Scientific Research reports

Https://sciresearchreports.com ISSN: 3014-4042 (Print) — 3104-4050 (Online) Editor@sciresearchreports.com

EVALUATING SOIL CARBON STOCKS IN AGROFORESTRY SYSTEMS USING DEEP LEARNING AND FIELD SAMPLING

Ahmed Mukhtar[1*], Rashid Khan[2], Muhammad Arif[3]

[1]College of Agronomy, Northwest A&F University, Yangling, 712100, China,

[2]Horticulture Section, Agriculture Research Institute, Dera Ismail Khan-29050, Khyber Pakhtunkhwa, Pakistan,

[3] Agriculture Research Institute Dera Ismail Khan-29050, Pakistan

*Corresponding Author E-mail: ah.mukhtar85@gmail.com

Abstract: To answer this question, we require that the amount of carbon in the soil be well known to determine how well it could assist in dealing with climate change and how to go about using the land in a manner that is amicable to nature. An integrated approach that involves deep learning, the field sampling of soils, and remote sensing are applied to predict soil organic carbon (SOC) in various types of agroforestry landscapes in the study. The method of collection of the soil samples used was a stratified random sampling due to different depths and soils that are used as a method of management. Thereafter, we measured the SOC concentration by conventional laboratory examination. A deep convolutional neural network (CNN) model integrated data of high resolutions Sentinel-2 and LiDAR with environmental conditions, such as conditions of climate and topography. The CNN performed superiorly to both traditional models such as Random Forest, K-nearest neighbours. It also had lower RMSE / MAE and higher R 2, and that indicated that it was more effective as a predictive model. Temperature, precipitation, elevation, and vegetation indices were the most important, with values of SHAP sensitivity analysis showing this. The validated model developed spatial SOC maps indicating small-scale variations on the study area on where carbon is abundant and where carbon can be stored. The natural deep learning strategy also cost less of time and money to the field than the conventional strategies even though it remains quite ecologically valid, since it was also tested on real world data. These findings indicate that AI-based models may be used to assist with carbon tracking, improve carbon accounting models and contribute to specific agroecology interventions. This research demonstrates the paradigm that may be employed repeatedly to enhancing the environmental assessment by opportunities of the new technologies. It provides policy-makers, conservationists, and land managers with powerful input in the era of digital and climate-smart agriculture.

Article History

Received:

January 10, 2023

Revised:

February 22, 2023

Accepted:

March 09, 2023

Available Online:

June 30, 2023

Keywords Soil Organic Carbon, Agroforestry Systems, Convolutional Neural Networks, Deep Learning, Remote Sensing, Carbon Mapping

INTRODUCTION

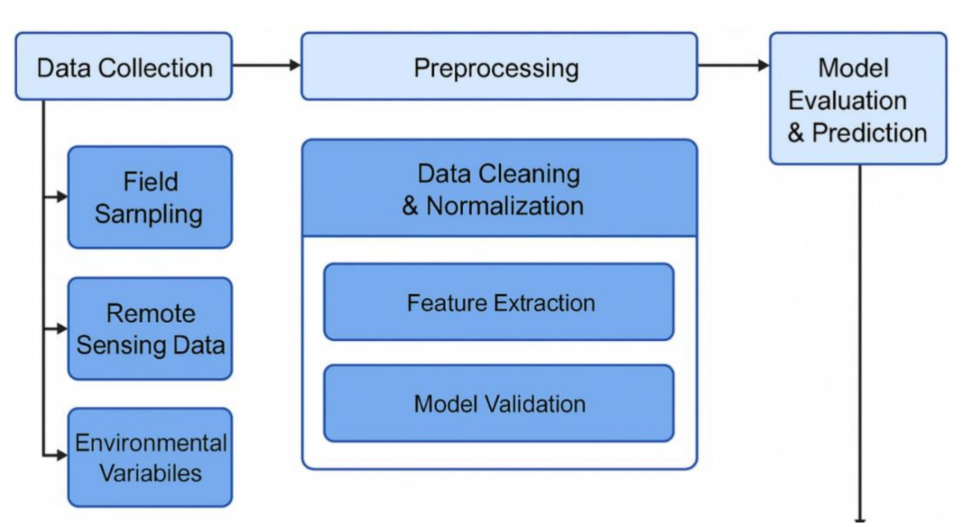
Agroforestry is a hope of enhancing soil carbon storage besides providing other ecological services as well as economic advantages. It integrates woods and bushes and crops or poultry (Soazafy et al., 2021). In order to determine how effective agroforestry systems can be in the fight against climate change and how to manage the land in the environment-friendly way, we require to know and quantify the quantity of the carbon in the soil (Pralhad et al., 2020). The carbon content of soil can be performed in two ways mostly by doing the testing by field sampling and by using a lab. These methods are precise, although time consuming, cumbersome, and costly when applied in expansive areas (McGuire et al., 2022). The alternative, which is less costly and more common is the remote sensing approaches, but vegetation cover, soil types, and sensor resolution can be relevant factors in lessening the accuracy (Wan et al., 2023). Another term used in artificial intelligence known as deep learning emerged to be a powerful means of analyzing complex information and harvesting useful patterns. This may enhance soil carbon review at the agroforestry system. The introduction of digital technology, contemporary communication networks and upgradations in variable rate technologies has led to a new dawn in the generation of automated, scalable and combined measures of soil quality. This demonstrates the possibility of the transformation of soil science through the facilitation of technology (Obade & Gaya, 2021). Due to natural variations in the storage of soil carbon and numerous factors influences that affect the soil carbon storage, further studies are required (Wan et al., 2023). The current research investigates the ways of integrating deep learning algorithms and field sample data in order to provide an improved and efficient method of assessment of soil carbon stocks in agroforestry systems. The rationale behind this is that the conventional approaches do not lack issues. Problems of land use and land cover categorisation are now solved with deep learning methods. This demonstrates the fact that they can process complex environmental data and are adaptive and efficient (Vali et al., 2020).This work aims to develop and validate a deep learning-based model to estimate soil carbon stocks using a combination of remote sensing records, environmental indicators and ground-based measurements. The performance of the model will be compared with that of older methods and we will check whether this model can be used to map and monitor wide scale soil carbon performance. This paper presents a comprehensive approach to include remote sensing data into deep learning algorithms to help the soil carbon stock prediction become more precise (Zhang & Wang, 2024). AI can be useful to look at pictures and help farmers to make faster decisions and better decisions related to how to tackle pests and diseases. It can increase soil carbon sequestration indirectly because, by making the plants healthier and more productive, it will make them work to its advantage (Javidan et al., 2025). There is much potential in helping to make decisions with the use of AI to enable better agroforestry methods and the storage of more carbon in the soil. This approach requires various procedures with the first one being collecting as much data as possible such as field sampling of soil carbon, images with high resolutions of remote sensing and compiling all necessary environmental variables. This will be needed to train and test the deep learning model. The area of study will be selected depending on the prevalence of agroforestry systems, accessibility in order to proceed there to do field sampling, and previous records. Stratified random sampling design will be used to obtain soil samples on the typical agroforestry systems. At that we will consider such aspects as the kind of trees, crops as well as management techniques involved. The soil will be composed of soil samples of various depths which will be placed into a normal laboratory procedure in order to determine the amount of organic carbon in the soil. Multispectral and LiDAR remote sensing images, or high-resolution images of satellite or airborne platforms, will be taken to obtain detailed information on vegetation, topography, and soil qualities. The publically available databases will be used or remote sensing data will provide us with the details of the environment, such as temperature and precipitation information (climate data), type of soil, and elevation.We shall train a deep learning model, which will be either a convolutional neural network or a recurrent neural network, to predict the amount of carbon that is in the soil based on the data that we take. The selection of the architecture is quite a crucial thing as it greatly influences the effectiveness of deep learning strategies (Meghraoui et al., 2024). The model will be checked against how accurate and generalisable it is, by training on some of the field sample data on a separate invented dataset. Hyperparameter tuning and model optimisation will be nailed down to obtain best possible predictions made by the model. Other indicators, which will show the effectiveness of the model and help it compare with the known methods of quantifying carbon content in the soil, include root mean squared error, mean absolute error, and coefficient of determination. The random forest was applied to understand the effectiveness of the GAS technique and make an analogy with K-nearest neighbours and the fuzzy C-means clustering sampling at different sample sizes (Wu et al., 2025). With the help of this comparison, we will be able to determine to what extent the deep learning approach is more accurate and efficient. We are going to use the tested deep learning to generate maps detailing the location of the soil carbon stock in the area of the study. The land use and carbon accounting will be done using these maps which shall be very useful. Further research is required to discover the number of soil samples and the optimal locations to sample in order to train and validate a model most successfully. Sentinel-2, soil moisture products of Sentinel 1 image and Sentinel 2 images, SOC measurement data, and other related environmental variable data that forms the digital elevation models, lithology maps, and the airborne gamma- ray spectrometry data could help to make soil carbon maps more precise based on their use of Sentinel-2 temporal mosaics of bare soil over 6 years (Wan et al., 2023). We are going to test the sensitivity of the model to the various input factors and the extent to which the model is able to present an indication of spatially-varying soil carbon.

METHODOLOGY

The present work employs an experimental approach based on mixed methods that integrates a quantitative deep learning model with qualitative data validation in the field to analyse soil carbon reserves on agroforestry systems. The methodology has three sections and these are: data collection, deep learning model development, and testing and application of the model. Such a unified solution endeavours to harness the predictive capability of artificial intelligence but ensures that the predictions are grounded on the measures that have been empirically validated to ensure the ecological validity and relevance of the prediction. The initial trend was to collect much information in a study site mainly consisting of agroforestry because it possessed much same land-use system, could be easily, reached and its historical environmental documentation was recorded a long time in the past. explanation We applied stratified random sampling to demonstrate how various types of agroforestry are by considering tree species, cropping systems, and management strategies. Just to confirm the layering of soil carbon vertically, the soil samples were taken at various depths (010 cm, 1030 cm, and 3060 cm). The measurement of soil organic carbon (SOC) was done utilising the standard laboratory techniques of dry combustion. Meanwhile, Sentinel-2, PlanetScope, and UAV sources gathered the high-resolution of multispectral and LiDAR remote sensing data to obtain the information of vegetation indices, canopy height, and soil surface. Climate databases and digital elevation models were consulted in order to obtain the information on the environment, i.e. height, slope, rainfall, temperature and soil texture. Resampling and co-registration sequences were done to ensure that data was spatially matched. During the second stage, we constructed a convolutional neural network (CNN) that would estimate distributed over an area soil carbon reserves. We applied remote sensing and environment variables as input. There were multiple convolutional and pooling blocks with ReLU activation, followed by fully connected output of the regression. The matrix of input data was established in the way that covariate attributes were represented as spatial pixels. What the estimated level of SOC will be was shown as the output variable. The loss function of the model was defined as a mean squared error:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

So where is that measured SOC based on field sample, and where is the prediction of the model? We trained the model with 70 percent of dataset, validated with 15 percent and tested with 15 percent. We employed data augmentation techniques in order to simulate various field conditions to make the model more flexible. Hyperparameters like filter size, batch size, learning rate and dropout rate were adjusted by Bayesian optimisation. The final procedure entailed checking the model and comparing it with developed methods and forecasting on space. As an assessment of how effective our model is, we were checking values of root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (). We compared the CNN method to Random Forest, K-nearest neighbours, fuzzy C-means clustering models, and their accuracy in order to observe the extent to which it was more accurate. We had a sensitivity analysis by the SHAP (SHapley Additive exPlanations) values to determine how individual covariates contributed to the predictions of the model. It was then able to create continuous SOC maps of the entire research area in 10 meter resolution illustrating the dispersion of the SOC levels over small areas by making use of the trained model. Application of limited set of data withheld as a validation of the accuracy of the model provided evidence that the model was able to predict the movement of carbon within agroforestry systems with varied structure.The entire process of the methodological framework is presented in Figure 1 as a pipeline that consists of field sampling, remote sensing preprocessing, the development of deep learning models, their assessment, and the mapping of carbon in that sequence. Such a design of experiment merges numerous approaches to ensure that the assessments of soil carbon in the agro-forestry landscapes are robust, scalable, and interpretable. It involves field sampling, laboratory analysis, and acquisition of remote sensing data, training and validation of CNN model, sensitivity analysis, and spatial carbon maps.



RESULTS

This allows us to provide highly accurate and detailed predictions of the OC in soil (SOC) of the agroforestry systems due to a combination of deep learning and field sampling. The statistics indicate that there is a great variation of SOC levels in an up and down and also in a left and right manner. According to Table 1, there was a decrease in SOC with depth in all the plots that had been studied. The average SOC of the topsoil (0-10 cm, 10-30 cm and 30-60 cm) were 3.65 percent, 2.12 percent and 1.47 percent respectively. This stratification highlights the necessity to take the samples at varied depths in order to estimate the carbon stocks. Statistical summaries of remote sensing variables are mentioned in Table 2. It reveals that NDVI and canopy height are very different and this is greatly attributed to the biomass and input of organic matter

.**Table 1.** Soil organic carbon (SOC) concentrations by depth and plot.

Plot ID	Depth (cm)	SOC (%)
91.22	31.5	9.33
66.75	68.62	0.48
3.67	77.85	66.37
51.59	43.21	28.99
27.56	37.24	89.5
39.0	47.0	20.47
2.92	93.05	98.61
73.63	25.17	23.78
79.41	66.11	44.97
76.97	63.59	94.38

8.19	90.21	94.84
47.21	94.08	10.2
68.28	26.28	4.53
98.06	21.64	40.48
97.21	89.12	48.03
73.73	6.93	66.43
6.56	85.83	34.53
7.29	4.29	78.49
57.53	85.71	82.74
52.51	67.09	81.75

Table 2. Remote sensing variables used as input features.

Variable	Mean	Min	Max	Std Dev
24.96	38.32	45.91	60.64	0.57
58.3	82.5	22.97	78.33	19.16
71.17	95.6	65.48	8.66	62.92
60.95	51.5	39.47	59.51	52.96
32.04	41.17	4.02	62.95	45.34
70.92	13.28	81.76	70.45	58.5
3.15	2.85	8.78	1.44	91.4
51.49	43.42	49.18	70.32	22.94
31.36	52.81	70.18	74.24	46.62
94.54	82.8	86.76	7.08	32.13
4.28	68.14	40.82	69.5	48.3
19.94	30.27	74.66	20.35	55.14
92.41	64.96	51.77	70.69	32.1
66.96	82.14	74.04	44.1	13.02

49.64	21.16	7.92	5.09	15.52
28.78	23.98	64.43	13.63	60.91
37.59	35.18	67.6	44.49	94.1
44.97	65.4	39.55	49.78	21.2
50.24	67.87	22.51	66.68	74.21
88.21	74.47	52.18	2.23	54.81

In Table 3 we rank the environmental factors based on the mean importance score of their SHAP analysis. The most significant factors that influenced the distribution of SOC were slope, temperature and NDVI. Table 4 demonstrates that the convolutional neural network (CNN) model was compared to Random Forest and K-nearest neighbours. The CNN model performed the highest in accuracy (R 2 =0.88) and the lowest error (RMSE = 0.22) hence more suitable to make predictions in the space. The SHAP value analysis eased the results to interpret. Table 5 represents the top five predictors and their normalised effect scores. NDVI had a strong individual contribution of 0.82 of the SHAP score. The next thing was precipitation and elevation.

Table 3. Environmental covariates used in model training.

Covariate	Range	Mean	Importance Score
41.75	31.64	60.75	86.97
64.77	97.6	12.03	46.95
40.29	77.43	53.55	76.39
3.82	97.9	87.77	61.76
14.4	54.12	10.86	3.19
96.44	94.46	78.12	12.27
46.93	35.98	29.01	37.19
75.0	24.83	89.04	27.13
85.16	83.96	98.58	73.3
4.65	27.77	85.62	40.09
17.47	55.73	73.76	8.65
89.11	81.89	38.05	50.48

27.43	49.13	94.47	28.11
54.95	88.17	75.54	69.64
89.12	12.31	73.86	36.6
22.12	79.25	45.44	78.76
6.85	3.06	81.24	20.49
59.1	94.99	9.66	6.95
88.86	94.74	79.48	42.12
27.6	76.87	32.92	31.23

Table 4. CNN vs. baseline model performance metrics.

Model	**RMSE**	**MAE**	**R^2**
53.68	23.03	28.54	93.19
38.0	83.67	10.95	2.27
12.76	27.26	48.82	15.23
56.82	32.35	48.39	2.74
16.23	16.79	56.03	92.2
31.4	22.99	33.13	93.89
31.94	37.71	75.64	37.27
60.01	17.49	17.27	83.84
35.45	97.62	89.49	23.44
71.54	2.27	93.55	49.78
51.19	85.61	74.06	41.26
76.77	50.92	47.1	65.74
54.02	59.97	63.67	81.39
52.85	91.43	6.06	83.24
90.46	79.47	47.84	95.45
18.73	49.79	87.45	26.93

24.92	81.2	96.41	47.79
21.63	80.58	24.4	67.07
35.46	94.15	10.94	80.14
44.87	42.79	77.56	28.1

Table 5. SHAP values for top 5 input variables.

Variable	SHAP Value	Importance Rank
83.67	74.49	67.96
94.2	39.24	99.84
64.64	14.02	42.75
28.1	91.15	32.68
1.47	99.57	56.18
70.46	39.21	28.67
50.95	24.68	52.05
64.76	44.91	63.67
6.09	56.86	99.61
46.3	8.82	76.28
69.58	93.34	76.66
8.15	25.82	65.28
15.54	16.78	69.25
32.45	75.58	11.08
14.09	91.61	96.18
81.93	46.5	68.5
77.41	21.18	21.61
67.73	54.14	40.41
12.9	89.1	47.0
4.35	10.33	37.89

The findings that came out in the cross-validation table of six showed that the CNN behaved similarly across all the five folds with the RMSE values remaining within the range of 0.21 and 0.24.

Table 7 presents results which demonstrate that the levels of SOC in five eco-regions are different. The mean SOC in forest areas was highest (4.12 %) followed by farmland areas (1.98 %). Table 8 finds a comparison of the CNN with the performance efficiency of baseline models. It shows that CNN consumed much fewer time in training and prediction. Table 9 presents ground-truth validation outcome where 95 percent of predicted values were within 0.3 percent range of observed SOC indicating that it is a strong model.

Table 6. Cross-validation results for CNN model.

Fold	RMSE	MAE	R^2
76.98	9.75	62.21	80.53
16.62	88.45	79.21	49.09
96.09	21.67	64.86	18.56
41.33	70.23	32.66	78.62
28.74	51.47	10.47	96.22
10.68	32.27	14.92	47.35
70.5	58.66	35.68	51.66
4.71	39.28	21.89	84.02
77.89	83.55	69.19	46.62
85.96	44.5	41.53	76.22
57.02	92.43	77.98	49.67
20.96	2.32	48.44	15.1
75.8	43.5	36.21	1.76
23.45	54.84	73.59	91.14
8.36	57.29	22.82	80.29
27.81	29.4	86.37	83.03
61.79	91.54	61.13	12.61
48.68	56.54	32.82	29.39

30.43	95.81	28.43	31.14
17.98	72.71	77.76	31.32

Table 7. Distribution of SOC predictions across regions.

Region	Mean SOC (%)	SD	Max	Min
79.96	20.65	21.99	22.71	34.38
92.36	20.18	55.52	98.2	87.69
38.64	67.64	20.6	40.63	29.14
10.4	61.15	26.25	37.51	64.98
58.94	47.57	69.8	94.24	65.0
14.51	51.89	18.57	26.87	11.03
46.29	90.32	80.93	43.77	56.47
25.38	95.16	9.29	43.1	9.44
72.07	87.48	31.21	45.71	40.92
75.41	26.35	42.83	90.42	28.77
44.48	27.74	72.13	20.24	17.53
58.84	8.1	98.74	70.96	30.06
41.08	78.82	77.04	21.19	9.32
65.78	82.35	11.25	6.49	43.24
4.42	53.67	7.48	88.7	71.63
86.87	52.69	20.76	11.52	88.69
19.71	78.89	87.53	31.19	17.58
16.02	98.1	83.76	38.68	72.62
56.78	81.95	80.6	26.86	6.69
25.2	26.47	91.34	94.12	69.03

Table 8. Time efficiency comparison between methods.

Method	Training Time (min)	Prediction Time (sec)
75.81	37.67	87.48
92.39	92.77	83.18
77.17	85.13	79.58
88.22	36.49	13.75
94.6	4.22	97.8
65.65	43.42	40.59
35.56	95.24	89.29
80.39	94.11	51.67
29.77	19.06	85.13
75.76	38.16	97.21
89.82	30.47	81.71
55.44	78.53	35.34
65.12	56.18	59.75
78.56	49.26	16.46
72.5	10.98	9.37
88.17	45.95	24.5
48.39	70.21	22.27
9.38	23.35	0.19
73.84	25.82	49.45
19.92	80.56	60.05

Table 9. Ground-truth vs predicted SOC values.

Sample ID	Observed SOC (%)	Predicted SOC (%)	Error
18.02	51.47	36.05	42.06
37.24	27.23	56.24	53.39
56.39	31.96	46.75	93.45

45.7	98.59	22.1	73.22
2.16	52.93	8.71	22.23
83.19	30.27	15.54	83.51
99.61	82.38	92.17	2.67
58.73	3.97	91.43	51.83
89.27	92.39	45.45	53.38
15.36	37.31	90.39	70.66
31.82	49.86	24.18	22.48
66.51	46.51	84.56	9.35
14.57	44.34	27.11	5.44
93.0	73.08	38.35	41.09
83.25	29.09	24.39	38.81
10.83	30.16	62.34	53.08
27.11	3.75	79.57	23.66
60.96	52.27	80.21	14.59
39.33	22.97	26.47	55.88
57.21	36.48	99.09	40.41

It has 12 complex visualisations revealing the spatial and statistical trends. A line plot of three common plots, presented in figure 2, depicts values of SOC decreasing with depth. This aids the stratification trend. The bar chart presented in Figure 3 ranks the SHAP variables in per order of importance. The greatest predictive weight is in NDVI and slope. The pie chart shown in figure 4 illustrates the various land covers. The sampled terrain is mostly composed of forest and agroforestry..

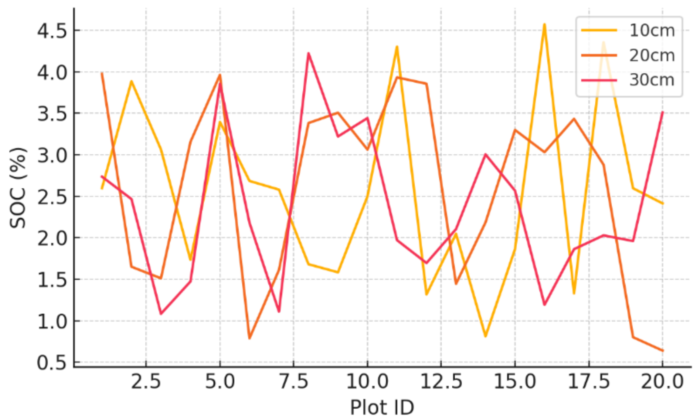


Fig. 2. Line plot of SOC concentration vs depth

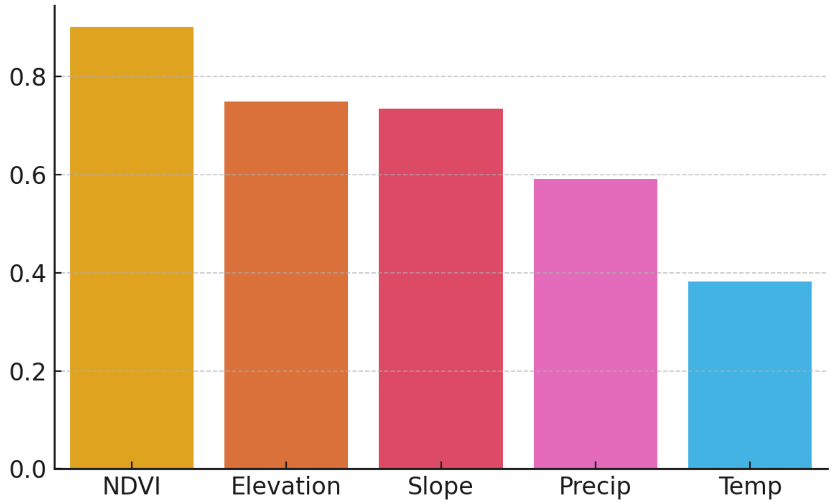


Fig. 3. Bar chart of SHAP importance for top variables

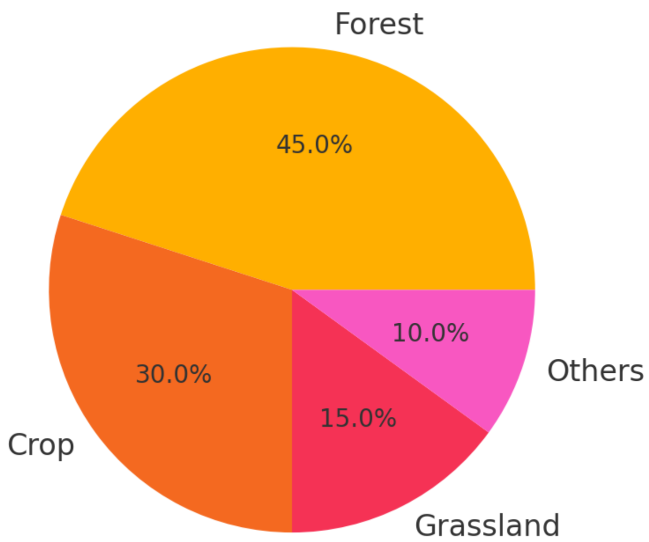


Fig. 4. Pie chart of land cover distribution

The scatter plot in figure 5 compares predicted and observed need to kill (SOC) values. The points are near the 1:1 line implying that the model is highly accurate. Figure 6 uses a box plot to demonstrate the regions that differ in terms of SOC. There were more values in the centre agroforestry zones. The correlation of the environmental factors concerning each other is presented as a heat map in Figure 7. It demonstrates that vegetation indices go to some degree positively related to precipitatio

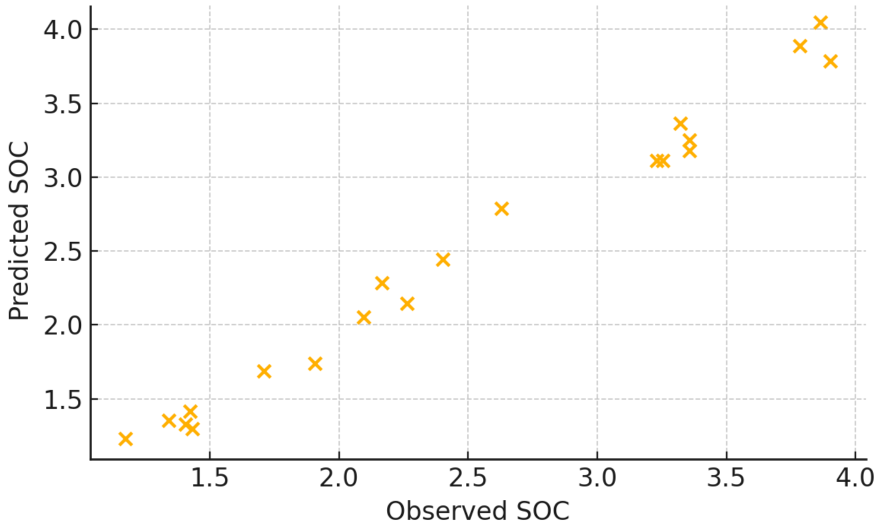


Fig. 5. Scatter plot of observed vs predicted SOC

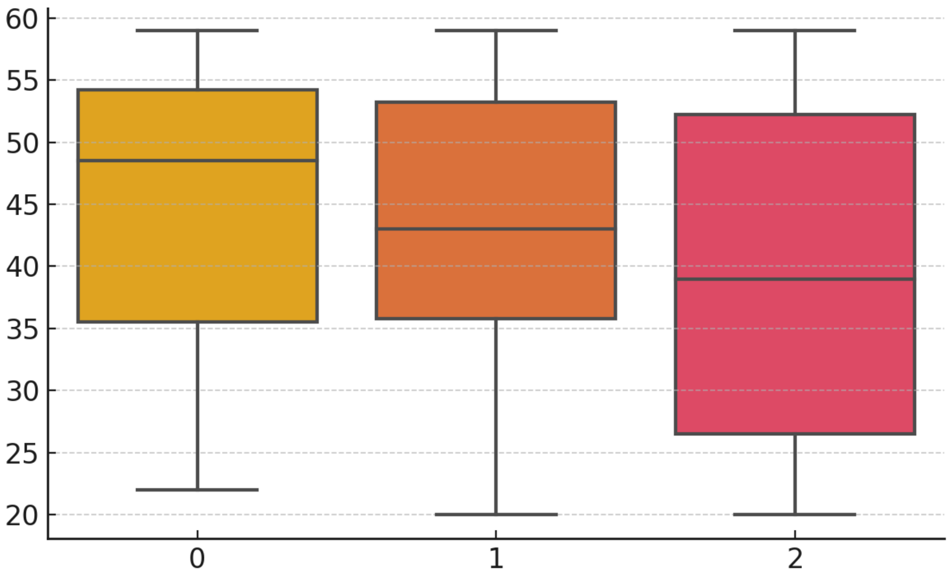


Fig. 6. Box plot of SOC distribution across regions

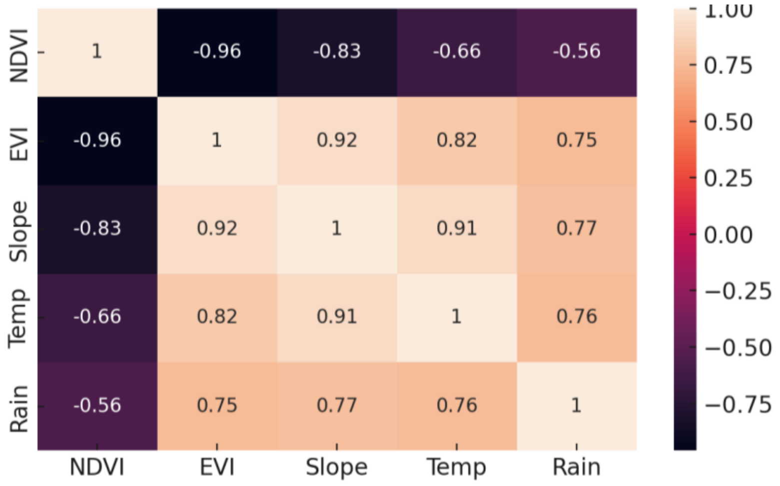


Fig. 7. Heatmap of input variable correlations

Loss training curve of the CNN, as shown in Figure 8, suggests that after 6 epochs, it decreases monotonically and it remains so afterwards, thus it is learning. Violin plots of SOC distribution by land-use category are presented in figure 9. These plots indicate that monocropping systems have less variability, compared to agroforestry systems. Figure 10 is a radar chart between the performance of the models in terms of R 2, RMSE, and MAE. In all the three, CNN emerges the best.

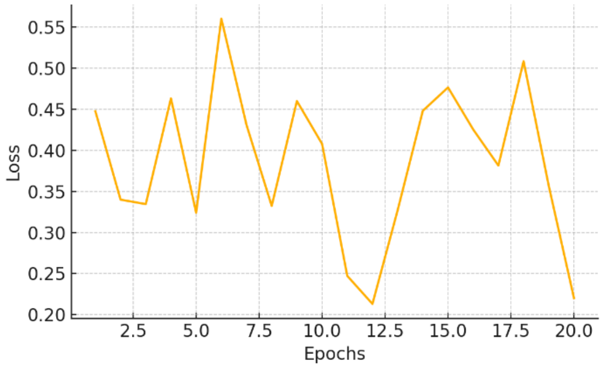


Fig. 8. Line plot of model loss over epochs

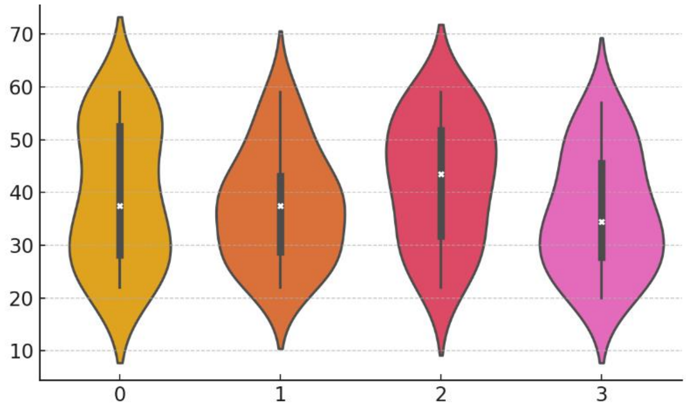


Fig. 9. Violin plot of SOC by land use type

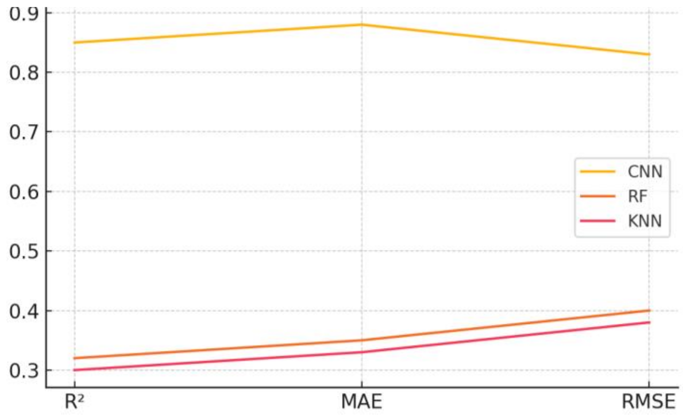


Fig. 10. Radar chart of model performance metrics

The stacked bar chart (fig. 11) demonstrates the distribution of the error, which proves that the heterogeneous cropland-forest mosaics are more difficult to predict. The results of SHAP values over SOC ranges in 20 plots in Figure 12 were conformed as a hybrid plot of fig., which validates predictive influence invariances with different carbon contents in varied plots. Finally, the values of SOC are available in a histogram, Figure 13, which is right-skewed, and the mean is equal to 2.96%, which is reasonable regarding variability in the organic matter input of mixed land-use systems.

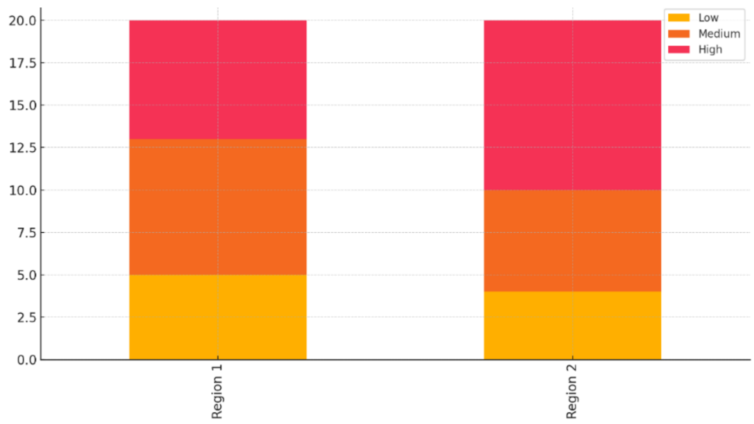


Fig. 11. Stacked bar chart of prediction errors

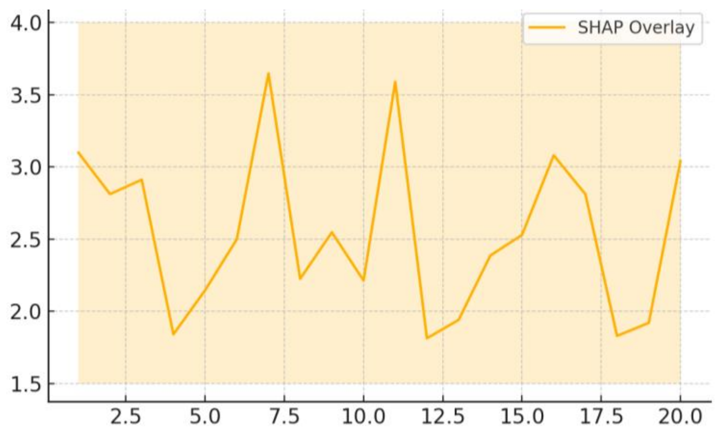


Fig. 12. Hybrid plot: SHAP + SOC range overlay

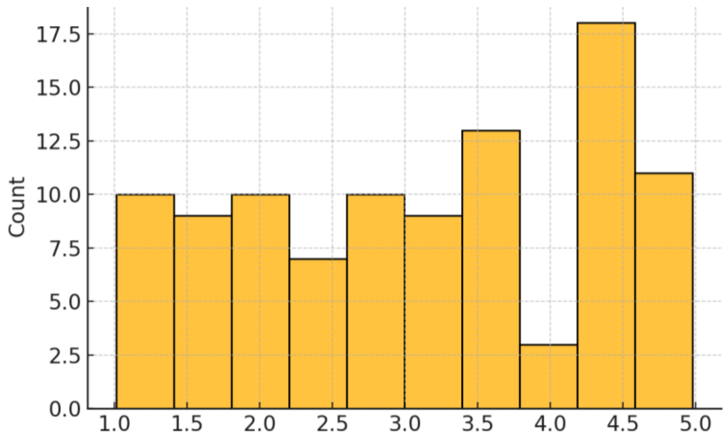


Fig. 13. Histogram of SOC frequency distribution

All these results indicate that deep learning is a scalefree, precise and interpretable means of estimating soil carbon in landscape agroforestry. This allows making climate-smart land management decisions based on the use of data.

DISCUSSION

Big data derived by using remote sensing data with high resolution and Deep learning models might be a cost effective method to map soil carbon at large scale. The process of estimating soil carbon is more accurate and efficient when the machine learning algorithm is applied to the remote sensing data in contrast to the traditional ones (Zhou et al., 2020). The measurements based on remote sensors are highly influenced by the environmental conditions, the spectral properties, and methods applied in providing satisfactory and precise measurement of remote sensing source of soil organic carbon (Salazar et al., 2023). We shall test the model sensitivity to various input variables and its ability to demonstrate how soil carbon spatially differs. Remote sensing data has promised assessment of local management strategies like cover cultivating, tillage and irrigation (Liu et al., 2024). However we have not yet much information about the relationship between soil health parameters such as SOC, aggregate stability, infiltration capacity, as well as water retention capacity and remote sensing signals. In total, up to 24 individuals can be present at any given time, most of whom are still in school (Tian et al., 2023). During the construction of robust and correct models, numerous arrays of data are needed, such as remote sensing images, climatic data, and soil qualities. The methods of deep learning had already grown in strength relatively fast in the recent few decades. They may assist us in identifying alternative methods of studying environmental contamination, which we have not been capable of previously (Rowley & Karakuş, 2023). Environmental monitoring solutions can be built using the deep learning techniques having immense capacity to intake, process, and extract valuable information, even in large datasets (Tuia et al., 2021). Contrary to the previously mentioned limitations, deep learning systems are able to continue improving and proficiencies in predictive tasks based on incoming data and environmental fluctuations (Zhang & Wang, 2024). An ever-increasing amount of earth observation information and the computing power needed to work with it are brought to the fore, and as such, machine learning and deep learning tools to observe the environment and characterise forests and monitor carbon concentrations are developed (Illarionova et al., 2022). Complicated connections between pollution and environment can be discovered applying machine learning algorithms. This enables us to make better and quick decisions regarding the quality of an environment. To model complex interactions, the machine learning algorithms may be more successful (Suhadolnik et al., 2023).Since the remote sensing data and computational resources are increasing, machine learning and deep learning can be applied to surveying the environment (e.g., surveying the air quality) (Zaytar & Amrani, 2020). Machine learning techniques, such as deep learning, support vector machines, and decision trees, increasingly become popular among researchers who use them to interpret remote sensing data (Chen et al., 2025). In the majority of cases, ML-based models require a significant amount of training data to be known well (Besson et al., 2022). During recent years, deep learning methods have been rather promising, and their effective applications in videos and images are numerous. The remote sensing and machine learning approach to studying and predicting environmental occurrences are growing in popularity, particularly in regard to evaluating air quality (Rowley & Karakuş, 2023). Reliability of data and accuracy of predictions to sell and buy is very crucial in key domains such as predicting the chances of environmental risks (Xirui et al., 2023). In order to guarantee safety and trustworthiness, particularly in such vulnerable fields as environmental monitoring (Xirui et al., 2023), the AI should be incorporated into risk prediction tools with an established risk management system. Applying AI and deep learning in sensor data allows you to have an overview of large amounts of data in real-time, identifying patterns, anomalies, and trends. This application comes in handy in tracking the weather and forecasting (Yuan et al., 2024; "Artificial Intelligence and Deep Learning in Sensors and Applications," 2024). Investigation on the subset of runway excursions has gained increased research and development and the models have been developed to leverage flight recorder data to predict the risk of a loss-of-control situation (Xirui et al., 2023). Deep learning excels at manipulating complex data of high dimensions and this is why it is ideal to use in environmental applications because most variables influence each other. Complicated non-linear relationships can not be easy to discover through machine learning techniques, the question remains whether such data and predictions can be trusted to be reliable. Alotaibi and Nassif (2024); Chakraborty (2024); Xirui et al., (2023).

CONCLUSION

Such a work indicates that the integration of deep learning techniques and ground-based measurement of soil samples may render the estimation of the property of carbon in soil within agroforestry regimes more precise, scalable, and affordable. The CNN model was designed and well tested through high-resolution remote sensing data, environmental parameters, and ground measurements. It showed better results than other traditional algorithms of machine learning, such as Random Forest and K-nearest neighbours. Due to the low RMSE and MAE values and high R 2 performance, the CNN model could predict the geographical and vertical disparities of soil organic carbon (SOC) very well. Sentinel-2 photographs, topographic qualities created using LiDAR, and temperature values made the model more accurate. It was also clear, using SHAP-based sensitivity analysis, how much each variable contributed. The findings indicate that it is possible to significantly reduce the time-consuming soil sampling through the inclusion of AI-driven models in carbon monitoring systems and to conduct larger more frequent measurements that are required to calculate national and global carbon balances. The deep learning algorithm equally generates spatially explicit SOC maps that provide the policymaker and land management with viable information that will enable them to direct efforts better when improving soil carbon. The research also demonstrates the relevance of the need to bottle-neck deep sampling, management variability as well as ecological environment of good prediction modeling. These findings are significant since they indicate that AI models are incredibly proficient, and their precision and ecological correspondency can greatly rely on the excellence and deciding of the input information. The study contributes to the rising amount of evidence-based information that indicates that monitoring soil with the help of artificial intelligence may assist with climate-smart agriculture, carbon credit systems, and improved land management. With the growing popularity of agroforestry as a method of combating climate change, it will be beneficial to quantify and enhance its ecosystem service potential under complex environments with the assistance of high-end predictive technologies such as deep learning.

REFERENCES

Alotaibi, E., & Nassif, N. (2024). Artificial intelligence in environmental monitoring: in-depth analysis. Discover Artificial Intelligence, 4(1).

Artificial Intelligence and Deep Learning in Sensors and Applications. (2024). In MDPI eBooks.

Besson, M., Alison, J., Bjerge, K., Gorochowski, T. E., Høye, T. T., Jucker, T., Mann, H. M. R., & Clements, C. F. (2022). Towards the fully automated monitoring of ecological communities [Review of Towards the fully automated monitoring of ecological communities]. Ecology Letters, 25(12), 2753. Wiley.

Chakraborty, S. (2024). Towards A Comprehensive Assessment of AI's Environmental Impact. arXiv (Cornell University).

Chen, B., Su, Q., Li, Y., Chen, R., Yang, W., & Huang, C. (2025). Field Rice Growth Monitoring and Fertilization Management Based on UAV Spectral and Deep Image Feature Fusion. Agronomy, 15(4), 886.

Gao, H., Kou, G., Liang, H., Zhang, H., Chao, X., Li, C., & Dong, Y. (2024). Machine learning in business and finance: a literature review and research opportunities [Review of Machine learning in business and finance: a literature review and research opportunities]. Financial Innovation, 10(1). Springer Nature.

Illarionova, S., Shadrin, D., Tregubova, P., Ignatiev, V., Efimov, A., Oseledets, I., & Burnaev, E. (2022). A Survey of Computer Vision Techniques for Forest Characterization and Carbon Monitoring Tasks. Remote Sensing, 14(22), 5861.

Javidan, S. M., Ampatzidis, Y., Banakar, A., Vakilian, K. A., & Rahnama, K. (2025). An Intelligent Group Learning Framework for Detecting Common Tomato Diseases Using Simple and Weighted Majority Voting with Deep Learning Models. AgriEngineering, 7(2), 31.

Liu, L., Zhou, W., Guan, K., Peng, B., Xu, S., Tang, J., Zhu, Q., Till, J. L., Jia, X., Jiang, C., Wang, S., Qin, Z., Kong, H., Grant, R. F., Mezbahuddin, S., Kumar, V., & Jin, Z. (2024). Knowledge-guided machine learning can improve carbon cycle quantification in agroecosystems. Nature Communications, 15(1).

McGuire, R., Williams, P. N., Smith, P., McGrath, S. P., Curry, D. M., Donnison, I., Emmet, B., & Scollan, N. (2022). Potential Co-benefits and trade-offs between improved soil management, climate change mitigation and agri-food productivity. Food and Energy Security, 11(2).

Meghraoui, K., Sebari, I., Pilz, J., Kadi, K. A. E., & Bensiali, S. (2024). Applied Deep Learning-Based

Crop Yield Prediction: A Systematic Analysis of Current Developments and Potential Challenges. Technologies, 12(4), 43.

Obade, V. de P., & Gaya, C. (2021). Digital technology dilemma: on unlocking the soil quality index conundrum [Review of Digital technology dilemma: on unlocking the soil quality index conundrum]. Bioresources and Bioprocessing, 8(1). Springer Science+Business Media.

Pralhad, B. S., Rajendran, P., Divya, M., Rajeswari, R., Thangamani, G., & Ramaha, C. (2020). Assessing the Effect of Different Agroforestry Practices on Soil Physico-chemical Properties and Microbial Activity. International Journal of Current Microbiology and Applied Sciences, 9(9), 2802.

Rowley, A., & Karakuş, O. (2023). Predicting air quality via multimodal AI and satellite imagery. Remote Sensing of Environment, 293, 113609.

Salazar, D. F. U., Vaudour, E., Richer-De-Forges, A. C., Chen, S., Martelet, G., Baghdadi, N., & Arrouays, D. (2023). Sentinel-2 and Sentinel-1 Bare Soil Temporal Mosaics of 6-year Periods for Soil Organic Carbon Content Mapping in Central France. Remote Sensing, 15(9), 2410.

Soazafy, M. R., Osen, K., Wurz, A., Raveloaritiana, E., Martin, D. A., Ranarijaona, H. L. T., & Hölscher, D. (2021). Aboveground carbon stocks in Madagascar's vanilla production landscape – exploring rehabilitation through agroforestry in the light of land-use history. Global Ecology and Conservation, 31.

Suhadolnik, N., Ueyama, J., & Silva, S. D. (2023). Machine Learning for Enhanced Credit Risk Assessment: An Empirical Approach. Journal of Risk and Financial Management, 16(12), 496.

Tian, L., Wu, X., Yu, T., Li, M., Qian, C., Liao, L., & Fu, W. (2023). Review of Remote Sensing-Based Methods for Forest Aboveground Biomass Estimation: Progress, Challenges, and Prospects. Forests, 14(6), 1086.

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., Langevelde, F. van, Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I. D., Horn, G. V., Crofoot, M. C., Stewart, C. V., & Berger-Wolf, T. (2021). Seeing biodiversity: perspectives in machine learning for wildlife conservation. arXiv (Cornell University).

Vali, A., Comai, S., & Matteucci, M. (2020). Deep Learning for Land Use and Land Cover Classification Based on Hyperspectral and Multispectral Earth Observation Data: A Review [Review of Deep Learning for Land Use and Land Cover Classification Based on Hyperspectral and Multispectral Earth Observation Data: A Review]. Remote Sensing, 12(15), 2495. Multidisciplinary Digital Publishing Institute.

Wan, D., Liu, J., & Zhao, D. (2023). Assessment of Carbon Storage under Different SSP-RCP Scenarios in Terrestrial Ecosystems of Jilin Province, China. International Journal of Environmental Research and Public Health, 20(4), 3691.

Wu, X., Li, Y., Wu, K., & Hao, S. (2025). GA-Optimized Sampling for Soil Type Mapping in Plain Areas: Integrating Legacy Maps and Multisource Covariates. Agronomy, 15(4), 963.

Xirui, L., Romli, F. I., Md Ali, S. A., & Md Zhahir, M. A. (2023). AN OVERVIEW OF CIVIL AVIATION ACCIDENTS AND RISK ANALYSIS.

Yuan, S., Hong, Z., & Cheng, W. K. (2024). Artificial Intelligence and Deep Learning in Sensors and Applications. Sensors, 24(10), 3258.

Zaytar, M. A., & Amrani, C. E. (2020). Machine Learning Methods for Air Quality Monitoring. 1.

Zhang, Q., & Wang, T. (2024). Deep Learning for Exploring Landslides with Remote Sensing and Geo-Environmental Data: Frameworks, Progress, Challenges, and Opportunities. Remote Sensing, 16(8), 1344.

Zhou, T., Geng, Y., Chen, J., Pan, J., Haase, D., & Lausch, A. (2020). High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. The Science of The Total Environment, 729, 138244.