INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 1233-1236
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

A REVIEW ON MACHINE LEARNING-BASED SOIL HEALTH MONITORING

Dr. P. Bastin Thiyagaraj[1], D. Suriya Prakash[2], R. Rashika[3], J. Sandhiya[4]

[1]Assistant Professor, Department of Information Technology, St. Joseph's College(Autonomous), Trichy-620002, TamilNadu, India.

[2 3 4]Department of Information Technology, St. Joseph's College(Autonomous), Trichy-620002, Tamil Nadu, India.

DOI: https://www.doi.org/10.58257/IJPREMS34742

ABSTRACT

Soil health is a critical factor in sustainable agriculture, food security, and ecosystem stability. Traditional laboratory-based soil analysis methods are accurate but remain costly, time-consuming, and spatially limited, creating the need for faster and scalable alternatives. In recent years, machine learning (ML) techniques integrated with remote sensing have emerged as promising solutions. This review examines three major approaches: the use of Unmanned Aerial Vehicles (UAVs) equipped with hyperspectral sensors for high-resolution nutrient mapping, ensemble ML models for robust soil organic carbon (SOC) prediction, and soil-science-informed ML approaches that embed domain knowledge into algorithms for improved interpretability. Studies such as [1] and [6] demonstrate the potential of UAV-based imaging for nutrient estimation, while [2] highlights the advantages of ensemble methods for SOC modeling across diverse landscapes. More recently, [3] emphasized soil-informed ML for balancing accuracy with scientific explainability in data-scarce regions. Through a comparative review of datasets, preprocessing techniques, and algorithms, this paper highlights the strengths, limitations, and trade-offs of these approaches. The study concludes that UAV hyperspectral imaging provides unmatched spatial detail, ensemble SOC prediction ensures robustness, and soil-informed ML enhances generalizability. A hybrid strategy integrating these methods is suggested as a practical direction for future soil health monitoring research.

Keywords: Soil Health, UAV Hyperspectral Imaging, Ensemble Machine Learning, Soil Organic Carbon, Precision Agriculture, Soil-Science-Informed ML.

1. INTRODUCTION

Soil health is fundamental for sustainable agriculture, food security, and ecosystem stability. Traditional laboratory-based soil analyses provide accurate results but are often time-consuming, costly, and limited in spatial coverage. With the increasing demand for rapid and scalable soil monitoring, machine learning (ML) integrated with remote sensing has emerged as a promising alternative. Current approaches, however, face several challenges. UAV hyperspectral imaging provides high-resolution data but requires complex preprocessing and calibration. Ensemble SOC prediction enhances robustness but may lack interpretability, while soil-science-informed ML integrates domain knowledge for improved generalization but relies heavily on high-quality soil datasets. A comprehensive comparison of these methods is still lacking, making it difficult for researchers to select the most suitable approach for specific contexts. Moreover, the lack of standardized evaluation metrics across studies further limits consistency in performance assessment and hinders practical adoption. This review analyzes and compares ML-based soil health monitoring techniques, focusing on UAV hyperspectral imaging, ensemble SOC prediction, and soil-science-informed ML. By summarizing existing studies, highlighting trends, and discussing limitations, it provides insights into the strengths and weaknesses of each approach and proposes directions for future research.

2. LITERATURE SURVEY

Researchers have explored ML and remote sensing for soil health, highlighting the strengths and limits of UAV imaging, ensemble methods, and soil-informed ML as a basis for future work.

Liu et al. [1] explored UAV hyperspectral imaging for soil nutrient monitoring. Their study demonstrated that UAVs equipped with hyperspectral sensors could generate high-resolution nutrient maps across heterogeneous fields. While the results confirmed accurate nutrient estimation, the approach required extensive preprocessing and calibration, which limited scalability and practical field deployment.

Zhou et al. [6] applied UAV-based hyperspectral imagery combined with regression models for predicting soil organic carbon (SOC). Their results suggested that UAV imaging could, in some cases, substitute for traditional laboratory measurements. However, their method struggled with variability in diverse landscapes, reducing generalizability across regions.

Mundada et al. [2] proposed ensemble machine learning techniques, including stacking and boosting, for SOC estimation. By integrating multiple algorithms, they achieved robust predictions and minimized overfitting across varied datasets. Despite these improvements, their models lacked interpretability, making it challenging for agronomists and soil scientists to understand or trust the decision-making process.

Minasny et al. [3] introduced soil-science-informed ML approaches that incorporated domain knowledge into algorithms through feature engineering and physics-based constraints. This enhanced model generalization and reliability in data-scarce environments. However, the method was heavily dependent on high-quality soil datasets, which may not always be available in practice.

Scope of the Literature: Across these studies, UAV hyperspectral imaging emerges as the most effective for capturing fine-scale spatial variability, ensemble machine learning provides robustness against data noise, and soil-informed ML enhances interpretability and generalization. At the same time, challenges remain in terms of preprocessing complexity, limited scalability, and data availability, suggesting that hybrid approaches integrating these methods may be the most promising future direction.

3. COMPARATIVE ANALYSIS

3.1 UAV Hyperspectral Imaging

Liu et al. [1] and Zhou et al. [6] demonstrated that Unmanned Aerial Vehicles (UAVs) equipped with hyperspectral sensors can capture high-resolution spectral data, enabling detailed mapping of soil nutrients and soil organic carbon (SOC). Their workflows typically involve UAV flight for data collection, preprocessing steps such as calibration, noise reduction, and normalization, followed by machine learning models like Random Forest and Partial Least Squares regression. The key advantage of this approach is the unmatched spatial detail and precision it provides in field-scale monitoring. However, limitations include the high cost of UAV equipment, requirement for complex preprocessing, and scalability challenges in larger or more diverse landscapes. The overall workflow is illustrated in Fig 1, which summarizes the sequential process from UAV flight to soil nutrient mapping.

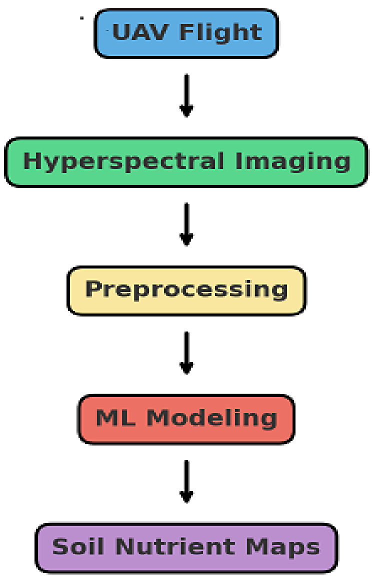

Fig 1. UAV Hyperspectral Workflow for Soil Nutrient Mapping

3.2 Ensemble SOC Prediction

Mundada et al. [2] and Adeniyi et al. [5] applied ensemble learning techniques such as stacking, bagging, and boosting to predict soil organic carbon (SOC) across varied landscapes. Their frameworks integrate multiple input features including UAV spectral data, satellite imagery, and Digital Elevation Models (DEM), combining them through algorithms like Random Forest, Gradient Boosting, and XGBoost. Ensemble methods improve robustness and reduce overfitting compared to single models, ensuring more reliable SOC predictions. However, a key drawback is their "black-box" nature, as interpretability is limited, making it difficult for soil scientists and agronomists to understand model decisions. The workflow of ensemble SOC prediction is depicted in Fig 2, showing how multiple learners are combined to improve prediction reliability.

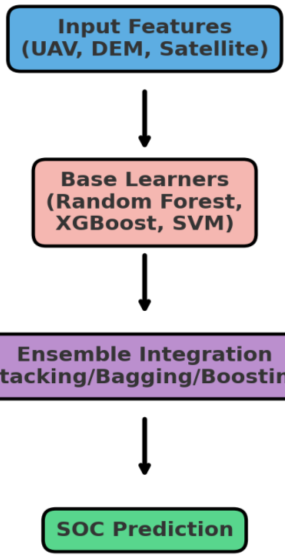

Fig 2. Ensemble SOC Prediction Workflow

3.3 Soil-Science-Informed ML

Minasny et al. [3] emphasized the integration of soil science knowledge into machine learning workflows to improve model interpretability and generalization in data-scarce environments. Their approach combines traditional soil data with domain-specific knowledge such as pedogenic processes, soil texture, and environmental covariates. By embedding this information into the model through feature engineering and constraints, the resulting predictions align more closely with scientific understanding. The key strength of this method is improved interpretability and trustworthiness compared to purely data-driven models. However, the success of soil-informed ML is heavily dependent on the availability and quality of soil datasets, which may not always be accessible in diverse regions. The overall framework is illustrated in Fig 3, showing how soil data and domain knowledge jointly contribute to interpretable ML predictions.

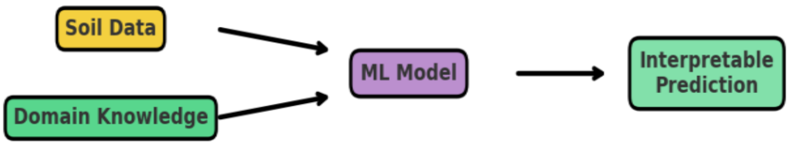

Fig 3. Soil-Science-Informed ML

3.4 Discussion

Across these three approaches, UAV hyperspectral imaging provides high spatial detail and accuracy, but it is constrained by high equipment cost, heavy preprocessing requirements, and limited scalability. Ensemble SOC prediction offers robustness and reliability by combining multiple base learners such as Random Forest, XGBoost, and SVM; however, these ensembles often behave as black boxes, making interpretability a challenge. Soil-science-informed ML adds interpretability and scientific alignment by embedding pedological knowledge into the learning process, yet it depends strongly on the availability and quality of soil datasets, which may not always be accessible across diverse regions.

In summary, UAV hyperspectral imaging excels in resolution, ensemble SOC prediction in robustness, and soil-science-informed ML in interpretability. A hybrid strategy that integrates the strengths of all three could provide a more balanced and scalable solution for soil health monitoring.

4. CONCLUSION

This review compared UAV hyperspectral imaging, ensemble SOC prediction, and soil-science-informed ML for soil health monitoring. UAV imaging provides high spatial detail, ensemble methods improve robustness, and soil-informed ML enhances interpretability. Each approach has trade-offs, such as preprocessing complexity, limited scalability, or reliance on domain knowledge, but a hybrid framework combining their strengths could offer scalable, accurate, and interpretable solutions for sustainable agriculture. In the future, integrating real-time UAV data with advanced ML models and cloud-based platforms could enable continuous soil monitoring, early detection of degradation, and precision intervention. Moreover, incorporating multi-source datasets, such as weather, crop type, and remote sensing imagery, may improve predictive accuracy and support decision-making for climate-resilient farming. Additionally, developing user-friendly decision-support tools for farmers and policymakers can bridge the gap between research outputs and on-ground adoption. Such innovations could transform soil health management, making it more proactive, data-driven, and environmentally sustainable.

5. REFERENCES

[1] K. Liu, Y. Wang, Z. Peng, X. Xu, J. Liu, Y. Song, H. Di, and D. Hua, “Monitoring soil nutrients using machine learning based on UAV hyperspectral remote sensing,” International Journal of Remote Sensing, 2024.

[2] S. Mundada, P. Jain, et al., “Predicting soil organic carbon with ensemble learning techniques,” Scientific Reports, Nature Publishing Group, 2025.

[3] B. Minasny, et al., “Soil Science-Informed Machine Learning,” Elsevier, 2024.

[4] Y. Yan, J. Yang, B. Li, C. Qin, W. Ji, Y. Xu, and Y. Huang, “High-Resolution Mapping of Soil Organic Matter at the Field Scale Using UAV Hyperspectral Images with a Small Calibration Dataset,” Remote Sensing, vol. 15, no. 5, p. 1433, 2023.

[5] O. D. Adeniyi, A. Brenning, A. Bernini, S. Brenna, and M. Maerker, “Digital Mapping of Soil Properties Using Ensemble Machine Learning Approaches in an Agricultural Lowland Area of Lombardy, Italy,” Land, vol. 12, no. 2, p. 494, 2023.

[6] J. Zhou, Y. Xu, X. Gu, et al., “High-Precision Mapping of Soil Organic Matter Based on UAV Imagery Using Machine Learning Algorithms,” Drones, vol. 7, no. 5, p. 290, 2023.

[7] N. Francos, P. Nasta, C. Allocca, et al., “Mapping Soil Organic Carbon Stock Using Hyperspectral Remote Sensing: A Case Study in the Sele River Plain in Southern Italy,” Remote Sensing, vol. 16, no. 5, p. 897, 2024.

[8] D. Triantakonstantis and A. Karakostas, “Soil Organic Carbon Monitoring and Modelling via Machine Learning Methods Using Soil and Remote Sensing Data,” Agriculture, vol. 15, p. 910, 2025.

[9] I. El-Jamaoui, M. J. M. Sánchez, C. P. Sirvent, and S. A. M. López, “A Comparative Assessment of Sentinel-2 and UAV-Based Imagery for Soil Organic Carbon Estimations Using Machine Learning Models,” Sensors, vol. 25, no. 17, p. 5281, 2025.

[10] R. Chaudhary, et al., “Monitoring Soil Health Using Machine Learning Techniques for Improved Crop Yield. Prediction,” Cureus Journal of Computer Science, 2025.