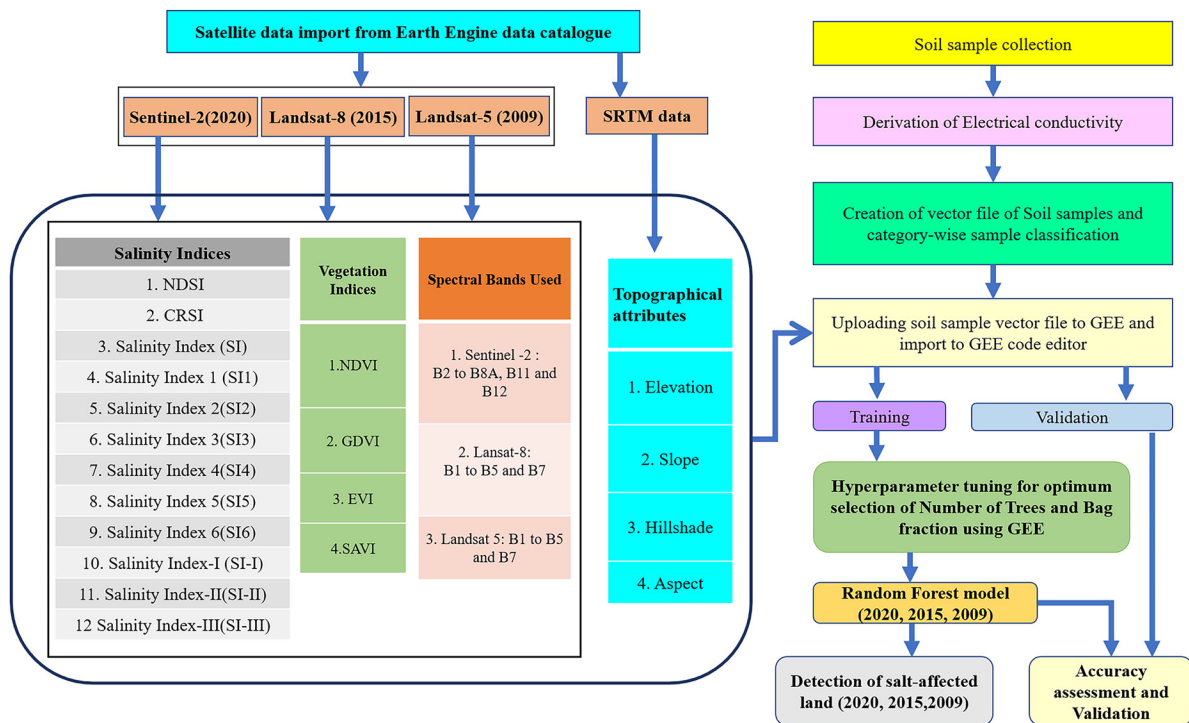
district experiences a semi-arid tropical climate, typically hot and dry. The average annual rainfall in this district is 661.6 m. The Thamirabarani River in Thoothukudi is highly influenced by seawater intrusion (Satheeskumar et al., 2021).

**Data used**

Satellite images of Sentinel-2 of 2020, Landsat 8 of 2015 and Landsat 5 TM of 2009 of May to August months having less than 20% cloud cover and Shuttle Radar Topography Mission (SRTM) 30-m resolution image (Fig. 2), available in the Google Earth Engine (GEE), were used in the study (Table 1). GEE possesses extensive geospatial datasets, including Sentinel, Landsat imageries and SRTM, and other ready-to-use products with the earth engine explorer web application, high-speed data processing and machine learning algorithm using Google’s computing infrastructure using application

**Fig. 1** Study area map showing soil sampling points in Thoothukudi District, Tamil Nadu, India





**Fig. 2** Methodology flow chart of automated delineation of SAL

programming interface library with a development environment using JavaScript and Python programming language which enables users to find, analyse and visualize large geospatial dataset without any supercomputer devices. The topographic maps of (58 K/3,4,7,8; 58H/9,13–15;58L/1–3,5; 58G/12,15,16) Survey of India (SOI) were used to demarcate the basic features like administrative boundary and rivers.

**Field sample collection and analysis**

A total of 593 soil samples were used in the present study. Of 593 soil samples, 258 soil samples were collected from different locations from 30 July to 5 August 2020 for EC analysis. The remaining samples were taken from the Indian Council of Agricultural soil database. The geographic coordinates of soil sampling locations were measured using GPS TDC 600

**Table 1** Satellite data used for the assessment of salt-affected lands

Satellite data used					
Sl.no	Year	Date	Data Used	Number of band used	Band details
1	2020	2020 May 01 to 2020 Aug 30	Sentinel-2	10	Blue (B2), green (B3), red (B4), red-edge1 (B5), red-edge 2 (B6), red-edge 3 (B7), near infrared (B8) Narrow near infrared (B8A), shortwave infrared 1 (B11) Shortwave infrared 2 (B12)
2	2015	2015 May 01 to 2015 Aug 30	Landsat-8	6	Blue (B2), green (B3), red (B4), near infrared (B5), shortwave infrared 1 (B6), shortwave infrared 2(B7)
3	2009	2009 Jan 05 to 2010 Aug 30	Landsat-5	6	Blue (B1), green (B2), red (B3), near infrared (B4), shortwave infrared 1 (B5), shortwave infrared 2(B7)

with a positioning accuracy of less than 2 m. At each sampling location, topsoils from four corners of quadrants were collected and mixed well. The samples were completely air-dried and passed through a 2-mm sieve to remove non-soil materials. Soil leachate was prepared at a soil/water ratio of 1:2.5, and then the EC of the soil was determined using a digital multi-parameter measuring apparatus (Systronics EC-TDS meter 308) at room temperature at 25 °C. The EC values were classified as non-saline (< 2 ds/m), slightly saline (2–4 ds/m), moderately saline (> 4, < 8 ds/m), highly saline (8–16 ds/m) and extremely saline (> 16) (Abrol et al., 1988; Ivushkin et al., 2019).

### Data processing

All selected satellite images from the GEE data catalogue were imported to the earth engine code editor section. The entire image collection was filtered using `script.filterDate()` based on the cloud-free nature of images between May to August for all 3 years. The study region shapefile was uploaded via ‘assets’ tool and imported into the code editor in GEE, and the extent of the study area was defined using `filterBounds()`. The median of all images was derived using `script.median()`, which produced the final image for the analysis. The use of median values between the date of interest of imageries reduces the data volume and ensures easy and fast analysis (Carrasco et al., 2019). As a preparatory step for analysis, the soil sampling point coordinates and measured EC values were used to make a shapefile in ArcGIS, and then imported into GEE.

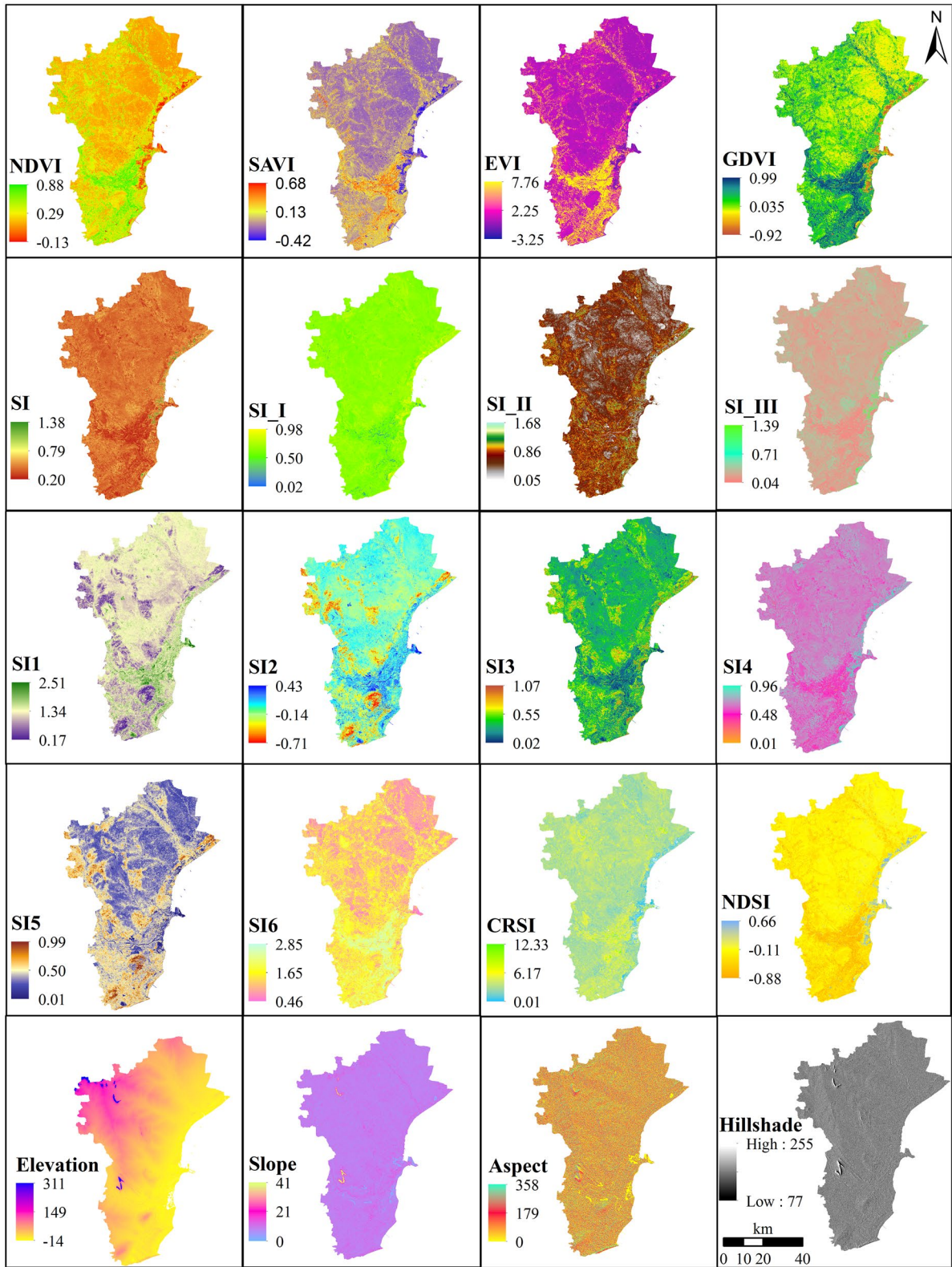
### Selection of predictors

Spectral indices are an effective method in arid and semi-arid regions to detect soil salinity (Fathizad et al., 2020; Gorji et al., 2017; Peng et al., 2019). The specific environmental conditions influence the selection of spectral indices (Wang et al., 2020). In this study, the commonly used soil salinity indicators have been selected to produce a powerful grouping in the soil salinity model and evaluate the comparison for further selection. Different spectral bands of Sentinel-2 (B2, B3, B4, B5, B6, B7, B8, B8a, B11, B12) were resampled to make the same spatial resolution of all satellite

images. Landsat 8 OLI (B2, B3, B4, B5, B6, B7) and Landsat 5 TM (B1, B2, B3, B4, B5 and B7) related to earth indicators were selected for the study (Table 1). Thermal images and other atmosphere-related bands were not considered in the analysis, as other atmospheric bands are not related to land degradation analysis, whereas thermal bands of Landsat 8 and 5 were of 100- and 60-m spatial resolution. The various soil salinity indices, vegetation indices and topographical attributes (Fig. 3) were integrated by various mathematical expressions of different band combinations as soil salinity indicators (Table 2). The bands are linked to their acronym using `select()`; the indices are calculated using the `script.expression()`. All selected 10 bands of Sentinel 2, sixteen indices and four topographical attributes were composed as predictors for the year 2020 using `addBands()`, whereas for the years 2015 and 2009, sixteen indices, six spectral bands and four topographical attributes were used.

### Random forest modelling

The RF classifier was trained using thirty variables for Sentinel-2 and 26 variables for Landsat 8 OLI, Landsat 5 TM and the EC shapefile. Of the measured EC values, 70% was used for training and the rest 30% was applied for validation using `script.filter()`. Of the 593 soil samples, 70% of the samples were used to train, and the remaining 30% samples were used for validation. The predictors selected training sets and their EC values have been integrated using script ‘`ee.Classifier.smileRandomForest()`’. The hyperparameter tuning was used to find the optimum number of trees and bag fraction with the highest training accuracy. With different settings from 1 to 500, the number of trees in the interval of 10 and the bag fraction varying from 0.1 to 0.9 in the interval of 0.1 was calculated to find the optimal number of trees and bag fraction with the highest training accuracy. Through the bag-fraction method, unused samples can participate in the decision tree-making process to assess the accuracy of each tree to improve model performance by considering the average accuracy value of all trees. The model was trained using `sampleRegions()`. The output was generated for the predictors, and their importance in the ranking was assessed using the script ‘`ee.Feature(null, ee.Dictionary().get(‘importance’))`’. The RF algorithm



**Fig. 3** Soil salinity predictors for the 2020 model for Thoothukudi District, India

**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 (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)	Khan et al. (2005)
	6.	Canopy response salinity index (CRSI)	$\sqrt{\frac{(\text{NIR} \times \text{Red}) - (\text{Green} \times \text{Blue})}{(\text{NIR} \times \text{Red}) + (\text{Green} \times \text{Blue})}}$ (6)	Scudiero et al. (2014)
	7.	Salinity index (SI)	$(\text{Blue} + \text{Red})^{0.5}$ (7)	Douaoui et al. (2006)
	8.	Salinity index 1 (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)	
Topography	16.	Salinity index VI (S-VI)	$(\text{SWIR1}/\text{SWIR2})$ (16)	
	17.	Elevation		Mehrjardi et al. (2014)
	18.	Slope		ArcGIS 10.6
	19.	Hillshade		ArcGIS 10.6
	20.	Aspect		ArcGIS 10.6

makes a group of decision trees and allows them ‘vote’ for the best likely class (Strobl et al., 2008).

The confusion matrix was calculated using the training dataset of the classified raster using script. confusionMatrix(), and thus, overall training accuracy was calculated from the classified raster using. accuracy() script. Similarly, for the validation dataset, error matrix and overall validation accuracy was calculated using.errorMatrix() and.accuracy() script. Fivefold cross-validation method, training and validation accuracy were calculated to validate the model performance. A very limited number of points in moderately, highly and extremely saline regions were selected because of fewer soil samples in these regions. The area of the classified raster image under each category was calculated using ee.Image.pixelArea() script.

## Results

### Soil sample assessment for electrical conductivity

Soil EC values have ranged from 0.31 to 72 ds/m, with a mean value of 5.94 ds/m. Both the training and validation samples ranged from 0.31 to 72 ds/m, to represent the whole dataset. The coefficient of variation of EC values was 2.46, which shows huge variability in the soil samples. Of the analysed samples, 19.76%, 6.59% and 73.64% had EC values of more than 4 ds/m, between 2 and 4 ds/m and less than 2 ds/m. The major part of the study area belongs to agricultural land. The mean and standard deviation (SD) of the training samples were 5.06 ds/m and 13.01 ds/m, respectively, while the validation sample mean and SD were 5.02 ds/m and 12.50 ds/m, respectively.

Machine learning with RF model

Hyperparameter was used to get the optimum value of the number of trees and bag fraction. The RF model was computed using the respective number of trees and bag fractions for the years 2020, 2015 and 2009. It was observed that 20 number of trees with a 0.6 bag fraction had the highest training accuracy of 99% in 2020. Similarly, the highest training accuracy was observed with 200 number of trees and 0.9 bag fraction in 2015; 0.5 bag fraction and 30 number of trees showed the highest training accuracy in 2009.

Variable of importance (VIMP) was derived using the RF model for 3-year study periods. Thirty variables were used in the 2020 RF model, whereas 26 variables were used in 2015 and in 2009. The importance of each variable was evaluated for each year's RF model. In 2020, the RF model's top 10 important variables were SI6, B8, NDSI, GDVI, NDVI, B2, B12, B8A, EVI, B5 and their VIMP scores were 6.6, 3.89, 3.84, 3.57, 3.48, 3.41, 3.39, 3.3, 3.16 and 2.93, respectively.

The RF model of 2015, the top 10 important variables were CRSI, SAVI, B5, SI4, NDVI, SI-II, B6, SI1, B2 and SI6, and their VIMP scores were 6.18, 6.15, 5.85, 5.58, 5.45, 5.14, 4.58, 4.33, 4.29 and 4.27 respectively. In 2009, the top 10 important variables were NDSI, SI1, B5, GDVI, B1, SI6, B7, NDVI, EVI and CRSI and their score of importance was 7.29,

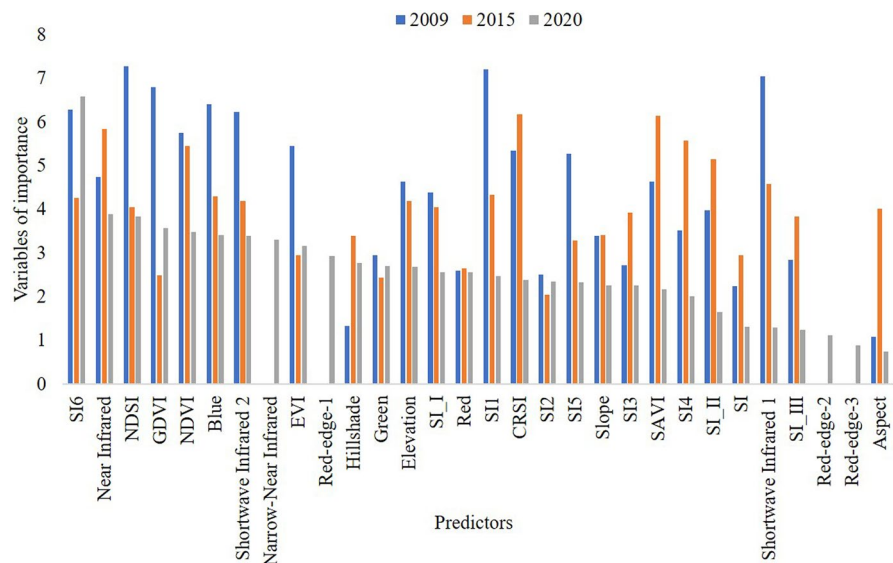
7.21, 7.05, 6.80, 6.42, 6.29, 6.23, 5.76, 5.45 and 5.35 respectively (Fig. 4).

The spatial extent of salt-affected lands

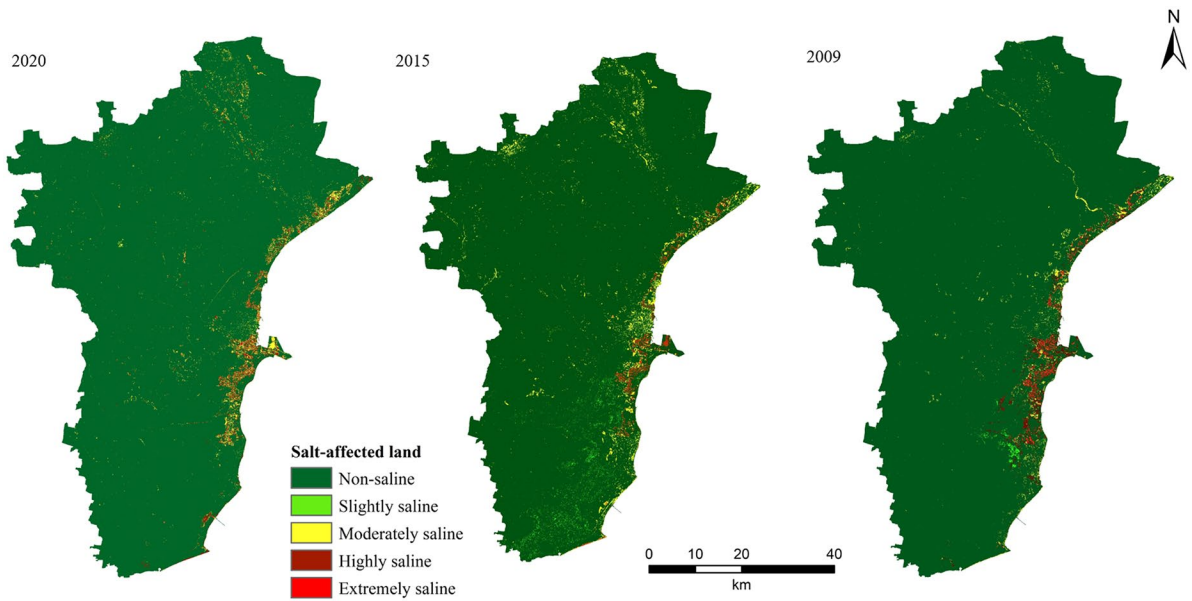
SAL of Thoothukudi District was identified using the RF model (Fig. 5). The total SAL in 2020, 2015 and 2009 was 134.4 sq. km, 128.42 sq. km and 120.41 sq. km respectively (Table 3). Of the SAL of 134.44 sq. km in 2020 (Fig. 6a), 75.94 sq. km was the moderately saline area, 42.97 sq. km was highly saline, and 15.53 sq. km was extremely high saline. The overall training accuracy was 99%, whereas validation accuracy was observed to be 91%. Also, the fivefold cross-validation result shows overall training and validation accuracy of 96% and 93%. As the number of soil samples was very less in moderately saline, highly saline and extremely saline region, very less samples were used in the training and validation process of the model. The accuracy assessment of soil salinity classes is given in Table 4. Classification errors mostly appear in the slightly saline and extremely saline regions (31 and 66% respectively).

Of the total SAL of 128.42 sq. km in 2015, moderately saline area occupies 82.51 sq. km, highly saline area 24.22 sq. km and 21.69 sq. km area as extremely high saline area (Fig. 6b). Of the total area of 120.41 sq. km in 2009, 43.35, 54.63 and 22.43 sq. km were moderately saline, highly saline and extremely high saline regions, respectively.

**Fig. 4** Variables of importance ranking of model predictors in 2009, 2015, and 2020





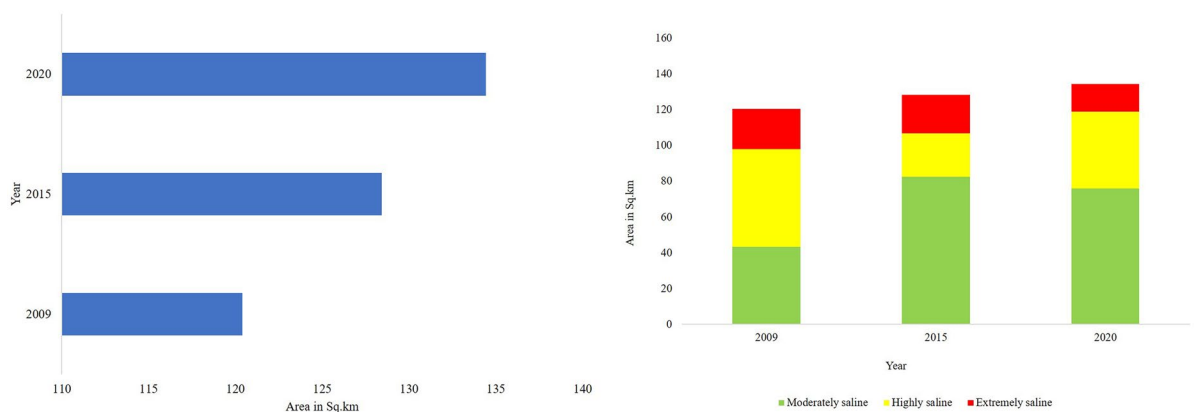


**Fig. 5** Soil salinity prediction of Thoothukudi District in 2009, 2015 and 2020

**Table 3** Spatial extent of various soil salinity classes using different satellite data

Extent of different levels of soil salinity (Sq. km)

Sl.no	Year	Data used	Non-saline	Slightly saline	Moderately saline	Highly saline	Extremely saline	Total salt-affected soil area(sq.km)
1	2020	Sentinel-2	4447.95	3.66	75.94	42.97	15.53	134.44
2	2015	Landsat-8 OLI	4399	31.52	82.51	24.22	21.69	128.42
3	2009	Landsat-5TM	4411.93	6.46	43.35	54.63	22.43	120.41



**Fig. 6** a Total salt-affected land in different years. b Classwise salt-affected land distribution in different years

**Table 4** Accuracy assessment of soil salinity classes using Sentinel 2 data of 2020

Salinity class	Predicted						Total sample	Producers' accuracy (%)
	Non-saline	Slightly saline	Moderately saline	Highly saline	Extremely saline			
Non-saline	134	0	0	0	0	134	100.00	
Slightly saline	10	5	0	1	0	16	31.25	
Moderately saline	1	0	13	0	0	14	92.86	
Observed Highly saline	0	0	1	7	1	9	77.78	
Observed Extremely saline	0	0	0	2	4	6	66.67	
Total sample	145	5	14	10	5	179		
User's accuracy (%)	92.41	100.00	92.86	70.00	80.00			

The performance of the RF model was evaluated based on the validation accuracy of the predicted soil salinity category with the collected soil sample salinity category. The overall validation accuracy of the model was 91.06% when tested with the sample data of 2020.

## Discussion

Optical remote sensing imagery and advanced machine learning technique have been used to identify areas of salt-affected land (SAL) of the coastal district in India as a model case study. Remote sensing data play a significant role in analysing EC because the influence of soil salt leads to specific reflections that form the basis for the prediction of SAL. Regions covered with white salt crust indicate highly SAL (Wang et al., 2020). However, in Sentinel-2 multispectral satellite data, each band's high spectral reflectance does not always indicate high salinity. This creates difficulties in using the multispectral band and its spectral indices in assessing soil salinity directly (Davis et al., 2019). However, the RF model evaluates VIMP, and it represents each factor's importance in model prediction. In the present study, we used the RF model and evaluated VIMP to represent the factors of high importance to the least importance on the model prediction accuracy. With this method, each tree grows separately without being pruned, and this does not overestimate the final model. Several other studies suggested the RF model as the best model with high accuracy level in soil

salinity monitoring (Fathizad et al., 2020; Yahiaoui et al., 2021).

In the present study, vegetation indices, salinity indices and topographical factors were used to assess the SAL. The spectral indices of vegetation and salinity were the most commonly used indices to assess soil salinity (Peng et al., 2019), but vegetation indices are usually more sensitive to soil salinity changes under high vegetation coverage (Peng et al., 2019). Vegetation spectral indices' and soil salinity spectral indices' responses to EC are influenced by many factors, including salt tolerance, vegetation cover, soil type and moisture (Metternicht & Zinck, 2003). The outcome may differ significantly under different environmental conditions. So far, no spectral indices can assess soil salinity in all environmental conditions (Allbed et al., 2014).

There are many strategies followed to get better accuracy results, such as (a) selection of existing spectral indices based on the environmental condition of the study area; (b) creation of new spectral indices based upon the local environmental conditions and (c) selection of sensitive spectral indices based upon vegetation coverage, for example in less vegetative area, spectral indices of salinity should be given priority and vegetative area spectral indices of vegetation should be considered.

Considering the study area environment in the present study, the RF model was applied with all the variables (indices of vegetation, soil salinity and topographical characteristics). As the major land use of the study area (Thoothukudi District) was

agricultural land, with saltpan, coastal plantation, industries, aquaculture, built-up, mudflat and sandy beach areas, we have used both vegetation and soil salinity indices to detect SAL over the district.

The important vegetation indices were NDVI and GDVI in 2020, SAVI and NDVI in 2015 and NDVI in 2009. Likewise, the important salinity indices were SI6 and NDSI in 2020. CRSI and SI4 in 2015, and NDSI and SI1 in 2009. Researchers have used vegetation and salinity indices to identify SAL worldwide (Fathizad et al., 2020; Gorji et al., 2017; Ijaz et al., 2020; Peng et al., 2019; Scudiero et al., 2014; Wu et al., 2014). Of the spectral bands of Sentinel 2, near-infrared (B8), blue, shortwave infrared 2, narrow near-infrared (B8A) and vegetation red-edge (B5) bands played an influencing role in the assessment of SAL. Also, near-infrared (B5), shortwave infrared 1 (B6) and blue (B2) were the important bands in Landsat 8. Of the Landsat 5 spectral bands, shortwave infrared-1 (B5) and blue (B1) and shortwave infrared-2 (B7) have influenced the assessment of SAL. Various studies worldwide suggested that near-infrared, blue and shortwave infrared bands were useful in detecting SAL (Nguyen et al., 2020; Khan & Sato, 2001). Among the topographical variables, hillshade, aspect and elevation were found to be not important for all 3 years, as the study region's topography is plain or flat terrain. Less rainfall limited the surface runoff region and has weakened topographical factors in the redistribution of soil salinity (Akramkhanov et al., 2011).

The prediction results of SAL from 2009 to 2020 show an increasing trend. The total SAL increased by 11.66% during the entire study period. Out of the total SAL in 2020, 56% were moderately saline, 32% highly saline, and 12% extremely saline respectively, whereas in 2009, 36% of the total SAL was moderately saline, 45% highly saline, and 19% extremely saline. It was noticed that there was no mix between the saline and non-saline categories. Also, the producer's accuracy of saline soil (98.27%) and non-saline (99.33%) class was improved which indicated the capacity of the model to differentiate the salt-affected and non-saline regions. Of the saline soil classes in 2020, the moderately saline class had the highest producer's accuracy (92.86%), followed by highly saline (77.78%) and extremely saline (66.67%).

In the north-eastern part of the district nearby the Vaippar river region, most of the salt-affected regions were non-saline during 2009 and were covered with vegetation, but from 2015 onwards, these regions' surface vegetation cover was reduced, and the surface energy and water balance were changed which increased surface albedo and soil salinity, and the surface soil gradually becomes bare land devoid of vegetation. The saline areas were mostly distributed in the mining surrounding regions, urban landscapes, nearby saltpan and river mouths of both the rivers Thamirabarani and Vaippar river regions. An increase in the SAL area over the study region was caused by many factors such as tsunamis, meteorological drought, shoreline change and seawater intrusion. In the present study, we found high to extreme soil salinity in and around the saltpan region and the Thamirabarani river surroundings. As the seawater flows through the Thamirabarani River to the inland coastal regions, the surrounding regions became moderate to highly saline. The salt deposition happening over the surface soil occurs due to saline groundwater and high evaporation rate, which turns the land into saline. Researchers have observed seawater dominance in the sub-surface water chemistry in the Thamirabarani delta region (Satheeskumar et al., 2021; Selvam et al., 2013) and the conversion of sandy beaches, dunes and mudflats to saltpan (Gangai & Ramachandran, 2010). According to Singaraja (2017), a lower concentration of electrical conductivity (EC) values was observed in groundwater samples of north-western and southern parts of the district, whereas very high EC values were found in the groundwater samples of north-eastern and central part of the district due to seawater intrusion, saltpan activities and other industrial activities. In addition, increasing EC trend along the groundwater flow direction indicates the leaching of secondary salts and anthropogenic impact by industrial activities, apart from seawater intrusion. During the ground truth verification, it was noticed that the bridge of about 2.5 km length was constructed on the southeastern side of the district to arrest the seawater flow from the Karumeni river mouth landward. Compared to the southeastern part, the north-eastern side is a low-lying region, and salt pans are present.

Soil salinity differs from place to place as the movement and salt accumulation are determined by geological, ecological, hydrological and climatic factors (Wang et al., 2020) which influence the

soil–water balance. Researchers worldwide argued that identifying salt-affected land through remote sensing techniques is very complex as the spectral range of mineral species that cause soil salinity does not have a single spectral signature to identify, and confused surface cover causes mixed spectral responses associated with salt deposits. Also, heterogeneous surface cover, soil EC values and sodicity prevented success in determining the surface soil salinity using moderate spatial resolution satellite images of Landsat 5 and 8 and the vegetation indices and soil salinity indices (Kilic et al., 2022). Long-term annual rainfall patterns can also be included for understanding the SAL over the study area in the future. The demarcation of SAL will help the local stakeholders to manage and make alternate livelihood options and restrict the extent of the saltpan area to protect other coastal ecosystems. The present study model can be applied to other coastal regions to demarcate the SAL at a state or national level to draw action plans to manage and control SALs.

## Conclusion

Advanced remote sensing and machine learning techniques coupled with field-level EC measurements have been used to measure the SALs of different categories from high to low saline. Various spectral indices of vegetation, soil salinity and topographical characteristics have been used as input variables for the RF model. Hyperparameter was used to calculate the optimum number of trees and bag fraction to improve the model accuracy assessment. The model can be applied at a different regional to national scale to draw policy measures to control and also make use of salt-affected regions for different purposes. Through soil salinity assessment using moderate spatial resolution satellite images such as Landsat 5 and 8 do not provide enough spatial resolution to reveal soil salinity in large regions. High-resolution images would be more appropriate to demarcate soil salinity areas. In addition to the indices to assess soil salinity, the spatial planning incorporating predicted regional level seawater rise, rainfall pattern and changes in land use can provide a better framework in the future to draw the policy measures to support the alternative livelihood options of the coastal population.

**Acknowledgements** Our sincere thanks to the Director, ICAR-Central Institute of Brackishwater Aquaculture, for providing support and facilities.

**Author contribution** S.Kabiraj — GIS analysis, sample collection, and manuscript writing. M. Jayanthi — conceptualization and manuscript writing. M. Samynathan — GIS analysis. S.Thirumurthy — sample collection.

**Funding** The present research work was funded by the Indian Council of Agricultural Research. Funding support provided for the project on resource mapping for aquaculture in Tamil Nadu by the Department of Fisheries, Government of Tamil Nadu, and ICAR-Extramural project.

**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.

**Code availability** The code used in the present study is available from the corresponding author on request.

## Declarations

**Competing interests** The authors declare no competing interests.

## References

- Abrol, I. P., Yadav, J. S. P., & Massoud, F. I. (1988). Salt-affected soils and their management (No. 39). FAO Soils Bulletin 39, Food & Agriculture Org.
- Akramkhanov, A., Martius, C., Park, S., & J., Hendrickx, J, M, H. (2011). Environmental factors of spatial distribution of soil salinity on flat irrigated terrain. *Geoderma*, 163(1–2), 52–55. <https://doi.org/10.1016/j.geoderma.2011.04.001>
- 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. *Advances in Space Research*, 69(2), 1072–1086. <https://doi.org/10.1016/j.asr.2021.10.024>
- Allbed, A., Kumar, L., & Sinha, P. (2014). Mapping and modelling spatial variation in soil salinity in the Al Hassa Oasis based on remote sensing indicators and regression techniques. *Remote Sensing*, 6(2), 1137–1157. <https://doi.org/10.3390/rs6021137>
- Arora, S., & Sharma, V. (2017). Reclamation and management of salt-affected soils for safeguarding agricultural productivity. *Journal of Safe Agriculture*, 1(1), 1–10.
- Carrasco, L., O’Neil, A., & W., Morton, R, D., Rowland, C, S. (2019). Evaluating combinations of temporally aggregated Sentinel-1, Sentinel-2 and Landsat 8 for land cover mapping with Google Earth Engine. *Remote Sensing*, 11(3), 288. <https://doi.org/10.3390/rs11030288>
- Das, R. S., Rahman, M., Sufian, N. P., Rahman, S. M. A., & Siddique, M. A. M. (2020). Assessment of soil salinity in the accreted and non-accreted land and its implication

- on the agricultural aspects of the Noakhali coastal region, Bangladesh. *Heliyon*, 6(9), e04926. <https://doi.org/10.1016/j.heliyon.2020.e04926>
- 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. *International Journal of Remote Sensing*, 40(16), 6134–6153. <https://doi.org/10.1080/01431161.2019.1587205>
- Douaoui, A., & E. K., Nicolas, H., Walter, 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>
- FAO. (2021). The world map of salt affected soil [WWW Document]. FOOD Agric. Organ, United Nations. Retrieved April 27, 2022, from <https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/global-map-of-salt-affected-soils/en/>
- Fathizad, H., Ardakani, M. A. H., Sodaiezadeh, H., Kerry, R., & Taghizadeh-Mehrjardi, R. (2020). Investigation of the spatial and temporal variation of soil salinity using random forests in the central desert of Iran. *Geoderma*, 365, 114233. <https://doi.org/10.1016/j.geoderma.2020.114233>
- Fatholouloumi, S., Vaezi, A. R., Alavipanah, S. K., Ghorbani, A., Saurette, D., & Biswas, A. (2020). Improved digital soil mapping with multitemporal remotely sensed satellite data fusion: a case study in Iran. *Science of the Total Environment*, 721, 137703. <https://doi.org/10.1016/j.scitotenv.2020.137703>
- Gangai, I., & P. D., Ramachandran, S. (2010). The role of spatial planning in coastal management—a case study of Tuticorin coast (India). *Land Use Policy*, 27(2), 518–534. <https://doi.org/10.1016/j.landusepol.2009.07.007>
- Gorji, T., Sertel, E., & Tanik, A. (2017). Monitoring soil salinity via remote sensing technology under data scarce conditions: A case study from Turkey. *Ecological Indicators*, 74, 384–391. <https://doi.org/10.1016/j.ecolind.2016.11.043>
- Habibi, V., Ahmadi, H., Jafari, M., & Moeini, A. (2021). Mapping soil salinity using a combined spectral and topographical index with artificial neural network. *PLoS One*, 16(5), e0228494. <https://doi.org/10.1371/journal.pone.0228494>
- Huete, A., & R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3), 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- Huete, A., Didan, K., Miura, T., Rodriguez, E., & P., Gao, X., Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 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)
- Ijaz, M., Ahmad, H., & R., Bibi, S., Ayub, M. A., Khalid, S. (2020). Soil salinity detection and monitoring using Landsat data: A case study from Kot Addu, Pakistan. *Arabian Journal of Geosciences*, 13(13), 1–9. <https://doi.org/10.1007/s12517-020-05572-8>
- Ivushkin, K., Bartholomeus, H., Bregt, A. K., Pulatov, A., Kempen, B., & de Sousa, L. (2019). Global mapping of soil salinity change. *Remote Sensing of Environment*, 231, 111260. <https://doi.org/10.1016/j.rse.2019.111260>
- Jiang, C., Chen, S., Pan, S., Fan, Y., & Ji, H. (2018). Geomorphic evolution of the Yellow River Delta: Quantification of basin-scale natural and anthropogenic impacts. *Catena*, 163, 361–377. <https://doi.org/10.1016/j.catena.2017.12.041>
- Jiang, H., & Shu, H. (2018). Optical remote sensing data based research on detecting soil salinity at different depth in an arid area oasis, Xinjiang, China. *Earth Science Informatics*, 12(1), 43–56. <https://doi.org/10.1007/s12145-018-0358-2>
- Khan, N., & M., Sato, Y. (2001). Monitoring hydro-salinity status and its impact in irrigated semi-arid areas using IRS-1B LISS-II data. *Asian Journal of Geoinform*, 1(3), 63–73.
- Khan, N. M., Rastoskuev, V., & V., Sato, Y., Shiozawa, S. (2005). Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators. *Agricultural Water Management*, 77(1–3), 96–109. <https://doi.org/10.1016/j.agwat.2004.09.038>
- Kilic, O. M., Budak, M., Gunal, E., Acir, N., Halbac-Cotoara-Zamfir, R., Alfarraj, S., & Ansari, M. J. (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, P. K., & Mittal, S. (2016). Demand vs supply of food in India-futuristic projection. *Proceedings of the Indian National Science Academy*, 82(5), 1579–1586. <https://doi.org/10.16943/ptinsa/2016/48889>
- Kumar, P., & Sharma, P. K. (2020). Soil salinity and food security in India. *Frontiers in Sustainable Food Systems*, 4, 533781. <https://doi.org/10.3389/fsufs.2020.533781>
- Kumar, S., Gautam, G., Saha, S., & K. (2015). Hyperspectral remote sensing data derived spectral indices in characterizing salt-affected soils: A case study of Indo-Gangetic plains of India. *Environ. Earth Science*, 73(7), 3299–3308. <https://doi.org/10.1007/s12665-014-3613-y>
- 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>
- Lu, W., Lu, D., Wang, G., Wu, J., Huang, J., & Li, G. (2018). Examining soil organic carbon distribution and dynamic change in a hickory plantation region with Landsat and ancillary data. *Catena*, 165, 576–589. <https://doi.org/10.1016/j.catena.2018.03.007>
- Machado, R., & Serralheiro, R. (2017). Soil salinity: effect on vegetable crop growth. Management practices to prevent and mitigate soil salinization. *Horticulturae*, 3(2), 30. <https://doi.org/10.3390/horticulturae3020030>
- Mandal, S., Raju, R., Kumar, A., Kumar, P., Sharma, P., & C. (2018). Current status of research, technology response and policy needs of salt-affected soils in India—a review. *Journal of the Indian Society of Coastal Agricultural Research*, 36(2), 40–53.
- Mehrjardi, R., & T., Minasny, B., Sarmadian, F., Malone, B., P. (2014). Digital mapping of soil salinity in Ardakan region, central Iran. *Geoderma*, 213, 15–28. <https://doi.org/10.1016/j.geoderma.2013.07.020>
- Metternicht, G., & I., Zinck, J. A. (2003). Remote sensing of soil salinity: Potentials and constraints. *Remote Sensing of Environment*, 85(1), 1–20. [https://doi.org/10.1016/S0034-4257\(02\)00188-8](https://doi.org/10.1016/S0034-4257(02)00188-8)

- Mushtak, T., & J., Zhou, J. (2012). Assessment of soil salinity risk on the agricultural area in Basrah Province, Iraq: Using remote sensing and GIS techniques. *Journal of Earth Science*, 23(6), 881–891. <https://doi.org/10.1007/s12583-012-0299-5>
- Nguyen, K. A., Liou, Y. A., Tran, H. P., Hoang, P. P., & Nguyen, T. H. (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. *Progress in Earth and Planetary Science*, 7(1), 1–16. <https://doi.org/10.1186/s40645-019-0311-0>
- Paliwal, A., Laborte, A., Nelson, A., Singh, R., & K. (2019). Salinity stress detection in rice crops using time series MODIS VI data. *International Journal of Remote Sensing*, 40(21), 8186–8202. <https://doi.org/10.1080/01431161.2018.1513667>
- 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>
- Periasamy, S., & Ravi, K. P. (2020). A novel approach to quantify soil salinity by simulating the dielectric loss of SAR in three-dimensional density space. *Remote Sensing of Environment*, 251, 112059. <https://doi.org/10.1016/j.rse.2020.112059>
- Qadir, M., Quillérou, E., Nangia, V., Murtaza, G., Singh, M., Thomas, R., Drechsel, P., & Noble, A. (2014). Economics of salt-induced land degradation and restoration. *In Natural Resources Forum*, 38(4), 282–95. <https://doi.org/10.1111/1477-8947.12054>
- Sahana, M., Rehman, S., Patel, P., & P., Dou, J., Hong, H., Sajjad, H. (2020). Assessing the degree of soil salinity in the Indian Sundarban Biosphere Reserve using measured soil electrical conductivity and remote sensing data-derived salinity indices. *Arabian Journal of Geosciences*, 13, 1289. <https://doi.org/10.1007/s12517-020-06310-w>
- Satheeskumar, V., Subramani, T., Lakshumanan, C., Roy, P. D., & Karunanidhi, D. (2021). Groundwater chemistry and demarcation of seawater intrusion zones in the Thamirabarani delta of south India based on geochemical signatures. *Environmental Geochemistry and Health*, 43, 757–770. <https://doi.org/10.1007/s10653-020-00536-z>
- Scudiero, E., Skaggs, T., & H., Corwin, D. L. (2014). Regional scale soil salinity evaluation using Landsat 7, western San Joaquin Valley, California, USA. *Geoderma Regional*, 2–3, 82–90. <https://doi.org/10.1016/j.geodrs.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. *Applied Water Science*, 3(1), 145–159. <https://doi.org/10.1007/s13201-012-0068-8>
- Sheik, M., & Chandrasekar, N. (2011). A shoreline change analysis along the coast between Kanyakumari and Tuticorin, India, using digital shoreline analysis system. *Geo-Spatial Information Science*, 14(4), 282–293. <https://doi.org/10.1007/s11806-011-0551-7>
- Sheik, M. (2011). A shoreline change analysis along the coast between Kanyakumari and Tuticorin, India, using digital shoreline analysis system. *Geo-Spatial Information Science*, 14(4), 282–293. <https://doi.org/10.1007/s11806-011-0551-7>
- Siebert, S., Henrich, V., Frenken, K., & Burke, J. (2013). Update of the digital global map of irrigation areas to version 5. In: Rheinische Friedrich-Wilhelms-Universität, Bonn, Germany and Food and Agriculture Organization of the United Nations, Rome, Italy.
- Singaraja, C. (2017). Relevance of water quality index for groundwater quality evaluation: Thoothukudi District, Tamil Nadu, India. *Applied Water Science*, 7, 2157–2173. <https://doi.org/10.1007/s13201-017-0594-5>
- Sparks, D. L. (2003). Environmental soil chemistry. Elsevier Academic Press, 352. Retrieved November 14, 2019, from [www.elsevier.com](http://www.elsevier.com) <https://doi.org/10.1016/C2009-0-02455-6>
- Strobl, C., Boulesteix, A., & L., Kneib, T., Augustin, T., Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1), 307. <https://doi.org/10.1186/1471-2105-9-307>
- Wang, F., Chen, X., Luo, G., & Han, Q. (2014). Mapping of regional soil salinities in Xinjiang and strategies for amelioration and management. *Chinese Geographical Science*, 25(3), 321–336. <https://doi.org/10.1007/s11769-014-0718-x>
- Wang, J., Ding, J., Yu, D., Teng, D., He, B., Chen, X., Ge, X., Zhang, Z., Wang, Y., Yang, X., Shi, T., & Su, F. (2020). Machine learning-based detection of soil salinity in an arid desert region, Northwest China: a comparison between Landsat-8 OLI and Sentinel-2 MSI. *Science of the Total Environment*, 707, 136092. <https://doi.org/10.1016/j.scitotenv.2019.136092>
- Wang, J., Ding, J., Yu, D., Ma, X., Zhang, Z., Ge, X., & Guo, Y. (2019). 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. <https://doi.org/10.1016/j.geoderma.2019.06.040>
- Wu, W., Zucca, C., Muhaimeed, A. S., Al-Shafie, W. M., Fadhil Al-Quraishi, A. M., Nangia, V., & Liu, G. (2018). Soil salinity prediction and mapping by machine learning regression in Central Mesopotamia, Iraq. *Land degradation & development*, 29(11), 4005–4014. <https://doi.org/10.1002/ldr.3148>
- Wu, W., Al-Shafie, W., & M., Muhaimeed, A. S., Ziadat, F., Nangia, V., Payne, W. B. (2014). Soil salinity mapping by multiscale remote sensing in Mesopotamia, Iraq. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, 7(11), 4442–4452. <https://doi.org/10.1109/JSTARS.2014.2360411>
- Yahiaoui, I., Bradaï, A., Douaoui, A., Abdennour, M., & A. (2021). Performance of random forest and buffer analysis of Sentinel-2 data for modelling soil salinity in the Lower-Chelif plain (Algeria). *International Journal of Remote Sensing*, 42(1), 148–171. <https://doi.org/10.1080/01431161.2020.1823515>

Zhu, S., Zhu, X., & M., Wang, Z, L., Liu, Z, U. (2012). Zeolite diagenesis and its control on petroleum reservoir quality of Permian in northwestern margin of Junggar Basin. *China. Science China Earth Sciences*, 55(3), 386–396. <https://doi.org/10.1007/s11430-011-4314-y>

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

Springer Nature or its licensor (e.g. a society or other partner) 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.