VOL 2, NO.2 (2022) 1-10
https://jagst.utm.my

Journal of Advanced Geospatial
Science & Technology

Machine Learning Models for Estimating Soil Salinity Using Sentinel-1 SAR and Landsat-8 OLI Data

Ghada Sahbeni* and Balázs Székely
Department of Geophysics and Space Science, Eötvös Loránd University,
Budapest, Pázmány Péter stny, 1/A, 1117, Hungary
*Corresponding author: gsahbeni@caesar.elte.hu

***Abstract* -** As Hungary has the largest expanse of naturally salt-affected soils in Europe with a continuous decrease in groundwater levels due to climate change, the expansion of saline soils to the detriment of arable lands has become a potential risk that requires continuous monitoring to sustain agricultural productivity and ensure food security. The study aims to estimate soil salinity in the Great Hungarian Plain, Eastern Hungary, using Sentinel-1 Synthetic Aperture Radar (SAR) C-band and Landsat-8 OLI data combined with three state-of-the-art machine learning models, i.e., Artificial Neural Network with feature extraction (PCANNET), Random Forest (RF) and Support Vector Machine (SVM). For this purpose, seventy-four soil samples provided by the Research Institute of Soil Sciences and Agricultural Chemistry (RISSAC) were collected in the Hungarian Soil Information and Monitoring System framework between September and October 2016. We compared the predictive performance of machine-learning-based models using the root mean square error (RMSE) and the correlation coefficient (r). The results revealed that the SVM-based model outperformed the other machine learning models with an RMSE equal to 0.24 g/kg and a correlation coefficient of 0.73. The study demonstrates the efficiency of machine learning techniques as valuable alternatives to estimate soil salinity and assist in land management planning with affordable costs.

Keywords *- Landsat-8 OLI, Machine learning, Sentinel-1 SAR, Soil salinity*

Article History: Received 14 January 2022, Accepted 25 March 2022, Published 31 August 2022

How to cite: Sahbeni, G. and Székely, B. (2022). Machine Learning Models for Estimating Soil Salinity Using Sentinel-1 SAR and Landsat-8 OLI Data. Journal of Advanced Geospatial Science & Technology. 2(2), 1-10.

1. Introduction

In the last decades, extensive progress has been made to predict soil salinity at regional and local scales using data retrieved from optical and radar sensors (Taghadosi et al., 2019, Szatmári et al., 2020, Sahbeni, 2021a). Nevertheless, salinity assessment at the farm scale can be challenging due to the lack of funding for extensive field surveys that enable a continuous monitoring system over the vulnerable zone. Furthermore, dealing with high-resolution multispectral and SAR data in terms of preprocessing, processing and storage depict a crucial aspect to be considered by experts before initiating such expensive research projects on bigger scales. Many scientists have studied multispectral data effectiveness to detect salinization inland based on spectral indices, including Land Surface Temperature (LST), Normalized Difference Vegetation Index (NDVI), albedo, Canopy Response Salinity Index (CRSI), and Vegetation Soil Salinity Index (VSSI) (Scudiero et al., 2017, Sahbeni, 2021b). Due to its operation under all weather conditions, radar sensors represent valuable instruments in soil salinization investigation by retrieving SAR intensity values from different polarizations (VV, VH, HH, and HV) (Grissa et al., 2011, Muhetaer et al., 2022). Fewer studies have exploited machine learning tools using radar and multispectral data to model this dynamic environmental process with acceptable accuracy (Zarei et al., 2021, Aksoy et al., 2021, Chen et al., 2020). Ma et al. (2021) have successfully mapped salinization over The Ogan-Kuqa River Oasis based on Sentinel-1 and Sentinel-2 data using machine learning algorithms. As the XGBoost model showed its superiority (R^2 = 0.68), the combination of topographic variables, spectral indices and SAR features effectively improved the approach accuracy, revealing the role of multi-source data combination in enhancing the output quality. Similarly, Hoa et al. (2019) used Sentinel-1 SAR data to spatially model climate-change-induced salinization in Ben Tre Province (Vietnam) while comparing the statistical performance of five machine learning models. The study has shown that the Gaussian Processes outperformed other methods with the highest correlation coefficient (R = 0.81), serving as a valuable tool for policymakers to implement sustainable agricultural systems locally to reduce climate change impacts.

Our study aims to estimate soil salinity using remotely sensed data derived from Landsat-8 OLI and Sentinel-1 SAR sensors and compare the predictive performance of three machine learning-based models: neural network with feature extraction, regression random forest, and support vector machine.

2. Materials and Methods

2.1 Study Area

The study area is located in Eastern Hungary. It covers around 8322 km^2 over the Great Hungarian Plain (Figure 1). It is distinct by a moderately warm-dry climate with a mean annual precipitation of 560 mm and a mean yearly evaporation of 900 mm. Meadow chernozems and humic sandy soils dominate the landscape with a large agricultural land cover (Pásztor et al., 2018). Seventy-four soil samples were collected between mid-September and mid-October 2016 in the Hungarian Soil Information and Monitoring System framework from vegetated and non-vegetated areas. Salt content is measured in the laboratory from saturated paste according to the Hungarian Standard MSZ- 08-0206/2-1978 (JRC-IES, 2013). Values range between 0 and 5.6 g/kg of soil.

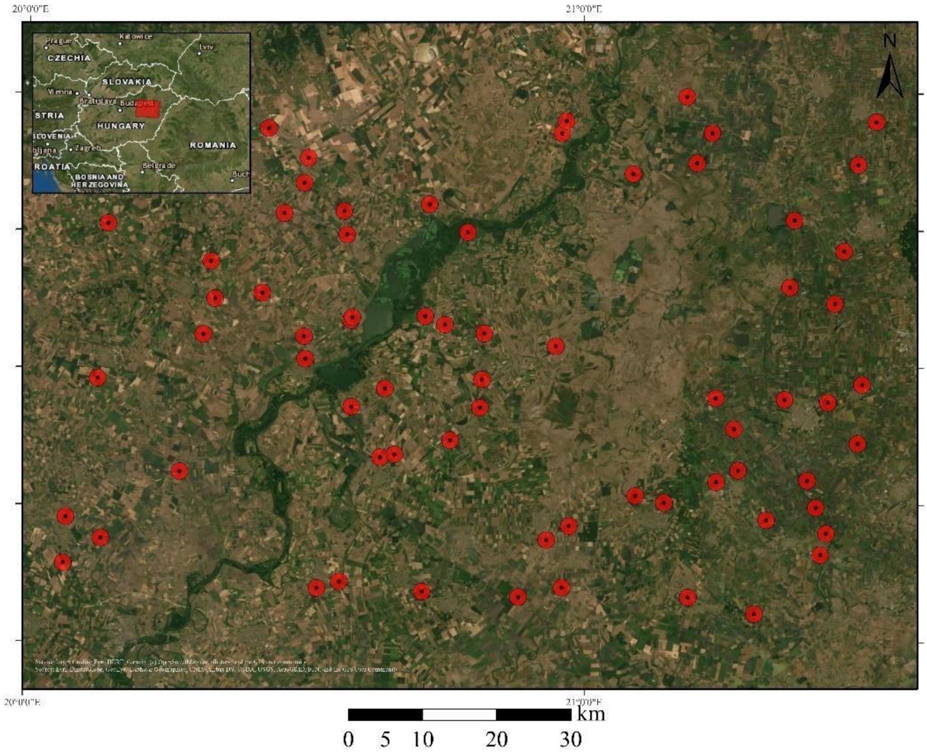

Figure 1. Location of sampling sites.

2.2 Dataset

Sentinel-1 SAR Ground Range Detected (GRD) product, acquired on 11 September 2016, was preprocessed using Sentinel-1 Toolbox. SAR data were radiometrically corrected, filtered from the speckle effect using a three-by-three Lee filter, and converted to a decibel scale. Then, Range Doppler Terrain Correction was applied.

Three Landsat-8 OLI Level-1 C products acquired on 20 May, 08 August, and 09 September 2016 were atmospherically and radiometrically corrected using the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm in ENVI IDL 5.3. Once multispectral data were preprocessed and stacked, spectral indices and principal component analysis (PCA) were derived to enhance data quality and reduce spectral dimensionality.

2.3 Methodology

We computed twenty spectral indices as presented in Table 1. Then spectral information was extracted from Landsat-8 OLI data using ArcMap 10.3. For SAR data, corresponding values were retrieved in intensity (dB) from the following features: VV, VH, VV – VH, and VV/VH. We used variables derived from Landsat-8 OLI and Sentinel-1 SAR to train three machine learning-based models, i.e., Neural network with feature extraction (PCANNET), Regression Random Forest (RF), and Radial Kernel Support Vector Machine (SVM). We used the following packages within R Studio to train and calibrate the models: ‘randomForest’ for RF, ‘e1071’ for SVM and ‘pcaNNET’ for PCANNET. An extensive explanation can be found in the R manuals. Then, we calculated the correlation coefficient (r) and root mean square error (RMSE) for accuracy assessment. Figure 2 summarizes the adopted methodology.

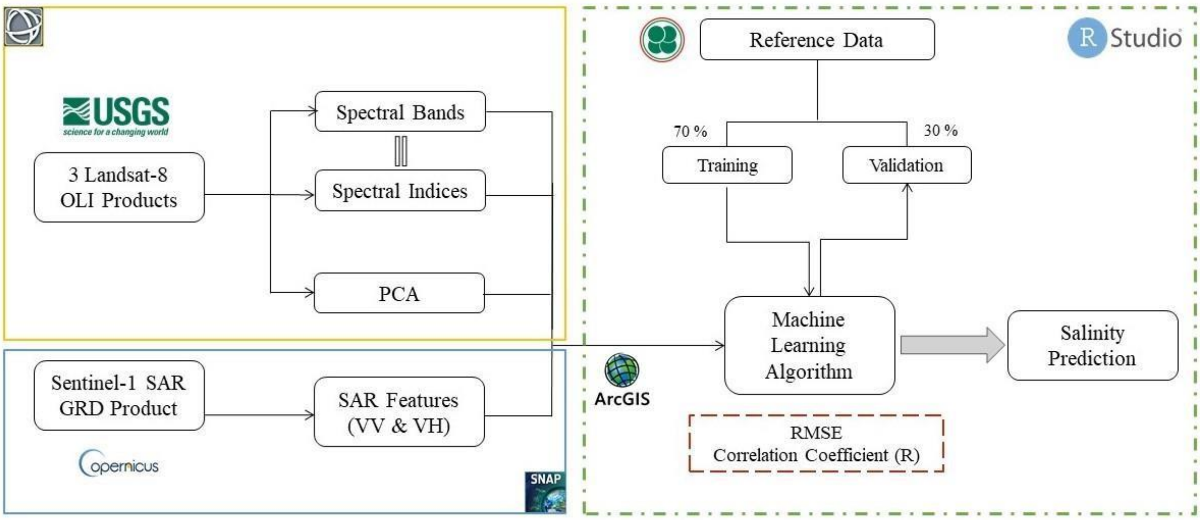

Figure 2. Methodology of the study.

Table 1. Spectral indices used in this study and their mathematical expressions.

Index	**Expression**
albedo	((0.356 × B) + (0.130 × R) + (0.373 × NIR) + (0.085 × SWIR1) + (0.072 × SWIR2) – 0.0018) /1.016 (Silva et al., 2016)
Differential Vegetation Index DVI	NIR – R (Basso et al., 2000)
Green Normalized Difference Vegetation Index GNDVI	(NIR – G) / (NIR + G) (Wu et al., 2014)
Intensity Index Int1	(G + R) / 2 (Bouaziz et al., 2011)
Intensity Index Int2	(G + R + NIR) / 2 (Bouaziz et al., 2011)
Normalized Difference Moisture Index NDMI	(NIR – SWIR1) / (NIR + SWIR1) (Wilson et al., 2002)
Normalized Difference Salinity Index NDSI	(R − NIR) / (R + NIR) (Khan et al., 2005)
Salinity Index 1 SI1	√ (G × R) (Douaoui et al., 2006)
Salinity Index 2 SI2	√ (NIR× R) (Dehni et al., 2012)
Salinity Index 3 SI3	√ (G^2 + R^2 + NIR^2) (Douaoui et al., 2006)
Salinity Index 4 SI4	√ (R^2 + G^2) (Yahiaoui et al., 2015)
Normalized Difference Vegetation Index NDVI	(NIR − R) / (NIR + R) (Rouse et al., 1974)
Soil Adjusted Vegetation Index SAVI	((NIR – R) / (NIR + R + L)) × (1 + L) (Huete et al., 1988)
Enhanced Vegetation Index EVI	2.5 × (NIR − R) / (NIR+C1 × R− C2 × B+L) (Huete et al., 2002)
Brightness Index BI	√ (R^2 + NIR^2) (Khan et al., 2005)
Bare Soil Index BSI	(G + NIR) / (G − NIR) (Li et al., 2013)
Normalized Pigment Chlorophyll Ratio Index NPCRI	(R − B) / (R + B) (Merzlyak et al., 1999)
Soil Salinity and Sodicity Index 1 SSSI1	SWIR1 − SWIR2 (Bannari et al., 2008)
Soil Salinity and Sodicity Index 2 SSSI2	(SWIR1 × SWIR2 – SWIR2 × SWIR2)/ SWIR1 (Bannari et al., 2008)
Normalized Difference Salinity Index VSSI	2 × G – 5 × (R + NIR) (Dehni et al., 2012)

3. Results

Table 2 illustrates the correlation coefficient (r) and RMSE results. The SVM-based model has the highest correlation coefficient (R = 0.73) and the lowest RMSE value (= 0.24). Figure 3 illustrates the relationship between measured normalized salt content (x) and predicted normalized salt content values (y). Its robustness can explain SVM superiority in cases of high dimensionality where the variables number is higher than the sample size. The advantage of using a radial basis kernel is eliminating overfitting issues caused by multicollinear variables. The results agree with studies by Jiang et al. (2019), Klibi et al. (2020), and Wang et al. (2021) regarding SVM algorithm effectiveness in soil salinity prediction. This research can be a scientific reference for determining the most suitable machine-learning algorithm to estimate soil salinity under semi-arid climates.

Table 2. Performance of three state-of-the-art machine learning models.

Model	**PCANNET**	**RF**	**SVM**
Correlation Coefficient (R)	0.65	0.71	0.73
RMSE (g/kg)	0.25	0.29	0.24

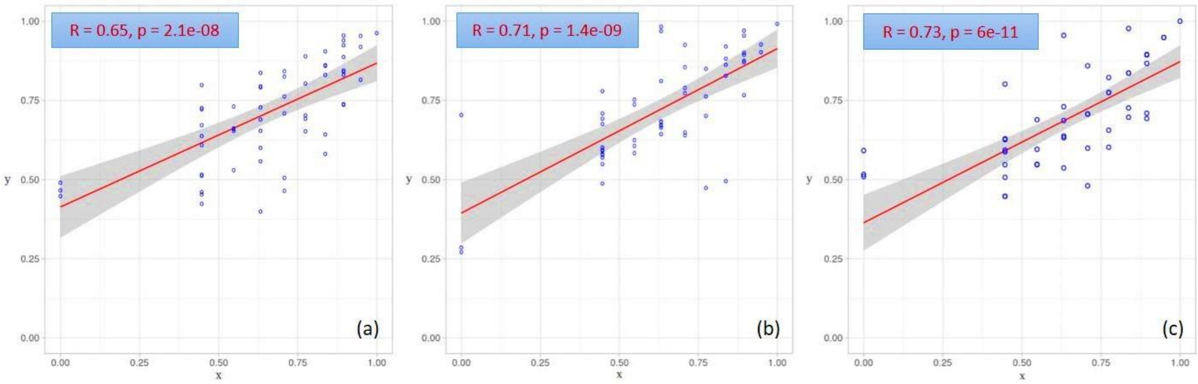

Figure 3. Relationship between measured (x) and estimated (y) Normalized Salt Content values using; (a) PCANNET-based model, (b) RF-based model, and (c) SVM-based model.

4. Conclusions

This study elucidates the potential of machine learning techniques combined with remote sensing tools in salinity prediction. Optical data retrieved from Landsat-8 OLI with Sentinel-1 SAR features have been used in this work. The SVM-based model showed superiority with a correlation coefficient of 0.73 and an RMSE of 0.24 g/kg. Nonetheless, more calibration is required to obtain optimal predictions. The preliminary findings revealed that combining data from multiple sources can be a promising approach in salinization monitoring with lower costs and a more sustainable environment at the early stages.

Acknowledgements

The authors would like to thank Prof. Pásztor L. for his support in acquiring field data from the Research Institute of Soil Sciences and Agricultural Chemistry (https://www.mta-taki.hu/hu). Landsat-8 OLI data were downloaded from the US Geological Survey (https://earthexplorer.usgs.gov) and Sentinel-1 SAR data from the Copernicus OpenAccessHub (https://scihub.copernicus.eu/dhus/#/home).

References

Aksoy, S., Yildirim, A., Gorji, T., Hamzehpour, N., Tanik, A. and Sertel, E. 2021. *"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. doi: 10.1016/j.asr.2021.10.024

Bannari, A., Guedon, A. M., El-Harti, A., Cherkaoui, F.Z. and El-Ghmari, A. 2008. *"Characterization Of Slightly And Moderately Saline And Sodic Soils In Irrigated Agricultural Land Using Simulated Data Of Advanced Land Imaging (EO-1) Sensor"*. Commun. Soil. Sci. Plant Anal., 39(19): 2795–811. doi: 10.1080/00103620802432717

Basso, F., Bove, E., Dumontet, S., Ferrara, A., Pisante, M., Quaranta, G. and Taberner, M. 2000. *"Evaluating Environmental Sensitivity At The Basin Scale Through The Use of Geographic Information Systems And Remotely Sensed Data: An Example Covering The Agri-Basin (Southern Italy)"*. Catena, (40): 19–35. doi: 10.1016/S0341- 8162(99)00062-4

Bouaziz, M., Matschullat, J. and Gloaguen, R. 2011. *"Improved Remote Sensing Detection of Soil Salinity From A Semi-Arid Climate In Northeast Brazil"*. CR Geosci, 343(11-12): 795–803. doi: 10.1016/j.crte.2011.09.003

Chen, Y., Qiu, Y., Zhang, Z., Zhang, J., Chen, C., Han, J. and Liu, D. 2020. *"Estimating Salt Content of Vegetated Soil At Different Depths With Sentinel-2 Data"*. PeerJ, 8, 10585. doi: 10.7717/peerj.10585

Dehni, A. and Lounis, M. 2012. *"Remote Sensing Techniques For Salt-Affected Soil Mapping: Application To The Oran Region of Algeria"*. Procedia Eng., (33): 188–98. doi: 10.1016/j.proeng.2012.01.1193

Douaoui, A. E. K., Nicolas, H., Walter, C. 2006. *“Detecting Salinity Hazards Within A Semi-Arid Context By Means of Combining Soil And Remote-Sensing Data”*. Geoderma, 134(1-2): 217. doi: 10.1016/j.geoderma.2005.10.009

Grissa,M., Abdelfattah, R., Mercier, G., Zribi, M., Chahbi, A. and Lili-Chabaane, Z. 2011. *“Empirical Model For Soil Salinity Mapping From SAR Data,”*. 2011 IEEE International Geoscience and Remote Sensing Symposium, 1099-1102. doi: 10.1109/IGARSS.2011.6049388

Hoa, P. V., Giang, N. V., Binh, N. A., Hai, L. V. H., Pham, T. D., Hasanlou, M. and Bui, D. T. 2019. *“Soil Salinity Mapping Using SAR Sentinel-1 Data And Advanced Machine Learning Algorithms: A Case Study At Ben Tre Province Of The Mekong River Delta (Vietnam)”*. Remote Sensing, 11(2). doi: 10.3390/rs11020128.

Huete, A. R. 1988. *“A soil-adjusted vegetation index (SAVI)”*. Remote Sens Environ., 25(3): 295–309. doi: 10.1016/0034-4257(88)90106-X

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X. and Ferreira, L. G. 2002. *“Overview of The Radiometric And Biophysical Performance of The MODIS Vegetation Indices”,* Remote Sens. Environ., 83: 195–213. doi: 10.1016/S0034-4257(02)00096-2

Jiang, H., Rusuli, Y., Amuti, T. and Qing, Q. 2019. *“Quantitative Assessment Of Soil Salinity Using Multi-Source Remote Sensing Data Based On The Support Vector Machine And Artificial Neural Network”*. International Journal of Remote Sensing, 40(1): 284-306. doi: 10.1080/01431161.2018.1513180

JRC- IES -European Commission Joint Research Centre Institute for Environment and Sustainability. 2013. *“European Hydro-pedological Data Inventory (EU-HYDI)”*.

Khan, N. M., Rastoskuev, V.V., Sato, Y. and Shiozawa, S. 2005. *“Assessment of Hydro Saline Land Degradation By Using A Simple Approach of Remote Sensing Indicators”*. Agricultural Water Management, 77(1-3): 96–109. doi: 10.1016/j.agwat.2004.09.038

Klibi, S., Tounsi, K., Rebah, Z. B., Solaiman, B. and Farah, I. R. 2020. *“Soil Salinity Prediction Using A Machine Learning Approach Through Hyperspectral Satellite Image”*. 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 1-6. doi: 10.1109/ATSIP49331.2020.9231870

Li, P., Jiang, L. and Feng, Z. 2013. *“Cross-Comparison Of Vegetation Indices Derived From Landsat-7 Enhanced Thematic Mapper Plus (ETM+) And Landsat-8 Operational Land Imager (OLI) Sensors”*. Remote Sensing, 6(1): 310–329. doi: 10.3390/rs6010310

Ma, G., Jianli, D., Han, L., Zhang, Z. and Ran, S. 2021. *“Digital Mapping of Soil Salinization Based on Sentinel-1 And Sentinel-2 Data Combined With Machine Learning*

Algorithms." Regional Sustainability, 2(2): 177-188.doi: 10.1016/j.regsus.2021.06.001

Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B. and Rakitin, V. Y. 1999. *"Non-Destructive Optical Detection of Leaf Senescence And Fruit Ripening"*, Physiol. Plant., 106(1): 135-141. doi: 10.1034/j.1399-3054.1999.106119.x

Muhetaer, N., Nurmemet, I., Abulaiti, A., Xiao, S. and Zhao, J. 2022. *"A Quantifying Approach To Soil Salinity Based On A Radar Feature Space Model Using ALOS PALSAR-2 Data"*. Remote Sensing, 14(2): 363. doi: 10.3390/rs14020363

Pásztor, L., Laborczi, A., Bakacsi, Z., Szabó, J. and Illés, G.2018. *"Compilation of A National Soil-Type Map For Hungary By Sequential Classification Methods".* Geoderma, 311: 93-108. doi: 10.1016/j.geoderma.2017.04.018

Rouse, J. W., Haas, R. H., Schell, J. A. and Deering, D. W. 1974. *"Monitoring vegetation systems in the Great Plains"*. S.C. Freden, E.P. Mercanti, M. Becker. Eds, with ERTS, Third Earth Resources Technology Satellite–1 Symposium, Vol. I: Technical Presentations, NASA SP-351, Washington, DC: NASA.

Sahbeni, G. 2021a. *"Soil Salinity Mapping Using Landsat 8 OLI Data and Regression Modeling In The Great Hungarian Plain"*. SN Applied Science, l.3(587): 1–13. doi: 10.1007/s42452-021-04587-4

Sahbeni, G. 2021b. *"A PLSR model to predict soil salinity using Sentinel-2 MSI data"*. Open Geosciences, 13(1): 977-987. doi:10.1515/geo-2020-0286

Scudiero, E., Corwin, D.L., Anderson, R.G., Yemoto, K., Clary, W., Wang Z. and Skaggs, T.H. 2017. *"Remote Sensing Is A Viable Tool For Mapping Soil Salinity In Agricultural Lands"*. Calif. Agriculture, 71(4): 231–238. doi:10.3733/ca.2017a0009

Silva, B. B., Braga, A. C., Braga, C.C., Oliveira de, L. M. M., Montenegro, S.M.G.L. and Barbosa Junior B. 2016. *"Procedures For Calculation Of The Albedo With OLI-Landsat 8 Images: Application To The Brazilian Semi-Arid"*. Revista Brasileira de Engenharia Agrícola e Ambiental, 20(1): 3-8. doi: 10.1590/1807-1929/agriambi.v20n1p3-8

Szatmári. G., Bakacsi. Z, and Laborczi, A. 2020. *"Elaborating Hungarian Segment of The Global Map of Salt-Affected Soils (Gssmap): National Contribution To An International Initiative".* Remote Sensing, 12(24): 4073. doi: 10.3390/rs12244073

Taghadosi, M., Hasanlou, M, and Eftekhari, K. 2019. "*Soil Salinity Mapping Using Dual-Polarized SAR Sentinel-1 Imagery"*. International Journal Remote Sensing, 40(1): 237-252. doi: 10.1080/01431161.2018.1512767

Wang, J., Peng, J., Li, H., Yin, C., Liu, W., Wang, T.and Zhang, H. 2021. *"Soil Salinity Mapping Using Machine Learning Algorithms With The Sentinel-2 MSI In Arid Areas, China,"*. Remote Sensing, 13(2): 305. doi: 10.3390/rs13020305

Wilson, E. H. and Sader, S. A. 2002. *"Detection Of Forest Harvest Type Using Multiple Dates Of Landsat TM Imagery"*. Remote Sensing Environment, 80(3): 385–396. doi: 10.1016/S0034-4257(01)00318-2

Wu, W., Mhaimeed, A.S, Al-Shafie, W.M., Ziadat, F., Dhehibi, B., Nangia ,V. and De Pauw, E. 2014. *"Mapping Soil Salinity Changes Using Remote Sensing In Central Iraq"*. Geoderma Reg, 2(3): 21–31. doi: 10.1016/j.geodrs.2014.09.002

Yahiaoui, I., Douaoui, A., Zhang, Q. and Ziane, A. 2015. *"Soil Salinity Prediction In The Lower Cheliff Plain (Algeria) Based On Remote Sensing And Topographic Feature Analysis".* J Arid Land., 7: 794–805. doi: 10.1007/s40333-015-0053-9

Zarei, A., Hasanlou, M. and Mahdianpari, M. 2021. *"A Comparison Of Machine Learning Models For Soil Salinity Estimation Using Multi-Spectral Earth Observation Data"*. ISPRS Ann. Photogramm. Remote Sensing Spatial Inf. Science, 3: 257. doi: 10.5194/isprs-annals-V-3-2021-257-2021