

PRECISION AGRICULTURE IN AL JOUF BASED ON MACHINE LEARNING AND REMOTE SENSING TECHNIQUES

Shaya A. Alshaya

Received on August 7, 2024

Presented by Ch. Roumenin, Member of BAS, on November 26, 2024

Abstract

In today's agricultural landscape, precision is crucial, utilizing advanced technologies like IoT, AI, aircraft, and satellite systems. Smart agriculture aims to revolutionize production by monitoring soil quality and employing data analytics through machine learning. This study enhances soil quality assessment in Saudi Arabia's Aljof region using remote sensing technology. Leveraging UAVs, remote sensing provides crucial data and imagery for olive cultivation. Cloud platforms with satellite data offer invaluable information, aiding public and private decision-making. Proposing an AI-powered application combined with remote sensing, this paper develops predictive models for soil moisture and time-series satellite image analysis. Experimental results show a 91% accuracy with decision tree and extra trees regressor models, highlighting their effectiveness. This research transforms agricultural productivity through advanced technology and data-driven methodologies.

Key words: remote sensing, precision agriculture, artificial intelligence, soil moisture, vegetation indices

I. Introduction. For decades, aircraft and satellites have been crucial in capturing extensive data and imagery of the Earth's surface, atmosphere, and oceans, benefiting urban and environmental monitoring. The launch of the first American civil satellite, Landsat, in 1972 marked a significant milestone in ground

surface monitoring, advancing civil space-based remote sensing. Precision agriculture has particularly benefited from these advancements, using remote sensing to improve productivity, reduce labour, and optimize irrigation planning, aiming to provide farmers with effective management tools [1].

Soil moisture is a critical variable in agriculture, offering insights into water content in fields and aiding in determining vegetation indices essential for precision agriculture. Remote sensing provides a significant advantage by collecting essential data across vast temporal and spatial scales, outperforming traditional methods. Soil moisture affects plant growth and crop yield by restoring moisture and regulating plant temperature. The plant's thermo-regulation process involves evaporating nearly 99% of acquired water and using only 0.2 to 0.5% for vegetative mass formation [2]. The World Meteorological Organization recognized the importance of soil moisture by including it among the 50 Essential Climatic Variables (ECV) for systematic observation in 2010, highlighting its dependence on various environmental, hydrogeological, and topological factors [3].

Accurate soil moisture measurement or prediction helps optimize sowing schedules, monitor soil moisture index (SMI) health, and prevent financial losses by minimizing excessive use of vegetation indices [4]. Vegetation indices in precision agriculture not only enhance customer experience but also lead to cost savings. Direct soil moisture measurement using in-soil moisture sensors is the most accurate method, with four main techniques: gravimeters, watermark sensors, capacitance sensors, and tension meters. However, these methods face challenges related to sensing area coverage, soil conditions, calibration, and accessibility [4].

These challenges have driven the development of remote sensing approaches, enabling soil moisture data acquisition without direct sensor contact [5]. Remote sensing methods are further enhanced by Artificial Intelligence (AI) algorithms, which rely on the performance of these algorithms [6]. Validating machine learning models that use remotely sensed data for soil moisture estimations often depends on publicly available long-term datasets from weather stations. International collaborations, such as the international soil moisture network, aim to improve global satellite products related to soil moisture through validation and enhancement [7].

Soil moisture remote sensing data extraction methods can be categorized into three main approaches: active remote sensing, passive remote sensing, and combined active and passive remote sensing.

Active remote sensing. This method uses electromagnetic microwave radiation with wavelengths ranging from 0.5 to 100 cm [8]. The NASA SAR sensor is a prominent active sensor for soil moisture assessment, known for its high-resolution data collection [9]. It operates based on sensor and soil parameters; sensor parameters influence signal backscatter variation, while soil parameters represent surface characteristics. The attenuation of signal backscatter through vegetation canopy and volume is calculated. There is a direct correlation between soil moisture and radar backscatter intensity: lower surface soil moisture increases radar backscat-

ter [10]. Techniques for active soil moisture sensors include backscattering models [11], statistical analysis techniques [12], and neural network applications [13].

Passive remote sensing. Passive sensors model soil moisture information and can delineate global soil moisture distribution regardless of vegetation canopy presence [14]. These sensors offer important properties such as the normalized difference vegetation index (NDVI) and examine the correlation between these indices and various soil and environmental parameters. Common methodologies for passive sensors include the universal triangular relationship method, brightness models, and statistical analysis techniques [15]. These methods are crucial for interpreting and deriving essential soil moisture data from passive remote sensing.

Combined active and passive remote sensing. Advances in remote sensor instruments and spacecraft technology have enabled the integration of active and passive remote sensing approaches. This combination seeks to mitigate the weaknesses and leverage the strengths of each method, aiming for higher spatial resolution from active methods and superior temporal resolution from passive techniques, collectively enhancing soil moisture estimation accuracy [15]. NASA's 2014 launch of the Soil Moisture Active Passive mission (SMAP) exemplifies this approach, providing global soil moisture data with resolutions sometimes as fine as 3 km per pixel [16]. Common methodologies for these sensors include microwave combined algorithms, statistical analysis techniques, and neural network applications.

The necessity for an accurate satellite-based monitoring system is evident, given these related works. This paper proposes a smart precision agriculture system grounded in machine learning approaches. Utilizing satellite-captured soil quality data in lieu of traditional local soil moisture enables coverage over extensive areas, supporting high-precision agriculture practices. The paper is structured as follows: Section II describes the proposed machine learning model based on remote sensing, while Section III details the conducted experimentation and findings. Finally, the conclusion is presented in the last section.

II. The proposed machine-learning model. The proposed model utilizes a publicly available dataset captured by the Landsat satellite. The training model encompasses four defined steps, as illustrated in Fig. 1.

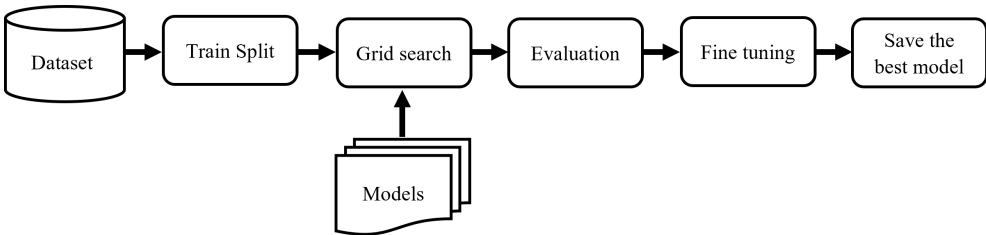


Fig. 1. Training model

Train split. This technique serves to assess the performance of a machine learning algorithm and is applicable to various supervised learning problems, including classification and regression tasks. It involves dividing the dataset into two subsets. The first subset, known as the training set, is employed to train the model. The second subset, the test set, remains independent of the model's training and is used to assess the model's predictions against actual expected values.

The size of both the training and test sets stands as a pivotal factor in configuring the train-test split. The train's size depends on a few factors, including the computational cost during both the training and evaluation phases, as well as the composition of the training and test sets. In our case, the test and split method were implemented using the scikit-learn library and particularly the `train_test_split` method with the following arguments:

- `x`: Dataset containing the spectral Indices after the feature selection procedure
- `y`: Corresponding soil moisture values
- `Test_size`: 0.2, meaning the split is 80% for the training and 20% for the test
- `random_state`: 0 to get the same training and testing set during different executions

After data cleaning, 69 samples remained from the initial 80. The dataset was split, with 55 samples (80%) used for training and 14 samples for testing. This configuration allows us to begin training our model for soil moisture prediction.

Grid search optimization. Each machine learning model relies on a set of parameters. To accurately map input data (independent variables) to targets (dependent variable), optimal hyperparameters must be selected. This complex task was automated using grid search via the `sklearn` Python module. Grid search explores and defines the best hyperparameters by exhaustively evaluating every position in the parameter grid, representing an n -dimensional search space.

The first step involves defining the search space, where each hyperparameter value represents a point. The optimization process aims to find a vector that maximizes model performance, such as achieving maximum accuracy or minimizing error. Among optimization algorithms, grid search and random search are the most recognized, with grid search systematically evaluating each point in the parameter grid.

Our objective is to estimate accurately the soil moisture based on spectral indices. The crucial step is selecting the best model using our data, which involves testing various regression models and evaluating their performance based on the best hyperparameters.

III. Results and experimentation. This section details the implementation and results of training a soil moisture dataset retrieved via the Google Earth Engine API using Python. The grid search optimization technique, applied

through GridSearchCV in sklearn, systematically evaluated various hyperparameters for each model. Grid search uses cross-validation, typically k-fold, to split data into subsets for training and validation, identifying hyperparameters that produce the best average performance.

The linear regression model achieved a low accuracy score of 58%, due to multicollinearity issues. To address this, regularization techniques were applied, specifically Ridge and ElasticNet models. Ridge uses L2 penalty, while ElasticNet combines L1 and L2 penalties. Both models showed similar accuracy improvements to around 70%. For ElasticNet, the best alpha hyperparameter value was 0.001, indicating a reliance on the L2 penalty, as the L1 penalty did not significantly enhance the prediction task.

The support vector machine (SVM) slightly outperformed Ridge and ElasticNet models, achieving a 72% score with optimal hyperparameters for the “C” coefficient and epsilon, using a linear kernel. The poly and RBF kernels showed similar performance. The Random Forest Regressor further improved the score to 79%, with the number of estimators (10, 50, or 100) being a critical factor, though higher estimators increase computational complexity. Decision Trees and Extra Trees regression models achieved approximately 91% accuracy using the squared_error criterion. These models differ in their splitting techniques: Decision Trees use bootstrap, while Extra Trees use a random search.

For data analysis, soil moisture scatter plots and box plots were employed as shown in Fig. 2. Scatter plots visualize relationships between variables, aiding regression analysis and machine learning projects. Box plots provide a standardized

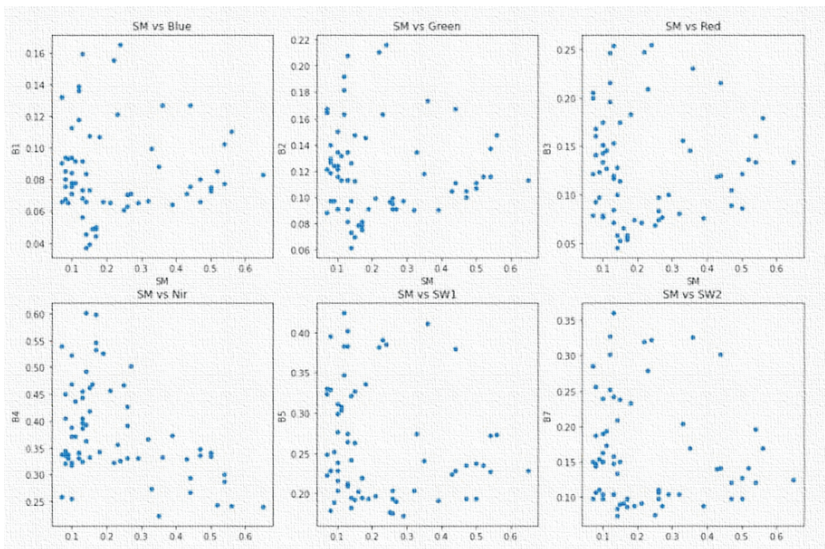


Fig. 2. Scatter plots of different variables in the dataset and the soil moisture variable

depiction of data distributions and identify outliers that can impact predictive performance. Data interpolation addresses sparsity issues, enhancing model performance. Methods like linear, polynomial, and spline interpolation were considered, with polynomial interpolation chosen for this dataset. Figure 3(b) demonstrates the effectiveness of polynomial interpolation with NDVI values, showing original data points and the interpolated continuous line.

$$(1) \quad NDVI = \frac{NIR - Red}{NIR + Red}.$$

This section explores and analyzes a time series dataset for NDVI, focusing on decomposition to identify seasonality and trends. Decomposition breaks down the data into components, revealing gradual increases or decreases (trend) and recurrent short-term cycles (seasonality). Figure 3(b) illustrates these time series components. The Holt-Winters method is applied for time series forecasting, incorporating both trend and seasonality. This technique extends exponential smoothing and uses three equations for forecasting: one for level (l_t), one for trend (b_t), and one for seasonality (s_t). This method effectively captures the dataset's inherent patterns for accurate predictions.

$$(2) \quad l_t = \alpha \cdot (y_t - s_{t-m}) + (1 - \alpha) \cdot (l_{t-1} + b_{t-1})$$

$$(3) \quad b_t = \beta \cdot (l_t - l_{t-1}) + (1 - \beta) \cdot b_{t-1}$$

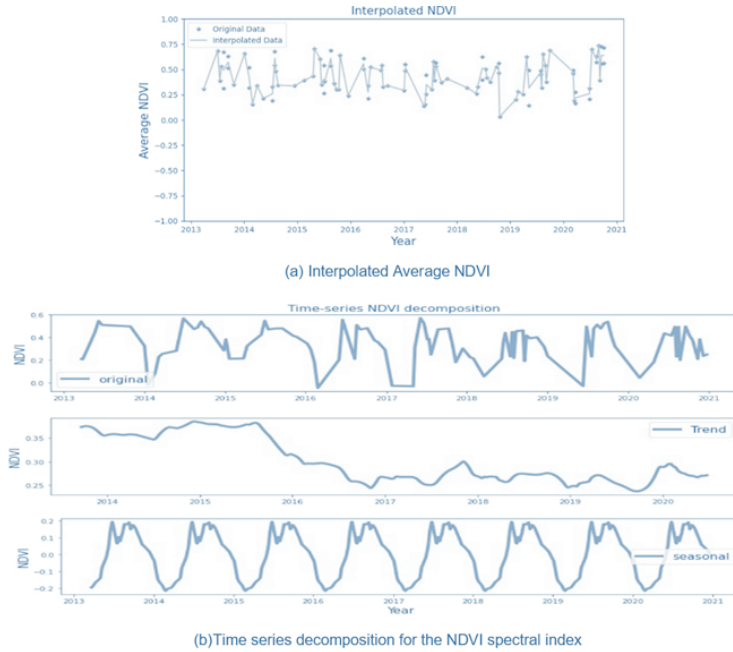


Fig. 3. NDVI results

$$(4) \quad s_t = \gamma \cdot (y_t - l_{t-1} - b_{t-1}) + (1 - \gamma) \cdot s_{t-m},$$

where:

- y_t is the observed value at time t ;
- l_t represents the level or smoothed value at time t ;
- b_t represents the trend component at time t ;
- s_t represents the seasonal component at time t ;
- m is the number of seasons in a cycle;
- α , β , and γ are smoothing parameters (often between 0 and 1) controlling the weights of the observed value, trend, and seasonal components, respectively.

From 2013 to 2016, the average NDVI value remained around 0.35 before declining, influenced by crop state and weather conditions, with notable seasonal fluctuations.

IV. Conclusions. This paper emphasizes the importance of soil moisture and vegetation indices in agricultural analysis, addressing data sparsity through machine learning models. The Extra Tree Regressor effectively forecasts vegetation indices from satellite imagery, focusing on the NDVI index, crucial for precision agriculture. Successful NDVI forecasting based on historical Google Earth Engine API data is demonstrated. The Decision Tree and Extra Tree Regressor models achieve approximately 91% accuracy in soil moisture estimation. Future improvements include augmenting the dataset with more soil moisture samples to enhance model performance.

REFERENCES

- [1] GLADIMA NISIA T., S. RAJESH (2022) Enhanced Feature Selection Method Using Wrapper-based Random Search Strategy and Mutual Information for Remote Sensing Image Classification, *C. R. Acad. Bulg. Sci.*, **75**(7), 1037–1044.
- [2] YANG L., Y. HAN, P. YANG et al. (2015) Effects of superabsorbent polymers on infiltration and evaporation of soil moisture under point source drip irrigation, *Irrig. Drain.*, **64**(2), 275–282.
- [3] DORIGO W., P. VAN OELEVEN, W. WAGNER et al. (2011) A new international network for in situ soil moisture data, *Eos Trans. Am. Geophys. Union*, **92**(17), 141–142.
- [4] NAVIDI M. N., J. SEYEDMOHAMMADI, S. A. SEYED JALALI (2022) Predicting soil water content using support vector machines improved by meta-heuristic algorithms and remotely sensed data, *Geomech. Geoengin.*, **17**(3), 712–726.
- [5] CHENG M., B. LI, X. JIAO et al. (2022) Using multimodal remote sensing data to estimate regional-scale soil moisture content: A case study of Beijing, China, *Agric. Water Manag.*, **260**, 107298.
- [6] GUO J., Q. BAI, W. GUO et al. (2022) Soil moisture content estimation in winter wheat planting area for multi-source sensing data using CNNR, *Comput. Electron. Agric.*, **193**(11), 106670.

- [7] YINGLAN A., G. WANG, P. HU et al. (2022) Root-zone soil moisture estimation based on remote sensing data and deep learning, *Environ. Res.*, **212**, 113278.
- [8] ARAB S. T., M. ISLAM, M. SHAMSUZZOLA et al. (2022) A Review of Remote Sensing Applications in Agriculture and Forestry to Establish Big Data Analytics, In: *Remote Sens. Appl.*, Chapter 1, 1–24.
- [9] ADELI S., B. SALEHI, M. MAHDIANPARI et al. (2021) Moving Toward L-Band NASA-ISRO SAR Mission (NISAR) Dense Time Series: Multipolarization Object-Based Classification of Wetlands Using Two Machine Learning Algorithms, *Earth Sp. Sci.*, **8**(11), e2021EA001742.
- [10] JOSEPHSON C., M. KOTARU, K. WINSTEIN et al. (2021) Low-cost in-ground soil moisture sensing with radar backscatter tags. In: *ACM SIGCAS Conference on Computing and Sustainable Societies*, 299–311.
- [11] XING M., L. CHEN, J. WANG et al. (2022) Soil Moisture Retrieval Using SAR Backscattering Ratio Method during the Crop Growing Season, *Remote Sens.*, **14**(13), 3210.
- [12] DASH S. K., R. SINHA (2022) Spatiotemporal dynamics and interrelationship between soil moisture and groundwater over the Critical Zone Observatory in the Central Ganga plain, North India, *Hydrol. Earth Syst. Sci. Discuss.*, 1–32.
- [13] YUAN Q., H. XU, T. LI et al. (2020) Estimating surface soil moisture from satellite observations using a generalized regression neural network trained on sparse ground-based measurements in the continental US, *J. Hydrol.*, **580**, 124351.
- [14] SURE A., O. DIKSHIT (2019) Estimation of root zone soil moisture using passive microwave remote sensing: A case study for rice and wheat crops for three states in the Indo-Gangetic basin, *J. Environ. Manage.*, **234**, 75–89.
- [15] EDOKOSSY K., A. CALABIA, S. JIN, I. MOLINA (2020) GNSS-reflectometry and remote sensing of soil moisture: A review of measurement techniques, methods, and applications, *Remote Sens.*, **12**(4), 614.
- [16] YE N., J. P. WALKER, X. WU et al. (2020) The soil moisture active passive experiments: Validation of the SMAP products in Australia, *IEEE Trans. Geosci. Remote Sens.*, **59**(4), 2922–2939.

Computer Science Department, College of Science, Al-Zulfi – 11932, Majmaah University, Saudi Arabia
e-mail: shaya@mu.edu.sa