Discover Applied Sciences

RESEARCH

Open Access

# Explainable XGBoost-based models of root-zone soil moisture profiling using coupled Sentinel-2 and IoT data in loam and silt loam soil

Tarek Alahmad[1], Miklós Neményi[1] and Anikó Nyéki[1*]

*Correspondence:
Anikó Nyéki
nyeki.aniko@sze.hu
[1]Albert Kázmér Mosonmagyaróvár Faculty of Agricultural and Food Sciences, Department of Bioengineering and Precision Technology, Széchenyi István University, Mosonmagyaróvár, Hungary

## Abstract

**Background** Accurate prediction of soil moisture content (SMC) is crucial for sustainable irrigation management and enhancing resilience against climate change. However, in-situ sensing offers accurate point-scale measurements but lacks spatial representativeness, while satellite-based offer spatial coverage but are either too coarse at field scale or indirect and cloud -sensitive. Integrating satellite observation with ground-based monitoring and IoT meteorological data could exploit complementary strengths by linking canopy conditions and atmospheric drivers to reliable in-field reference measurements.

**Method** This study predicts SMC at five depths (5 to 80 cm) for two soil texture classes (loam and silt loam) using Extreme Gradient Boosting (XGBoost) by integrating Sentinel-2 vegetation indices Normalized Difference Vegetation Index (NDVI) and Normalized Difference Moisture Index (NDMI) with Internet of Things (IoT)-derived meteorological data using two input scenarios and gravimetric SMC as reference for model training and evaluation.

**Results** The model trained with combined inputs achieved higher accuracy compared with using only vegetation indices as predictors in both soil textures and all depths. This was especially evident in loam soil at 5 and 20 cm depth, with $R^2$ values of 0.95 and 0.79 and RMSE values of 0.88% and 1.46%, respectively, compared to $R^2$ values of 0.70 and 0.69 and RMSE values of 2.26% and 1.77% when using vegetation indices only. The model achieved near-perfect accuracy in silt loam with $R^2 = 0.99$ at 5–20 cm and RMSE = 0.49–0.40% at same depths. SHapley Additive exPlanations (SHAP) analysis identified NDVI as the most influential predictor in surface soil layers (mean SHAP = 0.11–0.22), reflecting its strong sensitivity to canopy vigor. In contrast, solar radiation emerged as key determinant in deeper soil layers (60 and 80 cm; SHAP = 0.12–0.18), highlighting the importance of atmospheric evaporative demand in controlling subsoil moisture dynamics.

**Conclusions** The model's accuracy and interpretability enable depth-specific decision support for irrigation timing and water use efficiency under variable

Discover

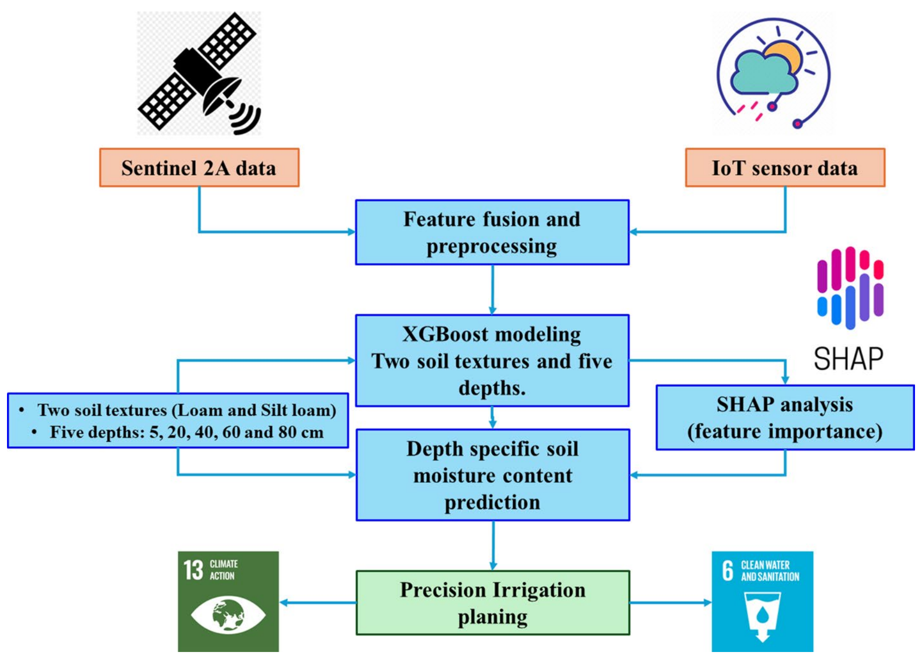
weather conditions, while providing actionable driver insights for climate-adaptive management aligned with SDGs 6 and 13. The approach is validated for loam and silt loam textures using optical Sentinel-2 indices, which are subject to cloud cover and revisit latency; therefore, the current framework is not suitable for real-time irrigation scheduling without accounting for these delays. Future integration with SAR and gap-filling strategies would be required for operational real-time applications.

### Article highlights

- XGBoost achieved high accuracy in predicting soil moisture across five depths.
- Integrating Sentinel-2-derived vegetation indices with IoT data improved prediction across five depths and two soil textures (loam and silt loam).
- Explainable AI (SHAP) analysis clarified depth-dependent drivers, with NDVI had the highest importance in shallower depths and solar radiation dominated in the deeper depths.
- The proposed approach supports depth specific irrigation decision-making and sustainable water management aligned with SDG 6 and 13.

**Keywords** Root-zone soil moisture, Depth-specific modeling, Multi-source data fusion, SHAP, XGBoost, Decision support

### Graphical abstract




Workflow for depth-specific soil moisture prediction using Sentinel-2 vegetation indices and IoT meteorological data with explainable XGBoost (SHAP).

## 1 Introduction

Soil moisture content is a critical factor for crop growth and agricultural productivity, directly influencing yield and food security [1, 2]. However, the significant spatial and temporal variation characterizing soil moisture variability, affected by diverse features including climate factors such as precipitation, solar radiation, temperature, and atmospheric evaporation demand, soil properties, and vegetation factors, presents a formidable challenge for accurate soil moisture forecasting [3–5]. Conventional methods for SMC prediction such as in-situ probes are limited in spatial coverage and often in accuracy [6]. Microwave remote sensing, particularly from satellites like Soil Moisture Active Passive (SMAP) and Soil Moisture and Ocean Salinity (SMOS), provides a more direct physical measurement of surface soil moisture and is less affected by atmospheric conditions. However, its coarse spatial resolution (often > 9 km) makes it unsuitable for field-scale precision agriculture applications. Furthermore, its signal is attenuated by dense vegetation canopies, limiting its accuracy in monitoring subsoil moisture dynamics in actively growing crops [7].

Remote sensing offers an attractive complementary data source: satellites like Sentinel-2 provide vegetation indices related to crop vigor and indirect soil moisture [8]. However, satellite observations typically represent near-surface conditions and may not capture subsoil moisture dynamics [9]. Furthermore, purely satellite-based indices might be influenced by factors such as crop development stage or soil backdrop [10]. *While microwave Synthetic Aperture Radar (SAR) missions such as Sentinel-1 can be more directly sensitive to the near surface soil moisture due to microwave penetration and reduce dependence on illumination [11], SAR backscatter an annual crop such as maize is strongly affected by canopy structure and surface roughness, and reliable soil moisture retrieval often require additional corrections or scattering-model parametrization [12, 13]. In this research, operational, low-complexity fusion of Sentinel-2 indices (NDVI and NDMI) with high frequence IoT meteorological data were used to evaluate weather vegetation conditions and atmospheric drivers could provide accurate and explainable depth-specific SMC prediction. On the other hand, Internet of Things (IoT) sensor networks (weather stations, soil probes) deliver local, real-time meteorological data that drive soil moisture changes [14, 15]. Integrating the high-resolution spatial context from satellites with direct, real-time forcing data from IoT sensors is a promising approach to overcome their individual limitations and improve SMC predictions across the soil profile [7]. In this research, Sentenil-2 and IoT data are combined through feature level fusion rather than pixel wide real-time mapping to support depth-specific SMC prediction. Sentinel-2 (NDVI and NDMI) are used as canopy state proxies, whereas IoT measurements provide high frequency meteorological forcing. However, Sentinel-2 observations may not coincide with field measurements, the sampling dates were paired with the closest cloud-free Sentinel-2 observations within a short-temporal window. Also, the revisit frequency and processing latency could limit the real-time regional deployment; operationally, prediction could be updated when new observations become available while meteorological demand remain continuous. It is important to clarify that Sentinel-2 data in this research are not used for spatial mapping of soil moisture across the field. Instead, vegetation indices are extracted precisely at sampling point locations to serve as proxies for crop canopy status. This captures the cumulative effect of soil moisture availability on plant vigor overtime information that is complementary to and cannot be obtained from real-time meteorological measurements alone. Thus, even without spatial mapping, the point-extracted indices provide valuable temporal information about vegetation response to soil water status.

Machine learning (ML) techniques have shown great success in modeling complex, non-linear relationships between environmental variables and soil moisture [16]. Unlike traditional linear regression approaches, ML models such as Support Vector Machines, Random Forest, and gradient-boosted trees can flexibly learn interactions between climate, vegetation, and soil properties [17]. XGBoost (eXtreme Gradient Boosting) is an efficient tree-boosting algorithm that often outperforms linear models in soil moisture prediction research [18]. Recent studies have applied ML to multi-source data. For instance, Emami et al. (2025) [19] developed BS-XGB, which combines backtracking search optimization with XGBoost, to optimize hyperparameters for soil moisture prediction at four depths: 10, 25, 50, and 100 cm. The model was trained and tested using in-situ meteorological and soil data including soil temperature, air temperature, solar radiation and relative humidity. The model achieved $R^2$=0.999 for training and 0.973 for testing. Soil temperature, humidity, and solar radiation were among the key predictors. The model is very accurate and efficient, exhibiting XGBoost's greatest potential in SMC prediction. Similarly, Zhu et al. (2024) [20] employed UAV imagery and ensemble learning (EL) for real-time SMC monitoring across Kiwifruit orchards root-zone depths. This study predicts SMC over multiple depth intervals: 0–10 cm (SMC10), 0–20 cm (SMC20), 0–30 cm (SMC30), 0–40 cm (SMC40), 0–50 cm (SMC50), 0–60 cm (SMC60). Gradient boosting decision tree (GBDT) excelled at SMC10 ($R^2$=0.963 ± 0.030, RMSE = 0.238 ± 0.111), while XGBoost outperformed for SMC20–SMC60 ($R^2$=0.963 ± 0.024, RMSE = 0.117 ± 0.053).

The accurate prediction of SMC and the understanding of SMC dynamics could support irrigation management, efficient water use, and drought mitigation. This is globally significant given increasing food demand and water scarcity; precision agriculture technologies have become essential to optimize resource use and enhance climate resilience. This pursuit directly supports global efforts towards the United Nations Sustainable Development Goals (SDGs), particularly SDG 6 (Clean Water and Sanitation) through efficient water use, and SDG 13 (Climate Action) by building adaptive capacity to climate extremes [21, 22]. While previous studies have utilized ML for SMC prediction, few have combined high-resolution optical satellite data with real-time, on-site IoT meteorological data in a depth-resolved framework for annual crops. More importantly, a significant gap exists in the interpretability of these complex models, which is crucial for building user trust and translating predictions into actionable irrigation advice. This study addresses these gaps by developing an interpretable XGBoost model that combines Sentinel-2 vegetation indices with IoT meteorological data to predict SMC at five depths (5–80 cm) in loam and silt loam soils; to compare the performance advantages of multi-source data fusion against the use of vegetation indices only (NDVI and NDMI), and to apply SHapley Additive exPlanations (SHAP) analysis to identify depth-specific SMC drivers in the two studied soil textures and translate these findings into irrigation recommendations. This work enables precision agriculture corresponding to SDGs 6 and 13 by combining model accuracy with actionable explanations.

## 2 Materials and methods

### 2.1 Study site

The field study was conducted in a 23 ha rainfed maize field near Mosonmagyaróvár, northwest Hungary (47°54’11.8"N 17°15’08.9"E) this rainfed field have been chosen to avoid the impact of irrigation on SMC variability driven primarily by natural meteorological conditions. The experimental field's soil is spatially heterogeneous and classified as loam and silt loam at the selected locations, based on the USDA soil texture classification (USDA, 2020) [23]. The field has a flat topography with a slight slope of 5%. Given the low-relief nature of the study site (mean slope ~ 5%), topographic variables such as slope and Topographic Wetness Index (TWI) are expected to have limited predictive power for soil moisture dynamics. This is consistent with findings from similar low-relief agricultural landscapes where terrain attributes were found to be poor indicators of spatial soil variation [24, 25]. During two growing seasons (2023–2024), in each season the maize was sowing in May and harvested in October climate conditions were typical for the regional climate, with precipitation from May to October each season ranged from 350 to 400 mm, and mean temperature during the season ranged between 18.5 °C and 21.3 °C, while the peak daily solar radiation were occurred in (June, July August and few days in September) with an average of 155 w/m$^2$ during the two seasons.

### 2.2 Soil sampling and gravimetric determination

Soil moisture samples were collected during two growing seasons (2023 and 2024) throughout the maize crop cycle. Sampling was conducted on 23 dates across both seasons, focusing on two selected soils textures: loam and silt loam (Fig. 1A) and the SMC variation across the two studied seasons shown in Fig. 2. Three representative points were chosen within each soil texture, resulting in a total of six sampling locations (three in the loam area and three in the silt loam area). At each location, soil samples were taken at five depths: 5 cm, 20 cm, 40 cm, 60 cm, and 80 cm. The shallowest depth (5 cm) represents the near-surface soil, while 20 cm and 40 cm correspond to the majority of root zone of maize (topsoil and mid-rooting depth) which could reach 150–160 cm depth [26], and 60–80 cm represent deeper subsoil layers potentially below the main rooting zone. Immediately after extraction, soil samples were sealed in airtight bags to prevent moisture loss and transported to the laboratory for gravimetric moisture determination, soil properties shown in (Table 1), and descriptive statistics (mean, standard deviation, and min–max range) of the gravimetrically measured SMC are reported in (Table 2) for each soil texture and depth.

### 2.3 Data collection

An on-site IoT sensor station Boreas LoRaWAN datalogger (EcoLogger BEL-06) connected to appropriate sensors was installed in the field to continuously monitor meteorological conditions (Fig. 1C). The station measured key meteorological variables including: air temperature (°C) and relative humidity (%) at ~2 m height (capacitive thermo-hygrometer, ±0.5 °C and ±3% accuracy); precipitation (mm) using a tipping-bucket rain gauge (±0.2 mm); wind speed (km/h) at 2 m (cup anemometer, ±0.5 m/s); and solar radiation (W/m$^2$) via a pyranometer (spectrum 300–1100 nm, ±5% accuracy). The IoT network operated constantly, and regular calibrations for sensors and checks were made to ensure the quality of the data. These sensors were part of a wireless IoT network transmitting data at 10-minute intervals to a cloud server. The overall date range covered two growing seasons, thus spanning diverse environmental conditions for model training. All meteorological measurements were quality-checked and aligned with the sampling timestamps and the measured SMC used as the reference target variable for model training and evaluation.

Sentinel-2 multispectral imagery was utilized to derive vegetation indices indicative of crop canopy status and indirectly soil moisture at specific sampling locations. Sentinel-2 Level-2 A products [27], which provide atmospherically corrected surface reflectance, were processed using QGIS software (version 3.36.2) [28]. While Google Earth

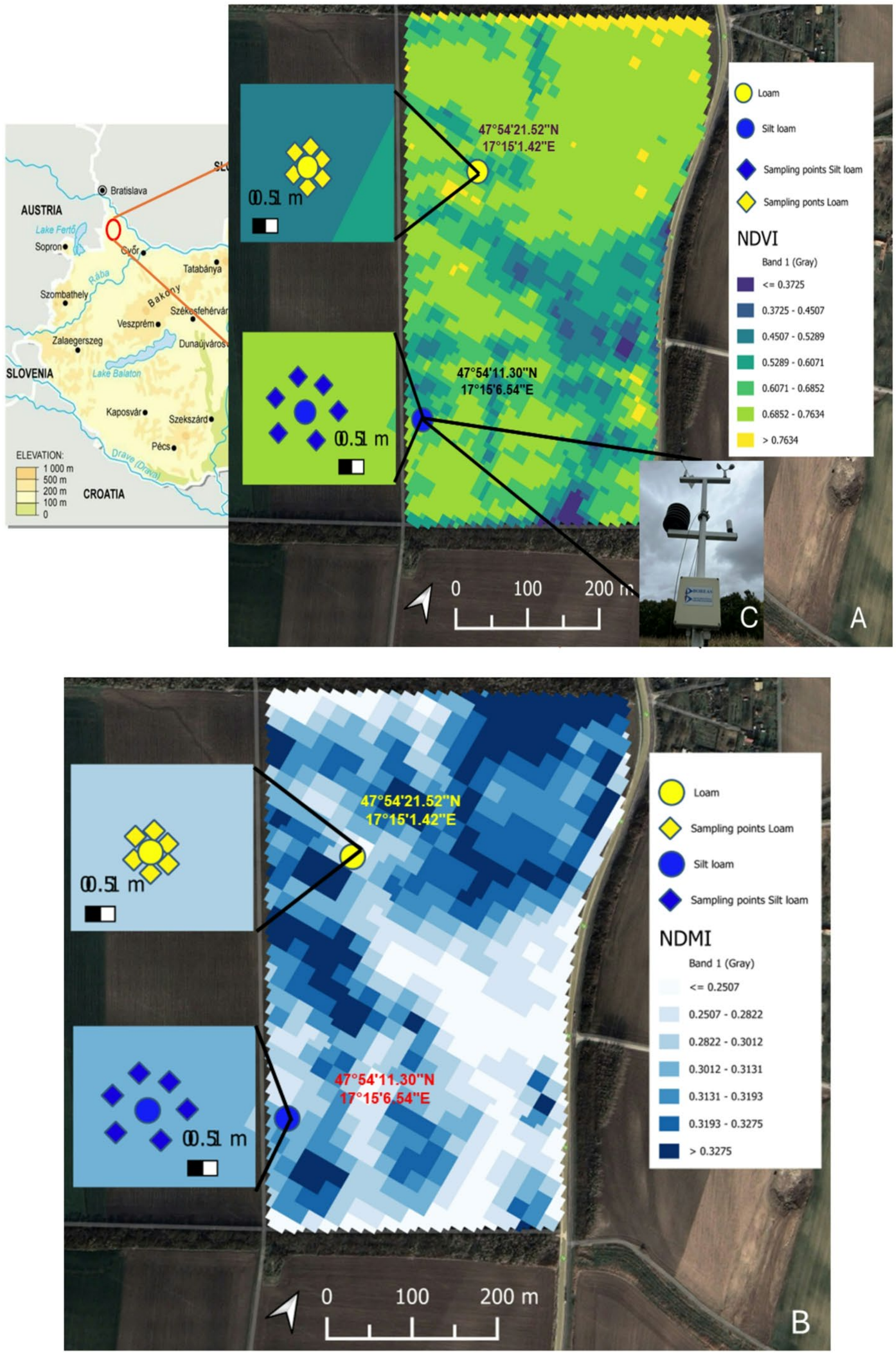



**Fig. 1** The field location from topographical map of Hungary, masked with (**A**) NDVI index and values in one date and (**B**) NDMI index and values downloaded from sentinel hub and visualized by QGIS, with sampling points and (**C**) IoT based sensor

Engine (GEE) offers powerful cloud-based processing capabilities for large-scale analyses, our point-based extraction workflow required: (i) precise control over index calculation at specific GPS coordinates, (ii) integration with locally stored laboratory data and field records, and (iii) reproducible extraction across 23 sampling dates with consistent per-sample quality control. QGIS allowed manual verification of each extraction against

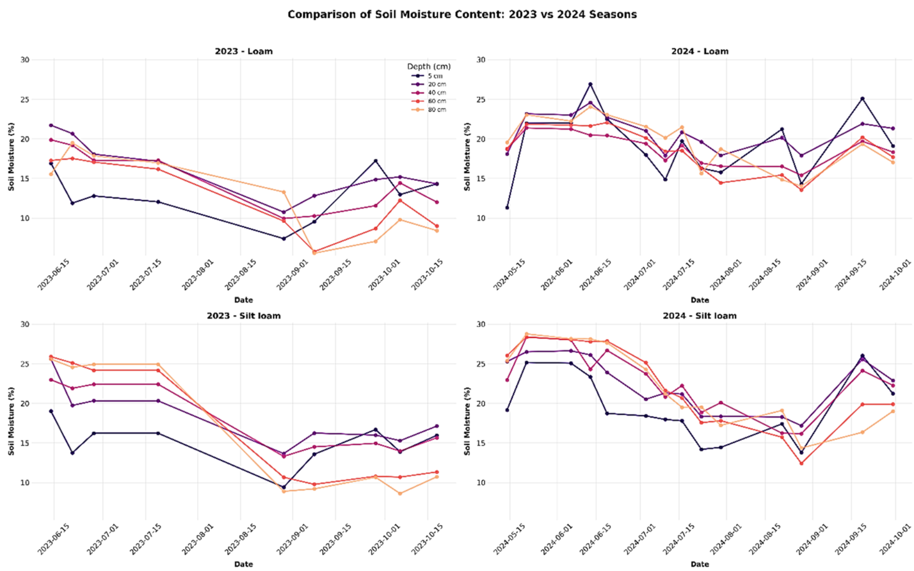



**Fig. 2** Soil moisture content variation across five depths in two soil textures (loam and silt loam) in 2023 and 2024 seasons with 23 soil sampling dates across both seasons

**Table 1** Soil properties across five depths based on laboratory analysis results

| Site | Depth (cm) | Sand% | Silt% | Clay% | Soil organic matter (%) | pH |
|---|---|---|---|---|---|---|
| Site A Loam soil | 5 | 40.99 | 44.41 | 14.5 | 1.17 | 7.43 |
| | 20 | 39.12 | 45.97 | 14.9 | 1.06 | 7.44 |
| | 40 | 44.79 | 41.04 | 14.2 | 0.61 | 7.64 |
| | 60 | 47.47 | 41.04 | 11.5 | 0.38 | 7.74 |
| | 80 | 49.01 | 40.91 | 10.1 | 0.28 | 7.82 |
| Site B Silt loam | 5 | 15.31 | 63.7 | 21 | 1.44 | 7.36 |
| | 20 | 14.36 | 64.6 | 20.9 | 1.4 | 7.36 |
| | 40 | 12.83 | 68.2 | 19 | 0.97 | 7.45 |
| | 60 | 12.36 | 69.93 | 17.7 | 0.6 | 7.52 |
| | 80 | 12.45 | 70.19 | 17.4 | 0.42 | 7.56 |

raw imagery and cloud masks a level of per-sample quality assurance that would have required custom scripting in GEE without adding methodological value.

For each soil sampling date, since the Sentinel-2 A imagery was not available at the same frequency of the meteorological data, NDVI and NDMI images were selected either on the same day or the closest available cloud-free date within ± 3 days and when a cloud-free image was not available close to the sampling date, the nearest available cloud-free acquisition was used (Appendix 1 and 2). Cloud masking was applied using the Sentinel-2 QA bands to ensure only clear-sky pixels were used. To ensure spatial correspondence between ground sampling points and satellite observations, each sampling location was recorded using a GPS receiver (± 1 m). Sentinel-2 vegetation indices were extracted at the sampling supports using a neighborhood-based approach to reduce geolocation mismatch and mixed-pixel effects. Specifically, sampling points were placed within visually homogeneous interior field areas (away from field boundaries), and NDVI/NDMI values were extracted as the means of a 3 × 3-pixel window centered on each sampling coordinate (NDVI at 10 m; NDMI derived from SWIR bands at 20 m resolution, using the corresponding pixel grid). Local homogeneity was checked by visual

**Table 2** Descriptive statistics (mean, standard deviation, and min–max range) of the gravimetrically measured SMC

| Year | Soil Texture | Depth (cm) | Mean (%) | SD (%) | Min (%) | Max (%) |
|---|---|---|---|---|---|---|
| 2023 | Loam | 5 | 12.80 | 3.08 | 7.3 | 17.56 |
| 2023 | Loam | 20 | 16.19 | 3.51 | 10.13 | 22.01 |
| 2023 | Loam | 40 | 14.66 | 3.76 | 9.89 | 21.36 |
| 2023 | Loam | 60 | 12.61 | 4.39 | 5.49 | 17.65 |
| 2023 | Loam | 80 | 12.68 | 4.94 | 5.49 | 20.77 |
| 2023 | Silt loam | 5 | 14.99 | 2.62 | 9.39 | 19.27 |
| 2023 | Silt loam | 20 | 18.27 | 3.51 | 12.36 | 25.7 |
| 2023 | Silt loam | 40 | 18.03 | 4.11 | 11.36 | 23.04 |
| 2023 | Silt loam | 60 | 16.97 | 7.23 | 9.65 | 27.23 |
| 2023 | Silt loam | 80 | 16.47 | 7.87 | 7.92 | 26.58 |
| 2024 | Loam | 5 | 19.23 | 4.29 | 11.11 | 27.29 |
| 2024 | Loam | 20 | 20.74 | 2.21 | 17.73 | 26.52 |
| 2024 | Loam | 40 | 18.69 | 1.86 | 15.26 | 21.57 |
| 2024 | Loam | 60 | 18.64 | 2.79 | 13.51 | 22.79 |
| 2024 | Loam | 80 | 19.64 | 3.17 | 13.77 | 24.38 |
| 2024 | Silt loam | 5 | 19.50 | 4.04 | 13.64 | 26.1 |
| 2024 | Silt loam | 20 | 22.31 | 3.35 | 17.1 | 27.81 |
| 2024 | Silt loam | 40 | 22.51 | 3.76 | 15.98 | 29.67 |
| 2024 | Silt loam | 60 | 22.08 | 5.08 | 12.41 | 28.87 |
| 2024 | Silt loam | 80 | 22.06 | 4.81 | 14.18 | 29.37 |

inspection of high-resolution basemaps and by verifying low variability of index values within the $3 \times 3$ window.

Two vegetation indices were calculated:

- Normalized Difference Vegetation Index (NDVI) = (NIR – Red) / (NIR + Red).

*(using Sentinel-2 Band 8 and Band 4) (Fig. 1A).*

- Normalized Difference Moisture Index (NDMI) = (NIR – SWIR1) / (NIR + SWIR1).

*(using Band 8 A and Band 11) (Fig. 1B).*

These indices were computed directly in QGIS using raster calculator tools. Critically, NDVI and NDMI values were extracted precisely at the GPS coordinates of each soil sampling location, ensuring that the vegetation indices corresponded spatially with the actual soil moisture measurements. We acknowledge the potential for NDVI saturation during peak vegetative stages but posit that its temporal trajectory still contains valuable information regarding crop water status. This approach allowed the modeling framework to capture localized variations in vegetation condition associated with each sampling point and its underlying soil texture and moisture status, rather than relying on field-averaged values.

### 2.4 Data integration and model development

Prior to modeling, the dataset was cleaned and preprocessed. The continuous features (NDVI, NDMI, weather variables) did not require normalization for tree-based models, but inspect distributions were conducted for anomalies. The XGBoost regressor (Python XGBoost library, v1.4) [29] was developed to predict soil moisture content using the collected features. Separate models were trained for each combination of soil texture and depth – effectively creating 10 distinct models. This approach allows the relationships between inputs and moisture content to differ by depth and soil, acknowledging that surface moisture and deep moisture respond differently to the used features, and that both soil textures have different hydrological behavior. For each soil/depth subset, two modeling scenarios in terms of input features were considered: (a) using only the vegetation indices (VIs) alone as predictors, and (b) using combined meteorological and vegetation indices as predictors. Scenario (a) tests the information content of remote sensing alone, while scenario (b) tests the added value of IoT sensor data. The same XGBoost algorithm was utilized in both scenarios to enable a performance comparison using different feature sets.

The hyperparameters of XGBoost were tuned using grid search and cross-validation. An extensive GridSearchCV was performed for each model, varying parameters such as maximum tree depth (tested values 3, 5, 7, 9), learning rate (0.01, 0.05, 0.1, 0.3), number of trees (100, 300, 500), column subsampling fraction (0.8, 1.0), and row subsampling (0.8, 1.0). The objective function was set to minimize mean squared error (reg: squared-error), and early stopping rounds were used to prevent overfitting during training. Five-fold cross-validation was employed on the training set for hyperparameter selection, with optimization primarily targeting the lowest root mean square error (RMSE). The best hyperparameter set was then used to train the final model for soil/depth scenario. In general, relatively shallow trees (max_depth ~ 5) with a moderate number of boosting rounds (100–300) and a learning rate around 0.1 gave the best results. These settings balance bias and variance, and the use of column/row subsampling in XGBoost inherently guards against overfitting. Due to their relevance in increasing model performance, all meteorological features were included in the combined scenario; none were removed through feature selection.

### 2.5 Model evaluation and SHAP analysis

Data for each soil/depth model were split into 80% for training and 20% for testing, using stratified random sampling such that each sampling date's data was roughly represented in both sets. This way, the test set evaluates model performance on unseen instances across the range of dates, though it is not a true hold-out of entire dates. The model performance was quantified using several standard evaluation metrics on the test set predictions Table 3:

**Table 3** Evaluation metrics used for model performance

| Metric | Equation | Reference |
|---|---|---|
| Mean Squared Error (MSE) | $MSE = {}^{1}/_{n} \sum_{t=1}^{n} \left(y_t - \widehat{y}t\right)^2$ | [30] |
| Root Mean Squared Error (RMSE) | $RMSE = \sqrt{{}^{1}/_{n} \sum_{t=1}^{n} \left(y_t - \widehat{y}t\right)^2}$ | [30] |
| Mean Absolute Error (MAE) | $MAE = {}^{1}/_{n} \sum_{t=1}^{n} \left\lvert y_t - \widehat{y}t\right\rvert$ | [30] |
| Coefficient of Determination ($R^2$)<br>*$R^2 = 1$ indicates a perfect fit, and $R^2 = 0$ indicates the model is no better than using the mean observed value | $R^2 = 1 - \frac{\sum_{t=1}^{n} \left(y_t - \widehat{y}t\right)^2}{\sum_{t=1}^{n} \left(y_t - \bar{y}\right)^2}$ | [31] |
| Nash–Sutcliffe Efficiency (NSE)<br>*NSE = 1 is perfect and NSE = 0 means no skill over the mean. | $NSE = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \widehat{y}_i\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}$ | [32] |

Where: $n$: the total number of observations in the dataset, $t$: The time index or sequence index of an observation, $y_t$: The observed value at time t, $\hat{y}t$: The predicted value by the model at time t, $y_i$ is the observed value at observation *i*; $\hat{y}_i$ is the predicted value at observation *i*; and $\bar{y}$ is the mean of observed values.

To interpret the fitted XGBoost models, Explainable AI – SHAP was applied. SHAP is optimised for tree-based models and calculates the contribution of each feature to the prediction for each sample. SHAP values represent the marginal effect of a feature on the model output, based on Shapley values from game theory [33]. For each depth/soil model, we computed the mean absolute SHAP value for each feature across all test samples. This provides a global ranking of feature importance – higher mean |SHAP| indicates the feature had a larger overall influence on model predictions (either positively or negatively). SHAP dependence plots were examined to see how features affected SMC predictions. All analysis was implemented in Python (using libraries: xgboost 1.4, scikit-learn 1.0, pandas 1.3, numpy 1.21, shap 0.41).

## 3 Results

### 3.1 Overall model performance

The XGBoost models demonstrated high accuracy in predicting SMC, with performance varying according to depth, soil texture, and input scenario.

Figure 3 summarizes the coefficient of determination ($R^2$), RMSE and other evaluation metrics for each soil texture and depth.

Using combined meteorological and vegetation inputs markedly improved accuracy compared to using VIs alone in nearly both soil textures and all depths. For instance, at 5 and 20 cm depths in loam soil, the vegetation-only model obtained $R^2$ of 0.70 and 0.69 with RMSE of 2.26% and 1.77 respectively, whereas integrating meteorological data increased $R^2$ to 0.95 and 0.79 and RMSE to 0.88% and 1.46% at the same depths respectively (Figs. 4 and 5). This is a reduction of >60% in RMSE, highlighting the value of

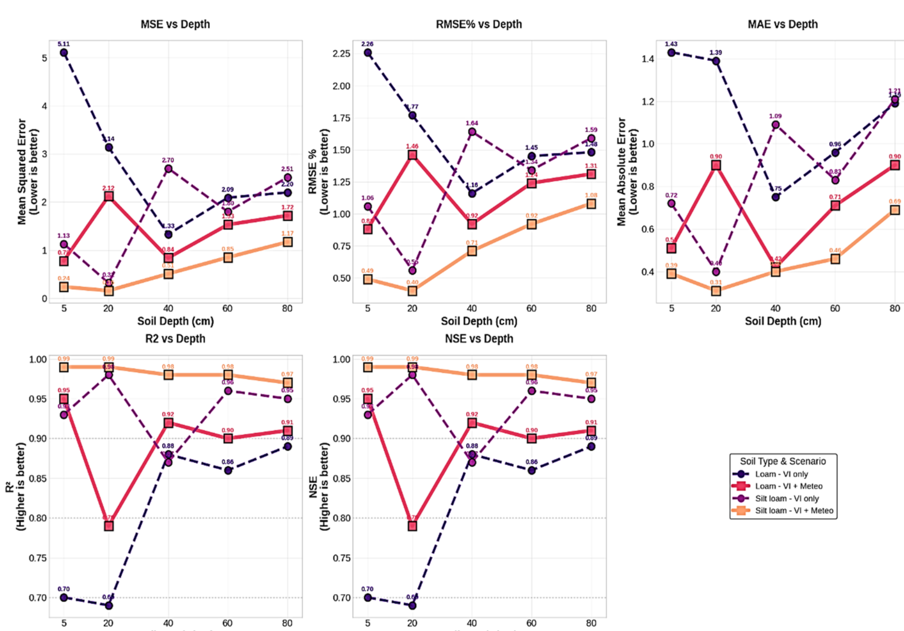



**Fig. 3** The evaluation metrics results of XGBoost model across all depths in two soil textures

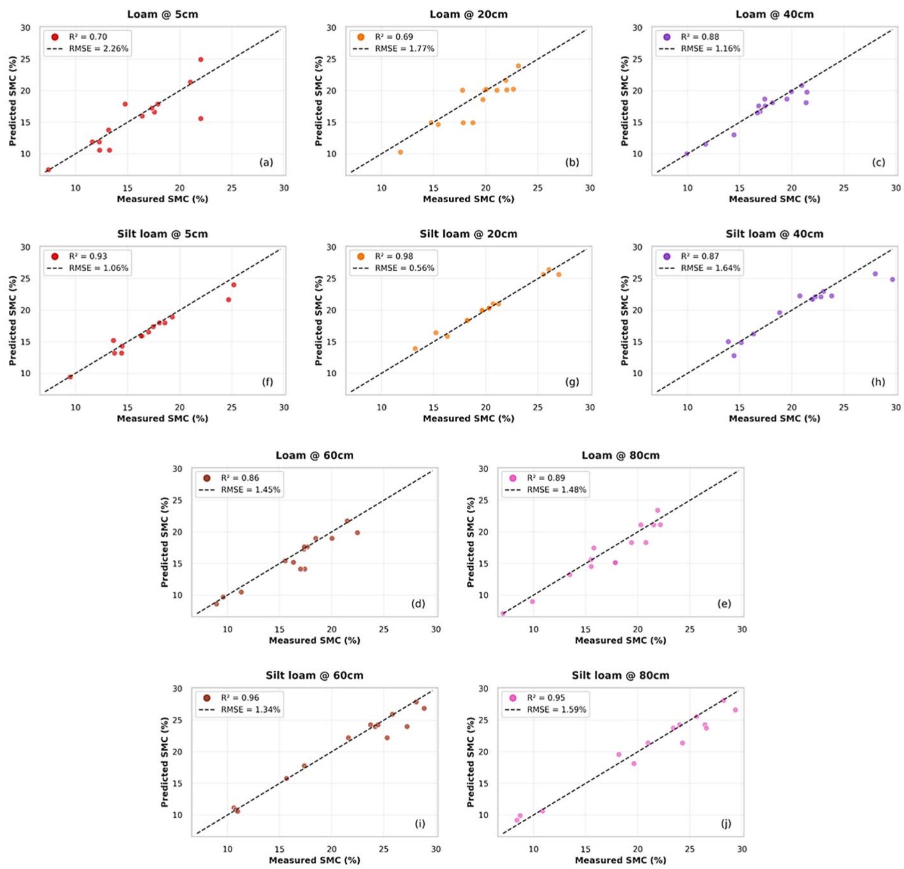



**Fig. 4** Distribution of measured vs. predicted SMC values for the test dataset (representing 20% of total data, $n = 14$ for each soil and depth, with $R^2$ and RMSE across all five depths in two soil textures with VIs only scenario

ground sensor data for topsoil moisture prediction. A similar trend is observed at other depths, where the second scenario consistently outperformed the “VIs only” scenario. This enhancement in predictive performance was most evident in loam soils at shallow depths. In contrast, vegetation-based models for silt loam soils demonstrated strong correlations with SMC even without meteorological inputs, particularly at 5 cm ($R^2$ of 0.93) and 20 cm ($R^2$ of 0.98). This suggests that maize canopy greenness serves as a robust indicator of subsurface moisture in this soil texture. However, the inclusion of weather parameters further improved model performance, achieving near-optimal precision with $R^2$ = 0.99 at 5 cm; $R^2$ = 0.99 at 20 cm. These results indicate that while vegetation indices alone may be adequate for silt loam under certain conditions, integrating meteorological data consistently refines SMC prediction across soil textures, with model fits approaching near-perfect agreement between predicted and observed values. At moderate to deeper depths (40 to 80 cm), performance of the model varied by soil texture. In loam, in the VIs only models performed adequately at 40 cm achieving $R^2$ of 0.88, and RMSE of 1.16%. However, meteorological integration still reduced errors by 21% to get RMSE of 0.92% and higher accuracy with $R^2$ of 0.92, indicating that meteorological features could explain even small moisture variations. Silt loam soil exhibited exceptional VI-only accuracy at 40 cm depth with $R^2$ = 0.87, RMSE = 1.64. The integration of meteorological data further improved the model performance, increasing $R^2$ to 0.98 and reducing the error by approximately 50%, resulting in an RMSE = 0.71%. In the deeper

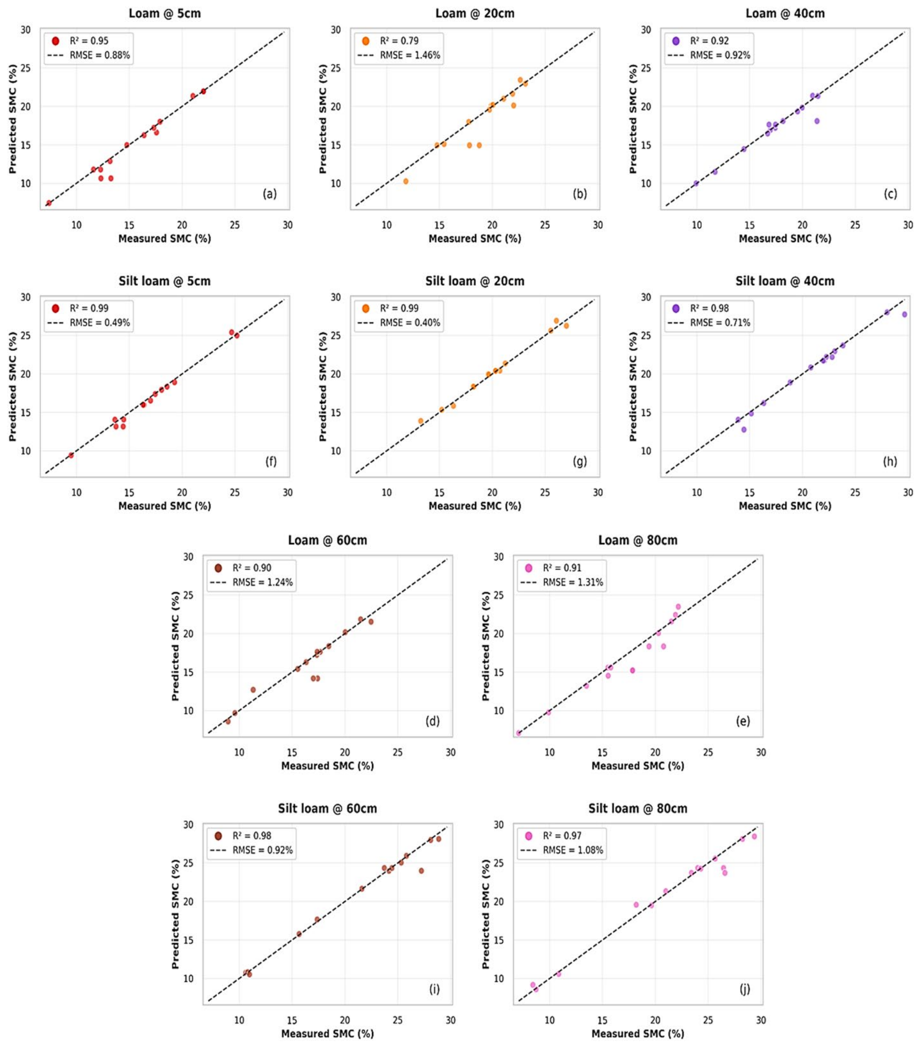


**Fig. 5** Distribution of measured vs. predicted SMC values for the test dataset (representing 20% of total data, $n = 14$ for each soil and depth, with $R^2$ and RMSE across all five depths in two soil textures with second scenario

layers, at 60 and 80 cm in loam soil, the model achieved adequate accuracy under the first scenario with ($R^2$ = 0.86–0.89 and RMSE = 1.45–1.48%, respectively), demonstrating that slower moisture dynamics in deeper layers correspond to cumulative vegetation health patterns. Similarly, in silt loam soil at 60–80 cm depths the model achieved an $R^2$ of 0.96–0.94 and RMSE of 1.34–1.59%, respectively. This reflects its capacity to retain moisture while preserving stable vegetation-moisture relationships. The second scenario significantly enhanced model performance, achieving near-perfect accuracy with $R^2$ = 0.98–0.98 and RMSE = 0.92–1.08% at 60 and 80 cm depths in silt loam soil. In loam soil, the same approach yielded $R^2$ = 0.90–0.91 and RMSE = 1.24–1.31%, highlighting the critical role of combined inputs in improving prediction accuracy.

### 3.2 SHAP analysis results

The SHAP analysis provided insight into which variables were driving the model predictions across soil textures, depths and input scenarios. (Figures 6, 7 and 8) illustrate the mean SHAP values for each feature in both scenarios.

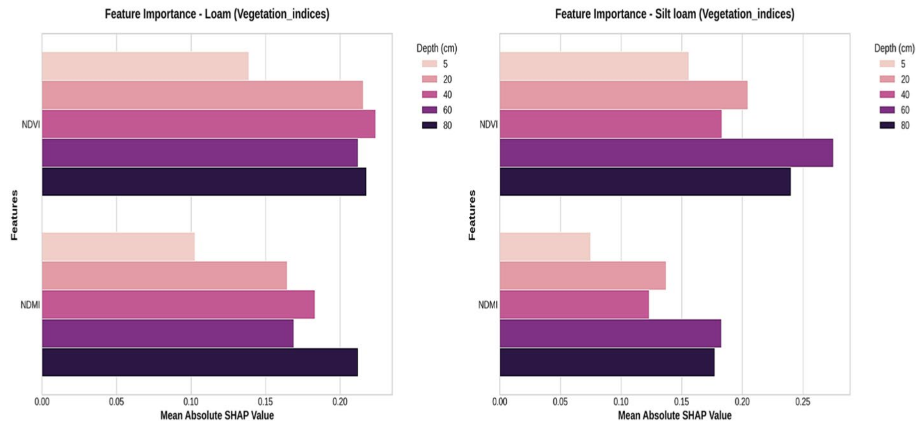



**Fig. 6** Mean SHAP values in loam and silt loam soil across the five depths under vegetation indices only scenarios

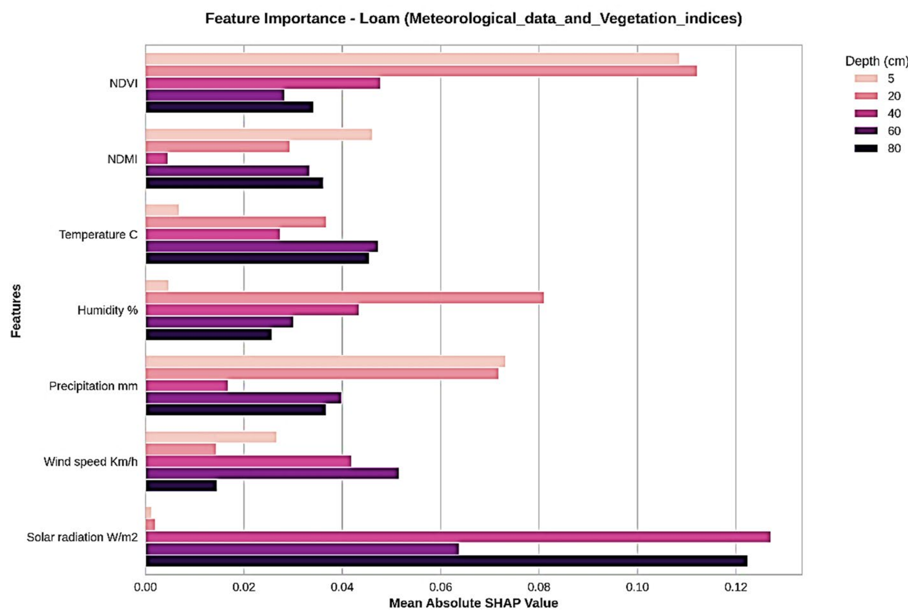



**Fig. 7** Mean SHAP values in loam soil across the five depths under combined scenarios (Meteorological and vegetation indices)

In the VI only scenario, NDVI consistently outperformed NDMI across all depths in both soil textures, with mean SHAP values ranging from 0.14 at 5 cm to 0.22 at 20, 40, and 80 cm depths (Fig. 6). This is likely due to NDVI's sensitivity to canopy vigour, which reflects the cumulative effects of moisture. NDMI, another vegetation index, was generally the second most important feature, although its contribution was typically lower than that of NDVI, with mean SHAP values ranging from 0.10 at 5 cm to 0.21 at 80 cm, where its influence on the model was similar to that of NDVI at this depth. In the silt loam, vegetation indices alone explained most of the variance, particularly at shallower depths, where the model achieved high accuracy with mean SHAP values of 0.16 and 0.2 at 5 and 20 cm depths respectively. This is likely because the fine texture of silt loam stabilizes the relationship between surface moisture and vegetation. In contrast, NDMI contribute less to SMC prediction at all depths with mean SHAP values ranged from 0.07 to 0.18. This confirms that greener, denser crop canopy (high NDVI) strongly

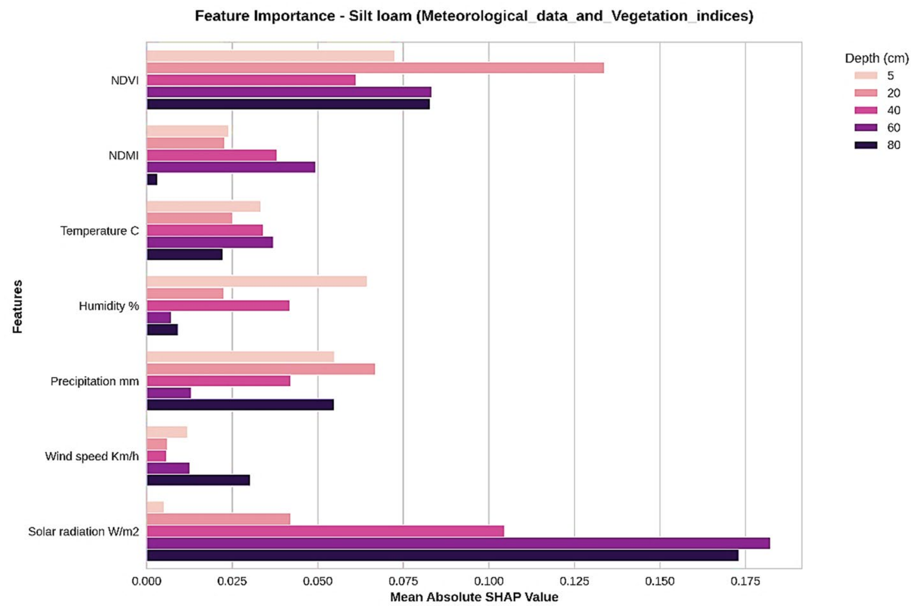



**Fig. 8** Mean SHAP values in silt loam soil across the five depths under combined scenarios (Meteorological and vegetation indices)

signals wetter topsoil conditions, which makes sense agronomically – sufficient soil moisture promotes vigorous growth and delays drought stress appearance. These findings align with other studies where vegetation indices were key predictors of surface soil moisture spatial variability [34].

The integration of meteorological features improved model performance across all depths and soil textures. The SHAP values also highlighted the contribution of these features to the prediction process at different depths. Solar radiation showed a high impact in both soil textures, especially at deeper layers, achieving mean SHAP values of 0.13, 0.06, 0.12 in loam soil and 0.10, 0.18 and 0.17 in silt loam at 40, 60 and 80 cm depths respectively (Figs. 6 and 7). The amount of energy available for evaporation and transpiration, or atmospheric evaporative demand, can be determined by solar radiation [35]. Deeper soil moisture will be lost due to plant transpiration and soil evaporation if high cumulative radiation over days is not replaced by precipitation [36]. On the other hand, precipitation was also among the most important features across depths especially at shallower levels. In loam soil, it had a mean SHAP value of 0.07 at both 5 and 20 cm depths, while in silt soil the values 0.05 and 0.07 at the same depths, respectively. Precipitation also played a notable role at 80 cm depth in both soil textures, with mean values of 0.04 in loam and 0.05 in silt loam. This reflects that recent precipitation—even on the sampling day or shortly before contributed to recharging subsoil moisture [37].

Temperature showed moderate importance at 60 and 80 cm depth in loam soil, with SHAP values of 0.05 at both depths. This is likely because higher temperatures are associated with increased evaporative demand and potentially faster soil drying, particularly given the sand content of the soil [38].

Wind speed showed some importance in the mid-layers of loam soil (40 to 60 cm), with SHAP values of 0.04 and 0.05, respectively. Relative humidity was notably important at 5 cm depth in silt loam (SHAP = 0.06) and at 20 cm depth in loam soil (SHAP = 0.08). This suggests that the silt loam surface lost moisture more quickly in drier air with low humidity, indicating a significant vapor pressure differential [39]. The SHAP analysis identifying NDVI as a key predictor in shallow layers, despite the known lag and indirect nature of the vegetation-soil moisture relationship, suggests that the model successfully learned the cumulative effect of soil water status on canopy development. In silt loam, which has higher water retention, this relationship is more stable and directly observable with mean SHAP values of 0.07 and 0.13, in 5 and 20 cm respectively. In coarser loam soils, where moisture stress manifests more rapidly, the real-time meteorological data was critical to correct for this lag and provide immediate forcing information, hence its greater importance in the model with mean SHAP value of 0.11 at both 5 and 20 cm depths.

## 4 Discussion

### 4.1 Model performance and texture-dependent dynamics

The findings of our study confirm the growing evidence that integrating remote sensing and environmental variables significantly improves the prediction of soil moisture content (SMC) prediction. Liu et al. [40] demonstrated the utility of GA-optimized Random Forest in citrus orchards using vegetation indices and soil and plant analyzer development (SPAD), showing strong depth-specific prediction accuracy. Our work similarly demonstrates the XGBoost model's high accuracy and interpretability, supported by SHAP analysis, offering actionable insights for precision agriculture, enabling effective soil moisture monitoring across depths using satellite imagery and IoT weather sensors. By integrating Sentinel-2 vegetation indices with localized meteorological data, farmers could optimize irrigation scheduling without the need for dense probe networks. For instance, consistently stable subsoil moisture at 60 to 80 cm depths in silt loam soil might allow delayed irrigation despite surface dryness, whereas rapid topsoil moisture depletion in loam soil, caused by its fast drainage, may require farmers to act earlier.

These results emphasized the ability of machine learning models in predicting SMC with high accuracy [41]. The variation in the prediction accuracy between loam and silt loam highlights the texture-dependent moisture-vegetation dynamics. The higher water-retention capacity of silt loam provides a more stable soil-crop moisture relationship, allowing NDVI/NDMI to reliably indicate SMC at all studied depths. The texture-dependent performance differences observed in this study align with those of Zeitoun et al. [42] who reported that loam soils exhibit faster watering-drying cycles due to their coarser texture and lower water retention capacity compared to silt loam. For instance, at a 20 cm depth, the VI-only model achieved an $R^2$ of 0.98 in silt loam, compared with just 0.69 in loam soil. This contrast can be explained by the faster watering-drying dynamics typically associated with the coarser texture of loam soil, where infiltration and drainage can cause rapid changes in near-surface moisture (5–40 cm) [42]. In such conditions, optical vegetation indices may not fully capture short-term soil moisture variation because they represent an indirect and partially lagged canopy response that is also influenced by phenology and other stressors [43]. Consequently, vegetation indices alone are less able to track the day-to-day variability of soil moisture content in loam soil. By integrating meteorological variables, the model considers the direct atmospheric impact, which helps represent the frequency and the contest of soil moisture changes and thereby closes the gap, enabling significantly higher accurate SMC prediction. This integration also allows the model to improve accuracy in silt loam. However, the models' dependence on predictions within observed dates could overestimate performance during extreme weather conditions, especially in loam soils, where meteorological inputs contribute to approximately 30% increase in accuracy at shallower depths (5 and 20 cm). This is consistent with the findings of Chen et al. (2017) [44], who found that interpolation-based soil moisture models produced larger errors when evaluated during extreme drought events outside the model training period.

The predictive performance achieved in this study ($R^2$ up to 0.99, RMSE as low as 0.40%) compares favorably with recent machine learning applications for soil moisture prediction. Emami et al. (2025) [19] developed a backtracking search-optimized XGBoost (BS-XGB) model using meteorological variables and soil temperature, achieving an $R^2$ of 0.973 in testing. However, their study did not integrate satellite vegetation indices, which our SHAP analysis identified as critical for shallow-layer prediction. Zhu et al. (2024) [20] employed UAV multispectral imagery with ensemble learning for kiwifruit orchards, achieving $R^2 = 0.963$ for 0–10 cm depth. Our approach extends such findings by demonstrating that comparable accuracy can be achieved with freely available Sentinel-2 data rather than requiring UAV campaigns, while also providing depth-resolved predictions down to 80 cm.

### 4.2 SHAP analysis and driver interpretability

The SHAP analysis revealed depth-dependent driver patterns that align with physical hydrological principles. NDVI dominance in shallow layers (5–20 cm) reflects the direct coupling between surface soil moisture and vegetation vigor, consistent with Chen et al. (2015) [34] who demonstrated that leaf area index integrates antecedent moisture conditions over weekly timescales. The shift to solar radiation dominance at deeper layers (60–80 cm) can be explained by the cumulative effect of atmospheric evaporative demand on subsoil moisture depletion, a process mediated by plant transpiration and deep root water uptake [35, 36]. This depth-dependent driver switching has practical implications: irrigation decisions for shallow-rooted crops should prioritize vegetation status, while deeper moisture management requires attention to radiation forecasts and cumulative evaporative demand.

### 4.3 Limitations and methodological considerations

A key limitation of the present work is the reliance on optical Sentinel-2 indices (NDVI/NDMI), which provide an indirect proxy for soil moisture. This reliance can be constrained by cloud cover, potential index saturation during peak canopy development, and revisit/processing latency, which together may reduce real-time applicability. In addition, we did not incorporate Sentinel-1 SAR observations, which are more directly sensitive to near-surface soil moisture and could improve robustness during cloudy periods; however, reliable SAR-based retrieval in maize typically requires additional corrections for canopy and roughness effects that were beyond the scope of this study.

Future work should therefore focus on (i) multisensor fusion (Sentinel-1/Sentinel-2/IoT) and gap-filling strategies (e.g., temporal interpolation or data assimilation) to strengthen resilience under data gaps and rapidly changing conditions; (ii) stricter temporal validation (e.g., hold-out by season or date blocks) to assess generalization to unseen weather extremes; and (iii) improving spatial representativeness by expanding sampling across within-field variability and by adopting stratified or Latin-hypercube sampling designs across texture/topography/vegetation gradients. Model performance may also benefit from an expanded covariate set (additional optical indices, terrain attributes, and soil properties) and from denser depth spacing in the upper soil profile to better resolve near-surface moisture gradients. Overall, these extensions would support broader operational transferability while maintaining the interpretability advantages of the proposed explainable XGBoost framework.

## 5 Conclusions

This study demonstrates the efficacy of explainable XGBoost models in predicting soil moisture content (SMC) across studied soil textures and depths utilizing Sentinel-2 vegetation indices with IoT meteorological data. The combined scenario achieved exceptional higher accuracy, with $R^2$ values up to 0.99 in slit loam and RMSE reductions exceeding 50% in loam soils, highlighting the significance of multi-source data integration for depth-resolved prediction. Explainability analysis (SHAP) showed that NDVI was the most important feature for predicting shallow-layer (5–20 cm) SMC, whereas Solar radiation and temperature were the most important features in deeper layers (60–80 cm), most likely due to their relationship with atmospheric evaporative demand and soil heat dynamics. Precipitation and relative humidity also had depth-specific impacts in both loam and silt loam, depending on soil texture.

Practically, this framework supports depth-specific irrigation decision making without requiring dense in-field moisture probe network. This approach is validated for two tested soil textures and relies on optical Sentinel-2 indices. Future research should prioritize temporal validation and integrate pixel-level and multi sensor extensions to improve robustness under cloud cover and extreme conditions, by balancing innovation with practical implementation, this approach has the potential to transform agricultural water management and support global efforts toward sustainable resource use and climate adaptation aligned with SDGs 6 and 13.

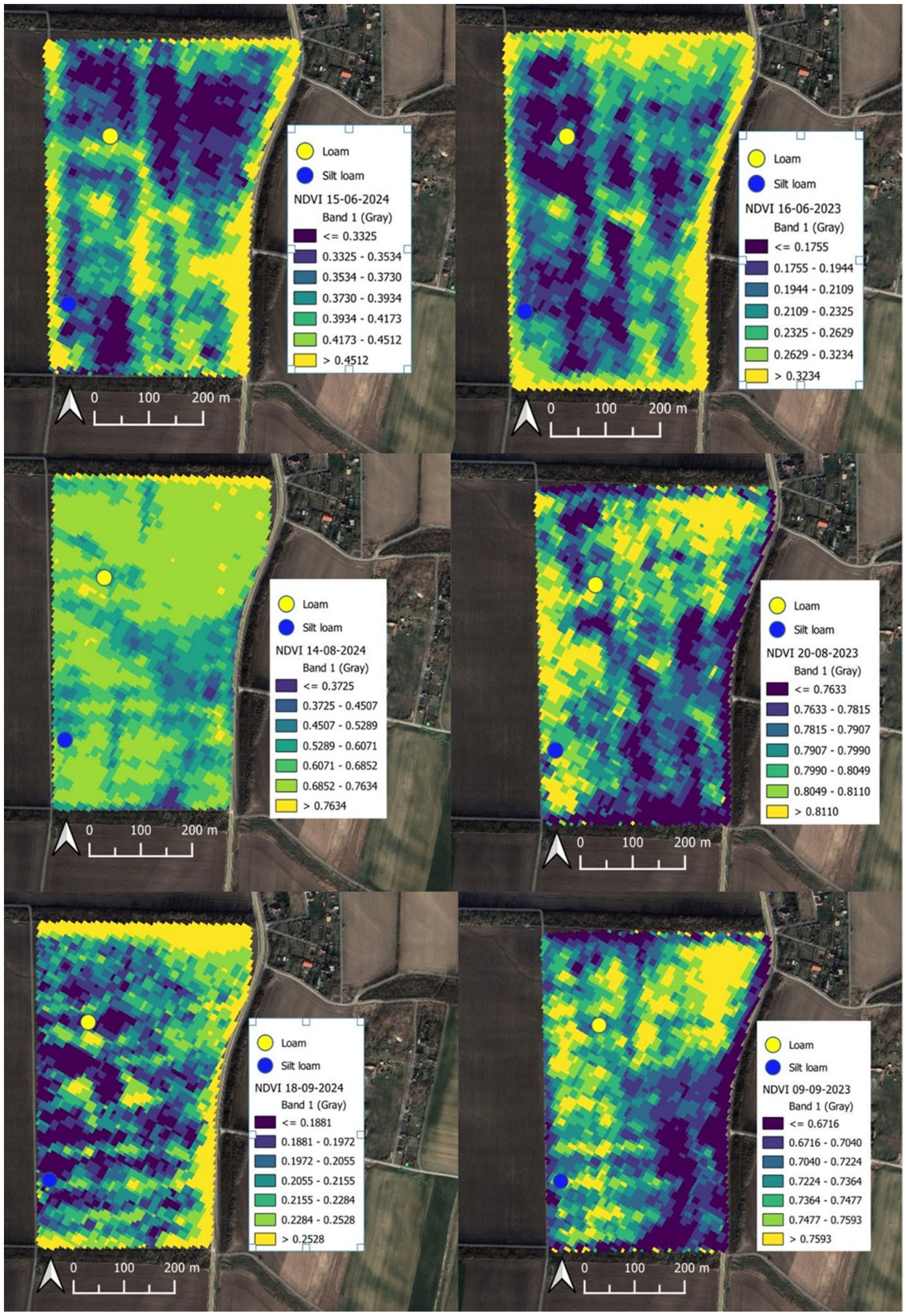
## Appendices

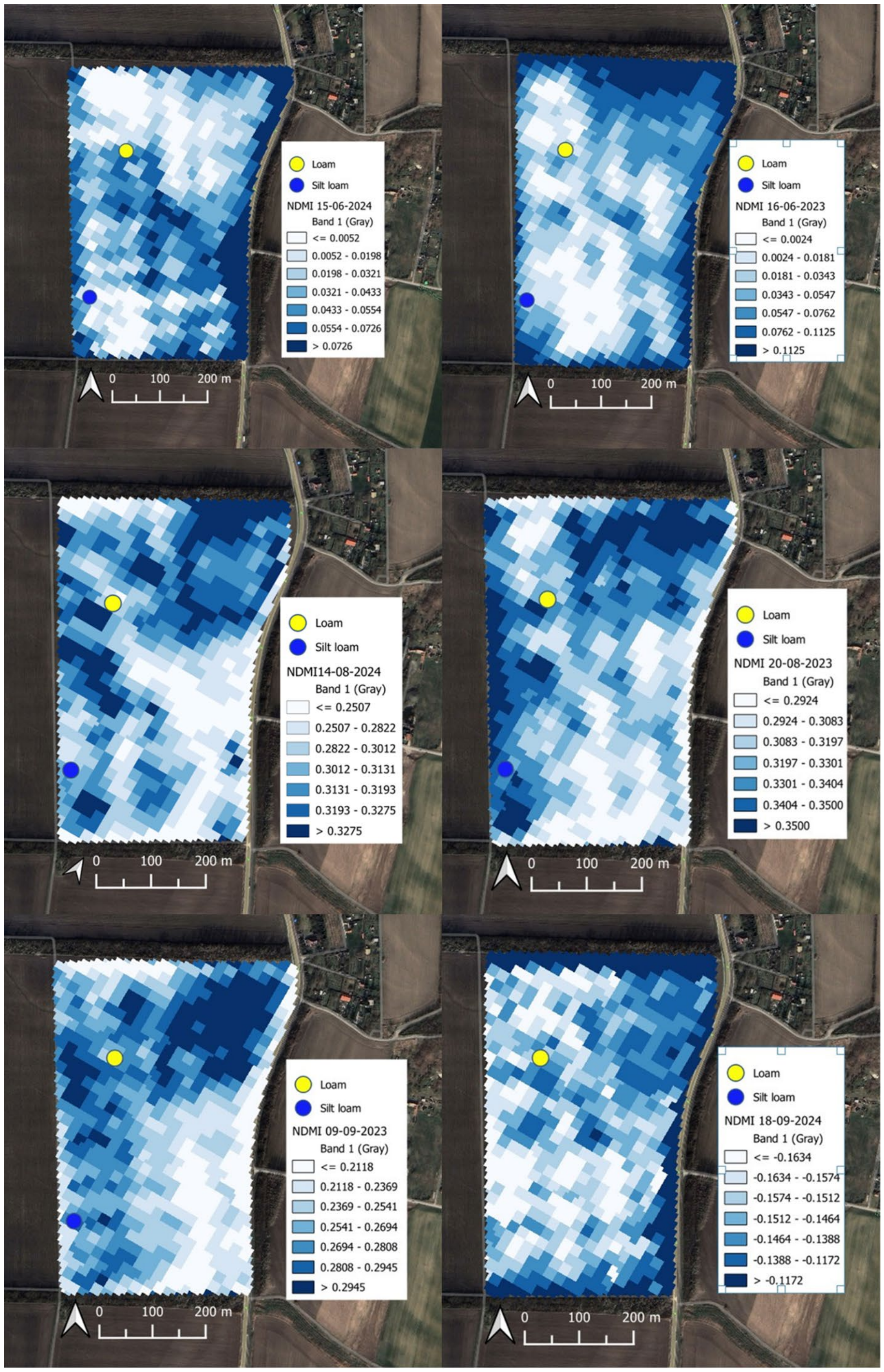
Appendix 1: The Sentinel-2 acquisition dates used to compute NDVI and NDMI for each soil sampling event.

| Season | SMC date | Satellite Date | Soil Texture | NDVI mean | NDVI std | NDMI mean | NDMI std |
|---|---|---|---|---|---|---|---|
| 2023 | 2023-06-14 | 2023-06-16 | loam | 0.193 | 0.013 | -0.077 | 0.007 |
| 2023 | 2023-06-14 | 2023-06-16 | silt loam | 0.215 | 0.013 | -0.095 | 0.004 |
| 2023 | 2023-06-21 | 2023-06-21 | loam | 0.299 | 0.027 | 0.023 | 0.011 |
| 2023 | 2023-06-21 | 2023-06-21 | silt loam | 0.306 | 0.027 | 0.029 | 0.014 |
| 2023 | 2023-06-28 | 2023-06-26 | loam | 0.485 | 0.039 | 0.075 | 0.017 |
| 2023 | 2023-06-28 | 2023-06-26 | silt loam | 0.515 | 0.032 | 0.104 | 0.002 |
| 2023 | 2023-07-19 | 2023-07-16 | loam | 0.764 | 0.019 | 0.311 | 0.013 |
| 2023 | 2023-07-19 | 2023-07-16 | silt loam | 0.783 | 0.003 | 0.329 | 0.013 |
| 2023 | 2023-08-29 | 2023-08-20 | loam | 0.793 | 0.007 | 0.329 | 0.009 |
| 2023 | 2023-08-29 | 2023-08-20 | silt loam | 0.809 | 0 | 0.345 | 0.008 |
| 2023 | 2023-09-08 | 2023-09-09 | loam | 0.722 | 0.003 | 0.271 | 0.006 |
| 2023 | 2023-09-08 | 2023-09-09 | silt loam | 0.75 | 0.004 | 0.269 | 0.011 |
| 2023 | 2023-09-28 | 2023-09-29 | loam | 0.512 | 0.026 | 0.043 | 0.018 |
| 2023 | 2023-09-28 | 2023-09-29 | silt loam | 0.581 | 0.003 | 0.034 | 0.005 |
| 2023 | 2023-10-06 | 2023-10-14 | loam | 0.31 | 0.015 | -0.154 | 0.016 |
| 2023 | 2023-10-06 | 2023-10-14 | silt loam | 0.335 | 0.007 | -0.164 | 0.018 |

| Season | SMC date | Satellite Date | Soil Texture | NDVI mean | NDVI std | NDMI mean | NDMI std |
|---|---|---|---|---|---|---|---|
| 2023 | 2023-10-16 | 2023-10-14 | loam | 0.31 | 0.015 | | |
| 2023 | 2023-10-16 | 2023-10-14 | silt loam | 0.335 | 0.007 | | |
| 2024 | 2024-05-14 | 2024-05-01 | loam | 0.141 | 0.006 | -0.142 | 0.005 |
| 2024 | 2024-05-14 | 2024-05-01 | silt loam | 0.147 | 0.005 | -0.137 | 0.003 |
| 2024 | 2024-05-21 | 2024-05-01 | loam | 0.141 | 0.006 | -0.142 | 0.005 |
| 2024 | 2024-05-21 | 2024-05-01 | silt loam | 0.147 | 0.005 | -0.137 | 0.003 |
| 2024 | 2024-06-06 | 2024-06-15 | loam | 0.39 | 0.008 | 0.055 | 0.015 |
| 2024 | 2024-06-06 | 2024-06-15 | silt loam | 0.356 | 0.009 | -0.003 | 0.017 |
| 2024 | 2024-06-13 | 2024-06-15 | loam | 0.39 | 0.008 | 0.055 | 0.015 |
| 2024 | 2024-06-13 | 2024-06-15 | silt loam | 0.356 | 0.009 | -0.003 | 0.017 |
| 2024 | 2024-06-19 | 2024-06-25 | loam | 0.669 | 0.005 | 0.3 | 0.018 |
| 2024 | 2024-06-19 | 2024-06-25 | silt loam | 0.526 | 0.002 | 0.258 | 0.012 |
| 2024 | 2024-07-03 | 2024-07-10 | loam | 0.794 | 0 | 0.369 | 0 |
| 2024 | 2024-07-03 | 2024-07-10 | silt loam | 0.787 | 0 | 0.33 | 0 |
| 2024 | 2024-07-10 | 2024-07-10 | loam | 0.794 | 0 | 0.369 | 0 |
| 2024 | 2024-07-10 | 2024-07-10 | silt loam | 0.787 | 0 | 0.33 | 0 |
| 2024 | 2024-07-16 | 2024-07-10 | loam | 0.794 | 0 | 0.369 | 0 |
| 2024 | 2024-07-16 | 2024-07-10 | silt loam | 0.787 | 0 | 0.33 | 0 |
| 2024 | 2024-07-23 | 2024-07-30 | loam | 0.81 | 0.001 | 0.357 | 0 |
| 2024 | 2024-07-23 | 2024-07-30 | silt loam | 0.798 | 0.002 | 0.355 | 0.006 |
| 2024 | 2024-07-30 | 2024-07-30 | loam | 0.81 | 0.001 | 0.357 | 0 |
| 2024 | 2024-07-30 | 2024-07-30 | silt loam | 0.798 | 0.002 | 0.355 | 0.006 |
| 2024 | 2024-08-21 | 2024-08-24 | loam | 0.709 | 0.012 | 0.302 | 0 |
| 2024 | 2024-08-21 | 2024-08-24 | silt loam | 0.703 | 0.017 | 0.306 | 0.001 |
| 2024 | 2024-08-28 | 2024-08-29 | loam | 0.564 | 0.022 | 0.121 | 0 |
| 2024 | 2024-08-28 | 2024-08-29 | silt loam | 0.543 | 0.031 | 0.104 | 0.003 |
| 2024 | 2024-09-19 | 2024-09-18 | loam | 0.217 | 0.013 | -0.148 | 0 |
| 2024 | 2024-09-19 | 2024-09-18 | silt loam | 0.211 | 0.008 | -0.147 | 0.002 |
| 2024 | 2024-09-30 | 2024-09-23 | loam | 0.192 | 0.003 | -0.188 | 0 |
| 2024 | 2024-09-30 | 2024-09-23 | silt loam | 0.192 | 0.009 | -0.187 | 0 |

Appendix 2: Sentinel-2 A Images from different dates through 2023 and 2024 seasons.






**Acknowledgements**
This research was supported by the EKÖP-25-3-I-SZE-23 University Research Fellowship Program of the Ministry for Culture and Innovation, funded by the National Research, Development and Innovation Fund; by the "Precision Bioengineering Research Group" supported by the "Széchenyi István University Foundation"; and by the János Bolyai Research Scholarship (Bo/00578/24) of the Hungarian Academy of Sciences3.

**Author contributions**
T.A.: conceptualization, data curation, methodology, writing original draft; A.N.: conceptualization, supervision, review and editing; M.N.: supervision.

**Funding**
Open access funding provided by Széchenyi István University (SZE). This study did not receive any funding from private or public organizations.

**Data availability**

The datasets used during the current study are available from the corresponding author on reasonable request.

## Declarations

**Ethical approval**

Not applicable. This article does not contain any studies with human participants or animals performed by any of the authors.

**Consent to participate**

Not applicable.

**Consent to publish**

Not applicable. All authors have read and approved the final manuscript and consent to its publication.

**Competing interests**

The authors declare no competing interests.

**Clinical trial number**

Not applicable.

Received: 14 November 2025 / Accepted: 13 April 2026

Published online: 22 April 2026

**References**

1. Zheng W, Zheng K, Gao L, et al. GRU–Transformer: A Novel Hybrid Model for Predicting Soil Moisture Content in Root Zones. Agronomy. 2024;14:432. https://doi.org/10.3390/agronomy14030432.
2. Milics G, Kovács AJ, Pörneczi A, et al. Soil moisture distribution mapping in topsoil and its effect on maize yield. Biol (Bratisl). 2017;72:847–53. https://doi.org/10.1515/biolog-2017-0100.
3. Thober S, Kumar R, Sheffield J, et al. Seasonal Soil Moisture Drought Prediction over Europe Using the North American Multi-Model Ensemble (NMME). J Hydrometeorol. 2015;16:2329–44. https://doi.org/10.1175/JHM-D-15-0053.1.
4. Cheng M, Li B, Jiao X, et al. Using multimodal remote sensing data to estimate regional-scale soil moisture content: A case study of Beijing, China. Agric Water Manag. 2022;260:107298. https://doi.org/10.1016/j.agwat.2021.107298.
5. Srivastava A, Yetemen O, Kumari N, Saco PM. (2018) Role of Solar Radiation and Topography on Soil Moisture Variations in Semiarid Aspect-Controlled Ecosystems. sat 1.
6. Yinglan A, Jiang X, Wang Y, et al. Study on spatio-temporal simulation and prediction of regional deep soil moisture using machine learning. J Contam Hydrol. 2023;258:104235. https://doi.org/10.1016/j.jconhyd.2023.104235.
7. Babaeian E, Sadeghi M, Jones SB, et al. Ground, Proximal, and Satellite Remote Sensing of Soil Moisture. Rev Geophys. 2019;57:530–616. https://doi.org/10.1029/2018RG000618.
8. Li Z-L, Leng P, Zhou C, et al. Soil moisture retrieval from remote sensing measurements: Current knowledge and directions for the future. Earth Sci Rev. 2021;218:103673. https://doi.org/10.1016/j.earscirev.2021.103673.
9. Babaeian E, Sadeghi M, Jones SB, et al. Ground, Proximal, and Satellite Remote Sensing of Soil Moisture. Rev Geophys. 2019;57:530–616. https://doi.org/10.1029/2018RG000618.
10. Pôças I, Calera A, Campos I, Cunha M. Remote sensing for estimating and mapping single and basal crop coefficientes: A review on spectral vegetation indices approaches. Agric Water Manag. 2020;233:106081. https://doi.org/10.1016/j.agwat.2020.106081.
11. Bulut Ü, Mohammadi B, Duan Z. Estimation of surface soil moisture from Sentinel-1 synthetic aperture radar imagery using machine learning method. Remote Sens Appl. 2024;36:101369. https://doi.org/10.1016/j.rsase.2024.101369.
12. Rahmati M, Balenzano A, Bechtold M, et al. Soil moisture retrieval from Sentinel-1: Lessons learned after more than a decade in orbit. Remote Sens Environ. 2026;333:115146. https://doi.org/10.1016/j.rse.2025.115146.
13. Davitt A, Winter JM, McDonald K. Integrated crop growth and radiometric modeling to support Sentinel synthetic aperture radar observations of agricultural fields. J Appl Remote Sens. 2020;14. https://doi.org/10.1117/1.JRS.14.044508.
14. Alahmad T, Neményi M, Nyéki A. Applying IoT Sensors and Big Data to Improve Precision Crop Production: A Review. Agronomy. 2023;13:2603. https://doi.org/10.3390/agronomy13102603.
15. Alahmad T, Neményi M, Nyéki A. Soil Moisture Content Prediction Using Gradient Boosting Regressor (GBR) Model: Soil-Specific Modeling with Five Depths. Appl Sci. 2025;15:5889. https://doi.org/10.3390/app15115889.
16. Nguyen TT, Ngo HH, Guo W, et al. A low-cost approach for soil moisture prediction using multi-sensor data and machine learning algorithm. Sci Total Environ. 2022;833:155066. https://doi.org/10.1016/j.scitotenv.2022.155066.
17. Ågren AM, Larson J, Paul SS, et al. Use of multiple LIDAR-derived digital terrain indices and machine learning for high-resolution national-scale soil moisture mapping of the Swedish forest landscape. Geoderma. 2021;404:115280. https://doi.org/10.1016/j.geoderma.2021.115280.
18. Ren Y, Ling F, Wang Y. Research on Provincial-Level Soil Moisture Prediction Based on Extreme Gradient Boosting Model. Agriculture. 2023;13:927. https://doi.org/10.3390/agriculture13050927.
19. Emami H, Emami S, Rezaverdinejad V. A backtracking search-based extreme gradient boosting algorithm for soil moisture prediction using meteorological variables. Earth Sci Inf. 2025;18:181. https://doi.org/10.1007/s12145-024-01674-z.
20. Zhu S, Cui N, Guo L, et al. Enhancing precision of root-zone soil moisture content prediction in a kiwifruit orchard using UAV multi-spectral image features and ensemble learning. Comput Electron Agric. 2024;221:108943. https://doi.org/10.1016/j.compag.2024.108943.
21. United Nation. (2024) UN-Water Annual Report 2023.
22. UN. (2023) UN 2023 Water Conference. New York.
23. USDA. ND. (2020) Natural Resources Conservation Service. Soil classification. Retrieved October 17, 2020.

24. Wang F, Zhang P, Chen S, et al. Spatial Prediction of Soil Texture in Low-Relief Agricultural Areas Using Rice and Wheat Growth Information with Spatiotemporal Stability. Remote Sens (Basel). 2025;17:1865. https://doi.org/10.3390/rs17111865.
25. Schönauer M, Asabere SB, Sauer D, Drollinger S. Topographic indices and ERA5-Land data to describe soil moisture variability in a Central European beech forest. J Hydrol Reg Stud. 2025;59:102456. https://doi.org/10.1016/j.ejrh.2025.102456.
26. Gao Y, Xie Y, Jiang H, et al. Soil water status and root distribution across the rooting zone in maize with plastic film mulching. Field Crops Res. 2014;156:40–7. https://doi.org/10.1016/j.fcr.2013.10.016.
27. Sentinel H. (2024) https://www.sentinel-hub.com
28. QGIS. (2024) QGIS.
29. Chen T, Guestrin C. (2016) XGBoost. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, USA, pp 785–794.
30. Willmott CJ, Ackleson SG, Davis RE, et al. Statistics for the evaluation and comparison of models. J Geophys Res Oceans. 1985;90:8995–9005. https://doi.org/10.1029/JC090iC05p08995.
31. Wright S. Correlation and causation. J Agric Res. 1921;20:557.
32. Nash JE, Sutcliffe JV. River flow forecasting through conceptual models part I — A discussion of principles. J Hydrol (Amst). 1970;10:282–90. https://doi.org/10.1016/0022-1694(70)90255-6.
33. Ergün S, EXPLAINING XGBOOST PREDICTIONS WITH SHAP VALUE: A COMPREHENSIVE GUIDE TO INTERPRETING DECISION TREE-BASED MODELS. New Trends Comput Sci. 2023;1:19–31. https://doi.org/10.3846/ntcs.2023.17901.
34. Chen M, Willgoose GR, Saco PM. Investigating the impact of leaf area index temporal variability on soil moisture predictions using remote sensing vegetation data. J Hydrol (Amst). 2015;522:274–84. https://doi.org/10.1016/j.jhydrol.2014.12.027.
35. Azorin-Molina C, Vicente-Serrano SM, Sanchez-Lorenzo A, et al. Atmospheric evaporative demand observations, estimates and driving factors in Spain (1961–2011). J Hydrol (Amst). 2015;523:262–77. https://doi.org/10.1016/j.jhydrol.2015.01.046.
36. Mu Y, Yuan Y, Jia X, et al. Hydrological losses and soil moisture carryover affected the relationship between evapotranspiration and rainfall in a temperate semiarid shrubland. Agric Meteorol. 2022;315:108831. https://doi.org/10.1016/j.agrformet.2022.108831.
37. Evaristo J, Jasechko S, McDonnell JJ. Global separation of plant transpiration from groundwater and streamflow. Nature. 2015;525:91–4. https://doi.org/10.1038/nature14983.
38. Zhai X, Liu Y, Hong Y, et al. Improved digital mapping of soil texture using the kernel temperature–vegetation dryness index and adaptive boosting. Ecol Inf. 2025;87:103083. https://doi.org/10.1016/j.ecoinf.2025.103083.
39. Nakanishi M, Onda Y, Takahashi J, et al. Changes in air dose rates due to soil moisture content in the Fukushima prefecture forests. Environ Pollut. 2023;334:122147. https://doi.org/10.1016/j.envpol.2023.122147.
40. Liu Q, Wu Z, Cui N, et al. Soil moisture content estimation of drip-irrigated citrus orchard based on UAV images and machine learning algorithm in Southwest China. Agric Water Manag. 2024;303:109069. https://doi.org/10.1016/j.agwat.2024.109069.
41. Alahmad T, Neményi M, Nyéki A. Soil Moisture Content Prediction in Loam Soil with RFR Model. Acta Agron Óváriensis. 2024;65:43–56. https://doi.org/10.17108/ActAgrOvar.2024.65.2.43.
42. Zeitoun R, Vandergeest M, Vasava HB, et al. In-Situ Estimation of Soil Water Retention Curve in Silt Loam and Loamy Sand Soils at Different Soil Depths. Sensors. 2021;21:447. https://doi.org/10.3390/s21020447.
43. Wang X, Xie H, Guan H, Zhou X. Different responses of MODIS-derived NDVI to root-zone soil moisture in semi-arid and humid regions. J Hydrol (Amst). 2007;340:12–24. https://doi.org/10.1016/j.jhydrol.2007.03.022.
44. Chen H, Fan L, Wu W, Liu H-B. Comparison of spatial interpolation methods for soil moisture and its application for monitoring drought. Environ Monit Assess. 2017;189:525. https://doi.org/10.1007/s10661-017-6244-4.


## Publisher's Note