Predictive Modeling of Saturated Hydraulic Conductivity using Machine Learning Techniques

Moussa S. Elbisy
Civil Engineering Department, College of Engineering and Architecture, Umm Al-Qura University, Makkah, Saudi Arabia
mselbisy@uqu.edu.sa (corresponding author)

Received: 13 January 2025 | Revised: 31 January 2025 | Accepted: 5 February 2025

 DOI: https://doi.org/10.48084/etasr.10225

ABSTRACT

The hydraulic conductivity of saturated soil is a critical parameter for understanding various engineering challenges related to groundwater. Machine learning techniques offer powerful methods to address complex nonlinear regression problems. This study developed three models, namely a Multilayer Perceptron Neural Network (MPNN), a Support Vector Machine (SVM), and a Tree Boost, to predict field saturated hydraulic conductivity using easily measurable soil properties, such as hydraulic conductivity, clay/silt ratio, soil saturation percentage, d90 of grains, liquid limit, plastic limit, soil pH, hydrocarbon anions, chloride ions, and calcium carbonate content. Soil samples were collected from two locations: the El-Nubaria and Sinai regions, located in the western delta of Egypt. To evaluate the performance of these models, five distinct metrics, namely Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Scatter Index (SI), and Correlation Coefficient (R), were employed along with a Taylor diagram. Among the models tested, the Tree Boost model demonstrated exceptional accuracy in predicting field-saturated hydraulic conductivity, having a lower SI (0.085) compared to the SVM (0.192) and MPN (0.226) models. Moreover, the Tree Boost model exhibited a higher R value (0.99) than SVM (0.981) and MPN (0.974). The Tree Boost results were compared with those of previous models. The findings highlight the effectiveness of the Tree Boost model and suggest its potential as a reliable tool for estimating field-saturated hydraulic conductivity and generating highly accurate predictions.

Keywords-saturated soil hydraulic conductivity; prediction; support vector machines; tree boost; multilayer perceptron neural network; soil properties

I. INTRODUCTION

Saturated hydraulic conductivity (K_{sat}) is a critical soil characteristic that affects water flow rate, soil quality, pollutant and chemical transport, nutrient availability, plant water absorption, and crop growth, and also indicates the geometry, dimensions, and connectivity of soil pores [1]. K_{sat} is a critical hydraulic property in agriculture, as it reflects the productive potential of soil [2]. Predicting heat and mass transport in soil, as well as in distributed hydrological modeling, is essential [3]. Therefore, accurate prediction of K_{sat} is crucial to reduce uncertainty in models, thus enhancing their practical applicability. Various methods, including empirical formulas, laboratory tests under steady or transient conditions in representative samples, tracer tests, auger hole tests, and pumping tests in wells, can be used to determine hydraulic conductivity in saturated zones [4]. A thorough examination of predictive techniques for saturated soils was presented in [5]. Due to the complexity of the phenomenon at the particle scale, establishing analytical relationships between the hydraulic conductivity of a specific soil and its governing characteristics is challenging, as these relationships must be simple, robust, and accurate.

The objective of predicting K_{sat} through easily observable parameters is to develop models that can reliably estimate it without the need for complex, time-intensive, and costly laboratory analyses. By utilizing readily available soil variables such as texture (clay, silt, and sand content), bulk density, organic carbon content, pH, porosity, and others, there is an ambition to develop dependable Pedotransfer Functions (PTFs) or machine learning models that can estimate K_{sat} with high accuracy. This approach enhances the efficiency of soil management practices by providing faster, more cost-effective, and reliable predictions of water movement in the soil, which is essential for effective water management, irrigation strategies, environmental assessments, and land use planning. In recent years, there has been a notable emergence of significant and commendable studies that focus on predicting K_{sat} through the application of machine learning techniques.

In [6], Support Vector Machines (SVM) and Multiple Linear Regression (MLR) were employed to predict the K_{sat} of sandy soil. This study focused on utilizing easily measurable

soil properties, such as the clay/silt ratio, liquid limit, and chloride ion content, to estimate K_{sat}. This study compared SVM with various kernel functions (linear, radial basis, and sigmoid) and evaluated the model's performance using genetic algorithms for parameter optimization. The results showed that the SVM model with the RBF kernel outperformed the MLR models, highlighting the effectiveness of SVM in soil property predictions. This study showed that SVM serves as an effective tool for estimating K_{sat}, providing improved accuracy compared to MLR methods, particularly in intricate datasets such as sandy soils.

In [7], the use of Artificial Neural Networks (ANNs) was investigated to enhance the estimation of K_{sat} in smectitic soils, particularly through the integration of fractal parameters obtained from soil particle and micro-aggregate size distributions. PTFs were developed for the estimation of K_{sat}, utilizing fractal parameters alongside conventional soil data, including bulk density, clay content, and sand content. The results showed that the integration of fractal parameters with conventional soil data can markedly improve K_{sat} prediction. ANN ensemble models, especially those utilizing fractal parameters, provided more accurate and reliable predictions than traditional methods or single ANN models. In [8], machine-learning-based PTFs were used to predict K_{sat} across various soil types. A large database (USKSAT), consisting of more than 18,000 soil samples, was used to compare four ML algorithms, namely K-Nearest Neighbors (KNN), SVM, Random Forest (RF), and Boosted Regression Trees (BRT), with BRT achieving the best performance. In [9], two modeling techniques, namely ANNs and MLR, were compared in predicting soil hydraulic conductivity (K). The results showed that ANNs were a superior method for predicting K, particularly across diverse soil types, due to their ability to model complex relationships. Although MLR models are less adept at capturing non-linear dependencies, they remain useful in scenarios where data relationships are more straightforward.

In [10], K_{sat} was predicted using hybrid machine learning models. Various machine learning algorithms were used to improve prediction accuracy compared to conventional models. Five distinct models were created utilizing the Soil Water Infiltration Global (SWIG) database, which included various soil predictors: percentages of clay, silt, and sand, organic carbon content, bulk density, and water content. MLP, RF, and SVM were compared, along with hybrid models that integrated these methods. Hybrid machine learning models significantly outperformed individual algorithms, highlighting the advantages of integrating various models to enhance the predictions of K_{sat}. In [11], hydraulic conductivity was predicted from soil grain size distribution utilizing a Supervised Committee Machine Artificial Intelligence (SCMAI) model. This model integrates three distinct AI models, namely Larsen Fuzzy Logic (LFL), Least Squares SVM (LSSVM), and Wavelet-ANN (WANN), to enhance predictions of hydraulic conductivity, an essential parameter for evaluating water flow in porous media. The SCMAI model proved to be an effective tool for predicting hydraulic conductivity, delivering better results than traditional individual models by combining their outputs.

In [12], the K_{sat} of agricultural soil was examined by applying different PTFs. These functions estimated K_{sat} using easily measurable soil properties, including particle size distribution, bulk density, and organic matter content. This study evaluated 10 PTF models, including machine learning algorithms such as RF and BRTs, in addition to neural network-based models such as Rosetta-SSC. Data were collected from a single agricultural field exhibiting various tillage treatments and differing spatial and temporal variability in K_{sat}. The results showed that machine learning-based PTF models, while providing estimates, generally underestimated the measured K_{sat} values, indicating a need for further refinement to enhance prediction accuracy. In [13], a new model was proposed for predicting K_{sat}, emphasizing the impact of particle size. This model extended the established Kozeny-Carman (KC) model by integrating the notion of equivalent particle size and addressing the influence of adsorbed water films on fine-grained soils. This approach sought to address the limitations of the KC model, which exhibits suboptimal performance for fine-grained soils due to the complexities associated with adsorbed water and the pore structure in these materials. The results showed that the proposed model was a more accurate and useful way to estimate K_{sat}, especially in fine-grained soils, because it takes into account the effects of water that has been absorbed and the structure of the pores.

In [14], ML models, including Radial Basis Function Neural Networks (RBFNNs) and MLPNNs, and hybrid models combining Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) were combined with Neural Networks (GA-NN and PSO-NN) for predicting K_{sat}. This study aimed to evaluate the predictive performance of these models using data from a field study at the Bajgah Agricultural Experimental Station in Iran, which measured soil properties such as bulk density, water content, and aggregate size. The MLR model performed poorly, showing low correlation and high error values. On the contrary, the PSO-NN model achieved the best accuracy, with a high correlation coefficient (0.958). This model outperformed RBFNNs, MLPNNs, and GA-NNs, which also showed promising results but with slightly higher errors. In [15], three machine learning models, namely RF, SVM, and Least-Squares SVM (LSSVM), were used to study K_{sat} prediction. K_{sat} was predicted using a range of soil characteristics, such as bulk density, porosity, pH, texture, salinity, and sodium adsorption ratio. Performance indicators showed that RF was the best model for predicting K_{sat}, surpassing both SVM and LSSVM models.

This study aimed to evaluate the effectiveness of various machine learning techniques, including MLP, ANN, SVM, and Tree Boost, in developing a more accurate and reliable model for predicting field saturated hydraulic conductivity (K_{field}). The findings of this study provide valuable insights for managing water resources, particularly in the context of groundwater.

II. MATERIALS AND METHODS

A. Study Area and Data

Soil samples were obtained from two locations in the El-Nubaria region (Sugar Beet Areas 1 and 2), located in the western delta of Egypt. The soil texture of Sugar Beet Area 1 ranges from sandy loam to loamy sand, but in Sugar Beet Area 2, it varies from sandy loam to sandy clay loam up to 2.0 meters below the surface. Soils in this region comprise a specific amount of fine gravel and gypsum at depths ranging from 0.40 to 0.60 m and are classified as calcareous soils. The second region is in Sinai, characterized by sandy and unstable soil. The chosen regions exhibit sandy soils that possess varying physical and chemical characteristics. 57 soil samples were collected from the El-Nubaria region and 28 soil samples from the Sinai region. Soil samples that showed disturbance were collected from designated areas and locations. Hydraulic conductivity was assessed at each site using the auger-hole method.

The soil samples were analyzed in the laboratory to determine their physical and chemical properties. The soil samples' physical properties encompassed silt, clay, and sand contents, d90 of the grains, Liquid Limit (LL), and Plastic Limit (PL). The chemical properties of the soil samples encompassed soil pH, and soil Saturation Percentage (SP), as well as the concentrations of chloride ions, bicarbonate anions (HCO_3), and calcium carbonate ($CaCO_3$). A permeameter apparatus was employed to measure the laboratory hydraulic conductivity (K_{lab}).

The soil's clay content ranged from 0.00% to 37.4%, while the K_{field} values varied from 0.07 m/day to 6.06 m/day. Soil samples from the Sinai region exhibited a higher salt concentration compared to those obtained from the El-Nubaria region. Nobaria samples were found to have a clay content varying between 0.10% and 27.00%, while Sinai samples were found to have a clay content varying between 0.00% and 37.4%. The d90 of the samples varied between 0.12 and 4.37 mm for the Nobaria samples and between 0.12 and 0.34 mm for the Sinai samples. The field hydraulic conductivity measured at sample locations in the Nobaria area varied between 0.07 m/day and 6.06 m/day, while it varied between 0.35 m/day and 1.39 m/day at the locations of Sinai samples. Hydraulic conductivity was determined in the laboratory using the permeameter setup and disturbed soil samples were collected from the two areas. Table I and Figure 1 present summaries of the laboratory analysis results for soil samples.

Correlation analysis was performed to enhance understanding of the relationships between these variables and their effects on one another. Correlation values ranged between -1 and 1, where the negative limit (-1) signifies a strong negative correlation, the positive limit (+1) signifies a strong positive correlation, and the mid-point (0) implies no correlation. Table II displays the correlation matrix for the variables. The correlation matrix reveals several noteworthy relationships between the variables. HCO_3 shows strong negative correlations with $CaCO_3$ (-0.59), PL (-0.51). $CaCO_3$ demonstrates moderate to strong positive correlations with LL(0.82), PL (0.81), and Soil Sp (0.72).

TABLE I. STATISTICAL ANALYSIS OF VARIOUS RANGES OF PHYSICAL AND CHEMICAL PROPERTIES.

	Mean	**Std Dev**	**Std Err Mean**	**Minimum**	**Maximum**	**Median**	**Mode**
K_{lab} (M/d)	19.629	21.337	3.254	0.070	93.410	11.490	7.610
Clay/Silt	1.346	1.250	0.191	0.000	5.100	0.940	0.000
d90	0.993	1.362	0.208	0.120	4.370	0.330	0.330
LL	15.695	14.485	2.209	0.000	50.600	20.300	0.000
PL	9.085	8.791	1.341	0.000	21.340	12.460	0.000
pH	7.502	0.222	0.034	7.100	8.100	7.500	7.500
HCO_3	4.577	1.251	0.191	2.310	7.000	4.190	4.380
Cl	320.551	583.511	88.985	1.700	2058.200	6.600	2.100
Soil Sp (%)	37.535	14.043	2.142	18.000	74.000	40.000	20.000
$CaCO_3$	20.935	18.246	2.783	1.570	54.890	19.720	3.140
K_{field}	1.020	1.202	0.183	0.070	6.060	0.650	0.600

TABLE II. CORRELATION MATRIX OF THE FEATURE VARIABLES.

	K_{lab}	**Clay/Silt**	**d90**	**LL**	**PL**	**pH**	**HCO_3**	**Cl**	**Soil Sp**	**$CaCO_3$**	**K_{field}**
K_{lab}	1.00	0.02	-0.20	-0.05	-0.09	0.08	0.00	0.13	0.26	-0.03	0.18
Clay/Silt	0.02	1.00	-0.20	-0.19	-0.18	-0.20	0.21	0.24	-0.32	-0.37	-0.11
d90	-0.20	-0.20	1.00	0.33	0.41	-0.05	-0.25	-0.31	0.28	0.46	-0.14
LL	-0.05	-0.19	0.33	1.00	0.89	0.35	-0.49	-0.33	0.68	0.82	0.08
PL	-0.09	-0.18	0.41	0.89	1.00	0.26	-0.51	-0.34	0.58	0.81	0.23
pH	0.08	-0.20	-0.05	0.35	0.26	1.00	-0.37	-0.23	0.51	0.38	0.30
HCO_3	0.00	0.21	-0.25	-0.49	-0.51	-0.37	1.00	0.66	-0.37	-0.59	-0.26
Cl	0.13	0.24	-0.31	-0.33	-0.34	-0.23	0.66	1.00	-0.10	-0.52	-0.01
Soil Sp	0.26	-0.32	0.28	0.68	0.58	0.51	-0.37	-0.10	1.00	0.72	0.26
$CaCO_3$	-0.03	-0.37	0.46	0.82	0.81	0.38	-0.59	-0.52	0.72	1.00	0.22
K_{field}	0.18	-0.11	-0.14	0.08	0.23	0.30	-0.26	-0.01	0.26	0.22	1.00

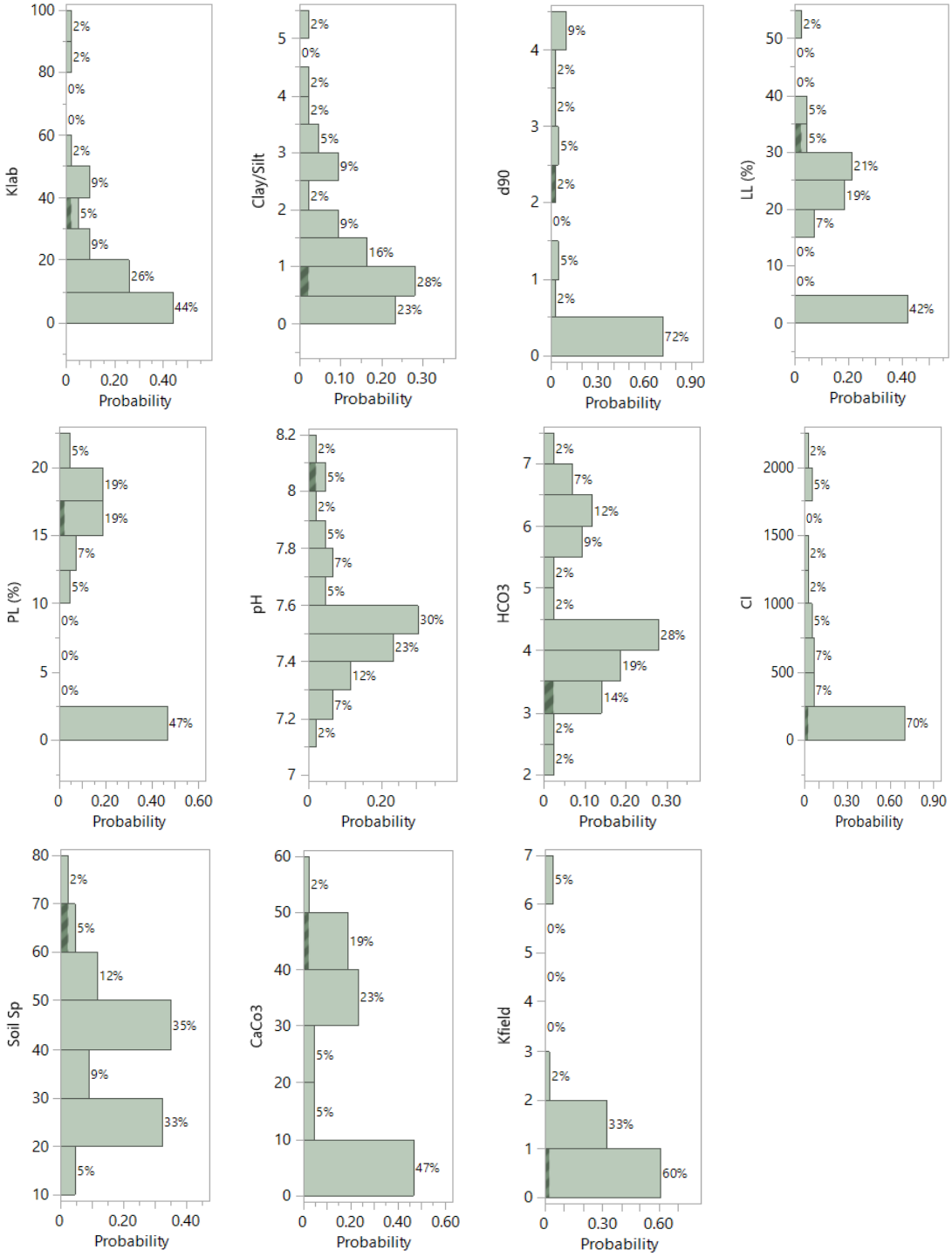

Fig. 1. Physical and chemical properties spectrum.

B. Methods

ML techniques were used to find precise approximations that provide a robust, computationally efficient, and cost-effective solution while reducing computational time [16].

1) Multilayer Perceptron Neural Network (MLPNN)

The MPNN is a frequently used model that can generate a known output using historical data, involving three fully interconnected layers, an input layer, at least one hidden layer, and an output layer. The MPNN connects all layers using a

straightforward relationship, which can be expressed as follows:

$$y_i = f\sum_{i=1}^{n}|xw + b|_i \quad (1)$$

where w denotes the weights, b denotes the biases, f denotes the operating function, x denotes the i^{th} input of an ANN, y denotes the j^{th} output of an ANN, and n denotes the number of inputs. This study used the conjugate gradient method to train the MPNN. To adjust the parameters, forward propagation (calculating loss) and backpropagation (calculating derivatives) were leveraged. Moreover, transfer functions, namely, sigmoid and linear functions, were used as activation functions for the hidden and output layers [17].

2) Support Vector Machine (SVM)

SVM was proposed to solve pattern recognition problems. With the introduction of the ε-insensitive loss function, the SVM method has been expanded to estimate nonlinear regression [18]. The basic concept of SVM regression is to nonlinearly map the original data x into a higher-dimensional feature space and to solve a linear regression problem in this feature space. The SVM regression function is:

$$f(x) = [w.\phi(x)] + b \quad (2)$$

where $\phi(x)$ is the nonlinear mapping function, w is the weight vector and b is the bias term. The values for w and b are estimated by minimizing the regularized risk function. The regression function is [18]:

$$f(x) = \sum_{i=1}^{l}(\delta_i - \delta_i^*)K(x_i, x) + b \quad (3)$$

where $K(x_i, x)$ is the kernel function,δ_i and δ_i^* are Lagrangian multipliers, and the data points corresponding to $\delta_i - \delta_i^* \neq 0$ are the support vectors. The radial basis kernel function (RBF) $K(x_i, x) = e^{(-\|x_i - x\|^2/2\partial^2)}$ was used, where δ^2 is the kernel parameter of the radial basis function kernel.

3) Tree Boost Method

The Tree Boost approach is a member of the boosted regression tree family of theoretical modeling techniques. Through the integration of a boosting technique, Tree Boost enhances the decision tree algorithm. Instead of constructing a single optimal model, the fundamental concept here is to merge a collection of weak models to produce a robust consensus model. The Tree Boost algorithm generates new decision trees sequentially by decreasing the residuals of the original trees. The procedure of constructing this sequential model is a form of functional gradient descent. This approach is capable of including both qualitative and quantitative factors in the regression analysis. Moreover, correlated predictive variables and missing data can be handled using it. Furthermore, it is considered to be resilient to the existence of outliers in the dataset, as well as to the use of irrelevant predictor variables. The Tree Boost model generally has three primary parameters: the learning rate (also known as the shrinkage parameter), the complexity of the tree, and the number of regression trees (tree size). Cross-validation techniques are employed to evaluate the generalization of the Tree Boost model and mitigate overfitting [19, 20].

Both the training subset, which was used to identify the optimal model parameters, and the validation subset were subsets of the training data. After the models were trained, they were subjected to a testing procedure to determine their efficacy in generalizing the acquired knowledge to previously unexplored cases. Approximately 70% of the entire dataset was randomly selected for model training, whereas the remaining 30% was used for model testing [21].

III. RESULTS AND DISCUSSION

The careful consideration of input and output variables is a fundamental part of the development of an ML model. A subset of 10 parameters (K_{lab}, Clay/Silt ratio, d90, LL, PL, pH, HCO_3, Cl, Soil Sp (%), and $CaCO_3$) was chosen to develop ML-based models to predict the K_{field} of sandy soil based on basic soil properties of saline and alkaline soil data. The ML-based models were evaluated using Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), correlation coefficient (R), and Scatter Index (SI). These statistical indicators are provided as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2 \quad (4)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|P_i - O_i| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2} \quad (6)$$

$$SI = \frac{RMSE}{\bar{O}} \quad (7)$$

$$R = \frac{\sum_{i=1}^{N}\left(P_i - \bar{P}\right)(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{N}\left(P_i - \bar{P}\right)^2 \sum_{i=1}^{N}(O_i - \bar{O})^2}} \quad (8)$$

where O_i, P_i, N, $\bar{O}$, and $\bar{P}$ denote the observed value, predicted value, total number of data points, observations mean value, and predictions mean value, respectively.

Using the MLP model, the results showed that MSE was 0.103, MAE was 0.226, RMSE was 0.321, SI was 0.226, and R was 0.974 for the predicted field saturated soil hydraulic conductivity (K_{field}) of sandy soil values. Figure 2 presents the scatter plot of the field-saturated soil hydraulic conductivity values measured and predicted by the MLP. When utilizing the SVM approach, the results were obtained using a 10-fold cross-validation based on the average results obtained for test data (10 folds). The cross-validation method was employed to train and test the model to ensure that all dataset instances were applied in both stages. The SVM model resulted in an MSE of 0.074, an MAE of 0.184, an RMSE of 0.272, and an SI of 0.192, with an R-value of 0.981. The statistical results for the Tree Boost were MSE of 0.014, MAE of 0.074, RMSE of 0.120, SI of 0.085, and R of 0.99 for the predicted K_{field} of the sandy soil.

Figure 3 shows the relationships between the actual and predicted K_{field} using various ML models, where a continuous diagonal line serves as the reference.

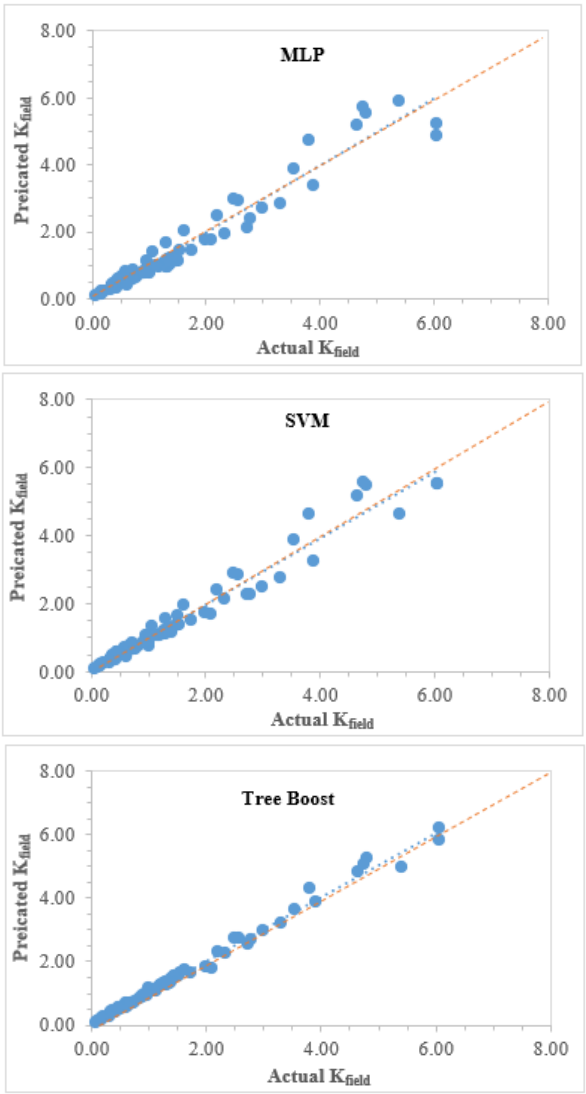

Fig. 2. Scatter plot of the measured and predicted field saturated soil hydraulic conductivity of sandy soil values for ML models.

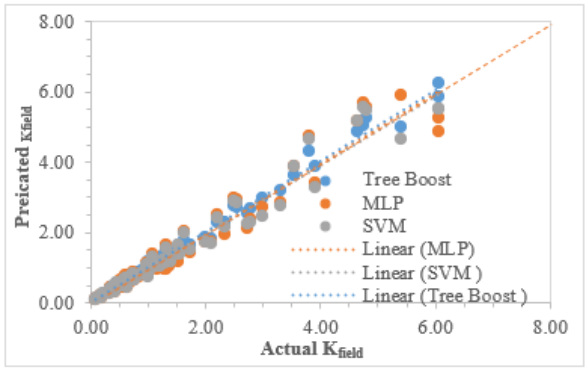

Fig. 3. Scatter plot of the measured and predicted field saturated soil hydraulic conductivity of sandy soil values for various ML models.

The actual K_{field} at any point along this reference line is identical to that predicted by the different models for a certain observation. If the predicted K_{field} values fall below this reference line, the model can be regarded to be conservative. However, if the predicted values lie above the reference line, the model can be considered to have overestimated K_{field}. To enhance the clarity of the results, a trend line can be drawn for comparison with the reference diagonal line. Generally, a model will be more accurate or more effective at predicting if the discrete points lie closer to the trend line. Here, the trend line for the Tree Boost model is proximate to the reference line, the trend lines for the MLP and SVM models being the second and third closest to the reference line, respectively.

The Taylor diagram provides a graphical representation of the suitability of various ANN and SVM models based on R^2, RMSE, and standard deviation. A Taylor diagram is a two-dimensional prediction space, and actual values are positioned according to their degree of coordination. The standard deviation, RMSE, and R^2 scores are each represented by the horizontal and vertical axes, circular lines, and radial lines, respectively. The accuracy of a model is measured by the proximity of each model to the actual value. The closer a prediction model is to the actual data, the more reliable it is. The GRNN model, which has a higher R^2 score and standard deviation and a lower RMSE than the other models, is closer to the actual values and, as shown in Figure 4, is marginally more reliable.

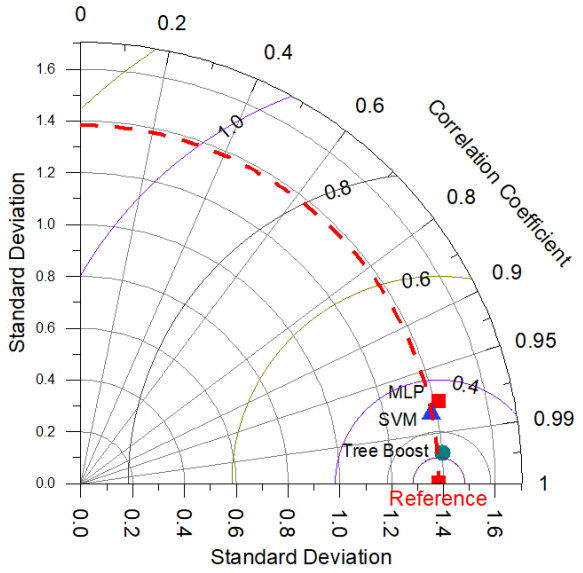

Fig. 4. Taylor diagram visualization of ML models performance for field saturated soil hydraulic conductivity prediction.

Compared to the other models, the Tree Boost model produced SI values that were 165.88%, and 125.87% lower than those of the MPNN, and SVM, respectively, as shown in Figure 5. The Tree Boost model exhibited superior prediction performance in terms of MSE, MAE, RMSE, SI, and R values, followed by SVM. In terms of predicting the K_{field} of sandy soil, the results demonstrate the Tree Boost model's performance to be superior to that of the other models. Additionally, the results indicate that the Tree Boost model

substantially reduces the overall error and accurately predicts the K_{field} of sandy soil.

To validate the performance of the Tree Boost model, it was compared with the SVM model with an RBF as proposed in [6]. The results of the RMSE and R values for the two models demonstrate the superior performance of the Tree Boost model, as shown in Figure 6.

Fig. 5. Comparison of MLP, SVM, and Tree Boost models.

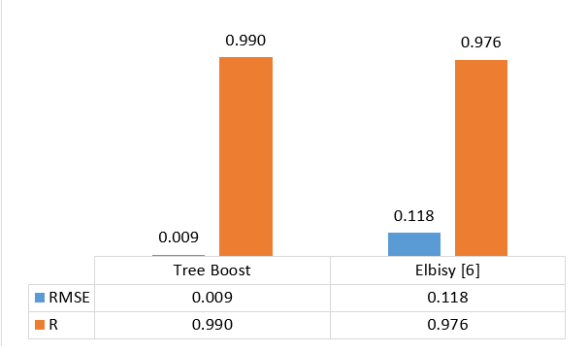

Fig. 6. Comparison of the Tree Boost model and the SVM-RBF model in [6].

IV. CONCLUSIONS

Accurate prediction of the hydraulic conductivity of saturated soil is crucial to addressing groundwater-related issues. When reliable data are available, machine learning techniques can be highly effective in generating precise predictions. This study developed three different models to predict the K_{field} of sandy soil, using basic soil properties that can be easily measured in the laboratory. Soil samples were collected from two locations, El-Nubaria and Sinai regions, located in the western delta of Egypt. The machine learning techniques selected for this study were MPNN, SVM, and Tree Boost. The accuracy of the predictions was assessed using five evaluation metrics, namely MSE, MAE, RMSE, SI, and R, along with a Taylor diagram.

The input features for the models included K_{lab}, clay/silt ratio, d90, LL, Pl, pH, HCo_3, Cl, Soil Sp (%), and $CaCo_3$. Among the models tested, the Tree Boost model demonstrated outstanding accuracy in estimating K_{field}. The SI of the Tree Boost model (0.085) was significantly lower than that of the SVM (0.192) and MLP (0.226) models. Furthermore, the R of the Tree Boost (0.99) exceeded that of the SVM (0.981) and MLP (0.974) models. Compared to the SVM model with an RBF in [6], the Tree Boost model significantly reduced the prediction error.

The Tree Boost model notably reduced the overall prediction error and accurately estimated the K_{field} of sandy soil based on easily measurable basic soil properties. These findings highlight the superior precision and reliability of the Tree Boost model compared to other algorithms. The results suggest that the Tree Boost model is a promising tool for estimating K_{field} with high accuracy.

REFERENCES

[1] V. Pot, X. Zhong, and P. C. Baveye, "Effect of resolution, reconstruction settings, and segmentation methods on the numerical calculation of saturated soil hydraulic conductivity from 3D computed tomography images," *Geoderma*, vol. 362, Mar. 2020, Art. no. 114089, https://doi.org/10.1016/j.geoderma.2019.114089.

[2] S. Rousseva *et al.*, "Chapter Two - Soil Water Characteristics of European SoilTrEC Critical Zone Observatories," in *Advances in Agronomy*, vol. 142, S. A. Banwart and D. L. Sparks, Eds. Academic Press, 2017, pp. 29–72.

[3] S. Jena, B. P. Mohanty, R. K. Panda, and M. Ramadas, "Toward Developing a Generalizable Pedotransfer Function for Saturated Hydraulic Conductivity Using Transfer Learning and Predictor Selector Algorithm," *Water Resources Research*, vol. 57, no. 7, 2021, Art. no. e2020WR028862, https://doi.org/10.1029/2020WR028862.

[4] J. W. Crawford, "The relationship between structure and the hydraulic conductivity of soil," *European Journal of Soil Science*, vol. 45, no. 4, pp. 493–502, 1994, https://doi.org/10.1111/j.1365-2389.1994.tb00535.x.

[5] F. Di Nunno and F. Granata, "Groundwater level prediction in Apulia region (Southern Italy) using NARX neural network," *Environmental Research*, vol. 190, Nov. 2020, Art. no. 110062, https://doi.org/10.1016/j.envres.2020.110062.

[6] M. S. Elbisy, "Support Vector Machine and regression analysis to predict the field hydraulic conductivity of sandy soil," *KSCE Journal of Civil Engineering*, vol. 19, no. 7, pp. 2307–2316, Nov. 2015, https://doi.org/10.1007/s12205-015-0210-x.

[7] A. Sedaghat, H. Bayat, and A. A. Safari Sinegani, "Estimation of soil saturated hydraulic conductivity by artificial neural networks ensemble in smectitic soils," *Eurasian Soil Science*, vol. 49, no. 3, pp. 347–357, Mar. 2016, https://doi.org/10.1134/S106422931603008X.

[8] S. N. Araya and T. A. Ghezzehei, "Using Machine Learning for Prediction of Saturated Hydraulic Conductivity and Its Sensitivity to Soil Structural Perturbations," *Water Resources Research*, vol. 55, no. 7, pp. 5715–5737, 2019, https://doi.org/10.1029/2018WR024357.

[9] C. G. Williams and O. O. Ojuri, "Predictive modelling of soils' hydraulic conductivity using artificial neural network and multiple linear regression," *SN Applied Sciences*, vol. 3, no. 2, Jan. 2021, Art. no. 152, https://doi.org/10.1007/s42452-020-03974-7.

[10] F. Granata, F. Di Nunno, and G. Modoni, "Hybrid Machine Learning Models for Soil Saturated Conductivity Prediction," *Water*, vol. 14, no. 11, Jan. 2022, Art. no. 1729, https://doi.org/10.3390/w14111729.

[11] M. Khalili-Maleki, R. V. Poursorkhabi, A. A. Nadiri, and R. Dabiri, "Prediction of hydraulic conductivity based on the soil grain size using supervised committee machine artificial intelligence," *Earth Science Informatics*, vol. 15, no. 4, pp. 2571–2583, Dec. 2022, https://doi.org/10.1007/s12145-022-00848-x.

[12] K. Báťková *et al.*, "Prediction of saturated hydraulic conductivity Ks of agricultural soil using pedotransfer functions," *Soil and Water Research*, vol. 18, no. 1, pp. 25–32, 2023, https://doi.org/10.17221/130/2022-SWR.

[13] J. Liu and X. Ren, "Hydraulic conductivity and particle size of soils: model and prediction." Research Square, Oct. 20, 2023, https://doi.org/10.21203/rs.3.rs-3447393/v1.

[14] A. A. Moosavi, M. A. Nematollahi, and M. Omidifard, "Comparing machine learning approaches for estimating soil saturated hydraulic

conductivity," *PLOS ONE*, vol. 19, no. 11, 2024, Art. no. e0310622, https://doi.org/10.1371/journal.pone.0310622.

[15] M. Farasati, M. Seyedian, and A. Fathaabadi, "Predicting soil hydraulic conductivity using random forest, SVM, and LSSVM models," *Natural Resource Modeling*, vol. 37, no. 4, 2024, Art. no. e12407, https://doi.org/10.1111/nrm.12407.

[16] M. A. E. Zareef, M. Ghalla, J. W. Hu, and A. M. Elbisy, "Machine learning approaches for estimating concrete shear strength in FRPreinforced members without shear reinforcement," *Steel and Composite Structures*, vol. 53, no. 3, pp. 327–341, Nov. 2024.

[17] M. S. Elbisy, "Machine learning techniques for estimating wave-overtopping discharges at coastal structures," *Ocean Engineering*, vol. 273, Apr. 2023, Art. no. 113972, https://doi.org/10.1016/j.oceaneng.2023.113972.

[18] V. Vapnik, S. Golowich, and A. Smola, "Support Vector Method for Function Approximation, Regression Estimation and Signal Processing," in *Advances in Neural Information Processing Systems*, 1996, vol. 9.

[19] J. H. Friedman and J. J. Meulman, "Multiple additive regression trees with application in epidemiology," *Statistics in Medicine*, vol. 22, no. 9, pp. 1365–1381, 2003, https://doi.org/10.1002/sim.1501.

[20] M. S. Elbisy, "Estimation of Wave Overtopping Discharges at Coastal Structures with Combined Slopes using Machine Learning Techniques," *Engineering, Technology & Applied Science Research*, vol. 14, no. 3, pp. 14033–14038, Jun. 2024, https://doi.org/10.48084/etasr.7175.

[21] A. Nassar and A. Elbisy, "Evaluation of Factors Leading to Time Delays and Cost Overruns in Marine Construction Projects," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 16095–16102, Oct. 2024, https://doi.org/10.48084/etasr.8116.