Real-time modeling of soil water properties in maize and soybean fields through the above canopy sensor and weather data fusion and machine learning

Nipuna Chamara[1], Geng Bai[1], Yufeng Ge[1 2]

[1]Department of Biological Systems Engineering, University of Nebraska-Lincoln, Lincoln, NE 68583, USA

[2]Center for Plant Science Innovation, University of Nebraska-Lincoln, Lincoln, NE 68588, USA

A paper from the Proceedings of the
16th International Conference on Precision Agriculture
21-24 July 2024
Manhattan, Kansas, United States

Abstract.
Real-time soil water content (SWC), soil water deficit (SWD), and available water content (AWC) estimation are vital for scheduling irrigation to prevent overwatering or underwatering. Various in-situ techniques, such as Time-domain reflectometry (TDR), frequency-domain reflectometry (FDR), capacitance, and resistive sensors, directly estimate SWC. While these methods offer high-frequency and accurate measurements, they are expensive and less installation-friendly in commercial agricultural fields, resulting in lower spatial resolution data. On the other hand, remote sensing-based real-time SWC estimation provides broader spatial coverage. In contrast, proximal remote sensing has emerged as a promising approach for real-time SWC estimation, gaining attention within irrigation communities. This research addresses the gap in estimating soil water properties using proximal remote sensing. The approach integrates proximal remote sensors positioned above ground with relevant weather data through artificial intelligence (AI) models. This integration is crucial for effective row crop irrigation management. This study used stationary sensor stations to collect proximal remote sensing data. Each of the three stationary proximal remote sensing units has an infrared radiometer, NDVI (Normalized Difference Vegetation Index) sensor, and PRI (Photochemical Reflectance Index) sensor. Variations of machine learning model types Gaussian process regression (GPR), linear regression (LR), neural network (NN), support vector machines (SVM), stepwise linear regression (SLR), Tree, kernel, and ensemble model types were evaluated. The GPR and neural NN model types demonstrated R-squared values surpassing 0.95 in accurately estimating SWC, SWD, and RWC. Adopting this solution with center pivot irrigation systems will improve irrigation water use efficiency.

Keywords. Soil water content, Soil water deficit, Total available water, Machine learning

Introduction

Agriculture accounts for 70% of all freshwater extraction and usage, while the rest is used for municipal and industrial purposes (FAO, 2021). Due to global phenomena such as climate change, reduction in arable lands, groundwater depletion, and population growth, improving water use efficiency in crop production is an essential target for the agriculture industry (Baztan et al., 2022). Irrigation is an indispensable agricultural management activity in regions where crop water demands exceed rainfall, directly affecting water use efficiency (Hatfield & Dold, 2019). Irrigation scheduling controls how much and how frequently water is applied (Eisenhauer et al., 2021). In arid and semi-arid regions, measuring or estimating SWC, SWD, and RWC have played essential roles in irrigation control or scheduling, increasing water use efficiency (Katimbo et al., 2023).

The methods of irrigation scheduling can be grouped into four categories: (1) evapotranspiration and soil water balance (ET-WB) or the simplified checkbook method, (2) soil water measurement method, (3) plant status indicators, and (4) process-based models (Eisenhauer et al., 2021; Gu et al., 2021; Katimbo et al., 2023). Time-domain reflectometry (TDR), frequency-domain reflectometry (FDR), capacitance, and resistive sensors are commercially available and allow direct, in-situ measurements of soil water content, leading to closed-loop irrigation control (Chamara et al., 2012). These sensors are popular among small-scale farms and gardening. Although they can provide frequent measurements with high accuracy, the labor and logistics required for their installation make them less favorable in large commercial fields. Typically, these sensors have below-ground sensing units and above-ground data logging and transmission components. Installation of the below-ground parts and retrieval before field preparation for the next season is time-consuming and expensive. Keeping the data logger or transmitter above the ground throughout the growing season poses practical challenges due to moving agricultural machinery for different operations (Chamara et al., 2022). Additionally, large commercial fields are often heterogeneous and require a high density of sensors to sufficiently capture the spatial variability of SWC, which makes them less economically viable.

Remote sensing-based estimation of soil water indicators can provide higher spatial coverage. In contrast, accuracy and temporal resolution are limited. Sensors operated at microwave (passive and active), thermal infrared, and optical ranges have been used to estimate SWC (Wang et al., 2023). Remote sensing can be separated into two main categories: distal and proximal. Cloud coverage negatively affects decisions in distal remote sensing from satellites or airborne platforms. Proximal remote sensing, on the other hand, uses sensors on fixed structures, pivot irrigation systems, ground robots, or unmanned aerial vehicles (UAVs) and has shown great potential in estimating SWC and irrigation decision-making (Mendes et al., 2023).

Volumetric SWC is one of the most critical parameters required for irrigation scheduling, along with the soil's volumetric water content at field capacity (θ_{fc}) and wilting point (θwp), crop type, and effective root depth (R_d), type of irrigation system, irrigation system efficiency (E_a), precipitation (P), evapotranspiration (ET), application uniformity (DU), conveyance efficiency (E_c). For instance, a center pivot irrigation system can determine the appropriate volume of irrigation water needed at a specific location by considering the Soil Water Depletion (SWD) value. This helps prevent under and over-water application (Eisenhauer et al., 2021). While under-application of water leads to crop water stress and yield reduction, over-application causes problems, including higher energy use for pumping water (Lo et al., 2015), excessive groundwater extraction, and increased deep percolation and nutrient loss through leaching (Abeygunawardena et al., 2023).

(Ehrler, 1971) established an inverse relationship between leaf-air temperature difference (ΔT =

$T_c - T_a$) and vapor pressure deficit (VPD), then a relationship associated with root zone water depletion and stress degree day index (SDD) was introduced by (Jackson et al., 1977). More recently (Singh et al., 2021) showed that a correlation existed between soil water depletion (D_{rw}) and canopy temperature differential (ΔT) when D_{rw} exceeds 170 mm for corn and 160 mm for soybean. However, they could not identify such a relationship when D_{rw} was lower. In these studies, a primary input was crop canopy temperature (T_c) measured by an infrared radiometer and crop management parameters such as cumulative growing degree days (CGD).

Based on the literature, we conclude that a single variable can not accurately explain SWC, SWD, and RWC. Still, multiple variables are required to represent soil, weather, farm management, and crop dynamic properties. Furthermore, developing empirical models including all these parameters is time-consuming and tedious (Adeyemi et al., 2018). However, AI-based models for SWC, SWD, and RWC modeling can identify associated relationships without much intervention from a human modeler. Several notable research outcomes have emerged in the context of soil moisture prediction for irrigation decisions. These include (1) a fuzzy neural network to determine the irrigation time and depth with real-time air temperature (T_a), relative humidity (RH), solar radiation (G), and wind speed (U) (Tsang & Jim, 2016), (2) dynamic neural networks for one day ahead volumetric SWC prediction based on past volumetric (SWC), precipitation (P), irrigation (I), climatic data (not explained the parameters), (3) a neural network for root zone SWC prediction via climatic data, rooting depth, hesternal soil moisture (Gu et al., 2021). These data-driven models contribute to better SWC predictions, aiding in informed irrigation practices.

However, as per the authors understanding, no data-driven AI model was developed to estimate SWC, SWD, and RWC in multiple subsoil layers, including the real-time weather data, farm ma nage ment parameters, and crop parameters excluding past SWC. Therefore, in this research, the authors tried to develop a method to explain the complex SWC, SWD, and RWC dynamics via multiple AI models using real-time weather, farm management, and crop parameters. Additionally, we utilize the practically reproducible parameters in a row crop production environment in this model development.

Therefore, the main objective of this research was to identify machine learning models that can estimate real-time total SWC (θ_v), total SWD, and /or total RWC in real-time with above-ground sensors. The secondary objective was to check the feasibility of estimating the real-time SWC, SWD, and RWC at 15cm, 45 cm, and 75 cm depths.

Materials And Methods

Study Area, Experimental Design, Data Collection, and Data Pre-processing

The data for this research were collected in the 2022 and 2023 growing seasons at the field phenotyping facility at the University of Nebraska-Lincoln, Eastern Nebraska Research, Extension and Education Center (ENREEC) near Mead, Nebraska, USA (41°08'42.6"N 96°26'24.1"W, 351 m above sea level) (Bai et al., 2019). Corn was a variety of DeKalb DKC51-40 RIB with a density of 30,000 plants per acre, and soybeans were Pioneer P26T23E with a density of 150,000 plants per acre. Each sensor station had three Teros 10 soil moisture sensors, an infrared radiometer, and NDVI and PRI SRS sensors. **Figure 1** shows the research plot arrangement for the 2022 growing season. We assigned two sensor nodes for the soybean and eight nodes for corn, similar to an experiment visualized in the research (Chamara, 2021) in 2022. Irrigation data are in **Table 1**. Both corn and soybean assigned groups were divided into rain-fed and irrigated treatments (**Figure 1**).

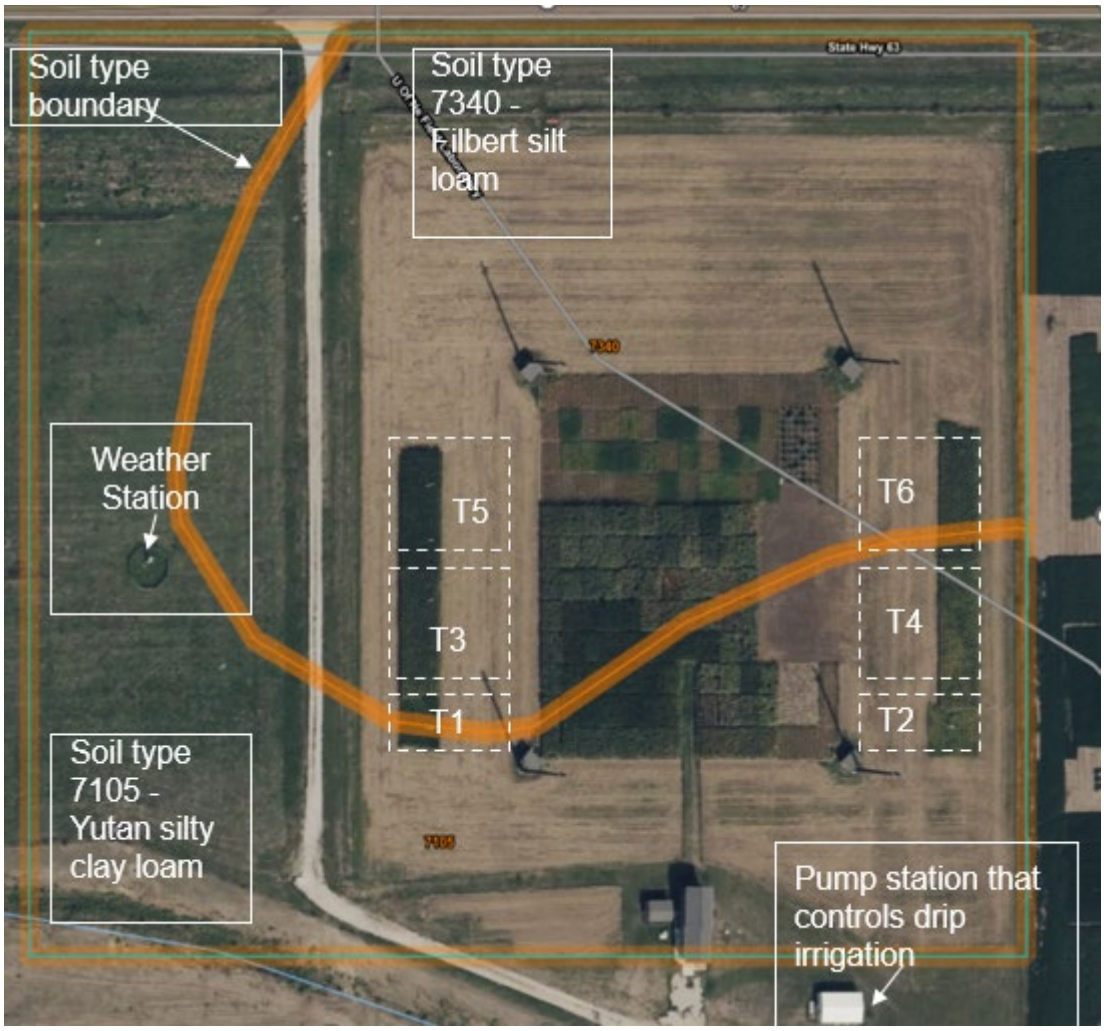


Figure 1 Experiment location, design, and soil properties [Soil map source:- (Soil Survey Staff, 2024)] T1- irrigated soybean. (T2- rainfed soybean, T3- corn with irrigated deficit nitrate applied, T4- corn with rainfed deficit nitrate applied, T5- corn with irrigated deficit nitrate applied, T6- corn with rainfed excess nitrate applied.) (Sensor station allocation – station 1 in T1, station 2 & 8 in T3, station 3 in T2, station 4 & 5 in T4, station 9 & 10 in T5, station 6 & 7 in T6)

Furthermore, there were plots with and without nitrate application within the corn category, regardless of irrigation status. However, due to the limited availability of NDVI, PRI sensors, and logistic issues, only data from three stations (Node 1 in T1, Node 2 in T2, and Node 3 in T3), consisting of irrigated soybean, rainfed soybean, and irrigated corn, in 2022, and data from 1 sensor station (Node 3 in T2) in 2023 were selected for this study. In 2023, only soybeans were grown in the research plots, and irrigated plots could not be maintained due to the breakdown of the irrigation system. Therefore, we considered the data collected in 2023 to be T2- rainfed soybean treatment.

Table 1. 2022 Irrigation data with irrigation period and amount

7/5 – 7/7	7/14 – 7/15	7/18 – 7/20	7/27 – 7/29	8/1 – 8/4	8/24 – 8/26	8/31 – 9/1	9/7	Season Total
1.162	1.008	0.912	1.115	1.498	0.834	1.260	0.864	8.65

Weather data (**Table** 2) were collected from a weather station approximately 30 m west of the field phenotyping facility (**Figure 1**) (Bai et al., 2019). Air temperature (T_a), relative humidity (RH), and incoming solar radiation (G) sensors collected data every minute at 2 m above the ground. Soil temperature (T_s) was collected below 10 cm, wind direction (U_d) wind speed (U) at 3 m height, atmospheric pressure (P_a) at 5 m height, and precipitation (P) at 1 m. These independent variables (**Table** 2) were selected to train the models that affect the dynamic variations of SWC, SWD, and RWC. The primary assumption in this experiment is that except for the parameters involved in the model training, others have a negligible effect on dynamic changes in volumetric water content. The windchill (WC) parameter depends on the air temperature and wind speed. The heat index (HI) parameter depends on air temperature and relative humidity, soil temperature of 10 cm under the soil near the weather station, and the dew point (DP) depends on the RH and T_a. Therefore, the derived parameters in the **Table 2**, such as HI, WC, and DP, were removed after analyzing the feature importance scores sorted using the RReliefF algorithm at early-stage model training.

A Campbell CR300 data logger powered by a solar panel collected the solar irradiance and canopy reflective radiance. Four two-band radiometers were used to measure the solar irradiance and canopy radiance. METERGroup, Inc.USA manufactured these radiometers. NDVI-hemispherical (N_i), NDVI-field stop (N_r), PRI-hemispherical (P_i), and PRI-field stop (Pr) are the four types of radiometers. The NDVI-hemispherical sensor can measure incoming solar radiation bands in red (650 nm) and NIR (810 nm) wavelengths, while the PRI-hemispherical measures blue (531 nm) and green (570 nm). NDVI-field-stop and PRI-field stop measures correspond to reflectance from the canopy. The sensor stations had to move out of the cropping area during planting and chemigation activities. The above-ground sensors were removed from the field to facilitate weedicide and pesticide application.

Table 1 Data set variables and definitions.

Data type	Parameter	Description (Averaged – 1 minute)	Measurement unit
Weather Parameters	Ta	Averaged Temperature	°C
	RH	Averaged Relative Humidity	%
	PC	Cumulative Precipitation	Inches
	DP	Averaged Dew Point	°C
	U	Averaged Wind Speed	m s−1
	G	Averaged Incoming Solar Radiation	W m−2
	HI	Averaged Heat Index	°C
	TS.	Averaged Soil Temperature	°C
	UC	Averaged Wind Chill	°C
	UD	Averaged Wind direction	degrees
	PA	Averaged Atmospheric Pressure	

Crop Canopy Parameters	Ir	Averaged Red Irradiance (650 nm)	W m-2
	Inir	Averaged Nir Irradiance (810 nm)	W m-2
	Ig	Averaged Green Irradiance (532 nm)	W m-2
	Iy	Averaged Yellow Irradiance (570 nm)	W m-2
	Rr	Averaged Red Radiance (650 nm)	W m-2 nm-1 sr-1
	Rnir	Averaged Near infrared Radiance (810 nm)	W m-2 nm-1 sr-1
	Rg	Averaged Green Radiance (532 nm)	W m-2 nm-1 sr-1
	Ry	Averaged Yellow Radiance (570 nm)	W m-2 nm-1 sr-1
	TC	Averaged Canopy Temperature	°C
	TSB	Averaged Sensor Body Temperature	°C
Soil Water Content	∑SWC0.15	Total soil water to 0.30 m of soil profile	m3/m3
	∑SWC0.45	Total soil water from 0.30 to 0.60 m of soil profile	m3/m3
	∑SWC0.75	Total soil water from 0.60 to 0.90 m of the soil profile	m3/m3
Soil Water Deficit	∑SWD0.15	Total soil water deficit to 0.30 m of soil profile	m3/m3
	∑SWD0.45	Total soil water deficit from 0.30 to 0.60 m of soil profile	m3/m3
	∑SWD0.75	Total soil water deficit from 0.60 to 0.90 m of soil profile	m3/m3
Remaining water capacity	∑RWC0.15	Available water capacity from 0 to 0.30 m of soil profile	m3/m3
	∑RWC0.45	Available water capacity from 0.30 to 0.60 m of soil profile	m3/m3
	∑RWC0.75	Available water capacity 0.60 to 0.90 m of soil profile	m3/m3
Farm Management Data	IC	Cumulative Irrigated Water Via Drip Irrigation	Inches
	CGD	Cumulative Growing Degree Days	days

Teros 10 soil water content sensors manufactured by METERGroup, Inc. were installed within two weeks of planting in the crop row with minimum disturbance to emerging plants at 15 cm, 45 cm, and 75 cm depths. The ZL6 data logger manufactured by METERGroup, Inc. logged the soil moisture every 10 minutes, and the data were downloaded monthly throughout the growing season. A subsurface drip irrigation system at 0.25 m depth applied irrigation to the irrigated treatment.

To calculate the SWD and RWC field capacity (θ_{fc}) and wilting point (θ_{wp}) in **Table** 3, Equations 1 to 4 and data in **Table 4** were used. **Table 4** soil profile data were obtained from the Web Soil Survey (WSS) platform (Soil Survey Staff, 2024). We considered T1, T2, and T4 plots on 7105

Yutan silty clay loam soil, and T3, T5, and T6 were on 7340 Filbert silt loam soil. Refer to **Table 4** for details on soil texture profile down to 80-inch depth.

Table 2 Soil water characteristics (Eisenhauer et al., 2021)

Soil Texture	θfc	θwp	AWC (Equation 1)
	cm3/cm3, in/in, or m/m		
Silt loam	0.3	0.12	0.18
Silty clay loam	0.38	0.22	0.16
Clay loam	0.4	0.25	0.15
Silty clay	0.4	0.27	0.13

Table 3 Soil properties (Soil Survey Staff, 2024)

Soil Type	Soil Texture Along the Soil Profile	Depth cm	θfc	θwp	Sensor Stations
7105 Yutan silty clay loam	Ap - 0 to 6 inches: silty clay loam	0-30	0.38	0.22	2022 Node 1, 3
	Bt1 - 6 to 13 inches: silty clay loam	30-60	0.38	0.22	
	Bt2 - 13 to 28 inches: silty clay loam BC - 28 to 43 inches: silt loam C - 43 to 79 inches: silt loam	60-90	0.33	0.15	
7340 Filbert silt loam	Ap - 0 to 7 inches: silt loam	0-30	0.30	0.12	2022 Node 2
	E - 7 to 15 inches: silt loam	30-60	0.38	0.22	
	Bt - 15 to 62 inches: silty clay Bt - 62 to 80 inches: silty clay loam	60-90	0.40	0.27	2023 Node 3

The SWD and RWC related to each sensor station were derived based on **Table 3** soil water characteristics, **Table 4** soil properties, and Equations 1, 2, and 3. During this calculation, we assume that the soil moisture sensor at 15 cm, 45 cm, and 75 cm depth represent the 0 – 30 cm, 30 – 60 cm, and 60 – 90 cm soil profiles.

Root depth is a critical parameter for plant water uptake. According to (Eisenhauer et al., 2021)The maximum effective rooting depths for fully grown soybeans and corn are 2.6-5.0 feet and 3.3-6.6 feet, respectively. Additionally, water extraction from the roots in shallow depths is higher than in deeper depths (**Figure 2**). These facts also change based on soil resistance to root penetration, management practices, and genotypes. The 0-90 cm (0 - 3 feet) soil profile covered by three soil moisture sensors at each sensor node is nearly equal to the 50% rooting depth of the soybean and corn maximum adequate rooting depth in this study. Therefore, we did not account for the rooting depth for RWC calculation in Equation 3.

In equations 1, 2, 3, and 4, θ_{fc} is field capacity, θ_{wp} wilting point, while θ_v is real-time volumetric water content. The L is the depth that each soil moisture sensor represents in the soil profile; in this study, it is 30 cm. The total SWD was calculated by adding the SWD at three depths.

$$AWC = (\theta_{fc} - \theta_{wp}) \quad (1)$$

$$SWD = (\theta_{fc} - \theta_v)L \quad (2)$$

$$RWC = (\theta_v - \theta_{wp})L \quad (3)$$

$$f_r = \frac{(\theta_{fc} - \theta_v)}{(\theta_{fc} - \theta_{wp})} \quad (4)$$

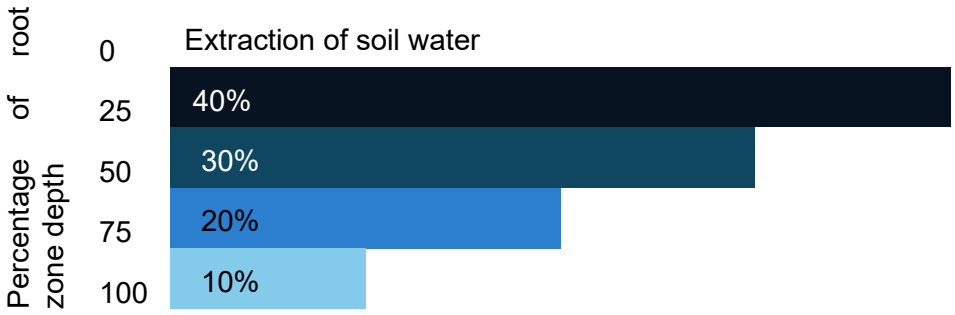


Figure 2 Soil water extraction by plant roots (Eisenhauer et al., 2021)

Maximum daily T_a and T_c were recorded, and a clear difference between Ta and Tc could only be seen at 10:00 AM and after 5:00 PM. Therefore, as the first step of data preprocessing, we filtered out the data points outside the 10:00 AM and after the 5:00 PM window. This step also helps eliminate sun sensor-dependent solar irradiance and canopy reflectance measurement inconsistencies. Information provided by (Chamara, 2021; Ehrler, 1971; Jackson et al., 1977; Singh et al., 2021) This decision is supported by the weather data, Campbell data logger, and meter group data, which were combined with the timestamp and added the cumulative growing degree days (CGD), cumulative irrigated water via drip irrigation, and cumulative precipitation (CP). Data preprocessing, visualization, and linear model development were performed using Rstudio 2023.06.1 software and the “ggplot2” and “ggpubr” libraries.

Discussion about Data

According to **Figure 3**, daily T_a, and rainfall data for 2022 and 2023. The maximum daily T_a was low in 2023 in July, while it exceeded the 2022 T_a in the middle of August. The rainfall in 2023 was significantly lower in August compared to 2022.

Figure 4, **Figure 5**, and **Figure 6** of the box and whisker plots summarize the soil's SWC, total $(\theta_v - \theta_{wp})$, and total $(\theta_{fcv} - \theta_v)$ under the selected sensor stations for this study. Based on Figures **Figure 4**, **Figure 5**, and **Figure 6** the SWC $(\theta_v - \theta_{wp})$ and $(\theta_{fcv} - \theta_v)$ value distributions these 4 nodes represent are significantly different from each other. The variance of SWC at 15 cm on node 1, 45 cm, and 75 cm soil sensors was higher than others.

According to **Figure 5** in 2022, Node 1 showed consistent water saturation in the soil layer at a 60-90 cm depth. Similarly, in 2023, Node 3 exhibited water saturation at 0-30 cm depth. The non-irrigated treatment soybean reached the wilting point per the **Figure 6** 2022 Node 3 box plot.

Therefore, machine learning model testing will be challenging as the models must predict unseen values during training. Therefore, we decided to train the models using 2022 Node 1 and Node 3 for soybeans. Then, 20% of the 2023 data will be used to perform transfer learning before testing the data in 2023. The model related to corn will be presented only with the 2022 Node 2 data.

As we are predicting continouse variables, we decided to use regression learning models in this experiment. The linear model summaries can be found in the Appendix. Additionally, we removed CGD, IC, and PC in model training to improve the independence of features and multicollinearity among features.

Figure 4 Volumetric water content box plots of each selected sensor station to proceed with machine learning model training, testing, and validation.

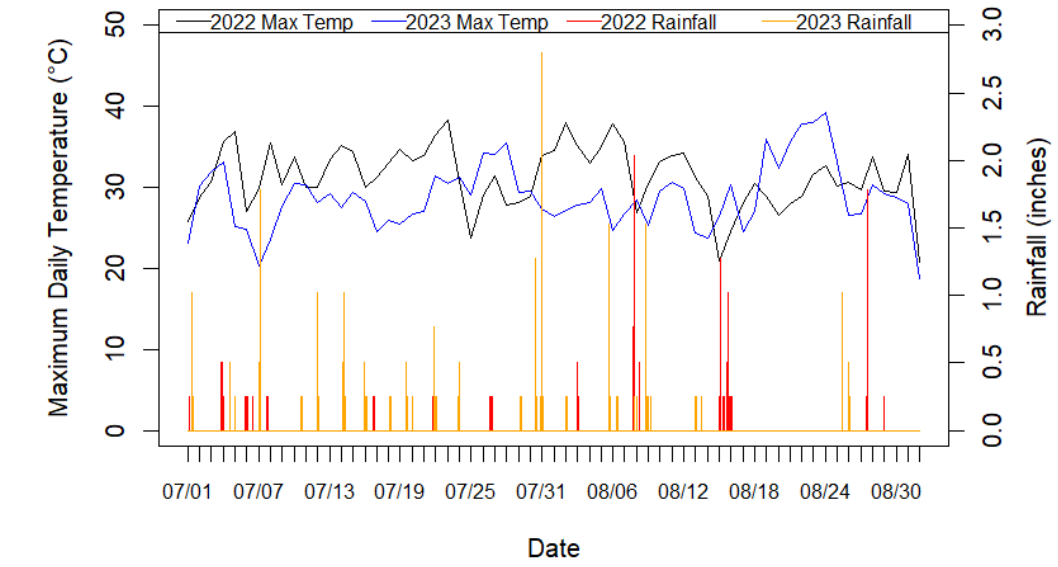
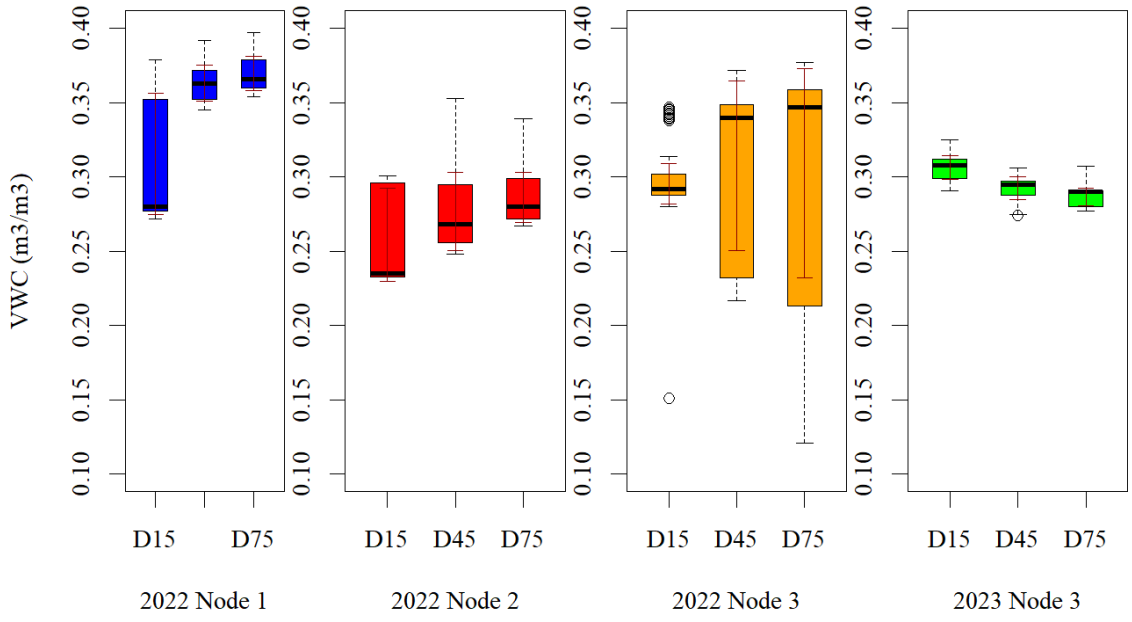


Figure 3 Maximum daily temperature and rainfall data in years 2022 and 2023



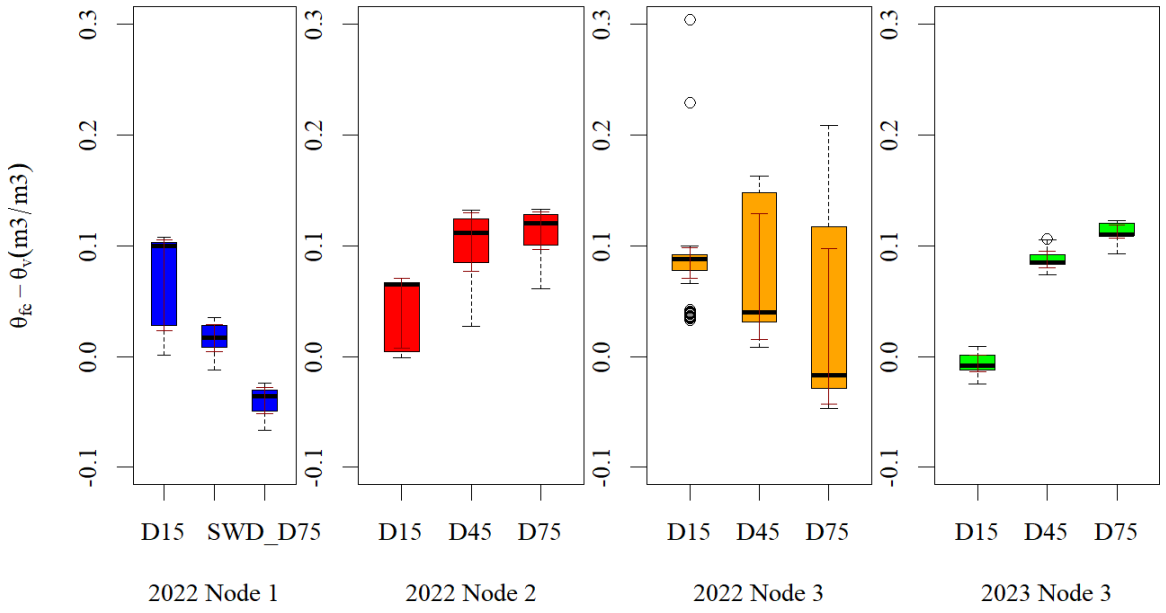


Figure 5 SWD box plots of each selected sensor station to proceed with machine learning model training, testing, and validation.

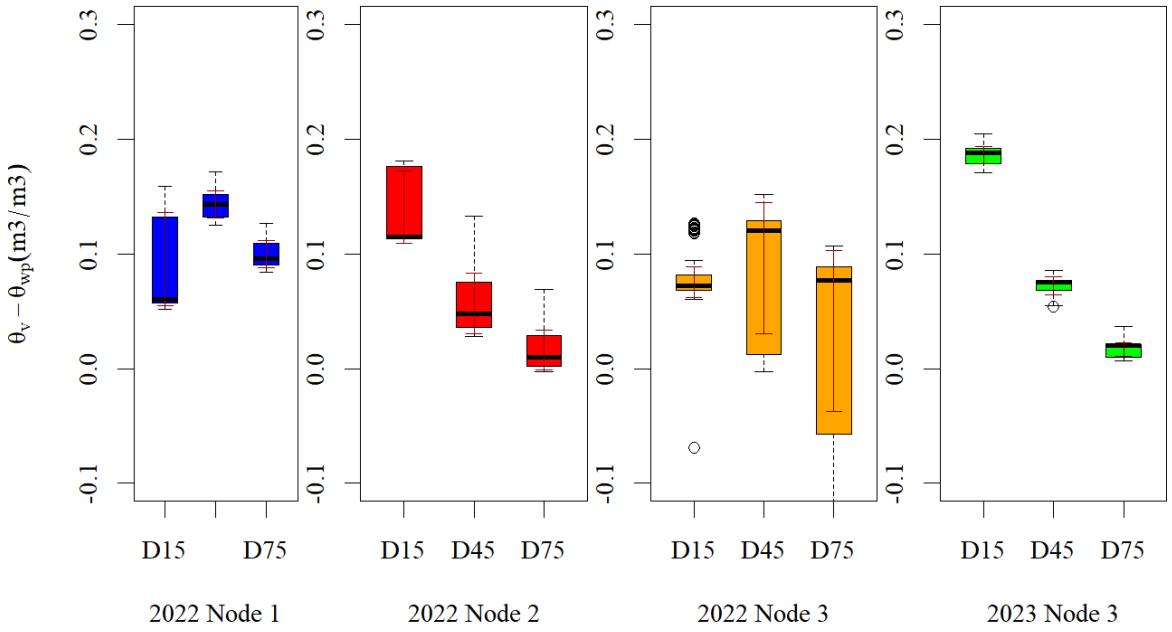


Figure 6 RWC box plots of each selected sensor station to proceed with machine learning model training, testing, and validation.

Machine Learning Model Training, Validation, and Testing

Machine learning model training, validation, and testing were performed in the MATLAB 2022 regression learner app (MATLAB version R2022b). Gaussian process regression (GPR), linear regression (LR), neural network (NN), support vector machines (SVM), stepwise linear regression (SLR), Tree, kernel, and ensemble models can be trained, validated, and tested in the MATLAB regression learner app. All the trained model variants and their hyperparameters are highlighted in **Table 5**. 10% of the data was set aside to evaluate the model. The app partitions the data into k = 10 folds for data not set aside for testing. It estimates the accuracy on each fold to protect the model training against overfitting. Bayesian optimization expected improvement per second plus acquisitions function, 30 iterations, and a training time limit of 600 seconds maximum were the selected model optimizer options during model training.

Table 4 Model type, default model number in MATLAB regression learner app, and hyperparameters

Model Number	Model Type	Hyperparameters
2.15	Ensemble	Minimum leaf size: 8; Number of learners: 30
2.14	Ensemble	Minimum leaf size: 8; Number of learners: 30; Learning rate: 0.1
2.19	Gaussian Process Regression	Basis function: Constant; Kernel function: Rational Quadratic; Use isotropic kernel: Yes; Kernel scale: Automatic; Signal standard deviation: Automatic; Sigma: Automatic; Standardize data: Yes; Optimize numeric parameters: Yes
2.17	Gaussian Process Regression	Basis function: Constant; Kernel function: Matern 5/2; Use isotropic kernel: Yes; Kernel scale: Automatic; Signal standard deviation: Automatic; Sigma: Automatic; Standardize data: Yes; Optimize numeric parameters: Yes
2.16	Gaussian Process Regression	Basis function: Constant; Kernel function: Squared Exponential; Use isotropic kernel: Yes; Kernel scale: Automatic; Signal standard deviation: Automatic; Sigma: Automatic; Standardize data: Yes; Optimize numeric parameters: Yes
2.18	Gaussian Process Regression	Basis function: Constant; Kernel function: Exponential; Use isotropic kernel: Yes; Kernel scale: Automatic; Signal standard deviation: Automatic; Sigma: Automatic; Standardize data: Yes; Optimize numeric parameters: Yes
2.25	Kernel	Learner: SVM; Number of expansion dimensions: Auto; Regularization strength (Lambda): Auto; Kernel scale: Auto; Epsilon: Auto; Iteration limit: 1000
2.26	Kernel	Learner: Least Squares Kernel; Number of expansion dimensions: Auto; Regularization strength (Lambda): Auto; Kernel scale: Auto; Iteration limit: 1000
2.2	Linear Regression	Terms: Interactions; Robust option: Off
2.1	Linear Regression	Terms: Linear; Robust option: Off
2.3	Linear Regression	Terms: Linear; Robust option: On
2.22	Neural Network	Number of fully connected layers: 1; First layer size: 100; Activation: ReLU; Iteration limit: 1000; Regularization strength (Lambda): 0; Standardize data: Yes
2.21	Neural Network	Number of fully connected layers: 1; First layer size: 25; Activation: ReLU; Iteration limit: 1000; Regularization strength (Lambda): 0; Standardize data: Yes
2.24	Neural Network	Number of fully connected layers: 3; First layer size: 10; Second layer size: 10; Third layer size: 10; Activation: ReLU; Iteration limit: 1000; Regularization strength (Lambda): 0; Standardize data: Yes

2.23	Neural Network	Number of fully connected layers: 2; First layer size: 10; Second layer size: 10; Activation: ReLU; Iteration limit: 1000; Regularization strength (Lambda): 0; Standardize data: Yes
2.2	Neural Network	Number of fully connected layers: 1; First layer size: 10; Activation: ReLU; Iteration limit: 1000; Regularization strength (Lambda): 0; Standardize data: Yes
2.4	Stepwise Linear Regression	Initial terms: Linear; Upper bound on terms: Interactions; Maximum number of steps: 1000
2.12	SVM	Kernel function: Gaussian; Kernel scale: 3.3; Box constraint: Automatic; Epsilon: Auto; Standardize data: Yes
2.1	SVM	Kernel function: Cubic; Kernel scale: Automatic; Box constraint: Automatic; Epsilon: Auto; Standardize data: Yes
2.11	SVM	Kernel function: Gaussian; Kernel scale: 0.83; Box constraint: Automatic; Epsilon: Auto; Standardize data: Yes
2.9	SVM	Kernel function: Quadratic; Kernel scale: Automatic; Box constraint: Automatic; Epsilon: Auto; Standardize data: Yes
2.13	SVM	Kernel function: Gaussian; Kernel scale: 13; Box constraint: Automatic; Epsilon: Auto; Standardize data: Yes
2.8	SVM	Kernel function: Linear; Kernel scale: Automatic; Box constraint: Automatic; Epsilon: Auto; Standardize data: Yes
2.5	Tree	Minimum leaf size: 4; Surrogate decision splits: Off
2.6	Tree	Minimum leaf size: 12; Surrogate decision splits: Off
2.7	Tree	Minimum leaf size: 36; Surrogate decision splits: Off

Model evaluation to identify the best models.

All the models were evaluated and averaged based on the equations (5), (6), (8), and (9). In the equations (5), (6), (7), and (8) P_i is the model prediction and O_i observed or the measured value and n is the number of observations.

$$r^2 = \frac{n\sum O_i P_i - (\sum O_i)(\sum P_i)}{\sqrt{[n\sum O_i^2 - (\sum O_i)^2][n\sum P_i^2 - (\sum P_i)^2]}} \quad (5)$$

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(P_i - O_i)| \quad (7)$$

$$MAPE = \left[\frac{1}{n}\sum_{i=1}^{n}|(P_i - O_i)|\right] \times 100\% \quad (8)$$

Results And Discussion

The correlation coefficient and the associated p values (see appendix) revealed significant correlations between almost all the parameters with SWC and SWD, supporting the finding of (Singh et al., 2021) relationship between ΔT and SWD. However, it shows no significant real-time

correlation between rain and the soil water content. We suggest that this is because it takes some time for the rainwater to penetrate the ground and increase the soil water content. **Figure** 7 summarizes the RSquared values of each model type on SWC, SWD, and RWC at 15 cm, 45 cm, and 75 cm depths and the TSWC, TSWD, and TRWC on the test data. One interesting outcome was that model type tree, gaussian process regression, and neural networks achieved above 0.95 RSquared values on the test data. This helps to conclude that the T_a, RH, U, G, T_S, U_D, P_A, and T_C parameters can explain the complex relationship with ∑SWC0.15, ∑SWC0.45, ∑SWC0.75, ∑SWD0.15, ∑SWD0.45, and ∑SWD0.75 with type tree, gaussian process regression (GPR), and neural networks with higher accuracy. Only three GPR models show significantly high R-squared values for all the predicted variables.

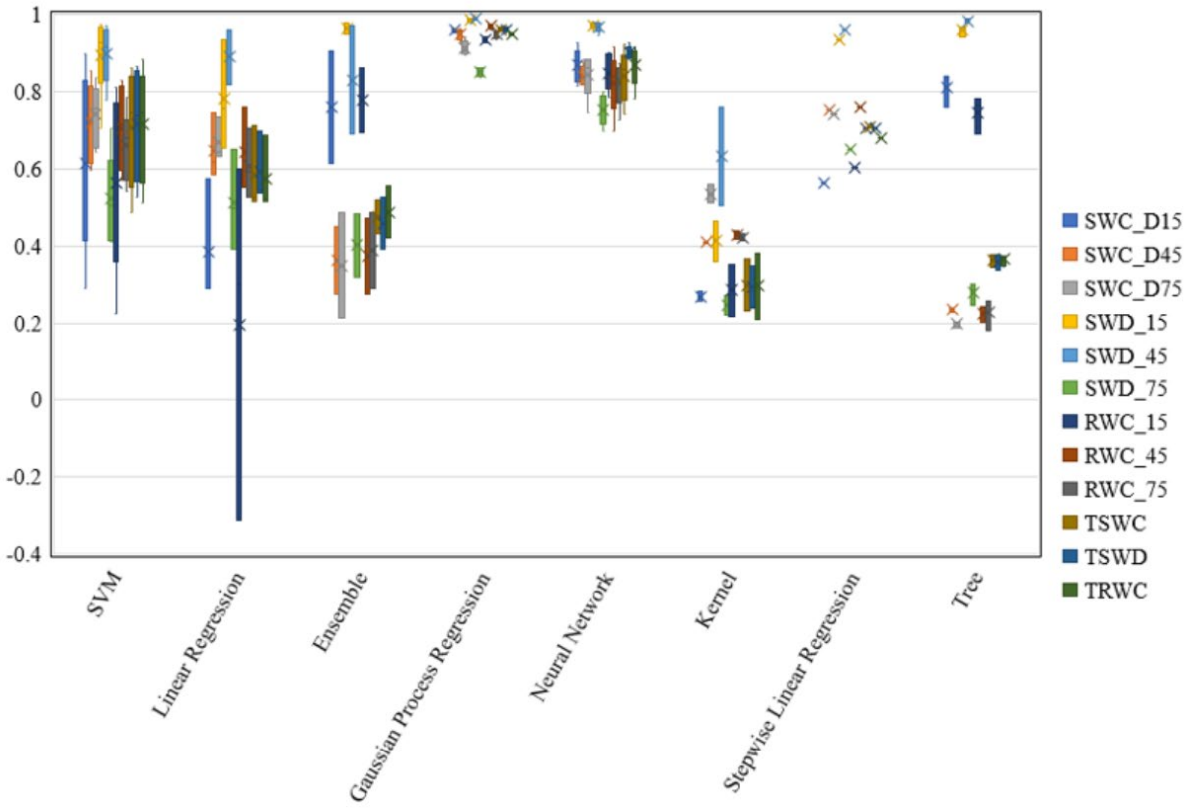


Figure 7 R squared values for different machine learning models for the test data set for soybean

SWD at 15 cm and 45 cm depth prediction models have the highest R squared value. Additionally, the SWC and SWD at 15 cm and 45 cm depth prediction performance on test data were better than the SWC and SWD at 75 cm depth. TSWC, TSWD, and TRWC prediction have the same prediction performance on all the model types.

Additionally, we have used two feature ranking algorithms to identify how the selected features contribute to the prediction of SWD. They are Minimum Redundancy, Maximum Relevance (MRMR) (Ding & Peng, 2005), and F-test. According to **Figure 8** and **Figure 9**, a critical outcome of this ranking is that atmospheric pressure was identified as a key feature in the model relevant to SWD prediction. However, the MRMR algorithm ranked PA (atmospheric pressure), RH (relative humidity), Rnir (Near-infrared Radiance (810 nm)), UD (wind direction), and Tc (canopy temperature) as the top 5 relevant to the prediction of SWD. In contrast, the F test ranked Tc, PA, Rr (Averaged Red Radiance (650 nm)), Ta (average air temperature), and Ry (Averaged Yellow

Radiance (570 nm)) as the top 5 relevant features for SWD prediction.

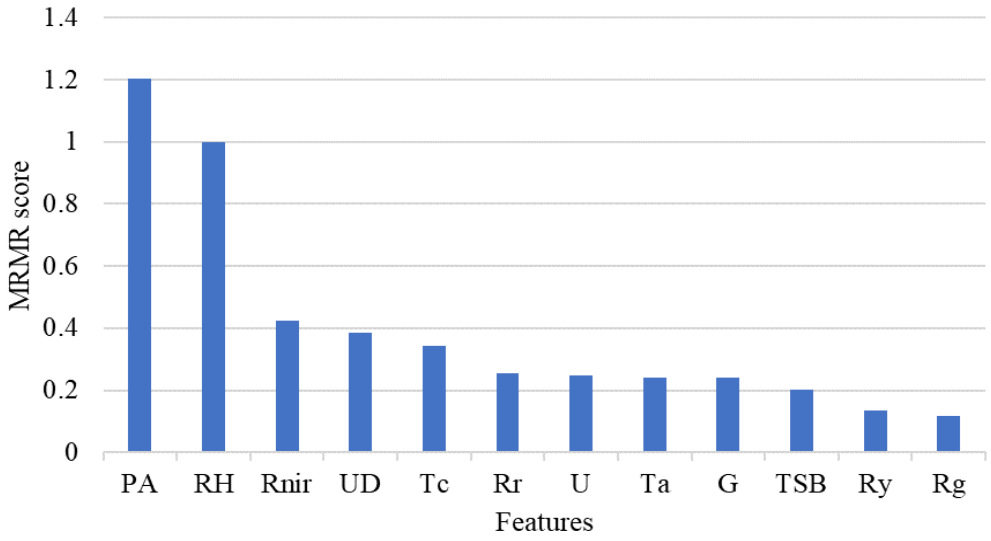


Figure 8 Feature importance ranking by the MRMR algorithm.

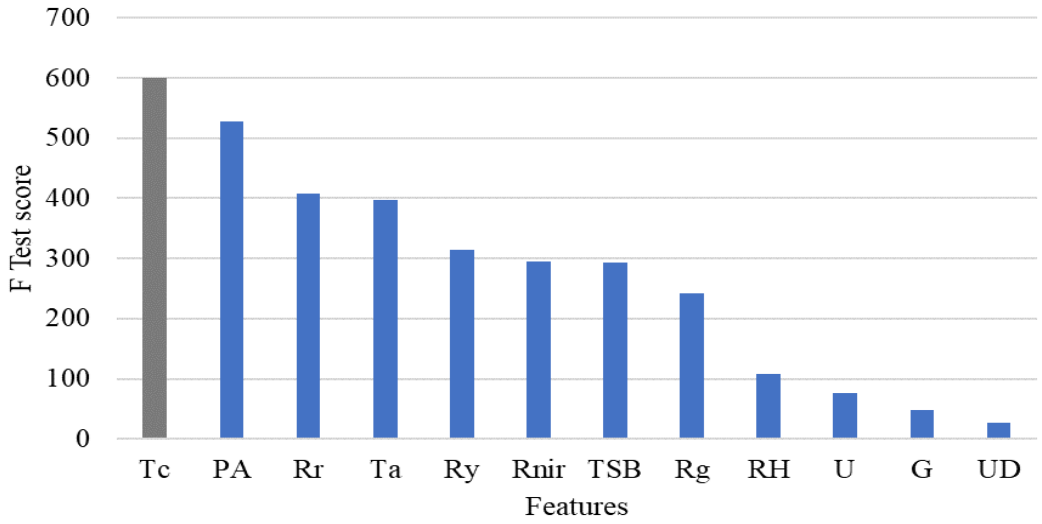


Figure 9 Feature importance ranking by the F Test (Tc has an infinity F score)

Generally, the relevant features identified by the machine feature ranking algorithms aligned with past research. However, these results lay the basis for using proximal remote sensing for soil water parameter prediction with machine learning models. But practically, we need more data to train the models for testing with real-time soil water parameter prediction. The data that represent different soil types and SWC spread from wilting point to field capacity (FC) are required to train or perform transfer learning for the best-performing models selected in this research for better

models. More data will help generalize the model and eliminate the effects of proximal remote sensing sensor placement or sensor agonistic issues.

According to the author's knowledge, this is the first time we have used the Ta, RH, U, G, TS, UD, PA, and TC parameters from proximal remote sensing data to estimate SWC, SWD, and RWC in soybean and corn fields.

This research outcome will be applied to develop proximal remote sensing tools for estimating soil water properties for irrigated field crops. Ground or aerial platforms will benefit from this research outcome as they can be used with NDVI, PRI, Canopy temperature, and weather data to accurately estimate the real-time SWD on field crops.

References

Abeygunawardena, T. S., Chamara, N., & Indeewara, A. (2023). *Use of Sensor Based Automated Irrigation for the Mitigation of Groundwater Depletion and Pollution Issues in Kalpitiya , Sri Lanka Use of Sensor Based Automated Irrigation for the Mitigation of Groundwater Depletion and Pollution Issues in Kalpitiya , Sri*. *September 2022*.

Adeyemi, O., Grove, I., Peets, S., Domun, Y., & Norton, T. (2018). Dynamic neural network modelling of soil moisture content for predictive irrigation scheduling. *Sensors (Switzerland)*, *18*(10). https://doi.org/10.3390/s18103408

Bai, G., Ge, Y., Scoby, D., Leavitt, B., Stoerger, V., Kirchgessner, N., Irmak, S., Graef, G., Schnable, J., & Awada, T. (2019). NU-Spidercam: A large-scale, cable-driven, integrated sensing and robotic system for advanced phenotyping, remote sensing, and agronomic research. *Computers and Electronics in Agriculture*, *160*(January), 71–81. https://doi.org/10.1016/j.compag.2019.03.009

Baztan, J., Bremer, S., da Cunha, C., De Rudder, A., Jaffrès, L., Jorgensen, B., Krauß, W., Marschütz, B., Peeters, D., Jensen, E. S., Vanderlinden, J. P., Wardekker, A., & Zhu, Z. (2022). Water and Climate Change. In *Water and Climate Change: Sustainable Development, Environmental and Policy Issues* (Issue July). https://doi.org/10.1016/B978-0-323-99875-8.00005-7

Chamara, N. (2021). Development of an Internet of Things (IoT) Enabled Novel Wireless Multi Sensor Network for Infield Crop Monitoring. In *University of Nebraska*.

Chamara, N., Amarasinghe, R., & Nandika, D. (2012). *Sensor Based Self Powered Smart Control System for Development of a Sensor Based Self Powered Smart Control System for*. *November*. https://doi.org/10.4038/jfa.v5i1-2.5180

Chamara, N., Islam, D., Frank, G., Shi, Y., & Ge, Y. (2022). Ag-IoT for crop and environment monitoring : Past , present , and future. *Agricultural Systems*, *203*(September), 103497. https://doi.org/10.1016/j.agsy.2022.103497

Ding, C., & Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of Bioinformatics and Computational Biology*, *3*(2), 185–205.

Ehrler, W. L. (1971). Cotton leaf temperatures as related to soil water depletion and meteorological factors. *Agronomy Journal*, *65*(3), 404–409.

Eisenhauer, D. E., L., D., Heeren, D. M., & Hoffman, G. J. (2021). Irrigation Systems. In *American Society of Agricultural and Biological Engineers (ASABE)*. https://doi.org/10.1201/b13110-5

FAO. (2021). *The State of the World's Land and Water Resources for Food and Agriculture—*

Systems at Breaking Point.

Gu, Z., Zhu, T., Jiao, X., Xu, J., & Qi, Z. (2021). Neural network soil moisture model for irrigation scheduling. *Computers and Electronics in Agriculture*, *180*(1), 105801. https://doi.org/10.1016/j.compag.2020.105801

Hatfield, J. L., & Dold, C. (2019). Water-use efficiency: Advances and challenges in a changing climate. *Frontiers in Plant Science*, *10*(February), 1–14. https://doi.org/10.3389/fpls.2019.00103

Jackson, R. D., Reginato, R. J., & Idso, S. B. (1977). Wheat canopy temperature: A practical tool for evaluating water requirements. *Water Resources Research*, *13*(3), 651–656. https://doi.org/10.1029/WR013i003p00651

Katimbo, A., Rudnick, D. R., Zhang, J., Ge, Y., DeJonge, K. C., Franz, T. E., Shi, Y., Liang, W. zhen, Qiao, X., Heeren, D. M., Kabenge, I., Nakabuye, H. N., & Duan, J. (2023). Evaluation of artificial intelligence algorithms with sensor data assimilation in estimating crop evapotranspiration and crop water stress index for irrigation water management. *Smart Agricultural Technology*, *4*(October 2022), 100176. https://doi.org/10.1016/j.atech.2023.100176

Lo, T. H., Heeren, D. M., Mateos, L., Luck, J. D., & Martin, D. L. (2015). *Potential Irrigation Reductions From Increasing Precipitation Utilization With Variable Rate Irrigation*.

Mendes, W. R., e Videira, A. M., Er-Raki, S., Heeren, D. M., Dutta, R., & Araújo, F. M. U. (2023). Development of a Fuzzy Variable Rate Irrigation Control System Based on Remote Sensing Data to Fully Automate Center Pivots. *IEEE Transactions on Automation Science and Engineering*, 1–17. https://doi.org/10.1109/tase.2023.3322120

Singh, J., Ge, Y., Heeren, D. M., Walter-Shea, E., Neale, C. M. U., Irmak, S., Woldt, W. E., Bai, G., Bhatti, S., & Maguire, M. S. (2021). Inter-relationships between water depletion and temperature differential in row crop canopies in a sub-humid climate. *Agricultural Water Management*, *256*. https://doi.org/10.1016/j.agwat.2021.107061

Soil Survey Staff. (2024). *Web Soil Survey*. Natural Resources Conservation Service, United States Department of Agriculture. https://websoilsurvey.nrcs.usda.gov/app/WebSoilSurvey.aspx

Tsang, S. W., & Jim, C. Y. (2016). Applying artificial intelligence modeling to optimize green roof irrigation. *Energy and Buildings*, *127*, 360–369. https://doi.org/10.1016/j.enbuild.2016.06.005

Wang, Y., Zhao, J., Guo, Z., Yang, H., & Li, N. (2023). Soil Moisture Inversion Based on Data Augmentation Method Using Multi-Source Remote Sensing Data. *Remote Sensing*, *15*(7). https://doi.org/10.3390/rs15071899