Science of the Total Environment 833 (2022) 155066

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Research Paper

A low-cost approach for soil moisture prediction using multi-sensor data and machine learning algorithm

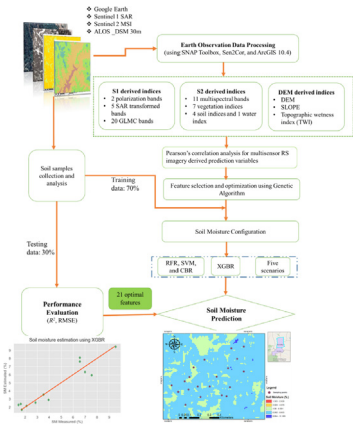
Thu Thuy Nguyen [a], Huu Hao Ngo [a,*], Wenshan Guo [a], Soon Woong Chang [b], Dinh Duc Nguyen [b], Chi Trung Nguyen [c], Jian Zhang [d], Shuang Liang [d], Xuan Thanh Bui [e], Ngoc Bich Hoang [f]

[a] Centre for Technology in Water and Wastewater, School of Civil and Environmental Engineering, University of Technology Sydney, Sydney, NSW 2007, Australia
[b] Department of Environmental Energy Engineering, Kyonggi University, 442-760, Republic of Korea
[c] Faculty of Science, Agriculture, Business and Law, UNE Business School, University of New England, Elm Avenue, Armidale, NSW 2351, Australia
[d] School of Environmental Science and Engineering, Shandong University, Qingdao 266237, China
[e] Key Laboratory of Advanced Waste Treatment Technology & Faculty of Environment and Natural Resources, Ho Chi Minh City University of Technology (HCMUT), Vietnam National University Ho Chi Minh (VNU-HCM), Ho Chi Minh City 700000, Viet Nam
[f] NTT Institute of Hi-Technology, Nguyen Tat Thanh University, Ho Chi Minh City, Viet Nam

HIGHLIGHTS

- Data fusion and machine learning are of importance for SM retrieval.
- The predictor variables derived from S-1, S-2, and ALOS DSM data were generated.
- 21 optimal features were identified using genetic algorithm.
- The extreme gradient boosting regression performed best in SM estimation.
- Inverted Red-Edge Chlorophyll Index was the most influential feature.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Damià Barceló

Keywords:
Soil moisture
Machine learning
Data fusion
Sentinel
ALOS

ABSTRACT

A high-resolution soil moisture prediction method has recently gained its importance in various fields such as forestry, agricultural and land management. However, accurate, robust and non- cost prohibitive spatially monitoring of soil moisture is challenging. In this research, a new approach involving the use of advance machine learning (ML) models, and multi-sensor data fusion including Sentinel-1(S1) C-band dual polarimetric synthetic aperture radar (SAR), Sentinel-2 (S2) multispectral data, and ALOS Global Digital Surface Model (ALOS DSM) to predict precisely soil moisture at 10 m spatial resolution across research areas in Australia. The total of 52 predictor variables generated from S1, S2 and ALOS DSM data fusion, including vegetation indices, soil indices, water index, SAR transformation indices, ALOS DSM derived indices like digital model elevation (DEM), slope, and topographic wetness index (TWI). The field soil data from Western Australia was employed. The performance capability of extreme gradient boosting regression (XGBR) together with the genetic algorithm (GA) optimizer for features selection and optimization for soil moisture prediction in bare lands was examined and compared with various scenarios and ML models. The proposed model (the XGBR-GA model) with 21 optimal features obtained from GA was yielded the highest performance (R^2 = 0. 891; RMSE = 0.875%) compared to random forest regression (RFR), support vector machine (SVM), and CatBoost gradient

* Corresponding author.
E-mail address: ngohuuhao121@gmail.com (H.H. Ngo).

http://dx.doi.org/10.1016/j.scitotenv.2022.155066
Received 13 December 2021; Received in revised form 30 March 2022; Accepted 2 April 2022
Available online 7 April 2022

boosting regression (CBR). Conclusively, the new approach using the XGBR-GA with features from combination of reliable free-of-charge remotely sensed data from Sentinel and ALOS imagery can effectively estimate the spatial variability of soil moisture. The described framework can further support precision agriculture and drought resilience programs via water use efficiency and smart irrigation management for crop production.

1. Introduction

Soil moisture (SM) has played vital roles in hydrological state and ecological processes which affects energy, water, and carbon cycles such as evaporation, transpiration, diversity and rainfall-runoff of various ecosystems (Ågren et al., 2021; efBabaeian et al., 2021; Robinson et al., 2008). Soil moisture is also a crucial predictor indicator for identify crop water stress, which helps agricultural drought monitoring. Thorough knowledge about the spatiotemporal patterns of SM is of essential importance for understanding water budgets in hydrological systems which helps prevent agricultural drought problems, water vulnerability, the issues of water shortage, and improve properly crop production across the world (Chaudhary et al., 2021; Tuller et al., 2019). Traditional ground techniques of soil moisture based on field experiments, in-situ soil sensing instrumentation, and geophysical and mobile sensing (Cheng et al., 2022; Robinson et al., 2008). The disadvantages of this method are high cost with small-scale monitoring. Remotely sensed measurements including active remote sensing and passive remote sensing recently have employed effectively for SM monitoring globally (Chaudhary et al., 2021; Cheng et al., 2022; Dubois et al., 2021; Prasad et al., 2018; Warner et al., 2021). At present, various satellite systems via microwave remote sensing like Soil Moisture Active Passive (SMAP) (Entekhabi et al., 2010), Advanced Scatter meter (ASCAT) (Wagner et al., 2013), and Soil Moisture and Ocean Salinity (SMOS) (Kerr et al., 2001) have been explored for global SM monitoring with spatial resolutions of 10 km, 50 km, and 35 km, respectively. With the low spatial resolution, SM data obtained from these aforementioned missions have not been used widely in farm scales for agricultural management.

Recent advances in earth observation technology such as using active and passive remote sensing (RS) imagery have been dedicated to solving the problems of SM dynamics retrieval on farming lands. Active remote sensing like Unmanned Aerial System (UAS) with highly flexible flight schedules and high spatial resolutions of images offer a great opportunity to estimate the SM for farm-scales (efBabaeian et al., 2021). The application of high-resolution about 2 m images from airborne LIDAR can accurately estimate the SM dynamics to support precision agriculture production (Ågren et al., 2021). However, the deployment of UAS and LIDAR have struggled with some obstacles such as limited fight time, high operation cost, and challenges with hyperspectral images processing which limits the application of active RS for SM monitoring (Gago et al., 2015).

Multispectral remote sensing sensors such as Sentinel 1 and Sentinel 2 datasets from European earth observation program Copernicus have employed recently to capture effectively the SM content in several agricultural areas across the world with the spatial resolution of 10-100 m (El Hajj et al., 2017; Georganos et al., 2018). The free-of-charge imagery from Sentinel date at high spatial and temporal resolutions are a proper solution to address the challenges of hyperspectral images in agricultural SM prediction. The C-band of the Sentinel-1A and-1B Synthetic Aperture Radar (SAR), and vegetation and soil indices from Sentinel -2A and -2B have been generated to estimate SM properties at high spatial resolution in the pilot scale (Aksoy et al., 2021; El Hajj et al., 2017; Karthikeyan and Mishra, 2021; Ma et al., 2021; Prasad et al., 2018; Schönauer et al., 2021; Senanayake et al., 2021). In addition, terrain indices from digital elevation (DEM) models such as slope, topographic wetness index (TWI), and death-to-water (DTW) index have also been used to predict the agricultural SM (Ågren et al., 2021; Murphy et al., 2008). Topo-hydrological indicators generated from high-resolution DEM data illustrated high correlations with soil properties and soil moistures by capturing the hydrological processes' characteristics of specific sites (Zhao et al., 2021; Zhou et al., 2020a, 2020b). According to Florinsky et al. (2002), soil properties including soil moisture have a significant relationship with topographic attributes, especially in agricultural landscapes.

Machine learning techniques are already commonly applied to handle diverse and large volumes of remote-sensing datasets, with very high performances (Carranza et al., 2021; Gómez et al., 2020; Gómez et al., 2021; Hosoda et al., 2020; Karthikeyan and Mishra, 2021; Ma et al., 2021; Prasad et al., 2018; Schmidt et al., 2020). Artificial intelligence techniques such as random forest regression (RFR), support vector machine (SVM), extreme gradient boosting regression (XGBR), CatBoost gradient boosting regression (CBR) have been employed widely to estimate soil moisture products with high prediction accuracy (Ågren et al., 2021; Carranza et al., 2021; Senanayake et al., 2021). The RFR algorithm performed well to predict the field-scale of soil moisture in China using unmanned aerial vehicle (UAV) imagery with coefficient determination (R^2) of 0.91 (Ge et al., 2019). The XGBR technique was used to estimate the SM dynamics in Swedish forest landscape using multiple LIDAR derived digital terrain indices with high performance values compared to RFR and SVM (Ågren et al., 2021). In general, ML algorithms provide a substantial potential for the SM estimation accurately. In this study, a new approach for soil moisture monitoring using the combination of three free-of-charge and high-resolution remote sensing datasets including Sentinel 1, Sentinel 2, and ALOS DSM was presented to estimate the soil moisture in field-scale. Four well-known ML algorithms including RFR, SVM, XGBR, and CBR were employed to test the performance of predictor variables from these datasets. The optimisation of hyper-parameters tuning and the selection of predictor variables during the construction phase of the ML techniques was applied to improve the performance of ML models. This study aims to: (1) assess the correlation of prediction indicators derived from multi-spectral images, SAR datasets, and ALOS DSM in the SM retrieval; (2) select and optimize features from these indicators using genetic algorithm (GA) and XGBR; (3) evaluate the prediction performance of the selected ML model (XGBR) with various scenarios of data-fusion level in the SM prediction while exploring the effectiveness of GA feature optimization on the ML model in mapping the SM content at 10 m spatial resolution; and (4) compare the estimation accuracy of XGBR model with other three well-known ML models using optimal features. The novel framework will be expanded to other field-scales or regional scales to build the SM map, which provides valuable data for different stakeholders like water managers, local authorities, and landholders to practice precision agriculture.

2. Materials and methods

2.1. Study area and soil sample collection

The study sites are located in the Wests, Goomalling shire (latitude coordinate: −31° 18′ S and longitude coordinate: 116° 49′ E), and Cookies area - Northam shire (latitude: −31° 39′ S, and longitude: 116° 39′ E) in the agricultural region of Western Australia (WA). The WA has a diverse type of agricultural production including vegetable industries which contributes a majority of total value of agricultural production in the region. Pastoral and cropping are two key agricultural practices in the WA (Kingwell et al., 2020). According to Australian Bureau of Agricultural and Resource Economics, high-rainfall, wheat-sheep, and pastoral zones are the main agricultural climatic zones in Australian (Salim and Islam, 2010). The type of climate in the WA is a Mediterranean climate where is hot and dry in summer, and cool and wet in winter seasons. The rainfall

season is from April and October which ranges between 300 and 600 mm (Kingwell et al., 2020).

Soil sampling points were selected from a binary land-use classification map which was produced by extreme gradient boosting classification from high spatial resolution Google Earth imagery and Sentinel 2 imagery. Nguyen et al. (2022) illustrated the detail of soil samples selection using the binary map. The active learning technique in remote sensing classification was used to assist in the selection of soil samples, which helps reduce the influence of vegetation on SM contents (Fu et al., 2010). Forty bare-soil sampling areas with a pixel (size of 10 m × 10 m) across the study locations (20 points for each plot) were identified from the binary map (Fig. 1). A Differential Global Positioning System (DGPS) was applied for accurately the samples' location identification with an accuracy of 1–3 cm (Michalski and Czajewski, 2004). The soil samples were taken in April 2021. There are four soil cores being taken in each sampling area to a depth of 7 cm from each experimental block by a standard tube of 7.3 cm in diameter. The soil samples were sent to the laboratory for soil moisture analysis by oven drying method from Standards Association of Australia.

2.2. Research framework

The research framework consists of five main steps (Fig. 2): (1) collecting surface soil dataset (0–10 cm) from the binary land-use classification map; (2) generating predictor indicators from optical (Sentinel 2), synthetic aperture radar (Sentinel 1), and terrain indices derived from ALOS DSM; (3) computing Pearson's correlation analysis and feature selection using genetic algorithm; (4) evaluating the performance of the XGBR model with five different scenarios developed from features derived from S1, S2, and ALOS DSM with 70% of SM measured dataset used for models' training and 30% for models' validation; and (5) comparing the performance of the XGBR model with other ML techniques using optimal features and building the SM dynamics map for the study areas.

2.3. Remote sensing data acquisition and image processing

2.3.1. Data acquisition

In this research, soil moisture predictor variables were computed from S-2 multispectral satellite data, S-1 C-band dual polarimetric SAR imagery, and ALOS DSM (Table 1). While Sentinel 1 and Sentinel 2 images were downloaded from the Copernicus Open Access Hub from European Space Agency (ESA), the ALOS DSM 30 m imagery was acquired from JAXA Earth Observation Research Centre. The SNAP Sentinel Application Platform toolbox were used for both Sentinel datasets processing, whereas ArcGIS 10.4 was employed to process ALOS imagery and compute the ALOS-DSM derived features. All images were resampled to a ground sampling distance (GSD) of 10 m and geocoded in the same projection of World Geodetic System (WGS84) - Universal Transverse Mercator (UTM) zone 50 South (50S).

2.3.2. Sentinel images processing

The S-2 image was processed via four main steps which presented in the Fig. 3. Ten multispectral bands were extracted for the study including B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12. Vegetation indices, soil indices, and water index were computed by thematic land processing function in the SNAP toolbox (Pasqualotto et al., 2019). Vegetation, soil and water indicators are presented as being sensitive to soil moisture content which recently was used for soil moisture properties estimation (Jin et al., 2017). Predictor variables derived from S-2 were illustrated in Table 2 below. A total of 22 indicators were computed from S-2 for the SM prediction.

A radar module in the SNAP toolbox was used to process the S-1 imagery. There are seven main stages involving the extraction of the SAR dataset for the SM monitoring. The first one is to convert the S-1 raw data to scale backscatter coefficient (σ^0) in decibel (dB) and correct the orbit file. The next steps are: (1) the removal of thermal and border noise; (2) radiometric

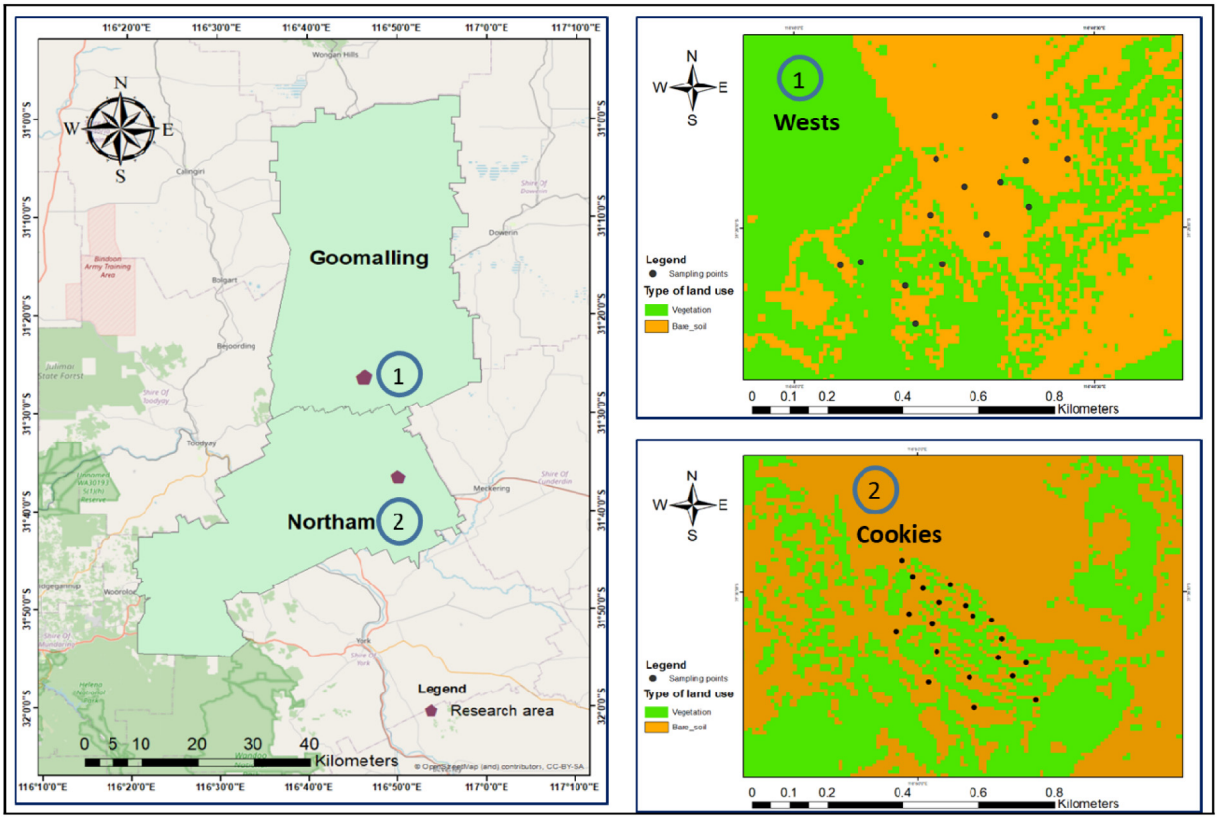


Fig. 1. Location of the study sites and sampling points in Wests and Cookies area.

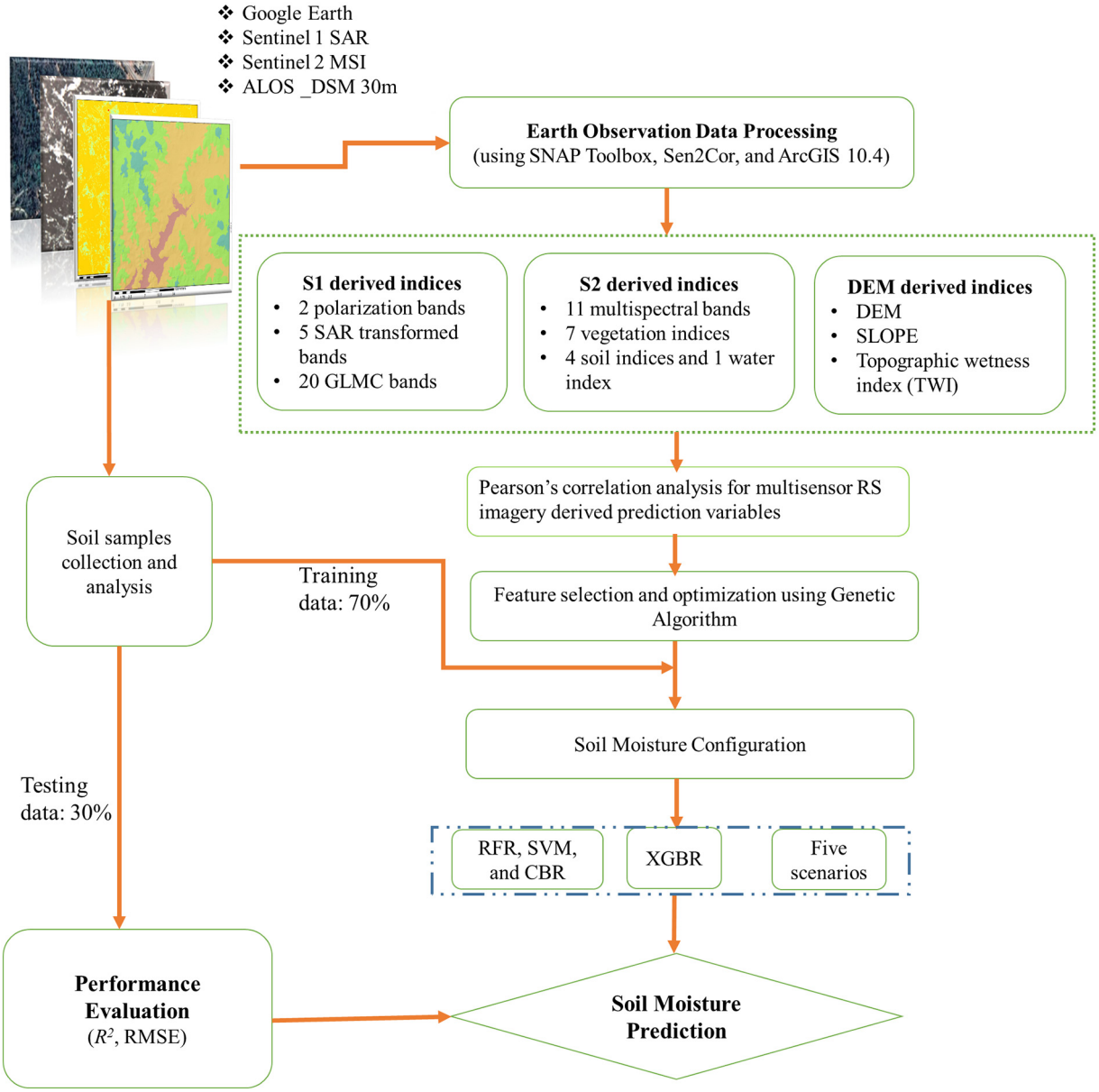


Fig. 2. A novel generated framework of SM estimation using multi-sensor data fusion and ML approach.

calibration; (3) speckle filtering; (4) the correction of range Doppler terrain; (5) normalized radar backscattering coefficient; and (6) the computation of SAR prediction variables including two original bands from dual polarization (VH and VV); the five transformed bands (VV/VH; VH/VV; VV-VH; VH-VV; (VV + VH)/2); and the 20 new indicators generated from VV and VH using the GLMC algorithm.

2.3.3. ALOS image processing

The Advanced Land Observing Satellite (ALOS) was launched by the Japan Aerospace Exploration Agency (JAXA) in 2006. JAXA recently provided the product of ALOS-DSM which is one of the newest remote sensing-based DEM. The ALOS-DSM has two kinds of resolution. ALOS-DSM with the resolution of 30 m is a free-of-charge dataset and higher prediction performance compared to Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM) ASTER GDEM and Shuttle Radar Topography Mission Digital Elevation Model (SRTM-DEM) (Nikolakopoulos, 2020). DEM and SLOPE derived indicators were generated by raster processing and calculation in ArcGIS 10.4. Fig. 4 shows the elevation of the study sites which ranges from 139 m to 480 m and slope is between 0 and 87 degree.

Table 1
Remote sensing data acquisition for the study areas.

Sensor	Scene / tile ID	Acquisition date (month/day/year)	Processing level	Spatial resolution (m)	Spectral band/polarization
S-2	50JML	04/17/2021	1C	10–20	13 multispectral bands
S-1	S1B_IW_GRDH1SDV	04/27/2021	GRD	10	Dual-polarization (VV and VH)
ALOS-DSM	AW3D30	04/01/2021		30 m	

Source: European Space Agency ESA and JAXA Earth Observation Research Centre.

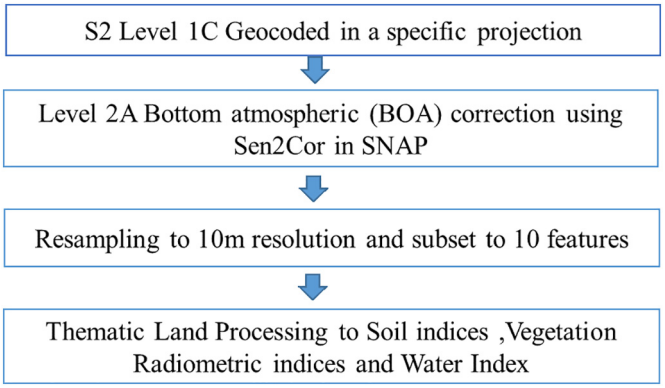


Fig. 3. The steps of Sentinel images processing using SNAP Toolbox.

Topographic Wetness Index (TWI) generated from digital elevation model (DEM) has been used for soil moisture estimation because TWI is helpful to identify the place where water is accumulated in the specific area with the differences of elevation. TWI highlights the terrain-driven balance of the catchment water supply and the water drainage of specific local areas. However, there are various algorithms such as a flow accumulation, a flow width, or a slope algorithm can be employed to compute TWI. It should be selecting the best one that the TWI obtains the high correlation with soil moisture content. The best TWI for soil moisture prediction is Freeman flow algorithm, local slope, and the equal cell size of flow width which was generated by the following equation (Kopecký et al., 2021) (Fig. 5).

$$\mathrm{TWI} = \ln \frac{{}^{Total\ catchment}/_{Flow\ Width}}{\tan(slope)} \quad (1)$$

2.3.4. Prediction scenarios

The prediction accuracy of ML techniques was tested with different scenarios which were developed based on the level of S-1, S-2, and ALOS DSM data fusion and the results from feature selection and optimization using GA. The five scenarios were presented in Table 3. While SC1 comprises of 22 indicators derived from S-2 and 3 indicators from ALOS-DSM, SC2 includes 27 S1 features, and three features derived from ALOS-DSM. SC3 consists of 22 S-2 predictor variables and 27 S1 variables. SC4 contains the total 52 features generated from both S-1, S-2, and ALOS DSM. The potential of 21 optimal features from GA selection for SM prediction was evaluated in SC5.The scenarios were presented in Table 3 below. The aim of scenarios development was to evaluate the impact of the level of different features combinations and the application of feature selection algorithm on how well the SM dynamic prediction went.

2.4. Machine learning algorithm

2.4.1. Extreme gradient boosting regression (XGBR)

Extreme gradient boosting technique is one of gradient tree boosting algorithms which developed by Chen and Guestrin (2016). It has a high performance for supervised learning to handle both regression and classification problems (Ha et al., 2021). The extreme gradient boosting regression (XGBR) algorithm is presented as a scalable end-to-end tree boosting which has widely used to address data mining issues (Chen and Guestrin, 2016). The XGBR is applied commonly because of its high execution speed with parallelization, out-of-core computation, and cache optimization. Data scientists prefer using the ML model due to its scalability in various scenarios, and its high performance for limited data studies. The XGBR model has a large of number of hyperparameters such as learning rate, max_depth, n_estimators, and gamma, which affects its performance. In this study, Optimal XGBR hyperparameters was explored by a Grid Search method with five-fold CV by testing the integration of hyperparameters.

2.4.2. Random forest regression (RFR)

The RFR algorithm is a well-known machine learning algorithm and easily applied effectively for various applications due to its simplicity to tune, train, and validate (Breiman, 2001). This ML model consists of a wide range of regression trees. Each regression tree is developed by bootstrapped training samples from the input dataset which can helps reduce the risk of overfitting issues of ML algorithms. In generally, the samples will be separated with about two-thirds of the dataset (in-bag data) for the training samples and the others for the validation samples (Out-Of-Bag (OBB data) (Pham et al., 2020). Three crucial parameters of RFR includes the number of regression trees generated from the bootstrap sample of the observation, the number of prediction features at each node, and the minimal size of the terminal nodes. The selection of these parameters can affect the performance of RFR.

2.4.3. Support vector machine (SVM)

The SVM algorithm was developed by Cortes and Vapnik (1995). It is one of the most popular supervised learning techniques using the kernel approach and statistical theory, which can handle for classification, regression, and outlier's detection problems (Cortes and Vapnik, 1995; Cristianini and Ricci, 2008). While the SVM can be applied effectively to death with non-linear problems, this technique does not obtain high

Table 2
Vegetation, soil, and water predictor variables derived from Sentinel 2.
(Modified from (Pham et al., 2020)).

Vegetation and soil index	Acronyms	S-2 band wavelengths	References
Ratio Vegetation Index	RVI	$\frac{NIR}{Red}$	(Tucker, 1979)
Normalized Difference Vegetation Index	NDVI	$\frac{NIR-Red}{NIR+Red}$	(Rouse et al., 1973)
Green Normalized Difference Vegetation Index	GNDVI	$\frac{NIR-Green}{NIR+Green}$	(Gitelson et al., 1996)
Normalized Difference Index using Bands 4 & 5 of S-2	NDI45	$\frac{RE1-Red}{RE1+Red}$	(Delegido et al., 2011)
Soil Adjusted Vegetation Index	SAVI	$(1+L)\left(\frac{NIR-Red}{NIR+Red+L}\right)$ L = 0.5 in most conditions	(Huete, 1988)
Inverted Red-Edge Chlorophyll Index	IRECl	$\frac{RE3-Red}{RE1/RE2}$	(Frampton et al., 2013)
Modified Chlorophyll Absorption in Reflectance Index	MCARI	[(RE1 − Red) − 0.2 × (RE1 − Green)] × (RE1 − NIR)	(Daughtry et al., 2000)
Brightness Index	BI	$\frac{\sqrt{(Red\times Red)+(Green\times Green)}}{2}$	(Escadafal, 1989)
Brightness Index 2	BI2	$\frac{\sqrt{(Red\times Red)+(Green\times Green)+(NIR\ x\ NIR)}}{2}$	(Escadafal, 1989)
Redness Index	RI	$\frac{Red\times Red}{Green\times Green\times Green}$	(Mathieu et al., 1998)
Color Index	CI	$\frac{Red-Green}{Red+Green}$	(Mathieu et al., 1998)
Normalized Difference Water Index	NDWI	(NIR − SWIR)/(NIR + SWIR)	(Gao, 1996)

Note: Band wavelengths of S-2: B2: Blue (492 nm), B3: Green (560 nm), B4: Red (665 nm), B5: Red-edge 1 (RE1) (704 nm), B6: Red-edge 2 (RE2) (740 nm), B7: Red-edge 3 (RE3) (783 nm), B8: near-infrared (NIR) (833 nm), B8A: Narrow-NIR (865 nm), B11: short-wavelength infrared (SWIR1) (1614 nm), and B12: SWIR2 (2202 nm).

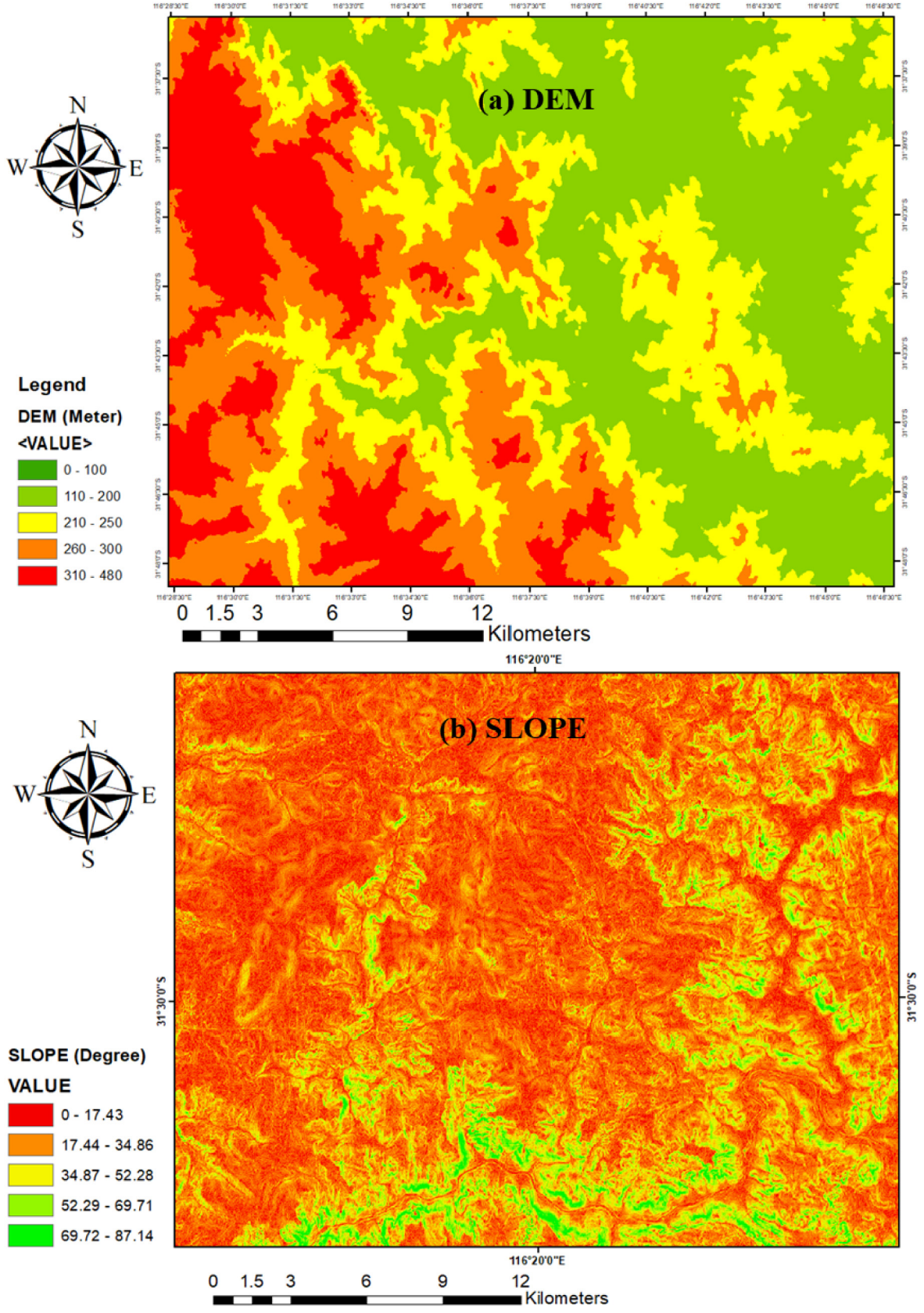


Fig. 4. Indices generated from ALOS DSM: (a) DEM and (b) SLOPE.

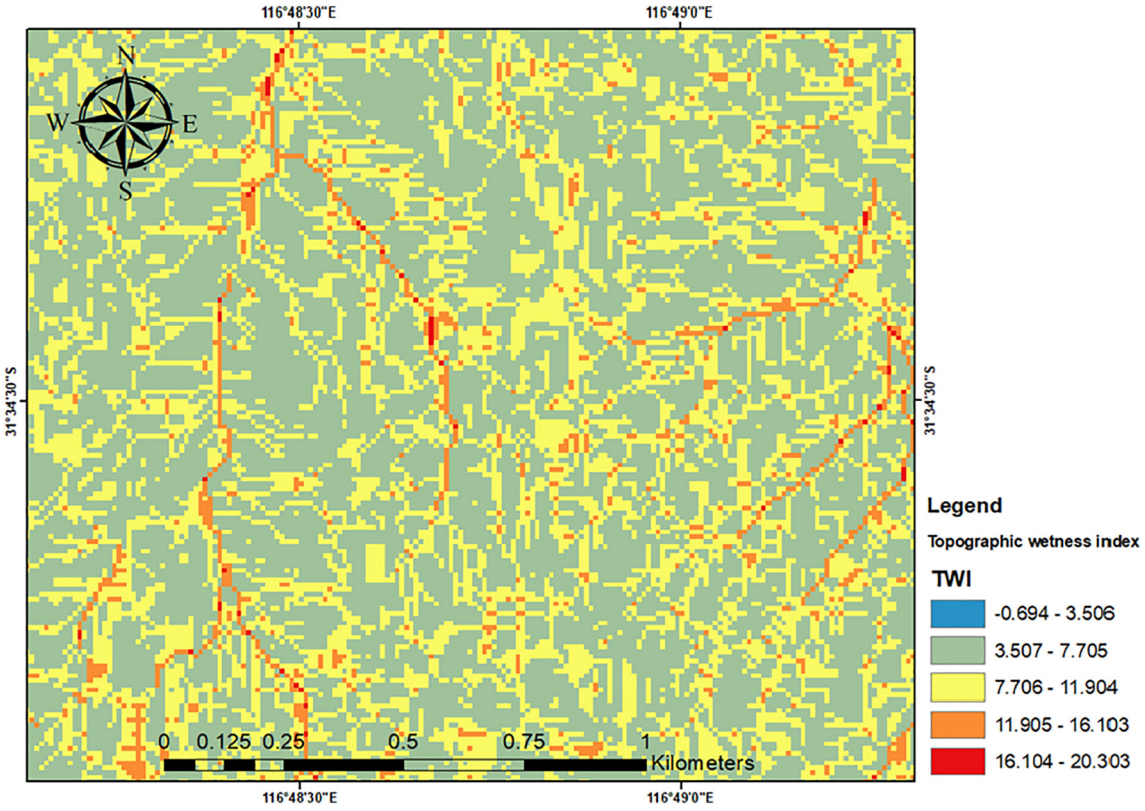


Fig. 5. TWI mapping in the study site.

performances with a noisy and overlapped dataset. Regularization parameter, the kernel function, and gamma controlling the overfitting are three main hyper-parameters of the SVM model which influences the prediction accuracy of this technique. Four types of kernel function for the SVM are polynomial, sigmoid, linear and radial basis function. The grid search with a five-fold CV was employed to identify the optimal hyper-parameters of each ML algorithm in Jupyter Notebook environment.

2.4.4. CatBoost gradient boosting regression (CBR)

CBR is known as a family member of gradient boosted decision trees (GBDT's). It is an interdisciplinary approach for classification and regression tasks in time-series and big data (Hancock and Khoshgoftaar, 2020). It can also solve and minimize the issue of over-fitting by identifying the best tree structure for the calculation of the leaf values (Dorogush et al., 2018). CBR have recently been employed for soil parameters and soil carbon estimation (Xu et al.). Max depth, learning rate, and the number of iterations is the key hyper-parameters of the CBR model. It is similar to XGBR, important hyper-parameters were tuned by hyperparameter tuning using grid search with five-fold CV to select optimal ones which helps improve the CBR model performance.

Table 3
Lists of developed scenarios for soil moisture estimation.

Scenario	Data fusion	Number of features
SC1	S-2 + DEM	25
SC2	S-1 + DEM	30
SC3	S-1 + S-2	49
SC4	S-1 + S-2 + DEM	52
SC5	S-1 + S-2 + DEM with feature selection	21

2.5. Genetic algorithm (GA) for feature selection

Features selection is vital for the ML model's performance. It also helps simplify the models, reduce the time for training and testing model, and address overfitting issues. A genetic algorithm with the XGBR method was employed to determine automatically optimal indicators for the SM content retrieval in the study from the total of 52 variables derived from selected RS missions. GA implementation includes the following stages: (1) population formation from soil samples; (2) generation of a mating pool based on the highest fitness individual values; (3) the selection of parents from the mating pool by random selection methods; and (4) the generation of parents' offspring using crossover and mutation operators. The selected generation with optimal features illustrates the lowest root mean squared error, RMSE. In this research, firstly, the prediction accuracy of the four ML algorithms including XGBR, CBR, RFR, and SVM were tested with all 52 generated features. The best predictive model was used with GA to select the optimal features. The prediction performance of selected ML models using the optimal features were then tested and compared. The prediction accuracy of ML models for soil properties can be improved with the use of the GA for the selection of predictor variables (Xie et al., 2015).

2.6. Model performance evaluation

To assess the model performance of the soil moisture estimation, two standard testing criteria were used to evaluate the performance of ML techniques with various scenarios including: the root mean square error (RMSE), and the coefficient of determination (R^2). Superior model performance illustrates the higher R^2 and lower RMSE. These criteria are assessed using the equations below:

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

Table 4
Pearson's correlation analysis of input variables and measured SM.

Input variables	Correlation coefficient (*r*)	Input variables	Correlation coefficient (*r*)
B2	0.005	VV-VH	0.076
B3	−0.046	VV/VH	−0.045
B4	0.087	VH/VV	0.045
B5	0.064	VH_Contrast	−0.073
B6	−0.155	VH_Dissimilarity	−0.045
B7	−0.247	VH_Homogeneity	0.022
B8	−0.279	VH_Angular Second Moment	−0.037
B8A	−0.355	VH_Energy	−0.002
B11	0.049	VH_Maximum Probability	−0.009
NDWI	−0.366	VH_Entropy	−0.014
B12	0.125	VH_GLCM Mean	−0.437
RVI	−0.389	VH_GLCM Variance	−0.440
NDVI	−0.402	VH_GLCM Correlation	0.042
GNDVI	−0.249	VV_Contrast	−0.328
NDI45	−0.055	VV_Dissimilarity	−0.382
SAVI	−0.499	VV_Homo-geneity	0.401
MCARI	−0.070	VV_Angular Second Moment	0.332
IRECI	−0.568	VV_Energy	0.352
BI	0.031	VV_Maximum Probability	0.311
BI2	−0.111	VV_Entropy	−0.377
CI	−0.329	VV_GLCM Mean	−0.415
RI	0.142	VV_GLCM Variance	−0.414
VH	−0.414	VV_GLCM Correlation	0.311
VV	−0.347	DEM	−0.616
(VH + VV)/2	−0.403	Slope	−0.495
VH-VV	−0.083	TWI	0.368

Table 5
Model performance of the XGBR technique in five scenarios.

Scenario (SC)	Number of features	R^2 testing (30%)	RMSE (%)
SC1	25	0.757	1.307
SC2	30	0.627	1.621
SC3	49	0.469	1.934
SC4	52	0.783	1.236
SC5	21	0.891	0.875

$$R^2 = \frac{\sum_{i=1}^{n}\left(P_i - \overline{O_i}\right)}{\sum_{i=1}^{n}\left(O_i - \overline{O_i}\right)} \quad (3)$$

where: n indicates the number of soil samples; P_i and O_i illustrate the predicted SM value and measured SM value of the i sample, respectively.

Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) indicator were employed to compare the performance of difference ML models for soil moisture estimation. Lower AIC and BIC values illustrated the better prediction accuracy of regression models (Claeskens and Hjort, 2008). These indicators were evaluated using Eqs. (4) and (5) below:

$$AIC = n * \log\ (SSE/n) + 2K \quad (4)$$

$$BIC = n * \log\ (SSE/n) + \log(n) * K \quad (5)$$

where: SSE illustrates the sum of squares errors; n indicates the number ò soil samples, and K presents the parameter's number.

3. Results and discussion

3.1. Correlation analysis of predictor indicators and measured SM

The relationship between the input features derived from S-1, S-2, ALOS-DSM and measured SM content was computed by Pearson's correlation coefficient method. According to Table 4, indicators derived from ALOS imagery have a higher correlation with the SM content compared to other indicators. While DEM and Slope obtained negative correlations, TWI illustrated a positive correlation with the SM. All vegetation indices generated from S-2 demonstrated negative correlation with the SM content. Some these indices revealed higher correlations with the SM content including NDVI, SAVI, and IRECI. Color index from soil indices had a higher correlation to the SM compared to another SIs. NDWI confirmed a negative and high relationship with the estimation of SM. Regarding to the S-1 derived indicators, most transformation bands obtained weak correlations with the SM content, however; VV, VH, and most GLCM textures from VV confirmed strong relationships with the measured SM.

3.2. Evaluation and comparison of scenarios and different ML models

The proposed XGBR model was trained and tested with different scenarios which were developed by various features extracted from S-1, S-2 and ALOS DSM (Table 5). The genetic algorithm procedure with extreme gradient boosting regression was implemented to select the best subset comprising of 21 optimal predictors out of 52 variables with the best accuracy of 0.75. The SC5 with optimal number of features including seven vegetation indices (NDWI, RVI, NDVI, GNDVI, SAVI, MCARI, IREC1), 11 S-1 derived indicators (VH, VV, MeanVHVV, VV_Contrast, VV_Dissimilarity, VV_Homogeneity, VV_Angular Second Moment, VV_Energy, VV_Maximum Probability, VV_Entropy, VV_GLMC Correlation), and both three variables from ALOS DSM yielded the highest SM estimation accuracy with the highest R^2 of 0.891 in the validation phase and the lowest RMSE of 0.875%, followed by SC4 with the maximum number of features extracted from selected sensors. A combination of S-2 and ALOS DSM derived predictor features illustrated a higher performance than the combination of S1 and DEM and S-1 and S-2 generated indicators.

Three well-known ML techniques including CBR, RFR, and SVM were employed to compare the accurate estimation of the SM content with the proposed XGBR-GA model using multi-source EO data fusion. The comparison of ML techniques was conducted with optimal features derived from S-1, S-2, and ALOS DSM. According to Table 6, gradient boosting regression algorithms (XGBR) outperformed RFR and SVM. While XGBR-GA achieved a highest prediction accuracy with R^2 = 0.891 and RMSE = 0.875, followed by CBR with R^2 = 0.789 and RMSE = 1.217 and SVM with R^2 = 0.493 and RMSE = 1.850. The RFR produced a lowest prediction performance with R^2 = 0.368 and RMSE = 2.491. Moreover, the XGBR-GA also presented lowest value of AIC and BIC compared to CBR, RFR and SVM.

Fig. 6 presents the scatter plots of the estimated versus measured the SM content from four well-known ML techniques in testing phase using optimal features. The XGBR with higher R^2 value and lower RMSE, AIC, and BIC yielded a better prediction with optimal variables extracted from these

Table 6
Performance comparison of ML algorithms on agricultural SM estimation.

No	Machine learning model	R^2 testing (30%)	RMSE (%)	AIC	BIC
1	Extreme gradient boosting regression with GA (XGBR-GA)	0.891	0.875	155.36	194.21
2	CatBoost gradient boosting regression (CBR)	0.789	1.217	157.72	198.46
3	Random Forest regression (RFR)	0.368	2.491	186.49	226.54
4	Support Vector Machine regression SVM)	0.493	1.850	179.58	218.50

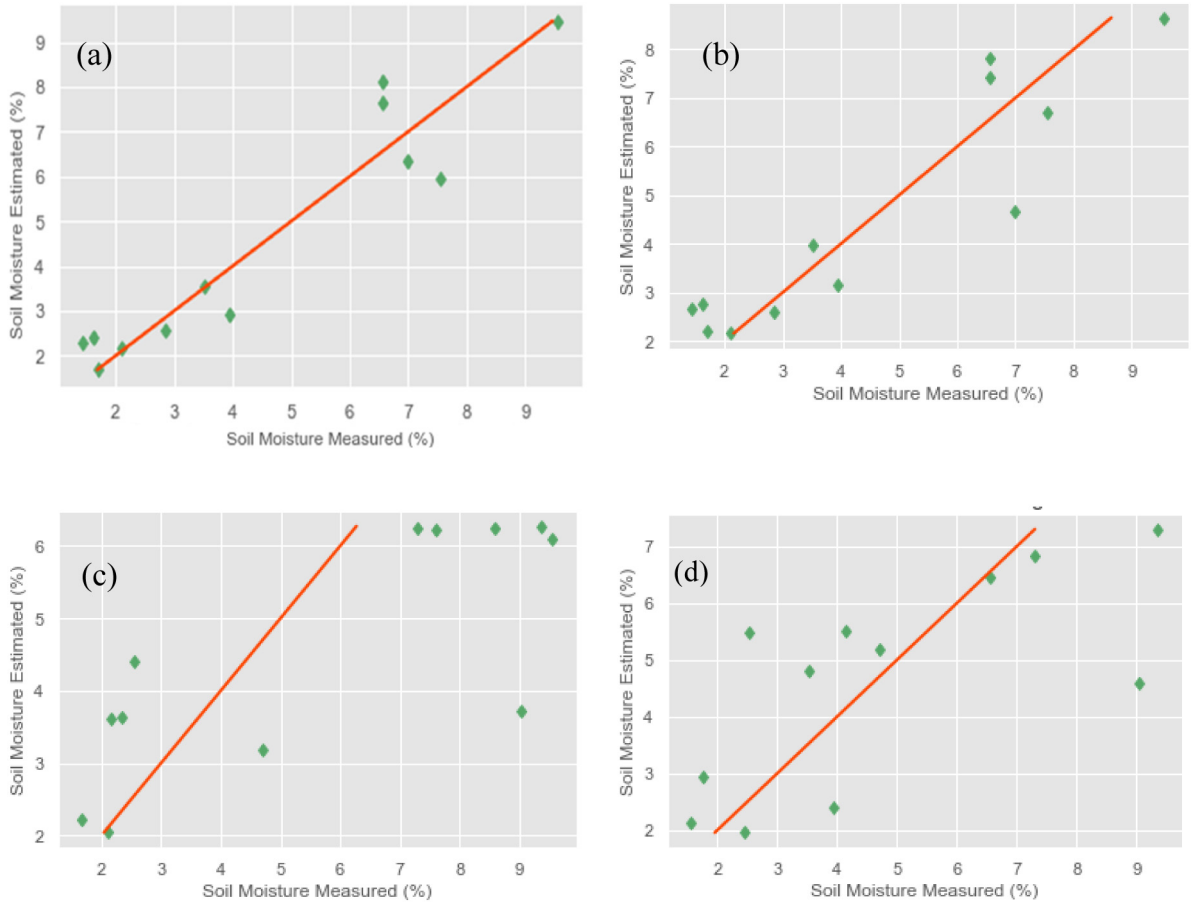


Fig. 6. Scatter plots of the measured SM and estimated SM using (a) XGBR, (b) CBR, (c) RFR, (d) and SVM.

multiple sensors using the genetic algorithm compared to CBR, RFR, and SVM. The proposed model using XGBR and GA indicates an R^2 value of 0.891, showing a higher prediction result compared to recent SM monitoring studies with R^2 reached 0.83 in SM prediction study using S1 and Landsat-7 data in Egypt (Mohamed et al., 2020) and R^2 of 0.72 in surface soil moisture estimation using S1 and S2 in India (Tripathi and Tiwari, 2020).

3.3. Spatial distribution patterns of SM maps

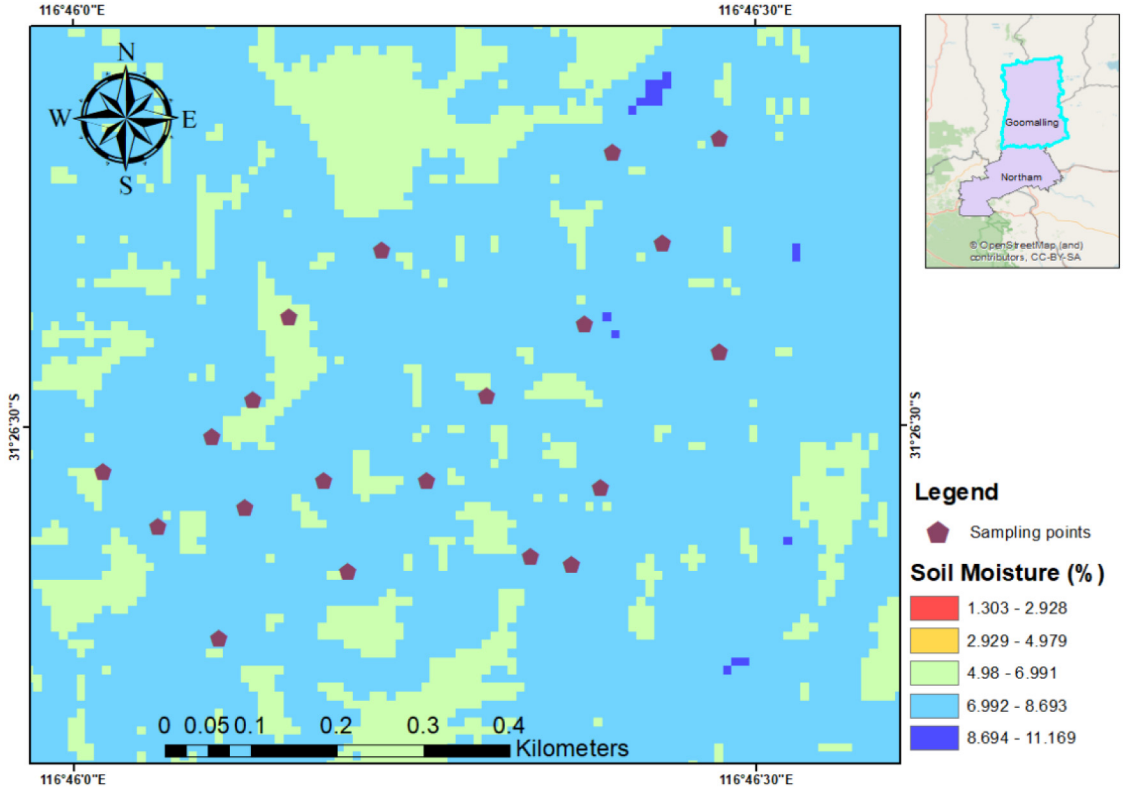
Based on scenario 5, the spatial dynamics of SM maps built for the Wests and Cookies areas using S-1, S-2, and ALOS DSM data fusion by the XGBR-GA model are revealed in Fig. 7. The XGBR model for the SM prediction in bare-soil pixels obtained the low level of uncertainty and stable prediction capabilities with the low standard deviation value. The proposed moisture prediction model using the XGBR-GA should be calibrated and tested with large-scale earth observation data, over several of land-use types, and various soil-depths.

3.4. Relative importance of SM prediction indicators

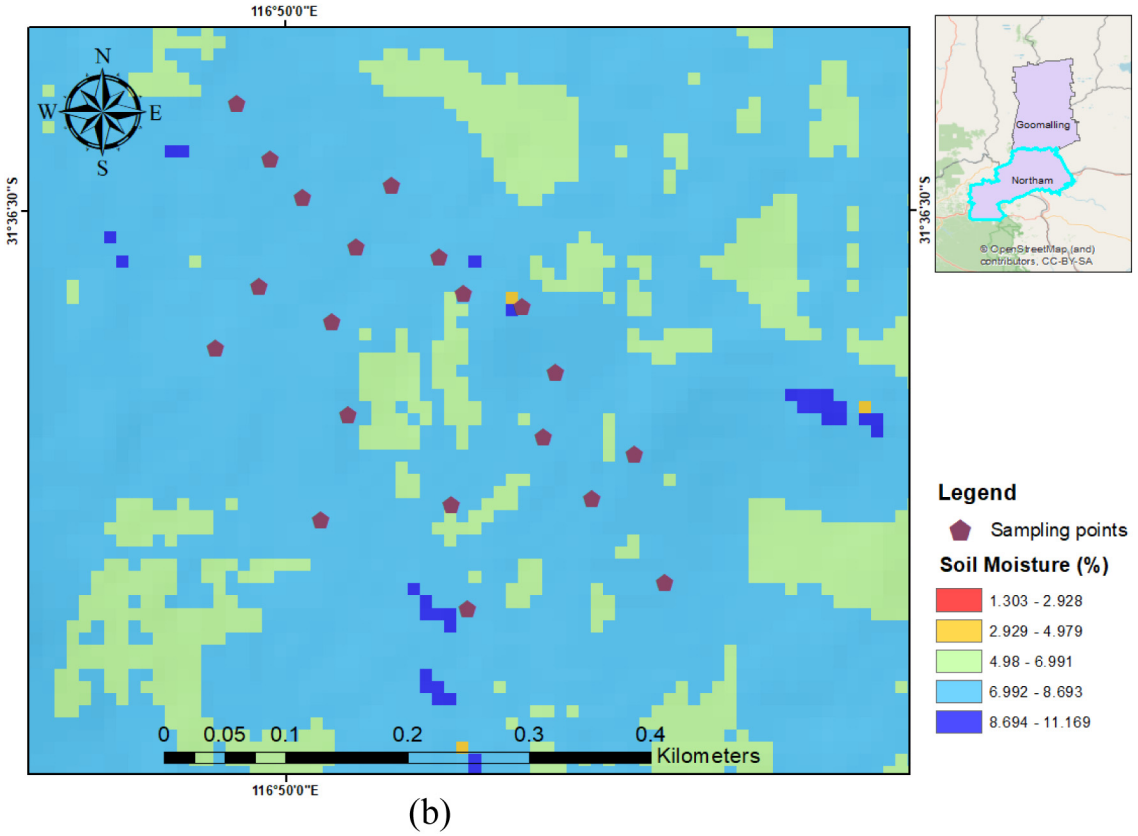
The estimation accuracy of the SM content has been greatly affected by predictor indicators selection and machine learning algorithm. The higher level of data fusion with optimal feature selection using the GA illustrated better prediction performance for retrieving the SM content. The XGBR had a higher capability to predict the SM pattern. The study also indicated that the GA could help improve the prediction accuracy of the SM estimation which is similar with the results from recent studies using ML models and GA for soil properties estimation (Xie et al., 2015). The successful application of ML models and big data from RS imagery in the SM prediction has been presented in much research at the regional, national and global scale (Carranza et al., 2021; Chaudhary et al., 2021; Cheng et al., 2022; Fang et al., 2021; Ma et al., 2021; Senanayake et al., 2021). The relative importance of optimal features using the GA is presented in Fig. 8. ALOS DSM-derived terrain indices played important roles in the SM prediction. Terrain variables were also mentioned as important indices for the SM prediction in previous studies (Ågren et al., 2021; Leempoel et al., 2015; Zhao et al., 2021). In addition, dual polarization VV, VH, and GLCM textures derived from S-1 are also crucial indices for the SM prediction. The SAR-based prediction indices can improve the estimation of soil moisture (El Hajj et al., 2017; Ma et al., 2020; Zhao et al., 2021). VH was illustrated as the most sensitive index for the SM retrieval in this study. Vegetation indices were selected as optimal features for the SM prediction such as the normalizer difference vegetation index (NDVI), and soil adjusted vegetation index (SAVI) which have been applied for not only vegetation classification, but also further indirectly the SM estimation (Kogan, 1995; Reza et al., 2020). Normalized difference water index (NDWI) from Sentinel 2 also highly correlated with the SM content (Ma et al., 2020). The soil moisture prediction model using the XGBR-GA should be calibrated and tested with large-scale earth observation data, over several of land-use types, and various soil-depths.

4. Conclusion

The present work presented a novel framework using the predictor variables from Sentinel datasets at 10 m and ALOS DSM at 30 m spatial resolution with a state-of-art machine learning technique (XGBR) and GA for the SM prediction. It is used for estimating the SM content in study sites of Western Australia. It can be seen that the combination of the selected remote sensing dataset illustrated to be very effective for the SM prediction.



(a)



(b)

Fig. 7. Maps of SM content in study areas: (a) Wests and (b) Cookies using XGBR-GA combined data fusion.

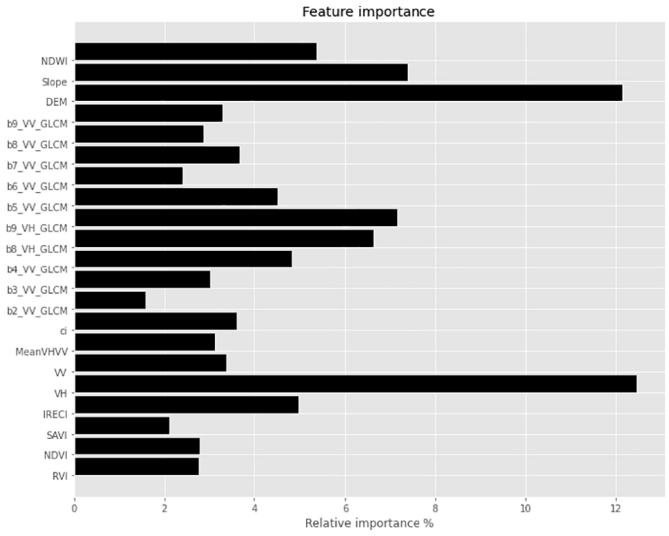


Fig. 8. Variable importance of optimal features derived from multi-source EO data.

High level of data fusion and the GA method for optimal features selection showed remarkably better prediction accuracy than single sensor derived features or scenarios without feature optimization. The XGBR model with 21 optimal prediction variables using genetic algorithm approach illustrated the highest prediction performance (R^2 = 0.891, RMSE = 0.875%). In addition, the proposed XGBR model combined with GA algorithm for variables selected can produce SM maps at 10 m spatial resolution using freely remote sensing datasets with a precise accuracy at different scales from field plots to region areas. VH and DEM had the highest relative importance in predicting the SM dynamics. The proposed model should be tested in large-scale areas with various land-use characteristics in further studies. In conclusion, this SM pattern monitoring approach can assist agricultural drought monitoring, the development of appropriate water management strategies, and precision agriculture in terms of climate change.

CRediT authorship contribution statement

Thu Thuy Nguyen: investigation, writing - original draft, methodology, formal analysis, data curation.

Huu Hao Ngo: supervision, investigation, project administration, conceptualization, review & editing.

Wenshan Guo: supervision, investigation, review & editing.

Soon Wang Chang: investigation, project administration, review & editing.

Dinh Duc Nguyen: methodology, formal analysis, resources, review.

Chi Trung Nguyen: methodology, data curation, resources.

Jian Zhang: methodology, data curation, resources.

Shuang Liang: methodology, resources, review.

Xuan Thanh Bui: methodology, data curation, review.

Ngoc Bich Hoang: methodology, resources, review.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the University of Technology, Sydney, Australia (UTS, RIA NGO; UTS SRS 2021), and the Australian Postgraduate Research Intern (APR. Intern), Astron Environmental Services Company.

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.

Aksoy, S., Yildirim, A., Gorji, T., Hamzehpour, N., Tanik, A., Sertel, E., 2021. Assessing the performance of machine learning algorithms for soil salinity mapping in Google earth engine platform using sentinel-2A and Landsat-8 OLI data. Adv. Space Res. 69, 1072–1086.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Carranza, C., Nolet, C., Pezij, M., van der Ploeg, M., 2021. Root zone soil moisture estimation with Random Forest. J. Hydrol. 593.

Chaudhary, S.K., Srivastava, P.K., Gupta, D.K., Kumar, P., Prasad, R., Pandey, D.K., Das, A.K., Gupta, M., 2021. Machine learning algorithms for soil moisture estimation using Sentinel-1: model development and implementation. Adv. Space Res. 69, 1799–1812.

Chen, T., Guestrin, C., 2016. XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.

Cheng, M., Li, B., Jiao, X., Huang, X., Fan, H., Lin, R., Liu, K., 2022. Using multimodal remote sensing data to estimate regional-scale soil moisture content: a case study of Beijing, China. Agric. Water Manag. 260.

Claeskens, G., Hjort, N.L., 2008. Model Selection and Model Averaging330. United Kingdom University Press Cambridge, Cambridge.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297.

Cristianini, N., Ricci, E., 2008. Support vector machines. In: Kao, M.-Y. (Ed.), Encyclopedia of Algorithms. Springer US, Boston, MA, pp. 928–932.

Daughtry, C.S.T., Walthall, C.L., Kim, M.S., de Colstoun, E.B., McMurtrey, J.E., 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. Remote Sens. Environ. 74 (2), 229–239.

Delegido, J., Verrelst, J., Alonso, L., Moreno, J., 2011. Evaluation of Sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. Sensors (Basel, Switzerland) 11 (7), 7063–7081.

Dorogush, A.V., Ershov, V., Gulin, A., 2018. CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363 https://doi.org/10.48550/arXiv.1810.11363.

Dubois, A., Teytaud, F., Verel, S., 2021. Short term soil moisture forecasts for potato crop farming: a machine learning approach. Comput. Electron. Agric. 180.

efBabaeian, E., Paheding, S., Siddique, N., Devabhaktuni, V.K., Tuller, M., 2021. Estimation of root zone soil moisture from ground and remotely sensed soil information with multisensor data fusion and automated machine learning. 260.

El Hajj, M., Baghdadi, N., Zribi, M., Bazzi, H., 2017. Synergic use of Sentinel-1 and Sentinel-2 images for operational soil moisture mapping at high spatial resolution over agricultural areas. Remote Sens. 9 (12), 1292.

Entekhabi, D., Njoku, E.G., Neill, P.E.O., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J., Kimball, J., Piepmeier, J.R., Koster, R.D., Martin, N., McDonald, K.C., Moghaddam, M., Moran, S., Reichle, R., Shi, J.C., Spencer, M.W., Thurman, S.W., Tsang, L., Zyl, J.V., 2010. The soil moisture active passive (SMAP) Mission. Proc. IEEE 98 (5), 704–716.

Escadafal, R., 1989. Remote sensing of arid soil surface color with Landsat thematic mapper. Adv. Space Res. 9 (1), 159–163.

Fang, B., Kansara, P., Dandridge, C., Lakshmi, V., 2021. Drought monitoring using high spatial resolution soil moisture data over Australia in 2015–2019. J. Hydrol. 594.

Florinsky, I.V., Eilers, R.G., Manning, G.R., Fuller, L.G., 2002. Prediction of soil properties by digital terrain modelling. Environ. Model Softw. 17 (3), 295–311.

Frampton, W.J., Dash, J., Watmough, G., Milton, E.J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. ISPRS J. Photogramm. Remote Sens. 82, 83–92.

Fu, X., Shao, M., Wei, X., Horton, R., 2010. Soil organic carbon and total nitrogen as affected by vegetation types in northern Loess Plateau of China. Geoderma 155 (1), 31–35.

Gago, J., Douthe, C., Coopman, R.E., Gallego, P.P., Ribas-Carbo, M., Flexas, J., Escalona, J., Medrano, H., 2015. UAVs challenge to assess water stress for sustainable agriculture. Agric. Water Manag. 153, 9–19.

Gao, B.-C., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. 58 (3), 257–266.

Ge, X., Wang, J., Ding, J., Cao, X., Zhang, Z., Liu, J., Li, X., 2019. Combining UAV-based hyperspectral imagery and machine learning algorithms for soil moisture content monitoring. PeerJ 7, e6926.

Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., Wolff, E., 2018. Very high resolution object-based land use-land cover urban classification using extreme gradient boosting. IEEE Geosci. Remote Sens. Lett. 15 (4), 607–611.

Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N., 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sens. Environ. 58 (3), 289–298.

Gómez, D., Salvador, P., Sanz, J., Casanova, J.L., 2020. Modelling desert locust presences using 32-year soil moisture data on a large-scale. Ecol. Indic. 117.

Gómez, D., Salvador, P., Sanz, J., Rodrigo, J.F., Gil, J., Casanova, J.L., 2021. Prediction of desert locust breeding areas using machine learning methods and SMOS (MIR_SMNRT2) near real time product. J. Arid Environ. 194.

Ha, N.T., Manley-Harris, M., Pham, T.D., Hawes, I., 2021. The use of radar and optical satellite imagery combined with advanced machine learning and metaheuristic optimization techniques to detect and quantify above ground biomass of intertidal seagrass in a New Zealand estuary. Int. J. Remote Sens. 42 (12), 4712–4738.

Hancock, J.T., Khoshgoftaar, T.M., 2020. CatBoost for big data: an interdisciplinary review. J. Big Data 7 (1), 94.

Hosoda, M., Tokonami, S., Suzuki, T., Janik, M., 2020. Machine learning as a tool for analysing the impact of environmental parameters on the radon exhalation rate from soil. Radiat. Meas. 138.

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 25 (3), 295–309.
Jin, X., Song, K., Du, J., Liu, H., Wen, Z., 2017. Comparison of different satellite bands and vegetation indices for estimation of soil organic matter based on simulated spectral configuration. Agric. For. Meteorol. 244–245, 57–71.
Karthikeyan, L., Mishra, A.K., 2021. Multi-layer high-resolution soil moisture estimation using machine learning over the United States. Remote Sens. Environ. 266.
Kerr, Y.H., Waldteufel, P., Wigneron, J., Martinuzzi, J., Font, J., Berger, M., 2001. Soil moisture retrieval from space: the soil moisture and ocean salinity (SMOS) mission. IEEE Trans. Geosci. Remote Sens. 39 (8), 1729–1735.
Kingwell, R., Islam, N., Xayavong, V., 2020. Farming systems and their business strategies in South-Western Australia: a decadal assessment of their profitability. Agric. Syst. 181, 102827.
Kogan, F.N., 1995. Application of vegetation index and brightness temperature for drought detection. Adv. Space Res. 15 (11), 91–100.
Kopecký, M., Macek, M., Wild, J., 2021. Topographic wetness index calculation guidelines based on measured soil moisture and plant species composition. Sci. Total Environ. 757, 143785.
Leempoel, K., Parisod, C., Geiser, C., Daprà, L., Vittoz, P., Joost, S., 2015. Very high-resolution digital elevation models: are multi-scale derived variables ecologically relevant? Methods Ecol. Evol. 6 (12), 1373–1383.
Ma, C., Li, X., McCabe, M.F., 2020. Retrieval of high-resolution soil moisture through combination of Sentinel-1 and Sentinel-2 data. Remote Sens. 12 (14), 2303.
Ma, G., Ding, J., Han, L., Zhang, Z., Ran, S., 2021. Digital mapping of soil salinization based on Sentinel-1 and Sentinel-2 data combined with machine learning algorithms. Region. Sustain. 2 (2), 177–188.
Mathieu, R., Pouget, M., Cervelle, B., Escadafal, R., 1998. Relationships between satellite-based radiometric indices simulated using laboratory reflectance data and typic soil color of an arid environment. Remote Sens. Environ. 66 (1), 17–28.
Michalski, A., Czajewski, J., 2004. The accuracy of the global positioning systems. IEEE Inst. Meas. Mag. 7 (1), 56–60.
Mohamed, E.S., Ali, A., El-Shirbeny, M., Abutaleb, K., Shaddad, S.M., 2020. Mapping soil moisture and their correlation with crop pattern using remotely sensed data in arid region. Egypt. J. Remote Sens. Space Sci. 23 (3), 347–353.
Murphy, P.N.C., Ogilvie, J., Castonguay, M., Zhang, C.-F., Meng, F.-R., Arp, P.A., 2008. Improving forest operations planning through high-resolution flow-channel and wet-areas mapping. For. Chron. 84 (4), 568–574.
Nguyen, T.T., Pham, T.D., Nguyen, C.T., Delfos, J., Archibald, R., Dang, K.B., Hoang, N.B., Guo, W., Ngo, H.H., 2022. A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion. Sci. Total Environ. 804, 150187.
Nikolakopoulos, K.G., 2020. Accuracy assessment of ALOS AW3D30 DSM and comparison to ALOS PRISM DSM created with classical photogrammetric techniques. Eur. J. Remote Sens. 53 (sup2), 39–52.
Pasqualotto, N., D'Urso, G., Bolognesi, S.F., Belfiore, O.R., Van Wittenberghe, S., Delegido, J., Pezzola, A., Winschel, C., Moreno, J., 2019. Retrieval of evapotranspiration from Sentinel-2: comparison of vegetation indices, semi-empirical models and SNAP biophysical processor approach. Agronomy 9 (10), 663.
Pham, T.D., Yokoya, N., Nguyen, T.T.T., Le, N.N., Ha, N.T., Xia, J., Takeuchi, W., Pham, T.D., 2020. Improvement of mangrove soil carbon stocks estimation in North Vietnam using Sentinel-2 data and machine learning approach. GISci. Remote Sens. 58 (1), 68–87.
Prasad, R., Deo, R.C., Li, Y., Maraseni, T., 2018. Soil moisture forecasting by a hybrid machine learning technique: ELM integrated with ensemble empirical mode decomposition. Geoderma 330, 136–161.
Reza, H., Davoud, Z., Mohammad Reza, N., Bakhtiar, F., Mehdi, R., 2020. Modification on optical trapezoid model for accurate estimation of soil moisture content in a maize growing field. J. Appl. Remote. Sens. 14 (3), 1–19.
Robinson, D.A., Campbell, C.S., Hopmans, J.W., Hornbuckle, B.K., Jones, S.B., Knight, R., Ogden, F., Selker, J., Wendroth, O., 2008. Soil moisture measurement for ecological and hydrological watershed-scale observatories: a review. Vadose Zone J. 7 (1), 358–389.
Rouse, J., Haas, R.H., Schell, J.A., Deering, D., 1973. Monitoring Vegetation Systems in the Great Plains With ERTS.
Salim, R.A., Islam, N., 2010. Exploring the impact of R&D and climate change on agricultural productivity growth: the case of Western Australia*. Aust. J. Agric. Resour. Econ. 54 (4), 561–582.
Schmidt, A., Mainwaring, D.B., Maguire, D.A., 2020. Development of a tailored combination of machine learning approaches to model volumetric soil water content within a Mesic forest in the Pacific northwest. J. Hydrol. 588.
Schönauer, M., Väätäinen, K., Prinz, R., Lindeman, H., Pszenny, D., Jansen, M., Maack, J., Talbot, B., Astrup, R., Jaeger, D., 2021. Spatio-temporal prediction of soil moisture and soil strength by depth-to-water maps. Int. J. Appl. Earth Obs. Geoinf. 105.
Senanayake, I.P., Yeo, I.Y., Walker, J.P., Willgoose, G.R., 2021. Estimating catchment scale soil moisture at a high spatial resolution: integrating remote sensing and machine learning. Sci. Total Environ. 776.
Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8 (2), 127–150.
Tuller, M., Babaeian, E., Jones, S., Montzka, C., Vereecken, H., Sadeghi, M., 2019. The paramount societal impact of soil moisture. Eos 100.
Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldaña, J., de Rosnay, P., Jann, A., Schneider, S., Komma, J., Kubu, G., Brugger, K., Aubrecht, C., Züger, J., Gangkofner, U., Kienberger, S., Brocca, L., Wang, Y., Blöschl, G., Eitzinger, J., Steinnocher, K., 2013. The ASCAT soil moisture product: a review of its specifications, validation results, and emerging applications. Meteorol. Z. 22 (1), 5–33.
Warner, D.L., Guevara, M., Callahan, J., Vargas, R., 2021. Downscaling satellite soil moisture for landscape applications: a case study in Delaware, USA. J. Hydrol. Region. Stud. 38.
Xie, H., Zhao, J., Wang, Q., Sui, Y., Wang, J., Yang, X., Zhang, X., Liang, C., 2015. Soil type recognition as improved by genetic algorithm-based variable selection using near infrared spectroscopy and partial least squares discriminant analysis. Sci. Rep. 5 (1), 10930.
Tripathi, A., Tiwari, R.K., 2020. Synergetic utilization of sentinel-1 SAR and sentinel-2 optical remote sensing data for surface soil moisture estimation for Rupnagar, Punjab, India. Geocarto Int. 1–22.
Zhao, Z., Yang, Q., Ding, X., Xing, Z., 2021. Model prediction of the soil moisture regime and soil nutrient regime based on DEM-derived topo-hydrologic variables for mapping ecosites. Land 10 (5).
Zhou, T., Geng, Y., Chen, J., Liu, M., Haase, D., Lausch, A., 2020a. Mapping soil organic carbon content using multi-source remote sensing variables in the Heihe River basin in China. Ecol. Indic. 114.
Zhou, T., Geng, Y., Chen, J., Pan, J., Haase, D., Lausch, A., 2020b. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. Sci. Total Environ. 729, 138244.