03
Research paper I Оригинални научни рад
DOI 10.7251/STP2215021A
ISSN 2566-4484

Uğur Avdan, Eskisehir Technical University, uavdan@eskisehir.edu.tr
Gordana Kaplan, Eskisehir Technical University, kaplangorde@gmail.com
Dilek Matci, Eskisehir Technical University, dkmatci@anadolu.edu.tr
Firat Erden, Eskisehir Technical University, firaterdem5@gmail.com
Zehra Yigit Avdan, Eskisehir Technical University, zyigit@eskisehir.edu.tr
Ece Mizik, Eskisehir Technical University, ecemizik@gmail.com
Ilknur Ozudogru, Eskisehir Technical University, ilknurozudogru@eskisehir.edu.tr

# EVALUATING MACHINE LEARNING MODELS FOR SOIL SALINITY ESTIMATION USING SATELLITE IMAGERY

***Abstract***

Salinity is one of the most critical problems for agricultural lands. Soil salinity should be monitored with fast, economical and accurate data and methods. In this study, soil salinity was estimated using remote sensing data and machine learning algorithms, where five different methods were used, and the results were compared. As a study area, Alpu, Turkey has been selected. Within the scope of the study, on-site measurements were made in cultivation areas where there are different agricultural products such as beets, wheat, tomatoes, and corn in the district. The results show that machine learning algorithms and Planetscope images successfully determine soil salinity. Future studies will evaluate the methods by taking samples from different product classes and wet/arid lands.

*Keywords: Remote Sensing, Soil salinity, Machine learning, PlanetScope.*

# ОЦЈЕНА МОДЕЛА МАШИНСКОГ УЧЕЊА ЗА ПРОЦЈЕНУ САЛИНИТЕТА ТЛА КОРИШТЕЊЕМ САТЕЛИТСКИХ СНИМАКА

***Сажетак***

Салинитет је један од најважнијих проблема за пољопривредна земљишта. Салинитет земљишта треба пратити брзим, економичним и тачним подацима и методама. У овом истраживању, салинитет земљишта је процијењен кориштењем података даљинске детекције и алгоритама машинског учења, гдје је кориштено пет различитих метода, а резултати су упоређени. За подручје истраживања одабран је град Алпу у Турској. У оквиру истраживања извршена су мјерења на лицу мјеста у подручјима узгоја различитих пољопривредних производа као што су репа, пшеница, парадајз и кукуруз. Резултати показују да алгоритми машинског учења и снимке сателитске констелације PlanetScope успјешно одређују салинитет земљишта. Будућа истраживања ће оцијенити ове методе узимањем узорака из различитих група производа и влажних/сушних земљишта.

*Кључне ријечи: даљинска детекција, салинитет земљишта, машинско учење, PlanetScope*

# 1. INTRODUCTION

Soil salinity is considered one of the most critical environmental problems, especially in arid and semiarid regions, as it causes land degradation and desertification [1,2]. This dynamic phenomenon can occur due to natural processes or human activities and can significantly threaten soil productivity and agricultural land [3]. Additionally, minimizing the risk of soil salinity decreases environmental issues and agricultural economic losses [4]. Due to the harmful effects of soil salinity on soil fertility and agricultural production, various practices are required to protect soil quality. The first step of these applications is to monitor the severity of soil salinity [5].

Salinity parameters, critical for ensuring soil sustainability, are often determined through on-site measurement and laboratory analysis. Electrical conductivity (EC) is used to determine soil salinity. It is measured using a saturated paste of soil samples, extracted in various water ratios, and on-site measuring methods. Collecting soil samples and conducting laboratory examinations is expensive and time-consuming [6]. Furthermore, using these approaches to detect salinity in large areas, dynamically monitor the temporal and spatial change of the salting process, and determine the salted zones is difficult [7]. Fast and cost-effective remote sensing technologies for determining and evaluating salinity-affected areas' geographical and temporal distribution have been developed in recent years [8,9].

Various salinity indices have been developed in recent years to detect salt-affected areas using satellite images, most of which are based on the spectral characteristics of soil salinity in various bands of satellite data [4,10]. With creating a regression model between these spectral indices and EC values, soil salinity maps can be obtained from satellite images. Traditional regression analysis methods and machine learning methods are the two primary regression approaches used to estimate soil salinity. Traditional regression analysis methods consist of least squares regression and partial least squares regression (PLSR) methods [11]. Machine learning methods are algorithms that learn solutions from data for decision-making and prediction about real-world problems. The most important feature that distinguishes it from traditional methods is that machine learning algorithms use a large amount of data to train the model and learn how to achieve tasks from the data [12].

Several attempts have been made to produce soil salinity maps using machine learning. Wang, *et al.* [13] have proposed a new spectral index and have used a neural network to produce soil moisture and salinity inversion. Similarly, Habibi, *et al.* [14] have investigated the quantitative assessment of soil salinity using remote sensing data based on the artificial neural network. Wang, *et al.* [15] have evaluated three different machine learning algorithms (Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN)) for soil salinity mapping with Sentinel-2 MSI data. In another study, Taghadosi, Hasanlou and Eftekhari [5] have used SVM to retrieve soil salinity from Sentinel-2 MSI images. Wu, *et al.* [16] have compared the SVM and RF methods for salinity mapping from a combined dataset consisting of Landsat 5 Thematic Mapper (TM) and ALOS L-band radar data. Wang, *et al.* [17] have integrated remote sensing and landscape characteristics to estimate soil salinity using machine learning methods. Wang, Chen, Wang and Li [11] have compared the performance of different machine learning algorithms for estimating the soil salinity of salt-affected soil using field spectral data.

In these studies, Landsat-8 OLI and Sentinel-2 MSI sensors have been used. However, since these sensors have medium spatial resolution and do not give frequent temporal data, the success of sensors with higher spatio-temporal resolution in determining soil salinity should be tested. Commercial satellite images, such as the PlanetScope, may be utilized to produce greater spatial resolution soil salinity maps. Daily high-resolution (3 m) Planetscope images can generate daily salinity indices because of their great geographical and temporal resolution. There has been no detailed investigation of soil salinity mapping from Planetscope images using different machine learning algorithms. This study set out to investigate the usefulness of different machine learning methods for soil salinity mapping from Planetscope images.

# 2. METHODS

## 2.1. STUDY AREA AND DATA

Eskişehir province is located northwest of the Central Anatolia Region of Turkey. This study was carried out in the Alpu district of Eskişehir Province, which has an area of 1,059.13 km$^2$ and an altitude of 700 m. When we examine the land distribution of the district, it is analyzed that 37.8% is agricultural land, 36.8% is forest land, 20.8% is meadow/pasture land and 4.6% is non-agricultural land. 37.8% of the study area is irrigated agricultural land and 62.5% is arid. The district's climate is a continental climate typical of the Central Anatolia Region, with hot and dry summers, rainy and cold winters, rainy and warm spring months, and deep and dry autumn months. The average annual precipitation is 398.1 kg, while the temperature in the district ranges from 30 to 38 ºC in summer and -5 to -2 ºC in winter (https://www.bebka.org.tr/admin/datas/sayfas/198/alpu-ilce-raporu_1568787633.pdf). The district has high agricultural crop productivity with its productive soils, and agricultural activities form the economic basis of the region [18,19].

Within the scope of the study, on-site measurements were made in cultivation areas where there are different agricultural products such as beets, wheat, tomatoes, and corn in the district. On 6 and 7 October 2020, the measurement was carried out using a random sampling method at soil surface in cultivated and harvested areas. By taking three different soil samples in the same point, the effects of the irrigation process on the electrical conductivity (EC) value were evaluated. During the measurement process, the coordinates of each measurement point were recorded using a handheld GPS device. The electrical conductivity values, which provide us with information about the salinity of the study area, were measured with the PNT 3000 COMBI+ device at the surveying points.

Planetscope consists of more than 120 nano-satellites manufactured by Planet Labs, Inc. The Planetscope system provides satellite images daily, high resolution (3 m) and 4-band (red, green, blue, and near-infrared).

## 2.2. METHODOLOGY

Soil samples of soil salinity were taken from different agricultural area types in the study area. The study made use of a PlanetScope imagery from October 7, 2020. Planetscope is a collection of more than 120 nano-satellites built by Planet Labs, Inc. The Planetscope system produces daily satellite photos with very high resolution (3 m) in four bands (red, green, blue, and near-infrared). Several spectral indices have been developed in the literature using remote sensing data for detecting and mapping salinity. Table 1 lists the spectral indices utilized in this research. While there are many salinity indices in the literature, because of the limitation of the spectral resolution of PlanetScope, we used the ones that only include RGB and NIR bands.

*Table 1. Details of the used salinity indices*

| Salinity Index (SI) | Formula | Reference |
|---|---|---|
| **SI – 1** | SI = $\sqrt{B \times R}$ | [20] |
| **SI – 2** | SI = $\sqrt{G \times R}$ | [20] |
| **SI – 3** | SI = $\sqrt{G^2 + R^2 + NIR^2}$ | [21] |
| **SI – 4** | SI = $\sqrt{G^2 + R^2}$ | [21] |
| **SI – 5** | SI = $\frac{B}{R}$ | [22] |
| **SI – 6** | SI = $\frac{B-R}{B+R}$ | [22] |
| **SI – 7** | SI = $\frac{G \times R}{B}$ | [22] |
| **SI – 8** | SI = $\frac{B \times R}{G}$ | [23] |
| **SI – 9** | SI = $\frac{NIR \times R}{G}$ | [23] |

Machine learning has been widely used in remote sensing application [24]. Decision tree algorithm have been resulted successfully in different studies. Decision tree learning is an approach often used in data mining. The objective is to develop a model that predicts the value of a target variable based on a number of input factors. Thus, this study aims at evaluating several decision tree algorisms for soil salinity prediction. Five tree algorithms have been evaluated in this study, namely, Decision Stump (DS), RF, M5P, Alternating Model Tree (AMT), Rapid Decision Tree (REPTree). Here we give a brief explanation of the used techniques.

DS are one-level decision trees primarily intended to be poor learners for boosting techniques. They are commonly called "one-rule" classifiers since they only predict class membership using a single characteristic. However, because the underlying metric for "best split" differs, executing these two methods will almost certainly result in somewhat different classifiers. The RF comprises many individual decision trees that work together as an ensemble. Each individual tree in the random forest produces a class prediction, and the class with the most votes becomes the prediction of our model. The M5P model tree reconstructs the M5 method [25], which is based on the standard decision tree but includes a linear regression function at the leaves nodes. The decision tree represents the algorithms in the form of a tree that has been educated using data to produce nodes [26]. In a single tree structure, alternating decision trees give the predictive capability of decision tree ensembles. They are a type of option tree, which are decision trees that have been enhanced with option nodes and produced via boosting [27]. RepTree is a rapid decision tree learner that constructs a decision/regression tree using information gain as the splitting criteria and prunes it with the reduced error pruning technique [28].

The five algorithms have been applied to our dataset. In order to investigate the classification algorithms, five statistical parameters have been evaluated, Correlation Coefficient (CC) indicating the specific measure that quantifies the strength of the relationship between the variables in a correlation analysis, Mean absolute error (MAE), errors between paired observations, Root mean squared error (RMSE), is the square root of the mean of the square of all of the error. RMSE is a good measure to compare prediction errors of different models, Relative absolute error (RAE) which is the magnitude of the difference between the exact value and the approximation, and Root relative squared error (RRSE) that is relative to what it would have been if a simple predictor had been used. The data in the dataset have been randomly divided into 70% training and 30% testing.

## 3. RESULTS

The results of the investigation are given in Table 2. According to the results, M5P resulted in the highest CC, while REPTree was most successful in the other parameters. The M5P was followed by AMT, DS, RF, and REPTree in the CC. Figure 2 and Eq (1) present the tree models of M5P and REPTree. While REPTree uses SI – 8, SI – 1, and SI – 5 in the tree decision, M5P uses SI – 3, SI – 5, and SI – 7 in the model equation, where SI – 5 has the highest weight. Using the two models, soil salinity prediction can be made and visualized.

*Table 2. Results of the investigated tree algorithms*

| Method | CC | MAE | RMSE | RAE | RRSR |
|---|---|---|---|---|---|
| RF | 0.83 | 3.23 | 4.27 | 70.56 | 83.31 |
| AMT | 0.88 | 3.72 | 4.05 | 81.39 | 79.04 |
| M5P | **0.92** | 3.44 | 3.73 | 75.29 | 72.81 |
| REPTree | 0.82 | **2.41** | **2.97** | **52.70** | **57.97** |
| DS | 0.84 | 3.40 | 4.43 | 74.28 | 86.33 |

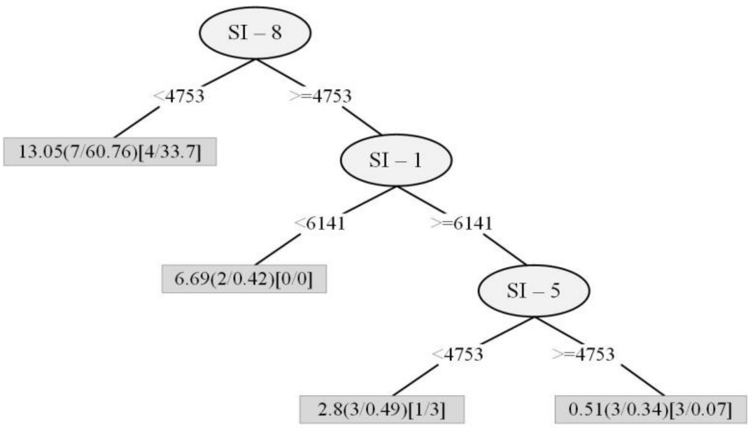


Figure 1. *Results of REPTree; tree construction*

The values on the lines connecting nodes reflect the splitting criterion depending on the parent node feature values. The value before the parentheses in the leaf node represents the categorization value. In addition, the first value in the first parenthesis represents the total number of cases from the training set in that leaf. In contrast, the second value represents the number of instances improperly categorized in that leaf. The first value in the second parenthesis, on the other hand, is the total number of occurrences from the pruning set in that leaf. The second value is the number of instances in that leaf that were mistakenly categorized.

The M5P model uses a smoothed linear model. With one rule, the results showed a significantly high correlation. The rule is presented in Eq. 1. Using Eq. 1, a soil salinity prediction map has been produced, shown in Figure 3.

$$EC = -0.0027 \text{ x SI3} - 68.8814 \text{ x SI5} + 0.0033 \text{ x SI7} + 80.1276 \quad (1)$$

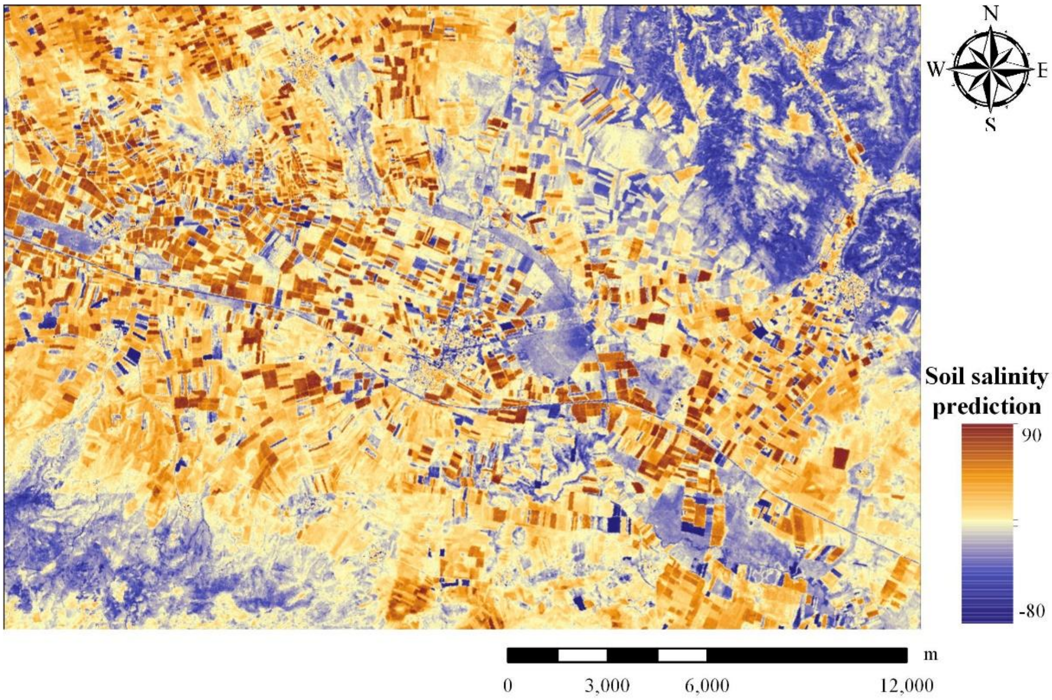


Figure 2. *Soil salinity prediction map*

## 4. DISCUSSION

Salinity is one of the most critical problems for agricultural lands lost due to wrong operations in our world. Especially in arid and semiarid regions, approximately half of the irrigated agricultural lands have salinity problems of varying degrees—good protection of limited resources and especially agricultural lands against the rapidly increasing world population. Therefore, soil salinity should be monitored with fast, economic and accurate data and methods. The fact that on-site measurements require much time and are not economical makes monitoring salinity with remotely sensed data advantageous.

In this study, soil salinity was estimated using remotely sensed data and machine learning algorithms, and the method with the best results was determined. In this direction, five different machine learning algorithms were used. Results show that the REPTree and M5P methods gave the best results. In other studies in the literature, it is seen that these algorithms give very successful results [29]. t is also stated that it is an advantage in that it is a flexible method and gives the best results in estimation processes in different fields [30].

In addition, it has been noticed that the SI5 index is used for estimation in both methods. SI5 can be used to estimate salinity in terms of both ease of calculation and performance. It is seen that this index, which is calculated by the ratio of the blue and green bands of the Planetscope satellite, also gives successful results in salinity studies calculated by statistical methods [31].

The results show that machine learning algorithms and Planetscope images successfully determine soil salinity. The methods will be evaluated in future studies by taking samples from different product classes and wet/arid lands.

## 5. CONCLUSION

Soil salinity caused by global climate change, which the Paris Climate Agreement often raises, has reached dangerous levels. In addition to its environmental effects, soil salinity has become a factor that causes soil fertility to decrease and restricts the economic gain obtained from agricultural activities. Accordingly, the studies conducted to determine, monitor, manage soil salinity in the Central Anatolia Region of Turkey, the economy based on agricultural activities, and restore the regions affected by soil salinity have gained significant importance. For this reason, it is very advantageous to monitor soil salinity with the use of remotely sensed data in a fast, economical, and not very time-consuming. In this study, soil salinity was estimated using machine learning algorithms with remotely detected data of agricultural areas of the Alpu district, where agricultural activities are carried out intensively in the Central Anatolia Region, and the methods that give the algorithms with remotely detected data of agricultural areas of the Alpu district, where agricultural activities are carried out intensively in the Central Anatolia Region, and the methods that give the best results were selected. Five different machine learning algorithms were used to determine soil salinity. Among these five algorithms, it has been determined that the REPTree and M5P methods, which have given very successful results in the literature, give the best results. In addition to the fact that the methods give better results in estimating soil salinity in different areas, they are also quite a flexible method. Additively the SI5 index, which is common to both methods used to estimate soil salinity, supports its availability both for the convenience of estimating and for performance. As a result, it was determined that machine learning algorithms and Planetscope images were successfully used to determine the soil salinity.

Further research is needed to test the potential of machine learning algorithms in estimating and mapping soil salinity. Also, machine learning algorithms used to estimate soil salinity should be studied in vast areas with different climates, geology, geomorphology, land use, and vegetation to evaluate their potential for use in different areas.

**ACKNOWLEDGMENT**

This study was supported by Eskisehir Technical University Scientific Research Projects Commission under grant no: 19ADP084.

**LITERATURE**

[1] Asfaw, E.; Suryabhagavan, K.; Argaw, M. Soil salinity modeling and mapping using remote sensing and gis: The case of wonji sugar cane irrigation farm, ethiopia. *Journal of the Saudi Society of Agricultural Sciences* 2018, *17*, 250-258.

[2] Li, H.; Webster, R.; Shi, Z. Mapping soil salinity in the yangtze delta: Reml and universal kriging (e-blup) revisited. *Geoderma* 2015, *237*, 71-77.

[3] Gorji, T.; Sertel, E.; Tanik, A. Monitoring soil salinity via remote sensing technology under data scarce conditions: A case study from turkey. *Ecological Indicators* 2017, *74*, 384-391.

[4] Allbed, A.; Kumar, L. Soil salinity mapping and monitoring in arid and semi-arid regions using remote sensing technology: A review. *Advances in remote sensing* 2013, *2013*.

[5] Taghadosi, M.M.; Hasanlou, M.; Eftekhari, K. Retrieval of soil salinity from sentinel-2 multispectral imagery. *European Journal of Remote Sensing* 2019, *52*, 138-154.

[6] El Harti, A.; Lhissou, R.; Chokmani, K.; Ouzemou, J.-e.; Hassouna, M.; Bachaoui, E.M.; El Ghmari, A. Spatiotemporal monitoring of soil salinization in irrigated tadla plain (morocco) using satellite spectral indices. *International Journal of Applied Earth Observation and Geoinformation* 2016, *50*, 64-73.

[7] Peng, J.; Biswas, A.; Jiang, Q.; Zhao, R.; Hu, J.; Hu, B.; Shi, Z. Estimating soil salinity from remote sensing and terrain data in southern xinjiang province, china. *Geoderma* 2019, *337*, 1309-1319.

[8] Corwin, D.; Scudiero, E. Review of soil salinity assessment for agriculture across multiple scales using proximal and/or remote sensors. *Advances in agronomy* 2019, *158*, 1-130.

[9] Liu, Y.; Zhang, F.; Wang, C.; Wu, S.; Liu, J.; Xu, A.; Pan, K.; Pan, X. Estimating the soil salinity over partially vegetated surfaces from multispectral remote sensing image using non-negative matrix factorization. *Geoderma* 2019, *354*, 113887.

[10] Elhag, M. Evaluation of different soil salinity mapping using remote sensing techniques in arid ecosystems, saudi arabia. *Journal of Sensors* 2016, *2016*.

[11] Wang, S.; Chen, Y.; Wang, M.; Li, J. Performance comparison of machine learning algorithms for estimating the soil salinity of salt-affected soil using field spectral data. *Remote Sensing* 2019, *11*, 2605.

[12] Ahmad, S.; Kalra, A.; Stephen, H. Estimating soil moisture using remote sensing data: A machine learning approach. *Advances in water resources* 2010, *33*, 69-80.

[13] Wang, J.; Wang, W.; Hu, Y.; Tian, S.; Liu, D. Soil moisture and salinity inversion based on new remote sensing index and neural network at a salina-alkaline wetland. *Water* 2021, *13*, 2762.

[14] Habibi, V.; Ahmadi, H.; Jafari, M.; Moeini, A. Quantitative assessment of soil salinity using remote sensing data based on the artificial neural network, case study: Sharif abad plain, central iran. *Modeling Earth Systems and Environment* 2021, *7*, 1373-1383.

[15] Wang, J.; Peng, J.; Li, H.; Yin, C.; Liu, W.; Wang, T.; Zhang, H. Soil salinity mapping using machine learning algorithms with the sentinel-2 msi in arid areas, china. *Remote Sensing* 2021, *13*, 305.

[16] Wu, W.; Zucca, C.; Muhaimeed, A.S.; Al-Shafie, W.M.; Fadhil Al-Quraishi, A.M.; Nangia, V.; Zhu, M.; Liu, G. Soil salinity prediction and mapping by machine learning regression in c entral m esopotamia, i raq. *Land degradation & development* 2018, *29*, 4005-4014.

[17] Wang, N.; Xue, J.; Peng, J.; Biswas, A.; He, Y.; Shi, Z. Integrating remote sensing and landscape characteristics to estimate soil salinity using machine learning methods: A case study from southern xinjiang, china. *Remote Sensing* 2020, *12*, 4118.

[18] Taşpınar, K.; Ateş, Ö.; Özge Pınar, M.; Yalçın, G.; Kızılaslan, F.; Fidantemiz, Y.F. Soil contamination assessment and potential sources of heavy metals of alpu plain eskişehir turkey. *International Journal of Environmental Health Research* 2021, 1-9.

[19] Derakhshandeh, M.; Tombul, M. Calibration of metric modeling for evapotranspiration estimation using landsat 8 imagery data. *Water Resources Management* 2021, 1-25.

[20] Khan, N.M.; Rastoskuev, V.V.; Sato, Y.; Shiozawa, S. Assessment of hydrosaline land degradation by using a simple approach of remote sensing indicators. *Agricultural Water Management* 2005, *77*, 96-109.

[21] Nicolas, H.; Walter, C. Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. *Geoderma* 2006, *134*, 217-230.

[22] Bannari, A.; Guedon, A.; El-Harti, A.; Cherkaoui, F.; El-Ghmari, A. Characterization of slightly and moderately saline and sodic soils in irrigated agricultural land using simulated data of advanced land imaging (eo-1) sensor. *Communications in soil science and plant analysis* 2008, *39*, 2795-2811.

[23] Khan, S.; Abbas, A. Using remote sensing techniques for appraisal of irrigated soil salinity. *Int. Congr. Model. Simul.(MODSIM), Model. Simul. Soc. Aust. New Zealand, Bright* 2007, 2632-2638.

[24] Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine* 2017, *5*, 8-36.

[25] Quinlan, J.R. In *Learning with continuous classes*, 5th Australian joint conference on artificial intelligence, 1992; World Scientific: pp 343-348.

[26] Blaifi, S.-a.; Moulahoum, S.; Benkercha, R.; Taghezouit, B.; Saim, A. M5p model tree based fast fuzzy maximum power point tracker. *Solar Energy* 2018, *163*, 405-424.

[27] Frank, E.; Mayo, M.; Kramer, S. In *Alternating model trees*, Proceedings of the 30th annual ACM symposium on applied computing, 2015; pp 871-878.

[28] Patel, R.R.; Aluvalu, R. A reduced error pruning technique for improving accuracy of decision tree learning. *International Journal of Engineering and Advanced Technology (IJEAT)* 2014, *3*, 8-11.

[29] Wang, F.; Shi, Z.; Biswas, A.; Yang, S.; Ding, J. Multi-algorithm comparison for predicting soil salinity. *Geoderma* 2020, *365*, 114211.

[30] Sepahvand, A.; Singh, B.; Sihag, P.; Nazari Samani, A.; Ahmadi, H.; Fiz Nia, S. Assessment of the various soft computing techniques to predict sodium absorption ratio (sar). *ISH Journal of Hydraulic Engineering* 2019, 1-12.

[31] Avdan, U.; Kaplan, G.; AVDAN, Z.Y.; MATÇI, D.K.; ERDEM, F.; MIZIK, E.; Ozudogru, I. Comparison of remote sensing soil electrical conductivity from planetscope and ground measured data in wheat and beet yields. 2021.