The Open Agriculture Journal
ISSN: 1874-3315
DOI: 10.2174/0118743315495282260915110324, 2026, 20, e18743315495282
1
RESEARCH ARTICLE
OPEN ACCESS

AI-enabled UAV-based Soil Organic Carbon Mapping in Arid Environments: A Pilot Study Protocol

Moath Awawdeh[1,*], Irfan Ahmed[1], Anees Bashir[1], Tarig Faisal[1], Nicky Putra[2], Almaha Jamal[1], Afra Rashed[1], Hamda Yousif[1] and Sarah Sadeq[1]

[1]*Department of Electrical Engineering, Higher Colleges of Technology, Al Dhanna Campus, Abu Dhabi, UAE*
[2]*Department of Chemical Engineering, Higher Colleges of Technology, Al Dhanna Campus, Abu Dhabi, UAE*

Abstract:

Introduction: Soil organic carbon (SOC) is an important indicator of soil health, agricultural productivity, and carbon sequestration potential. However, accurate and scalable SOC mapping in arid environments is constrained by high spatial heterogeneity and the limitations of conventional soil sampling. This study aims to develop a standardized UAV-enabled framework for high-resolution SOC mapping in arid agricultural environments.

Methods: A pilot-study protocol integrating UAV-based hyperspectral remote sensing with artificial intelligence and machine learning was developed. The workflow encompasses study-site selection, ground-reference sampling, UAV hyperspectral data acquisition, radiometric and geometric preprocessing, spectral feature extraction and selection, machine-learning model development, validation, uncertainty assessment, and performance evaluation using R^2, RMSE, and MAE. The protocol also incorporates assessment of environmental confounders, including soil moisture, surface roughness, and crop residues.

Results: The resulting framework provides a systematic and reproducible workflow for UAV-based SOC estimation, integrating field observations, hyperspectral features, predictive modelling, and uncertainty assessment. It establishes defined procedures for evaluating model robustness and transferability across varying field conditions.

Discussion: The framework addresses an important methodological gap in UAV-enabled SOC mapping by integrating remote sensing and AI within a standardized pilot-study design. Its emphasis on environmental confounders and uncertainty assessment can improve the reliability and comparability of SOC mapping studies. However, field validation across diverse arid environments remains necessary.

Conclusion: The proposed protocol provides a practical foundation for reproducible SOC mapping and subsequent field validation, supporting precision agriculture, sustainable soil management, and carbon monitoring, reporting, and verification (MRV) in arid regions.

Keywords: Soil organic carbon, Evaluation framework, Hyperspectral imaging, Unmanned aerial vehicles, Machine learning, Sustainable agriculture, Carbon sequestration.

Received: April 28, 2026
Revised: July 17, 2026
Accepted: August 03, 2026
Published: September 18, 2026

*Address correspondence to this author at the Department of Electrical Engineering, Higher Colleges of Technology, Al Dhanna Campus, Abu Dhabi, UAE; E-mail: mawawdeh@hct.ac.ae

Cite as: Awawdeh M, Ahmed I, Bashir A, Faisal T, Putra N, Jamal A, Rashed A, Yousif H, Sadeq S. AI-enabled UAV-based Soil Organic Carbon Mapping in Arid Environments: A Pilot Study Protocol. Open Agric J, 2026; 20: e18743315495282. http://dx.doi.org/10.2174/0118743315495282260915110324

1. INTRODUCTION

Soil organic carbon (SOC) constitutes a pivotal element of terrestrial carbon reservoirs and is extensively acknowledged as a critical metric of soil vitality, agricultural efficiency, and ecosystem robustness. SOC exerts a significant influence on soil architecture, nutrient cycling, and water retention, thereby fulfilling an essential function in the promotion of crop development and the stabilization of agricultural frameworks in the context of escalating climatic pressures. In addition to its agronomic significance, SOC serves as a substantial segment of the global carbon cycle, thereby establishing a connection between soil management methodologies and efforts aimed at climate mitigation and enduring environmental sustainability [1, 2]. As a result, the precise quantification and surveillance of SOC have emerged as primary aims within both agricultural management practices and climate-conscious land-use initiatives [3].

Increasing attention has been directed toward the role of SOC in sustainable land management and carbon stock enhancement. High-resolution SOC information supports precision agriculture practices, enables site-specific soil amendments, and provides a scientific basis to improve soil fertility and carbon storage through targeted interventions [4, 5]. In dryland and degradation-prone systems, SOC is often embedded within broader soil quality and degradation indices, where its spatial variability has been shown to correlate with crop productivity and land-use sustainability [6]. In parallel, imaging spectroscopy has begun to emerge as a supporting technology within measurement, monitoring, reporting, and verification (MRV) ecosystems for climate-smart agricultural practices, particularly in the context of tracking carbon-related benefits of management interventions [7].

Reliable SOC assessment remains difficult when using traditional soil sampling and laboratory analysis, despite its importance. Conventional methods are not suitable for routine monitoring in vast or diverse agricultural landscapes due to their labor-intensive nature, destructive nature, and intrinsic spatial coverage limitations [8]. Spatial variability in SOC frequently outweighs treatment effects at the field scale, necessitating extensive sampling to identify significant patterns-an approach that is rarely practical in practice [9]. Furthermore, the lack of agreement on the best SOC remote sensing techniques has been brought to light by synthesis studies and meta-analyses, highlighting ongoing ambiguity in model robustness and transferability [3].

Despite significant advances in UAV-based hyperspectral remote sensing, digital soil mapping, and artificial intelligence (AI) for soil organic carbon (SOC) estimation, existing studies have predominantly focused on developing predictive models and reporting site-specific mapping performance under particular environmental conditions. Comparatively limited attention has been given to establishing standardized, reproducible methodological protocols that comprehensively integrate field sampling design, UAV mission planning, hyperspectral data acquisition, preprocessing, feature engineering, AI-based model development, validation procedures, uncertainty assessment, and quality-control measures into a unified workflow. This methodological gap limits reproducibility, hinders comparisons across studies, and constrains the broader adoption of UAV-enabled SOC mapping, particularly in arid agricultural environments where soil heterogeneity, sparse vegetation cover, high surface reflectance, and environmental variability introduce additional complexities that require carefully standardized implementation procedures.

Unlike previous studies that primarily report completed experimental results, predictive model performance, or site-specific SOC distribution maps, this work presents a pilot study protocol designed to establish a scientifically rigorous and reproducible methodological framework for future implementation. Rather than evaluating the predictive accuracy of a particular machine learning algorithm, the proposed protocol standardizes the complete workflow from field sampling and UAV hyperspectral data acquisition to spectral preprocessing, AI model development, validation, uncertainty analysis, and reporting procedures. This protocol-oriented approach is intended to facilitate methodological consistency, enhance reproducibility, and provide a common reference framework for future pilot deployments and comparative studies in arid agricultural environments.

Hyperspectral remote sensing has demonstrated strong potential as a non-destructive alternative for SOC estimation due to its ability to capture subtle absorption features associated with organic matter and related soil constituents. Both airborne and spaceborne hyperspectral systems have been successfully used to retrieve SOC by exploiting spectral transformations, feature selection techniques, and chemometric or machine learning models [10–12]. However, the reliability of hyperspectral SOC estimation is highly sensitive to radiometric consistency, viewing geometry, and environmental noise. Studies employing bidirectional reflectance distribution function (BRDF) correction and spectral unmixing have shown that appropriate preprocessing can substantially improve SOC prediction accuracy and spatial coherence [13, 14]. Similarly, large-scale analyses have demonstrated that rigorous spectral quality control and outlier filtering are essential for improving model stability and generalization [14].

Advanced modeling techniques are required due to the large dimensionality and nonlinear nature of hyperspectral data. In order to overcome these difficulties, machine learning and deep learning methods are being used more frequently. This allows for the reliable inversion of SOC and associated soil characteristics from complex spectral signatures. It has been demonstrated that ensemble learning, convolutional neural networks, and gradient boosting frameworks perform better than single-model approaches, especially when paired with efficient feature selection and dimensionality reduction techniques [15]. While reducing redundancy and noise, competitive feature selection techniques, wavelet transformations, and fractional-order derivatives have further increased sensitivity to SOC-related spectral features [1, 16, 17].

The choice of sensor platform results in additional trade-offs from an operational standpoint. Due to their low spectral resolution and sensitivity to atmospheric impacts, satellite-based multispectral systems offer reliable and affordable coverage at regional scales [18, 19]. Unmanned aerial vehicle (UAV)- based hyperspectral sensing provides high spatial resolution and flexible deployment, enabling detailed characterization of soil organic carbon (SOC) variability at the field scale. Comparative studies indicate that UAV hyperspectral data can capture fine-scale spatial heterogeneity critical for precision soil management, whereas satellite-based observations typically provide broader coverage with greater temporal stability [20]. By incorporating supplementary data on crop growth and management, multi-temporal UAV captures have further shown the potential to overcome constrained bare-soil windows [21].

Real-world agricultural environments introduce additional complexity through vegetation cover, soil moisture variability, and mixed surface conditions. These factors can obscure soil spectral signatures and degrade SOC prediction performance if not explicitly addressed. Recent studies have shown that integrating data from remote sensing (RS), proximal soil sensing (PSS), and other sensor sources can improve SOC prediction performance, particularly when supported by robust calibration procedures, validation across diverse pedo-climatic conditions, and appropriate data processing and modeling frameworks [22]. UAV-based hyperspectral systems have also shown strong capability in retrieving soil carbon-related properties under moderately to densely vegetated conditions when supported by robust regression and validation strategies [23]. Data fusion approaches combining hyperspectral imagery with structural information, such as LiDAR, further improve model performance in complex environments [24].

The challenges associated with SOC estimation are particularly pronounced in arid and semi-arid regions, where low organic carbon content, high surface reflectance, salinity, and sparse vegetation complicate spectral interpretation. In such environments, SOC models often exhibit limited generalization, especially when restricted to topsoil layers or single land-use types [1]. Dryland studies have emphasized the sensitivity of SOC-related indicators to management and irrigation practices, highlighting the need for tailored preprocessing and modeling strategies [6, 25]. Recent work in the United Arab Emirates has demonstrated that, despite these challenges, reliable SOC estimation is achievable through the integration of hyperspectral and multispectral data combined with environment-specific preprocessing techniques, providing the first regionally validated evidence for SOC remote sensing in arid agricultural systems [26]. This pilot protocol is also timely in the context of the United Arab Emirates (UAE), where national policy is increasingly coupling climate action, food security, and measurable sustainability outcomes. The UAE's 'Net Zero by 2050' strategic direction emphasizes economy-wide decarbonization pathways and measurable progress tracking, while the National Food Security Strategy 2051 prioritizes sustainable domestic production enabled by advanced technologies. In parallel, COP28 in the UAE elevated food systems and agriculture within the global climate agenda through the adoption of the COP28 Declaration on Food and Agriculture. These directions strengthen the relevance of scalable, measurement-ready SOC monitoring approaches-particularly those compatible with future monitoring, reporting, and verification (MRV) workflows-because SOC is both an agronomic soil-health indicator and a carbon accounting variable for climate-smart land management [27–29]. From an implementation standpoint, the UAE has already demonstrated national interest in drone-enabled environmental and agricultural monitoring, including MOCCAE-led aerial mapping initiatives that applied remote sensing and aerial imagery to characterize agricultural areas and assets. Such efforts highlight the practical value of UAV-based sensing for operational data collection in UAE agricultural settings and support the rationale for a UAV hyperspectral pilot that is designed around traceable sampling, quality control, and reproducible reporting [30, 31].

The viability of using UAV-enabled hyperspectral imaging in conjunction with machine learning for SOC mapping in a dry agricultural setting is investigated in this research through a pilot-based study approach (see Fig. **1** for the Conceptual Overview). The work highlights methodological rigor, thorough preprocessing, and validated modeling to enable climate-resilient and carbon-aware agricultural decision-making, building on recent developments and regional evidence [26]. The design and evaluation protocol for a pilot study is given in this publication. For UAV-based hyperspectral SOC estimation in arid agricultural environments, it outlines a repeatable workflow and validation strategy; empirical findings and SOC maps will be presented in further work following field deployment and data collection.

The novelty of this study lies in the development of an integrated pilot protocol that combines UAV-based hyperspectral remote sensing, AI-driven predictive modelling, standardized validation procedures, uncertainty assessment, and quality assurance within a single reproducible framework tailored for arid agricultural systems. Accordingly, the objectives of this study are to: (i) develop a standardized protocol for UAV-based hyperspectral SOC mapping; (ii) establish a reproducible workflow encompassing field sampling, hyperspectral data acquisition, preprocessing, machine learning model development, and validation; (iii) identify key environmental and methodological considerations influencing protocol implementation in arid environments; and (iv) provide a transferable methodological framework capable of supporting future experimental studies and contributing to monitoring, reporting, and verification (MRV) initiatives for sustainable soil and carbon management.

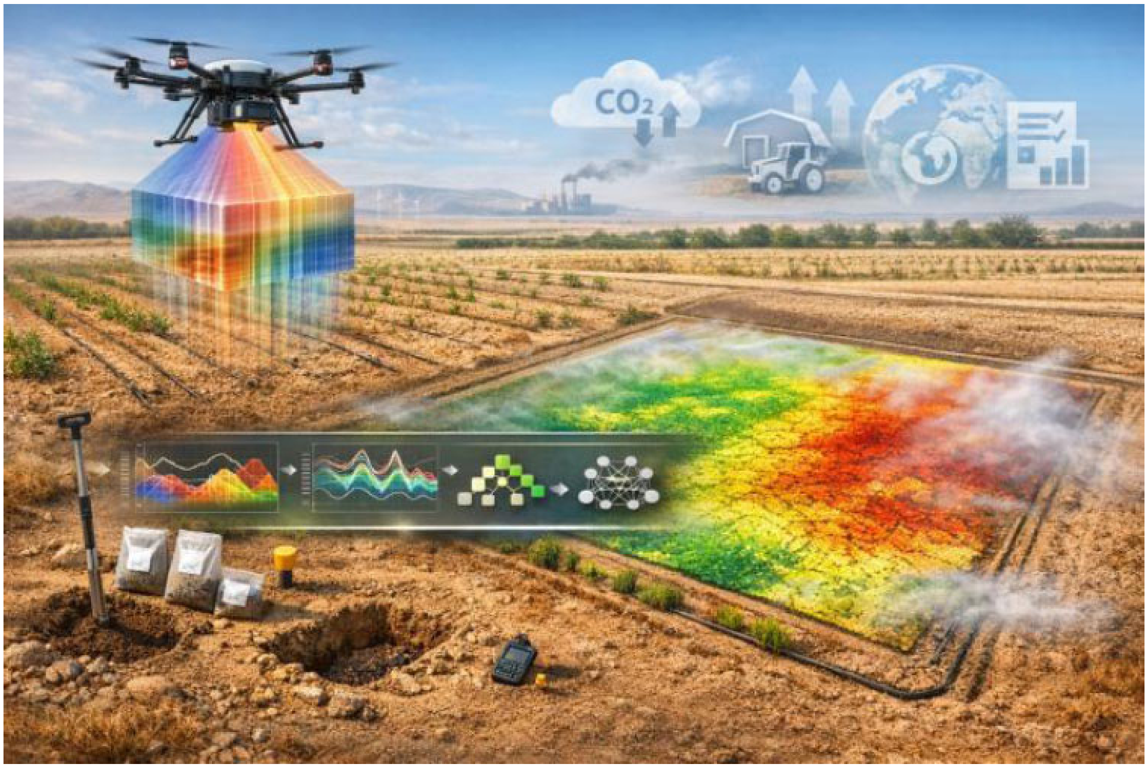

Fig. (1). Conceptual overview of the proposed pilot framework for AI-enabled UAV-based hyperspectral soil organic carbon (SOC) mapping in arid agricultural environments.

Contributions: The main contributions of this paper (protocol deliverables) include:

- A pilot-scale SOC mapping protocol tailored to arid agriculture, covering sampling design, reference measurement planning, UAV hyperspectral acquisition, preprocessing, feature engineering, model training, and validation.
- A standard evaluation plan (metrics, cross-validation strategy, transfer testing, uncertainty reporting) to ensure transparent and reproducible reporting.
- A deployment-ready reporting template, enabling MRV-oriented SOC monitoring and future decision-support integration.

2. RELATED WORK

The research shows that although SOC remote sensing has advanced significantly, issues with model robustness, environmental confounding factors, and practical deployment-particularly in arid environments-remain. The lack of a generally dependable SOC remote sensing solution is highlighted by meta-analytical evidence, which emphasizes the necessity of properly planned pilot studies with transparent validation and cautious result interpretation [3]. In this context, region-specific investigations that integrate hyperspectral sensing, machine learning, and ground validation are essential for advancing SOC monitoring toward practical application. The UAE's broader decarbonization momentum also reinforces the importance of high-integrity measurement systems. National initiatives are advancing monitoring and reporting infrastructure, and major UAE industry actors have publicly framed decarbonization as a measurable, technology-enabled pathway exemplified by ADNOC's net-zero operations ambition by 2045 and its participation in COP28-launched sector decarbonization commitments. While the present study is focused on agricultural SOC, the proposed protocol is intentionally structured to be MRV-compatible (transparent stages, auditable reporting artifacts, and uncertainty-aware evaluation), aligning with the UAE's growing emphasis on measurement-enabled sustainability [32–34]. Figure **2** presents a multi-perspective positioning of the proposed study across technology readiness, sensing capability, and innovation contribution dimensions. The technology readiness *versus* adoption landscape highlights that UAV-based hyperspectral SOC mapping currently resides in an early-to-intermediate adoption phase, particularly in arid agricultural contexts. While laboratory spectroscopy and satellite-based multispectral approaches demonstrate higher adoption rates, they also exhibit inherent limitations, either in spatial representativeness or spectral sensitivity, when deployed under real-world field conditions. The placement of this study within the "innovation zone" emphasizes that its contribution lies not in incremental model optimization, but in bridging high spectral fidelity with field-scale operational realism.

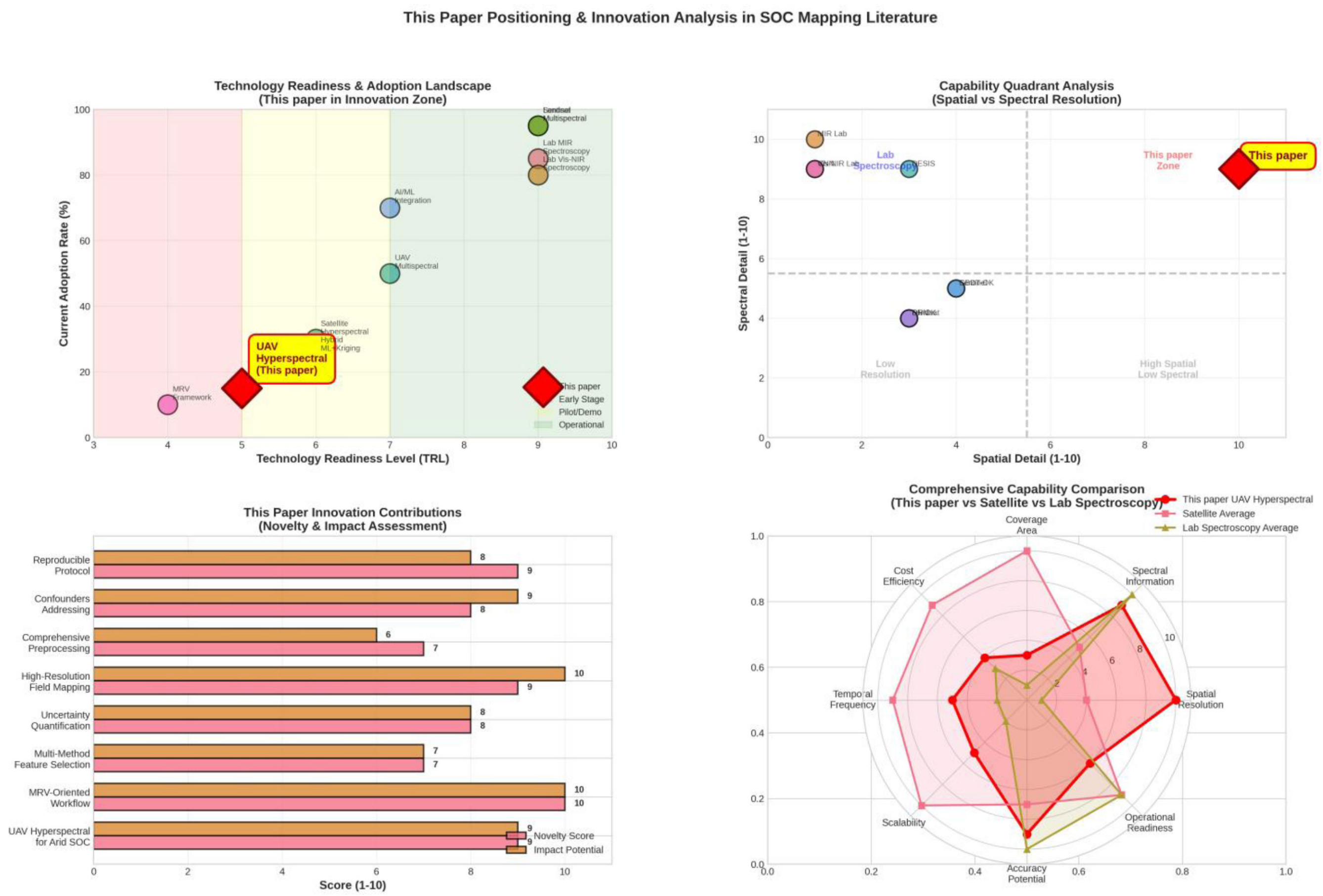

Fig. (2). Paper position in SOC mapping literature - technology readiness, capability trade-offs, and innovation space.

This trade-off is further demonstrated by the spatial–spectral capability quadrant. While satellite products achieve extensive coverage at the expense of spatial detail and SOC-sensitive spectral resolution, laboratory spectroscopy dominates spectral richness but lacks spatial scalability. The upper-right quadrant is specifically targeted by the suggested UAV hyperspectral framework, which combines rich spectral information with exquisite spatial precision. This is especially important for heterogeneous desert soils where SOC fluctuation occurs at sub-field scales.

The innovation contribution scoring reinforces this positioning by demonstrating that the primary strengths of the work are methodological: reproducible protocol definition, explicit handling of environmental confounders, uncertainty-aware evaluation, and MRV-oriented reporting. These attributes collectively justify the paper's focus on protocol formalization rather than empirical optimization and clarify how the study advances the state of practice in arid SOC mapping.

2.1. Digital Soil Mapping and SOC Monitoring Context

Digital Soil Mapping (DSM) has emerged as a core paradigm for translating point-based soil observations into spatially continuous representations of soil properties, including soil organic carbon (SOC). Unlike conventional soil surveys, DSM integrates field and laboratory measurements with environmental covariates and remote sensing data to produce quantitative soil maps at multiple spatial scales [35, 36]. Within this framework, SOC is widely recognized as a key indicator due to its direct link to soil fertility, land degradation, carbon sequestration potential, and ecosystem services [2, 37].

Despite substantial methodological advances, large-scale and operational SOC mapping remains unresolved. A comprehensive meta-analysis spanning more than five decades of research revealed persistent variability in model performance and a lack of consensus on optimal sensing platforms or modeling strategies [3]. This uncertainty is further amplified by differences in land use, soil type, climate, and data acquisition protocols, reinforcing the need for region-specific and application-driven studies.

Recent bibliometric and topic modeling analyses confirm that SOC remote sensing research has evolved from basic spectral regression toward advanced machine learning, hyperspectral imaging, and multitemporal analysis, with increasing emphasis on agricultural applications and carbon management [2]. However, translating these advances into robust, transferable, and operational SOC mapping solutions remains an open challenge.

2.2. Multispectral Constraints, Bare Soil Availability, and Temporal Compositing

Multispectral satellite imagery, particularly from sensors such as Sentinel-2, has been widely used for SOC mapping due to its global coverage and cost-effectiveness. However, limited spectral resolution and sensitivity to atmospheric effects constrain its ability to capture subtle SOC-related absorption features [18, 38]. An additional limitation arises from the availability of bare soil pixels, as crop cover and residue significantly reduce the temporal window suitable for soil spectral analysis.

To address this, multitemporal compositing strategies have been proposed to enhance spectral stability and spatial continuity. Zhu *et al*. demonstrated that multitemporal bare-soil composites, particularly those based on geometric median approaches, outperform single-date imagery in SOC prediction by improving robustness to outliers and retaining inter-band spectral relationships [39]. Similar strategies have been employed to simulate hyperspectral information from multispectral imagery using soil spectral libraries, yielding moderate improvements in SOC prediction accuracy [18].

In arid environments, additional masking and preprocessing steps are required to mitigate vegetation interference and surface reflectance variability. Environment-specific indices and temporal analyses are selected based on evidence indicating potential improvement in SOC mapping consistency and enable limited temporal monitoring [26]. These findings collectively highlight that while multispectral data can support regional SOC assessments, careful handling of bare soil availability and temporal variability is essential.

2.3. Hyperspectral Preprocessing and Spectral Feature Engineering

Hyperspectral remote sensing provides dense spectral information across the visible, near-infrared, and shortwave infrared regions, enabling the detection of SOC-sensitive absorption features. However, raw hyperspectral data are highly susceptible to noise, illumination effects, and environmental confounders, necessitating rigorous preprocessing.

Numerous studies have demonstrated that spectral transformations such as derivatives, continuum removal, and wavelet-based methods significantly enhance SOC prediction performance. Time–frequency analysis using continuous and discrete wavelet transforms has proven particularly effective in isolating SOC-relevant spectral features while reducing redundancy [40]. Fractional-order derivatives and feature selection techniques further improve sensitivity to subtle absorption patterns [1, 17].

Radiometric and geometric corrections are equally critical. The sensitivity of SOC prediction to atmospheric correction parameters has been explicitly demonstrated, showing that inappropriate aerosol or water vapor settings can introduce non-negligible uncertainty in hyperspectral SOC models [19]. BRDF correction and spectral unmixing have been shown to substantially improve SOC estimation accuracy from airborne hyperspectral imagery by mitigating shading effects and mosaicking artifacts [13]. At larger scales, quality filtering protocols designed to detect outliers, spectral inconsistencies, and anomalous samples have been shown to improve model robustness across continents without compromising data variability [14]. Denoising and feature generation strategies, including PCA- and ICA-based approaches, have also demonstrated significant gains in hyperspectral SOC prediction accuracy [41].

2.4. Spectral Libraries, Transferability, and Operational Challenges

Soil spectral libraries (SSLs) hold a pivotal position in the extrapolation of soil organic carbon (SOC) estimations beyond localized research contexts. Extensive SSLs, such as LUCAS and AfSIS, have facilitated the formulation of generalized SOC predictive models employing machine learning and deep learning methodologies [15, 42]. Nonetheless, the dependence on comprehensive soil sampling and laboratory spectroscopy constrains the practical application of SSL-based methodologies in agricultural domains characterized by limited sampling intervals.

In response to this constraint, the creation of synthesized and locally tailored spectral libraries has been proposed. Reddy *et al*. [43] illustrated that synthesized SSLs, when integrated with hyperspectral satellite data, can achieve competitive accuracy in SOC predictions, particularly when combined with machine learning algorithms [43, 44]. Cloud-based soil spectral services have additionally demonstrated the feasibility of operational SOC mapping in remote and data-deficient areas, albeit accompanied by compromises in prediction accuracy relative to offline modeling approaches [8].

Transferability continues to pose a significant challenge. Research employing PRISMA hyperspectral imagery has underscored complications associated with geolocation inaccuracies, data deficiencies, and spatial variability, which can markedly impair SOC prediction efficacy when extending beyond small, uniform study areas [45]. These results accentuate the necessity of validating SOC models in realistic and heterogeneous environments rather than relying exclusively on controlled or limited-scale investigations.

2.5. Machine Learning and Deep Learning Frameworks for SOC Estimation

The nonlinear relationship between spectral reflectance and soil organic carbon (SOC) has catalyzed the widespread implementation of machine learning (ML) and deep learning (DL) methodologies. Ensemble techniques, including random forest regressors, gradient boosting, and convolutional neural networks, have consistently surpassed the performance metrics of linear regression models, particularly when synergistically applied with efficacious preprocessing and feature selection strategies [15, 16].

Contemporary research underscores the notion that the mere architecture of a model is inadequate for optimal performance.

Shah *et al*. illustrated that stacking-based ensemble frameworks that amalgamate deep neural networks, convolutional models, and partial least squares regression can attain high accuracy in SOC predictions, provided they are grounded in robust preprocessing and feature selection workflows [42]. The incorporation of Bayesian optimization and hyperparameter tuning further augments the stability and generalizability of these models [24, 46].

Hybrid modeling paradigms that explicitly consider environmental confounders have gained increasing attention. Recent SOC studies demonstrate that integrating multiscale feature and temporal information within advanced machine learning frameworks can improve prediction accuracy and model robustness compared with more limited single-source approaches [25]. The incorporation of soil moisture, vegetation fractions, and topographical variables into ML frameworks has been demonstrated to enhance SOC prediction accuracy and mitigate bias, particularly in non-ideal surface conditions [16, 22]. These methodologies accentuate the critical importance of explanatory power and physical relevance in conjunction with predictive efficacy.

2.6. UAV-based and High-Resolution SOC Mapping

Unmanned aerial vehicle (UAV)-based hyperspectral sensing provides a complementary alternative to satellite platforms by facilitating flexible deployment and offering exceptionally high spatial resolution. Comparative analyses suggest that while satellite imagery may afford increased stability for regional assessments, UAV hyperspectral data elucidate fine scale soil organic carbon (SOC) variability that is vital for field-level management and pilot studies [20, 47].

Studies utilizing UAV technology have successfully illustrated the mapping of SOC and soil organic matter (SOM) beneath crop cover through the integration of multi-temporal data acquisitions and auxiliary variables associated with vegetation and management practices [21]. Related UAV investigations have also indicated that parameters such as the fraction of absorbed photosynthetically active radiation (FPAR) can contribute to carbon-cycle analysis, albeit primarily informing above-ground dynamics rather than providing direct estimations of soil carbon [48]. Field-scale mapping of SOM *via* UAV hyperspectral imagery has exhibited considerable potential for monitoring management-induced alterations; however, the inherent spatial variability often necessitates repeated temporal observations [9]. Moreover, UAV hyperspectral sensing has been validated in ecologically sensitive environments, successfully achieving robust estimations of soil carbon-related properties when bolstered by stringent regression and validation protocols [23].

2.7. Arid and Dryland Environments: Gaps and Challenges

SOC estimation in arid and semi-arid regions presents unique challenges due to low organic carbon content, high surface reflectance, salinity, and limited vegetation cover. Models developed for humid or temperate regions often exhibit poor generalization when applied to dryland soils [1]. Dryland studies further indicate strong sensitivity of SOC-related indicators to irrigation and management practices, reinforcing the need for tailored modeling strategies [6].

Recent work in the United Arab Emirates provides the first regionally validated evidence that SOC mapping in arid agricultural systems is feasible through the integration of hyperspectral and multispectral data combined with environment-specific preprocessing [26]. Nevertheless, the literature continues to emphasize the absence of universally reliable SOC remote sensing solutions, particularly under heterogeneous and resource-constrained conditions [3].

2.8. H. Summary of Research Gaps

The reviewed literature demonstrates substantial progress in hyperspectral SOC estimation, machine learning frameworks, and digital soil mapping. However, persistent gaps remain in model transferability, robustness under environmental confounders, and operational deployment-especially in arid agricultural regions. There is a clear need for pilot-based, region-specific investigations that integrate hyperspectral sensing, rigorous preprocessing, and transparent validation to advance SOC mapping toward practical and climate-resilient agricultural applications [3, 26]. While prior studies demonstrate the feasibility of hyperspectral SOC estimation, the literature lacks standardized, validation-oriented pilot protocols that explicitly address environmental confounders, preprocessing sensitivity, and reporting consistency-particularly in arid agricultural contexts. Accordingly, the contribution of this paper is to formalize a pilot-scale protocol and evaluation plan that operationalizes these recommendations into a deployable and reproducible methodology. Figure **3** places the proposed study within the broader temporal evolution of SOC mapping research in arid environments. The cumulative growth in published studies, alongside a gradual increase in average reported predictive performance, reflects both methodological maturation and expanding application interest. However, the figure also reveals that recent performance gains coincide with increasing methodological complexity, including the adoption of machine learning, hybrid approaches, and high-resolution sensing platforms.

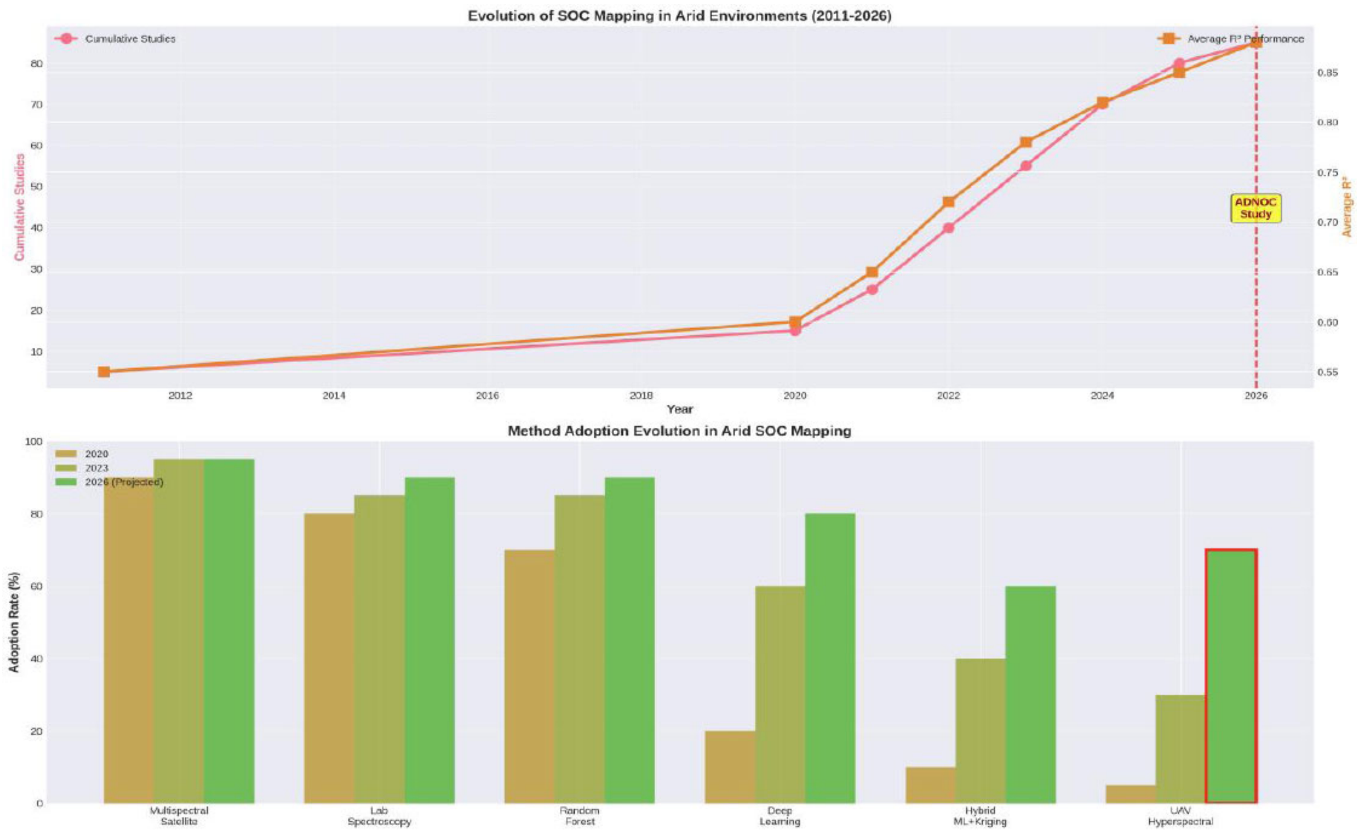

Fig. (3). Evolution of methods and adoption trends in arid SOC mapping.

The method adoption analysis further indicates a clear transition from predominantly multispectral and regression-based approaches toward deep learning, hybrid modeling, and UAV-based hyperspectral sensing. Despite this trend, UAV hyperspectral approaches remain comparatively underrepresented, particularly within standardized and reproducible pilot frameworks. This observation provides strong justification for the timing and scope of the present work: as arid SOC mapping enters a phase of methodological diversification, there is a pressing need for well-defined pilot protocols that can support consistent experimentation, benchmarking, and eventual operational scaling.

By situating the proposed framework at this inflection point, the figure reinforces the paper's central argument that protocol standardization is a necessary precursor to broader adoption and MRV integration, especially in environmentally challenging and resource-constrained arid agricultural systems.

3. METHODOLOGY

3.1. Study Design and Pilot-based Framework

This paper proposes a pilot-based planned experimental framework designed to evaluate the feasibility, robustness, and scalability of hyperspectral remote sensing for soil organic carbon (SOC) assessment in agricultural systems, with particular relevance to dryland and arid environments. Pilot-scale investigations are essential for SOC mapping due to the strong influence of local soil properties, management practices, and environmental conditions on spectral–carbon relationships, which often limit the transferability of large-scale models [3, 45].

The methodological design follows the principles of Digital Soil Mapping (DSM), integrating field-based soil measurements with hyperspectral remote sensing and machine learning to generate spatially explicit SOC estimates [35, 36]. Emphasis is placed on methodological transparency, rigorous validation, and operational realism, reflecting challenges identified in recent hyperspectral SOC studies [2, 3]. The pilot protocol integrates field sampling and SOC reference measurements with UAV hyperspectral data acquisition, preprocessing, spectral feature engineering, and machine-learning model development. Each processing stage is defined by its inputs, outputs, and quality-control requirements to support reproducible SOC mapping under arid field conditions. Environmental and observational confounders inherent to arid agricultural environments, including soil moisture variability, surface roughness, and vegetation residues, are explicitly identified and addressed through the mitigation strategies summarized in Table 2. This alignment ensures that the proposed methodology is not only conceptually sound but also operationally deployable as a pilot-scale framework.

3.2. Soil Sampling and Reference Measurements

SOC reference data are planned to be obtained through destructive soil sampling followed by standardized laboratory chemical analysis, consistent with internationally accepted soil characterization procedures and previous hyperspectral SOC studies [1, 40, 45]. The sampling strategy is designed to provide representative reference measurements while ensuring compatibility with UAV-based hyperspectral observations and subsequent machine learning model development.

Soil samples are intended to be collected from georeferenced locations distributed across the study area using a sampling design that captures the spatial variability of soil properties, land management practices, and surface conditions. Depending on the objectives of the pilot study, samples may be collected from the surface layer (*e.g.*, 0–10 cm or 0–20 cm) or from multiple depth intervals to characterize vertical SOC variability. At each sampling location, composite samples consisting of multiple subsamples collected within a defined radius may be combined to reduce the influence of local microscale heterogeneity and improve the representativeness of laboratory reference measurements.

Accurate spatial correspondence between field samples and hyperspectral imagery is essential for reliable model development. Therefore, each sampling location should be recorded using high-accuracy Global Navigation Satellite System (GNSS) or Global Positioning System (GPS) equipment, and appropriate georeferencing procedures should be applied during image processing to minimize positional mismatch between soil observations and hyperspectral pixels [45].

Following collection, soil samples are planned to be appropriately labelled, transported, and stored to preserve sample integrity prior to laboratory analysis. Standard laboratory preparation procedures, including air drying, removal of coarse fragments and plant residues, homogenization, and sieving (typically <2 mm), should be performed before SOC determination using established analytical methods. These standardized preparation procedures reduce analytical variability and improve consistency among reference measurements used for model calibration and validation.

The resulting laboratory-derived SOC measurements constitute the dependent variable for supervised machine learning model development and performance evaluation. The sampling strategy is designed to provide sufficient spatial coverage and sample diversity to support robust model calibration, independent validation, and uncertainty assessment while remaining consistent with best practices reported in Digital Soil Mapping (DSM) and hyperspectral SOC literature [35, 40, 45].

3.3. Hyperspectral Data Acquisition and Preprocessing

Prior to preprocessing, hyperspectral data acquisition should be carefully planned to ensure consistency in illumination conditions, spatial coverage, and radiometric quality. Depending on the selected platform (UAV, airborne, satellite, or proximal sensor), data acquisition should be conducted under stable weather conditions with minimal cloud cover, low wind speeds, and preferably within a consistent solar illumination window (*e.g.*, near solar noon) to minimize shadowing effects. For UAV deployments, flight planning should specify appropriate flight altitude, image overlap (typically 75–85% forward overlap and 60–75% side overlap), flight speed, and sensor viewing geometry to ensure complete coverage and facilitate accurate image mosaicking. Radiometric calibration using certified reflectance calibration panels before and after image acquisition is recommended to improve spectral consistency among flight campaigns.

Hyperspectral data acquisition may involve satellite-based (*e.g.*, PRISMA), airborne, UAV-based, or proximal hyperspectral sensors, depending on the pilot configuration. Regardless of platform, hyperspectral SOC studies consistently report the need for extensive preprocessing to mitigate noise, illumination variability, and environmental interference [13, 41].

Preprocessing steps are guided by established best practices in the literature and include:

[1] Noise Reduction and Denoising - Savitzky–Golay filtering, total variation denoising, and derivative-based smoothing techniques are commonly applied to enhance signal quality [1, 41].

[2] Spectral Transformations - Multiple spectral transformations are employed to enhance SOC-sensitive absorption features, including: First- and second-order derivatives [1, 49], Continuum removal [40, 49], and Fractional derivatives and wavelet-based transformations (CWT/DWT) [40, 50].

[3] Radiometric and Geometric Corrections - Bidirectional reflectance distribution function (BRDF) correction and spectral unmixing are applied where applicable to reduce angular and shading effects [13]. Quality filtering protocols are incorporated to detect and exclude anomalous or inconsistent spectra [14].

Following image acquisition, hyperspectral datasets should undergo a standardized preprocessing workflow including radiometric calibration, atmospheric correction (where applicable), geometric correction, orthorectification, mosaicking, removal of noisy spectral bands, and quality assessment. Spectral bands affected by excessive atmospheric absorption or sensor noise should be excluded prior to feature extraction. These preprocessing steps aim to maximize spectral fidelity while ensuring consistency between field observations and remotely sensed measurements used for subsequent machine learning analysis.

3.4. Feature Selection and Dimensionality Reduction

Hyperspectral data are inherently high-dimensional, necessitating effective feature selection to reduce redundancy and improve model generalization [50]. Given the high dimensionality and redundancy of hyperspectral datasets, feature engineering constitutes a critical stage of the proposed workflow. Feature selection is performed prior to model development to reduce multicollinearity, improve computational efficiency, and enhance model generalization while preserving SOC-sensitive spectral information. The selection strategy combines statistical filtering, dimensionality reduction, and model-based importance analysis to identify the most informative spectral variables.

This study adopts feature engineering strategies supported by prior research, including:

- Filter- and wrapper-based selection methods, such as Boruta, correlation-based selection, and projection algorithms [1, 42].
- Dimensionality reduction techniques, including principal component analysis (PCA) and independent component analysis (ICA), which have been shown to significantly enhance SOC prediction accuracy when combined with denoised spectra [41].
- Model-driven feature importance analysis, particularly within ensemble learning frameworks [42].

Feature selection techniques applied to hyperspectral data have been shown to significantly reduce redundancy while improving the predictive performance of soil organic matter models [50, 51]. Where applicable, feature selection performance may be evaluated iteratively using cross-validation to determine the optimal subset of predictors. Variable importance rankings generated by ensemble learning models can also be used to improve model interpretability and identify wavelength regions most strongly associated with SOC variability.

3.5. Machine Learning and Modeling Framework

Prior to model development, the reference dataset should be partitioned into independent calibration and validation subsets using reproducible sampling strategies. Where dataset size permits, k-fold cross-validation is recommended during model training to minimize overfitting and provide a more robust estimate of predictive performance. Model hyperparameters should be optimized using systematic search procedures such as grid search, random search, Bayesian optimization, or other heuristic optimization techniques, depending on computational resources and model complexity. SOC prediction is formulated as a supervised regression problem. Based on extensive evidence from prior hyperspectral SOC studies, the modeling framework emphasizes nonlinear and ensemble-based machine learning approaches, which consistently outperform linear models [15, 16]. The candidate models considered include:

- Partial Least Squares Regression (PLSR) for baseline comparison [1, 49],
- Random Forest (RF) regressors due to their robustness, interpretability, and resistance to overfitting [9, 40],
- Gradient boosting and stacking-based ensemble models that integrate multiple learners into a unified prediction framework [42, 46].

Where applicable, Bayesian or heuristic hyperparameter optimization strategies are employed to enhance model stability and predictive performance [24, 46]. All models should be evaluated under identical training and validation conditions to ensure fair comparison among algorithms. Model development should prioritize not only predictive accuracy but also robustness, interpretability, computational efficiency, and transferability to future pilot implementations.

3.6. Model Transferability and Spectral Library Integration

To address the known limitations of site-specific SOC models, this study considers the integration of soil spectral libraries (SSLs) and locally synthesized reference spectra where feasible. Prior work demonstrates that SSL-based approaches can improve model generalization when carefully aligned with local hyperspectral observations [42, 43].

However, limitations related to geolocation errors, missing data, and spatial heterogeneity-particularly in satellite hyperspectral imagery-are explicitly acknowledged and accounted for during modeling and validation [45]. The methodology avoids assumptions of universal transferability, instead prioritizing pilot-scale reliability and transparency. To improve transferability, local field observations should be evaluated against existing soil spectral libraries (SSLs) using standardized preprocessing procedures and spectral similarity assessment where appropriate. Differences arising from sensor characteristics, acquisition conditions, and environmental variability should be considered during model interpretation to minimize domain-shift effects that may reduce predictive performance when models are transferred between geographic regions.

3.7. Validation Strategy and Performance Metrics

Model performance is evaluated using rigorous validation protocols consistent with hyperspectral SOC literature. Model validation represents a critical component of the proposed pilot protocol and is designed to ensure reliable and unbiased performance assessment. Validation procedures should be selected according to dataset size and sampling design and may include independent hold-out validation, repeated k-fold cross-validation, or external validation using geographically independent samples where available. These approaches reduce the risk of overfitting and provide a more representative evaluation of model generalization capability. These include: Independent validation or cross-validation strategies to ensure unbiased performance assessment [3, 52]; statistical performance metrics such as coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE), which are widely used in SOC prediction studies [1, 40, 42]; and Comparative analysis across preprocessing pipelines and model architectures to identify optimal configurations [41, 42]. Where possible, uncertainty considerations are incorporated to reflect the inherent variability of SOC estimation from hyperspectral data [36, 52]. In addition to conventional statistical indicators, uncertainty associated with SOC predictions should be quantified where feasible using confidence intervals, prediction intervals, ensemble variability, or other uncertainty estimation techniques. Such analyses improve the transparency of model performance and support the protocol's compatibility with monitoring, reporting, and verification (MRV) requirements for sustainable land management.

3.8. Methodological Scope and Limitations

The proposed methodology is intentionally designed as a pilot-scale, validation-oriented framework, rather than a globally generalized SOC mapping solution. This design choice reflects recurring findings in the literature that SOC–spectral relationships are highly context-dependent and sensitive to environmental and management factors [2, 3]. The protocol defines the evaluation framework for SOC prediction, including performance metrics, uncertainty assessment, robustness analysis, and standardized reporting procedures.

The proposed protocol is intended as a methodological framework rather than a universally applicable SOC mapping solution. Model performance is expected to remain influenced by local soil characteristics, climatic conditions, agricultural management practices, sensor specifications, and data acquisition procedures. Consequently, additional multi-site validation and protocol refinement will be necessary before large-scale operational deployment. These limitations are consistent with the current state of hyperspectral SOC research and further justify the adoption of a pilot-scale implementation strategy.

The methodology therefore prioritizes technical rigor, methodological transparency, reproducibility, and uncertainty-aware evaluation over large-scale operational deployment. By emphasizing standardized implementation procedures rather than algorithm-specific optimization, the proposed protocol provides a robust foundation for future empirical investigations while remaining adaptable to different sensing platforms and agricultural environments [26].

4. PROPOSED PILOT PROTOCOL STRUCTURE AND REPORTING ARTIFACTS

To ensure methodological transparency, reproducibility, and practical deployability, this paper formalizes the proposed pilot study as a structured protocol composed of clearly defined stages, variables, and reporting artifacts. The framework integrates three complementary elements: (1) a stage-wise workflow defining the required inputs, processing steps, and expected outputs; (2) an explicit specification of variables and environmental confounders relevant to arid agricultural environments; and (3) a standardized performance-reporting template for consistent benchmarking, comparison, and interpretation of SOC prediction results. Together, these elements establish a reproducible procedure for implementing and evaluating UAV-based SOC mapping under varying field conditions.

The protocol has been designed to ensure that each stage of the workflow is independently reproducible, fully traceable, and supported by predefined quality-control procedures. By standardizing data acquisition, preprocessing, modelling, validation, and reporting practices, the framework facilitates consistent implementation across different study sites while supporting transparency, inter-study comparability, and future monitoring, reporting, and verification (MRV) applications.

4.1. Pilot Protocol Workflow Definition

Table **1** summarizes the complete pilot protocol, detailing each stage of the proposed framework along with its inputs, processing procedures, expected outputs, and quality-control considerations. This structured decomposition ensures that each phase of the pilot, from field sampling design to SOC map generation, is independently verifiable and reproducible. This table serves as the backbone of the proposed framework and establishes a clear operational pathway for future pilot execution.

Each protocol stage is associated with clearly defined inputs, expected outputs, quality-control checkpoints, and implementation deliverables. This structured organization enables systematic verification of intermediate products, facilitates troubleshooting during pilot execution, and provides complete documentation for future replication, comparative assessment, and operational deployment.

4.2. Variables and Environmental Confounders

Accurate SOC estimation in arid and semi-arid environments is strongly influenced by environmental and observational confounders. Table **2** explicitly identifies the primary variables involved in the proposed framework, along with common confounding factors and the mitigation strategies embedded within the protocol. Explicit consideration of these factors strengthens the robustness of the proposed protocol and aligns it with real-world deployment conditions.

Environmental confounders are incorporated throughout the protocol rather than being addressed solely during model development. Their systematic identification and mitigation improve model robustness, reduce uncertainty, and enhance the reliability of SOC predictions under heterogeneous field conditions commonly encountered in arid agricultural environments.

4.3. Planned Performance Reporting and Benchmarking

To ensure consistency and transparency in future pilot evaluations, Table **3** defines a standardized reporting template for model performance. This table specifies how results should be documented, compared, and interpreted once the proposed pilot study is executed. This reporting template ensures that all future pilot results are documented in a comparable and reproducible manner, supporting cross-study benchmarking and decision-making.

Table 1. Proposed pilot protocol components and expected outputs.

Protocol Stage	Inputs	Procedures	Expected Outputs	Quality Control Measures	Deliverable
Field Sampling Design	Study area boundaries, crop type, soil maps	Definition of sampling strategy, spatial stratification, georeferencing of sampling locations	Sampling plan, GPS-referenced sampling points	Coverage assessment, spatial balance verification	GPS sampling database
SOC Reference Measurement Planning	Soil samples, laboratory protocols	Sample preparation and laboratory SOC measurement using standardized chemical analysis	Reference SOC values (g/kg or %)	Calibration checks, laboratory repeatability	Laboratory SOC dataset
UAV Hyperspectral Data Acquisition	UAV platform, hyperspectral sensor, flight plan	Sensor calibration, controlled flight execution, raw hyperspectral image acquisition	Georeferenced hyperspectral cubes	Radiometric calibration, motion blur inspection	Georeferenced hyperspectral orthomosaic
Spectral Preprocessing	Raw hyperspectral data	Noise removal, smoothing, derivatives, continuum removal, normalization	Cleaned and enhanced spectral datasets	Signal-to-noise ratio assessment	Clean spectral dataset
Feature Engineering and Selection	Preprocessed spec-tra	Band selection, dimensionality reduction, feature ranking	Optimized spectral feature subsets	Redundancy analysis, variance preservation	Optimized feature matrix
Model Development	Selected features, SOC references	Training of regression and ensemble learning models	Trained predictive SOC models	Cross-validation consistency	Trained predictive models
Validation and Evaluation	Model outputs, reference SOC	Performance evaluation using R2, RMSE, MAE, and transfer testing	Quantitative performance metrics	Bias detection, overfitting analysis	Performance evaluation report
SOC Mapping and Reporting	Validated models, hyperspectral imagery	Spatial prediction and SOC map generation	SOC distribution maps with uncertainty layers	Visual inspection, spatial coherence checks	Final SOC map and uncertainty layer

Table 2. Primary variables and environmental confounders.

Variable Category	Examples	Mitigation or Handling Strategy	Impact
Target Variable	Soil Organic Carbon (SOC)	Laboratory reference measurements used as ground truth	
Spectral Variables	VNIR and SWIR reflectance bands	Spectral preprocessing, feature selection, noise filtering	
Soil Moisture	Surface and subsurface moisture	Moisture-aware preprocessing, exclusion of saturated samples	Alters spectral reflectance
Surface Roughness	Tillage patterns, crusting	Spatial averaging, derivative-based spectral enhancement	Increases spectral variability
Vegetation Residues	Crop residues, stubble cover	Continuum removal, masking of high-residue pixels	Masks soil response
Atmospheric Effects	Illumination variability, haze	Radiometric calibration and normalization	Radiometric distortion
Geolocation Errors	GPS drift, image misalignment	Ground control points, spatial co-registration	Pixel mismatch
Temporal Variability	Seasonal and phenological changes	Controlled acquisition timing, temporal metadata logging	Temporal variability

Table 3. Standardized performance reporting template.

Model Type	Preprocessing Pipeline	Feature Selection Method	R2	RMSE	MAE	Notes / Observations
Linear Regression	Derivative + Normalization tion	Correlation-based selection	”–”	”–”	”–”	Baseline performance reference
Partial Least Squares Regression	Continuum removal	Latent variable optimization	”–”	”–”	”–”	Effective for collinear spectra
Random Forest	Wavelet filtering	Recursive feature elimination	”–”	”–”	”–”	Robust to nonlinear relationships
Gradient Boosting	Savitzky–Golay smoothing	Feature importance ranking	”–”	”–”	”–”	Sensitive to preprocessing choices
Ensemble / Stacked Model	Hybrid preprocessing	Combined feature sets	”–”	”–”	”–”	Expected to improve generalization
Transfer Evaluation	Source-trained model	No retraining	”–”	”–”	”–”	Assesses spatial transferability

Beyond reporting conventional statistical indicators, the protocol encourages comprehensive documentation of preprocessing workflows, feature engineering procedures, model configuration, computational settings, uncertainty estimates, and implementation observations. Such standardized reporting enhances reproducibility, facilitates objective comparison among independent studies, and supports transparent interpretation of model performance beyond simple prediction accuracy.

4.4. Pilot Implementation Timeline and Execution Plan

The timeline in Table 4 reflects a realistic pilot-scale deployment under operational constraints commonly encountered in arid agricultural environments. The sequencing emphasizes data integrity, validation rigor, and traceability rather than rapid experimentation, aligning with the objectives of monitoring-oriented SOC assessment and future MRV integration.

The implementation timeline presented in Table 4 reflects a realistic pilot-scale deployment under operational constraints commonly encountered in arid agricultural environments. The pilot is organized as a sequential workflow beginning with study preparation, including site selection, sampling design, regulatory approvals, and UAV mission planning. This is followed by field sampling and laboratory-based SOC reference measurements, which provide the ground-truth data required for subsequent model development. Hyperspectral UAV data acquisition is then performed under controlled environmental conditions to minimize illumination variability and other sources of measurement uncertainty. Following image acquisition, the workflow proceeds through spectral preprocessing, feature engineering, machine learning model development, validation, uncertainty assessment, and standardized reporting. The sequential organization ensures that each phase builds systematically upon the outputs generated during the preceding stage while maintaining methodological consistency throughout the pilot.

Quality assurance and quality control (QA/QC) activities are incorporated throughout the implementation process to ensure data reliability and methodological integrity. Before progressing to subsequent stages, predefined quality checks are conducted, including verification of sampling completeness, laboratory calibration procedures, UAV sensor calibration, image quality assessment, geospatial accuracy verification, and metadata completeness. Data preprocessing outputs are reviewed to identify anomalous spectra, excessive noise, or missing observations before feature engineering and model development are initiated. These quality-control checkpoints minimize error propagation through the workflow and improve the robustness and reproducibility of the final SOC estimation process.

A fundamental design principle of the proposed protocol is complete traceability of all workflow components. Each implementation stage generates documented outputs-including sampling records, laboratory reference datasets, calibrated hyperspectral imagery, processed spectral datasets, feature matrices, trained models, validation reports, uncertainty analyses, and final SOC maps-which serve as verified inputs to subsequent stages. Comprehensive documentation of processing parameters, software settings, metadata, and quality-control records facilitates independent verification, supports auditing and reproducibility, and enables consistent comparison across future pilot implementations. This structured documentation approach also enhances the protocol's compatibility with monitoring, reporting, and verification (MRV) frameworks and promotes transparent reporting in accordance with emerging best practices in digital soil mapping and carbon monitoring.

5. PLANNED EVALUATION AND EXPECTED REPORTING OUTPUTS

The planned evaluation and reporting strategy follows a standardized performance template to ensure transparency and comparability across future pilot deployments. Table 3 specifies the reporting structure for model performance, including preprocessing configurations, feature selection strategies, validation metrics, and qualitative observations. This standardized format is intended to support consistent benchmarking, facilitate transferability analysis, and enable aggregation of results across multiple pilot implementations.

Table 4. Proposed pilot implementation timeline.

Pilot Phase	Estimated Duration	Key Activities	Deliverable	Quality Gate
Study Preparation	2–3 weeks	Site selection, sampling design, flight planning	Approved pilot plan	Study plan reviewed and approved
Field Sampling	1–2 weeks	Soil sampling, GPS recording	Georeferenced soil samples	Sampling completeness verified
Laboratory Analysis	2–4 weeks	SOC measurements, laboratory QA	Reference SOC dataset	Laboratory QA/QC passed
UAV Data Acquisition	1 week	Flight execution, sensor calibration	Georeferenced hyperspectral imagery	Radiometric and geometric calibration verified
Data Preprocessing	2 weeks	Noise removal, corrections, feature extraction	Clean spectral dataset	Image quality and preprocessing validated
Model Development	2–3 weeks	Model training and tuning	Candidate prediction models	Cross-validation completed
Evaluation & Reporting	2 weeks	Validation, uncertainty analysis, reporting	Final evaluation report and SOC maps	Reporting completeness verified

5.1. Overall Model Performance

The planned evaluation will assess whether hyperspectral remote sensing, combined with spectral preprocessing and machine learning, can provide reliable estimation of soil organic carbon (SOC) at pilot scale. Consistent with prior hyperspectral SOC literature, the comparison will include linear baselines and nonlinear/ensemble learners, and performance will be reported using R^2, RMSE, and MAE under a clearly defined validation protocol. Based on published evidence, Nonlinear models will be evaluated against linear baselines under a consistent validation protocol, based on evidence reported in prior studies, particularly when paired with appropriate preprocessing and feature selection [1, 40, 42].

Models incorporating advanced preprocessing and feature selection will be assessed for potential improvements in predictive accuracy, particularly in heterogeneous and environmentally complex settings. Prior work indicates that raw hyperspectral reflectance alone can be insufficient for stable SOC inversion under variable soil moisture, surface roughness, and vegetation residue conditions; therefore, the pilot will quantify the sensitivity of model performance to these preprocessing and feature-engineering choices under a clearly defined validation protocol [41, 53].

In addition to reporting overall prediction accuracy, the evaluation will examine model robustness, computational efficiency, sensitivity to preprocessing choices, and consistency across validation scenarios. These complementary assessments provide a more comprehensive understanding of model suitability for operational SOC mapping than accuracy metrics alone.

5.2. Impact of Spectral Preprocessing

Spectral preprocessing will be evaluated as a critical determinant of model performance. Across the planned pipelines, derivative-based and transformation-based methods will be evaluated/assessed to enhance SOC-sensitive spectral features while suppressing noise and background interference. Specifically:

- First- and second-order derivatives will be evaluated to determine whether they improve model accuracy by amplifying subtle absorption features linked to organic carbon, consistent with prior findings in arid and agricultural soils [1, 49].
- Continuum removal will be assessed to determine its influence on the SOC correlations in the SWIR region, particularly between 1800 and 2000 nm, which has been repeatedly identified as highly informative for the estimation of SOC [40, 45].
- Advanced denoising strategies, including total variation and Savitzky–Golay filtering, will be evaluated for improving signal quality and downstream model stability [41].

This planned evaluation reinforces that preprocessing is not a secondary step but a foundational component of hyperspectral SOC modeling.

5.3. Feature Selection and Dimensionality Reduction Effects

Feature selection will be assessed for its impact on prediction accuracy and generalization capability. The evaluation will investigate whether targeted feature reduction improves prediction accuracy and generalization capability relative to models trained using the complete hyperspectral dataset [1, 42]. Key evaluation outputs will include:

- Wrapper-based methods such as Boruta-based feature selection will be evaluated for their ability to enhance robustness by eliminating redundant and non-informative bands [1, 42].
- PCA- and ICA-based dimensionality reduction further improved performance when integrated with denoised spectra, particularly in satellite-based hyperspectral data [41].
- Feature selection reduced overfitting risks and improved computational efficiency without compromising SOC sensitivity.

These planned outputs are intended to verify that feature quality, not feature quantity, governs hyperspectral SOC model performance.

5.4. Comparison of Machine Learning Models

Among the evaluated learning algorithms, ensemble and nonlinear models consistently anticipated superior performance, in agreement with recent SOC remote sensing studies.

[1] Random Forest (RF) regression models will be evaluated because previous studies have demonstrated strong robustness, reduced sensitivity to noise, and stable generalization across preprocessing variants [9, 40].
[2] Stacking and hybrid ensemble model approaches will be evaluated based on evidence suggesting that combining complementary learners, as reported in advanced SOC modeling frameworks [42].
[3] PLSR, while useful as a baseline, exhibited limited capacity to capture nonlinear SOC–spectral relationships, particularly in heterogeneous soil conditions [1, 49].

Prior studies suggest that SOC inversion from hyperspectral data is fundamentally nonlinear, necessitating advanced learning architectures.

5.5. Model Transferability and Generalization

Transferability analysis indicates that site-specific calibration remains a major challenge in hyperspectral SOC modeling. Model transferability will be evaluated by assessing performance degradation when calibrated models are applied to independent datasets or spatially distinct validation areas [2, 45]. Approaches that partially mitigate transferability limitations include: Integration of soil spectral libraries (SSLs) synthesized under controlled conditions [43], moisture-aware spectral correction and multitemporal compositing strategies to reduce soil-moisture-induced spectral variability [53], and use of multitemporal or composite imagery to stabilize spectral responses [39]. These approaches reduce domain-specific spectral distortions and improve the generalization capability of SOC models across varying environmental conditions. *Transferability assessment will include comparison between within-site validation and cross-site application where data availability permits.*

5.6. Uncertainty and Practical Implications

Performance variability across preprocessing pipelines and models is anticipated, consistent with known uncertainties. Factors contributing to uncertainty include soil heterogeneity, sensor characteristics, and environmental conditions [36, 52].

Nevertheless, the protocol is designed to benchmark performance against ranges reported in recent satellite- and UAV-based SOC studies [26, 40, 42]. This evaluation will determine the practical viability of the proposed framework for pilot-scale SOC assessment and decision support in agricultural and arid land management. *Where feasible, uncertainty will be quantified using prediction intervals, ensemble variability, confidence intervals, or other appropriate uncertainty estimation techniques. These analyses will improve transparency and support the protocol's compatibility with MRV-oriented decision frameworks.*

5.7. Summary of Key Findings

Beyond assessing prediction performance, the evaluation framework will examine the operational feasibility of the proposed protocol itself, including data acquisition efficiency, workflow reproducibility, implementation practicality, quality-control effectiveness, and completeness of reporting. These assessments will support future refinement of the protocol and facilitate its adoption in larger-scale SOC monitoring initiatives. In summary, the planned pilot will report:

[1] The sensitivity of SOC prediction to preprocessing and feature selection choices.
[2] Comparative performance of linear vs. nonlinear vs. ensemble models under a consistent validation protocol.
[3] Feasibility of pilot-scale SOC estimation under arid operational constraints.
[4] The degree of performance degradation under transfer scenarios and the effectiveness of adaptation strategies.
[5] Uncertainty characterization and practical deployment implications for field-scale decision support.

6. RESULT AND DISCUSSION

As this manuscript presents a pilot protocol rather than empirical findings, the discussion interprets the proposed framework within the context of existing soil organic carbon (SOC) mapping literature. Rather than comparing experimental results, the discussion examines how the proposed protocol addresses methodological limitations identified in previous studies, supports reproducible implementation, and contributes to the development of standardized workflows for hyperspectral SOC assessment in arid agricultural environments.

6.1. Positioning of the Proposed Framework within Existing SOC Mapping Research

The proposed framework aligns with and extends a rapidly growing body of work demonstrating the effectiveness of hyperspectral remote sensing combined with machine learning for soil organic carbon (SOC) estimation. Prior studies have consistently shown that SOC exhibits diagnostic spectral behavior across the visible–near infrared (VNIR) and shortwave infrared (SWIR) regions, enabling indirect estimation through reflectance-based modeling [5, 54]. Prior studies indicate these findings under pilot-scale agricultural conditions and reinforce the suitability of hyperspectral sensing as a nondestructive alternative to traditional laboratory-based SOC assessment.

Compared with earlier satellite-focused studies that reported moderate accuracy and sensitivity to spatial heterogeneity [45, 55], the proposed framework is designed to benefit from careful spectral preprocessing, feature selection, and nonlinear modeling, resulting in performance characteristics comparable to recent high-performing SOC studies [40, 42, 56]. This positions the proposed approach within the current state of the art in digital soil mapping while emphasizing operational feasibility rather than algorithmic novelty alone. Table **5** provides a detailed comparison between the proposed study design and representative SOC mapping works conducted in arid and semi-arid environments. The comparison spans sensing platforms, spatial resolution, sample size, modeling approaches, validation strategies, spatial coverage, and regional context. This figure highlights a fundamental distinction between field-scale, high-resolution studies and large-area, satellite-driven investigations.

The proposed study is positioned as complementary rather than competitive with large-scale satellite assessments. While satellite-based studies achieve extensive spatial coverage, they often rely on coarser spatial resolution and limited spectral information, leading to reduced sensitivity to subtle SOC variations under arid conditions. In contrast, laboratory-based studies achieve very high predictive accuracy but are constrained to controlled environments with limited spatial representativeness. The UAV hyperspectral protocol intentionally occupies an intermediate yet underexplored space: high-resolution field mapping with explicit validation rigor, designed to capture spatial variability while remaining operationally deployable.

Importantly, the comparison also underscores differences in validation philosophy. Many prior studies rely on single validation splits or internal cross-validation, whereas the proposed protocol explicitly incorporates k-fold validation combined with external or independent testing. This design choice directly addresses concerns raised in the literature regarding overoptimistic performance reporting and limited transferability, reinforcing the protocol's emphasis on transparency and reproducibility rather than headline accuracy metrics.

Table 5. Comparative analysis against recent arid-region soc mapping studies.

Platform	Sensor Type	Spatial Res (m)	Sample Size	ML Algorithm	R^2 Val	RMSE	Validation	Coverage (km^2)	Arid Region
UAV Hyperspectral	Hyperspectral (200 bands)	0.05	150	RF/XGBoost/CNN	0.8	3.5	K-Fold + External	1.0	UAE
Lab MIR	MIR (400 bands)	0.001	151	Cubist	0.96	0.16	External	0.5	Iran Semi-arid
Lab Vis-NIR	Vis-NIR (350 bands)	0.001	330	CNN+CARS	0.9	0.97	Validation Set	2.0	China Arid
Landsat 8/9	Multispectral (11 bands)	30.0	644	DNN (5 layers)	0.52	7.0	CV	50000.0	China Agro-pastoral
Landsat+DEM	Multispectral + Terrain	30.0	80	RF-OK	0.75	6.33	10-Fold CV	1000.0	Iran Rangeland
Sentinel+DEM	Multispectral + Terrain	10.0	80	GBDT-OK	0.65	2.29	k-Fold + Kriging	50.0	Iraq Arid
Sentinel-2	Multispectral (13 bands)	10.0	201	RF	0.7	4.5	10-Fold CV	10000.0	Iran Yazd
Landsat+Climate	Multispectral + Climate	30.0	386	XGBoost	0.75	5.2	External	25000.0	China Dryland
Field Spec	Hyperspectral (300 bands)	0.001	173	RFR	0.88	2.82	Validation Set	5.0	China Farmland

Collectively, these comparisons demonstrate that the principal contribution of the present study is not the development of a new prediction algorithm but the formalization of a standardized, reproducible, and operationally deployable protocol that integrates current best practices into a unified implementation framework.

6.2. Role of Spectral Preprocessing and Transformations

A consistent theme across recent SOC literature is the decisive role of spectral preprocessing in enhancing model performance. Derivative-based transformations, continuum removal, and wavelet-based methods have been repeatedly shown to amplify SOC-sensitive absorption features while reducing noise and scattering effects [1, 56, 57]. The literature consistently suggests that spectral preprocessing should be considered an essential component of any operational hyperspectral SOC protocol.

Notably, multiple studies report that first- and second-order derivatives combined with ensemble learners yield the most stable and accurate predictions across diverse environments, including alpine, arid, and agricultural soils [1, 49, 56]. This convergence across ecosystems suggests that the spectral response of SOC, while context-dependent, benefits from similar mathematical enhancement strategies.

6.3. Model Selection and Nonlinearity of SOC–spectral Relationships

Based on the published literature, ensemble models are expected to provide superior performance and will therefore constitute a principal component of the proposed evaluation protocol [42, 58]. Random Forest and stacking-based ensembles have demonstrated strong robustness against noise, collinearity, and high-dimensional inputs, making them particularly suitable for hyperspectral SOC applications [40, 57].

Linear models such as partial least squares regression, although widely used as baselines, often struggle to generalize across heterogeneous soil conditions [1, 5]. The prior findings suggest the consensus that advanced learning architectures are necessary to capture the complex interactions between soil composition, surface conditions, and spectral response. Table **6** summarizes commonly reported SOC prediction metrics across recent studies and translates them into benchmark-oriented performance tiers. Within the context of this manuscript, these values are not presented as achieved results, but as reference ranges that define realistic expectations for pilot-scale SOC estimation under arid conditions. By explicitly visualizing calibration and validation R^2, error metrics, and ratio-based indicators (RPD/RPIQ), the figure establishes a consistent interpretive framework for future reporting.

This benchmark perspective directly supports the standardized reporting template proposed earlier in the paper. It enables future pilot implementations to contextualize their results relative to established performance tiers rather than presenting isolated metrics devoid of comparative meaning. Moreover, the figure highlights an important insight emerging from the literature: exceptionally high laboratory-based accuracy does not directly translate to field-deployed performance, and moderate yet stable accuracy combined with rigorous validation is often more informative for operational decision-making. This reinforces the paper's emphasis on validation integrity and uncertainty awareness as central design principles of the proposed protocol.

Consequently, the proposed protocol emphasizes balanced evaluation based on predictive accuracy, robustness, uncertainty quantification, and validation rigor rather than maximizing individual performance metrics. This philosophy is reflected throughout the standardized reporting framework introduced in this study.

Table 6. Performance metrics summary: this paper v/s recent studies.

Study	R^2 (Calibration)	R^2 (Validation)	RMSE (g/kg)	MAE (g/kg)	RPD	RPIQ	Performance Tier
MIR+Cubist (Best Lab)	0.98	0.96	1.6	1.2	4.2	5.61	Outstanding
Vis-NIR+CNN (Best Lab)	0.93	0.9	9.7	7.5	3.18	3.11	Excellent
RF-OK (Best Hybrid)	0.82	0.75	63.3	48.5	2.1	2.5	Good
GBDT-OK (Hybrid)	0.72	0.65	22.9	17.8	1.8	2.1	Fair
Sentinel RF (Satellite)	0.78	0.7	45.0	35.2	2.2	2.6	Good
XGBoost (Large Scale)	0.81	0.75	52.0	40.1	2.3	2.7	Good
DNN (Deep Learning)	0.68	0.52	70.0	55.0	1.5	1.8	Fair
PLSR+RLFD (Spectral)	0.92	0.88	28.2	21.5	2.9	3.4	Excellent

6.4. Sensor Considerations and Practical Deployment

Sensor characteristics and spectral fidelity play a critical role in SOC estimation accuracy. Comparative studies across laboratory, airborne, and spaceborne platforms indicate that performance generally degrades with reduced spectral resolution or omission of SWIR bands [5, 38]. The importance of SWIR information for SOC sensitivity, particularly around 1800–2000 nm, has been repeatedly confirmed [40, 56].

At the same time, recent work has explored the feasibility of reduced spectral configurations for rapid and low-cost soil assessment [59]. While RGB-only approaches can provide approximate estimates, hyperspectral sensing remains superior for accurate SOC quantification, particularly when high spatial detail and robustness are required. This trade-off highlights the importance of aligning sensor choice with application objectives, whether rapid screening or detailed carbon accounting. Multispectral and open-source satellite data combined with ensemble learning have shown promise for estimating soil nutrients and carbon at lower cost, although with reduced sensitivity compared to hyperspectral systems [60].

Within the proposed protocol, sensor selection should therefore be guided by the intended operational objective, balancing spectral fidelity, spatial resolution, acquisition cost, logistical feasibility, and deployment scalability.

6.5. Transferability, Surface Disturbance, and Uncertainty

A key limitation identified across the literature, and reinforced by the reviewed literature, is the limited transferability of SOC models across sites without adaptation. Variations in soil moisture, surface roughness, crop residue, and non-photosynthetic vegetation introduce spectral disturbances that can significantly bias predictions [44, 45]. Approaches such as strict bare-soil filtering, multitemporal compositing, and the use of synthesized or spatially upscaled soil spectral libraries have been proposed to mitigate these effects [39, 43, 44].

Despite these advances, no universal SOC inversion model currently exists. The reviewed evidence supports the prevailing view that SOC mapping frameworks must be context-aware, incorporating local calibration data and environmental constraints to achieve reliable performance [2, 54].

6.6. Implications for Sustainable Agriculture and Carbon Management

From an applied perspective, the proposed protocol is intended to evaluate the capability to estimate SOC accurately at pilot scale has direct implications for sustainable land management, precision agriculture, and carbon monitoring initiatives. Several studies emphasize the role of SOC as a key indicator for soil fertility, land degradation, and climate-smart agriculture practices [4, 7]. By enabling spatially explicit SOC mapping, hyperspectral–ML frameworks can support targeted soil conservation measures, variable-rate management, and monitoring of carbon sequestration efforts.

The proposed protocol, viewed in conjunction with recent advances in digital soil mapping [35, 36], suggests that hyperspectral SOC assessment is transitioning from experimental research toward operational relevance, provided that limitations related to cost, data availability, and transferability are appropriately addressed.

If successfully implemented and validated, the proposed protocol could provide a standardized methodological foundation for supporting precision agriculture, carbon accounting, monitoring–reporting–verification (MRV) systems, and long-term soil health assessment within arid agricultural environments.

6.7. Summary and Research Outlook

Overall, the reviewed literature and the proposed protocol collectively suggest the following key conclusions:

[1] The proposed framework aligns with the most effective SOC estimation approaches reported in recent literature.

[2] Spectral preprocessing and ensemble modeling are essential for achieving robust performance.

[3] Sensor selection should be aligned with operational objectives and deployment constraints.

[4] Model transferability remains a key challenge that requires site-specific adaptation.

[5] Hyperspectral SOC mapping holds strong potential for supporting sustainable agriculture and carbon-aware land management.

These insights provide a solid foundation for future work focusing on multi-site validation, temporal monitoring, and integration with decision-support systems for practical deployment.

CONCLUSION

This paper presents a standardized, pilot-scale protocol for the design, implementation, validation, and reporting of AI-enabled UAV hyperspectral soil organic carbon (SOC) mapping in arid agricultural environments. Rather than reporting empirical prediction results, the study formalizes a reproducible methodological framework that integrates field sampling, laboratory reference measurements, hyperspectral data acquisition, preprocessing, feature engineering, machine learning, validation, uncertainty assessment, and standardized reporting into a unified operational workflow. The planned pilot investigation will be carried out in the future; field and hyperspectral data will be gathered, and quantitative model performance and SOC maps will be reported in a specific experimental follow-up article.

The reviewed literature consistently indicates that soil organic carbon can be estimated using VNIR–SWIR hyperspectral reflectance when appropriate preprocessing techniques, feature engineering strategies, and machine learning algorithms are employed. Accordingly, the proposed protocol incorporates these evidence-based methodological components to maximize the reliability and reproducibility of future pilot implementations. Derivative-based and frequency-domain spectrum adjustments are intended to improve SOC-sensitive features, which will improve model robustness and predictive performance in accordance with recent literature [1, 40, 56]. The literature currently in publication supports the nonlinear character of SOC–spectral interactions by showing that ensemble and nonlinear models, especially tree-based and stacking techniques, regularly outperformed linear baselines [42, 58].

Hyperspectral sensing continues to have distinct advantages over reduced-band systems for accurate SOC estimation, particularly because of the contribution of SWIR wavelengths, according to comparative data from earlier studies and the suggested framework across laboratory, UAV, airborne, and satellite platforms [5, 38]. The reviewed literature consistently indicates that soil organic carbon can be estimated using VNIR–SWIR hyperspectral reflectance when appropriate preprocessing techniques, feature engineering strategies, and machine learning algorithms are employed. Accordingly, the proposed protocol incorporates these evidence-based methodological components to maximize the reliability and reproducibility of future pilot implementations [44, 54]. This work fills a crucial methodological vacuum in hyperspectral SOC research by formalizing a transparent and repeatable pilot methodology. This will allow future studies to present results in a consistent, comparable, and MRV-ready way.

The principal contribution of this work lies in the formalization of a transparent, reproducible, and validation-oriented pilot protocol rather than the development of a new predictive algorithm. By integrating current best practices in Digital Soil Mapping, hyperspectral remote sensing, machine learning, uncertainty-aware evaluation, and standardized reporting, the proposed framework provides a methodological foundation that can support consistent implementation, facilitate inter-study comparison, and promote reproducible research within the rapidly evolving field of UAV-based SOC mapping.

LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

Despite its potential for high-resolution SOC mapping, the proposed framework has several limitations. First, its transferability across locations may be constrained by site-specific calibration requirements. Differences in soil texture, mineral composition, moisture regimes, and land-management practices can alter spectral responses and reduce model generalization when calibration data from one site are applied to another. Multi-location calibration and independent validation are therefore necessary to assess the spatial robustness of SOC prediction. Second, surface and soil variability remains an important source of uncertainty. Although bare-soil filtering can reduce the influence of vegetation and crop residues, variations in surface roughness, soil moisture, and residual organic material can modify spectral characteristics and affect SOC estimation. These factors are particularly relevant in arid agricultural environments, where strong temporal and spatial variations in surface conditions may occur. Third, the practical scalability of hyperspectral UAV-based SOC mapping is influenced by data acquisition and operational costs. High-quality hyperspectral sensors and associated field operations may require substantial investment, while access to complementary airborne or spaceborne hyperspectral observations may be limited in some regions. These constraints can restrict the deployment of the framework over large areas.

Future research should prioritize multi-site and multi-season validation to quantify model generalization and temporal stability under different soil-moisture conditions, cropping patterns, and management practices. Such validation would provide stronger evidence of the framework's applicability beyond individual study sites. The integration of auxiliary data sources also represents an important research direction. Synthetic aperture radar (SAR), terrain attributes, soil spectral libraries, and other environmental variables could complement hyperspectral observations and improve model robustness under conditions where spectral information alone is insufficient. Hybrid modelling approaches that combine physically based soil-spectral relationships with data-driven machine-learning methods should also be investigated. Such approaches may improve model interpretability while reducing dependence on extensive site-specific ground calibration. Finally, future work should examine the translation of SOC estimates into decision-support tools for precision agriculture, sustainable soil management, and carbon accounting. Integration with emerging hyperspectral satellite missions could further support scalable SOC monitoring and contribute to standardized monitoring, reporting, and verification (MRV) frameworks for carbon-aware land management.

AUTHORS' CONTRIBUTIONS

The authors confirm their contributions to the paper as follows: M.M., I.A., A.B., T.F., and N.P.: Study conception and design; A.J., A.R., H.Y., and A.B.: Data collection; M.A., I.A., N.P., and S.S.: Analysis and interpretation of results; M.A., T.F., N.P., and A.P.: Draft manuscript. All authors reviewed the results and approved the final version of the manuscript.

LIST OF ABBREVIATION

AI	=	Artificial Intelligence
BRDF	=	Bidirectional Reflectance Distribution Function
CWT	=	Continuous Wavelet Transform
DL	=	Deep Learning
DSM	=	Digital Soil Mapping
DWT	=	Discrete Wavelet Transform
FPAR	=	Fraction of Absorbed Photosynthetically Active Radiation
GNSS	=	Global Navigation Satellite System
GPS	=	Global Positioning System
ICA	=	Independent Component Analysis
LiDAR	=	Light Detection and Ranging
MAE	=	Mean Absolute Error
ML	=	Machine Learning
MOCCAE	=	Ministry of Climate Change and Environment
MRV	=	Monitoring, Reporting, and Verification
OGDC	=	Oil & Gas Decarbonization Centre
PCA	=	Principal Component Analysis
PLSR	=	Partial Least Squares Regression
PSS	=	Proximal Soil Sensing
QA/QC	=	Quality Assurance/Quality Control
RF	=	Random Forest
RMSE	=	Root Mean Square Error
R^2	=	Coefficient of Determination
RPD	=	Ratio of Performance to Deviation
RPIQ	=	Ratio of Performance to Interquartile Distance
RS	=	Remote Sensing
SOC	=	Soil Organic Carbon
SOM	=	Soil Organic Matter
SSL	=	Soil Spectral Library
SWIR	=	Short-Wave Infrared
UAE	=	United Arab Emirates
UAV	=	Unmanned Aerial Vehicle
VNIR	=	Visible and Near-Infrared

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

Not Applicable.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

All data generated or analyzed during this study are included in this published article.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] Dai J, Yang H, Maimaitituersun A, Maimaitiaili K, Ouyang C. A general prediction model for organic matter at different soil depths in the urumqi river basin. Land Degrad Dev 2025; •••: 1-17.
http://dx.doi.org/10.1002/ldr.70392

[2] Li T, Cui L, Wu Y, *et al.* Soil organic carbon estimation *via* remote sensing and machine learning techniques: Global topic modeling and research trend exploration. Remote Sens 2024; 16(17): 3168.
http://dx.doi.org/10.3390/rs16173168

[3] McGuirk SL, Cairns IH. Soil carbon remote sensing: A meta-analysis and systematic review of published results from 1969-2022. Geotechnics 2025; 5(2): 33.
http://dx.doi.org/10.3390/geotechnics5020033

[4] Inoue Y, Yoshino K, Hosoi F, Iwasaki A, Hirayama T, Saito T. Hyperspectral sensing and mapping of soil carbon content for amending within-field heterogeneity of soil fertility and enhancing soil carbon sequestration. Precis Agric 2024; 25(6): 2758-75.
http://dx.doi.org/10.1007/s11119-024-10140-1

[5] Angelopoulou T, Chabrillat S, Pignatti S, *et al.* Evaluation of airborne hyspex and spaceborne prisma hyperspectral remote sensing data for soil organic matter and carbonates estimation. Remote Sens 2023; 15(4): 1106.
http://dx.doi.org/10.3390/rs15041106

[6] Majeed I, Roy S, Reddy NN, *et al.* Assessing soil degradation in agricultural landscapes of semi-arid tropics using proximal and remote sensing-based diffuse reflectance spectroscopy. Soil Use

Manage 2025; 41(2): e70101.
http://dx.doi.org/10.1111/sum.70101

[7] Jennewein JS, Hively W, Lamb BT, *et al.* Spaceborne imaging spectroscopy enables carbon trait estimation in cover crop and cash crop residues. Precis Agric 2024; 25(5): 2165-97.
http://dx.doi.org/10.1007/s11119-024-10159-4

[8] Novais JJM, Melo BMD, Neves Junior AF, *et al.* Online analysis of Amazon's soils through reflectance spectroscopy and cloud computing can support policies and the sustainable development. J Environ Manage 2025; 375: 124155.
http://dx.doi.org/10.1016/j.jenvman.2025.124155 PMID: 39848197

[9] Vavlas NC, Porre R, Meng L, *et al.* Cover crop impacts on soil organic matter dynamics and its quantification using UAV and proximal sensing. Smart Agricultural Technology 2024; 9: 100621.
http://dx.doi.org/10.1016/j.atech.2024.100621

[10] Han Y, Wang B, Yang J, Yin F, He L. Research on hyperspectral inversion of soil organic carbon in agricultural fields of the southern shaanxi mountain area. Remote Sens 2025; 17(4): 600.
http://dx.doi.org/10.3390/rs17040600

[11] Ribeiro SG, Teixeira AS, de Oliveira MRR, *et al.* Soil organic´ carbon content prediction using soil-reflected spectra: A comparison of two regression methods. Remote Sens 2021; 13(23): 4752.
http://dx.doi.org/10.3390/rs13234752

[12] Yang C, Feng M, Song L, *et al.* Study on hyperspectral estimation model of soil organic carbon content in the wheat field under different water treatments. Scientific Reports 2021; Vol. 11.

[13] Purushothaman NK, Premsagar A, Raj M, *et al.* Influence of bidirectional reflectance distribution function in estimating basic soil properties using airborne hyperspectral data. IEEE Trans Geosci Remote Sens 2025; 63: 1-11.
http://dx.doi.org/10.1109/TGRS.2025.3569059

[14] Bartsch BA, Demattê JAM, Tziolas N, *et al.* Soil laboratory and satellite spectral data filtering: A Spectral Quality Protocol (SQuaP). Remote Sens Environ 2026; 333: 115144.
http://dx.doi.org/10.1016/j.rse.2025.115144

[15] Jain S, Sethia D, Tiwari KC. TLM-stack: A deep learning-based novel framework for soil nutrients estimation using hyperspectral data. Comput Electron Agric 2026; 243: 111381.
http://dx.doi.org/10.1016/j.compag.2025.111381

[16] Luo Y, Li C, Huang J, Dong C, Wang J. Integrating fractional-order derivatives of soil and leaf hyperspectral reflectance for improved estimation of mangrove soil organic carbon. Geoderma 2025; 458: 117324.
http://dx.doi.org/10.1016/j.geoderma.2025.117324

[17] Wu M, Dou S, Lin N, *et al.* Hyperspectral estimation of soil organic matter based on fod-scars and machine learning algorithm. Guangpuxue Yu Guangpu Fenxi 2025; 45(1): 204-12.
http://dx.doi.org/10.3964/j.issn.1000-0593(2025)01-0204-09

[18] Hong Y, Chen Y, Chen S, *et al.* Bridging the gap between laboratory VNIR-SWIR spectra and Landsat-8 bare soil composite image for soil organic carbon prediction. Remote Sens Environ 2025; 328: 114874.
http://dx.doi.org/10.1016/j.rse.2025.114874

[19] Mruthyunjaya P, Shetty A, Umesh P, Gomez C. Impact of atmospheric correction methods parametrization on soil organic carbon estimation based on hyperion hyperspectral data. Remote Sens 2022; 14(20): 5117.
http://dx.doi.org/10.3390/rs14205117

[20] Song Q, Zhang W. Estimation and spatial distribution of soil organic carbon content in farmland using unmanned aerial vehicle hyperspectral remote sensing technology. Sci Rep 2026; 16(1): 5480.
http://dx.doi.org/10.1038/s41598-026-35096-2 PMID: 41545412

[21] Xia C, Zhang Y. Estimating field-scale soil organic matter in agricultural soils using uav hyperspectral imagery. AgriEngineering 2025; 7(10): 339.
http://dx.doi.org/10.3390/agriengineering7100339

[22] Ding Z, Liu K, Grunwald S, *et al.* Advancing soil organic carbon prediction: A comprehensive review of technologies, ai, process-based and hybrid modelling approaches. Adv Sci (Weinh) 2025; 12(31): e04152.
http://dx.doi.org/10.1002/advs.202504152 PMID: 40557741

[23] Nur N B, Bachmann C M, Tyler A C, *et al.* Mapping soil organic matter, total carbon, and total nitrogen in salt marshes using UAS-based hyperspectral imaging. J Geophys Res Biogeosci 2025; 130: e2024JG008421.
http://dx.doi.org/10.1029/2024JG008421

[24] Chen Y, Shi T, Li Q, *et al.* Mapping soil properties in tropical rainforest regions using integrated uav-based hyperspectral images and lidar points. Forests 2024; 15(12): 2222.
http://dx.doi.org/10.3390/f15122222

[25] Das A, Mishra MK, Chakraborty S, *et al.* Deep carbon: A multiscale feature-time fusion approach for field level digital soil organic carbon mapping. Eur J Soil Sci 2025; 76(4): e70161.
http://dx.doi.org/10.1111/ejss.70161

[26] Alsaleh ARS, Alcibahy M, Gafoor FA, *et al.* Estimation of soil organic carbon in arid agricultural fields based on hyperspectral satellite images. Geoderma 2025; 453: 117151.
http://dx.doi.org/10.1016/j.geoderma.2024.117151

[27] The UAE Net Zero by 2050 Strategic Initiative. 2023. Available from https://u.ae/en/about-the-uae/strategies-initiatives-and-awards/strategies-plans-and-visions/environment-and-energy/the-uae-net-zero-2050-strategy

[28] National Food Security Strategy 2051. 2023. Available from https://u.ae/en/about-the-uae/strategies-initiatives-and-awards/strategies-plans-and-visions/environment-and-energy/national-food-security-strategy-2051

[29] COP28 UAE Presidency. Food and Agriculture at COP28. 2023. Available from https://www.cop28.com/en/food-and-agriculture

[30] Ministry of Climate Change and Environment (MOCCAE), Aerial Mapping Project Using Drones. Aerial Mapping Project Using Drones. Dubai, UAE: Ministry of Climate Change and Environment. 2020. Available from https://moccae.gov.ae/Handlers/DownloadPDF.ashx?id=58470

[31] Mohammed Bin Rashid Reviews Agriculture Drone Survey. 2020. Available from https://mediaoffice.ae/news/2020/March/04-03/Mohammed-bin-Rashid-reviews-agriculture-drone-survery

[32] Abu Dhabi National Oil Company (ADNOC). Advancing Towards Net Zero: Delivering Progress, Abu Dhabi, UAE. 2023. Available from https://www.adnoc.ae/-/media/Adnoc-v2/Sub-Brands/Sustainability-and-Energy-Transition/Files/Advancing-TowardsNetZero---Delivering-Progress--JULY2023.ashx

[33] The Oil and Gas Decarbonization Charter Report 2024. 2024. Available from https://www.ogdc.org/wp-content/uploads/2024/11/The-Oil-Gas-Decarbonization-Charter-Report-2024.pdf

[34] Ministry of Climate Change and Environment (MOCCAE). Roadmap for the National Carbon Sequestration Project. 2022. Available from https://moccae.gov.ae/en/media-center/news/

[35] Kalambukattu JG, Kumar SK, Roy T. Digital soil mapping using geospatial data and machine learning techniques. Intelligence Systems for Earth, Environmental and Planetary Sciences. Elsevier 2024; pp. 335-78.
http://dx.doi.org/10.1016/B978-0-443-13293-3.00014-2

[36] Mitran T, Suresh JR, Sujatha G, *et al.* Digital soil mapping: A tool for sustainable soil management. In: Rahman MM, Biswas JC, Meena RS, Eds. Climate Change and Soil-Water-Plant Nexus: Agriculture and Environment. Springer Nature Singapore 2024; pp. 51-95.

[37] Maurya KV, Mahajan S, Chaube NR. Remote sensing techniques: mapping and monitoring of mangrove ecosystem: A review. Complex Intell Syst 2021; 7: 2797-818.
http://dx.doi.org/10.1007/s40747-021-00457-z

[38] Tsatsoulis T, Tsakiridis NL, Karyotis K, Zalidis GC. Assessing machine learning models for soil property prediction using resampled spectral data: Implications for hyperspectral spaceborne imaging. Workshop Hyperspectral Image and Signal Processing, Evolution in Remote Sensing. Athens, Greece, 2023, pp. 1-6

http://dx.doi.org/10.1109/WHISPERS61460.2023.10431009

[39] Zhu Y, Qi L, Wu Z, Shi P. Spectra-based predictive mapping of soil organic carbon in croplands: Single-date *versus* multitemporal bare soil compositing approaches. Geoderma 2024; 449: 116987. http://dx.doi.org/10.1016/j.geoderma.2024.116987

[40] Roy B, Sagan V, Alifu H, Saxton J, Ghoreishi D, Shakoor N. Soil carbon estimation from hyperspectral imagery with wavelet decomposition and frame theory. IEEE Trans Geosci Remote Sens 2024; 62: 1-12. http://dx.doi.org/10.1109/TGRS.2024.3461628

[41] Golkar Amoli M, Hasanlou M, Taghizadeh Mehrjardi R, Samadzadegan F. Exploring the potential of prisma satellite hyperspectral image for estimating soil organic carbon in marvdasht region, southern iran. Remote Sens 2024; 16(12): 2149. http://dx.doi.org/10.3390/rs16122149

[42] Shah A, Sah S, Singhal S, Jain S, Sethi D. dsep-xnet: A feature optimized ensemble framework for hyperspectral-based soil organic carbon prediction. 2024 International Conference on Augmented Reality, Intelligent Systems, and Industrial Automation (ARIIA). Manipal, India, 2024, pp. 1-8 http://dx.doi.org/10.1109/ARIIA63345.2024.11051496

[43] Reddy BS, Shwetha HR. Integrating soil spectral library and prisma data to estimate soil organic carbon in crop lands. IEEE Geosci Remote Sens Lett 2024; 21: 1-5. http://dx.doi.org/10.1109/LGRS.2024.3374824

[44] Milewski R, Abdelbaki A, Chabrillat S, Tziolas NV, van Wesemael B, Jacquemoud S. Simulation of spectral disturbance effects for improvement of soil property estimation. 2023 IEEE International Geoscience and Remote Sensing Symposium. Pasadena, CA, USA, 2023, pp. 1257-1260 http://dx.doi.org/10.1109/IGARSS52108.2023.10282492

[45] Salani G M, Lissoni M, Bianchini G, Brombin V, Natali S, Natali C. Soil organic carbon estimation in ferrara (northern italy) combining in situ geochemical analyses and hyperspectral remote sensing. Environments 2023; 10(10): 173. http://dx.doi.org/10.3390/environments10100173

[46] Alfin AA, Suciati N, Khotimah WN. Bayesian optimization of convolutional neural network hyperparameters for soil macronutrient estimation using hyperspectral imaging. 2025 International Conference on Smart Computing, IoT and Machine Learning (SIML). Surakarta, Indonesia, 2025, pp. 1-6 http://dx.doi.org/10.1109/SIML65326.2025.11080907

[47] Odebiri O, Mutanga O, Odindi J, Naicker R, Masemola C, Sibanda M. Deep learning approaches in remote sensing of soil organic carbon: a review of utility, challenges, and prospects. Environ Monit Assess 2021; 193(12): 802. http://dx.doi.org/10.1007/s10661-021-09561-6 PMID: 34778906

[48] He J, Guo Y, Zhang Y, Yang X, Liu T, Wang L. Dynamic estimation fpar of summer maize based on gf-1 satellite data. Nongye Jixie Xuebao 2022; 53(4): 164-72. http://dx.doi.org/10.6041/j.issn.1000-1298.2022.04.017

[49] Deng Y, Zhang X, Zhao X, Liu C, Jiang J, Zhang B. Estimation of soil organic carbon content based on ground hyperspectral remote sensing. Fifth International Conference on Geology, Mapping, and Remote Sensing (ICGMRS 2024). Wuhan, China, 2024 http://dx.doi.org/10.1117/12.3035450

[50] Liu J, Xie J, Meng T, Dong H. Organic matter estimation of surface soil using successive projection algorithm. Agron J 2022; 114(4): 1944-51. http://dx.doi.org/10.1002/agj2.20934

[51] Canero FM, Rodriguez-Galiano V, Aragones D. Machine Learning and Feature Selection for soil spectroscopy. An evaluation of Random Forest wrappers to predict soil organic matter, clay, and carbonates. Heliyon 2024; 10(9): e30228. http://dx.doi.org/10.1016/j.heliyon.2024.e30228 PMID: 38707402

[52] Pellikka P, Luotamo M, Sädekoski N, *et al.* Tropical altitudinal gradient soil organic carbon and nitrogen estimation using Specim IQ portable imaging spectrometer. Sci Total Environ 2023; 883: 163677. http://dx.doi.org/10.1016/j.scitotenv.2023.163677 PMID: 37105488

[53] Luo D, Xie Y, Tang J, *et al.* Improving the prediction accuracy of soil organic matter: Addressing the challenge of soil moisture variability. Ecol Indic 2025; 179: 114249. http://dx.doi.org/10.1016/j.ecolind.2025.114249

[54] van Wesemael B, Chabrillat S, Sanz Dias A, Berger M, Szantoi Z. Remote sensing for soil organic carbon mapping and monitoring. Remote Sens 2023; 15(14): 3464. http://dx.doi.org/10.3390/rs15143464

[55] Liu Q, He L, Guo L, *et al.* Digital mapping of soil organic carbon density using newly developed bare soil spectral indices and deep neural network. Catena 2022; 219: 106603. http://dx.doi.org/10.1016/j.catena.2022.106603

[56] Zhou W, Cao X, Wang K, *et al.* Hyperspectral modeling of soil organic carbon contenta case study in the three rivers source region qinghai tibet plateau. Bingchuan Dongtu 2023; 45(2): 823-32. http://dx.doi.org/10.7522/j.issn.1000-0240.2023.0063

[57] Zhang X, Meng X, Tang H, Liu H, Zhang X, Liu Q. Random forest prediction model for the soil organic matter with optimized spectral inputs. Nongye Gongcheng Xuebao (Beijing) 2023; 39(2): 90-9. [Transactions of the CSAE]. http://dx.doi.org/10.11975/j.issn.1002-6819.202207035

[58] Datta D, Paul M, Murshed M, Teng SW, Schmidtke L. Soil moisture, organic carbon, and nitrogen content prediction with hyperspectral data using regression models. Sensors 2022; 22(20): 7998. http://dx.doi.org/10.3390/s22207998 PMID: 36298349

[59] Datta D, Schmidtke L. Comparative analysis of machine and deep learning models for soil properties prediction from hyperspectral visual band. Environments 2023; 10(5): 77. http://dx.doi.org/10.3390/environments10050077

[60] Dhiman G, Bhattacharya J, Roy S. Soil textures and nutrients estimation using remote sensing data in North India - Punjab region. Procedia Computer Science 2182022; : 2041-8. http://dx.doi.org/10.1016/j.procs.2023.01.180