**SHORT DOCTORAL (PhD) DISSERTATION**

**TAREK ALAHMAD**

**MOSONMAGYARÓVÁR**

**2026**

**SZÉCHENYI ISTVÁN UNIVERSITY**
**ALBERT KÁZMÉR FACULTY OF AGRICULTURAL AND FOOD SCIENCES IN MOSONMAGYARÓVÁR**
**DEPARTMENT OF BIOENGINEERING AND PRECISION TECHNOLOGY**
**WITTMANN ANTAL MULTIDISCIPLINARY DOCTORAL SCHOOL OF PLANT, ANIMAL AND FOOD SCIENCES**

**HABERLANDT GOTTLIEB DOCTORAL PROGRAM FOR PLANT SCIENCE**

**DOCTORAL SCHOOL LEADER**
**PROF. DR. LÁSZLÓ VARGA, DSC**

**PROGRAM LEADER:**
**PROF. DR. GYULA PINKE DSC**

**SUPERVISOR(S):**

Dr. habil. Anikó Nyéki, Associate Professor

Prof. Dr. Miklós Neményi, Professor Emeritus, MHAS

Albert Kázmér Faculty of Agricultural and Food Sciences in Mosonmagyaróvár, Department of Bioengineering and Precision Technology.

**Real-Time Soil Moisture Prediction and Analysis for Precision Crop Production**

SUBMITTED BY

**TAREK ALAHMAD**

MOSONMAGYARÓVÁR

2026

Real-Time Soil Moisture Prediction and Analysis for Precision Crop Production

Written by

**TAREK ALAHMAD**

**Made in the framework of**

**Széchenyi István University**

**Albert Kázmér Faculty of Agricultural and Food Sciences in Mosonmagyaróvár**

**Wittmann Antal Multidisciplinary Doctoral School of Plant, Animal and Food Sciences**

**Haberlandt Gottlieb Doctoral Program for Plant Science**

**Supervisor(s):**

1. **Dr. habil. Anikó Nyéki, Associate Professor**
2. **Prof. Dr. Miklós Neményi, Professor Emeritus, MHAS**

**Albert Kázmér Faculty of Agricultural and Food Sciences in Mosonmagyaróvár**

**I recommend for acceptance (yes / no)**

**(signature)**

**The candidate reached ………% during the Complex (Comprehensive) Exam,**

**Mosonmagyaróvár, 2026 ……………**

**……..........…………………………**

**Chair of Complex Exam Committee**

**I recommend the dissertation to be accepted as a reviewer (yes / no):**

**First reviewer (Dr. ………………………………) yes / no**

**(signature)**

**Second reviewer (Dr. ………………………………) yes / no**

**(signature)**

**Maybe a third reviewer (Dr. ………………………………) yes / no**

**(signature)**

**The candidate reached ……………% at the dissertation defense and public debate,**

**Mosonmagyaróvár, ………………………………**

**Chair of Evaluation Committee**

**Doctoral (PhD) qualification: …………………**

**Head of Doctoral Council**

According to Annex 1 of the operation regulation of Wittmann Antal Multidisciplinary Doctoral School. Doctoral candidates have the option to write a short dissertation instead of the traditional PhD dissertation. The short dissertation must be based on first-authored full-length original research papers published or accepted for publication in Q1–Q3-ranked journals listed in the SCImago database (https://www.scimagojr.com/journalrank.php). These articles must collectively total at least 6 points, with at least one paper published (or accepted for publication) in a Q1- or Q2-ranked journal. The point system is as follows: Q1 article: 4 points, Q2 article: 2 points, Q3 article: 1 point. Additionally, PhD candidates must prepare and publish a review paper in a peer-reviewed scientific journal related to their research topic by the end of their first year of training.

Fulfilments of the conditions based on these publications:

- Applied Sciences, 15(11), 5889. (Q2, Impact Factor: 3.7) 2 points.
- Frontiers in Soil Science (Q1, Impact Factor: 3.7) 4 points.
- Discover Applied Sciences (Q2, Impact Factor: 2.8) 2 points.

**Abbreviation:**

| | |
|---|---|
| AI | Artificial Intelligence |
| CBR | CatBoost Regression |
| CNNs | Convolutional Neural Networks |
| CV% | Coefficient of Variation |
| DL | Deep Learning |
| FDR | Frequency-Domain Reflectometry |
| GBR | Gradient Boosting Regressor |
| GIS | Geographic Information System |
| GPS | Global Positioning System |
| IoT | Internet of Things |
| LoRaWAN | Long Range Wide Area Network |
| LSTM | Long-Short Term Memory |
| ML | Machine Learning |
| NDMI | Normalized Difference Moisture Index |
| NDVI | Normalized Difference Vegetation Index |
| OOB | out-of-bag |
| PA | Precision Agriculture |
| RFR | Random Forest Regression |
| SAR | Synthetic Aperture Radar |
| SHAP | SHapley Additive exPlanations |
| SMAP | Soil Moisture Active Passive |
| SMC | Soil Moisture Content |
| SMOS | Soil Moisture and Ocean Salinity |

| | |
|---|---|
| TDR | Time-Domain Reflectometry |
| UAV | Unmanned Aerial Vehicle |
| XAI | Explainable AI |
| XGBoost | eXtreme Gradient Boosting |

**Table of Contents:**

**Abstract:**

Precision agriculture (PA) requires precise and accurate information on soil moisture, optimized irrigation, and sustainable water management. This research focuses on the critical issue of accurate prediction of soil moisture content (SMC) in precision agriculture, specifically on optimizing irrigation and promoting sustainable water management. This research used a combination of gravimetric field measurements, IoT technology-based meteorological measurements, and Sentinel-2 vegetation indices using an explainable machine learning approach based on Random Forest Regression (RFR), Gradient Boosting Regressor (GBR), XGBoost, and Long Short-Term Memory (LSTM) networks, progressing from model evaluation to multi-source data fusion using SHAP interpretability and different feature importance analysis. The study conducted field measurements on a 23-ha maize field in Hungary, featuring different soil textures (loam, sandy loam, and silt loam) and depths (5-80 cm). The results showed texture- and depth-dependent soil moisture content variability, in 2023, soil moisture content in loam soil ranged from 12.61 ± 4.39% to 16.19 ± 3.51%, In the same year, silt loam soil showed higher moisture values, ranging from 14.99 ± 2.62% to 18.27 ± 3.51%. In 2024, soil moisture increased in both soil types: loam soil ranged from 18.64 ± 2.79% to 20.74 ± 2.21%, while silt loam soil recorded the highest values overall, ranging from 19.50 ± 4.04% to 22.51 ± 3.76%.

The current study employed machine learning algorithms, among which Random Forest Regression (RFR) showed the highest consistency in model performance ($R^2$ = 0.82-0.99), followed by

Gradient Boosting Regressor (GBR) algorithms, which showed high accuracy in predicting soil moisture content and identified precipitation and humidity as driving factors in loam, and solar radiation in silt loam. The integration of Sentinel-2 vegetation indices and meteorological measurements using the XGBoost algorithm showed the highest potential, especially in loam soil at shallow depths, reducing the root mean square error by more than 60%.

Overall, the results show that combining gravimetric measurements, IoT-based meteorological data, and Sentinel-2 vegetation indices substantially improves depth-specific soil moisture prediction and supports more precise irrigation scheduling under the variable climate conditions.

## 1. Introduction:

Precision Agriculture (PA), also known as precision crop production and site-specific crop management, has emerged as a main approach for increasing efficiency and environmental performance of modern farming by intentionally managing within-field variability (Alahmad et al., 2023). In PA, agronomic decisions are guided by spatially and temporally information on soil and crop conditions, allowing targeted intervention that meet crop demand and local site constrains. In this context, PA is commonly framed as an integration of site-adapted cultivation with remote sensing, geographic information system (GIS), mechanization development and information-based decision-support, including the comparison of yield-related data with soil and pest spatial related data to describe field heterogeneity (Nyéki, 2016).

Within the Department of Bioengineering and Precision Technology at Széchenyi István University, PA has been pursued as long term research has combined field experiments, innovative sensing, and machine learning (ML) to refine yield prediction, analyze soil-crop-atmosphere interactions, and facilitate sustainable crop management (Milics, 2008; Nyéki, 2016). The field configuration and its long-term use provide an analytical foundation for studying soil-crop-atmosphere interactions and testing precision approaches (Nyéki, 2016). Nyéki (2016) used precision field trails for maize yield prediction analysis and its sensitivity to soil and climate variables. Geospatial analysis techniques and remote sensing through the application of GIS were also presented as another method for enabling precision crop production, as presented by Milics (2008). Ambrus (2023) took a step forward in the practical application of the enabling of precision agriculture through the development of an autonomous small intelligent data collection and analysis robot that utilized environmental sensing and visual (RGB) data processing using ML techniques for precision agriculture applications. What is considered noteworthy is that the work presented how convolutional neural networks (CNNs) and machine vision techniques can be applied to agricultural monitoring applications such as segmentation-based yield prediction.

The impact of using data-driven techniques for precision agriculture is becoming difficult to ignore. For example, the study presented by Zsebő et al. (2024) which presented comparative Normalized Difference Vegetation Index (NDVI) measurement using handheld sensors and UAV cameras with multispectral cameras, highlighted the capabilities of integrated prediction models as well as the trade-offs between using different sensing technologies for crop monitoring applications. In essence, the application of sensing technologies, data analytics, and IoT technology that combines fixed monitoring stations with mobile sensing technologies has accelerated the development of PA.

Building on these advances, Neményi et al., (2022, 2023) developed IoT architectures for PA by integrating mobile sensing units, fixed monitoring stations, and data platforms. One outcome is the Mosonmagyaróvár Agro-IoT concept for smart, data-intensive agriculture that depends on interoperable sensing networks, digital databases, and AI to capture real-time soil and microclimate parameters. Practical applications, like the Field Monitoring Laboratory using (Long Range Wide Area Network) LoRaWAN connected IoT technology, crops, fuse soil, and microclimate sensing, which specifically include real-time environmental monitoring, like soil moisture measurements (Nyéki et al., 2025). These integrated methods provide a robust basis for data-driven PA. The current dissertation continues the work in the following ways, using AI and ML in real-time prediction of soil moisture, specifically targeting soil moisture content (SMC) as a vital aspect in soil, crops, and atmosphere interactions in precision crop production (Alahmad et al., 2025).

SMC plays a vital role in hydrological and ecological processes, influencing water availability, energy fluxes, and crop yields (Alahmad et al., 2025; Mane et al., 2024). However, traditional SMC measurements are limited in spatial resolution and frequency, while

IoT-based sensors provide high-frequency meteorological measurements and satellite indices offer periodic vegetation data, together enabling operational soil moisture monitoring (Tsai et al., 2024). Integrating of these data streams improves depth-resolved SMC modeling for precision irrigation and drought risk management. In most SMC prediction studies focus on shallow layers (0–20 cm), which is insufficient for irrigation decisions that require insight into root-zone water dynamics (Alahmad et al., 2025; Chen et al., 2025). Therefore, this dissertation advances depth-specific modeling using integrated IoT, gravimetric, and satellite-derived data to capture variability across soil profiles and textures.

### 1.1. Limits of conventional measurement and the need for integrated sensing–modeling approaches

Reliable SMC measurements can be obtained via laboratory oven-drying methods (gravimetric/volumetric) or *in-situ* sensors, but these approaches face limitations when high-frequency, depth-resolved, and spatially representative monitoring is required for PA (Yinglan et al., 2023). Gravimetric methods remain the standard for validation (Little M. and Smith, 1998; Singh et al., 2023). but IoT-based sensors now offer real-time, scalable solutions for environmental monitoring, irrigation planning, and weather tracking (Wu et al., 2023). Despite their advantages, sensor-based irrigation decisions can be complicated by spatial heterogeneity, calibration, and placement issues (Bicamumakuba et al., 2025). Remote sensing, such as Sentinel-2, provides spatially continuous vegetation indices but is limited to near-surface conditions and may be influenced by crop phenology and soil background (Babaeian et al., 2019; Pôças et al., 2020).

Microwave SAR such as Sentinel-1 provides direct sensitivity to near-surface SMC but is affected by crop canopy and surface roughness, which require more advance corrections (Rahmati et al., 2026). Therefore, integrating ground truth SMC, IoT meteorological data, and satellite vegetation indices provides a robust approach for predicting SMC in diverse fields (Babaeian et al., 2019). Sentinel-2 and IoT data are combined in this thesis through feature-level fusion rather than pixel-wide real time mapping to support depth-specific SMC prediction, where sentinel-2 indices NDVI and Normalized Difference Moisture Index (NDMI) serve as canopy state proxies and IoT measurements provide high-frequency meteorological impact.

### 1.2. Why ML for soil moisture prediction in precision crop production.

Soil moisture dynamics don't follow simple rules, the complexity of SMC dynamics driven by nonlinear interactions among meteorological factors, vegetation, and soil properties. However, it is important to capture this complexity in order to effectively manage water resources and alleviate drought situations (Tarolli and Zhao, 2023). Traditional statistical methods often have difficulty handling such complex relationships, whereas machine learning techniques such as Random Forest Regression (RFR), eXtreme Gradient Boosting (XGBoost), Gradient Boosting Regressor (GBR), and deep learning approaches like Long-Short Term Memory (LSTM) are better suited to deal with these interactions and learn from the data (Alahmad et al., 2025; Zheng et al., 2024).

Among the tree-based ensembles, XGBoost has been widely adopted as an effective boosting algorithm has outperformed linear baselines in SMC prediction applications (Ren et al., 2023). Recently, depth-focused research using multi-sources predictors demonstrated the practical significance of these approaches and boosting based models have achieved highest predictive accuracy and identified the physically interpretable factors such as soil temperature, humidity, and solar radiation as influential factors for SMC prediction (Emami et al., 2025; Zhu et al., 2024). Deep learning has also been used for SMC prediction using several predictors such as meteorological data and measured soil moisture values as inputs, highlighting the boarder trend toward data driven forecasting (Li et al., 2020). In this thesis, ML has been used as an enabling tool for depth specific and, soil types of aware prediction.

For precision crop production and irrigation predictive accuracy is highly important but the interpretability of the results in decision making is high essential. Therefore, this dissertation uses explainable AI (XAI) tools, such as SHapley Additive exPlanations (SHAP) within XGBoost, to quantify depth-dependent predictors, translating complex model outputs into actionable strategies for irrigation and crop management (Azorin-Molina et al., 2015). Nonetheless, gaps remain: most studies do not assess predictive performance across the full soil profile (Kheimi and Zounemat-Kermani, 2025) ; integrated, depth-specific frameworks using high-resolution satellite, IoT, and gravimetric data are rare; and many ML models lack interpretability, limiting their decision-support value.

This thesis addresses these gaps by (i) developing depth-specific, soil-aware AI-based ML models to predict SMC using IoT sensor meteorological data, gravimetric SMC measurements and two vegetation indices (NDVI and NDMI); (ii) applying explainable ML and feature importance approaches to identify the depth- dependent drivers; and (iii) translate these findings to enhance decision-making in precision irrigation and crop production.  The main objectives are:

- To build an integrated dataset for depth-specific soil moisture modelling by combining gravimetric SMC, IoT meteorology, and Sentinel-derived vegetation indices (NDVI/NDMI).
- To characterize spatio-temporal and depth-dependent SMC dynamics across three different soil texture and two growing seasons of maize.
- To evaluate and compare ML algorithms (RFR, XGBoost, GBR, and LSTM) across predictor scenarios - vegetation indices only, meteorological variables only, and combined inputs - using a consistent train-test protocol, assessing both depth-specific and soil-texture-specific modeling strategies against pooled approaches.
- To apply SHAP-based explainability and feature importance to quantify depth- and texture-dependent feature contributions, identify dominant environmental drivers of SMC variability, and translate these findings into actionable guidelines for precision irrigation and crop water management.

# 2. Literature Review:

## 2.1. Soil moisture concept relevant to crop production and irrigation planning

SMC is a vital variable in agricultural and hydrological systems, directly influencing crop physiology, nutrient transport, root development, and microbial activity (Bibek and Vivek, 2025). Typically measured gravimetrically (mass-based) or volumetrically, SMC, particularly when interpreted relative to field capacity, permanent wilting point, and other soil-water constants, determines the amount of plant-available water in the root zone and is therefore essential for irrigation planning.

Soil moisture dynamics are fundamentally governed by the field water balance, which describes the gains (precipitation, irrigation) and losses (evapotranspiration, drainage, runoff) of water within the root zone. Among these components, evapotranspiration (ET) is typically the dominant water-loss pathway in agricultural systems, accounting for up to 70–90% of total water inputs in temperate rainfed conditions (Allen et al., 1998; Nagy et al., 2013). Reference evapotranspiration ($ET_0$) estimated using approaches such as the FAO Penman-Monteith equation, integrates the effects of solar radiation, air temperature, relative humidity, and wind speed variables that collectively govern the atmospheric evaporative demand.

The relationship between SMC and crop water availability is intricate and non-linear, depending on soil texture, topography, organic matter, and environmental factors such as precipitation and evapotranspiration (Zheng et al., 2024). For instance, even though the maize roots can reach water at a depth of more than 150 cm, most of the water uptake occurs in the top 60 cm of the soil profile during the vegetative stage (Alahmad et al., 2025; Gao et al., 2014). Hence, it is important to measure soil moisture at different depths in the soil profile, as surface measurements do not provide a complete picture (Alahmad et al., 2025). The temporal and spatial variability in SMC is high because of the interplay between static (soil texture, topography, organic matter) and dynamic factors (precipitation, evapotranspiration, crop development).

### 2.2. Soil Moisture Measurement Approaches: Strengths and Limitations

SMC measurement techniques can be classified into three types: direct laboratory techniques, *in-situ* techniques, and remote sensing techniques. The gravimetric method, which include drying soil samples in an oven at 105°C, is still considered the standard method for achieving the highest accuracy (Dirksen, 1999; Singh et al., 2023). For this reason, many studies have collected gravimetric SMC data at different depths and across various soil types to validate other measurement approaches. *In-situ* techniques, such as capacitance probes, TDR, FDR, and neutron probes, are widely used to monitor soil moisture at multiple depths by measuring dielectric permittivity. Although these tools are essential for precision agriculture, their accuracy depends heavily on proper soil-specific calibration, and their performance can be affected by temperature changes and electromagnetic interference (Songara and Patel, 2022). Additionally, sensor networks provide measurements at specific points, which may not fully capture the spatial variability of soil moisture across an entire field unless sensors are deployed at high density (Alahmad et al., 2023), This limitation highlights the growing need for distributed, small, and smart data logging systems to better represent field-scale conditions.

Remote sensing allow for SMC predicts across a wide range of scales, from individual fields to global coverage. Common approaches include optical and infrared indices such as NDVI and NDMI, as well as microwave-based systems like Soil Moisture Active Passive (SMAP) and Soil Moisture and Ocean Salinity (SMOS). Although NDVI and NDMI offer useful, indirect information about canopy water status, their performance can be affected by cloud cover, surface conditions, and canopy density (Babaeian et al., 2019; Li et al., 2021). Microwave-based methods can provide more direct estimates of soil moisture, but they often operate at coarse spatial resolutions (greater than 9 km) and their signals can weaken in areas with dense vegetation (Babaeian et al., 2019). Synthetic Aperture Radar (SAR) missions, such as Sentinel-1, provide higher spatial resolution and the ability to penetrate cloud cover, yet their accuracy can still be influenced by factors like canopy structure and surface roughness (Rahmati et al., 2026). Despite these challenges, multispectral satellite imagery from platforms such as Sentinel-2 and Landsat continues to deliver reliable and cost-effective information on crop and soil conditions, typically available at regular intervals, often about once per month, making them essential tools for agricultural monitoring (Chen et al., 2015). However, it is important to highlight that vegetation indices derived from these satellites represent indirect, time-integrated signals of previous soil moisture conditions rather than real-time measurements of actual SMC (Alahmad et al., 2025).

### 2.3. ML for Soil Moisture Content Modelling

Conventional process-based and statistical models often have limitations when it comes to fully capturing the complex, multivariate, and non-linear behavior of SMC dynamics (Zheng et al., 2024). In contrast, machine learning approaches, especially tree-based ensemble models and deep learning networks, have shown significant improvements in prediction accuracy, often without the need for detailed physical parameterization (Cai et al., 2019; Nyéki et al., 2021). At the same time, advances in sensor networks now allow for the continuous collection of large volumes of environmental data from multiple sources. This steady flow of high-dimensional data further strengthens the capability of machine learning models to analyze patterns and deliver more reliable soil moisture predictions (Haddon et al., 2023).

A broad spectrum of ML approaches has been applied to SMC prediction, including RFR, Extreme Gradient Boosting Regression (XGBR), CatBoost Regression (CBR), and Long Short-Term Memory (LSTM) architectures (Ågren et al., 2021; Alahmad et al., 2024; Islam et al., 2023). Studies using direct in-situ sensor measurements as model targets have reported the highest ML performance values. For instance, Islam et al. (2023) working with ground-based sensor observations, found that stacking ensemble models combining RFR, XGBoost, CatBoost, and Extra Tree Regressor (ETR) achieved superior accuracy ($R^2 = 0.9982$) compared to blending ($R^2 = 0.9969$) and base models

(RFR $R^2$ = 0.9929). Feature importance analyses underscored the influence of humidity and wind, which varied by soil texture and irrigation regime, reinforcing the need for soil-specific modeling. Complementing these findings, Ren et al. (2023) reported XGBoost achieved an 88% accuracy (correlation coefficient = 0.69) for SMC prediction using meteorological predictors. Emami et al. (2025) further optimized XGBoost with backtracking search (BS-XGB), achieving near-perfect accuracy (training $R^2$ = 0.999, testing $R^2$ = 0.973) across multiple soil depths, with soil temperature, humidity, and solar radiation as key predictors. UAV-based ensemble learning approaches, such as those by Zhu et al. (2024), have also demonstrated high accuracy for depth-specific SMC prediction in orchards across various root-zone depths achieving $R^2 = 0.963 \pm 0.024$ for SMC20–SMC60.

Deep learning networks, including LSTM and Bidirectional LSTM (BLSTM), have shown $R^2$ values from 0.96 to 0.98 and low mean square error (MSE = 0.014–0.056), with LSTM outperforming others for daily reference evapotranspiration and SMC modeling (Alibabaei et al., 2021). Selection of loss functions, such as mean square error, further refined model performance.

### 2.4. Explainable AI for Environmental and Agro-hydrological Models

Although ML algorithms have demonstrated superior predictive results, the lack of transparency associated with such approaches remains a major hindrance to practical application in agriculture (Taheri et al., 2025). This issue has been resolved by incorporating XAI techniques, particularly SHAP, which helps to assess the contribution of input variables to model predictions. This helps to increase transparency in complex models (Hussein et al., 2024).

However, some limitations exist in incorporating soil textures and depths in SMC prediction models (Kheimi and Zounemat-Kermani, 2025). Most studies conducted on SMC prediction have not explored the full potential of variations in soil properties, depth, and meteorological conditions in influencing SMC. Furthermore, most existing studies do not distinguish between model performance achieved using direct gravimetric or in-situ observations versus remotely sensed retrievals or model-based estimates, making cross-study performance comparisons methodologically ambiguous.

For instance, studies have found that soils containing more silt and clay tend to retain more moisture compared to sandy soils. Similarly, an increase in depth results in a corresponding increase in SMC since more moisture remains in the soil due to reduced evaporation loss (Celik et al., 2022).

In addition, while some studies have successfully applied ML algorithms to SMC prediction, only a limited number have combined high-resolution satellite imagery with site-specific, high-frequency IoT-based meteorological data within a depth-resolved framework, particularly for annual crops. This research addresses these gaps by developing an interpretable XGBoost model that integrates vegetation indices derived from Sentinel-2 imagery with real-time IoT-based meteorological observations. The study also evaluates the predictive advantages of combining multiple data sources compared with models that rely only on vegetation indices such as NDVI and NDMI, while using SHAP analysis to provide greater transparency and understanding of the factors driving soil moisture predictions. By delivering transparent, depth-resolved insights, this work advances precision agriculture by harmonizing predictive accuracy with practical, interpretable recommendations aligned with SDGs 6 and 13.

# 3. Materials and methods:

## 3.1. Study area, soil texture classification, properties, and sampling.

Field experiments were conducted during the 2023 and 2024 maize growing seasons at a 23-ha rainfed field belonging to Széchenyi István University near Mosonmagyaróvár, Hungary (47°54'11.8"N, 17°15'08.9"E). The site has a temperate continental climate (mean annual precipitation ~600 mm). The terrain is nearly flat (elevation 133–138 m) with three soil textures according to USDA taxonomy (USDA, 2024): loam, sandy loam, and silt loam (Figure 1). Detailed soil properties are given in (Table 1, Article 1). The study covers (i) one growing season (June–October 2023) for the multi-model comparison (RFR, LSTM, and XGBoost), (ii) one growing season (June–October 2023) for the soil-specific GBR analysis (loam, silt loam), and (iii) two growing seasons (2023–2024, May–October each season) for depth- and soil-specific explainable XGBoost modelling integrating remote

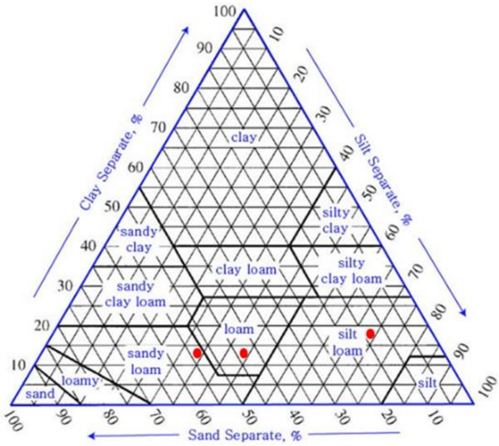



**Figure 1.** Soil textures classes (silt, clay and sand) for each studied soil.

Soil samples were collected at fixed points near IoT sensor stations (within 1 m²) at five depths (5, 20, 40, 60, 80 cm) with three replicates per point. Sampling occurred biweekly in 2023 (9 dates) and on 23 dates across 2023–2024 for the two-season dataset, totalling 1035 samples. Samples were sealed, transported to the laboratory, and oven-dried at 105°C to constant mass (Dirksen, 1999; Shukla et al., 2014) . Gravimetric SMC (%) was calculated as:

$$\boldsymbol{\theta = \frac{Mw}{Md} \times 100} \qquad \textbf{(1)}$$

Where:

- $\theta$ = Soil Moisture Content (%)
- *Mw* = Mass of water (g) = (Wet weight – Dry weight)
- *Md* = Mass of dry soil (g)

The three dissertation articles used different analytical subsets of the full dataset according to their specific objectives. Article 1 used the 2023 dataset including three soil textures, five depths, three replicates, and nine sampling dates, resulting in 405 samples. Article 2 used a soil-specific subset for loam and silt loam during the 2023 season, include in 300 samples. Article 3 used the full two-season dataset for loam and silt loam, include 690 samples.

## 3.2. Data Collection:

### 3.2.1. IoT-based meteorological data

An on-site IoT sensor station (Boreas Ltd., EcoLogger Boreas datalogger unit BEL-06) was installed in the field to continuously monitor meteorological conditions (Figure 2, C). The station continuously monitored: air temperature (°C), relative humidity (%), precipitation (mm), wind speed (km/h), and solar radiation ($W/m^2$). Data were transmitted via LoRaWAN at 10–15 min intervals and synchronized with soil sampling timestamps. These variables collectively represent the primary drivers of reference evapotranspiration ($ET_0$) as defined by the FAO Penman-Monteith approach (Allen et al., 1998; Nagy et al., 2013), and their inclusion was therefore justified both by their direct role in governing atmospheric water demand and their established relevance to SMC dynamics at multiple depths.

### 3.2.2. Satellite-derived vegetation indices

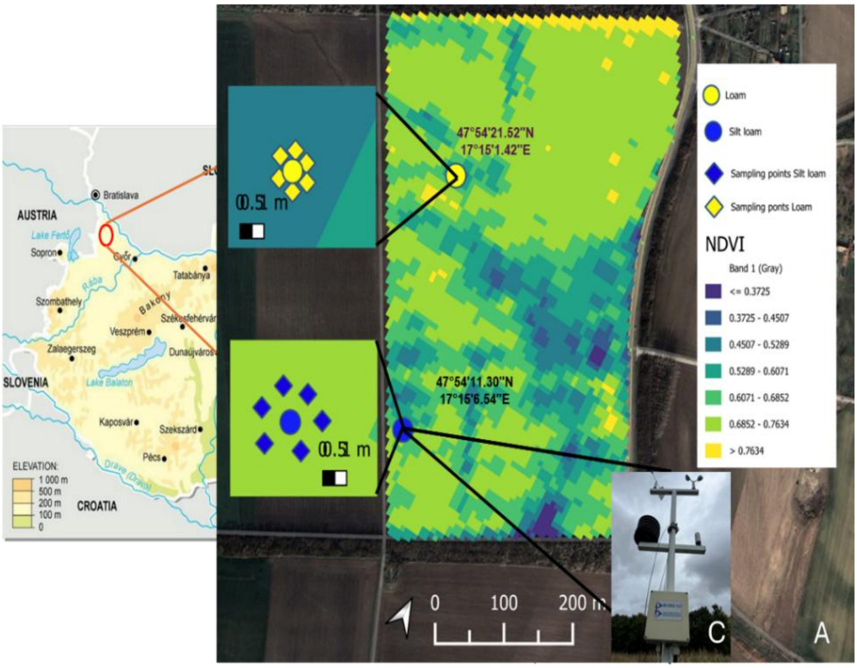
Sentinel-2 multispectral imagery was utilized to derive vegetation indices indicative of crop canopy status as proxies for soil moisture conditions. Sentinel-2 Level-2A images closest to each sampling date (within ±3 days, cloud-free) were processed in QGIS (Figure 2). (see Appendix 1 and 2 for complete acquisition dates and satellite images in Article 3). NDVI and NDMI were calculated using bands 4, 8, 8A, and 11, and extracted at sampling point coordinates.

$$NDVI = \frac{\text{NIR}-\text{Red}}{\text{NIR}+\text{Red}} \qquad (2)$$ (Strashok et al., 2022).

Using Sentinel-2 Band 8 (NIR, 842 nm) and Band 4 (Red, 665 nm).

$$NDMI = \frac{\text{NIR}-\text{SWIR1}}{\text{NIR}+\text{SWIR1}} \qquad (3)$$ (Strashok et al., 2022).

Using Sentinel-2 Band 8A (NIR narrow, 865 nm) and Band 11 (SWIR1, 1610 nm).



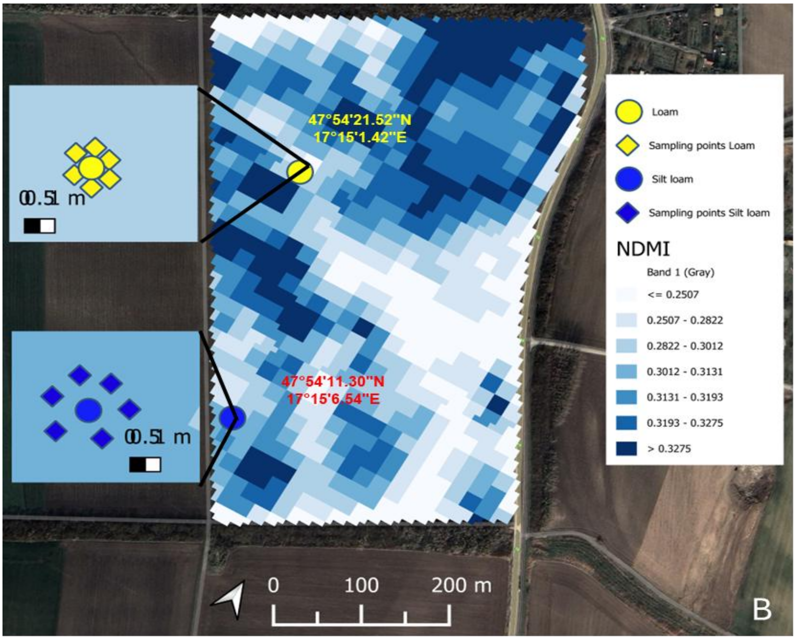



**Figure 2.** The field location from topographical map of Hungary, masked with (A)NDVI index and values in one date and (B) NDMI index and values downloaded from sentinel hub and visualized by QGIS, with sampling points and (C) IoT based sensor.

**Table 1.** Summary of datasets and modelling frameworks used in the dissertation.

| Article | Year | Soil textures | Depths | Samples used | Modelling approach |
|---|---|---|---|---|---|
| **Article 1** | 2023 maize season | Loam, sandy loam, silt loam | 5, 20, 40, 60, 80 cm | 405 | RFR, XGBoost, LSTM |
| **Article 2** | 2023 maize season | Loam, silt loam | 5, 20, 40, 60, 80 cm | 300 | Soil-specific GBR |
| **Article 3** | 2023–2024 maize seasons | Loam, silt loam | 5, 20, 40, 60, 80 cm | 690 | Explainable XGBoost with NDVI, NDMI, and meteorological data |

### 3.3. Data Preprocessing and Feature Engineering

Data preprocessing was performed using Python (version 3.10.12) (Python Software Foundation, 2024). Data were synchronized, checked for outliers, and missing values imputed (forward-fill for short gaps). Soil texture was encoded as a categorical variable. The final dataset included predictors: meteorological variables, depth, soil texture, and vegetation indices (NDVI, NDMI); target: SMC.

#### 3.3.1. Statistical characterization of SMC variability

To characterize SMC dynamics across depths and soil textures, descriptive statistics (mean, standard deviation, coefficient of variation) were calculated for each soil texture and depth using Python.

#### 3.3.2. Feature sets and Modelling frameworks:

Three modeling frameworks were developed corresponding to the three articles, all using Python (scikit-learn, TensorFlow/Keras, XGBoost).

- Multi-model comparison (Article 1)

RFR, XGBoost, and LSTM were trained on meteorological predictors, depth and soil texture to predict SMC across three textures and five depths. Data split: 80% training, 20% testing. Hyperparameters tuned via grid search (RFR, XGBoost) and early stopping (LSTM). LSTM architecture: two LSTM layers (50 units each) and dense output; input normalized to [0,1].

- For Article 2, soil-specific GBR models were developed for loam and silt loam separately.

Gradient Boosting Regressor (GBR): is an ensemble method that sequentially fits weak learners (typically shallow decision trees), where each subsequent tree is trained to correct the residuals of the previous ensemble (Friedman, 2001; Natekin and Knoll, 2013). Separate GBR models were developed for loam and silt loam using the same meteorological predictors and depth. Hyperparameters optimized with GridSearchCV (5-fold). Data split: 70% training, 30% testing.

- Explainable XGBoost with multi-source fusion (Article 3):

XGBoost models were built for each soil texture (loam, silt loam) and depth (5–80 cm) under two scenarios: (A) vegetation indices only (NDVI, NDMI); (B) combined (vegetation + meteorological). Hyperparameters tuned via GridSearchCV. SHAP analysis (TreeExplainer) applied to interpret feature contributions.

### 3.4. Model Evaluation and Performance Metrics

All models were evaluated using a consistent set of performance metrics. The following metrics were calculated on hold-out test sets:

- Mean Squared Error (MSE): Average of squared differences between predicted and observed values (Willmott et al., 1985).

$$MSE = {}^{1}/_{n} \sum_{t=1}^{n} (y_t - \hat{y}t)^2 \quad (4)$$

- Root Mean Squared Error (RMSE): Square root of MSE, measured in same units as original data, emphasizing larger deviations (Willmott et al., 1985).

$$RMSE = \sqrt{{}^{1}/_{n} \sum_{t=1}^{n} (y_t - \hat{y}t)^2} \quad (5)$$

- Mean Absolute Error (MAE): Average of absolute differences between predictions and observations (Willmott et al., 1985).

$$MAE = {}^{1}/_{n} \sum_{t=1}^{n} |y_t - \hat{y}t| \quad (6)$$

- Coefficient of Determination ($R^2$): Proportion of variance in observed data explained by the model (Wright, 1921). $R^2 = 1$ indicates perfect fit; $R^2 = 0$ indicates model performs no better than using the mean observed value.

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}t)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \qquad (7)$$

- Nash–Sutcliffe Efficiency (NSE): Hydrological metric indicating model skill relative to the mean observed value (Nash and Sutcliffe, 1970). NSE = 1 is perfect; NSE = 0 indicates no skill over the mean.

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (8)$$

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

**3.5. Feature importance analysis**

For improved model interpretability and to better understand the main factors driving SMC variability, feature importance analysis was performed using methods tailored to each specific model type:

- For RFR: Gini-based feature importance (also known as mean decrease in impurity) from the scikit-learn library was used. This method reflects the average reduction in prediction error contributed by each feature across all trees in the model (Nembrini et al., 2018).

- For XGBoost: Built-in gain-based feature importance was applied, which measures the total improvement in model accuracy achieved when splitting on a particular feature (Wang and Gao, 2025).
- For LSTM: A model-agnostic permutation importance approach was used. This method evaluates how much the model's prediction error (MSE) increases when a predictor variable is randomly shuffled in the validation dataset, effectively breaking its relationship with the target variable.
- For GBR: Built-in feature importance from scikit-learn was used, based on the reduction in impurity associated with each feature across all trees in the ensemble.

All feature importance analyses were conducted separately for each soil texture, depth, and model to ensure that the results captured context-specific patterns and driver hierarchies.

For Article 3, SHAP analysis was applied to interpret the trained XGBoost models. SHAP, which is particularly well suited for tree-based models, quantifies the contribution of each feature to individual predictions using Shapley values derived from cooperative game theory (Ergün, 2023). For each depth- and soil-specific model, the mean absolute SHAP values were calculated across all test samples to provide an overall ranking of feature importance. The SHAP analysis was implemented using the Python shap library (version 0.41).

# 4. Results

## 4.1. Statistical Analysis Results and SMC patterns for the Three Soil Textures.

In a study based on 405 SMC measurements collected from three soil textures silt loam, loam, and sandy loam across five depths during the 2023 maize growing season, clear soil texture- and depth-dependent differences were observed. Silt loam showed the highest average moisture levels, with mean SMC ranging from 14.99 – 18.27 ± 2.62-7.87%. Loam showed moderate SMC values, ranging from 12.61- 16.19 ± 3.10 - 4.39%. Sandy loam had the lowest average water retention, with mean SMC ranging from 12.43% to 14.98% ±2.15, reflecting its higher sand content and lower water-holding capacity. These results confirm that soil texture strongly influenced water retention, while depth affected both the mean SMC and its variability. (Figure 3).

Overall, deeper soil layers across all textures tended to hold more variable moisture. These findings emphasize the importance of accounting for both soil type and depth when planning irrigation strategies, as understanding these factors can improve water retention and support more efficient and sustainable water use (Guo et al., 2020).

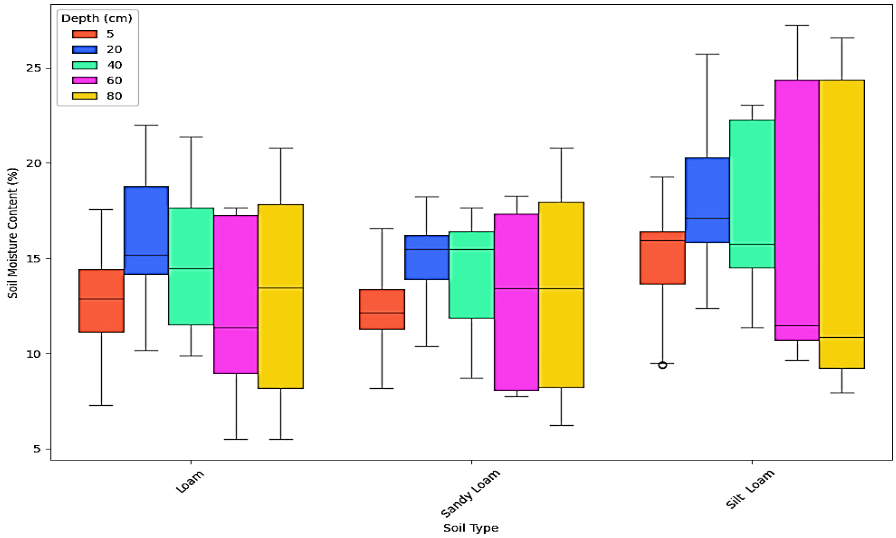



**Figure 3.** Soil moisture content variation in three soil textures at five different depths during 2023 season.

### 4.2.Statistical analysis of SMC variation across two soil textures (loam and silt loam).

The analysis of 705 gravimetric soil moisture content samples collected over two growing seasons revealed clear differences in moisture patterns depending on both soil texture and depth (Table 2). In loam soils, the average soil moisture content ranged from 12.6% at 60 cm depth to 16.2% at 20 cm depth, while the coefficient of variation (CV) increased from 24% at 5 cm to 39% at 80 cm.

In silt loam soils, overall moisture levels were higher, ranging from 15.0% to 22.5%, and variability also increased more sharply with depth, with CV rising from 17.5% at 5 cm to 47.8% at 80 cm. Statistical analysis using repeated measures ANOVA confirmed that soil depth had a significant effect on soil moisture content ($p < 0.05$) in both soil types. These findings support the need for depth-specific modeling and highlight how changes in soil structure and reduced infiltration rates can make moisture retention less predictable at deeper layers (ZHAO et al., 2018).

**Table 2**: Descriptive statistical analysis results for both soil textures at five depths.

| Soil texture | Depth (cm) | Mean SMC (%) | SD (%) | CV (%) |
|---|---|---|---|---|
| Loam | 5 | 12.8 | 3.08 | 24 |
| Loam | 20 | 16.19 | 3.51 | 21.7 |
| Loam | 40 | 14.66 | 3.76 | 25.6 |
| Loam | 60 | 12.61 | 4.39 | 34.8 |
| Loam | 80 | 12.68 | 4.94 | 39 |
| Silt loam | 5 | 14.99 | 2.62 | 17.5 |
| Silt loam | 20 | 18.27 | 3.51 | 19.2 |
| Silt loam | 40 | 18.03 | 4.11 | 22.8 |
| Silt loam | 60 | 16.97 | 7.23 | 42.6 |
| Silt loam | 80 | 16.47 | 7.87 | 47.8 |

### 4.3. Predictive performance across models and predictor scenarios.

#### 4.3.1. Performance evaluation across three soil textures and five depths.

The evaluation of three machine learning models RFR, XGBoost, and LSTM for predicting SMC across different soil textures and depths produced several key insights, as summarized in Figure 4. Across three soil textures and five depth levels, the Random Forest Regression (RFR) model delivered the most consistent and broadly applicable performance ($R^2 = 0.82 - 0.99$, RMSE < 1.5%). These results confirm that RFR is highly effective at capturing variations in soil moisture across a wide range of soil and depth conditions, consistent with findings from previous studies (Cheng et al., 2022; Draper et al., 2013; Lv et al., 2024; Ning et al., 2023). The XGBoost model performed particularly well at shallow to moderate depths (5–40 cm), achieving very low prediction errors, including an RMSE as low as 0.10% in sandy loam at 40 cm depth. However, its performance declined in deeper layers and in finer-textured soils. The LSTM model was especially effective at capturing time-related changes in soil moisture at shallow to intermediate depths (5–60 cm), reaching a peak performance of $R^2 = 0.99$ at 60 cm depth in loam soil. Nevertheless, its accuracy decreased at greater depths, and in some cases, the model showed signs of overfitting. Overall, the results demonstrate that all three models RFR, XGBoost, and LSTM are capable of effectively predicting soil moisture under varying soil textures and depths, supporting conclusions reported in earlier research (Filipović et al., 2022; Lv et al., 2024; Ning et al., 2023). Feature importance analysis and SHAP results consistently identified precipitation and temperature as the dominant predictors at shallow depths (5–20 cm), while the high temporal variability in SMC at these depths (CV up to 24% in loam and 17.5% in silt loam; Table 2) further supports the conclusion that surface soil layers respond rapidly to short-term meteorological forcing such as rainfall and temperature fluctuations, while deeper layers (40–80 cm) are influenced more by long-term factors like soil structure and inherent physical properties. This finding suggests that improving predictions for deeper soil layers will require incorporating additional variables that better represent these underlying soil characteristics (Babaeian et al., 2019).

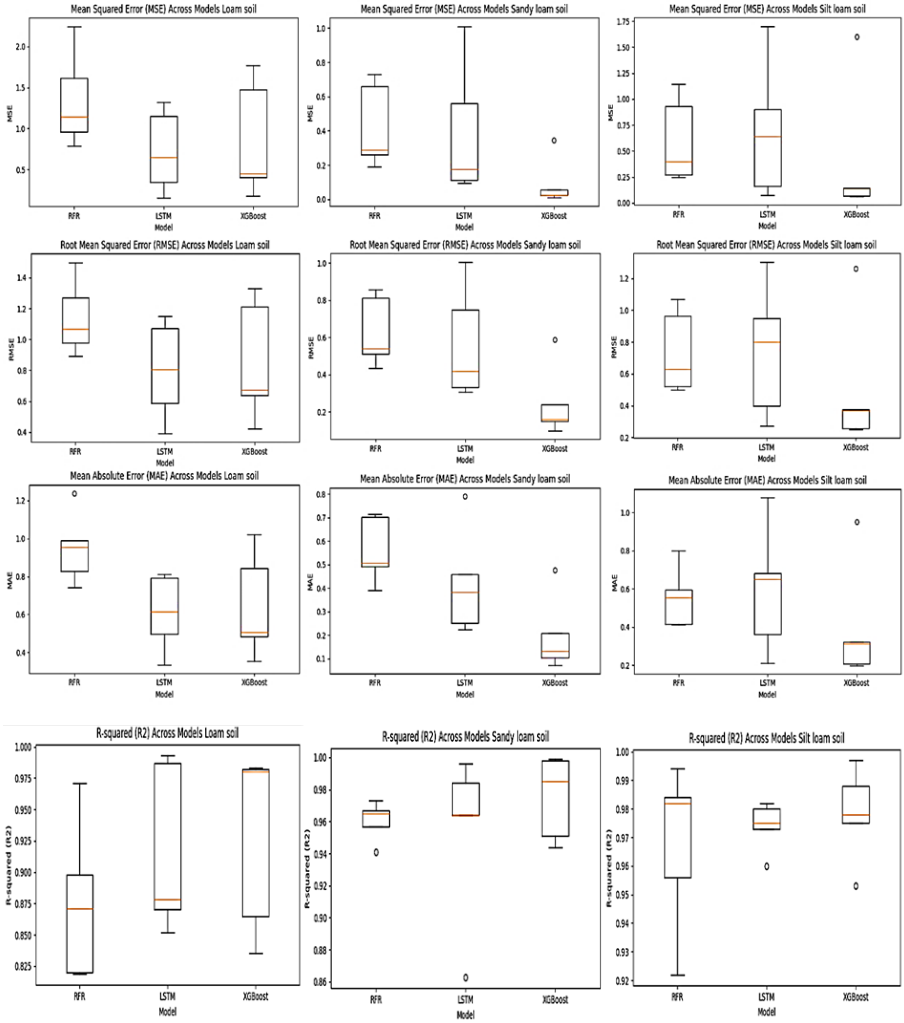



**Figure 4.** Predictive performance and validation results ($R^2$, MSE, RMSE, And MAE) for all soil textures with three models.

### 4.3.2. GBR Model Performance Results

GBR models demonstrated strong predictive performance for SMC across both silt loam and loam soils (Figure 5), as validated by 5-fold cross-validation. The model achieved higher accuracy in silt loam ($R^2$ = 0.98, MSE = 0.73, RMSE = 0.85%, MAE = 0.45%, NSE = 0.95) compared to loam ($R^2$ = 0.94, MSE = 0.94, RMSE = 0.97%, MAE = 0.73%, NSE = 0.94), largely due to the lower SMC variability and better water retention in silt loam. These results emphasize the critical role of soil texture in model performance and highlight the need for soil-specific models and the integration of additional physical soil properties to enhance prediction reliability across various soil textures (Kandala et al., 2024).

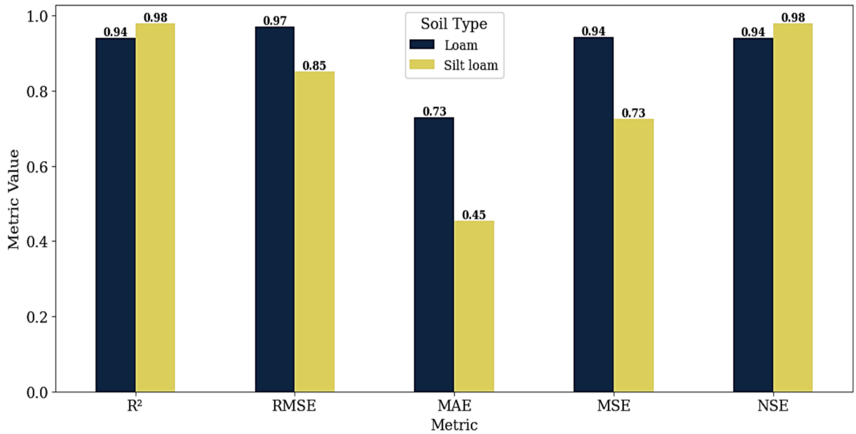



**Figure 5.** Evaluation metrics ($R^2$, MSE, RMSE, MAE, and NSE) for soil moisture content prediction in loam and silt loam soils using the GBR model.

#### 4.3.3. XGBoost Overall Model Performance with different inputs scenarios

Combining Sentinel-2 vegetation indices with IoT meteorological data consistently improved prediction accuracy over vegetation-only models across all depths and textures. The improvement was most pronounced in loam at shallow depths (5–20 cm) where RMSE reduction >60% (from 2.26% to 0.88% at 5 cm). In silt loam, vegetation-only models already performed well with $R^2$ up to 0.98 at 20 cm, but combined inputs still enhanced predictions to near-perfect accuracy ($R^2$ = 0.99). At deeper layers (40–80 cm), combined models maintained high accuracy with $R^2$ above or equal to 0.92 in loam, and around 0.98 in silt loam (Figure 6).

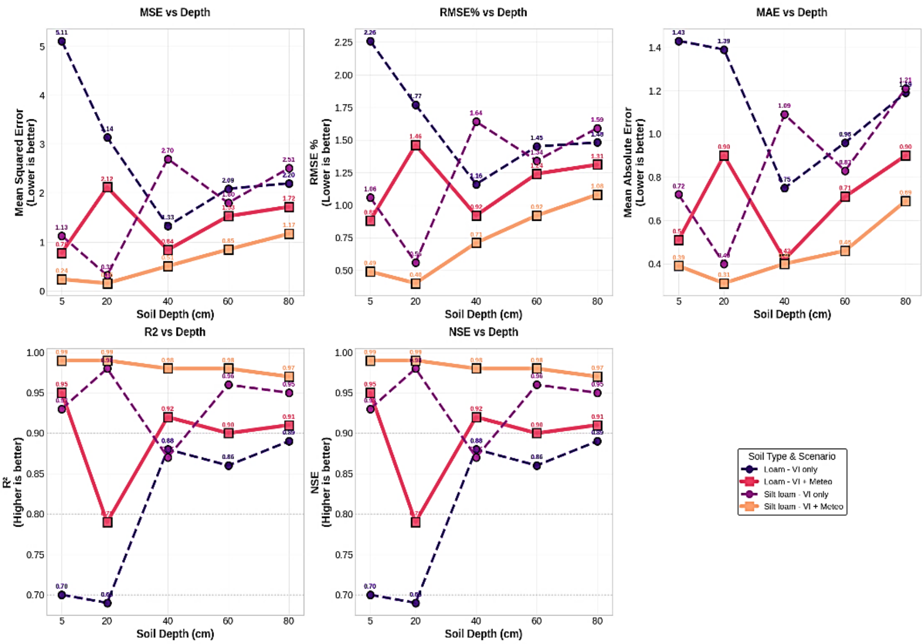



**Figure. 6** The evaluation metrics results of XGBoost model across all depths in two soil textures.

## 4.4. Feature importance analysis and Explainable AI approach results

### 4.4.1. Results of three models with three soil textures and 5 depths

Feature importance analysis of RFR, XGBoost, and LSTM models in SMC prediction revealed that solar radiation is the most influential variable, particularly at deeper soil layers (40–80 cm) (Article 1: Figure 12), with importance values ranging from 0.4 to 0.85 across all soil textures and depths especially in silt loam (RFR: 0.8, XGBoost: 0.79, LSTM: 0.7), underscoring the dominant role of evaporative demand (Han et al., 2023). At the surface layer (5 cm), precipitation emerged as one of the most influential predictors of soil moisture, with importance values of 0.14 for RFR, 0.40 for XGBoost, and 0.25 for LSTM, reflecting its direct role in replenishing surface moisture (Dai et al., 2022). Other environmental factors, including humidity, temperature, and wind speed, also played important roles, but their influence varied depending on the model used, the soil depth, and the soil texture. For example, humidity and wind speed were particularly important in the surface layers of loam and silt loam when modeled with RFR and LSTM, whereas temperature showed greater importance in sandy loam when using XGBoost, with importance values of 0.17 at 5 cm and 0.44 at 20 cm depth. These differences highlight the distinct strengths of each modeling approach. LSTM models are especially effective at identifying patterns that develop over time, which leads them to emphasize variables with delayed or cumulative effects. In contrast, RFR and XGBoost tend to prioritize variables that have a more immediate and direct impact on predictions. Despite these variations, solar radiation and precipitation consistently stood out as the most critical drivers across all models. This pattern aligns well with fundamental hydrological principles, where precipitation represents the primary source of water input, and solar radiation and temperature play central roles in controlling evaporation and evapotranspiration processes (Du et al., 2021).

### 4.4.2. Feature importance results of GBR model.

The study revealed that the main predictors of SMC vary depending on soil texture. In loam soils, precipitation and humidity were the most influential factors, each contributing about 31% to the predictions (Article 2: Figure 6), while soil depth also played a significant role (27%). These variables are closely linked to how moisture enters the soil and how water exchanges occur between the soil and the atmosphere (McColl and Tang, 2024). In contrast, for silt loam soils, solar radiation (46%) and wind speed (26%) emerged as the dominant drivers, highlighting the strong influence of radiation-driven evapotranspiration and wind processes in fine-textured soils. These findings are consistent with earlier research demonstrating the complex nature of soil–water interactions (Nyéki et al., 2022) and reinforce the importance of combining meteorological, soil, and vegetation data to achieve reliable and robust soil moisture predictions (Zheng et al., 2024).

### 4.4.3. SHAP analysis results for XGBoost model with two different scenarios.

Moreover, SHAP analysis showed that NDVI is the most important predictor in determining SMC across all soil textures and depths (Article 3, Figure 6), particularly in the surface layer, since a denser and healthier canopy is a sure sign of wetter soils. The strongest correlation between vegetation indices and SMC is observed in silt loam soils at shallow depths, where the dynamics between moisture and vegetation are stable. The addition of meteorological factors improved the accuracy of the model in all soils and depths (Article 3, Figures 7 and 8), with solar radiation being a key factor in the loss of moisture in deeper soils (Mu et al., 2022), while precipitation is a key predictor in the surface layer and in deeper soils where there is recharge. The strongest correlation is observed in silt loam, while in loam, due to the rapid changes in moisture, meteorological factors are important in accurately predicting SMC, according to Liu et al. (2024).

# 5. Discussion:

## 5.1. Depth dependent soil moisture dynamics across soil textures

The observed trend of increased SMC variability with depth, indicated by the maximum CV of 47.8% at 80 cm depth in silt loam soils, is consistent with the understanding that subsoil SMC is controlled by the cumulative effect of infiltration, root uptake, redistribution, etc., rather than changes in surface SMC (Guo et al., 2020). Significant depth effects at $p < 0.05$ confirm that depth-specific modeling is critical; this is one of the modeling gaps that need to be filled, as indicated in the literature (Kheimi and Zounemat-Kermani, 2025).

## 5.2. Comparative model performance and algorithm selection.

RFR was found to be the most robust across all scenarios, which aligns with its capacity for variance reduction through ensemble averaging (Breiman, 2001). On the other hand, XGBoost performed well at shallow depths with strong met signals, but its performance diminished at deeper depths, which could be a result of its focus on selecting the most important features (Chen and Guestrin, 2016). Finally, LSTM performed well with strong emphasis on regularization, but its performance at limited depths indicates that a recurrent network might need additional static features like soil properties for deep-layer predictions (Filipović et al., 2022). The above results provide a guide for selecting a model: RFR for general-purpose depth-resolved SMC predictions, XGBoost for surface-focused predictions, and LSTM when there are ample sequential data with moderate depths. The soil-specific GBR models presented in Article 2 resulted in high accuracy with $R^2$ = 0.94-0.98, with clear delineations between different feature importance, thus proving that texture-aware models are significantly better than generic ones. This further emphasizes that models need recalibration or retraining when ported from one soil type to another (Kandala et al., 2024).

### 5.3. Value of multi-source data fusion

Integrating optical satellite indices, such as NDVI and NDMI, with ground-based meteorological data significantly improves SMC prediction, especially in loam soils where rapid wetting and drying cycles occur (Monteiro et al., 2024). While vegetation-only models using satellite-derived optical indices as sole predictors can achieve high accuracy in stable fine-textured soils with $R^2$ up to 0.98 at 20 cm (Chen et al., 2015), they fall short in capturing fast changing moisture dynamics in loam, necessitating high-frequency meteorological data (Wang et al., 2007). Addressing the lack of field-scale, depth-specific frameworks, these findings highlight the importance of multi-source data fusion for reliable SMC estimation across diverse soil conditions (Alahmad et al., 2023).

### 5.4. Depth-dependent drivers and its interpretability.

The SHAP analysis indicated that with depth, the determining factor for SMC changes. In the top layers of the soil profile (5-20 cm), plant indices such as NDVI are most important (Srivastava et al., 2018). This indicates that there is a mutual relationship between healthy vegetation and top layers of soil moisture. In the middle layers of the profile, both plant indices and solar radiation are important, while in the deepest layers (60-80 cm), solar radiation is dominant, controlling deep soil moisture loss through root uptake and evaporation. The dominant role of solar radiation at deeper soil layers (40–80 cm) can be explained through its central role in driving evapotranspiration. High radiation increases $ET_0$, which intensifies root-zone water extraction by the crop, progressively depleting moisture at greater depths. This mechanism aligns with the Penman–Monteith framework, in which net radiation is the primary energy driver of ET (Allen et al., 1998). These are affected by soil texture; however, NDVI remains important at depth in silt loams, while meteorological factors are most important throughout the profile in loams, as predicted by hydrological theory (Zhai et al., 2025).

### 5.5. Implication for precision crop production and irrigation

This research uses precision depth-specific soil moisture prediction, feature importance, and SHAP analysis to bring complex models to life in irrigation optimization. In loam soils, the addition of short-term weather and surface NDVI allows for optimal irrigation, with the model's capacity to accurately predict deeper soil moisture preventing overwatering ($R^2$ value of 0.99 at 5-20 cm depth). In silt loam soils, the constant soil moisture in deeper depths and high NDVI correlation enable irrigation to be scheduled mainly by satellite imagery and weather, with solar radiation being a major driver. The recommended irrigation strategy is to use NDVI to detect early stress, weather to predict changes, direct soil testing periodically, especially deeper depths, and soil type-specific thresholds, which prioritize rapid response in loam and stable irrigation in silt loam. These strategies contribute to achieving the Sustainable Development Goals, specifically Goal 6 and Goal 13, by making irrigation more efficient and resilient, which is easily scalable with satellite and IoT imagery (United Nation, 2024, 2023).

### 5.6. Limitations and future research directions

This research offers important insights, but it has some limitations, such as the use of Sentinel-2 optical indices, which are limited by cloud cover, saturation, and delay in data availability (Babaeian et al., 2019). The addition of Sentinel-1 SAR imagery, which is less affected by atmospheric problems and is sensitive to soil moisture, may improve the results, although corrections for canopy and surface effects are necessary (Rahmati et al., 2026). The model's accuracy may be overestimated due to in-period predictions, while the addition of meteorological data improved accuracy in loam soils, although errors increase in drought conditions, indicating the importance of temporal validation (Chen et al., 2017). The study's single-field design may require new models to be developed to cover different areas, while the absence of specific soil, management, and topographic factors may affect the accuracy of the model, particularly in deeper layers. Future research should focus on integrating multi-sensor data, improving validation methods, expanding spatial and predictor coverage, and linking soil moisture prediction with yield and irrigation models for broader applicability and robustness. Due to soil moisture is temporally dependent, the use of random train–test splitting may increase the similarity between training and testing samples when observations collected close in time are included in both subsets. Therefore, the reported model performance should be interpreted as in-period predictive accuracy rather than fully independent future-season forecasting. This limitation was partly reduced by applying hold-out testing and cross-validation during model tuning, but future work should include stricter temporal validation, such as leave-one-date-out, blocked time-series validation, or training on one growing season and testing on another. Such validation would provide a stronger assessment of model transferability under contrasting meteorological conditions. Additionally, evapotranspiration was not explicitly incorporated as a predictor variable or used as a water balance framework to contextualize the observed SMC patterns. Future work should consider integrating $ET_0$ estimates derived from the FAO Penman–Monteith equation or remotely sensed energy balance models to improve both model accuracy and the interpretability of depth-dependent moisture dynamics.

## 6. Conclusions:

The aim of this dissertation was to improve soil moisture predictions in precision agriculture by developing and testing machine learning models that incorporate meteorological data, field data, and satellite data. The results show that it is not only possible to obtain accurate predictions of soil moisture at different depths in the soil profile, but also to make these predictions interpretable and actionable for irrigation management.

Three significant results were achieved. Firstly, the dissertation proves that depth-specific modeling is essential. This is due to soil moisture varies significantly with depth in the soil profile. The coefficient of variation was found to vary by up to 47.8% in the 80 cm depth in silt loam soils. Secondly, the dissertation proves that data fusion using both vegetation data and meteorological data was superior to vegetation data only in all soils investigated. The improvement in accuracy was found with RMSE reductions exceeding 60% in loam soils at shallow depths. Thirdly, the dissertation proves that the data fusion model was interpretable by applying SHAP analysis. The results show that the most important variable at shallow depths was NDVI, while at deeper depths solar radiation was the most important. Additionally, the dissertation proves that soil texture plays an important role in altering these effects.

This dissertation also opens the door to future breakthroughs. Firstly, the incorporation of Sentinel-1 SAR data would overcome the limitations of clouds cover. Also, the expansion of this research to other soils, crops, and regions would show how generalizable this research is. This dissertation emphasize that it is possible to obtain accurate predictions of soil moisture at different depths in the soil profile by integrating ground-truth measurements, IoT sensor networks, and satellite remote sensing within an interpretable machine learning framework.

**7. Theses:**

**7.1. Scientific Theses**

1. Soil moisture content varies significantly with both depth and soil texture throughout the maize growing season. In loam soils, gravimetric SMC ranged from 12.6% to 16.2% across depths, with coefficient of variation (CV) reaching 39% at 80 cm. In silt loam soils, SMC ranged from 15.0% to 18.3%, with CV reaching 47.8% at 80 cm. These differences confirm that depth- and texture-specific modeling is statistically necessary and cannot be replaced by a single generalized model.
2. Among the machine learning algorithms evaluated, Random Forest Regression (RFR) achieved the most consistent predictive accuracy across all soil textures and depths, with $R^2$ values ranging from 0.82 to 0.99 and RMSE below 1.5% under all conditions. XGBoost performed comparably at shallow depths (5–20 cm), while LSTM showed adequate performance at intermediate depths but degraded in deeper layers. GBR models trained separately for loam and silt loam soils demonstrated strong cross-validated performance, confirming that soil-specific modeling further improves accuracy over texture-pooled approaches.
3. The dominant environmental drivers of soil moisture content are texture- and depth-dependent, as quantified through feature importance and SHAP analysis. In loam soils, precipitation and relative humidity each contributed approximately 31% to model

predictions, reflecting the rapid hydrological response of coarser-textured profiles. In silt loam soils, solar radiation was the dominant driver (46%), followed by wind speed (26%), indicating stronger evaporative control. SHAP analysis further revealed a consistent depth gradient: NDVI was the primary predictor at surface layers (5–20 cm) across both textures, while solar radiation became dominant at deeper layers (60–80 cm), reflecting the progressive influence of cumulative evaporative demand on subsoil moisture.

4. Integrating Sentinel-2 vegetation indices (NDVI, NDMI) with IoT-based meteorological data consistently improved XGBoost model accuracy compared to vegetation-index-only inputs across all depths and soil textures. The improvement was most pronounced in loam soils at 5 cm depth, where RMSE decreased from 2.26% to 0.88% a reduction of more than 60%. This demonstrates that multi-source data fusion is a scientifically validated strategy for improving depth-resolved soil moisture prediction, particularly in soil textures with rapid surface moisture dynamics.
5. A depth-resolved, explainable machine learning framework integrating gravimetric field measurements, high-frequency IoT meteorological data, and Sentinel-2 satellite imagery was developed and validated across multiple soil textures and depths over two growing seasons. The framework combining RFR, GBR, and XGBoost models with SHAP interpretability produced reliable, physically interpretable predictions of soil moisture content and identified texture- and depth-specific driver hierarchies. This provides a scientifically grounded foundation for

developing precision irrigation decision-support systems tailored to site-specific soil and crop conditions.

**7.2. Practical Theses:**

1. Irrigation scheduling should be differentiated by soil texture. In loam soils, real-time meteorological inputs particularly precipitation and humidity are essential for capturing rapid moisture changes across the full profile. In silt loam soils, satellite-derived NDVI trends are sufficient for routine scheduling, with deep-layer verification recommended only during extended dry or extreme weather periods.
2. Solar radiation is the primary driver of deep soil moisture depletion (60–80 cm) through its role in driving evaporative demand. Irrigation decisions should account for cumulative radiation exposure over 3–5 consecutive days, not only surface layer status or rainfall records, particularly in loam soils where atmospheric signals propagate rapidly through the profile.
3. SHAP-derived decision rules can translate model outputs into actionable irrigation heuristics: (i) when surface NDVI is high and solar radiation is low, irrigation can be withheld even if surface soil appears dry; (ii) when cumulative radiation is high and no rainfall has occurred for 3–5 days, deep-profile depletion should be anticipated and irrigation triggered regardless of surface moisture readings. These rules are operable without specialized software.
4. The proposed monitoring architecture is deployable using existing low-cost infrastructure: free Sentinel-2 imagery, LoRaWAN IoT weather stations, and open-source Python libraries (scikit-learn,

XGBoost, SHAP). This makes the framework scalable across different farm sizes and regions without requiring specialized hardware investment.

5. The modeling framework enables proactive rather than reactive irrigation management. By combining real-time IoT sensor data with ML-based depth-specific predictions, soil moisture stress can be anticipated before it affects crop yield, allowing farmers to schedule irrigation at agronomically optimal timing rather than responding to visible crop stress.

8. **References**:

Ågren, A.M., Larson, J., Paul, S.S., Laudon, H., Lidberg, W., 2021. 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 404, 115280. https://doi.org/10.1016/j.geoderma.2021.115280

Alahmad, T., Neményi, M., Nyéki, A., 2025. Soil Moisture Content Prediction Using Gradient Boosting Regressor (GBR) Model: Soil-Specific Modeling with Five Depths. Applied Sciences 15, 5889. https://doi.org/10.3390/app15115889

Alahmad, T., Neményi, M., Nyéki, A., 2025. Soil moisture prediction with machine learning: combining vegetation indices and meteorological data, in: Precision Agriculture'25. Wageningen Academic, pp. 1319–1326.

Alahmad, T., Neményi, M., Nyéki, A., 2024. Soil Moisture Content Prediction in Loam Soil with RFR Model. Acta Agronomica Óváriensis 65, 43–56. https://doi.org/10.17108/ActAgrOvar.2024.65.2.43

Alahmad, T., Neményi, M., Nyéki, A., 2023. Applying IoT Sensors and Big Data to Improve Precision Crop Production: A Review. Agronomy 13, 2603. https://doi.org/10.3390/agronomy13102603

Alahmad, T., Neményi, M., Széles, A., Ali, N., Hijazi, O., Nyéki, A., 2025. Spatiotemporal prediction of soil moisture content at various depths in three soil types using machine learning algorithms. Frontiers in Soil Science 5. https://doi.org/10.3389/fsoil.2025.1612908

Alibabaei, K., Gaspar, P.D., Lima, T.M., 2021. Modeling Soil Water Content and Reference Evapotranspiration from Climate Data Using Deep Learning Method. Applied Sciences 11, 5029. https://doi.org/10.3390/app11115029

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Fao, rome 300, D05109.

Ambrus, B., 2023. Adatgyűjtési Technológiák Fejlesztése A Precíziós Növénytermesztésnél, Különös Tekintettel A Magajáró Berendezésekre. Széchenyi István Egyetem Albert Kázmér Mosonmagyaróvári Kar.

Azorin-Molina, C., Vicente-Serrano, S.M., Sanchez-Lorenzo, A., McVicar, T.R., Morán-Tejeda, E., Revuelto, J., El Kenawy, A., Martín-Hernández, N., Tomas-Burguera, M., 2015. Atmospheric evaporative demand observations, estimates and driving factors in Spain (1961–2011). J. Hydrol. (Amst). 523, 262–277. https://doi.org/10.1016/j.jhydrol.2015.01.046

Babaeian, E., Sadeghi, M., Jones, S.B., Montzka, C., Vereecken, H., Tuller, M., 2019. Ground, Proximal, and Satellite Remote Sensing of Soil Moisture. Reviews of Geophysics 57, 530–616. https://doi.org/10.1029/2018RG000618

Bibek, A., Vivek, S., 2025. Comparative Analysis of Soil and Water Dynamics in Conventional and Sod-Based Crop Rotation in Florida. Frontiers in Agronomy 7. https://doi.org/doi:10.3389/fagro.2025.1552425

Bicamumakuba, E., Reza, M.N., Jin, H., Samsuzzaman, Lee, K.-H., Chung, S.-O., 2025. Multi-Sensor Monitoring, Intelligent Control, and Data Processing for Smart Greenhouse Environment Management. Sensors 25, 6134. https://doi.org/10.3390/s25196134

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Cai, Y., Zheng, W., Zhang, X., Zhangzhong, L., Xue, X., 2019. Research on soil moisture prediction model based on deep learning. PLoS One 14, e0214508. https://doi.org/10.1371/journal.pone.0214508

Celik, M.F., Isik, M.S., Yuzugullu, O., Fajraoui, N., Erten, E., 2022. Soil Moisture Prediction from Remote Sensing Images Coupled with Climate, Soil Texture and Topography via Deep Learning. Remote Sens. (Basel). 14, 5584. https://doi.org/10.3390/rs14215584

Chen, H., Fan, L., Wu, W., Liu, H.-B., 2017. Comparison of spatial interpolation methods for soil moisture and its application for monitoring drought. Environ. Monit. Assess. 189, 525. https://doi.org/10.1007/s10661-017-6244-4

Chen, L., Wang, Y., Ren, J., Song, H., Zhang, G., Sun, J., Wang, L., Mu, J., Liu, C., Wang, M., 2025. Generating high accuracy multi-layer soil moisture at daily scale in the black soil region of China. Sci. Data 12, 1705. https://doi.org/10.1038/s41597-025-05986-7

Chen, M., Willgoose, G.R., Saco, P.M., 2015. Investigating the impact of leaf area index temporal variability on soil moisture predictions using remote sensing vegetation data. J. Hydrol. (Amst). 522, 274–284. https://doi.org/10.1016/j.jhydrol.2014.12.027

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. https://doi.org/10.1145/2939672.2939785

Cheng, M., Jiao, X., Liu, Y., Shao, M., Yu, X., Bai, Y., Wang, Z., Wang, S., Tuohuti, N., Liu, S., Shi, L., Yin, D., Huang, X., Nie, C., Jin, X., 2022. Estimation of soil moisture content under high maize canopy coverage from UAV multimodal data and machine learning. Agric. Water Manag. 264, 107530. https://doi.org/10.1016/j.agwat.2022.107530

Dai, L., Fu, R., Guo, X., Du, Y., Zhang, F., Cao, G., 2022. Soil Moisture Variations in Response to Precipitation Across Different Vegetation Types on the Northeastern Qinghai-Tibet Plateau. Front. Plant Sci. 13. https://doi.org/10.3389/fpls.2022.854152

Dirksen, C., 1999. Soil physics measurements. Catena Verlag.

Draper, C., Reichle, R., de Jeu, R., Naeimi, V., Parinussa, R., Wagner, W., 2013. Estimating root mean square errors in remotely sensed soil moisture over continental scale domains. Remote Sens. Environ. 137, 288–298. https://doi.org/10.1016/j.rse.2013.06.013

Du, M., Zhang, J., Elmahdi, A., Wang, Z., Yang, Q., Liu, H., Liu, C., Hu, Y., Gu, N., Bao, Z., Liu, Y., Jin, J., Wang, G., 2021. Variation Characteristics and Influencing Factors of Soil Moisture Content in the Lime Concretion Black Soil Region in Northern Anhui. Water (Basel). 13, 2251. https://doi.org/10.3390/w13162251

Emami, H., Emami, S., Rezaverdinejad, V., 2025. A backtracking search-based extreme gradient boosting algorithm for soil moisture prediction using meteorological variables. Earth Sci. Inform. 18, 181. https://doi.org/10.1007/s12145-024-01674-z

Ergün, S., 2023. Explaining Xgboost Predictions with Shap Value: A Comprehensive Guide To Interpreting Decision Tree-Based Models. New Trends in Computer Sciences 1, 19–31. https://doi.org/10.3846/ntcs.2023.17901

Filipović, N., Brdar, S., Mimić, G., Marko, O., Crnojević, V., 2022. Regional soil moisture prediction system based on Long Short-Term Memory network. Biosyst. Eng. 213, 30–38. https://doi.org/10.1016/j.biosystemseng.2021.11.019

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232.

Gao, Y., Xie, Y., Jiang, H., Wu, B., Niu, J., 2014. Soil water status and root distribution across the rooting zone in maize with plastic film mulching. Field Crops Res. 156, 40–47. https://doi.org/10.1016/j.fcr.2013.10.016

Guo, X., Fu, Q., Hang, Y., Lu, H., Gao, F., Si, J., 2020. Spatial Variability of Soil Moisture in Relation to Land Use Types and Topographic Features on Hillslopes in the Black Soil (Mollisols) Area of Northeast China. Sustainability 12, 3552. https://doi.org/10.3390/su12093552

Haddon, A., Kechichian, L., Harmand, J., Dejean, C., Ait-Mouheb, N., 2023. Linking soil moisture sensors and crop models for irrigation management. Ecol. Modell. 484, 110470. https://doi.org/10.1016/j.ecolmodel.2023.110470

Han, Q., Zeng, Y., Zhang, L., Wang, C., Prikaziuk, E., Niu, Z., Su, B., 2023. Global long term daily 1 km surface soil moisture dataset with physics informed machine learning.

Sci. Data 10, 101. https://doi.org/10.1038/s41597-023-02011-7

Hussein, E.E., Zerouali, B., Bailek, N., Derdour, A., Ghoneim, S.S.M., Santos, C.A.G., Hashim, M.A., 2024. Harnessing Explainable AI for Sustainable Agriculture: SHAP-Based Feature Selection in Multi-Model Evaluation of Irrigation Water Quality Indices. Water (Basel). 17, 59. https://doi.org/10.3390/w17010059

Islam, M.N., Logofatu, D., Haque, M.Z., 2023. A Comparative Study on Machine Learning Methods Through Evaluating the Impact of Contributing Factors on The Accuracy of Soil Moisture Prediction, in: 2023 International Conference on Innovations in Intelligent Systems and Applications (INISTA). IEEE, pp. 1–6. https://doi.org/10.1109/INISTA59065.2023.10310440

Kandala, R., Franssen, H.H., Chaudhuri, A., Sekhar, M., 2024. The value of soil temperature data versus soil moisture data for state, parameter, and flux estimation in unsaturated flow model. Vadose Zone Journal 23. https://doi.org/10.1002/vzj2.20298

Kheimi, M., Zounemat-Kermani, M., 2025. Conventional and advanced AI-based models in soil moisture prediction. Physics and Chemistry of the Earth 139. https://doi.org/10.1016/j.pce.2025.103944

Li, C., Zhang, Y., Ren, X., 2020. Modeling Hourly Soil Temperature Using Deep BiLSTM Neural Network. Algorithms 13, 173. https://doi.org/10.3390/a13070173

Li, Z.-L., Leng, P., Zhou, C., Chen, K.-S., Zhou, F.-C., Shang, G.-F., 2021. Soil moisture retrieval from remote sensing measurements: Current knowledge and directions for the future. Earth. Sci. Rev. 218, 103673. https://doi.org/10.1016/j.earscirev.2021.103673

Little M. and Smith, 1998. A comparison of three methods of soil water content determination. South African Journal of Plant and Soil 15, 80–89.

Liu, Q., Wu, Z., Cui, N., Zheng, S., Zhu, S., Jiang, S., Wang, Z., Gong, D., Wang, Y., Zhao, L., 2024. Soil moisture content estimation of drip-irrigated citrus orchard based on UAV images and machine learning algorithm in Southwest China. Agric. Water Manag. 303, 109069. https://doi.org/10.1016/j.agwat.2024.109069

Lv, C., Xie, Q., Peng, X., Dou, Q., Wang, J., Lopez-Sanchez, J.M., Shang, J., Chen, L., Fu, H., Zhu, J., Song, Y., 2024. Soil moisture retrieval over agricultural fields with machine learning: A comparison of quad-, compact-, and dual-polarimetric time-series SAR data. J. Hydrol. (Amst). 644, 132093. https://doi.org/10.1016/j.jhydrol.2024.132093

Mane, S., Das, N., Singh, G., Cosh, M., Dong, Y., 2024. Advancements in dielectric soil moisture sensor Calibration: A comprehensive review of methods and techniques.

Comput. Electron. Agric. 218, 108686. https://doi.org/10.1016/j.compag.2024.108686

McColl, K.A., Tang, L.I., 2024. An Analytic Theory of Near-Surface Relative Humidity over Land. J. Clim. 37, 1213–1230. https://doi.org/10.1175/JCLI-D-23-0342.1

Milics, G., 2008. A térinformatika és a távérzékelés alkalmazása a precíziós (helyspecifikus) növénytermesztésben. University of Pécs (Hungary).

Monteiro, A.T., Arenas-Castro, S., Punalekar, S.M., Cunha, M., Mendes, I., Giamberini, M., Marques da Costa, E., Fava, F., Lucas, R., 2024. Remote sensing of vegetation and soil moisture content in Atlantic humid mountains with Sentinel-1 and 2 satellite sensor data. Ecol. Indic. 163, 112123. https://doi.org/10.1016/j.ecolind.2024.112123

Mu, Y., Yuan, Y., Jia, X., Zha, T., Qin, S., Ye, Z., Liu, P., Yang, R., Tian, Y., 2022. Hydrological losses and soil moisture carryover affected the relationship between evapotranspiration and rainfall in a temperate semiarid shrubland. Agric. For. Meteorol. 315, 108831. https://doi.org/10.1016/j.agrformet.2022.108831

Nagy, V., Milics, G., Smuk, N., Kovács, A.J., Balla, I., Jolánkai, M., Deákvári, J., Szalay, K.D., Fenyvesi, L., Štekauerová, V., Wilhelm, Z., Rajkai, K., Németh, T., Neményi, M., 2013. Continuous field soil moisture content mapping by means of apparent electrical conductivity (ECa) measurement. Journal of Hydrology and Hydromechanics 61, 305–312. https://doi.org/10.2478/johh-2013-0039

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. J. Hydrol. (Amst). 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. Front. Neurorobot. 7. https://doi.org/10.3389/fnbot.2013.00021

Nembrini, S., König, I.R., Wright, M.N., 2018. The revival of the Gini importance? Bioinformatics 34, 3711–3718. https://doi.org/10.1093/bioinformatics/bty373

Neményi, M., Ambrus, B., Teschner, G., Alahmad, T., Nyéki, A., Kovács, A.J., 2023. Challenges of ecocentric sustainable development in agriculture with special regard to the internet of things (IoT), an ICT perspective. Progress in Agricultural Engineering Sciences 19, 113–122. https://doi.org/10.1556/446.2023.00099

Neményi, M., Kovács, A.J., Oláh, J., Popp, J., Erdei, E., Harsányi, E., Ambrus, B., Teschner, G., Nyéki, A., 2022. Challenges of sustainable agricultural development with special regard to Internet of Things: Survey. Progress in Agricultural Engineering Sciences 18, 95–114. https://doi.org/10.1556/446.2022.00053

Ning, J., Yao, Y., Tang, Q., Li, Y., Fisher, J.B., Zhang, Xiaotong, Jia, K., Xu, J., Shang, K., Yang, J., Yu, R., Liu, L., Zhang, Xueyi, Xie, Z., Fan, J., 2023. Soil moisture at 30 m from multiple satellite datasets fused by random forest. J. Hydrol. (Amst). 625, 130010. https://doi.org/10.1016/j.jhydrol.2023.130010

Nyéki, A., 2016. A precíziós növénytermesztés és a fenntartható mezőgazdaság kapcsolata.

Nyéki, A., Alahmad, T., Arshad, S., Gombkötő, N., Neményi, M., Mirzaei, M., Szilárd, S., Harsanyi, E., Al-Dalahmeh, M., Mohammed, S., 2025. Real-time monitoring of ammonia emissions from cereal crops using LoRaWAN-based sensing technology. Sci. Rep. 16, 1446. https://doi.org/10.1038/s41598-025-31661-3

Nyéki, A., Daróczy, B., Kerepesi, C., Neményi, M., Kovács, A.J., 2022. Spatial Variability of Soil Properties and Its Effect on Maize Yields within Field—A Case Study in Hungary. Agronomy 12, 395. https://doi.org/10.3390/agronomy12020395

Nyéki, A., Kerepesi, C., Daróczy, B., Benczúr, A., Milics, G., Nagy, J., Harsányi, E., Kovács, A.J., Neményi, M., 2021. Application of spatio-temporal data in site-specific maize yield prediction with machine learning methods. Precis. Agric. 22, 1397–1415. https://doi.org/10.1007/s11119-021-09833-8

Pôças, I., Calera, A., Campos, I., Cunha, M., 2020. Remote sensing for estimating and mapping single and basal crop coefficientes: A review on spectral vegetation indices approaches. Agric. Water Manag. 233, 106081. https://doi.org/10.1016/j.agwat.2020.106081

Python Software Foundation, 2024. Python Language Reference, version 3.10.

Rahmati, M., Balenzano, A., Bechtold, M., Brocca, L., Fluhrer, A., Jagdhuber, T., Karamvasis, K., Mengen, D., Reichle, R.H., Kim, S., Taghizadeh-Mehrjardi, R., Walker, J., Zhu, L., Montzka, C., 2026. Soil moisture retrieval from Sentinel-1: Lessons learned after more than a decade in orbit. Remote Sens. Environ. 333, 115146. https://doi.org/10.1016/j.rse.2025.115146

Ren, Y., Ling, F., Wang, Y., 2023. Research on Provincial-Level Soil Moisture Prediction Based on Extreme Gradient Boosting Model. Agriculture 13, 927. https://doi.org/10.3390/agriculture13050927

Shukla, A., Panchal, H., Mishra, M., Patel, P.R., Srivastava, H.S., Patel, P., Shukla, A.K., 2014. Soil moisture estimation using gravimetric technique and FDR probe technique: a comparative analysis. Am. Int. J. Res. Form. Appl. Nat. Sci 8, 89–92.

Singh, A., Gaurav, K., Sonkar, G.K., Lee, C.-C., 2023. Strategies to Measure Soil Moisture Using Traditional Methods, Automated Sensors, Remote Sensing, and Machine Learning Techniques: Review, Bibliometric Analysis, Applications, Research Findings, and Future Directions. IEEE Access 11, 13605–13635. https://doi.org/10.1109/ACCESS.2023.3243635

Songara, J.C., Patel, J.N., 2022. Calibration and comparison of various sensors for soil moisture measurement. Measurement 197, 111301. https://doi.org/10.1016/j.measurement.2022.111301

Srivastava, A., Yetemen, O., Kumari, N., Saco, P.M., 2018. Role of Solar Radiation and Topography on Soil Moisture Variations in Semiarid Aspect-Controlled Ecosystems. sat 1.

Strashok, O., Ziemiańska, M., Strashok, V., 2022. Evaluation and Correlation of Normalized Vegetation Index and Moisture Index in Kyiv (2017–2021). Journal of Ecological Engineering 23, 212–218. https://doi.org/10.12911/22998993/151884

Taheri, M., Bigdeli, M., Imanian, H., Mohammadian, A., 2025. An Overview of Machine-Learning Methods for Soil Moisture Estimation. Water (Basel). 17, 1638. https://doi.org/10.3390/w17111638

Tarolli, P., Zhao, W., 2023. Drought in agriculture: Preservation, adaptation, migration. The Innovation Geoscience 1, 100002. https://doi.org/10.59717/j.xinn-geo.2023.100002

Tsai, P.-H., Huang, Y., Tai, J.-H., 2024. Estimating soil water content from thermal images with an artificial neural network. Catena (Amst). 241, 108029. https://doi.org/10.1016/j.catena.2024.108029

United Nation, 2024. UN-Water Annual Report 2023 .

United Nation, 2023. UN 2023 Water Conference. New York.

USDA, 2024. Field Book for Describing and Sampling Soils V4.0 [WWW Document]. USDA . URL https://www.nrcs.usda.gov/resources/guides-and-instructions/field-book-for-describing-and-sampling-soils (accessed 6.11.25).

Wang, L., Gao, Y., 2025. Soil Moisture Inversion Using Multi-Sensor Remote Sensing Data Based on Feature Selection Method and Adaptive Stacking Algorithm. Remote Sens. (Basel). 17, 1569. https://doi.org/10.3390/rs17091569

Wang, X., Xie, H., Guan, H., Zhou, X., 2007. Different responses of MODIS-derived NDVI to root-zone soil moisture in semi-arid and humid regions. J. Hydrol. (Amst). 340, 12–24. https://doi.org/10.1016/j.jhydrol.2007.03.022

Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnell, J., Rowe, C.M., 1985. Statistics for the evaluation and comparison of models. J. Geophys. Res. Oceans 90, 8995–9005. https://doi.org/10.1029/JC090iC05p08995

Wright, S., 1921. Correlation and causation. J. Agric. Res. 20, 557.

Wu, Y., Yang, Z., Liu, Y., 2023. Internet-of-Things-Based Multiple-Sensor Monitoring System for Soil Information Diagnosis Using a Smartphone. Micromachines (Basel). 14, 1395. https://doi.org/10.3390/mi14071395

Yinglan, A., Jiang, X., Wang, Y., Wang, L., Zhang, Z., Duan, L., Fang, Q., 2023. Study on spatio-temporal simulation and prediction of regional deep soil moisture using machine learning. J. Contam. Hydrol. 258, 104235. https://doi.org/10.1016/j.jconhyd.2023.104235

Zhai, X., Liu, Y., Hong, Y., Yang, Y., Wang, P., Ye, Z., Liu, X., She, T., Wang, L., Xu, C., Zhang, L., Wang, Q., 2025. Improved digital mapping of soil texture using the kernel temperature–vegetation dryness index and adaptive boosting. Ecol. Inform. 87, 103083. https://doi.org/10.1016/j.ecoinf.2025.103083

ZHAO, W., FANG, X., DARYANTO, S., ZHANG, X., WANG, Y., 2018. Factors influencing soil moisture in the Loess Plateau, China: a review. Earth Environ. Sci. Trans. R. Soc. Edinb. 109, 501–509. https://doi.org/10.1017/S1755691018000671

Zheng, J., Jin, X., Li, Q., Lang, J., Yin, X., 2024. Soil moisture variation and affecting factors analysis in the Zhangjiakou–Chengde district based on modified temperature vegetation dryness index. Ecol. Indic. 168, 112775. https://doi.org/10.1016/j.ecolind.2024.112775

Zhu, S., Cui, N., Jin, H., Jin, X., Guo, L., Jiang, S., Wu, Z., Lv, M., Chen, F., Liu, Q., Wang, M., 2024. Optimization of multi-dimensional indices for kiwifruit orchard soil moisture content estimation using UAV and ground multi-sensors.

Agric. Water Manag. 294, 108705. https://doi.org/10.1016/j.agwat.2024.108705

Zsebő, S., Bede, L., Kukorelli, G., Kulmány, I.M., Milics, G., Stencinger, D., Teschner, G., Varga, Z., Vona, V., Kovács, A.J., 2024. Yield Prediction Using NDVI Values from GreenSeeker and MicaSense Cameras at Different Stages of Winter Wheat Phenology. Drones 8, 88. https://doi.org/10.3390/drones8030088

## 9. List of Articles that published as first Author or co-author:

1. Alahmad, T., Nyéki, A., Neményi, M. (2025*). Soil Moisture Content Prediction Using Gradient Boosting Regressor (GBR) Model: Soil-Specific Modeling with Five Depths*. Applied Sciences, 15(11), 5889. https://doi.org/10.3390/app15115889 (Q2, Impact Factor: 3.7)
2. Alahmad, T., Neményi, M., Széles, A., Ali, N., Hijazi, O., & Nyéki, A. (2025). *Spatiotemporal prediction of soil moisture content at various depths in three soil types using machine learning algorithms.* Frontiers in Soil Science https://doi.org/10.3389/fsoil.2025.1612908 (Q1, Impact Factor: 3.7)
3. Alahmad, T., Nyéki, A., Neményi, M. (2025*). Explainable XGBoost-based models of root-zone soil moisture profiling using coupled Sentinel-2 and IoT data*. Discover Applied Sciences https://doi.org/10.1007/s42452-026-08673-3 (Q2, Impact Factor: 2.8)

4. Alahmad, T., Neményi, M., Nyéki, A. (2023). *Applying IoT Sensors and Big Data to Improve Precision Crop Production: A Review.* Agronomy, 13(10), 2603. https://doi.org/10.3390/agronomy13102603
5. Alahmad, T., Neményi, M., Nyéki, A. (2024). *Soil Moisture Content Prediction in Loam Soil with RFR Model*. Acta Agronomica Óváriensis, 65(2), 43-56. https://doi.org/10.17108/ActAgrOvar.2024.65.2.43
6. Alahmad, T., Neményi, M., Nyéki, A. (2025*). Soil moisture prediction with machine learning: combining vegetation indices and meteorological data.* Precision agriculture '25, Brill. https://doi.org/10.1163/9789004725232_172
7. Alahmad, T., Neményi, M., Nyéki, A. (2025). *Enhancing Agricultural Sustainability using AI-Driven Soil Moisture Modeling: A Soil-Type and Depth Approach with SHAP Interpretability*. Acta Agronomica Óváriensis, https://doi.org/10.17108/ActAgrOvar.2025.66.2.5
8. Nyéki, A., Alahmad, T., Arshad, S., Gombkötő, N., Neményi, M., Mirzaei, M., Szabó, S., Endre, H., Al-Dalahmeh, M., Mohammed, S. (2025). *Real-time monitoring of ammonia emissions from cereal crops using LoRaWAN-based sensing technology*. Nature, Scientific Reports https://doi.org/10.1038/s41598-025-31661-3
9. Ambrus, B., Teschner, G., Kovács, AJ., Neményi, M., Helyes, L., Pék, Z., Takács, S., Alahmad, T., Nyéki, A. (2024). *Field-grown tomato yield estimation using point cloud segmentation with 3D shaping and RGB pictures from a field robot and digital single lens*

*reflex cameras*. Heliyon 10.20 (2024). DOI: https://doi.org/10.1016/j.heliyon.2024.e37997

10. Neményi, M., Ambrus, B., Teschner, G., Alahmad, T., Nyéki, A., Kovács, AJ. (2023). *Challenges of ecocentric sustainable development in agriculture with special regard to the internet of things (IoT), an ICT perspective*. Agricultural Engineering Sciences Journal, online pp 1-10 (2023). https://doi.org/10.1556/446.2023.00099
11. Nyéki, A., Daróczy, B., Neményi, M., Ambrus, B., Gergely, T., and Alahmad, T. (2025). *Predicting maize growth and biomass: Integrating gradient boosted trees with sentinel images and IoT*. Agricultural Engineering Sciences Journal DOI: https://doi.org/10.1556/446.2025.00202
12. Ali, N., Nyéki, A., Jalloul, A., Alahmad, T. (2024). *The Effect of Ascorbic Acid on Salt Tolerance and Seedling Performance in Triticum durum Defs. 'Douma 3' Under Salinity Stress in Syria*. Agronomy, 14(12), 2982; DOI: https://doi.org/10.3390/agronomy14122982

## 10. Appendix:

**Table A1.** Hyperparameter Search Ranges and Final Selected Values per Model and Article.

| **Article** | **Model** | **Parameter** | **Search Range** | **Final Selected Value** |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| **Article 1** | RFR | n_estimators | 10–200 (step 10) | 200 |
| **Article 1** | XGBoost | n_estimators | 100, 200, 300 | Best via CV |
| **Article 1** | XGBoost | learning_rate | 0.05, 0.10, 0.20 | Best via CV |
| **Article 1** | XGBoost | max_depth | 3, 5, 7 | Best via CV |
| **Article 1** | LSTM | Units (×2 layers) | Fixed | 50 |
| **Article 1** | LSTM | Epochs | Fixed | 1500 |
| **Article 1** | LSTM | Batch size | Fixed | 32 |
| **Article 1** | LSTM | Optimizer | Fixed | Adam |
| **Article 2** | GBR | n_estimators | 100, 200 | Best via CV (per soil type) |
| **Article 2** | GBR | learning_rate | 0.01, 0.05 | Best via CV (per soil type) |
| **Article 2** | GBR | max_depth | 4, 6 | Best via CV (per soil type) |
| **Article 2** | GBR | min_samples_split | 2, 4 | Best via CV (per soil type) |
| **Article 3** | XGBoost | n_estimators | 100, 200, 300 | 100 |
| **Article 3** | XGBoost | max_depth | 3, 5, 7 | 5 |

| | | | | |
|---|---|---|---|---|
| **Article 3** | XGBoost | learning_rate | 0.05, 0.10, 0.20 | 0.1 |
| **Article 3** | XGBoost | subsample | 0.7, 0.85, 1.0 | 1.0 |

frontiers | Frontiers in Soil Science

TYPE Original Research
PUBLISHED 30 September 2025
DOI 10.3389/fsoil.2025.1612908

Check for updates

OPEN ACCESS

EDITED BY
Jiban Shrestha,
Nepal Agricultural Research Council, Nepal

REVIEWED BY
Anurag Vidyarthi,
Graphic Era University, India
Yulin Cai,
Shandong University of Science and Technology, China

*CORRESPONDENCE
Tarek Alahmad
alahmad.tarek@sze.hu

RECEIVED 16 April 2025
ACCEPTED 09 September 2025
PUBLISHED 30 September 2025

CITATION
Alahmad T, Neményi M, Széles A, Ali N, Hijazi O and Nyéki A (2025) Spatiotemporal prediction of soil moisture content at various depths in three soil types using machine learning algorithms.
*Front. Soil Sci.* 5:1612908.
doi: 10.3389/fsoil.2025.1612908



# Spatiotemporal prediction of soil moisture content at various depths in three soil types using machine learning algorithms

Tarek Alahmad[1*], Miklós Neményi[1], Adrienn Széles[2], Nour Ali[3], Omar Hijazi[4] and Anikó Nyéki[1]

[1]Department of Biosystems Engineering and Precision Technology, Albert Kázmér Mosonmagyaróvár Faculty of Agricultural and Food Sciences, Széchenyi István University, Mosonmagyaróvár, Hungary, [2]Institute of Land Use, Engineering and Precision Farming Technology, Faculty of Agricultural and Food Sciences and Environmental Management, University of Debrecen, Debrecen, Hungary, [3]Laboratory of Crop Production and Multiplication, Field Crops Research Department, Agricultural Engineering Faculty, Damascus University, Damascus, Syria, [4]A Chair of Wood Science, Technical University of Munich, Freising, Germany

**Introduction:** Accurate prediction of soil moisture content (SMC) is crucial for agricultural systems as it affects hydrological cycles, crop growth, and resource management. Considering the challenges with prediction accuracy and determining the effect of soil texture, depth, and meteorological data on SMC variation and prediction capability of the used models, this research has been conducted.

**Methods:** Three machine learning (ML) models—random forest regression (RFR), eXtreme gradient boosting (XGBoost), and long short-term memory (LSTM)—were developed to predict SMC in three soil types (loam, sandy loam, and silt loam) at five depths of 5, 20, 40, 60, and 80 cm. The dataset was collected during the maize season in 2023, encompassing meteorological parameters collected using Internet of Things (IoT)-based sensors and SMC data calculated using the gravimetric method.

**Results:** The results showed variations in SMC in all studied soil types and depths, with silt loam exhibiting the highest variation in SMC. RFR demonstrated high accuracy at different depths and soil types, particularly in loam soil, at a depth of 80 with a root mean square error (RMSE) value of 0.89 and a mean absolute error (MAE) value of 0.74, and in silt loam at 40 cm depth with an RMSE value of 0.498 and an MAE of 0.416. LSTM performed effectively at shallower and moderate depths (60 and 20 cm) with RMSE values of 0.391 and 0.804 and MAE values of 0.335 and 0.793, respectively. In sandy loam soil at 5 cm depth, XGBoost displayed minimal errors and robust performance at the same depths with higher accuracy, achieving an RMSE of 0.025 and an MAE of 0.159. Analysis of training and validation loss revealed that the LSTM model stabilized and improved with more epochs, showing a more consistent decrease in MSE, while RFR and XGBoost exhibited higher performance with increased model complexity, shown in low MSE and RMSE values. Comparisons between measured and predicted SMC% values demonstrated the models' effectiveness in capturing soil moisture dynamics. Furthermore, feature importance analysis revealed that solar radiation and precipitation were the most influential predictors across all models, offering critical insights into dominant environmental drivers of soil moisture variability.

**Discussion:** By providing precise SMC predictions across different spatial and temporal scales, this study underscores the value of ML models for SMC prediction, which could have implications for improving irrigation scheduling, reducing water wastages, and enhancing sustainability.

KEYWORDS

machine learning, soil moisture content, RFR, LSTM, XGBoost, spatio-temporal prediction

## 1 Introduction

Current farming practices have already exceeded the Earth's carrying capacity (1). Therefore, the primary challenge is to enhance productivity and sustainably feed the world without depleting natural resources, particularly water, which is vital for crop production. Hence, it is imperative to implement more sustainable water management practices (2). With the increased demand for agricultural water resources, real-time monitoring of soil moisture is essential for developing a realistic irrigation schedule, leading to better water management (3, 4), and enhancing crop production (5). Soil moisture content (SMC) is a critical factor influencing farm yields and hydrological cycles, providing essential information on available water for vegetation growth requirements (6). Thus, understanding the spatial and temporal variation of SMC is crucial for various applications such as predicting floods, droughts (7), and forest fires, as well as for environmental and agricultural research (8, 9). Accurately predicting SMC is essential for the rational use and management of water resources (10, 11). However, the complex interplay of factors affecting soil water content makes prediction challenging, especially in spatiotemporal dynamics (12–14). Currently, direct methods for determining soil moisture that measure SMC directly from the soil samples include oven drying methods (both gravimetric and volumetric), while all automated systems for estimating soil moisture are classified as indirect methods (15). The gravimetric method is widely regarded as the most reliable and robust technique for estimating soil moisture. It quantifies soil moisture as the mass of water in a sample divided by the mass of dry soil, typically expressed in [kg/kg]. However, for comparative studies in earth sciences, it is often represented as a percentage (15). The gravimetric method has been utilized to validate data collected using indirect techniques (16, 17).

Existing prediction models face challenges related to accuracy, generalization, and processing multiple features, necessitating improvements in performance. Conventional prediction techniques frequently employ neural networks, linear regression, and empirical formulas (18, 19). Recently, advancements in sensor measuring technologies have enabled researchers to collect extensive, continuous, and reasonably accurate data at *in situ* monitoring sites (20, 21). Artificial intelligence-based models leveraging artificial neural networks (ANNs) represent a significant leap forward in machine learning (ML), enabling better extraction of hidden patterns within big data (22). Integrating artificial intelligence with prediction models enhances model performance and accuracy (23). However, current ANN models for SMC prediction tend to focus mainly on single data feature extraction, overlooking spatial and temporal factors, which limits the accuracy of predictions. Furthermore, existing soil moisture content (SMC%) prediction models often only forecast the average surface layer (0–20 cm) or a single depth, rendering them ineffective for practical agricultural irrigation decisions.

Various AI-based techniques, including random forest regression (RFR), support vector machine (SVM), extreme gradient boosting regression (XGBR), and CatBoost gradient boosting regression (CBR), have been utilized to overcome these limitations and more accurately forecast SMC (24–26). In their study, Ren et al. (27) utilized XGBoost to estimate soil moisture, demonstrating superior performance compared to other techniques with a correlation coefficient of 0.69 and an accuracy of 88%. The primary predictors were air relative humidity, maximum air temperature, and total precipitation. Consideration of soil characteristics, particularly during the 2022 drought, led to increased prediction accuracy. Additionally, Alibabaei et al. (28) discussed the use of deep learning algorithms, specifically long short-term memory (LSTM) and bidirectional LSTM (BLSTM), to model daily reference evapotranspiration and SMC for agricultural decision support systems. Their models achieved $R^2$ values between 0.96 and 0.98 and mean square error values between 0.014 and 0.056, with LSTM yielding the most favorable results. Furthermore, an analysis was conducted to assess how the loss function impacted the performance of the proposed models, revealing that the model employing mean square error as its loss function outperformed other models.

Considering the challenges of enhancing soil moisture prediction accuracy, and the lack of soil-specific and depth-specific modeling (29, 30), and to evaluate the efficiency of the dataset collected by Internet of Things (IoT)-based sensors in the models' training, this research aims to study how the soil type, depth, and meteorological data affect SMC variations; to develop and refine AI-based ML models including RFR, eXtreme gradient boosting (XGBoost), and LSTM, which consider effective models in

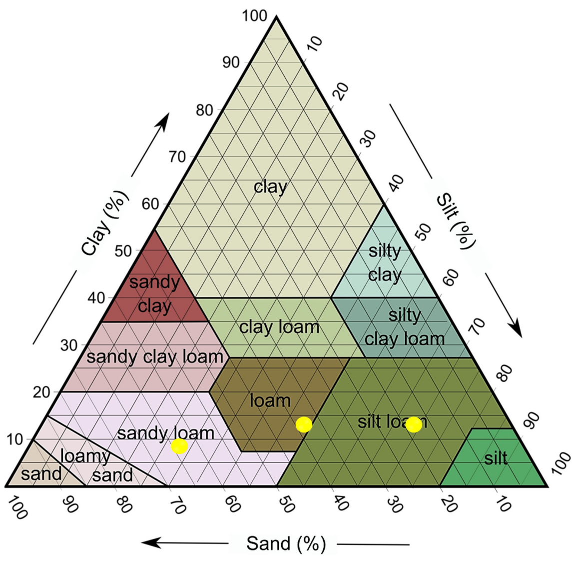



FIGURE 1
Soil texture classes (silt, clay, and sand) for each studied soil.

SMC prediction due to its ability to capture linear and non-linear relations between studied factors (29, 31); to predict SMC in various soil types and depths by combining meteorological variables with SMC data; and to evaluate models' performance using several performance metrics. The results will provide significant insights into the models' applicability for SMC prediction to enhance water use efficiency and enhance sustainability in agriculture by improving decision-making in irrigation management and enhancing sustainability.

# 2 Materials and methods

## 2.1 Experimental site and sensor setup

The research was conducted between June and October 2023 during the maize vegetation season at the 23-ha field, an agricultural site equipped with IoT sensors for data collection at Széchenyi István University in Mosonmagyaróvár, Hungary (32, 33). The area had an average annual precipitation of 580 mm in 2023, and an average annual temperature of 11.2°C, and during the maize vegetation season, the precipitation amount was approximately 400 mm with temperatures ranging between 18.5°C and 21.3°C. The soil types identified according to the USDA soil classification taxonomy (34) utilizing the soil texture triangle include loam, silt loam, and sandy loam (35) (Figure 1). The terrain has a slight slope of 5% with elevation varying between 133 and 138 m (23), soil pH ranged between 7.12 and 7.8, and other soil properties are shown in Table 1. Data collection from the field by sensors was carried out at 10- to 15-min intervals using LoRaWAN and Narrowband Internet of Things (NB-IoT).

## 2.2 Gravimetric technique for soil moisture content measurements

A total of 405 soil samples were collected from depths of 5, 20, 40, 60, and 80 cm on three soil types: loam, silt loam, and sandy loam, with 135 samples each. Sampling locations were determined according to the USDA's recommended procedures (36), ensuring representative and systematic coverage across the three soil textures, loam, sandy loam, and silt loam (Figure 2A). Each sampling point shown in the figure represents measurements taken at all five specified depths. The samples were collected using an auger on 10 different dates every 2 weeks (June 1–15, July 2–15, August 4–21, September 5–20, and October 3–18). They were collected from the field in the same sensor's locations from a 1-m² circle around the sensor. The samples were saved in plastic bags to maintain their moisture content before being moved to the laboratory. After that, they were placed in pre-weighed containers and weighed using a digital scale to record their initial weights. The samples were then transported to the laboratory and oven-dried at 105°C for 24 h (37, 38). After drying, the samples were weighed again to obtain their post-drying weights. Finally, the empty weights of the soil moisture containers were measured. SMC based on dry weight was calculated using Equation 1:

$$\theta = \frac{Mw}{Md} \times 100 \tag{1}$$

where:

- $\theta$ = soil moisture content (%)
- $Mw$ = mass of water (g) = (wet weight – dry weight)
- $Md$ = mass of dry soil (g)

TABLE 1 Soil property averages and range values in all studied soil types and depths with mean SMC values and standard deviation.

| Soil type | Number of samples | pH | Sand range | Silt range | Clay range | Mean SMC% range | SD of SMC% range | Bulk density average |
|---|---|---|---|---|---|---|---|---|
| Loam | 135 | 7.61 | 40.99–49.01 | 40.91–45.97 | 10.10–14.50 | 12.61–16.19 | 3.10–4.94 | 1.54 |
| Sandy loam | 135 | 7.71 | 53.8–59.99 | 32.04–36.27 | 8.02–9.92 | 12.43–14.98 | 2.15–5.27 | 1.57 |
| Silt loam | 135 | 7.45 | 12.36–15.31 | 63.70–70.19 | 17.40–21 | 14.99–18.27 | 2.62–7.87 | 1.52 |

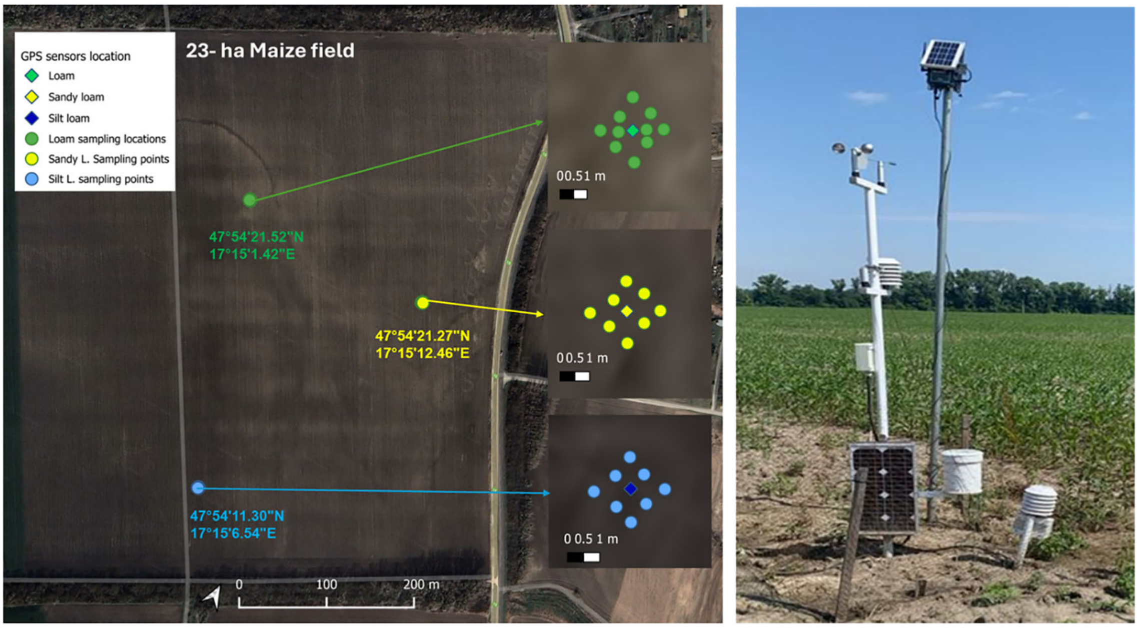



FIGURE 2
**(A)** Sensor station's locations at the field with soil types, GPS coordinates, and spatial distribution of sensor stations and corresponding sample locations within each identified soil type. **(B)** IoT sensor stations based in the field (wind sensor, precipitation sensor, solar radiation sensor, humidity sensor, and temperature sensor).

## 2.3 Data preparation and processing

The dataset was carefully prepared to address incorrect, missing, or hard-to-interpret values that could lead to overfitting or inaccurate, or deceptive results from the algorithm. Data preparation, including data transformation and statistical testing, was carried out using Python (version 3.10.12) (39) to ensure that there were no missing or zero values and to verify that the data types entering the model were accurate and consistent; in addition, basic statistical analyses including variation analysis and means calculations were conducted using Python. Additionally, dates and soil types were defined in the model to replace numerical values with appropriate date formats and soil type names. Following the data preparation, three ML algorithms—RFR, LSTM, and XGBoost—were employed to predict SMC in all soil types and depths. The dataset, which contained over 6,500 records collected from SMC measurements and meteorological data, was split into two parts: 80% for model training and 20% for testing. Each algorithm utilizes six inputs (SMC%, precipitation, humidity, temperature, solar radiation, and wind speed).

## 2.4 AI-based models

### 2.4.1 Random forest regression

The RFR is an approach based on classification trees. Belgiu and Drăguţ (40) described the primary phases of the RFR algorithm as follows:

1. Randomly selecting a subset of the original training set with replacement.
2. Using the subset to build the regression tree model.
3. Averaging the results of all trees to produce the final prediction.

The key hyperparameters used were as follows:

- n_estimators: defines the number of trees.
- random_state = 42: ensures reproducibility.
- max_leaf_nodes: limits the number of terminal nodes or leaves in each tree, which can reduce overfitting.

In general, the samples were divided, with approximately two-thirds of the dataset (in-bag data) for training samples and the remaining part for validation samples [out-of-bag (OOB) data] (Figure 3) (41). The RFR model was implemented using Python with the Pandas, NumPy, Matplotlib, and Scikit-learn libraries.

### 2.4.2 Long short-term memory architectures

LSTM networks are a type of recurrent neural network (RNN) architecture designed to address the vanishing gradient problem and effectively capture long-range dependencies in sequential data. In the RNN model, the output from the previous step is utilized as input for the next step. This model is well-suited for predicting SMC as it can model time-series data, such as meteorological variables, and their impact on soil moisture trends.

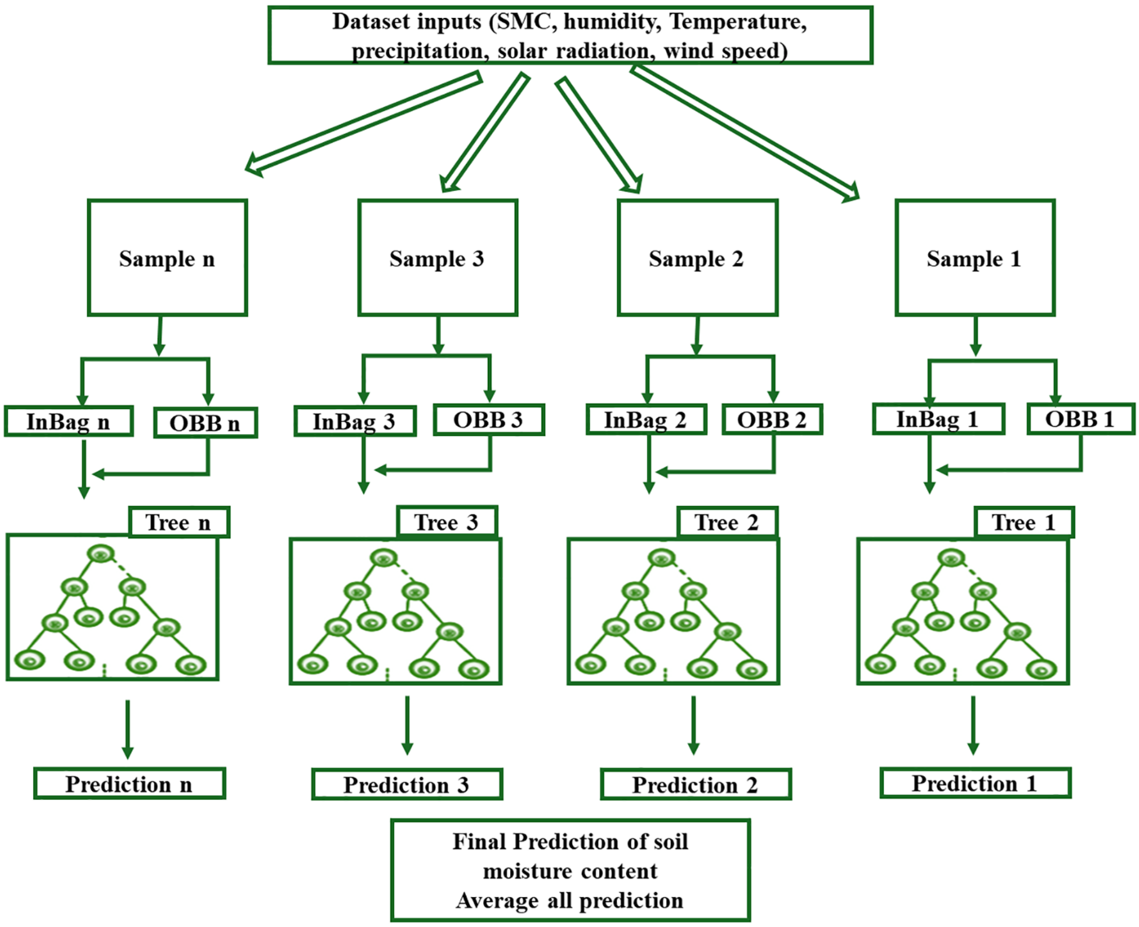



FIGURE 3
Random forest regression (RFR) model architecture.

The LSTM model used in this study comprises two layers (Figure 4): The first LSTM layer has 50 units and return_sequences=True, allowing the output to be passed to the second LSTM layer. The second LSTM layer also has 50 units. The hyperparameters utilized were epochs (number of training cycles) and batch_size. The optimizer used was Adam, with a mean_squared_error loss function. The output was generated by a dense layer with a single unit. The input data were normalized using the MinMaxScaler, ensuring all values were scaled between 0 and 1. The data were reshaped into a 3D array (samples, time steps, and features) before being input into the LSTM layers. The LSTM model was constructed and trained using Python with Pandas, NumPy, Matplotlib, and Scikit-learn libraries, in addition to TensorFlow for the LSTM layers.

### 2.4.3 XGBoost

The XGBoost technique, developed by Chen and Guestrin (42), is a powerful method for supervised learning that can handle regression and classification issues effectively. The XGBR algorithm is widely used in data mining due to its quick execution speed achieved by parallelization, out-of-core computation, and cache optimization. Data scientists appreciate its adaptability to different environments and excellent performance in small-scale data analysis. In a recent study, XGBoost was used to predict SMC using meteorological and soil moisture input features. The model's hyperparameters were fine-tuned using a grid search to identify the best configuration, including n_estimators, learning_rate, and max_depth. The optimal configuration was selected using fivefold cross-validation to ensure the model's ability to generalize well in studied soil types and depths. The flowchart for the XGBoost model is summarized in Figure 5. The XGBoost model was developed and trained using Python, along with libraries such as Pandas, NumPy, Matplotlib, and the XGBoost library.

## 2.5 Performance evaluation measures

Four indicators were calculated to quantify the performance of the different models.

Mean square error (MSE): The MSE is the average of the squared differences between projected and observed. In other words, the MSE represents the variance of the mistake (43). It is calculated using Equation 2:

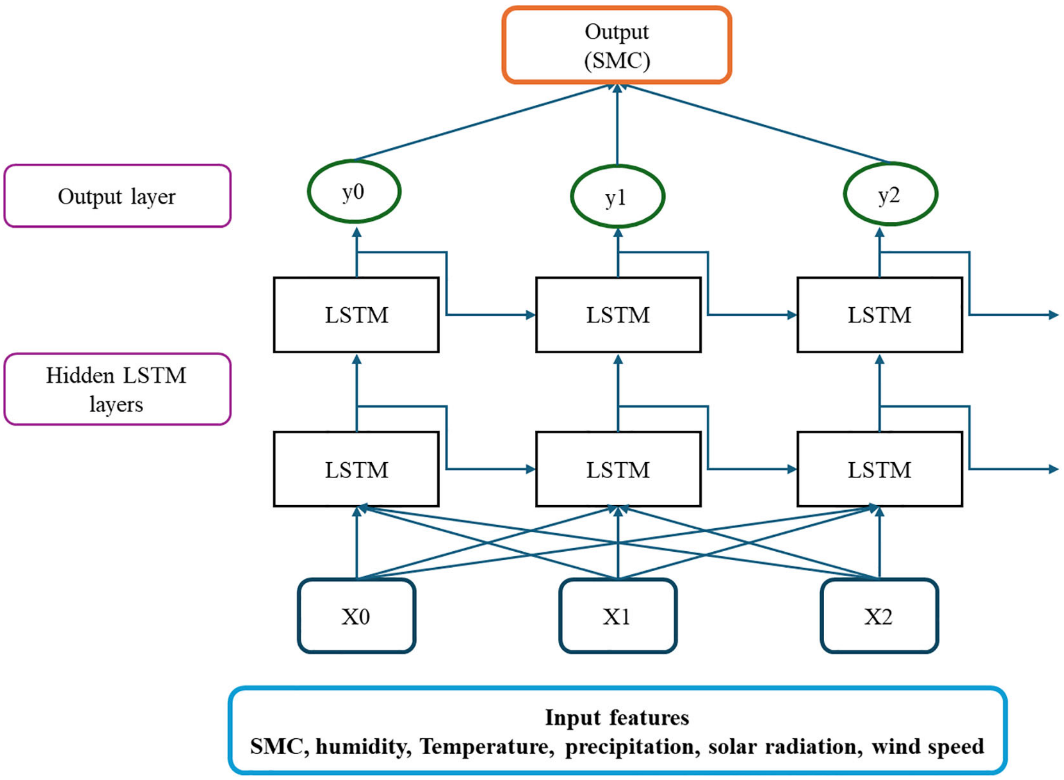



FIGURE 4
Diagram of LSTM layers.

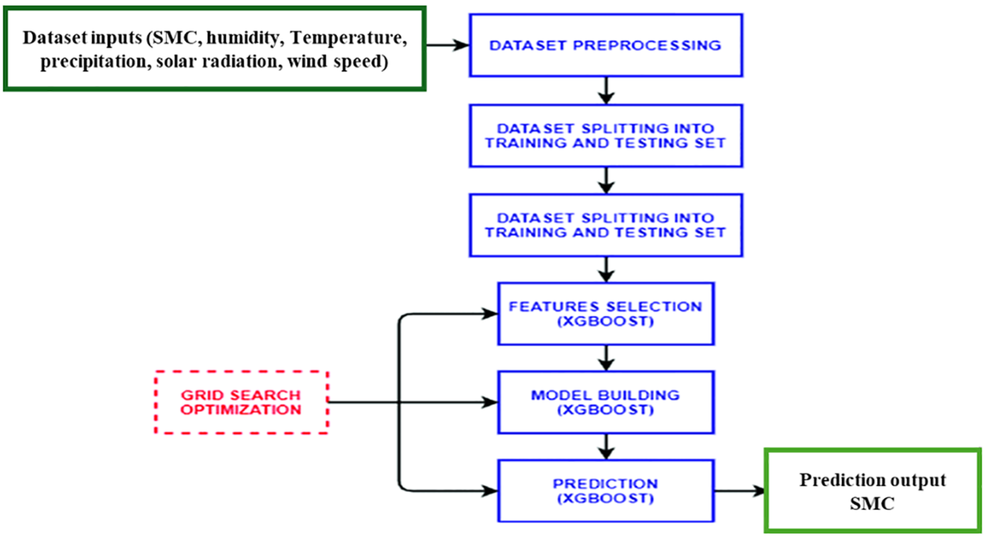



FIGURE 5
Flowchart of the XGBoost model.

$$\mathrm{MSE} = {}^{1}/_{n}\sum_{t=1}^{n}(y_t - \hat{y}t)^2 \tag{2}$$

where $y_t$ is the ground truth, and $\hat{y}$ represents the mean of the predicted values.

Root mean square error (RMSE): It captures the typical difference between predicted and observed values (43). It is calculated by taking the square root of the MSE. Unlike MSE, RMSE is measured in the same units as the original data, making it easier to interpret. Since

RMSE considers squared differences, it emphasizes larger deviations from predictions. This makes RMSE valuable when minimizing significant discrepancies is critical (Equation 3).

$$\mathrm{RMSE} = \sqrt{{}^{1}/_{n}\sum_{t=1}^{n}(y_t - \hat{y}\,t)^2} \quad (3)$$

Mean absolute error (MAE) (43): It calculates the average of the absolute differences between predictions and actual observations (Equation 4):

$$\mathrm{MAE} = {}^{1}/_{n}\sum_{t=1}^{n}|y_t - \hat{y}\,t| \quad (4)$$

*R*-squared ($R^2$): It is an indicator of statistical significance that compares the variation explained by the model to the total variance. Higher $R^2$ indicates less difference between real and projected values. It is calculated using Equation 5:

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}\,t)^2}{\sum_{t=1}^{n}(y_t - \underline{y})^2} \quad (5)$$

where $y_i$ is the predicted value, $y_t$ is the ground truth, and $\hat{y}$ represents the mean of the predicted values.

## 2.6 Feature importance analysis

Feature importance analysis was conducted for each ML model to evaluate and compare the contribution of studied meteorological variables to SMC prediction. For RFR, the Gini-based significance approach from scikit-learn was used to calculate the average reduction in error supplied by each feature across the forest (44, 45). In the LSTM model (an RNN model implemented in TensorFlow/Keras), feature importance is not inherently provided; thus, a model-agnostic permutation importance approach was applied and calculated by assessing the increase in MSE after shuffling each predictor variable in the validation dataset. Finally, with XGBoost, the built-in gain-based feature importance measure was utilized to calculate the cumulative improvement in the model's accuracy when splitting by each feature (46). All analyses were conducted in Python and all feature importance results were computed separately for each soil type, depth, and model.

# 3 Results and discussion

## 3.1 Statistical analysis results for the three soil types

After analyzing 405 SMC% measurements in three types of soil and five depths over the maize vegetation season, the results shown in Figure 6 demonstrate that silt loam soil has the highest mean SMC, ranging from 14.99% to 18.27% due to its balanced sand, silt, and clay particle texture, which maximizes water retention (47), with the highest SMC (27.23%) recorded at 60 cm depth and the highest

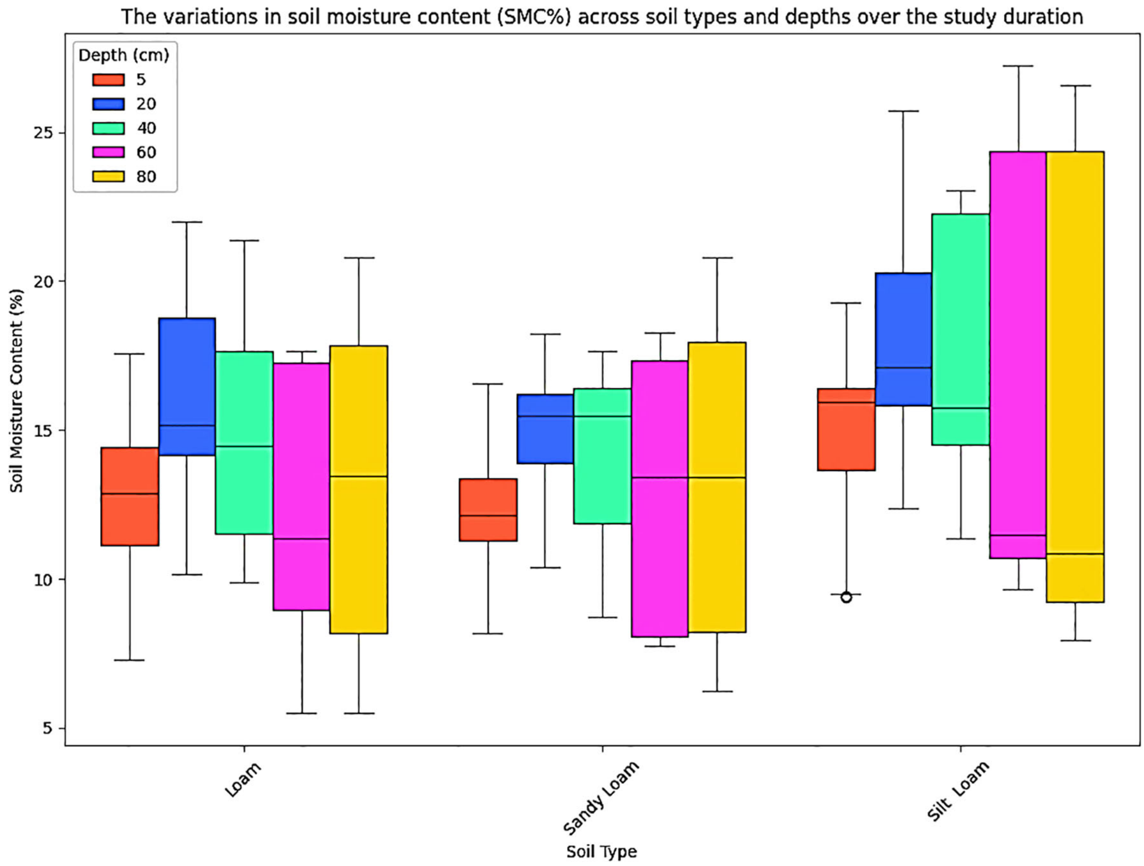



FIGURE 6
Soil moisture content variation in three soil types at five different depths.

variability in the deeper layers (60 and 80 cm). The mean moisture content of loam soil varied from 12.61% to 16.19% at the five depths, with larger variability shown at deeper layers (60 and 80 cm), and maximum SMC value was at 20 cm depth with 22.1%. Sandy loam soil had a lower mean moisture content than loam soil due to its higher sand content, which decreased water retention, and the mean SMC ranged between 12.43% and 14.98%, with the highest SMC (20.77%) recorded at 80 cm depth. The low moisture content of sandy loam resulted in little variation among the five depths due to the high sand content, which reduces the water holding capacity and effect on soil water movement and dynamics through the soil layers (48). The highest SMC values and variability in deeper soil layers across all soil types are attributed to higher water retention capacity at depth compared with surface layers in the studied area (49). This variation in SMC% highlights the need to consider soil types and depth when monitoring soil moisture dynamics, as these factors had a significant impact on SMC variation (50). These results could be used in optimizing irrigation management, which will reduce water waste and enhance sustainability.

## 3.2 Validation and predictive performance in studied soil types and depths

The performance of the three models (RFR, XGBoost, and LSTM) in predicting SMC in the studied soil types and depths resulted in several findings (Table 2).

Among the three models, the RFR model consistently demonstrated the most robust and generalizable performance in all studied soil types and depths. For instance, in loam soil, at depths of 80 and 20 cm, the model achieved the lowest RMSE values of 0.89 and 0.98 and MAE values of 0.74 and 0.89, respectively. Figure 7 highlighted the model's ability to highly accurately predict SMC at different layers in loam soil. In sandy loam soil, RFR performed best at depths of 20, 40, and 5 cm with RMSE values of 0.43, 0.51, and 0.54, respectively, and MAE values of 0.39, 0.49, and 0.51, respectively. Additionally, the model performed accurately in silt loam soil at 40- and 60-cm depths with RMSE values of 0.49 and 0.52, and MAE values of 0.42 and 0.41, respectively. These results align with the results of Cheng et al. (10) and Draper et al. (51) and suggest that RFR accurately captures the variation in SMC in all studied soil types and depths (52).

On the other hand, the LSTM models also showed promising results, particularly at shallow to moderate depths (20, 40, and 60 cm) and at all soil types where the meteorological factors have more impact on SMC variation at these layers, while it performed less consistently compared with RFR. For instance, in loam soil, at depths of 60 and 20 cm, the model achieved the lowest RMSE values of 0.39 and 0.80 and MAE values of 0.34 and 0.79, respectively. Figure 7 highlighted the model's ability to accurately predict SMC at different layers in loam soil. In sandy loam soil, LSTM had the best performance at depths of 20 and 5 cm with RMSE values of 0.31 and 0.42, respectively, and MAE values of 0.22 and 0.38, respectively. In addition, in silt loam soil, the lowest errors were at 20- and 5-cm depths with RMSE values of 0.27 and 0.34, and MAE values of 0.21 and 0.36, respectively (Figure 7), aligning with the results of Filipović et al. (3) and Park et al. (53). These findings indicate that the LSTM model demonstrates promising performance in predicting SMC, and the ability to accurately predict SMC, especially at shallower to moderate depths where the impact of environmental factors is higher than deeper layers (28).

The XGBoost models that offered a high performance in capturing complex and nonlinear relations performed well in different soil types and depths especially the shallower depths. The model showed a high accuracy at 40- and 60-cm depths in loam soil with the lowest RMSE values of 0.422 and 0.638, and MAE values of 0.354 and 0.507, respectively (Figure 7). In sandy loam, the best errors were at depths of 40 and 5 cm where XGBoost had RMSE values of 0.099 and 0.159, and MAE values of 0.072 and 0.132, respectively, which considers higher performance in shallow depths compared with the results by Ren et al. (27), where the highest RMSE was 11.11 and MAE was 4.87. Similarly, in silt loam soil, it performed well at 5 and 20 cm with RMSE values of 0.249 and 0.256, and MAE values of 0.199 and 0.206, respectively. The acceptable threshold for RMSE value is 10% (54). During the experiment period, the models predicted SMC with RMSE values ranging from 0.24% to 0.9%. These low MSE and RMSE values, along with a high $R^2$ value, suggest that these models can effectively predict SMC as an alternative to classic empirical equations (55).

The variation in the measured parameters between deeper layers and upper layers is caused by different influencing factors. Shallower depths (5–20 cm) are directly affected by weather conditions such as precipitation and humidity, making it easier for the model to predict SMC accurately with high-quality data. On the other hand, deeper layers (40–80 cm) are influenced by factors like surface water dynamics, soil structure, and other soil properties, making it more complex for the model to accurately predict SMC in these layers (56, 57) and emphasizing the need to consider more features in model training for deeper soil layers.

The RFR exhibited consistent prediction accuracy with high $R^2$ values. In loam soil, RFR achieved $R^2$ values ranging from 0.820 to 0.971, with the highest $R^2$ at 80 cm depth. Similarly, in sandy loam soil, RFR performed well, with $R^2$ values ranging from 0.941 to 0.973, peaking at 20 cm depth. Moreover, in silt loam soil, the model had $R^2$ values ranging from 0.956 to 0.994, with the highest $R^2$ at 60 cm depth (Figure 8). These results align with the findings of Cheng et al. (10) and Zhao et al. (58) and demonstrate that the RFR model offered the best generalization in the studied soil types and depths, especially the deeper layers where the impact of meteorological factors is less, and SMC is affected by other factors such as soil properties and components (11). On the other hand, the LSTM model showed promising accuracy results, especially at shallower depths where the model achieved $R^2$ values in loam soil ranging from 0.852 to 0.987, with the highest $R^2$ found at 60 cm depth. In sandy loam soil, LSTM performed well with $R^2$ values ranging from 0.964 to 0.996, peaking at 80 cm depth. Similarly, in silt loam soil, $R^2$ values varied from 0.982 to 0.980 (Figure 8), with the best $R^2$ found at 60 cm depth; similar results were achieved in other research studies (28, 59). These findings indicate that LSTM models can accurately represent the temporal dynamics of SMC in various soil

10.3389/fsoil.2025.1612908

TABLE 2 Performance metrics evaluation results of the three used models across studied soil types and depths.

| Soil type | Depth (cm) | Model | MSE | RMSE | MAE | $R^2$ |
|---|---|---|---|---|---|---|
| Loam | 5 | RFR | 1.614 | 1.27 | 0.953 | 0.82 |
| Loam | 20 | RFR | 0.957 | 0.978 | 0.827 | 0.819 |
| Loam | 40 | RFR | 1.141 | 1.068 | 0.99 | 0.871 |
| Loam | 60 | RFR | 2.247 | 1.499 | 1.238 | 0.898 |
| Loam | 80 | RFR | 0.792 | 0.89 | 0.74 | 0.971 |
| Loam | 5 | LSTM | 1.322 | 1.15 | 0.811 | 0.852 |
| Loam | 20 | LSTM | 0.646 | 0.804 | 0.793 | 0.878 |
| Loam | 40 | LSTM | 1.149 | 1.072 | 0.615 | 0.87 |
| Loam | 60 | LSTM | 0.153 | 0.391 | 0.335 | 0.993 |
| Loam | 80 | LSTM | 0.343 | 0.585 | 0.496 | 0.987 |
| Loam | 5 | XGBoost | 1.472 | 1.213 | 0.843 | 0.835 |
| Loam | 20 | XGBoost | 1.768 | 1.33 | 1.023 | 0.865 |
| Loam | 40 | XGBoost | 0.178 | 0.422 | 0.354 | 0.98 |
| Loam | 60 | XGBoost | 0.407 | 0.638 | 0.507 | 0.982 |
| Loam | 80 | XGBoost | 0.45 | 0.671 | 0.482 | 0.983 |
| Sandy loam | 5 | RFR | 0.29 | 0.539 | 0.508 | 0.941 |
| Sandy loam | 20 | RFR | 0.189 | 0.434 | 0.39 | 0.967 |
| Sandy loam | 40 | RFR | 0.26 | 0.51 | 0.493 | 0.965 |
| Sandy loam | 60 | RFR | 0.658 | 0.811 | 0.716 | 0.957 |
| Sandy loam | 80 | RFR | 0.731 | 0.855 | 0.702 | 0.973 |
| Sandy loam | 5 | LSTM | 0.176 | 0.42 | 0.384 | 0.964 |
| Sandy loam | 20 | LSTM | 0.095 | 0.308 | 0.223 | 0.984 |
| Sandy loam | 40 | LSTM | 1.01 | 1.005 | 0.791 | 0.863 |
| Sandy loam | 60 | LSTM | 0.559 | 0.748 | 0.458 | 0.964 |
| Sandy loam | 80 | LSTM | 0.111 | 0.333 | 0.251 | 0.996 |
| Sandy loam | 5 | XGBoost | 0.025 | 0.159 | 0.132 | 0.951 |
| Sandy loam | 20 | XGBoost | 0.057 | 0.239 | 0.207 | 0.944 |
| Sandy loam | 40 | XGBoost | 0.01 | 0.099 | 0.072 | 0.998 |
| Sandy loam | 60 | XGBoost | 0.346 | 0.588 | 0.477 | 0.985 |

*(Continued)*

TABLE 2 Continued

| Soil type | Depth (cm) | Model | MSE | RMSE | MAE | $R^2$ |
|---|---|---|---|---|---|---|
| Sandy loam | 80 | XGBoost | 0.023 | 0.151 | 0.105 | 0.999 |
| Silt loam | 5 | RFR | 0.396 | 0.629 | 0.555 | 0.956 |
| Silt loam | 20 | RFR | 0.932 | 0.965 | 0.596 | 0.922 |
| Silt loam | 40 | RFR | 0.248 | 0.498 | 0.416 | 0.984 |
| Silt Loam | 60 | RFR | 0.272 | 0.521 | 0.411 | 0.994 |
| Silt loam | 80 | RFR | 1.145 | 1.07 | 0.801 | 0.982 |
| Silt loam | 5 | LSTM | 0.159 | 0.399 | 0.363 | 0.982 |
| Silt loam | 20 | LSTM | 0.073 | 0.271 | 0.21 | 0.975 |
| Silt loam | 40 | LSTM | 0.64 | 0.8 | 0.651 | 0.96 |
| Silt loam | 60 | LSTM | 0.901 | 0.949 | 0.682 | 0.98 |
| Silt loam | 80 | LSTM | 1.702 | 1.305 | 1.079 | 0.973 |
| Silt loam | 5 | XGBoost | 0.062 | 0.249 | 0.199 | 0.953 |
| Silt loam | 20 | XGBoost | 0.066 | 0.256 | 0.206 | 0.978 |
| Silt loam | 40 | XGBoost | 0.137 | 0.37 | 0.312 | 0.988 |
| Silt loam | 60 | XGBoost | 0.141 | 0.375 | 0.322 | 0.997 |
| Silt loam | 80 | XGBoost | 1.599 | 1.264 | 0.953 | 0.975 |

types and depths, particularly at shallow to intermediate depths, compared to other models (60). However, LSTM performance decreased as depth increased in all soil types.

In loam soil, XGBoost had $R^2$ values ranging from 0.835 to 0.983, with the highest $R^2$ at 40 cm depth, while in sandy loam soil, $R^2$ values ranged from 0.951 to 0.999, with the highest value at 40 cm depth. Similarly, in silt loam soil, $R^2$ values varied from 0.953 to 0.997, with the best $R^2$ found at 60 cm depth (Figure 8). These findings indicate that XGBoost models may accurately capture the complicated interactions between SMC and environmental variables in a variety of soil types and depths. However, XGBoost's performance decreased at deeper depths in loam and silt loam soils (27, 61). The results showed that the RFR model had the most consistent performance in all soil types and depths in comparison with XGBoost that performed best at shallower depths (5 to 40 cm) and LTSM in shallow to moderate layers (5, 20, 40, and 60 cm). In addition, RFR offered the best generalization capacity. These findings highlighted the need for further improvements for the models to increase accuracy in the deep oil layers where factors other than meteorological data have a significant impact on soil moisture dynamics.

## 3.3 The training and validation loss behaviors of models in studied soil types and depths

The learning curves of the LSTM model indicated a decrease in loss values (MSE) with increasing epochs, particularly between 600 and 800,

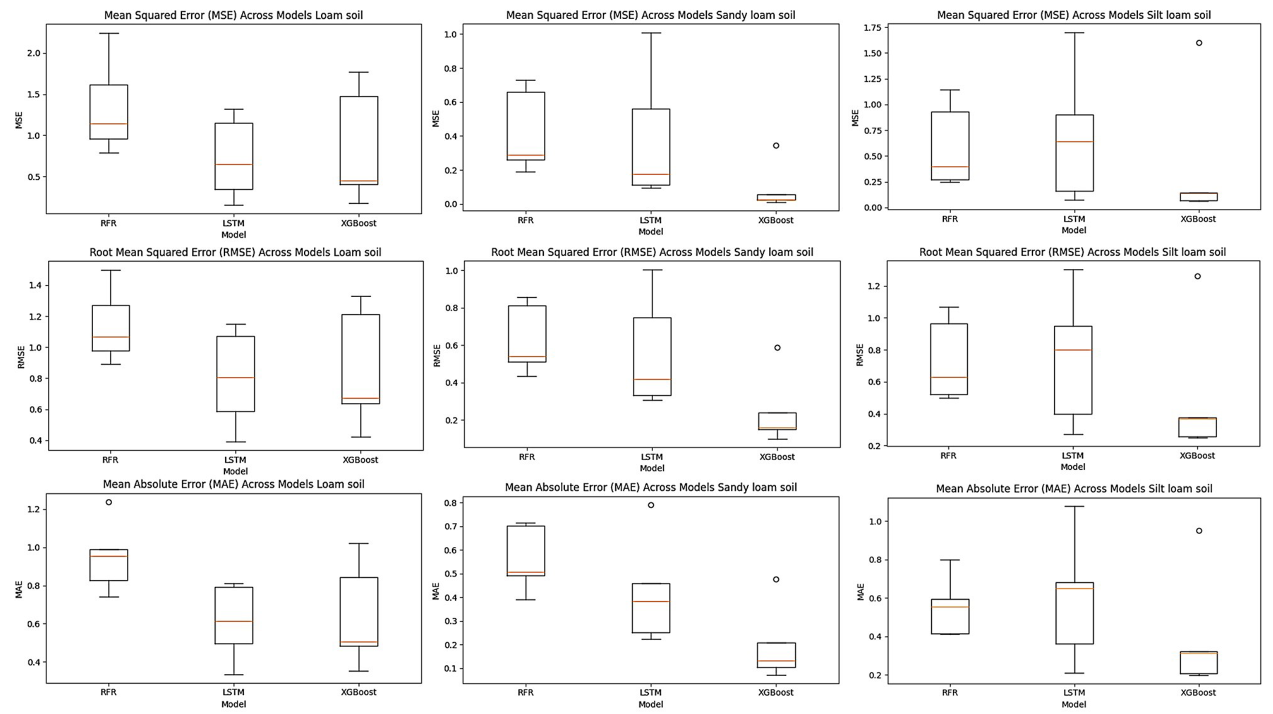



FIGURE 7
Predictive performance and validation results (MSE, RMSE, and MAE) for all soil types.

suggesting improved model generalization. However, there were slight variations in loss curves, indicating potential moderate overfitting at specific depths and soil types, such as 60 and 80 cm in loam soil (Figure 9) and 20, 40, and 80 cm in silt loam soil (Figure 10) (62). On the other hand, when the number of trees increased to 200, both the RFR and XGBoost models demonstrated decreasing loss values, showing improved model performance with increased complexity (63). However, deviations from this trend were observed, particularly in XGBoost models for sandy loam soil at 5 and 40 cm depth (Figure 11) and silt loam soil at 20 and 40 cm depth (Figure 10), where loss values slightly increased with the number of trees, indicating potential instability or poor performance under certain conditions. Overall, while all models showed improved performance with a higher number of epochs for LSTM (62) and greater model complexity for RFR and XGBoost (63), the LSTM model consistently outperforms and stabilizes in all studied soil types and depths.

## 3.4 Comparison of measured and predicted SMC% values

The comparison between measured and predicted SMC% values at different depths in all soil types provides insight into the models' predictive performance. The results show the model's ability to accurately represent underlying patterns and variability in SMC at various depths and soil types. The close agreement between measured and predicted values indicates that the modeling approaches accurately predict SMC% (see Figure 12), and the mean

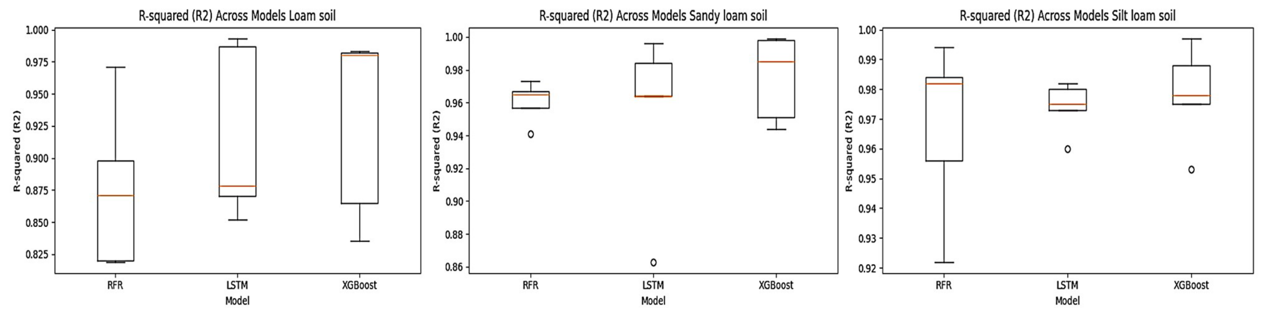



FIGURE 8
Accuracy performance evaluation of the RFR, LSTM, and XGBoost in all soil types.

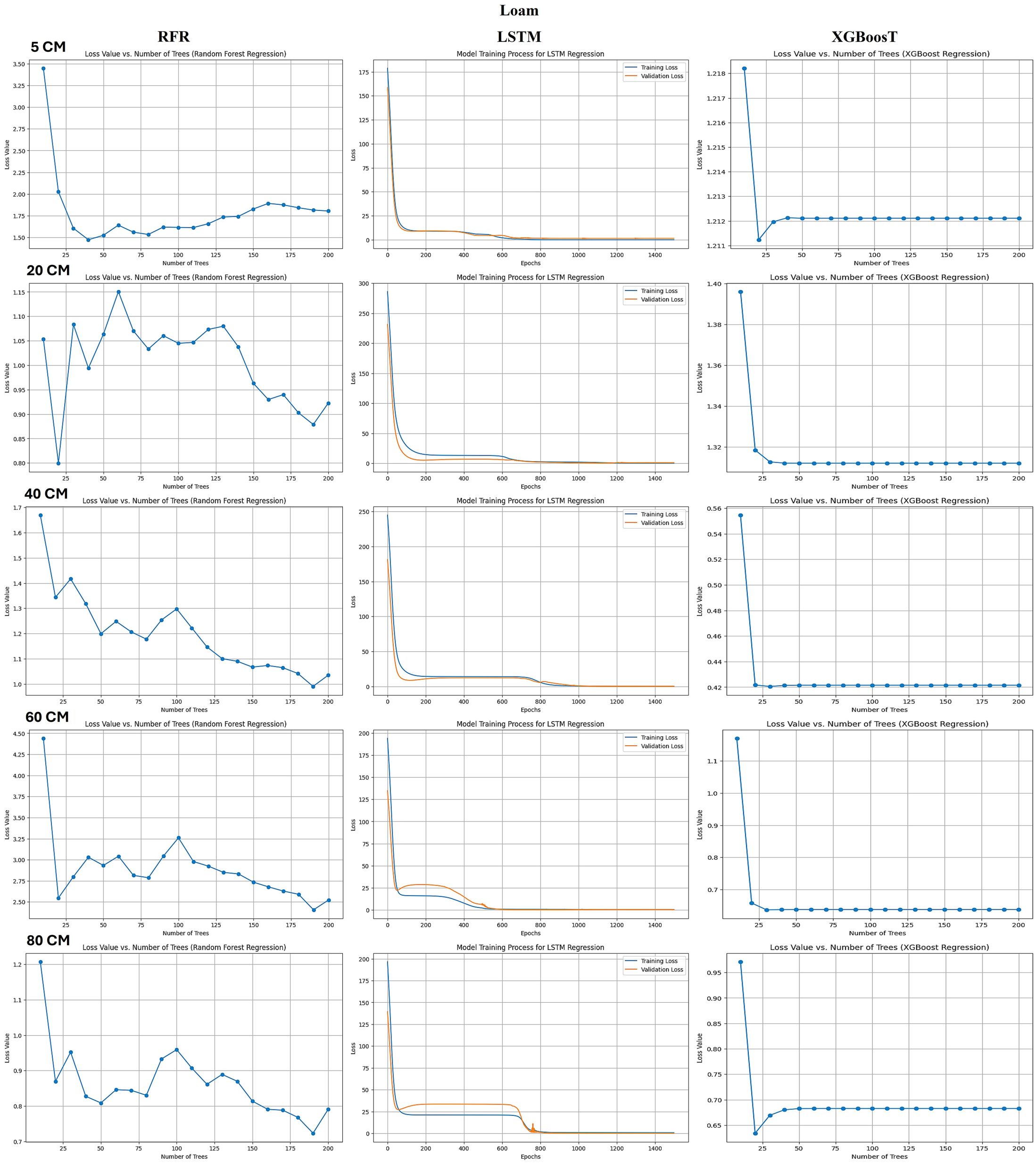



FIGURE 9
The loss value vs. the number of trees and epochs for RFR, XGBoost, and LSTM at various depths in loam soil.

measured and predicted values are provided in Appendix A. Compared with previous research by Alibabaei et al. (28) and Ren et al. (27), which developed different models to predict SMC, this research combines multiple soil layers' SMC prediction with three AI-based ML models in three soil types, with the use of gravimetric data for more accurate validation and learning; the use of *in situ* IoT sensors data and high temporal resolution allows for more robust performance of the model than the use of only satellite images for model training. Thus, the soil-specific and depth-specific modeling achieved in this research is vital for precision irrigation management and drought monitoring. Variations between measured and predicted values, especially at certain depths or soil types, suggest many potentials for model development especially for deep soil layers (52, 55). Enhancing models' generalizability and applicability could be done by considering extending the inputs with more variables such as different locations instead of only one location in this research, irrigations, evapotranspiration, vegetation indices from more vegetation seasons, and utilizing more soil properties such as soil texture and organic matter content.

## 3.5 Feature importance analysis results

The feature importance analysis revealed both common patterns and significant differences in how the three models prioritize the environmental variables. Overall, the studied meteorological data

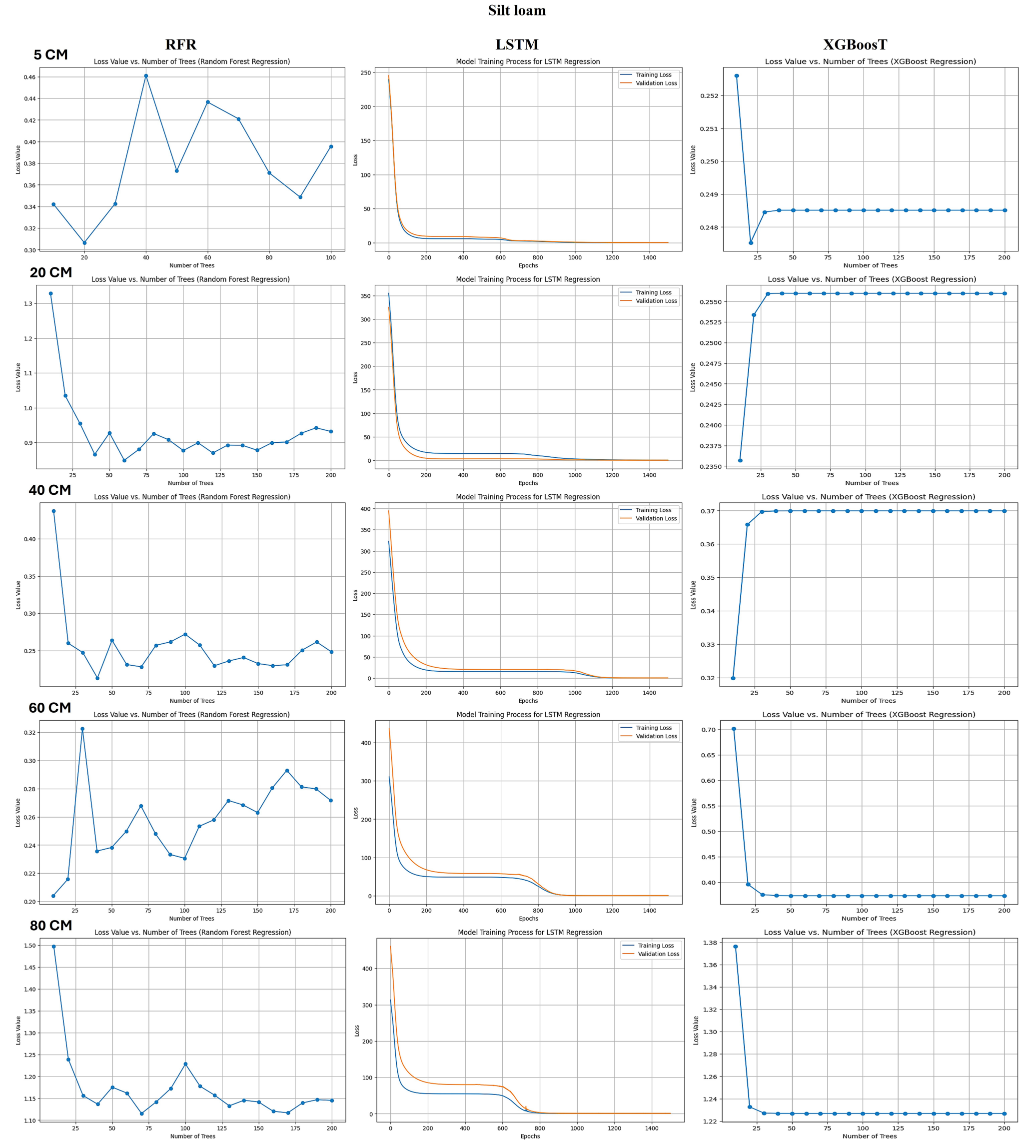



FIGURE 10
The loss value vs. the number of trees and epochs for RFR, XGBoost, and LSTM at various depths in silt loam soil.

have contributed to SMC prediction, but solar radiation consistently showed as the most effective predictor, particularly for SMC in deeper soil layers (40 to 80 cm), achieving average values ranging from 0.4 to 0.85 across three soil types and five depths with highest values of impact on silt loam with three models (0.8 with RFR, 0.79 with XGBoost, and 0.7 in LSTM) (Figure 13). This result demonstrates that evaporative demand caused by solar energy input is a dominating component driving soil moisture variability at depth, which is consistent with previous research indicating that solar radiation and temperature strongly influence soil water evaporation and drying (64, 65). On the other hand, precipitation had an important role in SMC prediction, especially at surface layer 5 cm in the three soil types with average values of 0.14 in RFR, 0.4 in XGBoost, and 0.25 in LSTM. These results emphasize the role of precipitation in increasing the SMC in surface layers (66). Humidity, temperature, and wind speed had varied importance among models, soil types, and depths. While analysis results showed the relatively high importance of humidity and wind speed in loam and silt loam soils, particularly at surface layers 5 and 20 cm with RFR and LSTM models, temperature had high importance value in sandy loam soil with the XGBoost model achieving values of 0.17 and 0.44 at 5- and 20-cm depths, respectively.

These variations represent each algorithm's learning method. LSTMs, which are designed to capture temporal patterns and lag effects, may prioritize characteristics such as temperature that represent long-term environmental influences. RFR and XGBoost, on the other hand, use threshold-based decision rules and prioritize features that bring quick advances in predictive accuracy, such as recent precipitation or current wind speeds. Despite these model-specific variations, solar radiation and precipitation were the two most important variables impacting SMC across depths, soil types, and models. These findings are consistent with hydrological SMC

10.3389/fsoil.2025.1612908

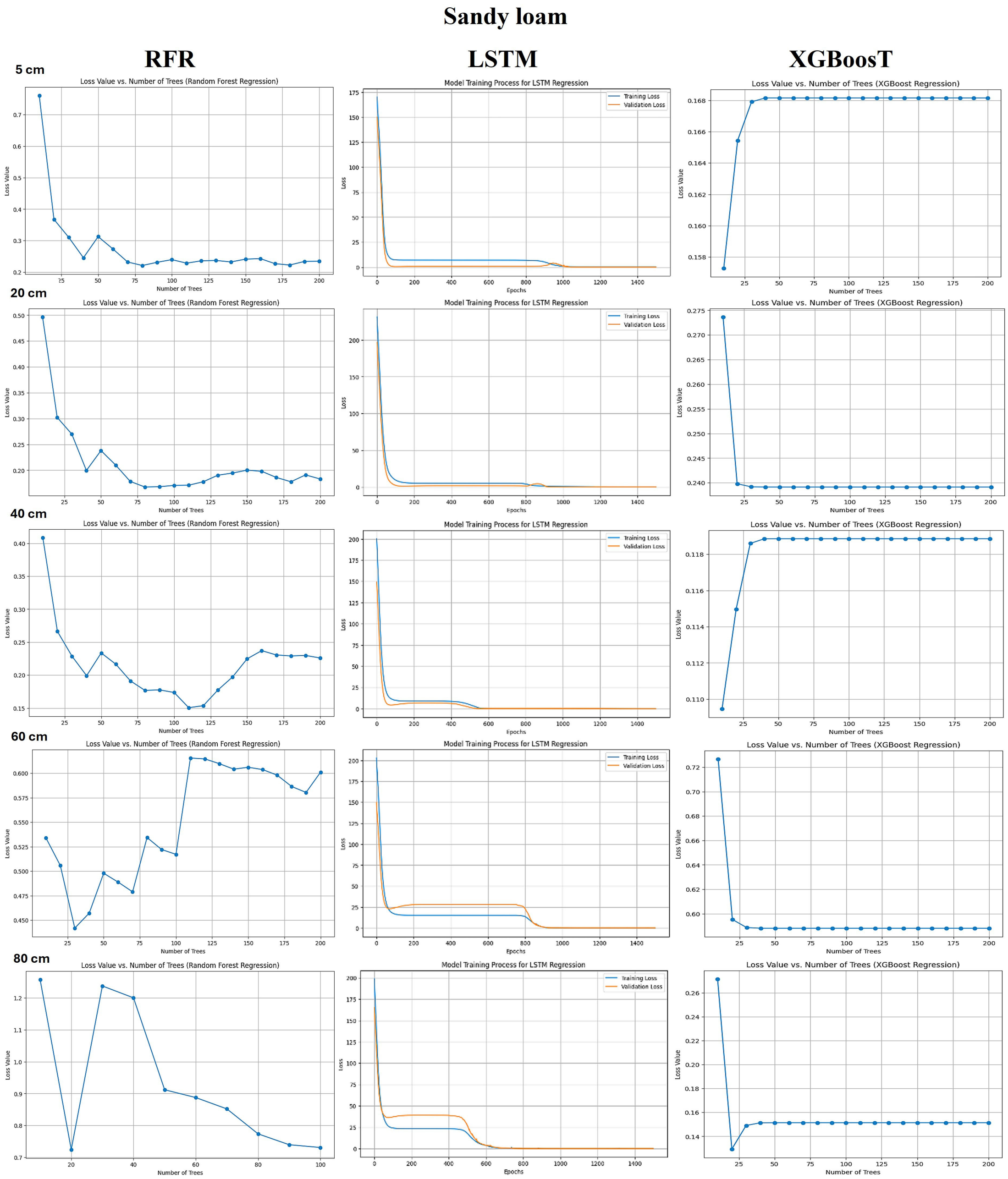



FIGURE 11
The loss value vs. number of trees and epochs for RFR, XGBoost, and LSTM at various depths in sandy loam soil.

principles, in which precipitation determines water input and solar radiation with temperature causing evapotranspiration and moisture loss (67). Overall, the feature importance results showed that the ML models effectively captured the key hydrological factors that influenced soil moisture dynamics. In particular, the ability to quantify and rank the impact of different predictors provides a solid base for optimizing irrigation systems. By identifying the key variables impacting SMC at certain depths and soil types, this study allows for data-driven decisions that support site-specific irrigation scheduling, reduce water use waste, and improve resource use efficiency. Additionally, feature importance analysis not only enhances model interpretability but also directly contributes to the objectives of sustainable water management and precision crop production. However, it is important to acknowledge that the influence of environmental predictors varies notably across soil types and depths. This highlights the need to develop soil-specific and depth-specific models that account for these variations by selecting key drivers tailored to each scenario. Additionally, the current study was limited to data from a single location and one growing season, which may affect the generalizability of the results

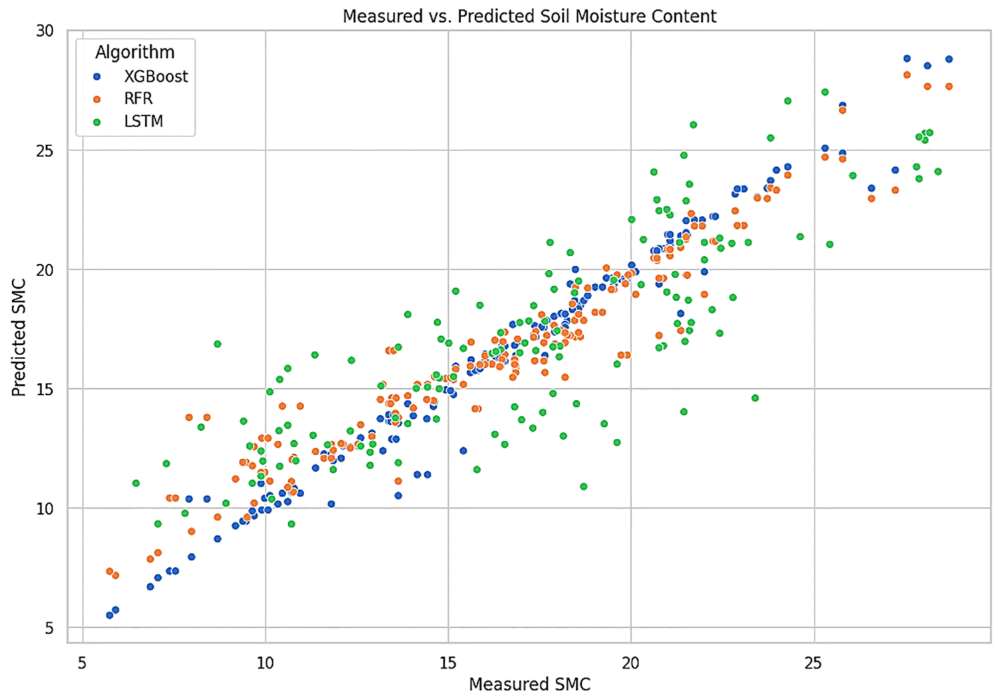



FIGURE 12
Measured and predicted soil moisture content results by the three ML models (RFR, XGBoost, and LSTM).

to other regions or climatic conditions. Incorporating multi-season and multi-location datasets in future research would improve the robustness and transferability of the models. Furthermore, integrating intrinsic soil properties such as the percentages of sand, silt, and clay, along with other relevant variables like vegetation indices and organic matter into the predictive framework, could enhance model accuracy and adaptability. Expanding the comparison to include additional ML algorithms and larger datasets will also be essential to strengthen performance and ensure scalability for broader precision irrigation applications.

# 4 Conclusions

The research introduced soil-specific and depth-specific modeling for SMC prediction at three different soil types and five depths (5, 20, 40, 60, and 80 cm) using three ML models (RFR, XGBoost, and LSTM). The results showed that SMC varied in the studied soil types, with sandy loam soil exhibiting less variance and silt loam and loam soils showing higher variability across depths. This emphasizes the importance of considering both soil type and depth when monitoring SMC.

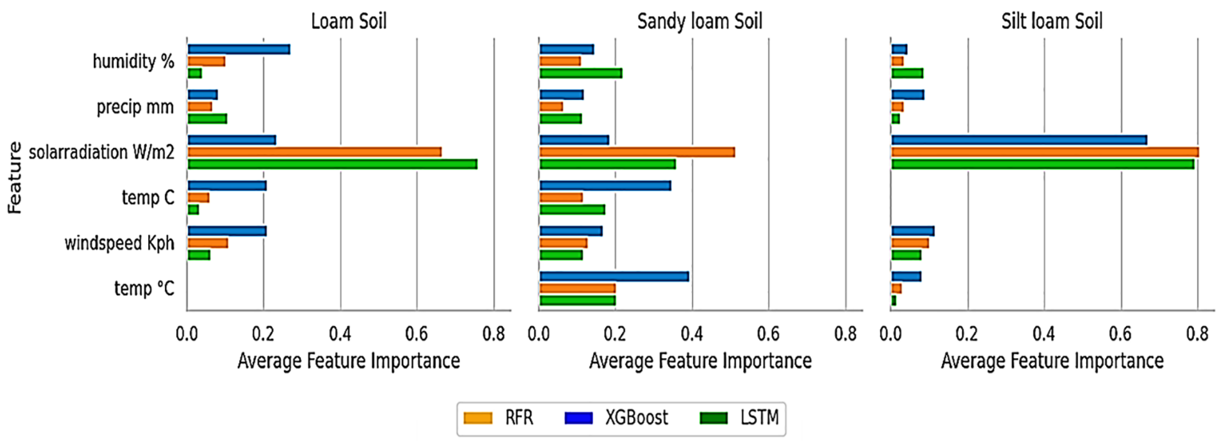



FIGURE 13
Average feature importance value results by the three ML models (RFR, XGBoost, and LSTM) per soil type across all depths.

RFR performed the best in all studied soil types and depths, particularly in capturing soil moisture variability in deeper soil layers (60 and 80 cm). LSTM performed well at shallower to moderate depths (5 to 60 cm) with less accuracy in deeper soil layers. On the other hand, XGBoost achieved high accuracy in predicting SMC in shallower depths, particularly in sandy loam soil due to its ability to model complex interactions between meteorological inputs and soil moisture dynamics. The differences between measured and predicted values provide opportunities for model modification and validation, especially for deeper layers and the need to enhance the generalizability and applicability of models by considering increasing inputs for model training with more variables such as different locations, irrigations, evapotranspiration, vegetation indices, and soil properties, as well as highlighting the fact that each soil type and depth could be modeled with different algorithms. The feature importance analysis results further demonstrated the dominant role of solar radiation and precipitation, emphasizing their significant influence on soil moisture dynamics.

These findings highlight the practical importance of these models in agricultural and environmental management. By accurately predicting SMC, these models could enhance water-use efficiency, optimize irrigation schedule, and improve understanding of soil moisture dynamics for ecosystem management. The study also demonstrates the potential for expanding these models to predict SMC in different regions, offering valuable insights for decision-makers in agriculture, hydrology, and enhancing the sustainability that aligned with Sustainable Development Goal 6 (SDG6) (sustainable management of water). While the models demonstrated strong predictive performance, their application is currently limited to a single site and one growing season. Future research should focus on expanding datasets across multiple locations and seasons, integrating additional soil properties and vegetation indices, and testing a wider range of algorithms to enhance accuracy, adaptability, and scalability for broader precision irrigation applications.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Author contributions

TA: Writing – review & editing, Writing – original draft, Formal Analysis, Methodology. MN: Writing – review & editing, Supervision. AS: Validation, Writing – review & editing. NA: Writing – review & editing. OH: Writing – review & editing. AN: Methodology, Conceptualization, Writing – review & editing, Writing – original draft.

## Funding

The author(s) declare financial support was received for the research and/or publication of this article. The research was supported by the EKÖP-24-3-I-SZE-107 University Research Fellowship Program of the Ministry for Culture and Innovation from the Source of the National Research, Development and Innovation Fund; the “Precision Bioengineering Research Group” supported by the “Széchenyi István University Foundation”; and the János Bolyai Research Scholarship (Bo/00578/24) of the Hungarian Academy of Sciences. TKP2021-NKTA-32 has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher’s note



## References

1. Sundmaeker H, Verdouw C, Wolfert S, Freire LP. Internet of food and farm 2020. In: *Digitising the industry internet of things connecting the physical, digital and virtualWorlds*. New York, USA: River Publishers (2022). p. 129–51.

2. Kulmány IM, Bede-Fazekas Á, Beslin A, Giczi Z, Milics G, Kovács B, et al. Calibration of an Arduino-based low-cost capacitive soil moisture sensor for smart agriculture. *J Hydrology Hydromechanics*. (2022) 70:330–40. doi: 10.2478/johh-2022-0014

3. Filipović N, Brdar S, Mimić G, Marko O, Crnojević V. Regional soil moisture prediction system based on Long Short-Term Memory network. *Biosyst Eng*. (2022) 213:30–8. doi: 10.1016/j.biosystemseng.2021.11.019

4. Stroobosscher ZJ, Athelly A, Guzmán SM. Assessing capacitance soil moisture sensor probes' ability to sense nitrogen, phosphorus, and potassium using volumetric ion content. *Front Agron*. (2024) 6:1346946. doi: 10.3389/fagro.2024.1346946

5. Bibek Acharya VS. COMPARATIVE ANALYSIS OF SOIL AND WATER DYNAMICS IN CONVENTIONAL AND SOD-BASED CROP ROTATION IN FLORIDA. *Front Agron*. (2025) 7:1552425. doi: 10.3389/fagro.2025.1552425

6. Luo W, Xu X, Liu W, Liu M, Li Z, Peng T, et al. UAV based soil moisture remote sensing in a karst mountainous catchment. *Catena (Amst)*. (2019) 174:478–89. doi: 10.1016/j.catena.2018.11.017

7. Wang X, Liu H, Sun Z, Han X. Soil moisture inversion based on multiple drought indices and RBFNN: A case study of northern Hebei Province. *Heliyon*. (2024) 10: e37426. doi: 10.1016/j.heliyon.2024.e37426

8. Seneviratne SI, Corti T, Davin EL, Hirschi M, Jaeger EB, Lehner I, et al. Investigating soil moisture–climate interactions in a changing climate: A review. *Earth Sci Rev*. (2010) 99:125–61. doi: 10.1016/j.earscirev.2010.02.004

9. Mulla DJ. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosyst Eng*. (2013) 114:358–71. doi: 10.1016/j.biosystemseng.2012.08.009

10. Cheng M, Jiao X, Liu Y, Shao M, Yu X, Bai Y, et al. Estimation of soil moisture content under high maize canopy coverage from UAV multimodal data and machine learning. *Agric Water Manag*. (2022) 264:107530. doi: 10.1016/j.agwat.2022.107530

11. Lv C, Xie Q, Peng X, Dou Q, Wang J, Lopez-Sanchez JM, et al. Soil moisture retrieval over agricultural fields with machine learning: A comparison of quad-, compact-, and dual-polarimetric time-series SAR data. *J Hydrol (Amst)*. (2024) 644:132093. doi: 10.1016/j.jhydrol.2024.132093

12. Holsten A, Vetter T, Vohland K, Krysanova V. Impact of climate change on soil moisture dynamics in Brandenburg with a focus on nature conservation areas. *Ecol Modell*. (2009) 220:2076–87. doi: 10.1016/j.ecolmodel.2009.04.038

13. Gao L, Wang Y, Geris J, Hallett PD, Peng X. The role of sampling strategy on apparent temporal stability of soil moisture under subtropical hydroclimatic conditions. *J Hydrology Hydromechanics*. (2019) 67:260–70. doi: 10.2478/johh-2019-0006

14. Meißl G, Klebinder K, Zieher T, Lechner V, Kohl B, Markart G. Influence of antecedent soil moisture content and land use on the surface runoff response to heavy rainfall simulation experiments investigated in Alpine catchments. *Heliyon*. (2023) 9: e18597. doi: 10.1016/j.heliyon.2023.e18597

15. Singh A, Gaurav K, Sonkar GK, Lee C-C. Strategies to measure soil moisture using traditional methods, automated sensors, remote sensing, and machine learning techniques: review, bibliometric analysis, applications, research findings, and future directions. *IEEE Access*. (2023) 11:13605–35. doi: 10.1109/ACCESS.2023.3243635

16. Little M, Smith. A comparison of three methods of soil water content determination. *South Afr J Plant Soil*. (1998) 15:80–9. doi: 10.1080/02571862.1998.10635121

17. Wagner W, Lemoine G, Rott H. A method for estimating soil moisture from ERS scatterometer and soil data. *Remote Sens Environ*. (1999) 70:191–207. doi: 10.1016/S0034-4257(99)00036-X

18. Cai Y, Zheng W, Zhang X, Zhangzhong L, Xue X. Research on soil moisture prediction model based on deep learning. *PloS One*. (2019) 14:e0214508. doi: 10.1371/journal.pone.0214508

19. Cheng M, Li B, Jiao X, Huang X, Fan H, Lin R, 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. doi: 10.1016/j.agwat.2021.107298

20. Zheng W, Zhangzhong L, Zhang X, Wang C, Zhang S, Sun S, et al. *A review on the soil moisture prediction model and its application in the information system*. Cham, Switzerland (2019). pp. 352–64. pp. 352–64. doi: 10.1007/978-3-030-06137-1_32.

21. Haddon A, Kechichian L, Harmand J, Dejean C, Ait-Mouheb N. Linking soil moisture sensors and crop models for irrigation management. *Ecol Modell*. (2023) 484:110470. doi: 10.1016/j.ecolmodel.2023.110470

22. Liakos K, Busato P, Moshou D, Pearson S, Bochtis D. Machine learning in agriculture: A review. *Sensors*. (2018) 18:2674. doi: 10.3390/s18082674

23. Nyéki A, Kerepesi C, Daróczy B, Benczúr A, Milics G, Nagy J, et al. Application of spatio-temporal data in site-specific maize yield prediction with machine learning methods. *Precis Agric*. (2021) 22:1397–415. doi: 10.1007/s11119-021-09833-8

24. Ågren AM, Larson J, Paul SS, Laudon H, Lidberg W. 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. doi: 10.1016/j.geoderma.2021.115280

25. Carranza C, Nolet C, Pezij M, van der Ploeg M. Root zone soil moisture estimation with Random Forest. *J Hydrol (Amst)*. (2021) 593:125840. doi: 10.1016/j.jhydrol.2020.125840

26. Senanayake IP, Yeo I-Y, Walker JP, Willgoose GR. Estimating catchment scale soil moisture at a high spatial resolution: Integrating remote sensing and machine learning. *Sci Total Environ*. (2021) 776:145924. doi: 10.1016/j.scitotenv.2021.145924

27. Ren Y, Ling F, Wang Y. Research on provincial-level soil moisture prediction based on extreme gradient boosting model. *Agriculture*. (2023) 13:927. doi: 10.3390/agriculture13050927

28. Alibabaei K, Gaspar PD, Lima TM. Modeling soil water content and reference evapotranspiration from climate data using deep learning method. *Appl Sci*. (2021) 11:5029. doi: 10.3390/app11115029

29. Paul S, Singh S. Soil moisture prediction using machine learning techniques. In: *Proceedings of the 2020 3rd international conference on computational intelligence and intelligent systems*. ACM, New York, NY, USA (2020). p. 1–7. doi: 10.1145/3440840.3440854

30. Kheimi M, Zounemat-Kermani M. Conventional and advanced AI-based models in soil moisture prediction. *Phys Chem Earth Parts A/B/C*. (2025) 139:103944. doi: 10.1016/j.pce.2025.103944

31. Sahour H, Gholami V, Torkaman J, Vazifedan M, Saeedi S. Random forest and extreme gradient boosting algorithms for streamflow modeling using vessel features and tree-rings. *Environ Earth Sci*. (2021) 80:747. doi: 10.1007/s12665-021-10054-5

32. Neményi M, Kovács AJ, Oláh J, Popp J, Erdei E, Harsányi E, et al. Challenges of sustainable agricultural development with special regard to Internet of Things: Survey. *Prog Agric Eng Sci*. (2022) 18:95–114. doi: 10.1556/446.2022.00053

33. Alahmad T, Neményi M, Nyéki A. Applying ioT sensors and big data to improve precision crop production: A review. *Agronomy*. (2023) 13:2603. doi: 10.3390/agronomy13102603

34. USDA. ND. *Natural Resources Conservation Service. Soil classification* (2020). Available online at: https://www.nrcs.usda.gov/wps/portal/nrcs/main/soils/survey/class/ (Accessed July 29, 2024).

35. Milics G, Kovács AJ, Pörneczi A, Nyéki A, Varga Z, Nagy V, et al. Soil moisture distribution mapping in topsoil and its effect on maize yield. *Biol (Bratisl)*. (2017) 72:847–53. doi: 10.1515/biolog-2017-0100

36. USDA. *Field book for describing and sampling soils V4.0*. United States Department of Agriculture, Washington DC, USA: USDA (2024). Available online at: https://www.nrcs.usda.gov/resources/guides-and-instructions/field-book-for-describing-and-sampling-soils.

37. Dirksen C. *Soil physics measurements*. Stuttgart, Germany: Catena Verlag (1999).

38. Shukla A, Panchal H, Mishra M, Patel PR, Srivastava HS, Patel P, et al. Soil moisture estimation using gravimetric technique and FDR probe technique: a comparative analysis. *Am Int J Res Form. Appl Nat Sci*. (2014) 8:89–92. doi: 10.13140/RG.2.2.28776.70405

39. Python Software Foundation. *Python Language Reference, version 3.10* (2024). Available online at: https://www.python.org (Accessed November 11, 2024).

40. Belgiu M, Drăguţ L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J Photogrammetry Remote Sens*. (2016) 114:24–31. doi: 10.1016/j.isprsjprs.2016.01.011

41. Loan Pham T, Huy Nguyen V, Le Thu Hoa T, Tien Ha TT, Phan CN, Vu XD, et al. Application of chemical fertilizers and plant spacing improves growth and root yield of rehmannia glutinosa libosch. *Asian J Plant Sci*. (2020) 19:68–76. doi: 10.3923/ajps.2020.68.76

42. Chen T, Guestrin C. XGBoost. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, New York, NY, USA (2016). p. 785–94. doi: 10.1145/2939672.2939785

43. Willmott CJ, Ackleson SG, Davis RE, Feddema JJ, Klink KM, Legates DR, et al. Statistics for the evaluation and comparison of models. *J Geophys Res Oceans*. (1985) 90:8995–9005. doi: 10.1029/JC090iC05p08995

44. Nembrini S, König IR, Wright MN. The revival of the Gini importance? *Bioinformatics*. (2018) 34:3711–8. doi: 10.1093/bioinformatics/bty373

45. Arya PK, Sur K, Dhote S, Siral H, Kundu T, Mehta BS, et al. Integrating multi-source satellite imagery and socio-economic household data for wealth-based poverty assessment of India: A GIS and machine learning based approach. *Soc Indic Res*. (2025) 179:653–76. doi: 10.1007/s11205-025-03614-w

46. Wang L, Gao Y. Soil moisture inversion using multi-sensor remote sensing data based on feature selection method and adaptive stacking algorithm. *Remote Sens (Basel)*. (2025) 17:1569. doi: 10.3390/rs17091569

47. Zhang J, Amonette JE, Flury M. Effect of biochar and biochar particle size on plant-available water of sand, silt loam, and clay soil. *Soil Tillage Res*. (2021) 212:104992. doi: 10.1016/j.still.2021.104992

48. Bajpai A, Kaushal A. Soil moisture distribution under trickle irrigation: a review. *Water Supply*. (2020) 20:761–72. doi: 10.2166/ws.2020.005

49. Yang L, Wei W, Chen L, Jia F, Mo B. Spatial variations of shallow and deep soil moisture in the semi-arid Loess Plateau, China. *Hydrol Earth Syst Sci*. (2012) 16:3199–217. doi: 10.5194/hess-16-3199-2012

50. Guo X, Fu Q, Hang Y, Lu H, Gao F, Si J. Spatial variability of soil moisture in relation to land use types and topographic features on hillslopes in the black soil (Mollisols) area of northeast China. *Sustainability*. (2020) 12:3552. doi: 10.3390/su12093552

51. Draper C, Reichle R, de Jeu R, Naeimi V, Parinussa R, Wagner W. Estimating root mean square errors in remotely sensed soil moisture over continental scale domains. *Remote Sens Environ*. (2013) 137:288–98. doi: 10.1016/j.rse.2013.06.013

52. Ning J, Yao Y, Tang Q, Li Y, Fisher JB, Zhang X, et al. Soil moisture at 30 m from multiple satellite datasets fused by random forest. *J Hydrol (Amst)*. (2023) 625:130010. doi: 10.1016/j.jhydrol.2023.130010

53. Park S-H, Lee B-Y, Kim M-J, Sang W, Seo MC, Baek J-K, et al. Development of a soil moisture prediction model based on recurrent neural network long short-term memory (RNN-LSTM) in soybean cultivation. *Sensors*. (2023) 23:1976. doi: 10.3390/s23041976

54. O'Callaghan JR, Menzies DJ, Bailey PH. Digital simulation of agricultural drier performance. *J Agric Eng Res*. (1971) 16:223–44. doi: 10.1016/S0021-8634(71)80016-1

55. Behroozi-Khazaei N, Nasirahmadi A. A neural network based model to analyze rice parboiling process with small dataset. *J Food Sci Technol*. (2017) 54:2562–9. doi: 10.1007/s13197-017-2701-x

56. Babaeian E, Sadeghi M, Jones SB, Montzka C, Vereecken H, Tuller M. Ground, proximal, and satellite remote sensing of soil moisture. *Rev Geophysics*. (2019) 57:530–616. doi: 10.1029/2018RG000618

57. Wang L, Fang S, Pei Z, Zhu Y, Khoi DN, Han W. Using fengYun-3C VSM data and multivariate models to estimate land surface soil moisture. *Remote Sens (Basel)*. (2020) 12:1038. doi: 10.3390/rs12061038

58. Zhao W, Li A, Huang P, Juelin H, Xianming M. Surface soil moisture relationship model construction based on random forest method. In: *2017 IEEE international geoscience and remote sensing symposium (IGARSS)*. Fort Worth, TX, USA: IEEE (2017). doi: 10.1109/IGARSS.2017.8127378

59. Li C, Zhang Y, Ren X. Modeling hourly soil temperature using deep biLSTM neural network. *Algorithms*. (2020) 13:173. doi: 10.3390/a13070173

60. Basir S, Noel S, Buckmaster D, Ashik-E-Rabbani M. Enhancing subsurface soil moisture forecasting: A long short-term memory network model using weather data. *Agriculture*. (2024) 14:333. doi: 10.3390/agriculture14030333

61. Nguyen TT, Ngo HH, Guo W, Chang SW, Nguyen DD, Nguyen CT, et al. A low-cost approach for soil moisture prediction using multi-sensor data and machine learning algorithm. *Sci Total Environ*. (2022) 833:155066. doi: 10.1016/j.scitotenv.2022.155066

62. Santosa S, Santosa YP, Lambang Goro G, Wahjoedi -, Mahbub J. Computational of Concrete Slump Model Based on H2O Deep Learning framework and Bagging to reduce Effects of Noise and Overfitting. *JOIV : Int J Inf Visualization*. (2023) 7:370. doi: 10.30630/joiv.7.2.1201

63. Zhang Y, Han W, Zhang H, Niu X, Shao G. Evaluating soil moisture content under maize coverage using UAV multimodal data by machine learning algorithms. *J Hydrol (Amst)*. (2023) 617:129086. doi: 10.1016/j.jhydrol.2023.129086

64. Srivastava A, Yetemen O, Kumari N, Saco PM. Role of solar radiation and topography on soil moisture variations in semiarid aspect-controlled ecosystems. *sat*. (2018) 1. doi: 10.13140/RG.2.2.28776.70405

65. Han Q, Zeng Y, Zhang L, Wang C, Prikaziuk E, Niu Z, et al. Global long term daily 1 km surface soil moisture dataset with physics informed machine learning. *Sci Data*. (2023) 10:101. doi: 10.1038/s41597-023-02011-7

66. Dai L, Fu R, Guo X, Du Y, Zhang F, Cao G. Soil moisture variations in response to precipitation across different vegetation types on the northeastern qinghai-tibet plateau. *Front Plant Sci*. (2022) 13:854152. doi: 10.3389/fpls.2022.854152

67. Du M, Zhang J, Wang Y, Liu H, Wang Z, Liu C, et al. Evaluating the contribution of different environmental drivers to changes in evapotranspiration and soil moisture, a case study of the Wudaogou Experimental Station. *J Contam Hydrol*. (2021) 243:103912. doi: 10.1016/j.jconhyd.2021.103912

*applied sciences*

Article

# Soil Moisture Content Prediction Using Gradient Boosting Regressor (GBR) Model: Soil-Specific Modeling with Five Depths

**Tarek Alahmad *, Miklós Neményi and Anikó Nyéki**

Department of Biosystems Engineering and Precision Technology, Albert Kázmér Mosonmagyaróvár Faculty of Agricultural and Food Sciences, Széchenyi István University, 9200 Mosonmagyarovar, Hungary; nemenyi.miklos@sze.hu (M.N.); nyeki.aniko@sze.hu (A.N.)
* Correspondence: alahmad.tarek@sze.hu; Tel.: +36-302184633

**Abstract:** Monitoring soil moisture content (SMC) remains challenging due to its spatial and temporal variability. Accurate SMC prediction is essential for optimizing irrigation and enhancing water use efficiency. In this research, a Gradient Boosting Regressor (GBR) model was developed and validated to predict SMC in two soil textures, loam and silt loam, using meteorological data from Internet of Things (IoT) sensors and gravimetric SMC field measurements collected from five different depths. The statistical analysis revealed significant variation in SMC across depths in loam soil ($p < 0.05$), while silt loam exhibited more stable moisture distribution. The GBR model demonstrated high performance in both soil textures, achieving $R^2$ values of 0.98 and 0.94 for silt loam and loam soils, respectively, with low prediction errors (RMSE 0.85 and 0.97, respectively). Feature importance analysis showed that precipitation and humidity were the most influential features in loam soil, while solar radiation had the highest impact on prediction in silt loam soil. Soil depth also showed a significant contribution to SMC prediction in both soils. These results highlight the necessity for soil-specific modeling to enhance SMC prediction accuracy, optimize irrigation systems, and support water resources management approaches aligning with SDG6 objectives.

**Keywords:** soil moisture content; gradient boosting regressor (GBR); IoT sensors; soil specific modeling; crop water management

check for updates

Academic Editor: José Miguel Molina Martínez

Received: 5 March 2025
Revised: 15 May 2025
Accepted: 22 May 2025
Published: 23 May 2025

**Citation:** 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



## 1. Introduction

Soil moisture content (SMC) plays a crucial role in hydrological and ecological processes that affect energy, water, and carbon cycles, including evaporation [1], transpiration, diversity [2], and rainfall–runoff in various ecosystems [3]. Understanding the spatiotemporal patterns of SMC is essential for water resources management, mitigating the effects of agricultural drought [4], addressing water scarcity, and improving crop production [5,6]. Moreover, the collection of high-quality soil moisture data is significant for scientific research purposes, such as weather forecasting, flood prediction, drought monitoring [7,8], precision agriculture [9], and irrigation management [10].

As the demand for high-precision soil moisture data increases, there is a growing need to develop reliable in situ soil moisture sensor networks that balance cost, complexity, and low power consumption for effective soil moisture monitoring [11]. The advancement of sensing technologies, including Internet of Things (IoT), has enabled real-time data exchange between connected devices using a wireless sensor network, allowing for constant monitoring of various environmental parameters in real time [12]. Recent studies have demonstrated the applicability of IoT-based approaches in agriculture for environmental monitoring [13], controlling greenhouse gas emissions [14], irrigation optimization, weather monitoring, product traceability, and intelligent equipment management [15]. Notably, IoT sensors address the limitations of traditional soil moisture measurement techniques, which are often time consuming and lack real-time monitoring capabilities, by providing real-time, low-maintenance, and scalable solutions [16]. In addition to in situ measurements, remote sensing approaches utilizing satellite images have been widely used for SMC monitoring and predicting model development [17]. The integration of meteorological data and soil properties monitoring in SMC prediction provides a better understanding of its variability and the types of interactions between SMC and these parameters.

Accurate SMC prediction is important for enhancing irrigation management and optimizing crop production [18]. Traditional statistical methods often struggle to capture the complex, nonlinear, and multivariate interactions among soil, environmental, and meteorological factors that influence SMC [19]. In contrast, machine learning (ML) algorithms can handle these complexities and offer high potential for accurate SMC prediction [20,21]. ML models have the capability to utilize the high-resolution dataset provided by IoT-based sensors to enhance prediction accuracy of SMC. Among these, the Gradient Boosting Regressor (GBR) algorithm has demonstrated strong performance in SMC prediction due to its ability to combine multiple weak learners (typically decision trees) to iteratively minimize loss functions, thereby improving both accuracy and interpretability [22]. For instance, Zambudio et al. [23] utilized two years of data to train a GBR model and artificial neural network (ANN) to predict SMC in top surface layers. The results showed that the GBR model performed effectively in predicting SMC with a 0.01 learning rate, 5 max depth, and 350 estimators, achieving a mean square error (MSE) of 0.027 and a 20% difference in maximum between measured and predicted values. Similarly, Ren et al. [24] utilized XGBoost to predict soil moisture, highlighting its high performance in comparison with other models, with a correlation coefficient of 0.69 and an accuracy of 88%. The primary predictors were air relative humidity, maximum air temperature, and total precipitation. Consideration of soil characteristics, particularly during the 2022 drought, led to improved prediction accuracy.

However, existing research had a lack of soil texture-specific and depth-specific modeling, which considers significant factors regarding their strong impact on SMC. For instance, soils with higher silt and clay content generally exhibit higher water-holding capacity compared to sandy soils, which leads to higher SMC [25]. Additionally, soil depth has a crucial impact on SMC, as deeper layers often retain more moisture due to lower evaporation and higher water retention. Fu et al. [26] demonstrated this in their study utilizing the water balance equation, showing that predictive model performance varied with depth, achieving lower RMSE values at 100 cm and 200 cm compared to shallow depths.

In response to these developments, international frameworks such as the UN-Water 2023 Annual Report and the EU's "Farm to Fork" strategy have emphasized the significance of adaptive modern water resource management approaches that align with Sustainable Development Goal 6 (SDG6): Water and Sanitation for All [27]. The integration of IoT and ML in precision crop production contributes to these goals by allowing for better resource management, enabling data-driven decision making [28].

This research developed and validated soil texture-specific Gradient Boosting Regressor (GBR) models to predict SMC in two distinct soil textures, loam and silt loam. These soils differ in their moisture retention characteristics, where the loam soil exhibits moderate water-holding capacity due to its balanced sand, silt, and clay proportions, while silt loam retains more moisture because of its finer texture and higher silt content [29].

High-resolution meteorological data were collected using calibrated IoT-based sensors for model training. The key objectives of this research are to evaluate GBR model performance using multiple evaluation metrics; to identify optimal model hyperparameters by performing 5-fold cross-validation and grid search; and to identify the most influential predictors of SMC in each soil texture utilizing feature importance analysis. These will allow for better understanding of soil moisture dynamics in each soil texture, as well as better optimization of irrigation schedules and enhanced water resource use efficiency, aligning with SDG6 goals for sustainable water management.

## 2. Materials and Methods

### 2.1. Experimental Site and Data Collection

This field experiment was conducted during the maize vegetation season at a 23-hectare field belonging to Széchenyi István University, located in Mosonmagyaróvár, Hungary. The site lies within the Little Hungarian Plain, an alluvial plain of the Leitha River, characterized by relatively flat topography with a slight slope of approximately 5% and elevation ranging from 133 to 138 m above sea level. This gradient allows for natural surface drainage while minimizing erosion risks. The region experiences a temperate climate with moderate precipitation, typical of Central Europe. During the study period, from May to October, total precipitation was approximately 400 mm, and the average air temperature ranged from 18.5 °C to 21.3 °C. Based on USDA soil classification taxonomy [30] and laboratory particle size distribution analysis using the soil texture triangle, three primary soil textures were identified in the field: loam, sandy loam, and silt loam (Figure 1). These soil textures represent more than 60% of the soils in this region [29]. The field is an alluvial plain of the Leitha River located within the little Hungarian plain and is characterized by relatively flat topography. The field has a slight slope of 5%, with elevation varying between 133 and 138 m above sea level. This slight gradient allows natural surface drainage while minimizing erosion risks. For this study, two soil textures—loam and silt loam—were selected for detailed analysis. Their specific physical and chemical properties are presented in Table 1. The field is managed under a cereal-based crop rotation system, which includes winter wheat, spring barley, maize, and soybeans.

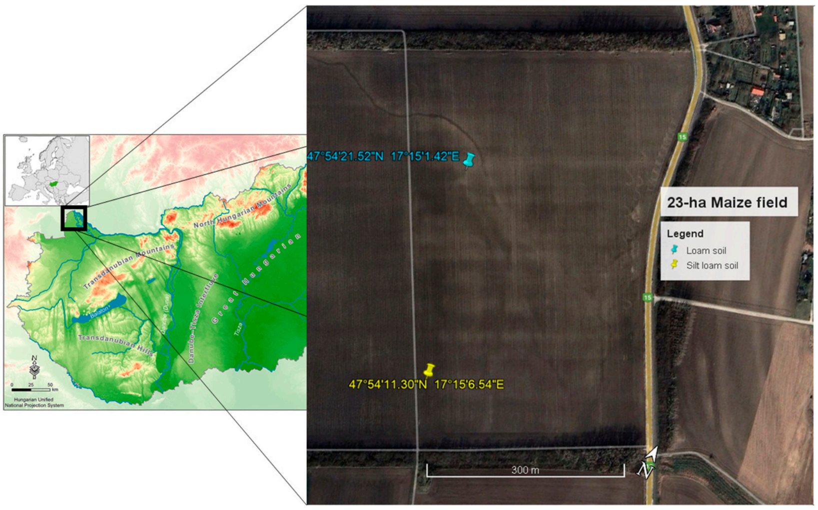



**Figure 1.** Soil texture in the 23 ha maize field in Mosonmagyaróvár, Hungary, with GPS points of the sampling locations.

**Table 1.** Soil properties averages and ranges of soil content of sand, silt, and clay in the five depths in the two studied soil textures with mean SMC.

| Soil Texture | pH | Sand % Range | Silt % Range | Clay % Range | Mean SMC % | Std of SMC % | Bulk Density |
|---|---|---|---|---|---|---|---|
| Loam | 7.614 | 40.99–49.01 | 40.91–45.97 | 10.10–14.50 | 13.79 | 3.94 | 1.538 |
| Silt loam | 7.45 | 12.36–15.31 | 63.70–70.19 | 17.40–21 | 16.95 | 5.07 | 1.52 |

### 2.2. Soil Sampling and Gravimetric Measurements

From the two selected soil textures, 300 soil samples were collected between June and October 2023 using a hand auger (Figure 2B). Sampling was conducted at five depths (5, 20, 40, 60, and 80 cm) from three sampling points near the sensor installations, following a random sampling grid approach. Soil samples were collected at 10 different dates spaced two weeks apart, resulting in a total of 150 samples from each soil texture.


**Figure 2.** (**A**) IoT-based meteorological sensor station within the field. (**B**) Collecting soil samples using a hand auger from sampling points around the sensor stations.

Soil moisture content was calculated using the gravimetric method based on dry weight [31]. The samples from each depth and soil texture were placed in pre-weighed aluminum containers and dried in an oven at 105 °C until a consistent weight was achieved. The containers were then weighed again to calculate dry weight of the soil. The soil moisture content (θ) was calculated using the following Equation (1):

$$\theta = \frac{Mw}{Md} \times 100 \tag{1}$$

where

$\theta$ = Soil Moisture Content (%);
$Mw$ = Mass of water (g) = (Wet weight − Dry weight);
$Md$ = Mass of dry soil (g).

*2.3. Meteorological Data Collection Utilizing IoT Sensor*

Meteorological data were collected using an IoT-based meteorological sensor (Boreas Ltd., EcoLogger Boreas datalogger unit BEL-06, BHS-06 for temperature and humidity, BWS/s-06 Wind Speed Sensor, BIS-06 global solar radiation, and BPD-06 for precipitation) (Figure 2A). The manufacturer claims an accuracy of ±0.2 °C for temperature, ±2% for humidity, and ±0.1 mm for precipitation. To minimize the measurement error that could affect SMC prediction, the meteorological sensor was calibrated to ensure the accuracy of data. The data were collected at 10–15-min intervals using LoRaWAN, a low power consumption device with the capability of transmitting sensor signals up to 30–40 km away on level terrain. Meteorological parameters, including precipitation (mm), temperature (°C), humidity (%), wind speed (km/h), and solar radiation ($W/m^2$), were used to examine how these meteorological factors influence variations in soil moisture content, considering the effects of both soil depth and soil texture. These features were selected based on their influence on soil moisture dynamics, where precipitation is considered the primary source of soil water and SMC increased with increasing precipitation [32]. Humidity, on the other hand, reflects the air moisture content, influencing the evaporation rate from the surface soil where increasing the humidity reduces the moisture losses by evaporation [33]. Solar radiation also has a significant impact on SMC depletion by its effect on available energy for evaporation and transpiration. Temperature also plays a crucial role in SMC variation; evaporation increases at higher temperatures. Also, wind speed has a role in enhancing moisture losses in the soil surface due to the increase in air movement [34,35]. The data collected represent the average of the field and were assigned to all sampling locations. Also, the data were synchronized with the dates of soil sampling to ensure accurate pairing of meteorological and SMC observations.

*2.4. Data Preprocessing and Feature Engineering*

Data preprocessing was performed utilizing Python (version 3.10.12) [36], and meteorological data were synchronized with the SMC data. Categorical variables such as soil texture were encoded, and missing data were handled by imputation. Data preparation, such as transformation, cleaning, and statistical testing to ensure uniformity and prevent inaccurate or missing values, was conducted using Python. Six predictive features were selected to train the model including depth (cm), precipitation (mm), temperature (°C), humidity (%), wind speed (km/h), and solar radiation ($W/m^2$), and the target variable is SMC%. These features were chosen based on their availability from field sensors and their well-documented influence on soil moisture dynamics as reported in previous studies, highlighting their agricultural relevance in irrigation management, crop water requirements, and soil water balance assessments [37,38].

To determine whether SMC varied significantly across the five depths while accounting for repeated sampling at the same times across time, Repeated Measures ANOVA (Equation (2)) was performed independently for each soil texture using the statsmodels library's AnovaRM. The subject was identified using sampling points, and the within-subject element was depth (cm). The statistical significance level was set to $p < 0.05$.

Using Python, a descriptive statistical analysis was conducted to determine the mean soil moisture content (SMC%), standard deviation (SD), and coefficient of variation (CV%) for each soil texture and depth.

$$Y_{ijk} = \mu + \alpha_i + S_j + \varepsilon_{ijk} \quad (2)$$

where

$Y_{ijk}$ = SMC measurement at depth i, samplisng point j, time k.

M = Overall mean, $\alpha_i$ = Fixed effect of depth, $S_j$ = Random effect of sampling point.

$\varepsilon_{ijk}$ = Residual error.

The coefficient of variation (CV%) was calculated using the following formula:

$$\mathbf{CV(\%)} = \frac{SD}{Mean} \times \mathbf{100} \tag{3}$$

### 2.5. Development and Optimization of GBR Model

A Gradient Boosting Regressor (GBR) model was used to predict SMC due to its ability to capture the nonlinear interactions between studied features. GBR considers the form of ML algorithms that produce prediction models by integrating multiple weak learners, mainly decision trees, in a sequential manner. The procedure begins with an initial model, and the succeeding models are trained to optimize the previous models' errors, which effectively minimizes the loss function utilizing gradient descent approaches. This allowed us to develop a robust model with the ability to process complex patterns in data and enhance accuracy [39,40].

The model was trained for the two soil textures (loam and silt loam) separately to increase soil-specific prediction; the models were developed using the scikit-learn library; and the models were optimized using GridSearchCV for optimizing hyperparameters, including n_estimators, learning_rate, max_depth, and min_sample_split. A 5-fold cross-validation was utilized for performance validation to reduce overfitting and enhance robustness of the models. The dataset was split into 70% training and 30% testing sets. Feature importance analysis was carried out to determine the most important features that impact the SMC prediction in each soil texture. All modeling, analysis, and visualization were performed using Python libraries such as pandas, numpy, scikit-learn, seaborn, and matplotlib. The methodology used was developed to ensure the model's robustness, explainability, and applicability.

### 2.6. Performance Evaluation Measures

Five metrics were calculated both on the hold-out test set and by cross-validation to quantify the performance of the models.

1- MSE is the average of the squared differences between predicted and observed values [41].

$$MSE = 1/n\sum_{t=1}^{n}\left(y_t - \hat{y}t\right)^2 \tag{4}$$

2- The root means squared error (RMSE) [41] is as follows:

$$RMSE = \sqrt{1/n\sum_{t=1}^{n}\left(y_t - \hat{y}t\right)^2} \tag{5}$$

3- Mean Absolute Error (MAE) calculates the average of the absolute differences between predictions and actual observations, as shown in Equation (6) [41]:

$$MAE = 1/n\sum_{t=1}^{n}\left|y_t - \hat{y}t\right| \tag{6}$$

where $n$ is the total number of observations in the dataset, t is the time index or sequence index of an observation, $y_t$ is the observed value at time t, and $\hat{y}t$ is the predicted value by the model at time t.

4- The Nash Sutcliffe efficiency metric (NSE) [42] is as follows:

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \quad (7)$$

where $n$ is the total number of observations in the dataset; $y_i$ is the observed value at observation $i$; $\hat{y}_i$ is the predicted value at observation $i$; and $\overline{y}$ is the mean of observed values.

5- R-squared ($R^2$) is a statistical metric that represents the proportion of the total variance in the observed data that is explained by the model. Higher $R^2$ indicates less difference between observed and predicted values [43].

$$R^2 = 1 - \frac{\sum_{t=1}^{n}\left(y_t - \hat{y}t\right)^2}{\sum_{t=1}^{n}(y_t - y\_)^2} \quad (8)$$

## 3. Results and Discussion

### *3.1. Statistical Analysis Results for the Two Soil Textures*

The descriptive statistical analysis (Table 2) shows that in loam soil, the highest mean SMC (16.19%) was found at 20 cm with the least variability (CV = 21.7%), while the deepest layer (80 cm) had the most variability (CV = 39%) ranging from 5.49% to 20.77%. In silt loam soil, SMC was generally higher at all depths, with the highest mean SMC at 20 cm (18.27%). However, the deepest layers (60 cm and 80 cm) demonstrated higher variations (CV = 42.6% and 47.8%, respectively) with SMC ranging from 7.92% to 26.58%, indicating less consistent moisture retention at depth.

**Table 2.** Descriptive statistical analysis results for both soil textures at five depths.

| Soil Texture | Depth (cm) | Mean SMC (%) | SD (%) | CV (%) |
|---|---|---|---|---|
| Loam | 5 | 12.8 | 3.08 | 24 |
| Loam | 20 | 16.19 | 3.51 | 21.7 |
| Loam | 40 | 14.66 | 3.76 | 25.6 |
| Loam | 60 | 12.61 | 4.39 | 34.8 |
| Loam | 80 | 12.68 | 4.94 | 39 |
| Silt loam | 5 | 14.99 | 2.62 | 17.5 |
| Silt loam | 20 | 18.27 | 3.51 | 19.2 |
| Silt loam | 40 | 18.03 | 4.11 | 22.8 |
| Silt loam | 60 | 16.97 | 7.23 | 42.6 |
| Silt loam | 80 | 16.47 | 7.87 | 47.8 |

The repeated measures ANOVA revealed substantial variation in SMC across depths in both soil textures ($p < 0.05$). The variation in SMC across depths (Figure 3) highlights the more stable behavior of silt loam soil in the upper layers and the higher variability observed in loam soil at deeper layers.

The observed patterns confirmed the significant impact of depth on SMC variability, where deeper layers in both soil textures exhibited higher CV%, indicating less predictable moisture retention likely due to soil structural changes and lower infiltration rates. These findings highlight the impact of soil depth on SMC variability, emphasizing the importance of considering both soil texture and depth during monitoring [44,45] and the need for considering soil texture and depth in irrigation planning and management to optimize water resources and increase water use efficiency.

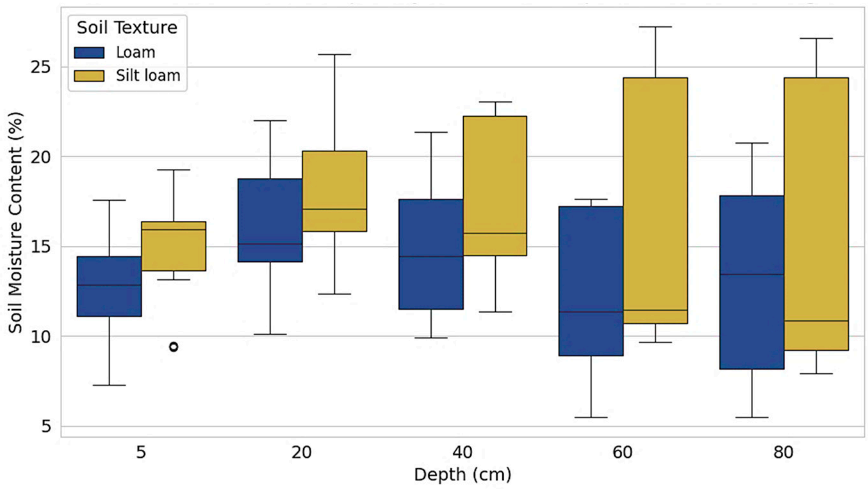



**Figure 3.** Variations in soil moisture content (SMC%) across five depths and two soil textures during the studied vegetation period, based on repeated measures ANOVA.

### *3.2. Model Performance Results*

The results show that GBR model performance was highly effective in predicting SMC in both soil textures (Figure 4). The model performance was evaluated using test data and validated by using 5-fold cross-validation. In silt loam soil, the model achieved the highest accuracy, with an $R^2$ of 0.98, MSE of 0.73, RMSE of 0.85, MAE of 0.45, and NSE of 0.95 (Figure 5), reflecting the model's consistency and generalization capability. In loam soil, the model performance was slightly lower, likely due to the higher variation in SMC values compared with silt loam, with an $R^2$ of 0.94, MSE of 0.94, RMSE of 0.97, MAE of 0.73, and NSE of 0.94. These results highlight the impact of soil texture on predictive model performance, where less prediction accuracy was shown in loam soil due to its higher variability in SMC compared with silt loam soil where its balanced structure and high water retention had a significant role in enhancing model accuracy. These results emphasize the need for soil-specific modeling to enhance SMC prediction [46] and the necessity for integrating more physical soil properties into models for improving prediction reliability across different soil textures.

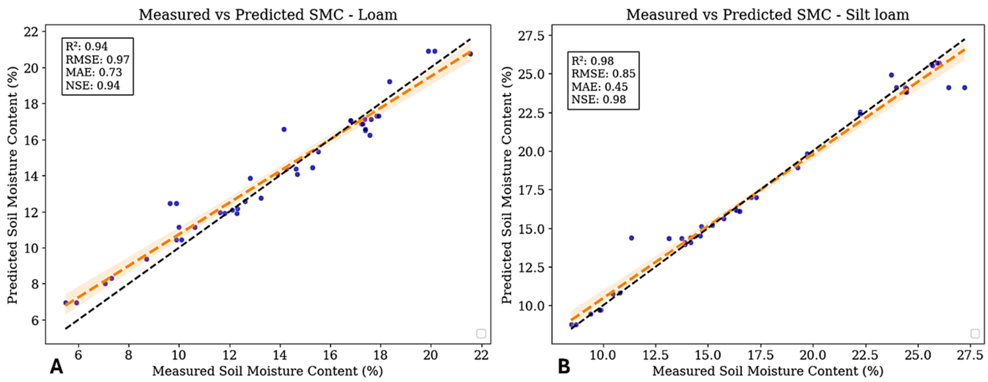



**Figure 4.** Model performance and comparison between SMC predicted and measured values in (**A**) loam and (**B**) silt loam soil.

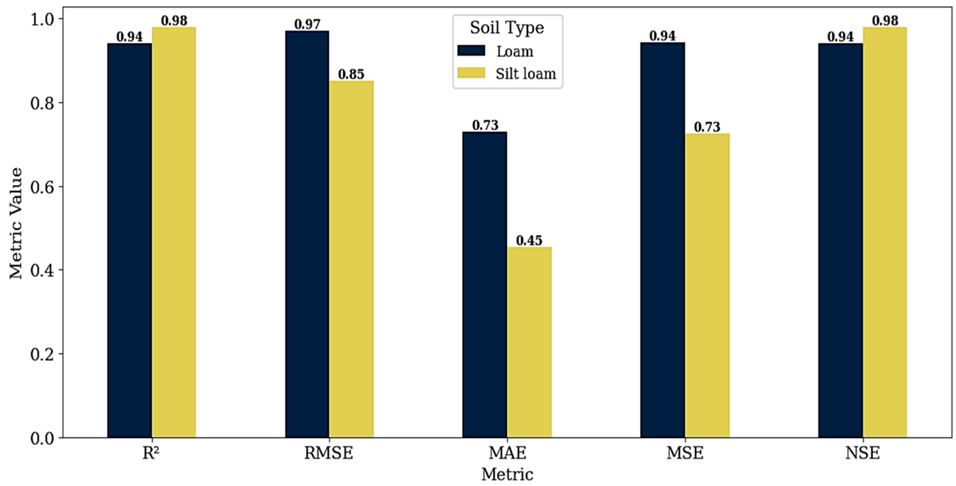



**Figure 5.** Evaluation metrics ($R^2$, MSE, RMSE, MAE, and NSE) for soil moisture content prediction in loam and silt loam soils using the GBR model.

### *3.3. Feature Importance Analysis*

To better understand the influence of meteorological features and soil depth on SMC prediction in the two examined soil textures (loam and silt loam), as well as to improve the model's interpretability, feature importance analysis was performed for both soil textures using the GBR model. The results reveal that the impact of the analyzed variables varied with soil texture. In loam soil, precipitation and humidity were the most influential features in prediction, with values of 31% for both (Figure 6). Soil depth had a substantial impact on SMC, with an importance value of 27%, as loam soil retains water effectively due to its balanced particle size distribution. Precipitation and humidity also played crucial roles in influencing SMC in loam soil. In contrast, wind speed and solar radiation had minimal impact, both with importance values below 6%. These results emphasize the significance of considering precipitation, humidity, and soil depth when monitoring and predicting SMC in loam soils, particularly as precipitation directly contributes to replenishing soil moisture [47], and humidity indirectly affects soil air water dynamics [48]. Conversely, in silt loam soil, the results emphasize the importance of solar radiation in SMC prediction with a value of 46%, followed by wind speed and depth with values of 26% and 24%, respectively, highlighting the role of radiation-driven evapotranspiration in controlling moisture variation in silt loam soils due to their balanced texture and its high water-holding capacity compared with loam soil and the role of wind in increasing evapotranspiration.

These findings are consistent with previous research, highlighting the critical role of environmental variables such as precipitation and solar radiation in influencing soil moisture dynamics and demonstrating the complexity of soil–water interactions in the studied soils [49]. Additionally, integrating the effects of soil texture and depth provided further insights into their significance in driving the fluctuations and dynamics of SMC [29].

The results also emphasize the importance of incorporating multiple features in model training for SMC prediction across different soil textures, including vegetation indices, which offer valuable information on soil–plant interactions and their impact on soil moisture patterns [19,50]. Increasing the number of sampling locations across the field would further improve the spatial representativity of the dataset, thereby enhancing the model's robustness and generalizability. The model effectively captured the relationships between the selected input variables and SMC in both soil textures, highlighting its potential to reduce the need for extensive field sampling and support cost-efficient soil moisture monitoring.

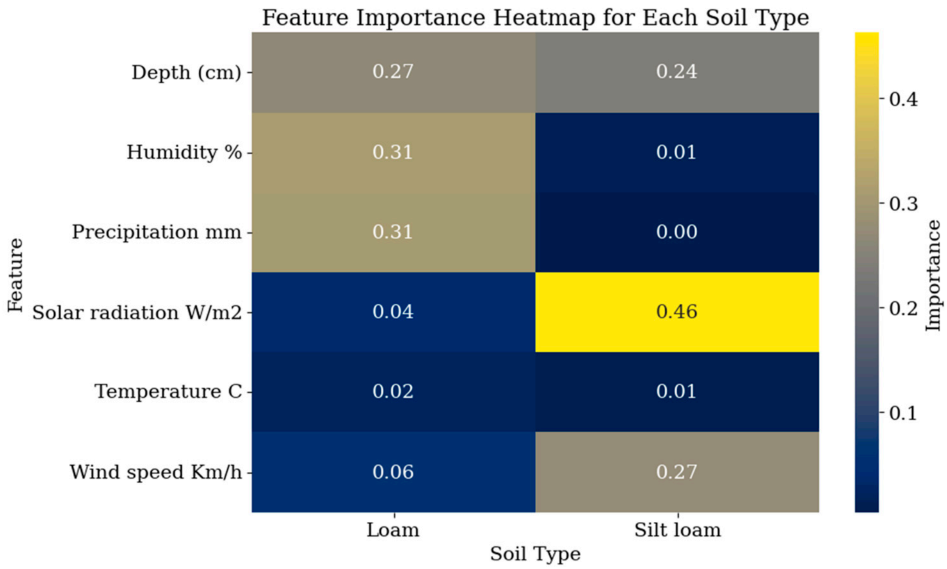



**Figure 6.** Feature importance analysis results of GBR model in loam and silt loam soils.

These findings show the complex interactions of meteorological conditions and depth with SMC [51]. In contrast, satellite-based models frequently have higher spatial resolution but limited soil penetration and temporal alignment with ground conditions, since they utilize passive microwaves signals and downscaling approaches to predict SMC in the surface soil layers [52]. These findings highlight the significance of soil-specific water management systems and real-time monitoring to improve crop production precision.

## 4. Conclusions

This study demonstrated that the Gradient Boosting Regressor (GBR) model is a reliable and accurate approach for predicting SMC in two distinct soil textures, loam and silt loam, using soil depth and meteorological data as input features. The GBR models achieved high prediction accuracy in both soil textures, with $R^2$ values exceeding 0.94, and the highest performance observed in silt loam soil. These findings emphasized the robustness and generalization capability of the GBR model in capturing nonlinear interactions between SMC, depth and meteorological parameters, despite the limited sample size and spatial coverage.

The repeated measures ANOVA confirmed significant variation in SMC across depths, particularly in loam soil, where deeper layers exhibited higher variability. In contrast, silt loam soil exhibited higher overall SMC with more stable distribution across depths. The feature importance analysis provided further insights, revealing that precipitation and humidity were the most influential variables in loam soil (both contributing approximately 31%), while solar radiation was the dominant factor in silt loam soil (46%). Soil depth also had a significant effect on SMC prediction in both soil textures, emphasizing the need to consider depth-specific patterns in modeling and monitoring efforts.

These findings underscore the importance of incorporating both meteorological variables and soil-specific characteristics in SMC modeling. The model's ability to achieve high accuracy with limited data supports its potential to reduce the need for extensive field sampling, contributing to more efficient data-driven decision making for improving water use efficiency and irrigation management, in alignment with SDG 6 goals on sustainable water use and ecosystem protection.

Additionally, the study emphasizes the necessity of soil texture-specific modeling combined with depth-sensitive predictions and explainable feature analysis. However, the study acknowledges the limitation regarding the assumption of independence in statistical analysis, which was addressed by applying a repeated measures design. Future research should further strengthen the modeling framework by incorporating more repeated sampling, integrating additional soil physical properties, and considering the impact of crop growth stages through remote sensing data to enhance the accuracy and applicability of SMC prediction models.

**Author Contributions:** Conceptualization, A.N. and T.A.; methodology, T.A. and A.N.; validation, T.A., M.N. and A.N.; formal analysis, T.A.; investigation, T.A.; data curation, T.A.; writing—original draft preparation, T.A.; writing—review and editing, A.N. and M.N.; visualization, T.A.; supervision, M.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The datasets for this study would be available by the authors upon request.

**Acknowledgments:** The research was supported by the EKÖP-24-3-I-Sze-107 University Research Fellowship Program of the Ministry for Culture and Innovation from the source of the national research, development, and innovation fund, and 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 Sciences.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Babaeian, E.; Paheding, S.; Siddique, N.; Devabhaktuni, V.K.; Tuller, M. Estimation of Root Zone Soil Moisture from Ground and Remotely Sensed Soil Information with Multisensor Data Fusion and Automated Machine Learning. *Remote Sens. Environ.* **2021**, *260*, 112434. [CrossRef]
2. Ågren, A.M.; Larson, J.; Paul, S.S.; Laudon, H.; Lidberg, W. 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. [CrossRef]
3. Mane, S.; Das, N.; Singh, G.; Cosh, M.; Dong, Y. Advancements in Dielectric Soil Moisture Sensor Calibration: A Comprehensive Review of Methods and Techniques. *Comput. Electron. Agric.* **2024**, *218*, 108686. [CrossRef]
4. Tarolli, P.; Zhao, W. Drought in Agriculture: Preservation, Adaptation, Migration. *Innov. Geosci.* **2023**, *1*, 100002. [CrossRef]
5. Tsai, P.-H.; Huang, Y.; Tai, J.-H. Estimating Soil Water Content from Thermal Images with an Artificial Neural Network. *Catena* **2024**, *241*, 108029. [CrossRef]
6. Acharya, B.; Sharma, V. Comparative analysis of soil and water dynamics in conventional and sod-based crop rotation in Florida. *Front. Agron.* **2025**, *7*, 1552425. [CrossRef]
7. Markham, A.; Trigoni, N. Magneto-Inductive Networked Rescue System (MINERS). In Proceedings of the 11th International Conference on Information Processing in Sensor Networks, Beijing China, 16–20 April 2012; ACM: New York, NY, USA, 2012; pp. 317–328. [CrossRef]
8. Basara, J.B. Soil Moisture Observations for Flash Flood Research and Prediction. In *Coping with Flash Floods*; Springer: Dordrecht, The Netherlands, 2001; pp. 231–241. [CrossRef]
9. Zhang, X.; Andreyev, A.; Zumpf, C.; Negri, M.C.; Guha, S.; Ghosh, M. Thoreau: A Subterranean Wireless Sensing Network for Agriculture and the Environment. In Proceedings of the 2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Atlanta, GA, USA, 1–4 May 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 78–84. [CrossRef]
10. Stroobosscher, Z.J.; Athelly, A.; Guzmán, S.M. Assessing Capacitance Soil Moisture Sensor Probes' Ability to Sense Nitrogen, Phosphorus, and Potassium Using Volumetric Ion Content. *Front. Agron.* **2024**, *6*, 1346946. [CrossRef]
11. Gurung, S.; Thakur, S.; Smithers, B.; Acevedo, M. Wireless Soil Moisture Sensor Networks for Agriculture. In Proceedings of the 2020 Waste-Management Education Research (WERC), Las Cruces, NM, USA, 12–15 April 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–9. [CrossRef]
12. Neményi, M.; Ambrus, B.; Teschner, G.; Alahmad, T.; Nyéki, A.; Kovács, A.J. Challenges of Ecocentric Sustainable Development in Agriculture with Special Regard to the Internet of Things (IoT), an ICT Perspective. *Prog. Agric. Eng. Sci.* **2023**, *19*, 113–122. [CrossRef]

13. Alahmad, T.; Neményi, M.; Nyéki, A. Applying IoT Sensors and Big Data to Improve Precision Crop Production: A Review. *Agronomy* **2023**, *13*, 2603. [CrossRef]
14. Maraveas, C.; Piromalis, D.; Arvanitis, K.G.; Bartzanas, T.; Loukatos, D. Applications of IoT for Optimized Greenhouse Environment and Resources Management. *Comput. Electron. Agric.* **2022**, *198*, 106993. [CrossRef]
15. Wu, Y.; Yang, Z.; Liu, Y. Internet-of-Things-Based Multiple-Sensor Monitoring System for Soil Information Diagnosis Using a Smartphone. *Micromachines* **2023**, *14*, 1395. [CrossRef] [PubMed]
16. Songara, J.C.; Patel, J.N. Calibration and Comparison of Various Sensors for Soil Moisture Measurement. *Measurement* **2022**, *197*, 111301. [CrossRef]
17. Li, Z.-L.; Leng, P.; Zhou, C.; Chen, K.-S.; Zhou, F.-C.; Shang, G.-F. Soil Moisture Retrieval from Remote Sensing Measurements: Current Knowledge and Directions for the Future. *Earth Sci. Rev.* **2021**, *218*, 103673. [CrossRef]
18. Zhu, S.; Cui, N.; Jin, H.; Jin, X.; Guo, L.; Jiang, S.; Wu, Z.; Lv, M.; Chen, F.; Liu, Q.; et al. Optimization of Multi-Dimensional Indices for Kiwifruit Orchard Soil Moisture Content Estimation Using UAV and Ground Multi-Sensors. *Agric. Water Manag.* **2024**, *294*, 108705. [CrossRef]
19. Zheng, J.; Jin, X.; Li, Q.; Lang, J.; Yin, X. Soil Moisture Variation and Affecting Factors Analysis in the Zhangjiakou–Chengde District Based on Modified Temperature Vegetation Dryness Index. *Ecol. Indic.* **2024**, *168*, 112775. [CrossRef]
20. Chandrappa, V.Y.; Ray, B.; Ashwatha, N.; Shrestha, P. Spatiotemporal Modeling to Predict Soil Moisture for Sustainable Smart Irrigation. *Internet Things* **2023**, *21*, 100671. [CrossRef]
21. Li, B.; You, Z.; Ni, K.; Wang, Y. Prediction of Soil Compaction Parameters Using Machine Learning Models. *Appl. Sci.* **2024**, *14*, 2716. [CrossRef]
22. Islam, M.N.; Logofatu, D.; Haque, M.Z. A Comparative Study on Machine Learning Methods Through Evaluating the Impact of Contributing Factors on The Accuracy of Soil Moisture Prediction. In Proceedings of the 2023 International Conference on Innovations in Intelligent Systems and Applications (INISTA), Hammamet, Tunisia, 20–23 September 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6. [CrossRef]
23. Zambudio Martínez, M.; da Silveira, L.H.M.; Marin-Perez, R.; Gomez, A.F.S. Development and Comparison of Artificial Neural Networks and Gradient Boosting Regressors for Predicting Topsoil Moisture Using Forecast Data. *AI* **2025**, *6*, 41. [CrossRef]
24. Ren, Y.; Ling, F.; Wang, Y. Research on Provincial-Level Soil Moisture Prediction Based on Extreme Gradient Boosting Model. *Agriculture* **2023**, *13*, 927. [CrossRef]
25. Celik, M.F.; Isik, M.S.; Yuzugullu, O.; Fajraoui, N.; Erten, E. Soil Moisture Prediction from Remote Sensing Images Coupled with Climate, Soil Texture and Topography via Deep Learning. *Remote Sens.* **2022**, *14*, 5584. [CrossRef]
26. Fu, R.; Xie, L.; Liu, T.; Zheng, B.; Zhang, Y.; Hu, S. A Soil Moisture Prediction Model, Based on Depth and Water Balance Equation: A Case Study of the Xilingol League Grassland. *Int. J. Environ. Res. Public Health* **2023**, *20*, 1374. [CrossRef] [PubMed]
27. United Nations. UN-Water Annual Report 2023. 2024. Available online: https://www.unwater.org/publications/un-water-annual-report-2023 (accessed on 4 November 2024).
28. Popp, J.; Oláh, J.; Neményi, M.; Nyéki, A. Global Challenges and the 'Farm to Fork' Strategies of the European Green Deal: Blessing or Curse. *Prog. Agric. Eng. Sci.* **2024**, *20*, 101–111. [CrossRef]
29. Nyéki, A.; Daróczy, B.; Kerepesi, C.; Neményi, M.; Kovács, A.J. Spatial Variability of Soil Properties and Its Effect on Maize Yields within Field—A Case Study in Hungary. *Agronomy* **2022**, *12*, 395. [CrossRef]
30. USDA. *Natural Resources Conservation Service. Soil Classification*; Retrieved 17 October 2020; 2020. Available online: https://www.nrcs.usda.gov/wps/portal/nrcs/main/soils/survey/class/ (accessed on 29 July 2024).
31. Dirksen, C. *Soil Physics Measurements*; Catena Verlag: Stuttgart, Germany, 1999; ISBN 3923381433.
32. Ehtasham, L.; Sherani, S.H.; Nawaz, F. Acceleration of the Hydrological Cycle and Its Impact on Water Availability over Land: An Adverse Effect of Climate Change. *Meteorol. Hydrol. Water Manag.* **2024**, *12*, 1–21. [CrossRef]
33. Moghadas, D.; Jadoon, K.Z.; Vanderborght, J.; Lambot, S.; Vereecken, H. Effects of Near Surface Soil Moisture Profiles During Evaporation on Far-Field Ground-Penetrating Radar Data: A Numerical Study. *Vadose Zone J.* **2013**, *12*, 1–11. [CrossRef]
34. Granata, F. Evapotranspiration Evaluation Models Based on Machine Learning Algorithms—A Comparative Study. *Agric. Water Manag.* **2019**, *217*, 303–315. [CrossRef]
35. Du, M.; Zhang, J.; Wang, Y.; Liu, H.; Wang, Z.; Liu, C.; Yang, Q.; Hu, Y.; Bao, Z.; Liu, Y.; et al. Evaluating the Contribution of Different Environmental Drivers to Changes in Evapotranspiration and Soil Moisture, a Case Study of the Wudaogou Experimental Station. *J. Contam. Hydrol.* **2021**, *243*, 103912. [CrossRef]
36. Python Software Foundation. Python Language Reference, Version 3.10. 2024. Available online: https://www.python.org (accessed on 11 November 2024).
37. Hussain, M.; Sharmin, N.; Shafiul, S.B. Estimation of Soil Moisture with Meteorological Variables in Supervised Machine Learning Models. In Proceedings of the 2023 International Conference on Electrical, Computer and Communication Engineering (ECCE), Chittagong, Bangladesh, 23–25 February 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6. [CrossRef]

38. Wang, S.; Fu, G. Modelling Soil Moisture Using Climate Data and Normalized Difference Vegetation Index Based on Nine Algorithms in Alpine Grasslands. *Front. Environ. Sci.* **2023**, *11*, 1130448. [CrossRef]
39. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
40. Natekin, A.; Knoll, A. Gradient Boosting Machines, a Tutorial. *Front. Neurorobot.* **2013**, *7*, 21. [CrossRef]
41. Willmott, C.J.; Ackleson, S.G.; Davis, R.E.; Feddema, J.J.; Klink, K.M.; Legates, D.R.; O'Donnell, J.; Rowe, C.M. Statistics for the Evaluation and Comparison of Models. *J. Geophys. Res. Ocean.* **1985**, *90*, 8995–9005. [CrossRef]
42. Nash, J.E.; Sutcliffe, J.V. River Flow Forecasting through Conceptual Models Part I—A Discussion of Principles. *J. Hydrol.* **1970**, *10*, 282–290. [CrossRef]
43. Wright, S. Correlation and Causation. *J. Agric. Res.* **1921**, *20*, 557.
44. Guo, X.; Fu, Q.; Hang, Y.; Lu, H.; Gao, F.; Si, J. Spatial Variability of Soil Moisture in Relation to Land Use Types and Topographic Features on Hillslopes in the Black Soil (Mollisols) Area of Northeast China. *Sustainability* **2020**, *12*, 3552. [CrossRef]
45. Zhao, W.; Fang, X.; Daryanto, S.; Zhang, X.; Wang, Y. Factors Influencing Soil Moisture in the Loess Plateau, China: A Review. *Environ. Sci. Trans. R. Soc. Edinb.* **2018**, *109*, 501–509. [CrossRef]
46. Kandala, R.; Franssen, H.H.; Chaudhuri, A.; Sekhar, M. The Value of Soil Temperature Data versus Soil Moisture Data for State, Parameter, and Flux Estimation in Unsaturated Flow Model. *Vadose Zone J.* **2024**, *23*, e20298. [CrossRef]
47. Sehler, R.; Li, J.; Reager, J.; Ye, H. Investigating Relationship Between Soil Moisture and Precipitation Globally Using Remote Sensing Observations. *J. Contemp. Water Res. Educ.* **2019**, *168*, 106–118. [CrossRef]
48. McColl, K.A.; Tang, L.I. An Analytic Theory of Near-Surface Relative Humidity over Land. *J. Clim.* **2024**, *37*, 1213–1230. [CrossRef]
49. Leng, P.; Li, Z.-L.; Duan, S.-B.; Gao, M.-F.; Huo, H.-Y. A Practical Approach for Deriving All-Weather Soil Moisture Content Using Combined Satellite and Meteorological Data. *ISPRS J. Photogramm. Remote Sens.* **2017**, *131*, 40–51. [CrossRef]
50. Ines, A.V.M.; Das, N.N.; Hansen, J.W.; Njoku, E.G. Assimilation of Remotely Sensed Soil Moisture and Vegetation with a Crop Simulation Model for Maize Yield Prediction. *Remote Sens. Environ.* **2013**, *138*, 149–164. [CrossRef]
51. Zhang, Z.; Chen, X.; Pan, Z.; Zhao, P.; Zhang, J.; Jiang, K.; Wang, J.; Han, G.; Song, Y.; Huang, N.; et al. Quantitative Estimation of the Effects of Soil Moisture on Temperature Using a Soil Water and Heat Coupling Model. *Agriculture* **2022**, *12*, 1371. [CrossRef]
52. Meng, X.; Zeng, J.; Yang, Y.; Zhao, W.; Ma, H.; Letu, H.; Zhu, Q.; Liu, Y.; Wang, P.; Peng, J. High-Resolution Soil Moisture Mapping through Passive Microwave Remote Sensing Downscaling. *Innov. Geosci.* **2024**, *2*, 100105. [CrossRef]

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.

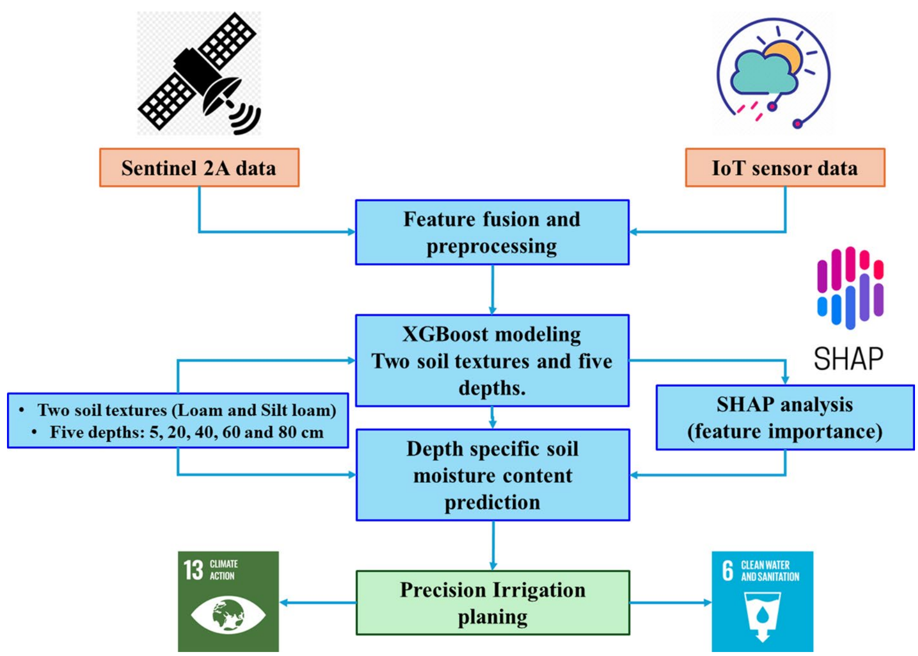
**Conclusions** The model's accuracy and interpretability enable depth-specific decision support for irrigation timing and water use efficiency under variable weather conditions, while providing actionable driver insights for climate-adaptive management aligned with SDGs 6 and 13. The approach is validated 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/m²) 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

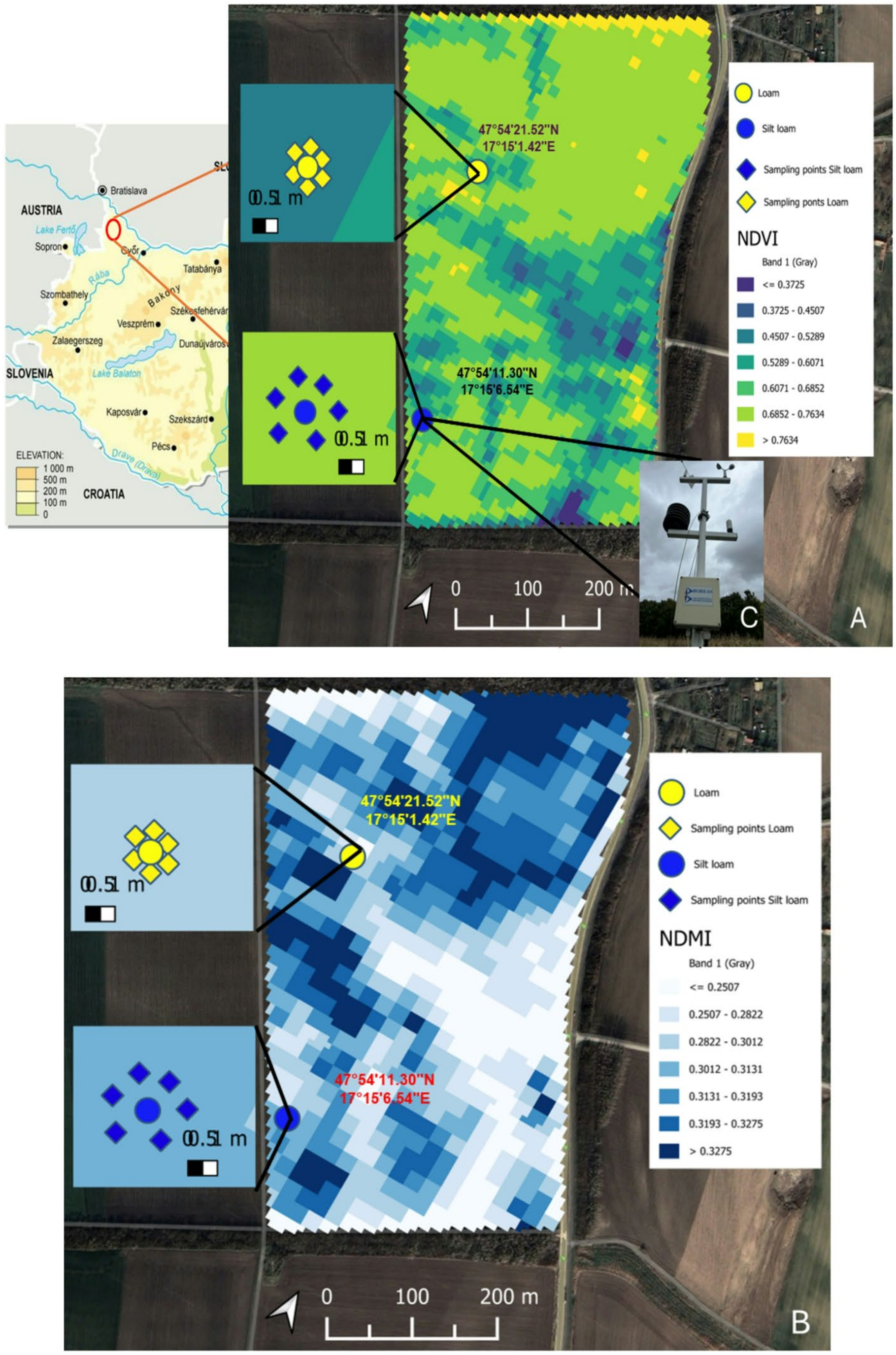



**Fig. 1** The field location from topographical map of Hungary, masked with (**A**) NDVI index and values in one date and (**B**) NDMI index and values downloaded from sentinel hub and visualized by QGIS, with sampling points and (**C**) IoT based sensor

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

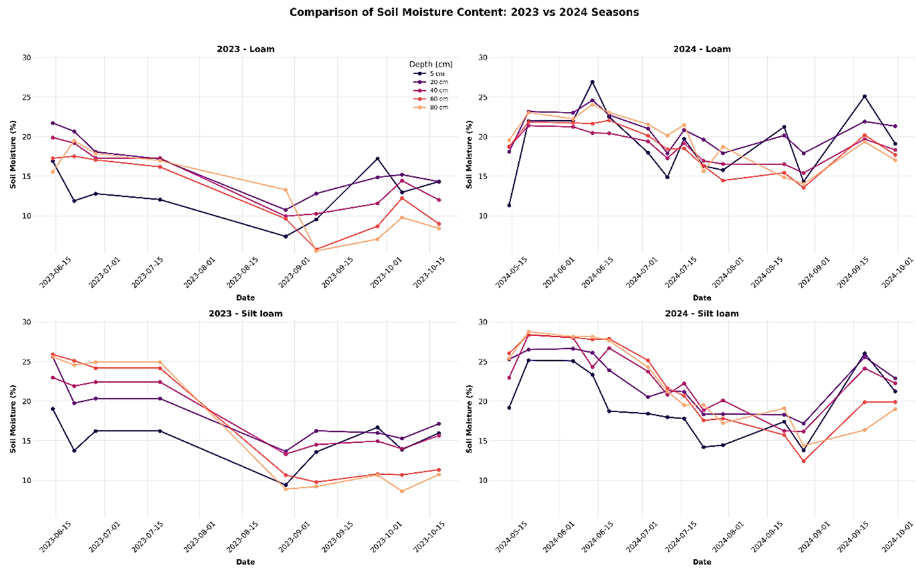



**Fig. 2** Soil moisture content variation across five depths in two soil textures (loam and silt loam) in 2023 and 2024 seasons with 23 soil sampling dates across both seasons

**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) | $MAE = {}^{1}/_{n} \sum_{t=1}^{n} \left\vert y_t - \widehat{y}t\right\vert$ | [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, $\widehat{y}t$: The predicted value by the model at time t, $y_i$ is the observed value at observation *i*; $\widehat{y}_i$ is the predicted value at observation *i*; and $\overline{y}$ 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

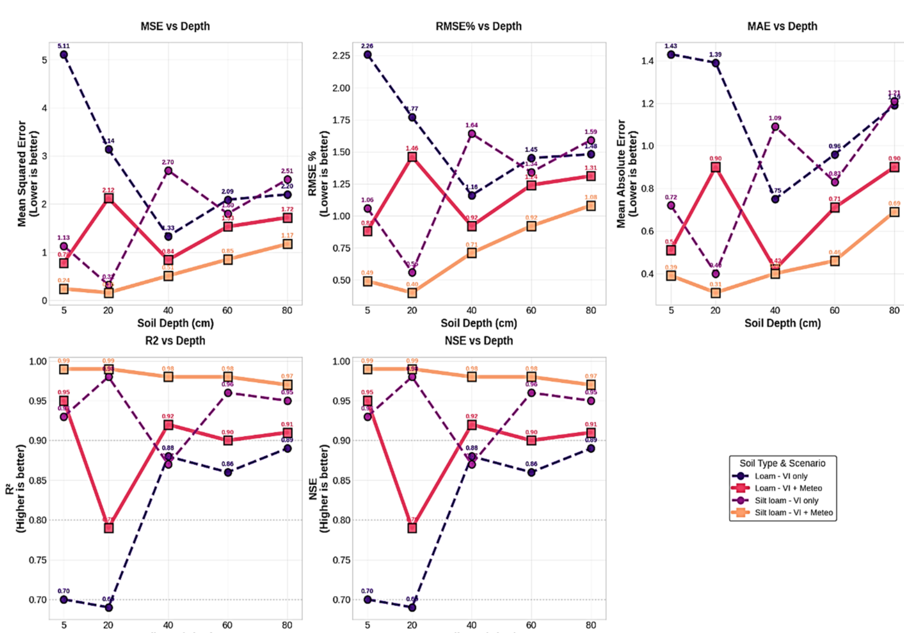



**Fig. 3** The evaluation metrics results of XGBoost model across all depths in two soil textures

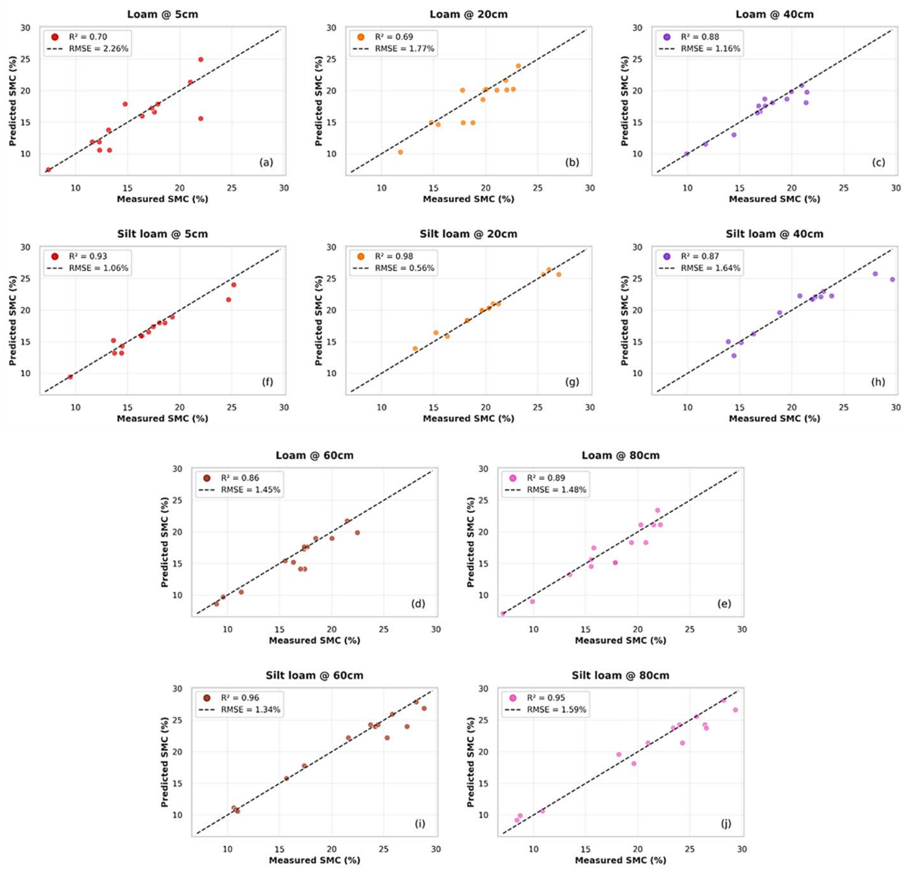



**Fig. 4** Distribution of measured vs. predicted SMC values for the test dataset (representing 20% of total data, $n = 14$ for each soil and depth, with $R^2$ and RMSE across all five depths in two soil textures with VIs only scenario

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

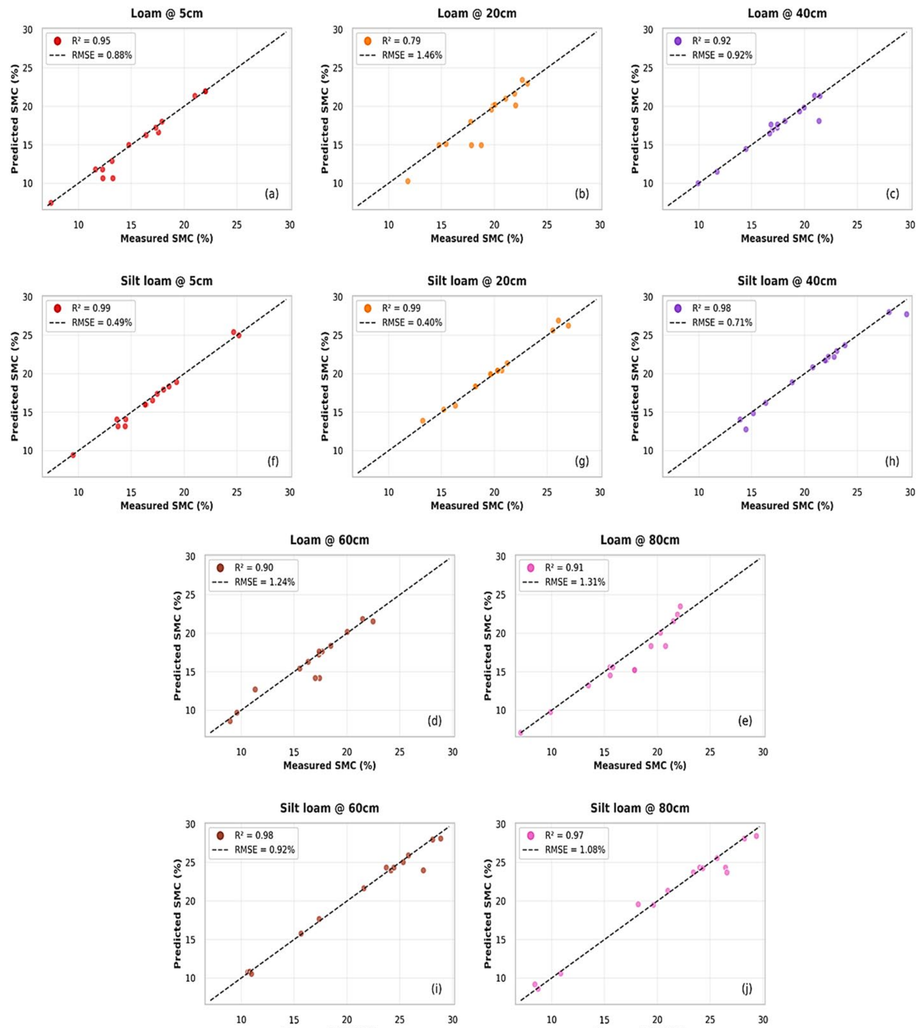



**Fig. 5** Distribution of measured vs. predicted SMC values for the test dataset (representing 20% of total data, $n = 14$ for each soil and depth, with $R^2$ and RMSE across all five depths in two soil textures with second scenario

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.

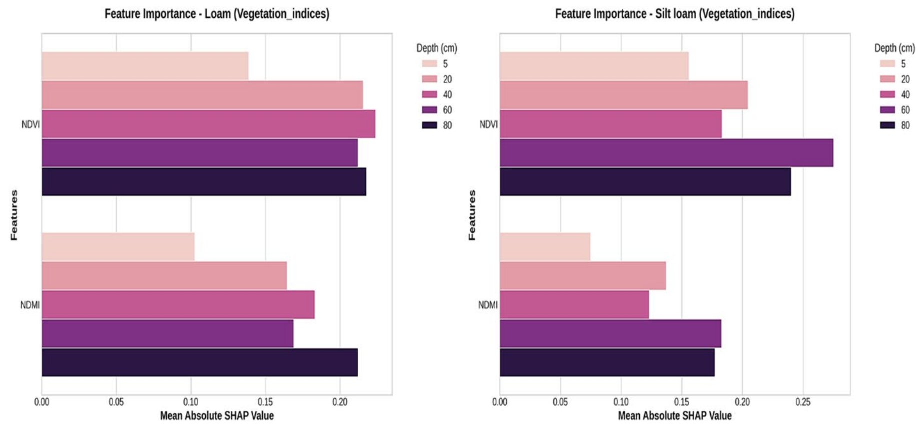



**Fig. 6** Mean SHAP values in loam and silt loam soil across the five depths under vegetation indices only scenarios

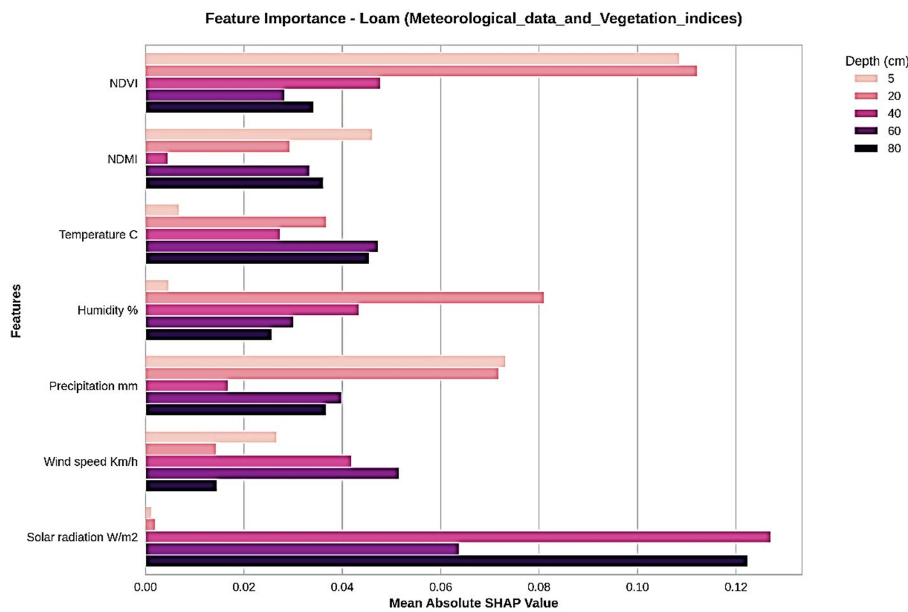



**Fig. 7** Mean SHAP values in loam soil across the five depths under combined scenarios (Meteorological and vegetation indices)

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

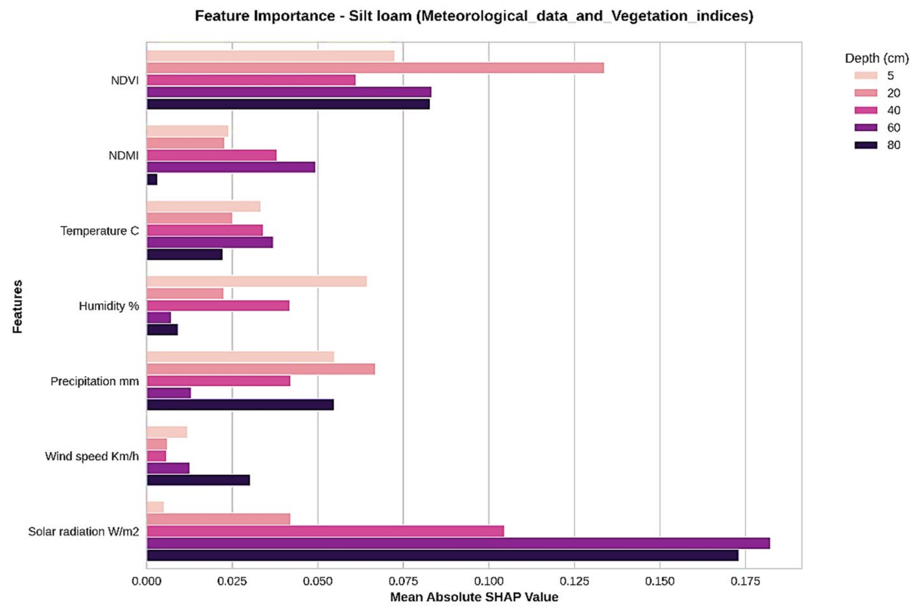



**Fig. 8** Mean SHAP values in silt loam soil across the five depths under combined scenarios (Meteorological and vegetation indices)

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.

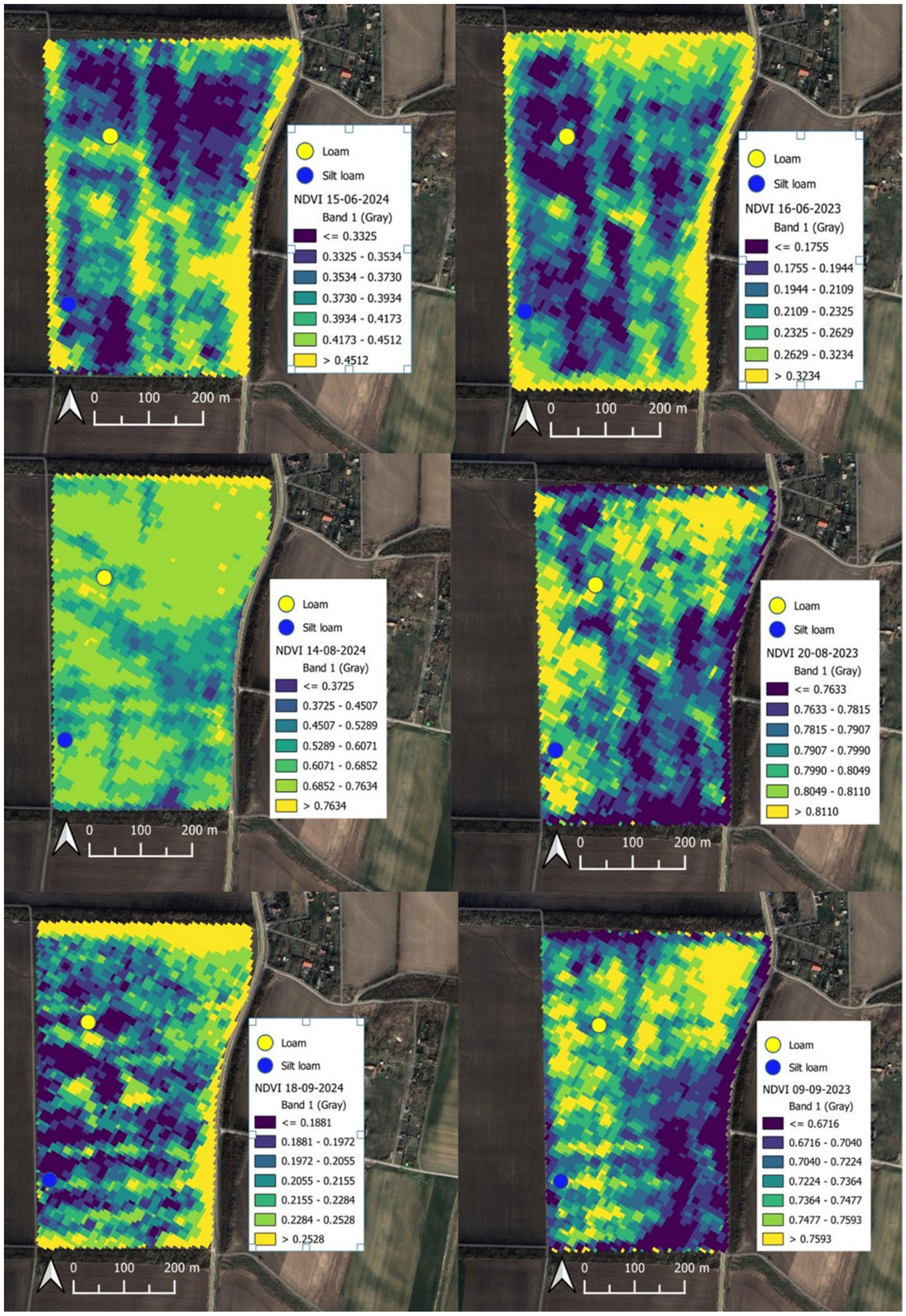
## Appendices

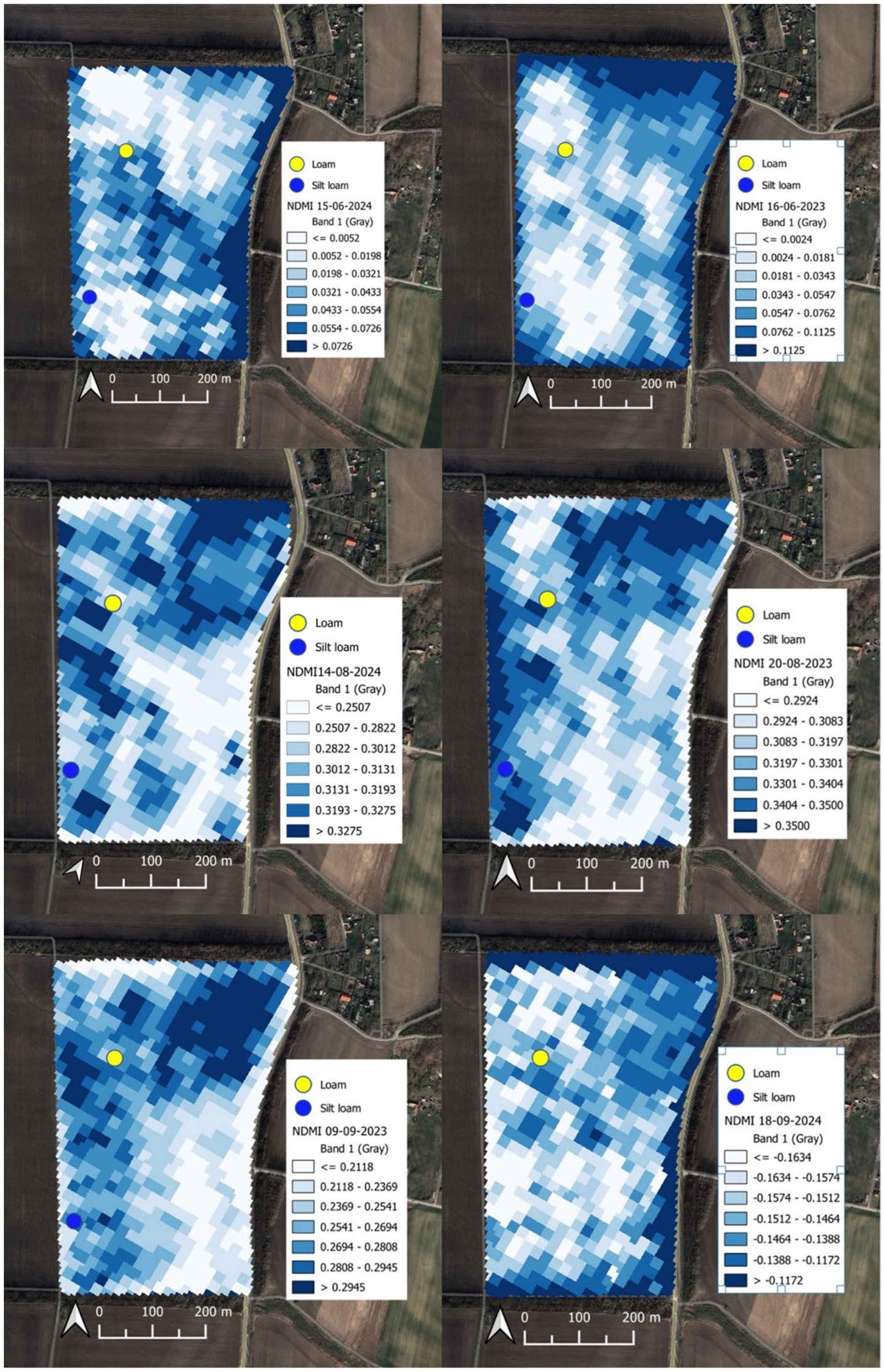
Appendix 1: The Sentinel-2 acquisition dates used to compute NDVI and NDMI for each soil sampling event.

| 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