Developing Machine Learning Models to Predict Water Retention Properties of Soils of Sri Lanka

K.A.I.L Kasthuri Arachchi[1], D.N. Vidana Gamage,[2]*, R.B. Mapa[2]

[1]Postgraduate Institute of Agriculture[2] , Department of Soil Science, University of Peradeniya, 24000, Sri Lanka
**dumindavidana@agri.pdn.ac.lk*

Measuring soil water retention properties, such as volumetric water content (VWC) at field capacity (FC) and permanent wilting point (PWP), is a labor-intensive process, resulting in insufficient data for accurate irrigation scheduling. This study aimed to develop machine learning (ML) models to predict VWC at 10 (VWC_{10}), 33, (VWC_{33}), and 1500 kPa (VWC_{1500}) of Sri Lankan soils using a data set of an extensive soil survey. Soil properties such as soil texture, bulk density (BD), soil organic carbon (SOC) were used with Multiple Linear Regression (MLR), K-Nearest Neighbor (KNN), Random Forest (RF), and Support Vector Regression (SVR) models. The predictive accuracy was evaluated using the coefficient of determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Square Error (RMSE) across the entire dataset and various textural classes. Among the models tested, MLR outperformed ML models, achieving a satisfactory accuracy ($R^2 = 0.63$) in predicting VWC_{10} and VWC_{33} with the sand, silt, and bulk density combination. Furthermore, the combinations of sand and VWC_{10} and sand and VWC_{33} notably boosted the predictive accuracy of VWC_{1500} across all tested models (RMSE 3-4 %), with SVM and RF emerging as the top performers with R^2 values of 0.85 and 0.90, respectively. Nevertheless, when silt, SOC, and BD were added to the sand and VWC_{10} or sand and VWC_{33} combination, there was only a slight improvement in the predictive accuracy of VWC_{1500}. Results of the study revealed that the SVM and RF could be used to predict the VWC_{1500} of soils of Sri Lanka with higher accuracy using combinations of easily measured parameters like sand content and VWC_{10} or sand content and VWC_{33}. Future research will aim to enhance the prediction models for VWC_{10} and VWC_{33} by including a larger data set covering diverse soil types.

Keywords: Machine learning, Field capacity, Permanent wilting point, Water retention

Acknowledgment: University Research Grant (No. URG/2019/45/Ag) for financial assistance. Soil Science Society of Sri Lanka for the permission to use the data.