engineering proceedings

MDPI

Proceeding Paper

Edge-AI-Based Multi-Modal Soil Organic Matter Estimation Using RGB, Spectral, and Munsell Color Representations †

Mario Fahim [1], Dionysia Varvarigou [2], Aliki Panagiotarou [2], Pantelis Barouchas [3] and Paris Kitsos [1,*]

[1] Department of Electrical and Computer Engineering, University of Peloponnese, 26334 Patras, Greece; me@mariofahim.com
[2] Department of Electrical and Computer Engineering, University of Patras, 26504 Patras, Greece; dvarvarigou@upatras.gr (D.V.), apanagiotarou@upatras.gr (A.P.)
[3] Department of Agriculture, University of Patras, 26504 Patras, Greece; pbar@upatras.gr
* Correspondence: kitsos@go.uop.gr
† Presented at the 2nd International Conference on Engineering, Technology, and Applied Science Innovations (ICETASI 2026), Zakynthos, Greece, 24–26 September 2026.

Abstract

Accurate estimation of Soil Organic Matter (SOM) is essential for sustainable agricultural practices and precision farming. However, conventional laboratory-based analyses are costly, time-consuming, and require destructive sampling, limiting their applicability for frequent field measurements. Although machine learning approaches have demonstrated promising results for SOM prediction, many existing solutions rely on laboratory instruments or computational environments that limit their deployment in real-world agricultural scenarios. This paper presents a low-cost edge-AI framework for non-destructive SOM estimation based on a Raspberry Pi 5 platform. The proposed system integrates four complementary prediction models: a spectral ensemble model, an RGB image-based model, a Munsell colour-based model, and a hybrid RGB–Munsell model. The evaluation, performed on a dataset of 6167 soil samples, showed the spectral ensemble model achieving the highest performance with an R^2 of 0.845, an MAE of 0.416% SOM, and an RPD of 2.54. A lightweight web interface enables image acquisition, data input, and local inference without requiring cloud connectivity. The framework addresses practical deployment challenges, including camera-based feature extraction, model compatibility, and execution on resource-constrained hardware. The implemented prototype demonstrates the feasibility of executing the complete SOM estimation pipeline entirely on an embedded platform, enabling portable and autonomous soil analysis.

Keywords: Soil Organic Matter estimation; precision agriculture; edge artificial intelligence; machine learning; RGB imaging; soil spectroscopy; Raspberry Pi

Check for updates

Academic Editor: Konstantinos Poirazidis

Published: 6 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Soil Organic Matter (SOM) is a fundamental component of soil quality that directly influences nutrient availability, water retention, soil structure, and ecosystem sustainability. Its preservation is also closely linked to global challenges such as food security and climate change mitigation [1]. Consequently, accurate and frequent SOM assessment is essential to support sustainable agricultural practices and precision agriculture applications [2]. However, conventional laboratory-based measurements, although accurate, require sampling, chemical preparation, and specialized equipment, making them time-consuming and difficult to apply for large-scale, real-time field monitoring [3].

 https://doi.org/10.3390/engproc2026167012

Recent advances in sensing technologies and machine learning have created new opportunities for rapid, non-destructive soil analysis. In particular, visible and near-infrared spectroscopy has demonstrated strong potential for estimating soil properties by exploiting relationships between spectral responses and soil composition [4]. Similarly, soil colour has been investigated as an indirect indicator of organic matter content, since colour variations are related to soil composition and organic matter characteristics [5]. Nevertheless, these approaches often rely on laboratory instruments or controlled acquisition conditions, limiting their direct deployment in low-cost field environments.

Machine learning techniques provide an effective way to exploit complex relationships between soil measurements and SOM content. Various regression and ensemble methods have been successfully applied in agricultural applications, enabling the combination of heterogeneous features and improving prediction capabilities [6,7].

In summary, there are three main challenges in developing an accessible SOM estimation system: (i) the integration of different input representations, including spectral, colour-based, and manually measured soil descriptors, (ii) the development of accurate machine learning models while avoiding unreliable evaluation caused by data leakage or inappropriate validation strategies, and (iii) the deployment of the trained models into a portable and autonomous field device. Existing studies mainly focus on individual sensing approaches or laboratory-oriented prediction models and do not provide an integrated, low-cost solution that supports different measurement scenarios.

To address these challenges, this paper presents a portable machine-learning-based system for non-destructive SOM estimation, implemented on a Raspberry Pi 5 (Raspberry Pi Ltd., Cambridge, UK) platform. The proposed approach combines four complementary prediction models: (i) a spectral ensemble model based on soil reflectance measurements, (ii) an RGB camera-based model using soil colour features, (iii) a Munsell colour-based model using manual soil colour description, and (iv) a hybrid model combining RGB and Munsell information. The proposed system integrates data processing, model inference, and user interaction into a local web application, enabling operation without external computational infrastructure.

The proposed methodology is evaluated through a complete implementation workflow, from dataset preparation and model training to deployment and execution on the embedded device. The results demonstrate the feasibility of performing SOM estimation locally using a low-cost computing platform, while highlighting the advantages and limitations of each sensing modality. The main contributions of this work are threefold: (i) the development of a multimodal SOM estimation framework integrating spectral, RGB imaging, and Munsell colour representations; (ii) the implementation of the complete prediction pipeline on a Raspberry Pi 5 edge platform, enabling autonomous local inference without cloud dependency; and (iii) the comparative evaluation of different sensing modalities to investigate their suitability for low-cost precision agriculture applications.

The remainder of this paper is structured as follows. Section 2 presents the related work regarding SOM estimation, colour-based analysis, spectroscopy, and machine learning approaches. Section 3 describes the proposed methodology, including the developed prediction models and system architecture. Section 4 presents the implementation, deployment, and evaluation results. Finally, Section 5 concludes the paper and presents future research directions.

2. Related Work

Existing research has demonstrated the potential of different sensing and modelling approaches for estimating SOM. However, most existing solutions focus on a single measurement modality or laboratory-based analysis, limiting their direct applicability in low-cost

and autonomous field environments. To the best of our knowledge, there is a lack of an integrated methodology that combines spectral, image-based, and colour-based soil representations into a portable system capable of performing local SOM estimation.

2.1. Soil Organic Matter Estimation Methods

Traditional SOM estimation methods are mainly based on laboratory chemical analysis. Methods such as Walkley–Black and loss-on-ignition provide reliable estimations of organic matter content but require sample collection, laboratory equipment, and time-consuming procedures [3,8]. Although these approaches remain widely used as reference methods, their operational limitations make them unsuitable for continuous and large-scale field monitoring.

To overcome these limitations, non-destructive sensing techniques have been investigated, with spectroscopy becoming one of the most widely explored approaches for soil characterization. Visible and near-infrared spectroscopy has shown the ability to estimate several soil properties by exploiting relationships between spectral responses and chemical composition [4]. Furthermore, diffuse reflectance spectroscopy has shown potential for simultaneously predicting multiple soil properties, reducing the need for conventional laboratory analysis [9].

Another research direction exploits soil colour as an indirect indicator of soil properties. Several studies have demonstrated correlations between soil colour parameters and organic matter content [5]. More recent approaches have extended colour-based analysis through digital sensors and Munsell-based measurements, enabling more accessible soil characterization [10]. Nevertheless, colour measurements remain sensitive to illumination conditions, sensor characteristics, and calibration requirements, creating challenges for direct deployment in uncontrolled environments.

2.2. Machine Learning for Soil Prediction

The increasing availability of sensing data has enabled the application of machine learning methods for agricultural prediction tasks. Machine learning models can capture complex relationships between input measurements and soil properties, providing improved prediction capabilities compared with traditional statistical approaches [6]. Several algorithms, including support vector regression, partial least squares regression, random forests, and ensemble methods, have been successfully applied in soil spectroscopy and agricultural applications [11,12].

Recent studies highlight the importance of combining multiple algorithms to improve robustness and prediction accuracy. Ensemble learning approaches exploit the complementary behaviour of different models and have been widely adopted for regression problems involving complex and high-dimensional data [12]. However, most existing machine learning-based SOM estimation approaches are developed and evaluated independently for a specific sensing modality, without considering how different representations of soil information can be combined into a unified operational system.

Additionally, reliable evaluation of machine learning models requires careful experimental design. Improper validation strategies or data leakage can lead to overly optimistic performance estimates and poor generalization to unseen samples [13]. Therefore, a complete SOM estimation solution should not only provide accurate prediction models but also ensure reliable training, validation, and deployment procedures.

2.3. Portable and Edge-Based Agricultural Systems

The deployment of machine learning models directly on low-cost computing platforms has become increasingly relevant for precision agriculture applications. Edge computing

enables local data processing, reducing dependency on external computational resources and improving system autonomy. Low-cost embedded platforms, such as Raspberry Pi devices, provide sufficient computational capabilities for executing lightweight machine learning workflows while maintaining portability and affordability [14].

However, existing studies mainly focus either on developing prediction models or on analysing individual sensing techniques. There is still a need for a complete end-to-end system integrating data acquisition, machine learning inference, and user interaction into a portable device suitable for practical agricultural use. Based on these observations, this work addresses the limitations of existing approaches by introducing a unified SOM estimation framework combining spectral, RGB camera, Munsell colour, and hybrid representations. The proposed system integrates multiple trained models into a Raspberry Pi-based platform, enabling autonomous and low-cost soil assessment directly in the field.

3. Proposed Methodology

This work proposes a multimodal machine-learning framework for non-destructive SOM estimation, integrating spectral, image-based, and colour-description-based representations. The objective is to develop a complete pipeline capable of supporting different sensing scenarios, ranging from high-information spectral measurements to low-cost RGB camera acquisition suitable for field deployment. The proposed methodology combines four complementary prediction models: (i) a spectral ensemble model based on soil reflectance measurements, (ii) an RGB model based on colour features extracted from soil images, (iii) a Munsell colour model using manually recorded soil colour descriptors, and (iv) a hybrid RGB–Munsell model combining image-derived and manually obtained colour information.

The complete workflow consists of five main stages: (i) dataset preparation and feature construction, (ii) generation of the different input representations, (iii) machine learning model training and validation, (iv) integration of the trained models into a Raspberry Pi 5 edge platform, and (v) local SOM inference through a web-based interface. Unlike approaches that focus exclusively on prediction performance, this work considers the complete transition from model development to practical deployment by addressing data consistency, model compatibility, and execution constraints on embedded hardware. The overall workflow of the proposed framework is illustrated in Figure 1.

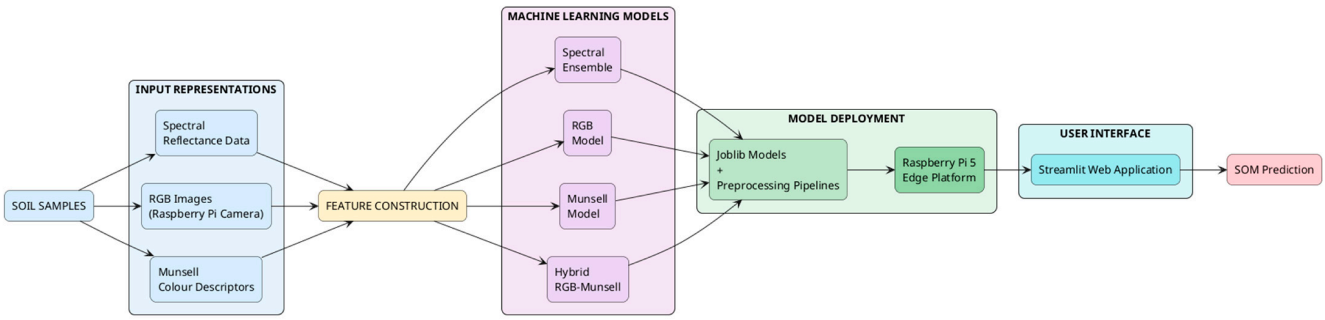


Figure 1. Proposed multimodal Soil Organic Matter estimation framework.

The main design objective is not only to achieve accurate SOM prediction but also to transform the developed models into a deployable field-oriented solution where different sensing modalities can be selected depending on the available acquisition mechanism.

3.1. System Overview

The complete methodology follows a model-development-to-deployment workflow. During the development phase, the soil dataset is transformed into different feature representations and used to train independent prediction models. During the deployment phase, the trained models and their preprocessing components are transferred to the Raspberry Pi 5, where the corresponding inference pipeline is executed locally.

The proposed system architecture consists of four main components: the data representation component, the machine learning component, the deployment component, and the user interaction component. The data representation component supports three complementary sources of soil information: spectral measurements, RGB images acquired through a Raspberry Pi Camera Module 3 (Raspberry Pi Ltd., Cambridge, UK), and Munsell colour descriptors. Each representation is processed independently to generate the corresponding feature vectors used by the prediction models. The proposed system is depicted in Figure 2.

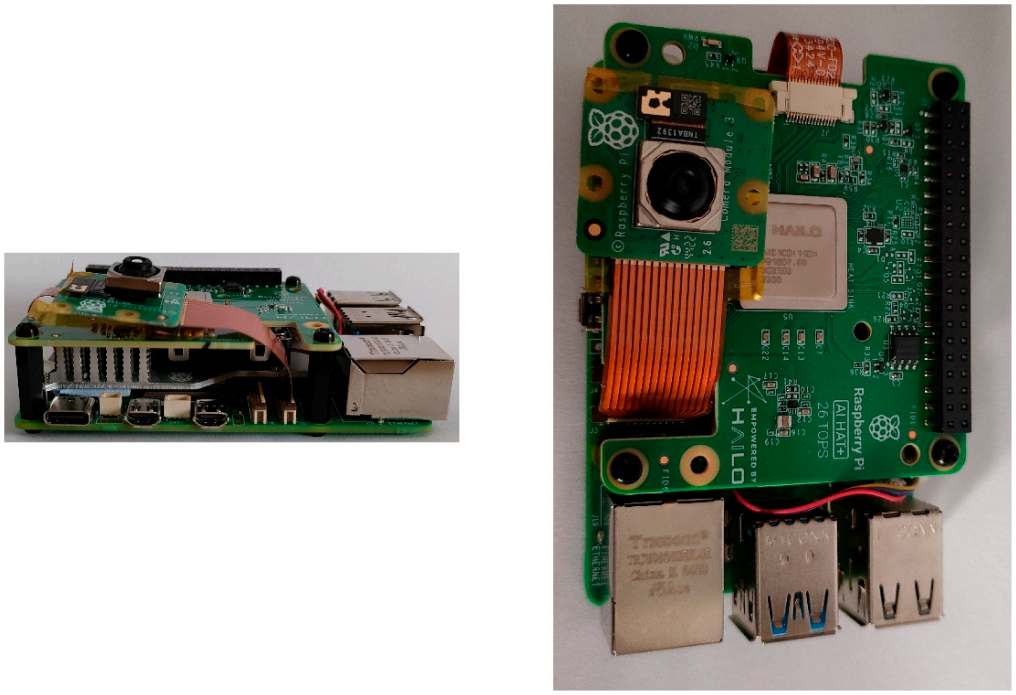


Figure 2. The proposed system.

The machine learning component includes four trained regression models. The spectral model employs an ensemble approach combining multiple regression algorithms after spectral preprocessing, while the RGB, Munsell, and hybrid models exploit complementary colour representations for SOM estimation. The trained models, together with their required preprocessing pipelines, are exported using the Joblib format and transferred to the Raspberry Pi 5 for local inference.

The deployment component implements the complete prediction pipeline on the embedded device, including image acquisition, feature extraction, input reconstruction, and SOM prediction. A lightweight Streamlit-based interface provides access to the different prediction modes, enabling interaction through a local web browser without requiring cloud connectivity or external computational resources.

3.2. Data Acquisition and Preprocessing

The first stage of the proposed methodology focuses on the acquisition and preprocessing of soil data. The dataset used in this work contains soil samples characterized

through multiple representations, including spectral measurements, RGB-derived colour features, and Munsell colour descriptors, together with their corresponding experimentally measured SOM values. Each sample is associated with the experimentally measured SOM value, which is considered the prediction target. The same samples are reused across the different modelling approaches, allowing a consistent comparison between spectral, RGB, Munsell, and hybrid representations.

The dataset used in this study consists of soil samples, each characterised by laboratory reference measurements, visible–near-infrared reflectance spectra, and a Munsell colour description. The reference variable used throughout this work is soil organic matter (SOM), determined in the laboratory and used as the prediction target for all four models. It is noted that SOM and soil organic carbon (SOC) are distinct quantities, conventionally related by a factor of approximately 1.724; all values reported in this paper refer to SOM.

The dataset used in this study is not publicly available. It was compiled within the authors' research groups and consists of soil samples for which SOM had been determined by the Walkley–Black method [8], together with the corresponding reflectance spectra and Munsell colour descriptions. As shown in Table 1, the hue letter is strongly imbalanced, with the R category represented by only 13 samples.

Table 1. Descriptive statistics of the dataset.

Variable	Unit	N	Mean	SD	Min	Max
Soil organic matter	%	6167	3.79	1.35	0.80	10.25
Munsell HUE	–	6167	3.59	3.49	0.10	10.00
Munsell VALUE	–	6167	4.31	0.40	1.10	5.90
Munsell CHROMA	–	6167	2.82	0.45	0.60	4.80
Munsell LETTER	–	6167	3 categories: Y (4478), YR (1676), R (13)			

Reflectance spectra were acquired with a Shimadzu UV-3600 spectrophotometer with an MPC-3100 Multi-Purpose Large-Sample Compartment integrating sphere, spectrometer (Shimadzu Corporation, Kyoto, Japan) over the range 250–2500 nm at 5 nm intervals, yielding 449 spectral bands per sample. The Munsell description comprises the categorical hue letter (LETTER) together with the numerical components hue (HUE), value (VALUE) and chroma (CHROMA). Samples with missing or invalid reference values were discarded, and SOM values were restricted to the physically plausible interval $(0.1, 15]$ %. Samples with incomplete spectral or Munsell records were likewise removed, so that each modelling scenario uses only fully characterised samples. Table 1 summarises the resulting dataset. The dataset was partitioned into training and test subsets using an 80/20 split.

Before model training, the dataset is divided into training and testing subsets. All preprocessing operations, feature transformations, hyperparameter selection, and ensemble weight estimation are performed exclusively using the training partition to prevent information leakage and obtain realistic estimates of model generalization. The same dataset is used to construct the four prediction scenarios, enabling a direct comparison between different sensing modalities.

For the spectral representation, reflectance values across the available wavelength range are used as input variables after the applied preprocessing pipeline. For the RGB representation, colour features are extracted from reconstructed image-based measurements. The extracted RGB values are transformed into additional colour spaces, including CIELAB, and combined with derived colour indices. Since the available camera does not provide near-infrared information, the RGB-derived vegetation index feature is fixed to zero during deployment.

For the Munsell representation, the categorical soil colour descriptor (LETTER) and numerical components (HUE, VALUE, CHROMA) are used as input variables. The hybrid representation combines RGB-derived features with Munsell descriptors to investigate whether the complementary information provided by visual imaging and human colour classification improves SOM estimation.

3.3. Colour and Feature Extraction

For RGB-based prediction, images acquired from the Raspberry Pi camera are processed by selecting a central region of interest to reduce the influence of borders, background elements, and shadows. The mean RGB values are calculated and transformed through the same colour conversion pipeline used during model development, including conversion from sRGB to linear RGB, XYZ, and CIELAB colour space. The resulting feature vector consists of eight features: the three RGB channels, the three CIELAB components, SCI and an RGB-derived vegetation index descriptor.

Since the RGB model is trained using colour features reconstructed from spectral information, while deployment obtains these features directly from RGB camera images, the two acquisition processes are not identical. Therefore, the deployed camera-based models represent a practical approximation of the original sensing process. This approximation is introduced because the available dataset provides spectral-derived colour information, whereas the deployed system relies on direct image acquisition through an RGB camera. Consequently, some differences may exist between the feature distributions during model training and inference. Factors such as camera sensor characteristics, illumination conditions, image acquisition geometry, and environmental variability can influence the extracted colour features. Nevertheless, this approach reflects the intended low-cost deployment scenario, where RGB imaging provides an accessible alternative to laboratory-grade spectral acquisition.

Additional colour descriptors are calculated from the extracted RGB values, including the soil colour index (SCI). The SCI is computed as:

$$SCI = (R - G)/(R + G)$$

Although the training dataset includes spectral information that enables calculation of vegetation-related indices, the deployed RGB camera does not capture infrared wavelengths. Therefore, the RGB-derived vegetation index descriptor is set to zero during inference. This design choice introduces a difference between training and deployment feature distributions, which is considered during the interpretation of camera-based predictions. Consequently, the RGB model is intended as a low-cost relative estimation approach rather than a replacement for laboratory-grade measurements.

3.4. Machine Learning-Based SOM Prediction

The SOM estimation problem is formulated as a supervised regression task, where the objective is to learn the relationship between soil representations and the experimentally measured SOM percentage. Four complementary models are developed to investigate the contribution of different sensing modalities.

The spectral model uses reflectance measurements across the available wavelength range as input. A preprocessing pipeline is applied to the spectral data before training, and multiple regression models are combined through an ensemble strategy to exploit the complementary behaviour of different algorithms.

The RGB model is designed specifically for deployment with the Raspberry Pi camera. It uses an eight-dimensional feature vector consisting of RGB values, CIELAB components,

SCI, and RGB-derived vegetation index information. This model represents the lowest-cost acquisition scenario, requiring only an RGB camera.

The Munsell model uses manually recorded soil colour descriptors consisting of HUE, LETTER, VALUE, and CHROMA. Since LETTER represents a categorical variable, it is encoded through a OneHotEncoder integrated into the preprocessing pipeline. The categories are extracted directly from the trained model during deployment to ensure consistency between training and inference.

The hybrid RGB–Munsell model combines automatically extracted colour features from the camera with manually obtained Munsell descriptors. The objective is to investigate whether the combination of objective image-based measurements and standardized colour classification improves SOM estimation.

After training, the complete models, including preprocessing pipelines and required metadata, are serialized using Joblib and transferred to the Raspberry Pi 5. During deployment, the corresponding model is loaded locally and used for SOM prediction through the developed web interface.

3.5. Model Deployment Preparation

To enable execution on the Raspberry Pi 5, the trained models were exported together with their preprocessing pipelines and metadata required during inference. The exported Joblib files preserve the complete machine learning workflow, including feature transformations, categorical encoding, and regression models. During deployment, the same software environment used during development is maintained to avoid compatibility issues between training and inference stages. Furthermore, the deployment environment uses the same scikit-learn version as the training environment to ensure compatibility with serialized preprocessing pipelines. All models were developed in Python (version 3.14.4; Python Software Foundation, Wilmington, DE, USA) using scikit-learn (version 1.8.0) and Joblib (version 1.5.3), and the user interface was implemented with Streamlit (version 1.57.0; Snowflake Inc., Bozeman, MT, USA).

The inference application reconstructs the required input format dynamically based on the loaded model configuration. This approach ensures compatibility between different model versions and avoids manual modification of feature names or categorical variables when switching between prediction modes.

4. System Validation and Demonstration

The developed system was validated through a combination of predictive performance evaluation and prototype-level deployment validation. The validation considered both the predictive capability of the developed models and the feasibility of executing the complete SOM estimation workflow on a Raspberry Pi 5 edge platform.

4.1. Prototype Deployment Scenario

Figure 3 illustrates the operational workflow of the proposed Soil Organic Matter estimation system. The developed prototype demonstrates how image acquisition, colour feature extraction, and machine learning inference are integrated into a single embedded platform capable of providing portable soil analysis. The process begins with the collection of a soil sample or direct observation of the soil surface, which is placed within the acquisition area of the device. A Raspberry Pi 5 equipped with a camera module is responsible for capturing the soil image under controlled conditions. The acquired image is then processed through the preprocessing stage, where the image is processed to extract the relevant soil colour information while reducing the influence of background regions and shadows.

Following preprocessing, colour features are extracted from the soil image. These features represent the visual characteristics of the soil and are used as input variables for the machine learning models. The extracted information is subsequently provided to the SOM prediction module, which estimates the organic matter content based on the trained models. The developed system supports multiple prediction approaches, including the RGB-based model designed for camera acquisition, the Munsell colour-based model, the hybrid RGB–Munsell model, and the spectral ensemble model. This enables the evaluation of different input representations while maintaining the same deployment framework. The final SOM estimation output provides the user with an immediate estimation of Soil Organic Matter content, which can support decisions related to soil monitoring and precision agriculture management.

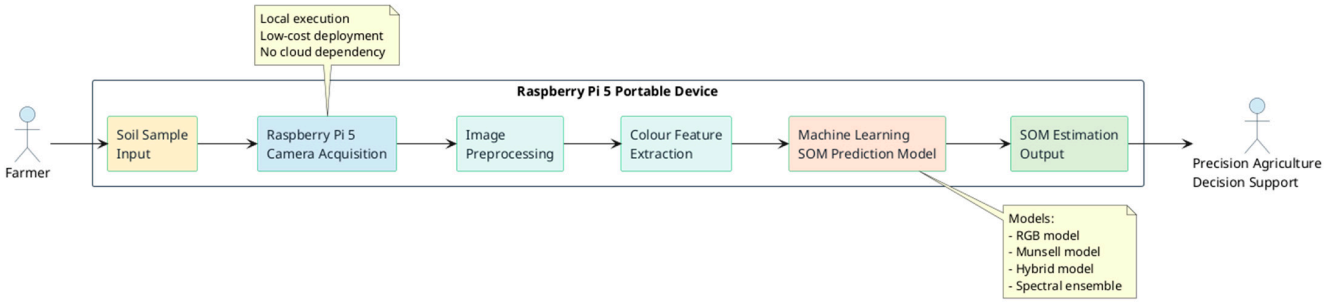


Figure 3. End-to-End Workflow of the Proposed SOM Estimation System.

4.2. Model Performance Evaluation

The predictive performance of the four developed models was evaluated using the same independent test set to enable a fair comparison between different sensing modalities. Three evaluation metrics were considered: coefficient of determination (R^2), mean absolute error (MAE), and ratio of performance to deviation (RPD). R^2 measures the proportion of variance explained by the model, MAE quantifies the average prediction error in SOM percentage units, and RPD provides an indicator of model applicability by comparing prediction error with the variability of the reference measurements.

The ratio of performance to deviation (RPD) was interpreted according to commonly adopted classification criteria for soil spectroscopy applications. In general, RPD values below 1.4 indicate that the model is not suitable for quantitative prediction, values between approximately 1.4 and 2.0 indicate moderate predictive ability, values between 2.0 and 2.5 indicate good predictive performance, and values above 2.5 indicate excellent prediction capability [15]. According to this classification, the spectral ensemble model (RPD = 2.54) achieves excellent predictive capability, while the RGB (RPD = 1.14), Munsell (RPD = 1.09), and hybrid RGB–Munsell (RPD = 1.18) models remain below the commonly accepted threshold for reliable quantitative prediction. These results highlight the trade-off between prediction performance and acquisition cost, where lower-cost colour-based approaches provide increased accessibility but reduced predictive capability compared with spectral measurements.

As seen in Table 2, the spectral ensemble achieved the highest performance, obtaining an R^2 value of 0.845 and an MAE of 0.416% SOM. This result is expected since spectral measurements provide richer information related to soil composition. In contrast, RGB, Munsell, and hybrid approaches achieved lower predictive performance but require significantly cheaper acquisition mechanisms. This behaviour is also reflected by the RPD values, where the colour-based approaches remain below the threshold commonly associated with reliable quantitative prediction.

Table 2. Predictive performance comparison of the developed SOM estimation models.

Model	R^2	MAE (% SOM)	RPD
Spectral	0.845	0.416	2.54
RGB	0.233	0.951	1.14
Munsell	0.157	0.997	1.09
Hybrid	0.277	0.919	1.18

Among the camera-oriented approaches, the hybrid RGB–Munsell model achieved slightly higher performance than the individual RGB and Munsell models. However, the observed difference in R^2 values is relatively small, and no statistical significance analysis was performed; therefore, the results should be interpreted as an indication of potential complementary information rather than as a statistically confirmed improvement. These results should be interpreted considering that the RGB-based models are designed for low-cost field deployment rather than replacing laboratory-grade spectral measurements.

4.3. Edge Deployment Validation

Beyond the predictive evaluation, the developed system was validated through the deployment of the complete SOM estimation workflow on the Raspberry Pi 5 edge platform. The objective of this validation was to demonstrate that the trained models could be transferred from the development environment and executed locally on a low-cost embedded device without requiring external computational resources.

The trained models, together with their corresponding preprocessing pipelines and required metadata, were exported from the development environment using the Joblib format and transferred to the Raspberry Pi 5. The deployment preserved the complete inference workflow, including input preparation, feature extraction, preprocessing operations, and model prediction. To ensure compatibility between the training and execution environments, the same machine learning library versions and software dependencies were maintained during deployment.

During operation, the user accesses the developed web interface through a local browser connected to the Raspberry Pi. The Streamlit-based application is hosted directly on the embedded device and is accessible through the local network (e.g., using the Raspberry Pi IP address and the application port 8501), allowing interaction from a smartphone or laptop without requiring a dedicated display or external computer. For image-based estimation, the soil image is acquired directly from the Raspberry Pi Camera Module through server-side capture, ensuring that the captured image corresponds to the intended acquisition setup. This approach avoids inconsistencies that may occur when using external cameras, such as the camera of the user's mobile device, and enables repeatable image acquisition conditions.

The deployed application supports the execution of all four prediction scenarios within the same framework: spectral prediction through uploaded spectral measurements, RGB-based prediction through camera images, Munsell-based prediction through manual colour descriptors, and hybrid prediction combining RGB and Munsell information. Therefore, the deployment validation demonstrates the successful integration of data acquisition, feature extraction, machine learning inference, and user interaction into a single autonomous edge-AI platform.

The results confirm the feasibility of executing the complete SOM estimation pipeline locally on a Raspberry Pi 5, providing a portable solution that does not depend on cloud services or external computing infrastructure.

4.4. Discussion

The obtained results highlight a clear trade-off between prediction accuracy and deployment cost in SOM estimation systems. The spectral ensemble achieved the highest predictive performance across all three metrics, with an R^2 of 0.845, an MAE of 0.416% SOM and an RPD of 2.54, placing it in the highest performance class of the classification adopted in Section 4.2. This outcome is expected, since reflectance spectra capture wavelength-dependent responses across the visible and infrared regions that are directly related to soil composition. However, this predictive advantage comes at the cost of specialised acquisition equipment, which restricts the applicability of the spectral approach in low-cost, field-oriented scenarios and motivates the investigation of simpler sensing modalities.

The lower performance of the RGB, Munsell and hybrid approaches can be attributed to the substantially reduced information content of colour-based representations. RGB imaging samples only three broad channels within the visible range and is further affected by illumination conditions, sensor characteristics, soil moisture and acquisition geometry. Munsell descriptors provide a standardised and reproducible description of soil appearance, but likewise summarise the visual impression of the sample into a small number of attributes. Neither representation captures the infrared response, which carries a substantial part of the compositional information exploited by the spectral model. A feature-importance analysis of the hybrid model supports this interpretation quantitatively: the soil colour index and the RGB-derived vegetation index descriptor together account for approximately 28% of the total feature importance, yet these are precisely the descriptors that the deployed camera either derives from broadband channels rather than narrow spectral bands, or cannot compute at all.

Among the camera-oriented approaches, the hybrid RGB–Munsell model achieved a marginally higher R^2 than the individual RGB and Munsell models. Although this may indicate that image-derived features and standardised colour descriptors carry partly complementary information, the difference cannot be considered statistically established on the present evaluation, which is based on a single test split without confidence interval analysis. From a deployment perspective, the implemented Raspberry Pi 5 platform demonstrates that a complete machine-learning workflow for SOM estimation can be transferred from a development environment to a portable edge device. Integrating image acquisition, feature extraction, model inference and user interaction within a single embedded system reduces the complexity of field operation and removes the need for continuous connectivity to external computational resources.

Several limitations remain. The camera-based models are sensitive to variations in illumination, soil moisture and acquisition geometry, and the difference between the features available during training and those obtained during real camera acquisition introduces a distribution shift that should be taken into account when interpreting predictions. Consequently, the performance reported in Table 2 for the RGB model should be regarded as an upper bound on the accuracy attainable under field conditions. Future work should therefore prioritise the acquisition of a wider range of soil samples, the introduction of a colour-calibration procedure based on a reference target, the integration of additional soil properties, and the optimisation of the models for efficient edge execution. These developments would further strengthen the suitability of the proposed platform as a practical decision-support tool for precision agriculture.

5. Conclusions

This paper presented a portable and low-cost edge-AI system for Soil Organic Matter estimation based on multiple sensing representations and machine learning techniques. The proposed framework integrates four complementary prediction approaches: a spectral

ensemble model, an RGB image-based model, a Munsell colour-based model, and a hybrid RGB–Munsell model. The complete workflow, including data processing, feature extraction, model inference, and user interaction, was implemented on a Raspberry Pi 5 platform, enabling local operation without cloud dependency. The experimental evaluation demonstrated the different capabilities of each sensing modality. The spectral ensemble achieved the highest predictive performance, obtaining an R^2 value of 0.845 and an MAE of 0.416% SOM, while the RGB, Munsell, and hybrid models provided lower-cost alternatives suitable for portable deployment scenarios. The hybrid RGB–Munsell approach achieved slightly higher performance compared with the individual colour-based models, suggesting that combining complementary soil colour representations may provide additional information. However, further evaluation using larger datasets and statistical analysis is required to confirm this observation.

The deployment validation further demonstrated the feasibility of executing the complete SOM estimation pipeline on an embedded edge platform, including image acquisition, feature extraction, machine learning inference, and user interaction through a local web interface. This implementation provides a practical foundation for portable and accessible soil monitoring solutions in precision agriculture applications. Despite the demonstrated feasibility, further work is required to improve robustness under different environmental conditions and soil characteristics. Future developments will focus on expanding the dataset, improving colour calibration mechanisms, integrating additional soil properties, and optimizing machine learning models for efficient edge execution. Overall, the proposed system demonstrates the potential of combining embedded systems, computer vision, and machine learning to develop sustainable and accessible tools for soil assessment.

Author Contributions: Conceptualization, M.F. and P.K.; methodology, M.F. and P.K.; investigation, M.F.; writing—original draft preparation, M.F. and D.V.; writing—review and editing, M.F. and D.V.; visualization, M.F.; supervision, A.P., P.B. and P.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset supporting the reported results is not publicly available. It was compiled within the authors' research groups and may be made available by the corresponding author upon reasonable request.

Acknowledgments: During the preparation of this manuscript, the authors used OpenAI ChatGPT (GPT-5.5 Thinking, accessed July 2026) to assist with manuscript structuring, language editing, formatting support, figure-layout refinement, and bibliography-positioning checks. The tool was not used to collect empirical data, generate empirical results, or perform statistical analysis. The authors reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lal, R. Soil carbon sequestration impacts on global climate change and food security. *Science* **2004**, *304*, 1623–1627. [CrossRef] [PubMed]
2. McBratney, A.; Whelan, B.; Ancev, T.; Bouma, J. Future directions of precision agriculture. *Precis. Agric.* **2005**, *6*, 7–23. [CrossRef]
3. Nelson, D.W.; Sommers, L.E. Total carbon, organic carbon, and organic matter. In *Methods of Soil Analysis: Part 3—Chemical Methods*; Sparks, D.L., Ed.; SSSA Book Series 5; Soil Science Society of America: Madison, WI, USA, 1996; pp. 961–1010.
4. Nocita, M.; Stevens, A.; van Wesemael, B.; Aitkenhead, M.; Bachmann, M.; Barthès, B.; Ben Dor, E.; Brown, D.J.; Clairotte, M.; Csorba, A.; et al. Soil spectroscopy: An alternative to wet chemistry for soil monitoring. *Adv. Agron.* **2015**, *132*, 139–159. [CrossRef]

5. Viscarra Rossel, R.A.; Minasny, B.; Roudier, P.; McBratney, A.B. Colour space models for soil science. *Geoderma* **2006**, *133*, 320–337. [CrossRef]
6. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [CrossRef] [PubMed]
7. Alpaydin, E. *Introduction to Machine Learning*, 4th ed.; MIT Press: Cambridge, MA, USA, 2020.
8. Walkley, A.; Black, I.A. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Sci.* **1934**, *37*, 29–38. [CrossRef]
9. Viscarra Rossel, R.A.; Walvoort, D.J.J.; McBratney, A.B.; Janik, L.J.; Skjemstad, J.O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **2006**, *131*, 59–75. [CrossRef]
10. Swetha, R.K.; Dasgupta, S.; Chakraborty, S.; Li, B.; Weindorf, D.C.; Mancini, M.; Silva, S.H.G.; Ribeiro, B.T.; Curi, N.; Ray, D.P. Using Nix color sensor and Munsell soil color variables to classify contrasting soil types and predict soil organic carbon in Eastern India. *Comput. Electron. Agric.* **2022**, *199*, 107192. [CrossRef]
11. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
12. Dietterich, T.G. Ensemble methods in machine learning. In *Multiple Classifier Systems*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2000; Volume 1857, pp. 1–15.
13. Kaufman, S.; Rosset, S.; Perlich, C.; Stitelman, O. Leakage in data mining: Formulation, detection, and avoidance. *ACM Trans. Knowl. Discov. Data* **2012**, *6*, 15.
14. Raspberry Pi Ltd. Raspberry Pi 5 Product Brief and Documentation. 2023. Available online: https://www.raspberrypi.com/documentation/ (accessed on 1 August 2026).
15. Williams, P.C.; Sobering, D.C. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. *J. Near Infrared Spectrosc.* **1993**, *1*, 25–32. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.