CHILEAN JOURNAL
of AGRICULTURAL RESEARCH

CHILEANJAR
CHILEAN JOURNAL of AGRICULTURAL RESEARCH

RESEARCH ARTICLE

Salinization evaluation in arable land using unmanned aerial vehicles, multispectral images, and machine learning

Patricia Paulina Hernández-Victoria[1], Héctor Flores-Magdaleno[1*], Abel Quevedo-Nolasco[1], Jorge Flores-Velázquez[1], and Gustavo Cruz-Cárdenas[2]

[1]Colegio de Postgraduados, Campus Montecillo, Departamento de Hidrociencias, Texcoco 56264, México.
[2]Instituto Politécnico Nacional, CIIDIR Unidad Michoacán, COFAAA, Jiquilpan, Michoacán 59510, México.
*Corresponding author (mhector@colpos.mx).
Received: 5 November 2025; Accepted: 6 February 2026, doi:10.4067/S0718-58392026000300369

ABSTRACT

Soil salinity reduces crop productivity and promotes desertification. Traditional methods to assess soil salinity are costly and require a lot of time and effort. Salinity diagnosis could be more efficient by processing aerial images. This work aimed to predict the spatial distribution of soil salinity and sodicity as a function of electrical conductivity (EC) and the exchangeable Na percentage (ESP) in agricultural lands. Three hundred soil samples were collected randomly in the study area to determine in a laboratory the EC and the ESP. The reflectance of multispectral images obtained with a camera transported in an unmanned aerial vehicle was used to calculate several reflectance indices which were used to predict EC and ESP categories. Five automated learning algorithms: Neural Networks, Random Forest, Support Vector Machine, Nearest Neighbors, and Decision Trees were trained and validated. The results of the Random Forest model indicated an accuracy of 0.69, recall of 0.69, F1 of 0.69, and kappa of 0.39 to predict EC categories. The Support vector machine was the most satisfactory model for predicting the ESP, obtaining an accuracy of 0.73, recall of 0.65, F1 of 0.69, and kappa of 0.41. We recommend monitoring the study area at different year times, by identifying and mapping areas with salinity problems, farmers could implement more effective management practices, which could increase crop productivity.

Key words: Artificial neural network, decision tree, random forest, soil salinity, support vector machine, unmanned aerial vehicles.

INTRODUCTION

Soil is a natural resource of public interest under increasing environmental pressure. It must be managed sustainably for the benefit of future generations since it is a non-renewable resource, i.e., once lost, it cannot be recovered (Bouslihim et al., 2021). Current problems include soil salinization and sodification, which cause a yield decrease of 10%-25% for many crops, can prevent harvesting altogether when severe, and lead to desertification. Salinity problems occur under all climatic conditions; globally, about 2000 ha of arable land are lost daily (Shahid et al., 2018).

Soil salinity is the second leading cause of land degradation after erosion; it has been a cause of the decline of agricultural societies for 10 000 yr. Historical records of migrations caused by the salinization of arable soil showcase the failure of civilizations due to the increased salinity of agricultural fields. The best-known case is Mesopotamia, now Iraq (Leidi and Pardo, 2002).

Salinity has damaged approximately 20% of the world's land, and according to some estimates, this area is rapidly increasing due to irregular climate changes and anthropogenic activities (Arora and Bhatla, 2017). Excess salts in the soil significantly threaten agricultural productivity, environmental health, and financial well-being. Considering the extent of salt-affected soils, the cost of land degradation in 2013 was US$441 per hectare, and the estimated annual economic losses are US$27 billion dollars at present (Shahid et al., 2018). Corwin (2020) indicated that the consequences of climate change have been overlooked, and it will be necessary to monitor and map changes in the degree of soil salinity.

Monitoring soil salinity and sodicity is imperative to bring science, practice, and policy together to review the status and challenges associated with their management. Conventional methods for monitoring soil salt accumulation involve periodic field visits to collect samples and multiple laboratory analyses, which require workforce, time, and money (Vermeulen and Van Niekerk, 2017). Another disadvantage of these methods is data-scarce regions or inaccessible areas; in these cases, soil salinity has been quantified using remote sensing data and automated learning algorithms (Gorji et al., 2017). Salinity mapping helps improve existing and potential soil salinization regions through land management. For soil salinity assessment, high spatial resolution data are preferred (Ge et al., 2022).

The application of unmanned aerial vehicles (UAVs) implies an increase in efficiency in the agro-morphological characterization process, reducing time in obtaining parameters in the field, the UAVs in conjunction with multispectral cameras are very useful tools for calculating vegetation indices (Cruz-Grimaldo et al., 2025).

Recently, automated learning has been used in digital soil mapping to obtain the relationship between spatial changes in soil properties and environmental variables and infer the spatial distribution of soil properties. This approach can more effectively assess the nonlinear relationships between soil and environmental covariates (Ding et al., 2020). Many automated learning algorithms have been evaluated for soil property prediction, including Support Vector Machines, Regression Trees, Decision Trees, Neural Networks, Random Forests, and Stochastic Gradient Algorithms (Fathololoumi et al., 2020).

The objective of this study was to predict the spatial distribution of soil salinity as a function of electrical conductivity (EC) and the exchangeable Na percentage (ESP) in agricultural lands of the Chapala Marsh in the State of Michoacán. The reflectance of multispectral images obtained with a camera transported in an unmanned aerial vehicle was used to calculate several reflectance indices which were used to train and validate five automated learning algorithms in the prediction of EC and ESP categories.

MATERIALS AND METHODS

Study area

The study area comprises 100 ha of soil located at an altitude of 1540 m in the municipality of Villamar, Michoacan, Mexico. The site is characterized by a temperate climate, with rainfall in summer and an average temperature between 12 and 18 °C. The coldest month is between -3 and 18 °C, the hottest month is below 22 °C, and the average annual rainfall is 900 mm. The main crops are corn (*Zea mays* L.), sorghum (*Sorghum* spp.), onion (*Allium cepa* L.), alfalfa (*Medicago sativa* L.), and tomato (*Solanum lycopersicum* L.), while the predominant soils are Vertisols and Fluvisols (Estrada et al., 2023).

During the dry season, 300 simple soil samples were collected randomly at surface level (0-30 cm) according to NOM-021-RECNAT-2000 between 13 and 15 May 2022 (SEMARNAT, 2002) (Figure 1). The coordinates of each sampling point were recorded with a portable global positioning system (GPS). Samples were air-dried, ground, and sieved with a 2 mm aperture. For sample processing, we used the laboratory analytical method according to Rice et al. (2012) to determine the following salinity parameters in the saturation extract: Electric conductivity (EC), Ca^{2+}, Mg^{2+} and Na^{+}.

Sodium adsorption ratio (SAR) was calculated using Equation 1, which analyzes the effect of Na concentration on soil physical properties (Aceves, 2011) and uses cation exchange reactions between polyvalent ions. The Na^{+}, Ca^{2+}, and Mg^{2+} ions concentration is expressed as mEq L^{-1}.

$$\mathrm{SAR} = \frac{\mathrm{Na}^{+}}{\sqrt{\frac{\mathrm{Ca}^{2+} + \mathrm{Mg}^{2+}}{2}}} \qquad (1)$$

The exchangeable Na percentage (ESP) in relation to the SAR was also calculated according to Equation 2. This parameter was calculated to define the position and relevance of the Na^{+} ion within the exchange complex (Pérez, 2019):

$$\mathrm{ESP} = \frac{100(-0.126 + 0.01475\ \mathrm{SAR})}{1(-0.126 + 0.01475\ \mathrm{SAR})} \quad (2)$$

Salinity-affected soil was classified according to the USDA (1975): Normal (EC < 4 dS m^{-1} and ESP < 15%), saline (EC ≥ 4 dS m^{-1} and ESP < 15%), sodic (EC < 4 dS m^{-1} and ESP ≥ 15%) and saline-sodic (EC ≥ 4 dS m^{-1} and ESP ≥ 15%). Categories 1 and 2 (saline or non-saline) were obtained for both EC and ESP. The values were modeled to predict the corresponding category in the whole study area and not only in the sampling points, using different applications of Machine Learning (ML) algorithms.

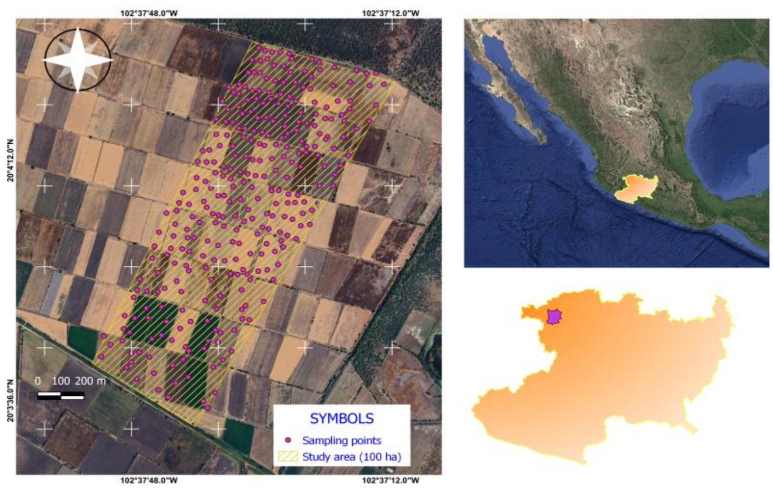

Figure 1. Study area and spatial distribution where 300 samples were taken in Villamar Michoacán, México.

Multispectral camera images and an unmanned aerial vehicle (UAV)

The study area of 100 ha of bare soil was flown over with a drone (eBee, SenseFly, Santiago, Chile) with real time kinematic (rtk) global positioning system (GPS). The geographic coordinate points were recorded using a GPS-Global Navigation Satellite System (GNSS) receiver station (Hi-Target, Guangzhou, China) with a vertical accuracy range of 0.5 m, a horizontal accuracy of 3 cm in "x" and "y". A Sequoia multispectral camera (Parrot, Paris, France) with a 16 Mpx RGB sensor was attached to the drone, also incorporating four single band sensors with a spectral bandwidth in green (530-570 nm), red (640-680 nm), red edge (730-740 nm) and near-infrared (NIR) (770-810 nm) of 1.2 Mpx, and a 35 g sun sensor for real-time illumination correction. The multispectral images were processed photogrammetrically and radiometrically to obtain a reflectance map for each band using the PIX4Dmapper program (Pix4d SA, Lausanne, Switzerland). The four image bands were used to calculate several spectral indices to model soil salinity, images with a resolution of 22 cm per pixel were obtained.

Spectral indices

An 8-band layer stack image was generated using several soil salinity spectral indices to model and predict the soil salinity categories. These indices were derived from the reflectance values of the multispectral images obtained with the green, red-edge, red, and NIR bands (Table 1). Figure 2 shows the original orthomosaic of the study area with the pertinent corrections. The four bands captured by the multispectral camera, green, red-edge, red, and NIR, are also shown, as well as the spectral indexes calculated: SI14, soil adjusted vegetation index (SAVI), normalized difference vegetation index (NDVI), normalized difference salinity index (NDSI).

Table 1. Spectral indices used to model and map soil salinity.

Resource	Index number	Spectral index	Equation
Multispectral images	1	Green	G
	2	Red-edge	RE
	3	Red	R
	4	Near infrared	NIR
	5	Salinity index 14	$\mathrm{SI14} = \frac{\mathrm{R} \times \mathrm{NIR}}{\mathrm{G}}$
	6	Soil adjusted vegetation index	$\mathrm{SAVI} = \frac{(\mathrm{NIR} - \mathrm{R})}{(\mathrm{NIR} + \mathrm{R} + \mathrm{L} \times (1 + \mathrm{L}))}$
	7	Normalized differential vegetation index	$\mathrm{NDVI} = \frac{(\mathrm{NIR} - \mathrm{R})}{(\mathrm{NIR} + \mathrm{R})}$
	8	Normalized differential salinity index	$\mathrm{NDSI} = \frac{(\mathrm{R} - \mathrm{NIR})}{(\mathrm{R} + \mathrm{NIR})}$

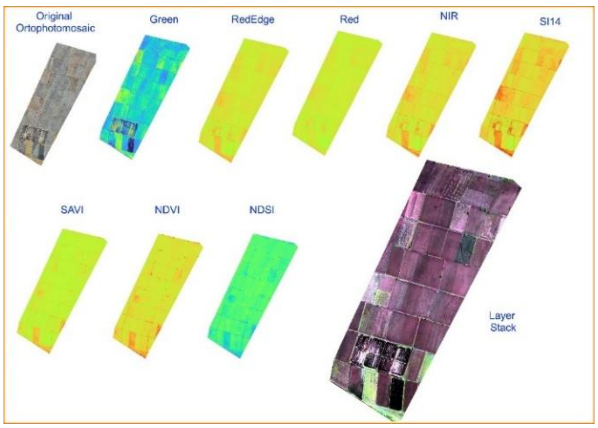

Figure 2. Layers and spectral soil salinity indices used to model the soil salinity (Layer stack). The R, G and B channel bands are 1, 2 and 3 respectively. NIR: Near infrared; SI14: salinity index 14; SAVI: soil adjusted vegetation index; NDVI: normalized differential vegetation index; NDSI: normalized differential salinity index.

Soil salinity modeling using machine learning algorithms

Five machine learning (ML) algorithms were used to model the soil salinity, which were trained and validated: Decision Trees (DT), Random Forest (RF), Artificial Neural Network (ANN), Super Vector Machine (SVM), and K-Nearest Neighbors (k-NN).

The DTs are classified as non-parametric methods. A tree represents a set of hierarchically organized conditions applied successively from the root to a terminal node or leaf of the tree (Breiman, 1984). Every decision tree starts with a node that includes all the sample cases to be classified; this node is called root. Internal nodes and terminal nodes, which are also called leaves, can also be distinguished.

The RF is a non-parametric supervised machine learning algorithm (Breiman, 2001) with a well-known high classification and regression accuracy. This algorithm generates multiple DTs that are constructed based on randomly selected subsets by bootstrap aggregation (bagging) (Breiman, 1996) on a training dataset; the results are combined to obtain a single model more robust than the results of each separate tree (Lizares, 2017).

A special case of ML algorithms is ANNs. They are similar to human neurons in terms of their output capacity. In the specific case of Deep Learning, these algorithmic structures allow models with multiple processing layers to learn data representations with multiple abstraction levels; these models perform a series of linear and nonlinear transformations that generate an output close to the expected one from the input data. In this case, supervised learning consists of obtaining the parameters of these transformations (the weights and bias) and ensuring that these transformations are optimal, i.e., the produced output and the expected output differ very little (Torres et al., 2018).

The SVMs are an example of supervised learning models that minimize a learning model's expected error and reduce the overfitting problem (Yu et al., 2016). An SVM first maps the input points to higher dimensional feature space (i.e., if the input points are in two dimensions, they are mapped by the SVM to three dimensions) and then finds a hyperplane that separates them and maximizes at margin m between classes in this space (Betancourt, 2005).

The k-NN algorithm is one of the most widely used algorithms in machine learning. Due to its practicality and efficiency, it makes no assumptions about data distribution and trains quickly (Lantz, 2019). It is a non-parametric algorithm, does not require prior knowledge about the dataset, and assumes that instances in the datasets are independently and identically distributed; therefore, instances close to each other have the same classification (Pooja, 2017).

Validation of results

The reflectance indices were used to train and validate five automated learning algorithms in the prediction of electrical conductivity (EC) and exchangeable Na percentage (ESP) categories. Once the categories of EC and ESP were predicted, a map salinity was generated according to the parameters established by the USDA.

This was done using the QGIS 3.28.7 program (QGIS, Grüt, Switzerland), with the Orfeo Toolbox software, which implements the different algorithms to train the used models. Each model dataset was assigned a calibration percentage (75%) and a validation percentage (25%). To measure the detection performance and compare the five classification methods, we used the criteria of Flores et al. (2021) the three most common metrics for measuring the visualization effectiveness of detection algorithms corresponding to Equations 3-5 were calculated:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

$$\text{F1} = \frac{2}{\frac{1}{\text{prec}} + \frac{1}{\text{recall}}} = \frac{\text{TP}}{\text{TP} + \frac{\text{FN} + \text{FP}}{2}} \tag{5}$$

where TP is the number of true positive detections, i.e., those in which salinity was detected and confirmed by laboratory measurements; FP is the number of false positives, i.e., cases in which salinity was detected, but the laboratory results indicate otherwise, and FN corresponds to the not detected.

"Precision" or positive predicted value (PPV) is the percentage of correct detections (Equation 3); "Recall" or true positive rate (TPR) measures the sensitivity of the entire positive class, including true and false, i.e., it shows the percentage of correctness (Equation 4). "F1" measures the overall classification performance as it combines "Accuracy" and "Recall", and high accuracy will only be obtained if both "Recall" and "Accuracy" are high (Equation 5). These data are presented in a confusion matrix, an analytical tool that evaluates the thematic accuracy of a given classification (as a response to the classification model under evaluation) against a reference classification assumed to be the true value (obtained from data taken in situ during field sampling). This matrix measures the degree of agreement or coincidence between the evaluated and reference classifications (Porskamp et al., 2018; Rajendran et al., 2018). Another indicator resulting from the confusion matrix is the Cohen's kappa statics (Equation 6); this is used to evaluate algorithms (Sinha and Kumar, 2013). Kappa is widely used, especially when evaluating change, since change detection can be seen as a simple classification problem with only two categories:

$$\text{kappa} = \frac{\text{P}_0 - \text{P}_\text{e}}{1 - \text{P}_\text{e}} \tag{6}$$

where P_0 is the total agreement probability, and P_e is the agreement probability due to chance. Kappa values range between 0 and 1, and values closer to 1 indicate a higher matching strength. Table 2 shows the categories proposed by Landis and Koch (1977).

Finally, Figure 3 shows a general diagram summarizing the processes carried out to achieve a spatial salinity prediction in the total study area. The parameter settings for the used model are presented in Table 3.

Table 2. Parameter tuning and standardization of models for electrical conductivity (EC) and exchangeable Na percentage (ESP) prediction. DT: Decision tree; RF: Random Forest; SVM: Super Vector Machine; ANN: Artificial Neural Network; k-NN: k Nearest Neighbors.

Model to predict ESP and EC	Standardization
DT	Maximum depth of the tree: 10
	Minimum number of samples in each node: 10
RF	Maximum depth of the tree: 10
	Minimum number of samples in each node: 10
	Maximum number of trees in the forest: 200
ANN	Number of neurons in each intermedia layer: 32
SVM	Training classificator: libSVM, kernel type radial basis function (rbf), and model csvc
K-NN	Number of neighbors: 32

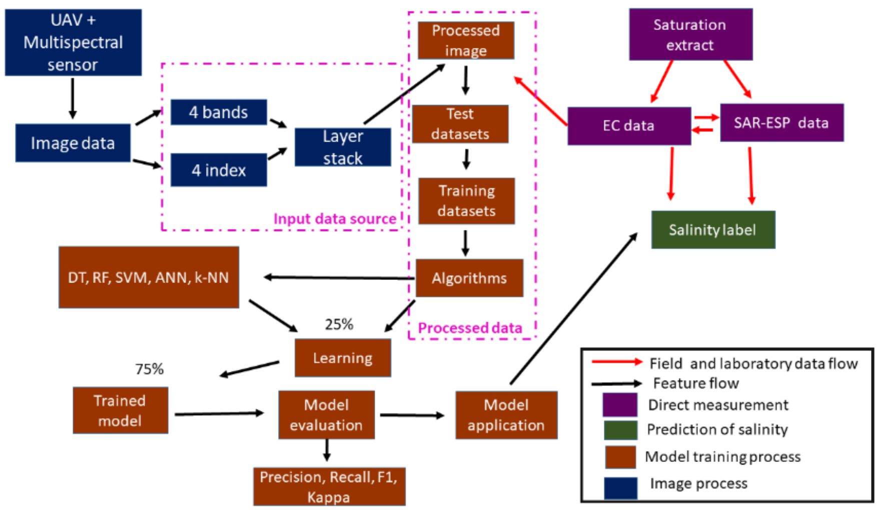

Figure 3. Experimental flow diagram and general process of Machine Learning application for the salinity prediction dataset. DT: Decision tree; RF: Random Forest; SVM: Super Vector Machine; ANN: Artificial Neural Network; k-NN: k Nearest Neighbors; EC: electrical conductivity; SAR: Na adsorption ratio; ESP: exchangeable Na percentage.

Table 3. Classification of Kappa coefficient categories (Landis and Koch, 1977).

Kappa coefficient	Matching strength
Poor	< 0.00
Slight	0.01-0.20
Acceptable	0.21-00.40
Moderate	0.41-0.60
Good	0.61-0.80
Very good	0.81-1.00

RESULTS AND DISCUSSION

Degree of salinization and sodification in the study area

According to the results obtained in the 300 simple soil samples the electrical conductivity (EC) ranged from 0.62 to 38.8 dS m^{-1}, with a mean of 3.48 ± 4.18 dS m^{-1}. The results show great heterogeneity since they are in the intervals corresponding to the classification of categories from non-saline to extremely saline soils. Manual 60 defines several general categories of soil salinity: 0-2 dS m^{-1} not saline, 2-4 dS m^{-1} slightly saline, 4-8 dS m^{-1} moderately saline, 8-16 dS m^{-1} strongly saline and > 16 dS m^{-1} extremely saline (Richards, 1954). Other researchers have already reported this high variability in soil salinity, indicating that high soil conductivity is due to natural and anthropogenic issues (Silva et al., 2006). In this case, one of the natural causes for high conductivity is location: The soils are located in a geothermal zone known as "Los Negritos". The second cause is that some soils are irrigated with water from deep wells, which has various salinity values ranging from 1.0 to 4.5 dS m^{-1}, leading to a salinization process over time (Silva et al., 2002). The values obtained for Na adsorption ratio (SAR) and exchangeable Na percentage (ESP) indicate values ranging from non-sodic to extremely sodic soils, as shown in Table 4.

Table 4. Statistical summary of the soil properties: Electrical conductivity (EC), Na adsorption ratio (SAR) and exchangeable Na percentage (ESP).

Parameter	Minimum	Maximum	Average	First quartile	Second quartile	Third quartile	Standard deviation
EC, dS m^{-1}	0.62	38.80	4.64	1.62	3.48	6.57	4.18
SAR	0	42.00	5.76	1.60	4.50	8.52	5.83
ESP	0	37.77	6.38	1.08	4.53	10.14	6.54

Model performance

During validation, the models for ESP prediction achieved precision scores of 0.61-0.76, recall scores of 0.56-0.65, and F1-scores of 0.62-0.69. While all models were acceptable according to the Kappa index, the Support Vector Machine (SVM) model provided the best fit, categorized as 'moderate'. The SVM model yielded an accuracy of 0.73, a recall of 0.65, an F1-score of 0.69, and a Kappa index of 0.42 (Figure 4).

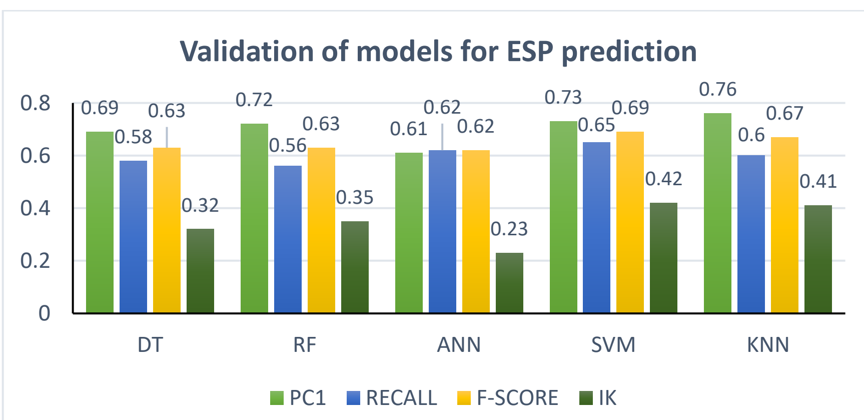

Figure 4. Comparison of different classifiers to evaluate the quality of the algorithms used to predict exchangeable Na percentage (ESP). DT: Decision tree; RF: Random Forest; SVM: Super Vector Machine; ANN: Artificial Neural Network; k-NN: k Nearest Neighbors; IK: Kappa index.

For EC prediction, the same models demonstrated precision between 0.57-0.69, recall between 0.64-0.79, and F1-scores between 0.62-0.70. All models were again classified as acceptable. The Random Forest (RF) algorithm was the top performer for EC, with a precision of 0.68, recall of 0.69, F1-score of 0.69, and a Kappa index of 0.37 (Figure 5).

Several digital soil mapping studies have been performed worldwide, the most common are those derived from satellite images; however, the use of unmanned aerial vehicle (UAV) images has emerged recently. For example, Merembayev et al. (2022) used ML techniques and a comprehensive set of remote sensing data to predict soil salinity; they obtained an accuracy, recall, and F1 of 0.67 for RF, while for DT they were 0.67, 0.44, and 0.53 respectively. Of note is that these authors did not perform in situ sampling, but they used additional features such as elevation and temperature points to increase the model score.

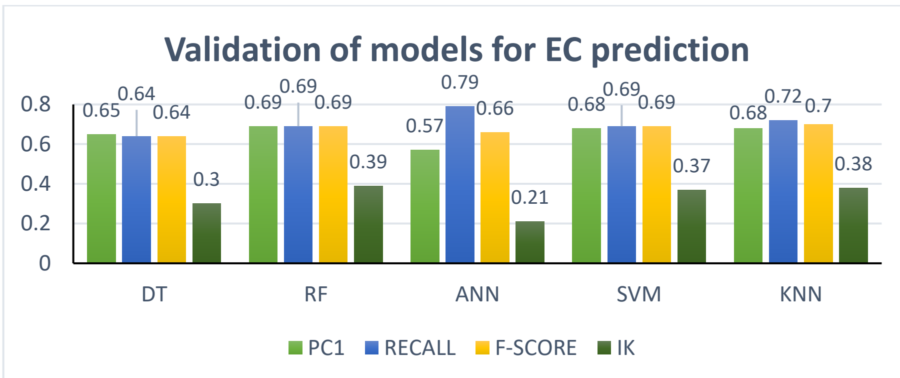

Figure 5. Comparison of different classifiers to evaluate the quality of the algorithms used to predict electrical conductivity (EC). DT: Decision tree; RF: Random Forest; SVM: Super Vector Machine; ANN: Artificial Neural Network; k-NN: k Nearest Neighbors; IK: Kappa index.

Conversely, Haq et al. (2023) collected 195 samples in Pakistan and simultaneously obtained spectral data from the same region. They applied different algorithms, and RF also outperformed the algorithms with an accuracy of 0.60, recall of 0.57, and F1 of 0.57. The authors indicate that the combination of remote sensing technology and machine learning models is effective for assessing soil types and salinity at the local scale and that this dataset could be expanded with soil samples. Likewise, Gu et al. (2022) studied data obtained from Landsat 8 to extract saline soil with different salinity indices with a U2 Network; the authors took 159 composite soil samples and found recall, accuracy, and F1 values between 0.90 and 0.95. The experimental results were different depending on the different salinity indices, which also agrees with the study of Allbed et al. (2014). These authors observed that adding salinity indices to the model produced a higher agreement (recall, precision and F1) than multiple bands alone. Overlapping the salinity index does not necessarily produce higher accuracy. Hypothetically, accumulating more salinity indices allows the model to bear more information; however, the results of this study and others (Allbed et al., 2014; Gu et al., 2022) do not support that hypothesis. Instead, overlapping excessive salinity indices could reduce the accuracy of the classification.

Other soil salinity mapping studies using satellite images have calculated kappa indices similar to those found in this work. For example, Mehrjardi et al. (2008) obtained a kappa of 0.47 with a linear regression model applied to 16 900 ha with 47 samples collected in Iran. Conversely, Meza and Acuña (2013) found a kappa of 0.92 with backpropagation artificial neural networks (ANNs) on the coast of Peru, with 190 samples collected. Foronda and Colinet (2023) collected 135 samples in a basin in Bolivia and obtained good predictions of EC using Super Vector Machine (SVM) (kappa: 0.58); they also found that ESP is better predicted with Random Forest (RF) (kappa: 0.82). These results also agree with Wang et al. (2015), who described that RF worked better than SVM for predicting salinity from spectral data and field-measured parameters.

Xie et al. (2022) worked in a coastal area of China; they collected 174 samples (for both bare soil and vegetation) in 3.83 km^2 using multispectral cameras coupled to a UAV and satellite images. Coupling data from both sources, the best models were RF and SVM with kappa of 0.88. The best-fitted and most repeated models were provided by SVM and RF, at least for soil salinity; these results agree with the present study, although we also found that K-Nearest Neighbors (k-NN) yields satisfactory predictions for both EC and ESP.

Compared to previous studies, the present work has more samples taken in situ in a smaller area, indicating a greater sampling effort. It should be emphasized that the success of digital soil mapping depends on several factors, such as high quality of appropriate predictors, representativeness of field observation, calibration of soil-landscape relationships, and rationality of model validation methods. However, selecting appropriate environmental factors remains a challenge (Wei et al., 2021). Each model has its own advantages and disadvantages, and a single model is usually not perfect. One possible approach is to combine different models to benefit from their advantages in modeling (Abedi et al., 2020). The choice of model and variables importantly influences the prediction of soil salinity. Direct monitoring of very slight and minor salinities using multispectral imaging is limited because different soil salinities have different spectral properties (Ge et al., 2022).

Spatial distribution of EC and ESP

Quantifying the uncertainty of the resulting maps is essential and confirms their applicability for management decision-making procedures. The SVM was used as the best predictive model for mapping the spatial distribution of ESP, and EC was the best for RF because they obtained the highest values in the validation metrics. The spatial distribution of the predicted EC and ESP is shown in Figure 6, where the highest sodicity is observed in the southern area of the study; however, EC has a greater spatial and heterogeneous distribution.

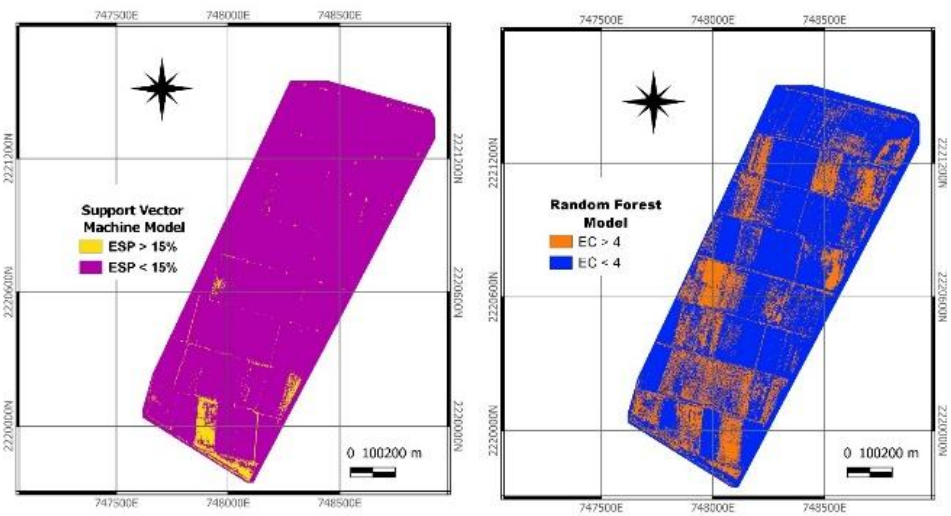

Figure 6. Spatial distribution of soil salinity obtained from the best predictive models. On the left side, the algorithm used is Support Vector Machine for exchangeable Na percentage (ESP) prediction, while on the right side, it is Random Forest for electrical conductivity (EC) prediction.

According to the USDA, the 100 ha property was classified into four categories: 66 ha were classified as normal soil, 29.77 ha as saline, 2.44 ha as sodic, and 1.52 ha as saline-sodic (Figure 7). Using this classification, the prediction generated by the best models indicates the Na ratio is intensifying in the southern zone but improved towards the north of the study area. Moreover, the prediction of high EC is heterogeneous but has a wider distribution than ESP in the area since it is found throughout the property but in different proportions. The distribution of salts in the soil is heterogeneous because salinization is a complex and variable process in space and time at different observation scales.

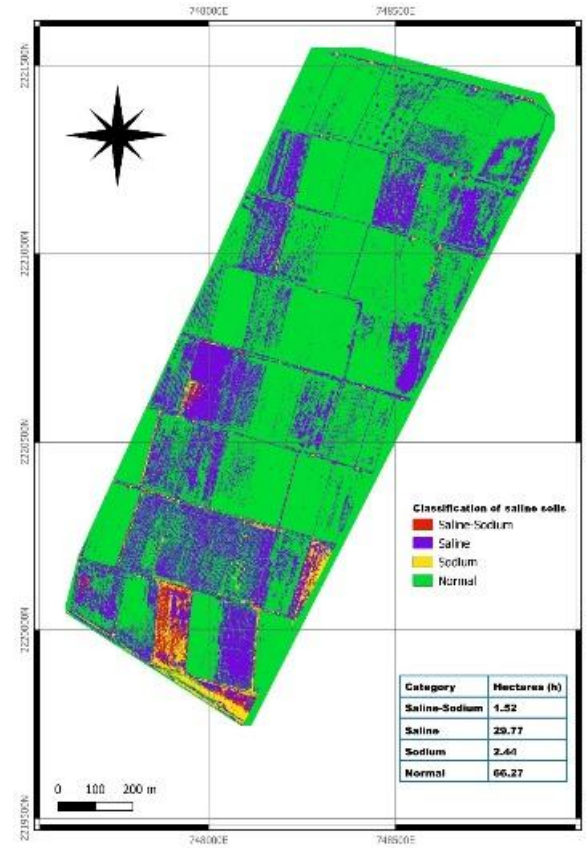

Figure 7. Final prediction of the spatial distribution of four salinity categories according to the USDA in the agricultural plots of Villamar, Michoacan.

CONCLUSIONS

The performance of five machine learning algorithms for modeling and mapping electrical conductivity (EC) and exchangeable Na percentage (ESP) in agricultural plots in the state of Michoacán was evaluated. The results indicated moderate to acceptable performance in the five models, where Random Forest and Support Vector Machine outperformed the others in predicting EC and ESP respectively. There is a high heterogeneity of salt distribution in the field from non-saline to extremely saline and from non-sodic to extremely sodic; however, the southern zone was characterized by higher salinity and sodicity. Future studies should carry out sampling during the rainy season to observe changes in soil salt values

This study demonstrates that the combination of unmanned aerial vehicle imagery and machine learning algorithms is an effective tool for diagnosing and monitoring soil salinity.

Author contribution

Conceptualization: P.H-V., H.F-M. Data curation: P.H-V. Formal analysis: P.H-V., H.F-M. Funding acquisition: H.F-M. Investigation: P.H-V., H.F-M. Methodology: P.H-V., H.F-M. Project administration: H.F-M. Software: P.H-V., H.F-M. Supervision: H.F-M. Validation: P.H-V. Visualization: P.H-V. Writing-original draft: P.H-V. Writing review and editing: P.H-V., H.F-M., A.Q-N., J.F-V., G.C-C. All co-authors reviewed the final version and approved the manuscript before submission.

Acknowledgements

The authors would like to thank the CONAHCYT for the financial support provided and to the College of Postgraduados, for the resources provided to carry out this investigation.

References

Abedi, F., Amirian, A., Faraji, M., Taghizadeh, R., Kerry, R., Razmjoue, D., et al. 2020. Salt dome related soil salinity in southern Iran: Prediction and mapping with averaging machine learning models. Land Degradation and Development 32:1540-1554.

Aceves, E. 2011. El ensalitramiento de los suelos bajo riego. Colegio de Postgraduados. 2nd ed. Mundi Prensa, México.

Allbed, A., Kumar, L., Aldakheel, Y.Y. 2014. Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial resolution imageries: Applications in a date palm dominated region. Geoderma 230-231:1-8.

Arora, D., Bhatla, S. 2017. Melatonin and nitric oxide regulate sunflower seedling growth under salt stress accompanying differential expression of Cu/Zn SOD and Mn SOD. Free Radical Biology and Medicine 106:315-328.

Betancourt, G.A. 2005. Las máquinas de soporte vectorial (SVMs). Scientia et Technica 1:67-72.

Bouslihim, Y., Rochdi, A., Paaza, N.E.A. 2021. Machine learning approaches for the prediction of soil aggregate stability. Heliyon 7:e06480.

Breiman, L. 1984. Classification and regression trees. Chapman and Hall/CRC, New York, USA.

Breiman, L. 1996. Bagging predictors. Machine Learning 26:123-140.

Breiman, L. 2001. Random forests. Machine Learning 45:5-32.

Corwin, D.L. 2020. Climate change impacts on soil salinity in agricultural areas. European Journal of Soil Science 72:842-862.

Cruz-Grimaldo, C., Nieves, M., Vera, E., Duran, M., Morales, A., Salazar, W., et al. 2025. Yield predictions of ‘Del Cerro’ cotton (*Gossypium hirsutum* L.) germplasm by multispectral monitoring in the north coast of Peru. Chilean Journal of Agricultural Research 85:15-26. doi:10.4067/S0718-58392025000100015.

Ding, J., Yang, S., Shi, Q., Wei, Y., Wang, F. 2020. Using apparent electrical conductivity as indicator for investigating potential spatial variation of soil salinity across seven oases along Tarim River in Southern Xinjiang, China. Remote Sensing 12:2601.

Estrada, F., Cruz, G., Ochoa, S., Teodoro, J. 2023. Cartografía digital de suelos con regresión-Kriging y datos de sensores remotos. Terra Latinoamericana 41:e1617.

Fathololoumi, 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.

Flores, D., González, I., Lozano, R., Vazquez, J.M., Hernandez, J.L. 2021. Automated agave detection and counting using a convolutional neural network and unmanned aerial systems. Drones 5:4.

Foronda, D., Colinet, G. 2023. Prediction of soil salinity/sodicity and salt-affected soil classes from soluble salt ions using machine learning algorithms. Soil Systems 7:47.

Ge, X., Ding, J., Teng, D., Wang, J., Huo, T., Jin, X., et al. 2022. Updated soil salinity with fine spatial resolution and high accuracy: The synergy of Sentinel-2 MSI, environmental covariates and hybrid machine learning approaches. Catena 212:106054.

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.

Gu, Q., Han, Y., Xu, Y., Ge, H., Li, X. 2022. Extraction of saline soil distributions using different salinity indices and deep neural networks. Remote Sensing 14:4647.

Haq, Y.U., Shahbaz, M., Asif, S., Ouahada, K., Hamam, H. 2023. Identification of soil types and salinity using MODIS Terra Data and machine learning techniques in multiple regions of Pakistan. Sensors 23:8121.

Landis, J.R., Koch, G.G. 1977. The measurement of observer agreement for categorical data. Biometrics 33:159-174.

Lantz, B. 2019. Machine learning with R. 3rd ed. Packt Publishing, Birmingham, UK.

Leidi, E.O., Pardo, J.M. 2002. Tolerancia de los cultivos al estrés salino: qué hay de nuevo. Revista de Investigaciones de la Facultad de Ciencias Agrarias-UNR 2(2):70-91.

Lizares, M. 2017. Comparación de modelos de clasificación: regresión logística y árboles de clasificación para evaluar el rendimiento académico. Licenciatura. Universidad Nacional Mayor de San Marcos, Lima, Perú.

Mehrjardi, R.T., Mahmoodi, S.H., Taze, M., Sahebjalal, E. 2008. Accuracy assessment of soil salinity map in Yazd-Ardakan Plain, Central Iran, based on Landsat ETM+ imagery. American-Eurasian Journal of Agricultural and Environmental Science 3:708-712.

Merembayev, T., Amirgaliyev, Y., Saurov, S., Wójcik, W. 2022. Soil salinity classification using machine learning algorithms and radar data in the case from the south of Kazakhstan. Journal of Ecological Engineering 23:61-67.

Meza, R.R.B., Acuña, J.R. 2013.Clasificación de la salinidad del suelo mediante imágenes de satélite y las redes neuronales artificiales. Revista ECIPerú 10(1):4-8.

Pérez, A. 2019. Riesgo de sodicidad en los suelos de Cantón Milagro, Guayas-Ecuador en época de estiaje. Revista Politécnica 42:15-22.

Pooja, R.A. 2017. Review of various KNN Techniques. International Journal for Research in Applied Science and Engineering Technology 5:1174-1179.

Porskamp, P., Rattray, A., Young, M., Lerodiaconou, D. 2018. Multiscale and hierarchical classification for benthic habitat mapping. Geosciences 8:145.

Rajendran, S., Nasir, S., El-Ghali, M.A., Alzebdah, K., Salim Al-Rajhi, A., Al-Battashi, M. 2018. Spectral signature characterization and remote mapping of Oman exotic limestones for industrial rock resource assessment. Geosciences 8:145.

Rice, E.W., Baird, R.B., Eaton, A.D., Clesceri, L.S. 2012. Standard methods for the examination of water and wastewater. American Public Health Association (APHA), Washington D.C., USA.

Richards, L.A. 1954. Diagnosis and improvement of saline and alkali soils. USDA Agriculture Handbook, US Department of Agriculture, Washington, DC, USA.

SEMARNAT. 2002. Norma Oficial Mexicana NOM-021-RECNAT-2000. Diario Oficial de la Federación. Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT), Ciudad de México, México. Available at https://www.dof.gob.mx

Shahid, S.A., Zaman, M., Heng, L. 2018. Soil salinity: Historical perspectives and a world overview of the problem. p. 43-53. In Zaman, M., Shahid, S.A., Heng, L. (eds.) Guideline for salinity assessment, mitigation and adaptation using nuclear and related techniques. Volume 1. Springer Cham, Gewerbestrasse, Switzerland.

Silva, J., Castillo, R., Ochoa, S., López, S. 2002. Lake Chapala and the Cienega aquifer: Chemical evidence of hydraulic communication. Geofísica Internacional 41:63-73.

Silva, J.T., Ochoa, S., Cristóbal, D., Estrada, F. 2006. Calidad química del agua subterránea de la Ciénega de Chapala como factor de degradación del suelo. Terra Latinoamericana 24:503-513.

Sinha, P., Kumar, L. 2013. Independent two-step thresholding of binary images in inter-annual land cover change/no-change identification. ISPRS Journal of Photogrammetry and Remote Sensing 81:31-43.

USDA. 1975. Soil taxonomy. A basic system of soil classification for making and interpreting soil surveys. Soil Survey Staff, USDA, Washington D.C., USA.

Torres, J., Viñals, J., Torres, T. 2018. Deep learning. Introducción práctica con Keras. 3rd ed. Watch this Space, Barcelona, España.

Vermeulen, D., Van Niekerk, A. 2017. Machine learning performance for predicting soil salinity using different combinations of geomorphometric covariates. Geoderma 299:1-12.

Wang, F., Chen, X., Luo, G., Han, Q. 2015. Mapping of regional soil salinities in Xinjiang and strategies for amelioration and management. Chinese Geographical Science 25:321-336.

Wei, Y., Ding, J., Yang, S., Wang, F., Wang, C. 2021. Soil salinity prediction based on scale-dependent relationships with environmental variables by discrete wavelet transform in the Tarim Basin. Catena 196:104939.

Xie, L., Feng, X., Zhang, C., Dong, Y., Huang, J., Cheng, J. 2022. A framework for soil salinity monitoring in coastal wetland reclamation areas based on combined unmanned aerial vehicle (UAV) data and satellite data. Drones 6:257.

Yu, L., Fu, H., Wu, B., Clinton, N., Gong, P. 2016. Exploring the potential role of feature selection in global land-cover mapping. International Journal of Remote Sensing 37:5491-5504.