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:

PA Precision Agriculture

GIS Geographic Information System GPS Global Positioning System ML Machine Learning

AI Artificial Intelligence

CNNs Convolutional Neural Networks IoT Internet of Things

LoRaWAN Long Range Wide Area Network NDVI Normalized Difference Vegetation Index

NDMI Normalized Difference Moisture Index UAV Unmanned Aerial Vehicle SMC Soil Moisture Content

RFR Random Forest Regression XGBoost eXtreme Gradient Boosting GBR Gradient Boosting Regressor

DL Deep Learning

LSTM Long-Short Term Memory XAI Explainable AI

SHAP SHapley Additive exPlanations SAR Synthetic Aperture Radar TDR Time-Domain Reflectometry FDR Frequency-Domain Reflectometry SMAP Soil Moisture Active Passive

SMOS Soil Moisture and Ocean Salinity CBR CatBoost Regression

CV% Coefficient of Variation OOB out-of-bag

Table of Contents:

Abstract:

Precision agriculture (PA) requires precise and accurate information on soil moisture, optimized irrigation, and sustainable water management. The current PhD dissertation 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, progressing from model evaluation towards data fusion using SHAP interpretability. 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, which were confirmed using statistical

analysis.

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%.

The current dissertation has successfully demonstrated the potential of accurate, interpretable, and action-oriented soil moisture content prediction, providing a scientific basis for precision crop water management and sustainable agriculture under climate change.

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 Gábor, 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 in (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., 2025c).

SMC plays a vital role in hydrological and ecological processes, influencing water availability, energy fluxes, and crop yields

(Alahmad et al., 2025c; Mane et al., 2024). However, traditional SMC measurements are limited in spatial resolution and frequency, while IoT-based sensors and satellite indices provide scalable, real-time alternatives (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., 2025a; 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., 2025b; 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 different predictor scenarios: (i) vegetation indices only (NDVI/NDMI), (ii) meteorological variables only, and (iii) combined predictors for depth-specific SMC prediction accuracy.
- To benchmark and optimize models by comparing machine learning approaches under a consistent train–validation–test protocol, assessing depth- and soil-specific strategies versus pooled modelling to improve generalizability.
- To apply explainable machine learning techniques (SHAP and permutation importance) to quantify global and depth-dependent feature contributions, and to translate these results into agronomically interpretable driver mechanisms for SMC variability.
- To synthesize predictive and explainability findings into actionable implications for precision crop production and irrigation decision support align with Sustainable Development Goals (SDGs) 6 (Clean Water and Sanitation) and 13 (Climate Action).

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 Acharya, 2025). Typically measured gravimetrically (mass-based) or volumetrically, SMC determines the amount of plant-available water and is thus essential for irrigation planning. 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., 2025a; 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., 2025c). 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., 2025b).

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). For instance, Islam et al. (2023) 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±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. 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, real-time 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

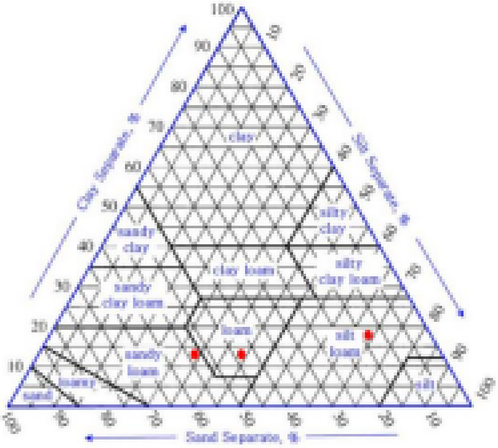


Figure 1. Soil textures classes (silt, clay and sand) for each studied soil. Soil samples were collected at fixed points near IoT sensor stations (within 1 m²) at five depths (5, 20, 40, 60, 80 cm) with three replicates per point. Sampling occurred biweekly in 2023 (9 dates) and on 23 dates across 2023–2024 for the two-season dataset, totaling 705

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:

�� =����

����× ������ (1)

Where:

- θ = Soil Moisture Content (%)
- Mw = Mass of water (g) = (Wet weight – Dry weight) •

Md = Mass of dry soil (g)

The total number of gravimetric samples collected across all studies was 705, comprising: Article 1: 405 samples (3 soil textures × 5 depths × 3 replicates × 9 dates) | Article 2: 300 samples (2 soil textures × 5 depths × 3 replicates × 10 dates) | Article 3: 705 samples (2 soil textures

× 5 depths × 3 replicates × 23 dates across two seasons).

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.

3.2.2. Satellite-derived vegetation indices

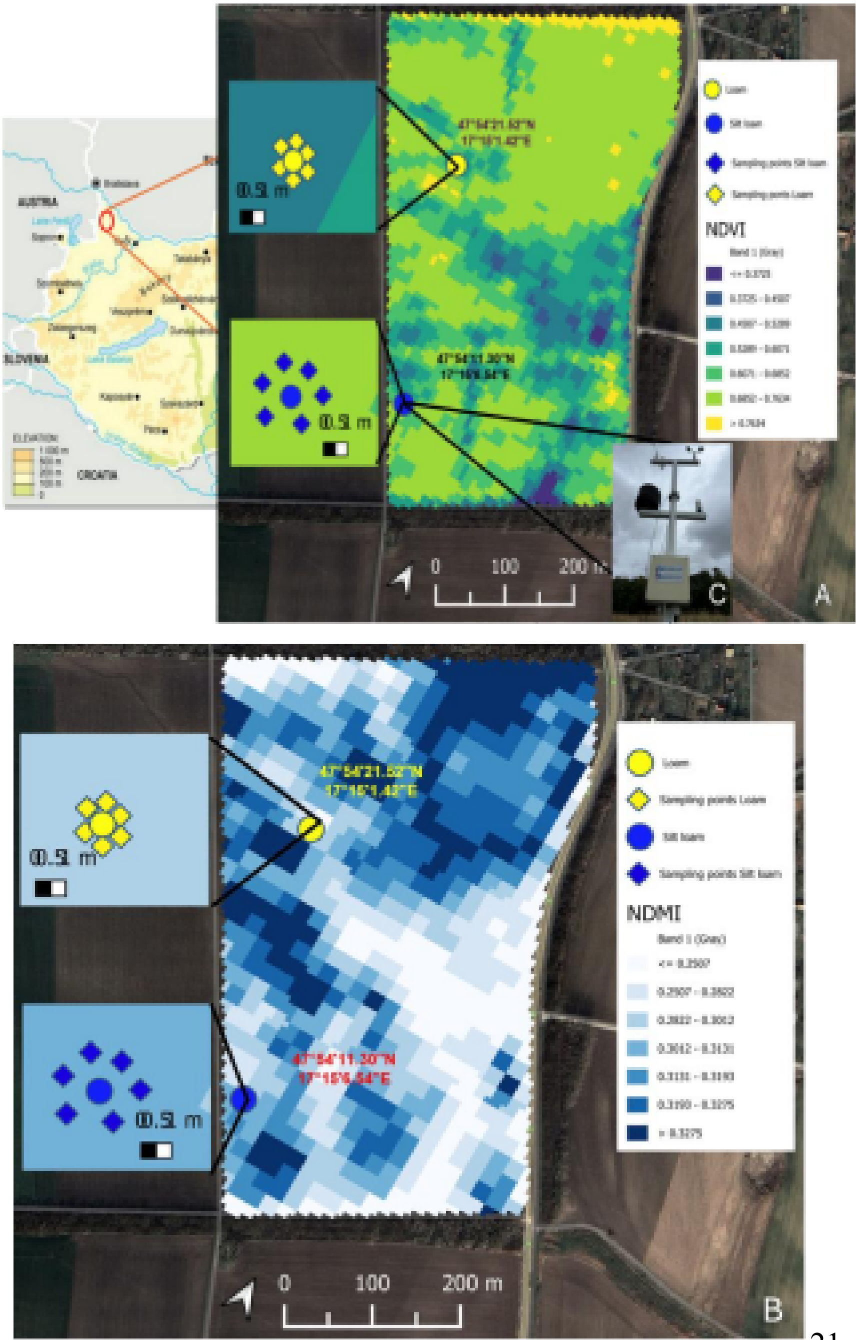
Sentinel-2 multispectral imagery was utilized to derive vegetation indices indicative of crop canopy status as proxies for soil moisture conditions. Sentinel-2 Level-2A images closest to each sampling date (within ±3 days, cloud-free) were processed in QGIS (Figure 2). (see Appendix 1 and 2 for complete 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.

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

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

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

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



NDVI
0.5 m
0 100 200 m
C
A
NDMI
47°54'21.52"N
17°15'1.42"E
47°54'11.30"N
17°15'6.54"E
0.5 m
0 100 200 m
B

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.

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 (Article 1: Table 1and Article 3: Table 2).

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).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

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

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

- 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.

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

- 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.

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

Where: n: the total number of observations in the dataset, i: The time index or sequence index of an observation, y_i: The observed value at time t, ����: The predicted value by the model at time t, ����is the

observed value at observation ��; ����is the predicted value at observation ��; and �� 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, the results clearly showed that both soil texture and depth play a major role in determining how moisture is retained in the soil (Figure 3). Silt loam consistently recorded the highest moisture levels, particularly in the deeper layers. Loam soils showed moderate moisture levels, with the most favorable conditions observed around 20 cm depth. In contrast, sandy loam had the lowest overall moisture content, reaching its highest values at 80 cm depth and showing relatively little variation across depths.

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).

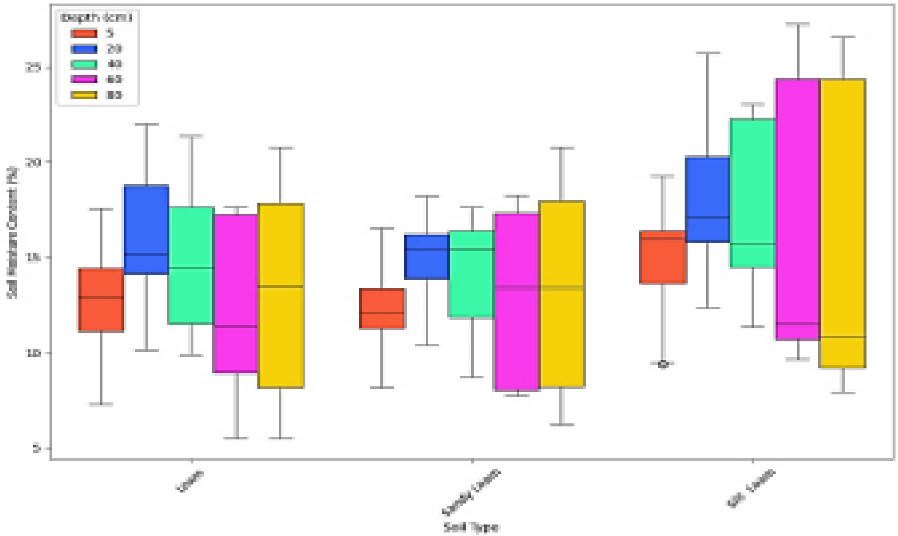


Figure 3. Soil moisture content variation in three soil textures at five different depths during 2023 season.

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 1). 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 1: 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). The study also highlighted that shallow soil layers (5–20 cm) respond quickly to meteorological factors such as rainfall and temperature, 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).

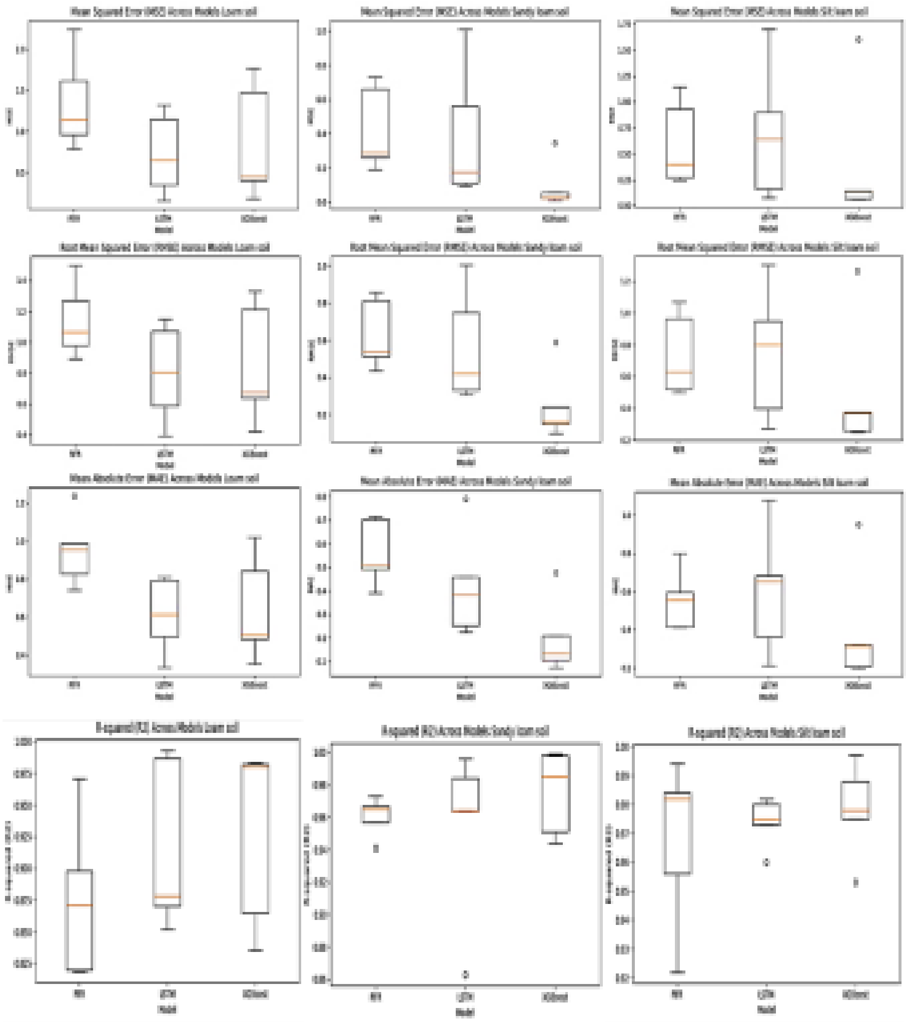


Figure 4. Predictive performance and validation results (R^2, MSE, RMSE, And MAE) for all soil textures with three models.

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).

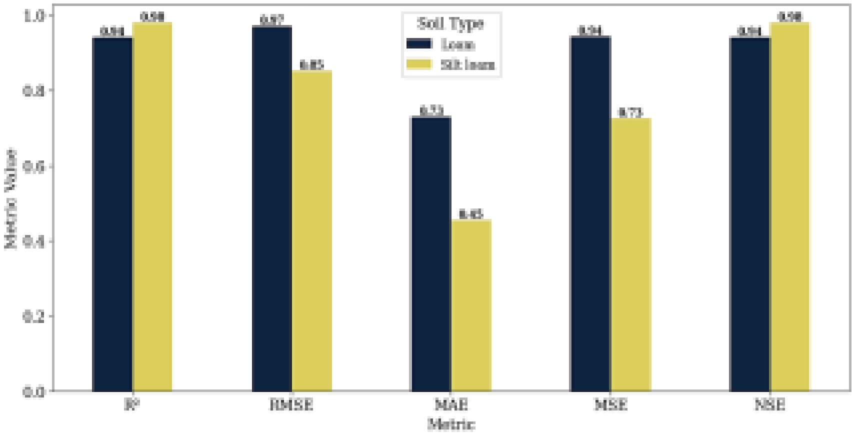


Figure 5. Evaluation metrics (R^2, MSE, RMSE, MAE, and NSE) for soil moisture content prediction in loam and silt loam soils using the GBR model.

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 10).

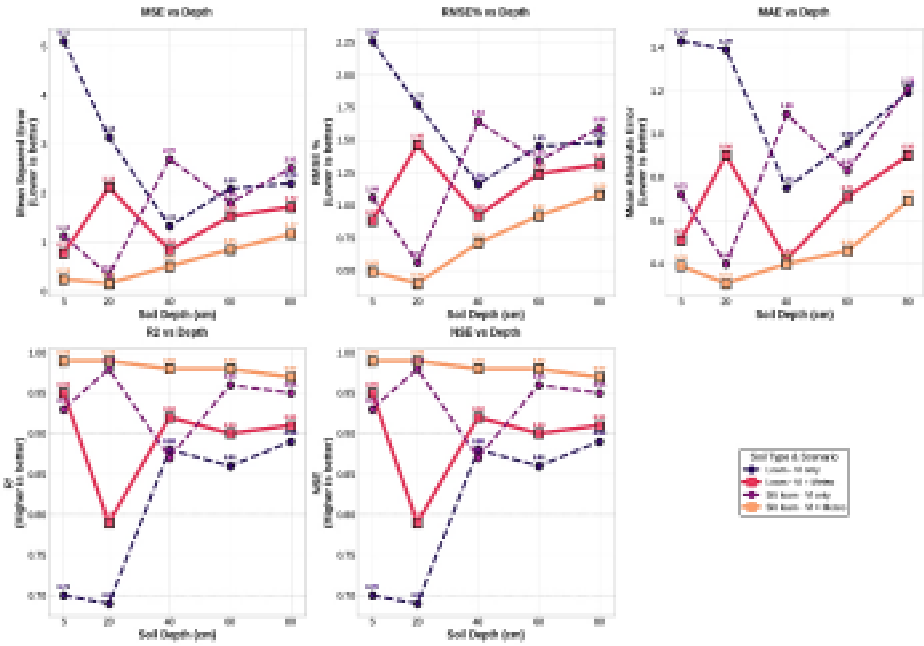


Figure. 10 The evaluation metrics results of XGBoost model across all depths in two soil textures.

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 achieve high accuracy in stable, fine-textured silt loam 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. 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 (UN, 2023; United Nation, 2024).

5.6. Limitations and future research directions

5.7.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.

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. Since the variation in soil moisture with depth is significantly affected by soil texture, the model should vary with depth and soil texture rather than using a one-size-fits-all approach.
2. Among all the algorithms, the best performing general model was found to be the Random Forest Regression model in predicting soil moisture levels at different depths and soil textures. This model was able to obtain high R^2 values ranging from 0.82 to 0.99 and low RMSE values less than 1.5% under all conditions. The best model to use in predicting soil moisture levels at different depths and soil textures was found to be the Random Forest model. However, the best model to use depends on the depth at which one wants to carry out the predictions. XGBoost was also seen to perform well in the shallower zones, and LSTM was seen to perform well in the upper and middle layers. However, these two algorithms were seen to lose their efficiency in deeper layers.
3. The primary environmental factors affecting soil moisture vary

quantitatively in different soil textures. In loam soil, the two factors affecting soil moisture levels were found to be precipitation and humidity. Both these factors were seen to contribute 31% to the predictions. Similarly, in the case of silt loam soil, solar radiation was seen to dominate in affecting soil moisture levels by having the highest importance at 46%, followed by wind speed at 26%.

4. The integration of Sentinel-2 vegetation indices with IoT meteorological data enhanced the accuracy of predictions significantly compared to vegetation index predictions alone. The

improvement was more pronounced in loam soils at shallow depth, with an improvement in model accuracy by more than 60%, reducing the RMSE by 2.26% to 0.88% at 5 cm depth.

5. Depth-dependent factors influencing soil moisture content were quantified using explainable AI. Results indicated texture dependent hierarchies in influencing factors. For loam soils, precipitation and humidity were the dominant factors at 31%, while in Silt Loam soils, solar radiation was the dominant factor at 46%, followed by wind speed at 26%. SHAP analysis indicated that in both soils, NDVI was the most important predictor near the surface, while solar radiation was the dominant predictor at depth (60-80 cm).

6. By integrating gravimetric SMC data with high-frequency meteorological data and satellite vegetation data, a strong foundation exists to develop depth-dependent ML models that can capture both high-frequency surface dynamics and deeper soil layer responses.

7. The integration of explainable machine learning with field data,

meteorological data, and satellite data creates a strong scientific foundation to develop depth-dependent irrigation recommendations in precision crop production.

7.2. Practical Theses:

1. Irrigation scheduling for loam soils should incorporate real-time meteorology in addition to canopy and deep layer prediction to differentiate between whole-profile depletion and surface drying. For silt loam soils, the close relationship between vegetation

indices and soil moisture allows for scheduling of irrigation mainly using satellite-derived NDVI trends, with some deep layer verification under extreme weather events.

2. Solar radiation is another important factor that influences the loss of deep soil moisture. During times of increased evaporative demand, the scheduling of irrigation should take into account the total amount of solar radiation exposure rather than relying solely on rainfall or surface layer information. This is especially true for loam soils because of the rapid propagation of atmospheric information through the profile.

3. Soil-type-specific approaches to monitoring will enable maximum efficiency in terms of cost and accuracy. For loam soils, there is a need to use integrated sensor-satellite systems with real-time meteorology to capture rapid changes in soil moisture. For silt loam soils, there is a possibility of using satellite-derived monitoring with some validation under extreme weather events.

4. The outputs of Explainable AI can be converted into simple decision rules. The SHAP tool can be used to develop heuristics, e.g., (i) when surface NDVI is high, and solar radiation is low, do

not irrigate even if surface soil is dry; (ii) when there have been 3-5 consecutive days of high solar radiation without appreciable rain, anticipate soil dehydration in the deep soil profile and irrigate regardless of surface soil status. These heuristics can be used by farmers.

5. Precision irrigation systems require not only technology but also expertise on the part of farmers. High-level expertise in agronomy

is always needed to interpret the outputs of precision systems, understand when expert advice can be overridden, and develop general rules based on specific situations. Precision technology can only be valuable when farmers have the expertise needed to use it critically.

6. A scalable architecture can be built on existing infrastructure. The proposed framework can be implemented using (i) existing free Sentinel-2 satellite images, (ii) existing low-cost IoT weather stations using LoRaWAN technology, (iii) open-source Python libraries (scikit-learn, XGBoost, SHAP) for machine learning, and (iv) cloud or edge computing technology for prediction delivery. This architecture can be used to provide expert decision support automatically, as well as being interpretable and understandable by farmers.

7. The modeling framework provides a proactive approach, not a reactive approach, in managing irrigation schedules. The system can predict moisture stress in the soil using real-time IoT sensors and weather forecasts, thus allowing farmers to schedule irrigation at times when it is most beneficial.

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.

Alahmad, T., Neményi, M., Nyéki, A., 2025a. Soil Moisture Content Prediction Using Gradient Boosting Regressor (GBR) Model:

Soil-Specific Modeling with Five Depths. Applied Sciences 15, 5889.

Alahmad, T., Neményi, M., Nyéki, A., 2025b. 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.

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

Alahmad, T., Neményi, M., Széles, A., Ali, N., Hijazi, O., Nyéki, A., 2025c. Spatiotemporal prediction of soil moisture content at various depths in three soil types using machine learning algorithms. Frontiers in Soil Science 5.

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.

Ambrus Bálint, 2023. ADATGYŰJTÉSI TECHNOLÓGIÁK

FEJLESZTÉSE A PRECÍZIÓS NÖVÉNYTERMESZTÉSNÉL, KÜLÖNÖS TEKINTETTEL A MAGAJÁRÓ BERENDEZÉSEKRE. Széchenyi István Egyetem Albert Kázmér Mosonmagyaróvári Kar.

Anikó, N., Tarek, A., Arshad, S., Nóra, G., Miklós, N., 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.

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.

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.

Bibek Acharya, V.S., 2025. COMPARATIVE ANALYSIS OF SOIL AND WATER DYNAMICS IN CONVENTIONAL AND SOD BASED CROP ROTATION IN FLORIDA. Frontiers in Agronomy 7.

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.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32.

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.

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.

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.

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.

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.

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.

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.

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.
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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

Kheimi, M., Zounemat-Kermani, M., 2025. Conventional and advanced AI-based models in soil moisture prediction. Physics and Chemistry of the Earth 139.

Li, C., Zhang, Y., Ren, X., 2020. Modeling Hourly Soil Temperature Using Deep BiLSTM Neural Network. Algorithms 13, 173. 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.

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.

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.

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.

McColl, K.A., Tang, L.I., 2024. An Analytic Theory of Near-Surface Relative Humidity over Land. J. Clim. 37, 1213–1230. Milics Gábor, 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.

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.

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.

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.

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.

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.

Nyéki, A., 2016. A precíziós növénytermesztés és a fenntartható mezőgazdaság kapcsolata. DOKTORI (PhD) ÉRTEKEZÉS TÉZISEI

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.

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.

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.

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.

Ren, Y., Ling, F., Wang, Y., 2023. Research on Provincial-Level Soil Moisture Prediction Based on Extreme Gradient Boosting Model. Agriculture 13, 927.

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.

Songara, J.C., Patel, J.N., 2022. Calibration and comparison of various sensors for soil moisture measurement. Measurement 197, 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.

Taheri, M., Bigdeli, M., Imanian, H., Mohammadian, A., 2025. An Overview of Machine-Learning Methods for Soil Moisture Estimation. Water (Basel). 17, 1638.

Tarolli, P., Zhao, W., 2023. Drought in agriculture: Preservation, adaptation, migration. The Innovation Geoscience 1, 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.

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

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

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.

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.

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.

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.

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.

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.

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.

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.

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.

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. Tarek Alahmad Miklós Neményi, Adrienn Széles, Nour Ali, Omar Hijazi, Anikó Nyéki (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 (Q2, 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 Anikó, Alahmad Tarek, Sana Arshad, Gombkötő Nóra, Neményi Miklós, Morad Mirzaei, Szabó Szilárd, Endre Harsanyi, Main Al-Dalahmeh, Safwan Mohammed. (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. B Ambrus, G Teschner, AJ Kovács, M Neményi, L Helyes, Z Pék, S Takács, T Alahmad, A Nyéki (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 Miklós, Ambrus Bálint, Teschner Gergely, Alahmad Tarek, Nyéki Anikó, Kovács Attila J. (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. Anikó Nyéki, Bálint Daróczy, Miklós Neményi, Bálint Ambrus, Gergely Teschner, and Tarek Alahmad. (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. Nour Ali, Anikó Nyéki, Aïda Jalloul, Tarek Alahmad. (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

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, Neme´ nyi M, Sze´ les A, Ali N, Hijazi O and Nye´ ki 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

TYPE Original Research
PUBLISHED 30 September 2025
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*], Miklo´ s Neme´nyi [1], Adrienn Sze´ les [2], Nour Ali [3], Omar Hijazi [4] and Aniko´ Nye´ ki [1]

[1]Department of Biosystems Engineering and Precision Technology, Albert Ka´ zme´ r Mosonmagyaro´ va´r Faculty of Agricultural and Food Sciences, Sze´ chenyi Istva´ n University, Mosonmagyaro´ va´r, 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

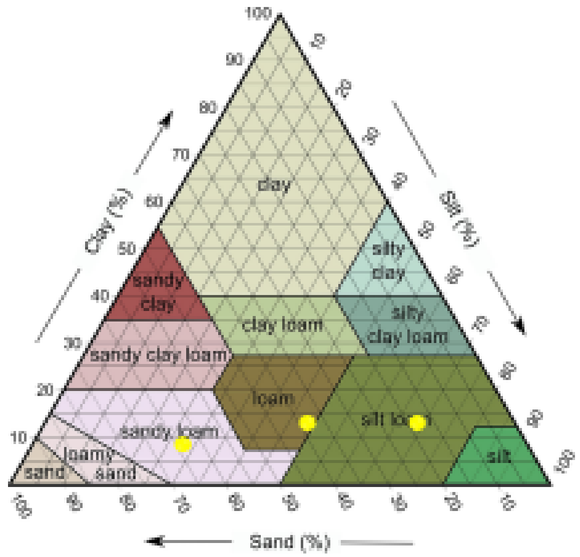


FIGURE 1
Soil texture classes (silt, clay, and sand) for each studied soil.

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:

$$q = \frac{Mw}{Md} \, 100 \quad (1)$$

where:

- q = 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.

Loam	135	7.61	40.99–49.01	40.91–45.97	10.10–14.50	12.61–16.19	3.10–4.94
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
Silt loam	135	7.45	12.36–15.31	63.70–70.19	17.40–21	14.99–18.27	2.62–7.87

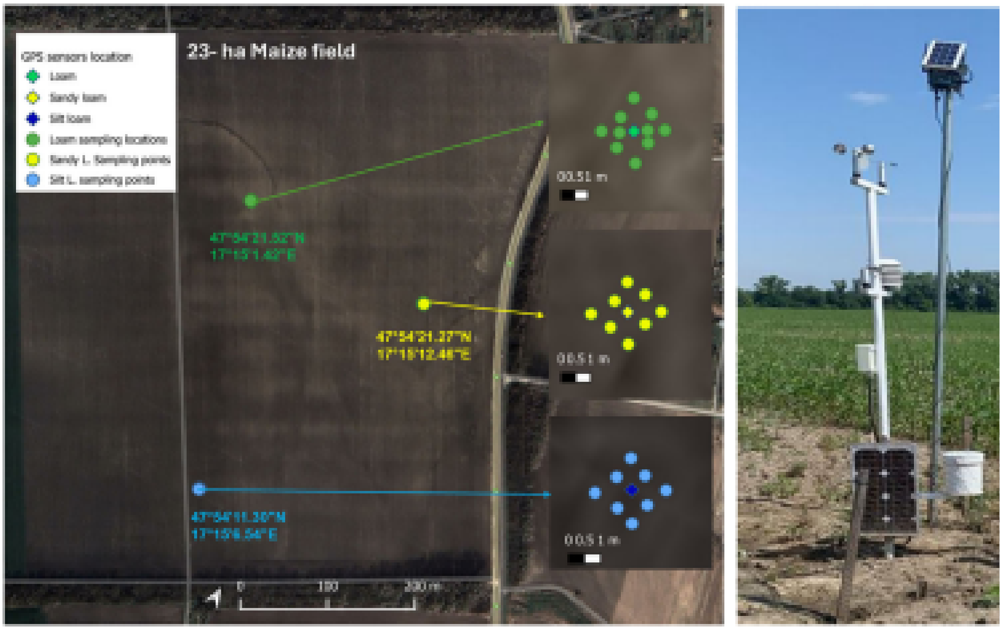


FIGURE 2
(A) Sensor station's locations at the field with soil types, GPS coordinates, and spatial distribution of sensor stations and corres locations within each identified soil type. (B) IoT sensor stations based in the field (wind sensor, precipitation sensor, solar radia sensor, and temperature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

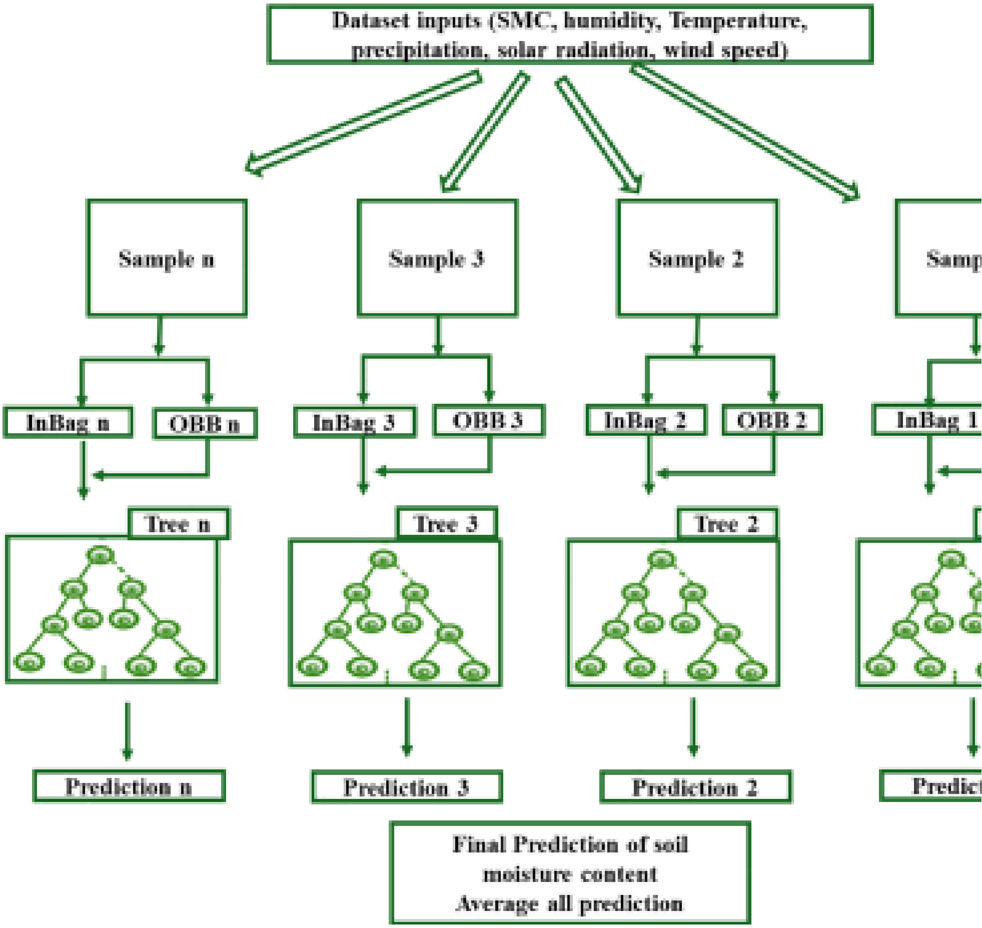


FIGURE 3
Random forest regression (RFR) model architecture.

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:

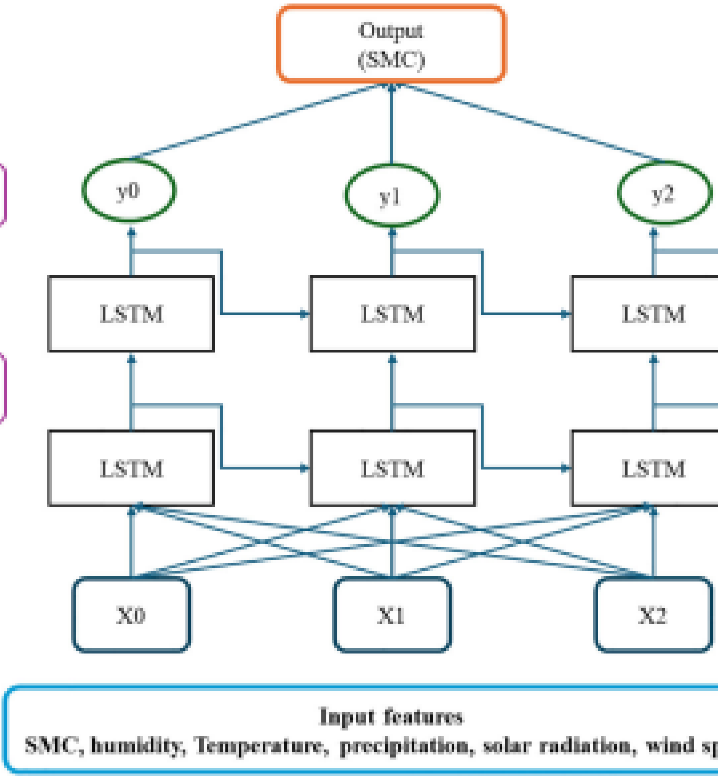


FIGURE 4
Diagram of LSTM layers.

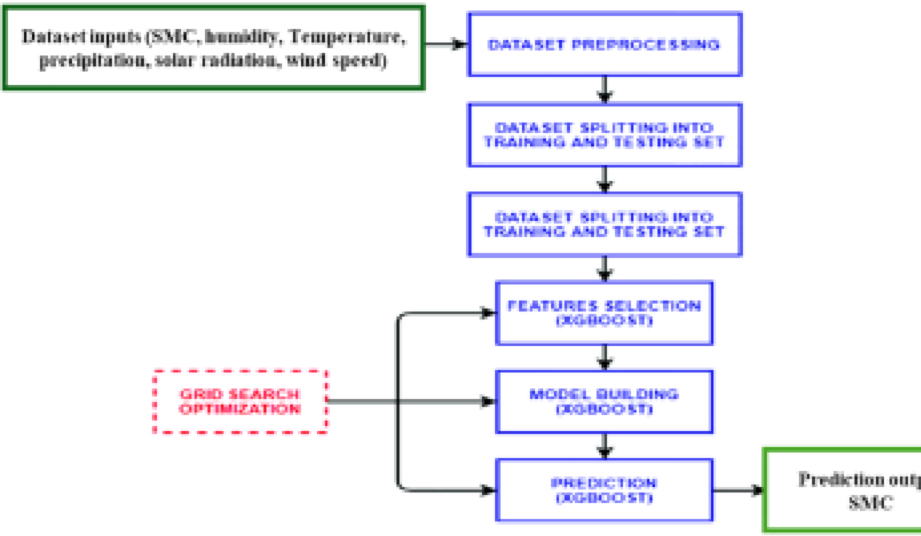


FIGURE 5
Flowchart of the XGBoost model.

Root mean square error (RMSE): It captures the typical difference

$MSE = {}^{1}=^{n} O^{n}{}_{t=1}(y_t - yt\ \hat{}\ )^2$ (2)

between predicted and observed values (43). It is calculated by taking
where y_t is the ground truth, and y^ represents the mean of the
the square root of the MSE. Unlike MSE, RMSE is measured in the
predicted values.
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).

q
ffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffi
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

RMSE =

${}^{1}=^{n} O^{n}{}_{t=1}(y_t - yt\ \hat{}\ )^2$

importance is not inherently provided; thus, a model-agnostic
(3)
permutation importance approach was applied and calculated by
Mean absolute error (MAE) (43): It calculates the average of the absolute differences between predictions and actual observations (Equation 4):

$MAE = {}^{1}=^{n} O^{n}{}_{t=1} jj\ y_t - yt\ \hat{}$ (4)

R-squared (R²): 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:

$R^2 = 1 - \underline{O}^{n}{}_{t=1}(y_t - yt\ \hat{}\ )^2$

$O^{n}{}_{t=1}(y_t - {}^{y})^{\cancel{2}}$ (5)

where y_i is the predicted value, y_t is the ground truth, and 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 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

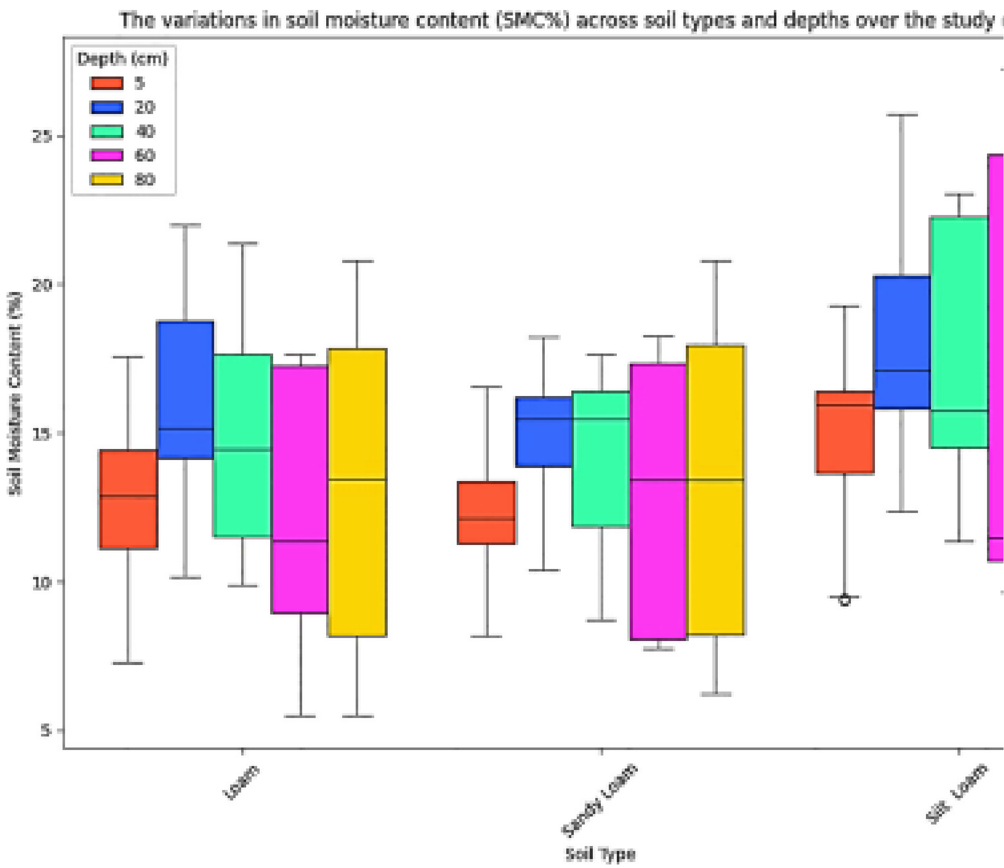


FIGURE 6
Soil moisture content variation in three soil types at five different depths.

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

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

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

Sandy loam	80	XGBoost	0.023
Silt loam	5	RFR	0.396
Silt loam	20	RFR	0.932
Silt loam	40	RFR	0.248
Silt Loam	60	RFR	0.272
Silt loam	80	RFR	1.145
Silt loam	5	LSTM	0.159
Silt loam	20	LSTM	0.073
Silt loam	40	LSTM	0.64
Silt loam	60	LSTM	0.901
Silt loam	80	LSTM	1.702
Silt loam	5	XGBoost	0.062
Silt loam	20	XGBoost	0.066
Silt loam	40	XGBoost	0.137
Silt loam	60	XGBoost	0.141
Silt loam	80	XGBoo	1.59

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

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.

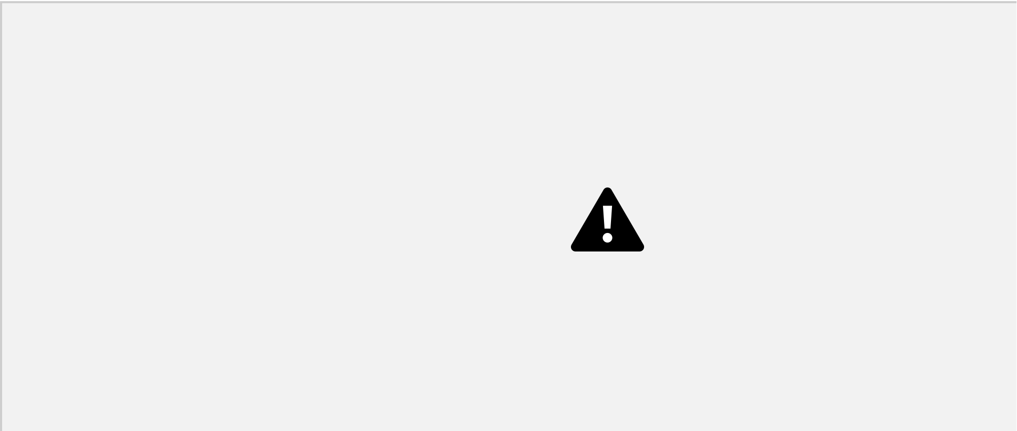


FIGURE 13
Average feature importance value results by the three ML models (RFR, XGBoost, and LSTM) per soil type across all depths.

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.

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.
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

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, DrăguţL. 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