Unveiling Soil-Vegetation Interactions: Reflection Relationships and an Attention-Based Deep Learning Approach for Carbon Estimation

Dristi Datta[1, 2, *], Manoranjan Paul[1, 2], Manzur Murshed[3], Shyh Wei Teng[2, 4], and Leigh M. Schmidtke[5]

[1]School of Computing, Mathematics, and Engineering, Charles Sturt University, Bathurst NSW 2795, Australia
[2]Cooperative Research Centre for High Performance Soils, Callaghan NSW 2308, Australia
[3]School of Info Technology, Deakin University, Burwood VIC 3125, Australia
[4]Institute of Innovation, Science and Sustainability, Federation University, Mount Helen VIC 3350, Australia
[5]Gulbali Institue, Charles Sturt University, Wagga Wagga NSW 2678, Australia
[*]Correspondence: ddatta@csu.edu.au

***Abstract*—Estimating soil organic carbon (SOC) from satellite imagery, particularly in areas with both bare soil and vegetation, poses significant challenges. Traditional approaches often overlook the complex interactions between soil and vegetation. Addressing this gap, our study introduces an innovative method that leverages novel correction of hyperspectral reflections to adjust for vegetation levels, enhancing SOC estimation accuracy. Moreover, we propose an attention-based deep neural network that dynamically prioritizes spectral features crucial for SOC prediction. This mechanism significantly improves the model's ability to detect significant features for accurate SOC estimation. Comparative experiments with traditional models on a benchmark dataset demonstrate our method's effectiveness in reducing vegetation influence and accurately estimating SOC across mixed landscapes. Our findings represent a notable advancement in SOC estimation from satellite imagery, highlighting the potential of advanced learning-based techniques with attention-driven feature weighting for SOC estimation.**

***Index Terms*—Soil carbon estimation, LUCAS dataset, remote sensing, attention-based deep learning, Landsat 8**

I. Introduction

Soil organic carbon (SOC) is vital for agricultural productivity and ecosystem health, yet traditional field sampling methods are time-consuming and impractical for large-scale analysis. Remote sensing, particularly lab-based hyperspectral imaging (HSI), offers an efficient alternative for SOC estimation, providing detailed spatial data that enables accurate soil property analysis without direct sampling [1]. Machine learning models leveraging HSI data have shown promising results in SOC estimation [2], [3]. Despite HSI's advantages, its high cost and accessibility issues prompt some researchers to explore visual band multispectral data, which also achieves comparable SOC estimations at a lower cost [4]. HSI and visual band cameras capture information in specific land portions, leading researchers to focus on satellite image data for large-scale SOC estimation, enabling nation-scale forecasts.

Multi-spectral satellite data like Sentinel 2 (S2) and Landsat 8 (L8) provide broad coverage and accessibility for SOC estimation; however, they face challenges in accuracy compared to hyperspectral data, primarily due to environmental noise and the need for extensive pre-processing [5]–[8]. While some studies have made progress in soil property mapping using machine learning under varied agricultural practices, they mainly focus on broad attributes without delving into the nuanced relationship between bare and vegetated soil, relying on traditional models [9]. Most research concentrates on bare soil SOC estimation, with minimal vegetation cover significantly biasing reflectance characteristics in satellite images, thus hindering accurate SOC estimations by learning-based models [10].

This study focuses on estimating SOC in fields with vegetated crops, mainly on barley-covered lands. To the best of our knowledge, this is the first time we have explored the reflection relation from bare and barely vegetated crops, and learning-based models have been introduced to estimate SOC levels. The proposed approach begins by unveiling the reflection characteristics of L8 satellite images, capturing various SOC levels in both bare soil and barley-vegetated crops. Introducing a correction coefficient derived from these reflections mitigates the impact of vegetation on soil surfaces, enabling satisfactory estimations by learning-based models. This study used the publicly available Land Use and Coverage Area Body Survey (LUCAS) 2018 dataset [11] and corresponding L8 satellite images.

A number of learning-based models, such as random forest (RF), support vector regressor (SVR), CatBoost regressor (CBR), and one-dimension convolution neural network (1dCNN), commonly described in previous literature [2], [4], have been used to analyze the estimation performance of SOC from L8 reflectance data including different vegetation and soil indices. Due to their inherent structures, RF and CBR may face challenges capturing intricate spatial dependencies and patterns within L8 data. SVR may struggle with satellite data's high-dimensional and complex nature, especially when spatial relationships are pivotal. Although proficient in capturing sequential patterns, 1dCNNs may only partially exploit the inherent spatial relationships present in satellite images. These limitations underscore the need for more advanced approaches to comprehend better the nuanced spatial and spectral characteristics inherent in satellite data for satisfactory

TABLE I
DESCRIPTIVE STATISTICAL PARAMETERS FOR THE SOC DATASET INVESTIGATED IN THE STUDY

Soil type	NDVI range	S. No.	Min	Max	Mean	Median	Std.	CV(%)
Bare soil	$0 < NDVI < 0.25$	378	2.2	29.9	12.57	11.60	6.35	50.51
Unhealthy Barley	$0.25 < NDVI < 0.33$	89	3.9	29.9	13.09	12.20	6.29	48.05
Moderate Healthy Barley	$0.33 < NDVI < 0.66$	201	2.5	29.5	14.10	12.80	6.56	46.52
Healthy Barley	$0.66 < NDVI < 1$	70	2.7	29.7	15.67	14.80	7.03	44.86
All Barley	$0.25 < NDVI < 1$	360	2.5	29.9	14.15	12.85	6.63	46.85
Bare soil + Barley	$0 < NDVI < 1$	738	2.2	29.9	13.34	12.30	6.53	48.95

SOC estimation.

To address the limitations of the conventional learning-based models, this research also proposes an attention-based deep neural network for estimating SOC from L8 data. This network's motivation is how humans selectively focus on specific information when processing complex inputs. The purpose of an attention layer is to specifically focus on different input features, assigning different importance weights to each element. This attention-based approach is expected to improve the accuracy and robustness of SOC estimations, particularly in areas with varying vegetation cover and environmental conditions. Experimental results show that utilizing the proposed attention-based deep neural network with proposed correction coefficients can reduce vegetation impacts and estimate SOC from barley-vegetated crops satisfactorily.

II. DATASET PREPARATION

In 2001, LUCAS was established by the Statistical Office of the European Union (EUROSTAT) to create a comprehensive pan-European database focused on landscape parameters crucial for assessing agricultural and environmental coverage [11]. The LUCAS 2018 database contains 18,984 soil sample records across Europe. The LUCAS-SOIL-2018 CSV file includes a unique identification code (PointID) and eight corresponding soil properties: pH(CaCl2), pH(H2O), electrical conductivity (EC), SOC, carbonate content (CaCO3), phosphorus (P), total nitrogen (N), extractable potassium (K). The file named LUCAS-SOIL-2018.shp provides the theoretical coordinates of the soil samples. For this analysis, we focus on SOC for bare soil samples and barley vegetation at 0-20 cm depth. The statistical description of the considered dataset is provided in Table I.

For satellite data, we utilized L8 imagery, available through the USGS Earth Explorer website (https://earthexplorer.usgs.gov/). Images corresponding to the soil sample locations and dates within a 15-day window and under 10% cloud cover were selected, covering 11 bands as detailed in [12]. Although L8 covers 11 bands, only 7 bands were considered for analysis due to their established relevance to SOC estimation, as demonstrated in prior research [7]. The remaining four bands, which include thermal and cirrus bands, were excluded because they are less pertinent to SOC prediction and more suited to other applications such as temperature measurement and atmospheric studies [13]. These images underwent radiometric calibration, atmospheric correction, and pansharpening in ENVI software, following published protocols [14]. Reflectance values were then extracted via ArcGIS Pro for analysis.

III. PROPOSED CORRECTION COEFFICIENT

Remote sensing involves measuring the object's reflectance to understand the materials' characteristics. This study divided the dataset into six classes depending on the SOC range for bare soil and barley vegetation. We also classify different stages of barley depending on the NDVI indexing reported in Table I for a better understanding of the reflection relationships.

The average reflectance of the same class SOC for bare soil and barley vegetated soil samples has been drawn in Fig. 1. If the SOC values lie in the lower range, higher reflection is noted, and vice versa. A noticeable pattern has been observed from the reflectance of these four sub-figures, from bare to vegetated soil. L8 reflectance has the most variation due to the barley vegetation, especially from band four (B4) to band seven (B7), on the top of the soil. Therefore, to estimate SOC from satellite data, these vegetation effects need to be handled before learning-based training.

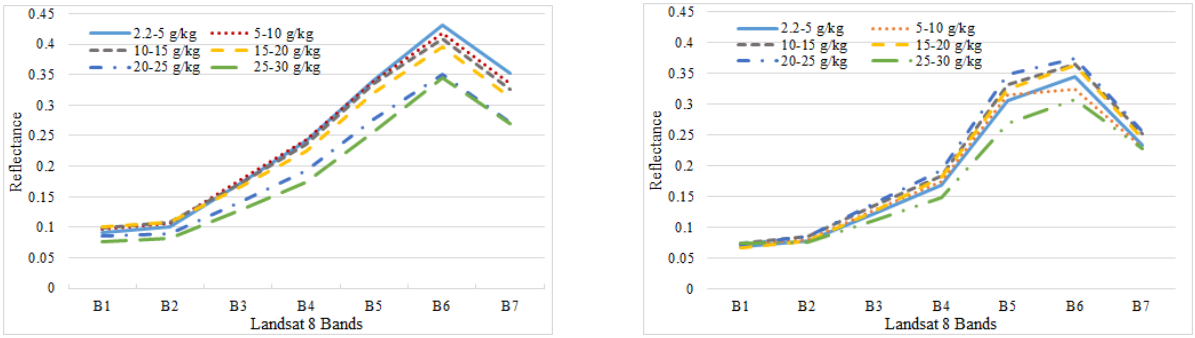
A number of methods have been considered to minimize the vegetation effect. First, we tried to apply linear regression to minimize the vegetation impacts; however, this method did not work satisfactorily as the data was not linearly separable, especially for Fig. 1c. In the second step, the polynomial regression curve fitting approach has been considered up to four degrees of polynomial equations. This approach helps capture the trend of complex data structures. However, this method fails to differentiate the change from moderate to healthy vegetation (Figs. 1c to 1d). This approach becomes more complex and has a good chance of overfitting the data.

After observing those characteristics and minimizing the complexity, a simple and straightforward approach that is more suited to minimizing vegetation effects has been adopted. A novel correction coefficient matrix has been proposed using a difference model to minimize the vegetation effects on the top of the soil from Equation 1.

$$R_b = R_v + \epsilon \tag{1}$$

where R_b and R_v are the average bare soil and vegetation reflection, and ϵ is the correction coefficient, respectively.

According to Equation 1, our observations illustrated in Fig. 2 are considered correction coefficient matrices for different stages of barley vegetation. After that, according to the SOC range and vegetation scenario, the reflectance of seven bands



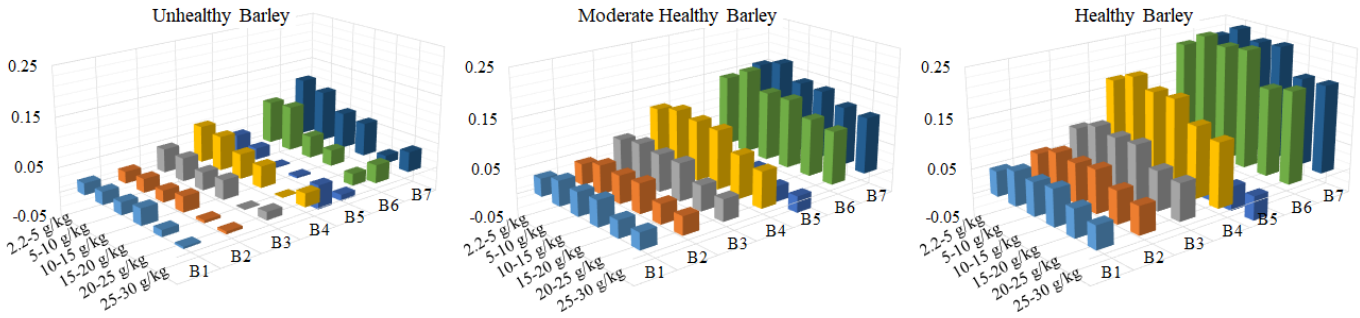
(a) Bare soil

(b) Unhealthy barley

(c) Moderate healthy barley

(d) Healthy barley

Fig. 1. Average reflection for bare soil and barley vegetation in different carbon range variation.



(a) Unhealthy barley

(b) Moderate healthy barley

(c) Healthy barley

Fig. 2. Correction coefficient for unhealthy, moderate healthy, and healthy barley vegetation.

of L8 has been updated according to the correction coefficient, and vegetation impact has been minimized.

The proposed correction coefficient matrix values have six different SOC classes and three vegetation scenarios, and the NDVI index can only classify the condition of vegetation. That may need to be clarified when choosing the proper correction index. To address the issue of selecting the most appropriate correction index for samples, the nearest bare soil location within an image can be identified, and a learning-based model can estimate the SOC values from bare soil and help choose a proper correction index.

IV. Proposed Attention-Based Network

Our proposed attention-based network, meticulously structured to process and interpret L8 satellite imagery for SOC estimation, is specifically designed to tackle the intricacies of SOC estimation. Utilizing an attention mechanism, the model adeptly navigates the diverse data layers of L8 imagery, which include multi-spectral imagery across seven bands, alongside derived soil and vegetation indices and the TCT, each providing unique insights into the Earth's surface. This approach dynamically prioritizes features directly relevant to SOC content, enabling the identification and exploitation of crucial spatial dependencies and subtle patterns often overlooked by traditional methods. By concentrating on the most significant features, our model aims to significantly enhance the accuracy and interpretability of SOC predictions, establishing a new benchmark for satellite-based SOC estimation techniques, as illustrated in Fig. 3 and the network architecture unfolds as

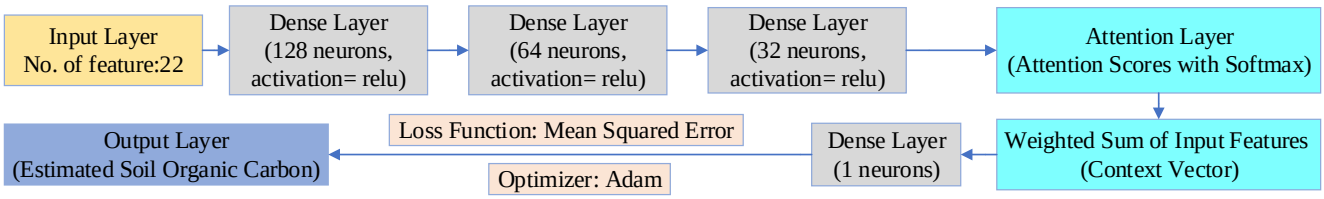


Fig. 3. Schematic representation of the proposed attention mechanism in the neural network architecture.

follows:

Input Layer: This initial layer receives the L8 reflectance bands, vegetation indices, and environmental features as input. It forms the foundation of our network, setting the stage for complex pattern recognition and analysis.

Dense Layers: Following the input layer, the network architecture includes a sequence of densely connected layers with ReLU (Rectified Linear Unit) activation functions. These layers, arranged hierarchically with 128, 64, and then 32 neurons, are designed to process increasingly complex representations of the input data. The output of each layer, D_j, is computed as:

$$D_j = \text{ReLU}(W_j D_{j-1} + b_j), \quad j = 1, 2, 3 \tag{2}$$

where D_j is the output, W_j the weights, and b_j the bias of the j^{th} layer, allowing the model to effectively uncover intricate patterns essential for SOC estimation. This hierarchical design enhances the network's ability to discern complex spatial dependencies and attributes within the data.

Attention Layer: This layer utilizes an attention mechanism to selectively emphasize the most relevant features for SOC estimation. By calculating attention scores using a dot-product operation between the input features from the last dense layer, $D_{n=3}$ (where n is the number of layers), and a trainable query vector, $\boldsymbol{q}$, the network dynamically focuses on critical information:

$$\boldsymbol{\alpha} = \text{softmax}(\boldsymbol{q}^T D_{n=3}) \tag{3}$$

The computed scores, $\boldsymbol{\alpha}$, effectively determine the importance of each feature, enabling the model to prioritize and process bands or indices crucial for SOC estimation with enhanced accuracy. This mechanism significantly improves the model's ability to make data-driven decisions based on feature relevance.

Context Vector and Output: The attention mechanism's outcome is synthesized into a context vector $\mathbf{C}$, a weighted sum of pivotal features $\mathbf{A}$ and the last dense layer's output D_3, effectively capturing the most relevant information for SOC estimation:

$$C = \mathbf{A} + D_{n=3} \tag{4}$$

Leveraging this comprehensive context, a final dense layer with a linear activation function computes the SOC estimation, $\hat{y}$, as follows:

$$\hat{y} = W_o C + b_o \tag{5}$$

In this equation, W_o and b_o represent the weight and bias parameters of the output layer, respectively. These parameters are fine-tuned during the training process to accurately map the context vector $\mathbf{C}$ to the estimated SOC values. This final step ensures the model's estimation are directly influenced by the most salient features identified through the attention mechanism, enhancing the accuracy and interpretability of SOC estimations.

Optimization and Loss Function: To facilitate efficient learning, the model employs the Adam optimizer, renowned for its effectiveness in handling sparse gradients and adaptive learning rate adjustments. Coupled with this optimization strategy is the use of Mean Squared Error (MSE) as the loss function, which quantifies the difference between the actual SOC values and the model's estimation. The MSE is mathematically represented as:

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 \tag{6}$$

where m is the number of samples, y_i denotes the actual SOC value for the i^{th} sample, and $\hat{y}_i$ is the corresponding estimated SOC value by the model. This loss function ensures that the training process is geared towards minimizing the estimation errors, refining the model's accuracy over time.

The incorporation of these optimization and loss calculation techniques into our model's architecture specifically addresses the challenges posed by SOC estimation from L8 satellite data. By introducing key innovations such as dynamic feature weighting through an attention mechanism, capturing complex spatial dependencies, and fine-tuning the model with the Adam optimizer and MSE loss function, our approach significantly enhances prediction accuracy and model interpretability. These advancements mark a notable progression beyond the capabilities of traditional deep learning and machine learning models, highlighting our model's sophistication and potential in providing accurate SOC estimations.

V. Methodology

This section presents a streamlined methodology for SOC estimation using the LUCAS 2018 dataset and L8 satellite imagery (Fig. 4). Initially, L8 images corresponding to bare and barley soil samples were processed through radiometric calibration, atmospheric correction, and pansharpening to a 15m resolution. Reflectance values were analyzed to observe the relationship between bare soil and barley vegetation, leading to the development of three correction matrices based on vegetation health. These matrices were applied to adjust for vegetation impact.

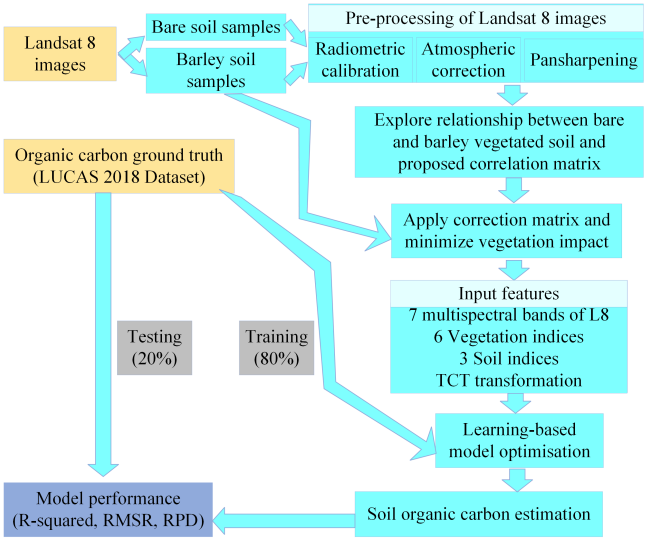


Fig. 4. Illustration of the regression framework employed in the proposed investigations.

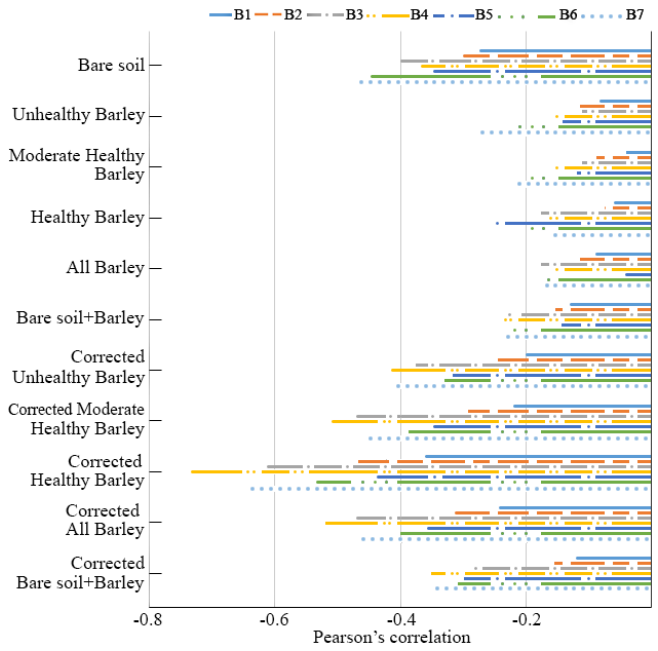


Fig. 5. Pearson's correlation analysis comparing raw and corrected Landsat 8 satellite data reflections across the seven bands (B).

Input features included seven L8 bands, six vegetation indices (RVI, NDVI, GNDVI, EVI, SAVI, MSAVI), three soil indices (BI, SI, CI), and TCT-transformed bands, totaling 22 features for model development. An attention-based deep neural network was proposed, and its SOC estimation performance was compared against three ML models (RF, SVR, CBR) and a 1dCNN. Model implementation and hyperparameter optimization were conducted using scikit-learn and a five-fold grid-search method. The dataset was split into training (80%) and testing (20%) sets, with model performance evaluated using R^2, $RMSE$, and RPD metrics, detailed mathematically in R. Reda et al. [15].

VI. Results and Discussion

This section presents the SOC estimation results derived from L8 satellite data and corresponding ground truth from the LUCAS 2018 dataset. Our investigation commenced with a comprehensive Pearson's correlation matrix analysis to decipher the intricate relationships between each L8 band and SOC. Fig. 5 visually presents the correlation findings, revealing a negative correlation trend with SOC across raw bands. Notably, bare soil exhibited significantly higher correlation values than barley plants' various stages. However, Pearson's correlation has significantly increased when the different stages of barley have transformed with the proposed correction coefficients matrix (Fig. 5), suggesting the potential of learning-based models for SOC estimation even in barley-vegetated fields.

In the second stage of analysis, we compared the estimation performance of our novel attention-based deep learning (AM) model with that of established counterparts (Table II), including 1dCNN and conventional ML models (RF, SVR, and CBR). To make the learning-based models robust, we considered a five-fold cross-validation technique to eliminate underfitting and over-fitting data and understand the estimation performance of unseen data. The paper [7] introduced novel indices and used those indices and other features to estimate SOC for bare soil. As it is a relevant and recent method, we have also compared the performance of the proposed method against this technique. However, our proposed AM model significantly outperforms the method in [7], where $R^2 = 0.54, RMSE = 4.42, RPD = 1.43$.

The estimation performance of learning-based models significantly decreases even with minimal vegetation on the top of the soil, as the reflectance of the L8 satellite is highly impacted due to vegetation and exhibits $R^2 = 0$, indicating that learning-based models do not provide any explanatory power for the observed variability in the dependent variable. Therefore, estimating SOC from the remote sensing data with the vegetation becomes challenging. Despite a performance decline across models, our proposed correction matrix was pivotal in maintaining a comparable estimation performance for all considered models. The AM model provides the best estimation results for the different stages of barley plants (corrected bands). The estimation performance of AM decreases slightly when all types of corrected barley soil samples are considered together ($R^2 = 0.51, RMSE = 5.07, RPD = 1.41$). Finally, when corrected bare soil and all barley soil samples are considered, AM provides the best SOC estimation with $R^2 = 0.46, RMSE = 5.45, RPD = 1.35$ compared to other models.

Implementing our novel correction coefficient matrix significantly improved SOC estimation accuracy by mitigating vegetation impact, highlighting our attention-based neural network's unique ability to assign weights to crucial features dynamically. This dynamic feature weighting is a distinctive aspect of our model that contributed to its superior perfor-

TABLE II
REGRESSION RESULTS (R^2, $RMSE$, RPD) OF DL AND ML MODELS IN ESTIMATING SOIL ORGANIC CARBON

Soil property	AM	1dCNN	RF	SVR	CBR
Bare soil	**0.54, 4.42, 1.43**	0.52, 4.43, 1.42	0.49, 4.54, 1.40	0.45, 4.60, 1.38	0.52, 4.45, 1.42
Bare soil (R^2, $RMSE$) [7]	—	—	0.37, 5.29 [7]	0.45, 5.79 [7]	0.52, 5.25 [7]
Unhealthy Barley	0, 5.83, 0.97	0, 7.35, 0.77	0, 6.56, 0.86	0, 5.87, 0.96	0, 6.23, 0.91
M. Healthy Barley	0, 6.02, 0.90	0, 7.30, 0.74	0, 6.29, 0.86	0, 6.36, 0.85	0, 6.17, 0.88
Healthy Barley	0, 7.22, 0.94	0, 8.03, 0.85	0, 8.74, 0.78	0, 7.54, 0.90	0, 9.79, 0.70
All Barley	0, 7.25, 0.98	0, 7.90, 0.90	0, 7.25, 0.98	0, 7.22, 0.98	0, 7.39, 0.96
Bare soil and Barley	**0.19, 6.35, 1.05**	0.13, 6.61, 1.01	0.17, 6.44, 1.03	0.14, 6.53, 1.02	0.11, 6.31, 1.05
C. Unhealthy Barley	**0.34, 4.47, 1.23**	0.30, 4.51, 1.18	0.28, 4.81, 1.18	0.33, 4.64, 1.22	0.29, 4.79, 1.17
C. M. Healthy Barley	**0.44, 4.32, 1.37**	0.40, 4.46, 1.29	0.35, 4.64, 1.24	0.43, 4.35, 1.32	0.38, 4.52, 1.27
C. Healthy Barley	**0.53, 4.63, 1.47**	0.48, 4.95, 1.38	0.43, 5.14, 1.33	0.50, 4.67, 1.42	0.48, 4.93, 1.38
C. All Barley	**0.51, 5.07, 1.41**	0.43, 5.49, 1.32	0.41, 5.42, 1.31	0.45, 5.39, 1.35	0.33, 5.96, 1.22
C. Bare soil and Barley	**0.46, 5.45, 1.35**	0.41, 5.59, 1.30	0.38, 5.67, 1.28	0.36, 5.71, 1.25	0.31, 5.77, 1.22

M. = Moderate, C. = Corrected.

mance over conventional DL and ML models. Acknowledging the necessity of a critical examination, we employed a rigorous five-fold cross-validation to affirm our model's robustness, ensuring its reliability amidst vegetation cover challenges. Despite its success, our study's dependence on the LUCAS 2018 dataset and specific preprocessing methods might affect its broader applicability. By thoroughly documenting our approach, we aim to enable reproducibility and inspire future research to broaden our findings across diverse environmental conditions and datasets, pushing forward the precision and utility of SOC estimation in vegetated landscapes.

VII. CONCLUSION

This study explores the possibility of estimating SOC from the barley-vegetated cover fields from the L8 satellite data with a ground truth LUCAS 2018 dataset. The proposed correction coefficient matrix successfully minimizes the vegetation effects on the top of the soil, and satisfactory estimation performance is observed. A comparable study of popular DL and ML models has been investigated to understand the SOC estimation performance. An attention-based deep neural network has also been proposed and is crucial in assigning dynamic weighting to the most significant feature bands. Our proposed model offers a promising framework for SOC estimation, addressing the challenges posed by the variability in L8 reflectance data. This work can be further extended to understand different vegetation's reflectance behavior and estimate other soil contents.

ACKNOWLEDGMENT

This work has been supported by the Cooperative Research Centre for High Performance Soils whose activities are funded by the Australian Government's Cooperative Research Centre Program and Charles Sturt University, Australia.

REFERENCES

[1] Y. Li, C. Chang, Z. Wang, and G. Zhao, "Remote sensing prediction and characteristic analysis of cultivated land salinization in different seasons and multiple soil layers in the coastal area," *International Journal of Applied Earth Observation and Geoinformation*, vol. 111, p. 102838, 2022.

[2] D. Datta, M. Paul, M. Murshed, S. W. Teng, and L. Schmidtke, "Soil moisture, organic carbon, and nitrogen content prediction with hyperspectral data using regression models," *Sensors*, vol. 22, no. 20, p. 7998, 2022.

[3] X. Ge, J. Ding, X. Jin, J. Wang, X. Chen, X. Li, J. Liu, and B. Xie, "Estimating agricultural soil moisture content through uav-based hyperspectral images in the arid region," *Remote Sensing*, vol. 13, no. 8, p. 1562, 2021.

[4] D. Datta, M. Paul, M. Murshed, S. W. Teng, and L. Schmidtke, "Comparative analysis of machine and deep learning models for soil properties prediction from hyperspectral visual band," *Environments*, vol. 10, no. 5, p. 77, 2023.

[5] T. Zhou, Y. Geng, C. Ji, X. Xu, H. Wang, J. Pan, J. Bumberger, D. Haase, and A. Lausch, "Prediction of soil organic carbon and the c: N ratio on a national scale using machine learning and satellite data: A comparison between sentinel-2, sentinel-3 and landsat-8 images," *Science of the Total Environment*, vol. 755, p. 142661, 2021.

[6] C. B. Pande, S. A. Kadam, R. Jayaraman, S. Gorantiwar, and M. Shinde, "Prediction of soil chemical properties using multispectral satellite images and wavelet transforms methods," *Journal of the Saudi Society of Agricultural Sciences*, vol. 21, no. 1, pp. 21–28, 2022.

[7] D. Datta, M. Paul, M. Murshed, S. W. Teng, and L. M. Schmidtke, "Novel dry soil and vegetation indices to predict soil contents from landsat 8 satellite data," in *2023 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2023, pp. 152–159.

[8] O. Yuzugullu, N. Fajraoui, A. Don, and F. Liebisch, "Satellite-based soil organic carbon mapping on european soils using available datasets and support sampling," *Science of Remote Sensing*, vol. 9, p. 100118, 2024.

[9] F. Kaya, A. Keshavarzi, R. Francaviglia, G. Kaplan, L. Başayiğit, and M. Dedeoğlu, "Assessing machine learning-based prediction under different agricultural practices for digital mapping of soil organic carbon and available phosphorus," *Agriculture*, vol. 12, no. 7, p. 1062, 2022.

[10] N. Tziolas, N. Tsakiridis, S. Chabrillat, J. A. Demattê, E. Ben-Dor, A. Gholizadeh, G. Zalidis, and B. Van Wesemael, "Earth observation data-driven cropland soil monitoring: A review," *Remote Sensing*, vol. 13, no. 21, p. 4439, 2021.

[11] A. Orgiazzi, C. Ballabio, P. Panagos, A. Jones, and O. Fernández-Ugalde, "Lucas soil, the largest expandable soil dataset for europe: a review," *European Journal of Soil Science*, vol. 69, no. 1, pp. 140–153, 2018.

[12] D. P. Roy, M. A. Wulder, T. R. Loveland, C. E. Woodcock, R. G. Allen, M. C. Anderson, D. Helder, J. R. Irons, D. M. Johnson, R. Kennedy *et al.*, "Landsat-8: Science and product vision for terrestrial global change research," *Remote sensing of Environment*, vol. 145, pp. 154–172, 2014.

[13] J. A. Barsi, J. R. Schott, S. J. Hook, N. G. Raqueno, B. L. Markham, and R. G. Radocinski, "Landsat-8 thermal infrared sensor (tirs) vicarious radiometric calibration," *Remote Sensing*, vol. 6, no. 11, pp. 11 607–11 626, 2014.

[14] D. Phiri, J. Morgenroth, C. Xu, and T. Hermosilla, "Effects of pre-processing methods on landsat oli-8 land cover classification using obia and random forests classifier," *International journal of applied earth observation and geoinformation*, vol. 73, pp. 170–178, 2018.

[15] R. Reda, T. Saffaj, B. Ilham, O. Saidi, K. Issam, L. Brahim *et al.*, "A comparative study between a new method and other machine learning algorithms for soil organic carbon and total nitrogen prediction using near infrared spectroscopy," *Chemometrics and Intelligent Laboratory Systems*, vol. 195, p. 103873, 2019.