






Article

Comparative Analysis of Machine and Deep Learning Models for Soil Properties Prediction from Hyperspectral Visual Band

Dristi Datta ^{1,*}, Manoranjan Paul ¹, Manzur Murshed ², Shyh Wei Teng ³ and Leigh Schmidtke ⁴

¹ School of Computing, Mathematics, and Engineering, Charles Sturt University, Bathurst, NSW 2795, Australia; mpaul@csu.edu.au

² School of Information Technology, Deakin University, Burwood, VIC 3125, Australia

³ Institute of Innovation, Science and Sustainability, Federation University Australia, Gippsland, VIC 3842, Australia

⁴ Gulbali Institute, Charles Sturt University, Wagga Wagga, NSW 2650, Australia

* Correspondence: ddatta@csu.edu.au

Abstract: Estimating various properties of soil, including moisture, carbon, and nitrogen, is crucial for studying their correlation with plant health and food production. However, conventional methods such as oven-drying and chemical analysis are laborious, expensive, and only feasible for a limited land area. With the advent of remote sensing technologies like multi/hyperspectral imaging, it is now possible to predict soil properties non-invasive and cost-effectively for a large expanse of bare land. Recent research shows the possibility of predicting those soil contents from a wide range of hyperspectral data using good prediction algorithms. However, these kinds of hyperspectral sensors are expensive and not widely available. Therefore, this paper investigates different machine and deep learning techniques to predict soil nutrient properties using only the red (R), green (G), and blue (B) bands data to propose a suitable machine/deep learning model that can be used as a rapid soil test. Another objective of this research is to observe and compare the prediction accuracy in three cases i. hyperspectral band ii. full spectrum of the visual band, and iii. three-channel of RGB band and provide a guideline to the user on which spectrum information they should use to predict those soil properties. The outcome of this research helps to develop a mobile application that is easy to use for a quick soil test. This research also explores learning-based algorithms with significant feature combinations and their performance comparisons in predicting soil properties from visual band data. For this, we also explore the impact of dimensional reduction (i.e., principal component analysis) and transformations (i.e., empirical mode decomposition) of features. The results show that the proposed model can comparably predict the soil contents from the three-channel RGB data.

Keywords: learning-based algorithms; RGB band; rapid soil test; empirical mode decomposition; principal component analysis



Citation: Datta, D.; Paul, M.; Murshed, M.; Teng, S.W.; Schmidtke, L. Comparative Analysis of Machine and Deep Learning Models for Soil Properties Prediction from Hyperspectral Visual Band. *Environments* **2023**, *10*, 77. <https://doi.org/10.3390/environments10050077>

Academic Editors: Sutie Xu, Virginia L. Jin, Navreet Kaur Mahal and Jing Hu

Received: 24 March 2023

Revised: 18 April 2023

Accepted: 2 May 2023

Published: 4 May 2023



Copyright: © 2023 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Soil moisture (SM), soil organic carbon (SOC), and nitrogen content (NC) are fundamental soil properties that describe soil quality and contribute to plant growth [1,2]. SM is necessary for the absorption of nutrients by the roots and contributes to evaporation rates that influence the activity of microorganisms in the soil, which are responsible for breaking down organic matter and releasing nutrients that plants can use [3]. Adequate SM also supports the transport of nutrients within the plant and helps regulate plant temperature through transpiration [4]. SOC helps improve soil structure, water-holding capacity, and nutrient retention as it serves as a source of nutrients and energy for soil microorganisms essential for nutrient cycling and soil health [5]. NC creates chlorophyll that helps plants' metabolism and structure and increases plant protein through incorporation into amino acids and peptides [6]. Additionally, NC levels can affect the quality of plant products, such as the protein content of grains and the flavor of fruits and vegetables [7]. Accurate

estimation of these soil properties is essential for ecosystem management, such as soil quality assessments, climate change, and a long-term prediction that contributes to taking necessary steps to maintain and improve soil quality. Therefore, regularly monitoring these properties is vital for forest and agricultural management and maintaining nitrogen and carbon cycles [8].

Traditional measuring methods, such as laboratory analysis with chemical SM, SOC, and NC treatments, involve lots of time, labor, and costs. SM can traditionally measure by thermogravimetric method, microelectromechanical system (MEMS), heat flux sensors [9], time-domain reflectometry (TDR) methods [10], etc. SOC estimations performed by the Walkley–Black, dry combustion, humic matter colorimetry, etc. [11]. In contrast, the Kjeldahl digestion method can traditionally measure NC [12]. These methods are slow and can cover only a tiny piece of land. Remote sensing imaging can be a promising solution to overcome these limitations and efficiently measure soil properties over larger areas [13]. Therefore, developing an accurate estimation method that can quickly and precisely predict SM, SOC, and total NC is crucial. Hence, some research has been performed using remote sensing imaging as it can cover vast areas, and a short time is required for soil quality estimation.

Hyperspectral imaging (HSI) technology has become increasingly prevalent in recent times. It can capture vast amounts of spectral and spatial data organized in multiple layers, resulting in highly detailed and precise images of a given location. HSI sensors are typically installed on satellites or aircraft. They are continually updated, allowing real-time Earth surface monitoring and capturing a wide expanse in a single image. This technology has numerous applications in fields such as agriculture [14], environmental monitoring [15], food quality analysis [16], water management [17], mineral exploration [18], forensic analysis [19], medical diagnosis [20,21], urban planning [22], military applications [23], mineralogical mapping [24], crime imaging analysis [25], etc. Additionally, SM, SOC, and NC prediction from HSI is becoming an important field of research.

With the rapid developments of hyperspectral remote sensing technology with advanced sensors, it has become possible to predict SM, SOC, and NC in a non-invasive way (without physically testing any soil samples). A wide range of machine learning (ML) models with different feature combinations is presented in paper [26]. The results show that SM, SOC, and NC can predict accurately from HS data. Some other studies [27–34] also proposed different ML models that can predict SM effectively from HS data. On the other hand, due to the availability of the Land Use and Coverage Area body Survey (LUCAS) dataset, many researchers proposed good models to predict SOC, NC, and also other soil parameters from a wide range of HS datasets (400–2500 nm). The study [35,36] proposed a large-scale SOC mapping and prediction approach. Similarly, another study [37–44] proposed different ML models to predict NC from HS data. The lab-based HS remote sensing prediction model provides the best prediction performance. The data is almost free from environmental noise and can accommodate large HS bands with small spatial resolutions. However, the hyperspectral camera and sensors are expensive and not readily available.

Some researchers try to predict different soil properties from satellite data to overcome this problem. Multi-spectral satellite data, such as Sentinel 2, Landsat 8, etc., are freely available and cover vast land with a single image. In paper [45–49], the authors predict SM from different Satellite data. Other authors [40,50–55] studied the possibilities of predicting SOC and NC from Satellite data. The soil properties predicting accuracy from Satellite data is comparably lower than the average predicting accuracy from lab-based HS data. This is because satellite data comes up with many environmental noises; therefore, extreme pre-processing and massive computation power is needed to extract information from Satellite data. On the other hand, visual range/digital cameras are everywhere and available in a wide range.

Some research has been performed to predict different soil properties from digital images. In paper [56], authors predict SM from digital images by using two machine learning models: multilayer perceptron (MLP) and a support vector regression (SVR). Similarly, the possibility of predicting soil organic matter from smartphone images is

studied in research [57]. Another article [58] studied the prediction performance of SM and organic matter from mobile captured images. Some other studies [59,60] also predict soil texture from digital images. Other articles describe Munsell soil color chart [61] and predicting soil carbon [62–64]. Although satisfactory research progress has already been achieved in predicting SM and SOC from visual band images; however, the possibility of predicting NC still needs to be discovered. Regarding our knowledge and understanding, no literature shows the potential of predicting NC from visual band images.

Therefore, the principle aim of this study is to propose a model that can also predict different soil properties from mobile captured images or traditional cameras. Hence, in this study, only the three-channel of the visual band (VB), i.e., blue (450 nm), green (550 nm), and red (650 nm), have been considered and explore the possibilities of predicting SM, SOC, and NC. Although digital cameras can capture only three bands: red (R), green (G), and blue (B), there are several research studies [65–67] that suggest that using only the average of the three RGB bands can produce comparable results to using the entire spectral range of 400–700 nm. We also explore the multi-spectral VB range (400–700 nm) to understand the prediction performance.

Empirical mode decomposition (EMD) is a popular method to decompose the time-series signal into different intrinsic mode functions (IMFs) based on frequency differences. This technique is also helpful in filtering out some specific signal components from the original signal in the presence of noise [68]. In this study, we have applied EMD in the spectral signal data to filter out some specific spectral signals and extract new features from them to be used in machine learning techniques to improve classification performance. Based on our knowledge, this is the first time EMD is applied in the spectral domain to see its effectiveness in the classifications, as the hyperspectral data sometimes have noise for soil analysis. In this study, we have also applied principal component analysis (PCA) on the hyperspectral reflection data and the newly extracted features from the EMD to observe the classification performance.

The contributions of this research are given as follows:

- To explore the possibility of prediction soil properties (SM, SOC, NC) from a 3-channel RGB band and the full spectrum of visual band data.
- To provide a prediction comparison in three cases: i. full range of HS band ii. VB, and iii. only RGB band.
- Effectively use EMD to eliminate data noise and improve the soil property prediction accuracy.
- To reduce large data dimension using PCA, minimize computational time and cost and improve prediction accuracy.
- Performance of a wide range of ML and DL models is evaluated to understand the best prediction accuracy depending on various soil properties.

The paper is structured as follows: In Section 2, we provide a description of the HS remote sensing data and the ground truth SM, SOC, and NC data. Section 3 outlines the step-by-step workflow employed in this study. The results obtained from using different algorithms to predict SM, SOC, and NC are presented in Section 4, along with a comparative analysis that includes validation and evaluation. A critical discussion of the research outcomes is given in Section 5. Section 6 concludes the paper by offering a methodology for predicting SM, SOC, and NC for visual band data that can be used as a guide.

2. Datasets

For this study, we have used three different hyperspectral datasets. A brief overview of these datasets is given below.

2.1. Soil Moisture Dataset

The SM dataset was captured during the five-day field camping that spanned over two weeks in May 2017 in Karlsruhe, Germany, which is freely available for research [69]. The measurement setup centered around an undisturbed soil sample consisting of bare

soil without vegetation. The soil sample was composed of strongly clayey silt with a 15 cm radius and 20 cm height, irrigated according to a defined schema. The resulting soil moisture ranged from 25% to 42%, as measured by TDR sensors installed in various depths from 2 cm to 18 cm. However, for this study, the uppermost sensor's readings, situated 2 cm deep, were considered as they provided the most accurate estimate of subsurface soil moisture.

After that, HS remote sensing data was captured using a Cubert1 UHD 285 HS snapshot camera, with 50×50 pixel images, each characterized by 125 spectral channels ranging from 450 nm to 950 nm with a 4 nm spectral resolution. The camera was placed on a tripod at 1.70 m height and captured the entire soil sample. Each image's mean spectrum of the soil surface was calibrated with the spectralon spectrum to produce a single data point. The dataset comprises 679 high-dimensional data points, including 125 hyperspectral bands and one soil-moisture value as ground truth.

2.2. Soil Organic Carbon and Nitrogen Dataset

The LUCAS database is a large-scale dataset managed by the European Commission's Joint Research Centre [70]. It contains information on land use, land cover, soil properties, and topsoil characteristics from over 300,000 locations across 33 European countries. The database is updated every three years starting from 2006 [70]. Its purpose is to provide policymakers, researchers, and the public with valuable information for environmental assessments, resource management, and decision-making.

The LUCAS database collects data through a standardized sampling approach, where soil samples and measurements are taken at randomly selected locations across different land use types. The database contains a wide range of information, including soil texture, pH, organic carbon, nitrogen, and other nutrient properties. The LUCAS database is freely available to the public, and users can access the data through the European Soil Data Centre (ESDAC) portal [71]. The database has been used for various purposes, including environmental monitoring, soil health assessments, land-use planning, etc. [72,73].

We have used the LUCAS 2015 dataset for SOC and NC prediction. This sampling consists of about 20,000 soil samples collected using a multi-level stratified random sampling technique to represent Europe's various land use types. For each sampling point, a composite sample consisting of five topsoil samples (0–20 cm) was created and analyzed for physical, chemical, and reflectance properties using standardized methods in a single laboratory.

After sample collection, the FOSS XDS Rapid Content Analyzer (FOSS NIRSystems Inc., Denmark) was used to record the absorbance of each soil sample in the 400–2499.5 nm range with 0.5 nm intervals [72]. The dataset comprises 28 countries' data in approximately 21,782 soil sample data. From the LUCAS 2015 dataset, for this research, we only considered the Swedish dataset that comprises 1891 soil samples. Figure 1 shows the point sampling location in Sweden obtained from the given .shp file in ArcGIS Pro software. The dataset includes point IDs associated with soil properties, including SOC and NC. The value of SOC ranges from 2 to 543.1 g/kg, and the NC range from 0.2 to 36.7 g/kg.

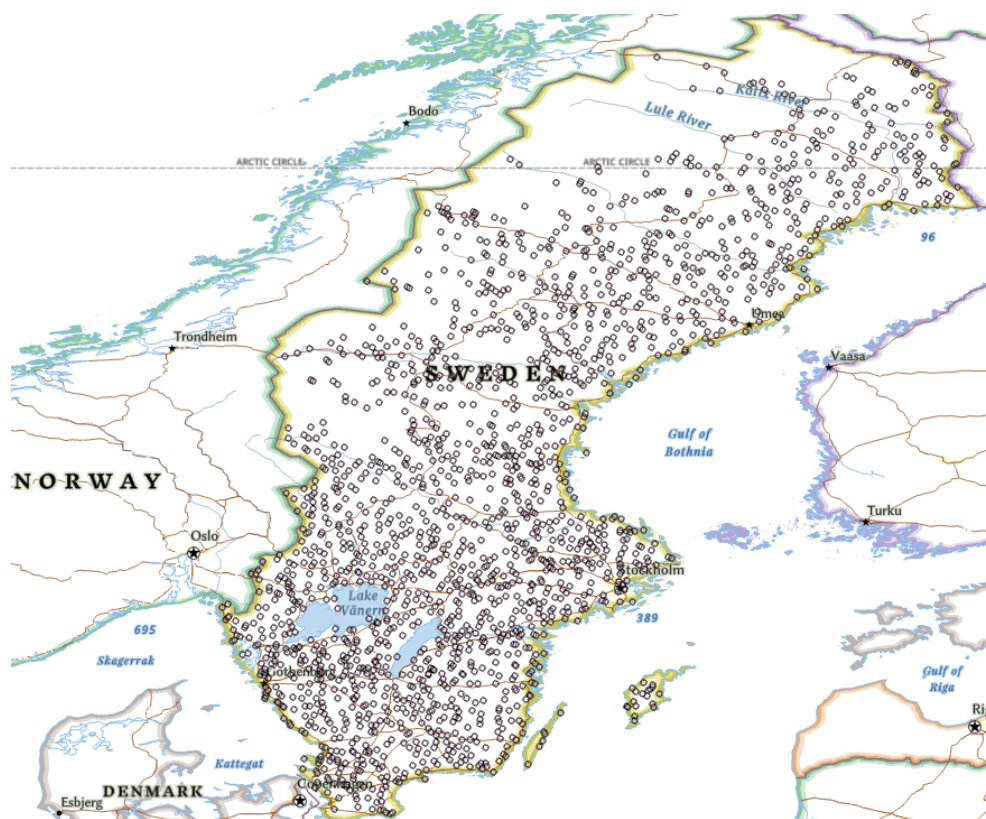


Figure 1. Study area for soil organic carbon and nitrogen prediction where the points show the sampling locations.

3. Methodology

This study is divided into two processes: data preprocessing and regression model to predict SM, SOC, and NC from three HS datasets. Figure 2 presents the work methodology for this research.

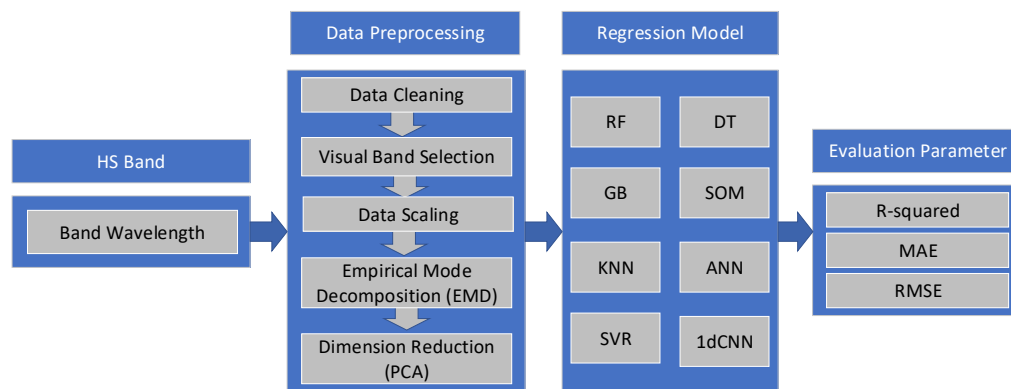


Figure 2. Schematic illustration of the proposed regression framework using different machine learning and deep learning models.

3.1. Data Prepossessing

In order to manage the high dimensionality of the HS data, five consecutive preprocessing steps were examined in this study which are given as follows:

3.1.1. Data Cleaning and Filtering

Data cleaning and filtering are vital to building an excellent machine and deep-learning model. For all three datasets, some data samples with missing ground truth data are manu-

ally removed to minimize the inhomogeneity of data. Some other non-significant features are also removed to make the dataset more usable and convenient for training purposes.

3.1.2. Visual Band Selection

This study aims to propose a model that can predict SM, SOC, and NC from general digital camera images. Therefore, all other band reflection/absorbance values are manually removed for the three different datasets, making the dataset only visual band range. For SM data, the useable band values are 454 to 702 nm, and for SOC and NC data, 400–700 nm has been considered. After that, we are making the three datasets that contain only three channels, i.e., red (R), green (G), and blue (B). To do this, we take the mean reflectance/absorbance of the blue band (400–500 nm), green band (500–600 nm), and red band (600–700 nm). These reflectance/absorbance values are used as input features to predict the performance of SM, SOC, and NC.

3.1.3. Data Scaling

In the third step of data preprocessing, feature scaling was performed to standardize all the input features before training the model. The motivation behind the feature scaling is to make all the input features appropriately scaled to balance their contributions. In this study, the standard scaling approach has been considered that mathematically works according to the Formula (1).

$$z_{score} = (x_{training} - u_{mean})/s.d \quad (1)$$

where z_{score} is the standard score, $x_{training}$ is the training sample, and u_{mean} and $s.d$ are the mean and standard deviation of the training sample, respectively.

3.1.4. Empirical Mode Decomposition

EMD is a technique that converts the input bands into physically meaningful additives [74]. This study passes all three VB range data through the EMD process to obtain the maximum variance from the reflectance/absorbance data. The first IMF is then taken as an input feature. Figure 3 shows the one sample SM reflectance curve, and the corresponding first IMF is given in Figure 4. The mathematical expression of EMD is given as follows:

$$I_{signal} = \sum_{n=1}^N IMF_n + RES_n \quad (2)$$

where, I_{signal} is the reflectance/absorbance signal, IMF_n is the N^{th} intrinsic mode functions, and RES_n is the residual to the corresponding N^{th} intrinsic mode, respectively.

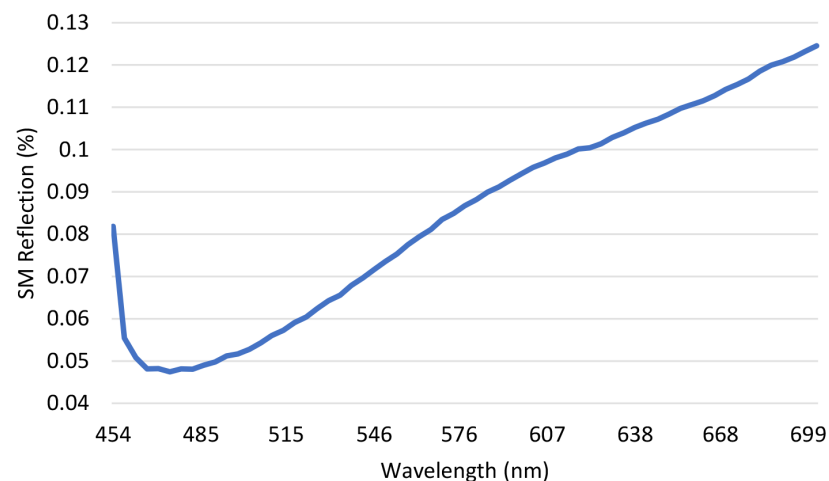


Figure 3. Soil moisture reflectance curve of visual band range.

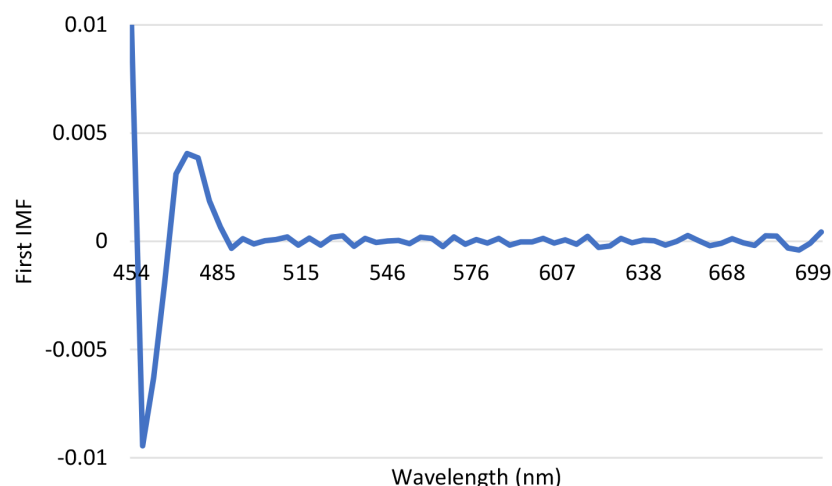


Figure 4. Soil moisture reflectance curve corresponding first IMF.

3.1.5. Dimension Reduction

Finally, the Principal Component Analysis (PCA) technique has been adopted to address non-linear high dimensional data units and effectively reduce data dimension [75] in the visual band range. In this study, we depend on the primary seven major additives that extract nearly 99.99% features from all three VB datasets.

3.2. Regression Model

To predict SM, SOC, and NC from three VB datasets, seven different well-known ML regressor models have been adopted to understand the prediction performance, and they are Random Forest (RF) [76], Decision Tree (DT) [77], Gradient Boosting (GB) [78], Self Organizing Map (SOM) [27], K Nearest Neighbors (KNN) [79], Artificial Neural Network (ANN) [80] and Support Vector Regression (SVR) [81]. The performance of a one-dimensional Convolution Neural Network (1dCNN) [82] is also evaluated. Most ML regressors are designed in scikit-learn library packages, Susi library packages implement the SOM model, and the PyTorch library develops 1dCNN. A brief discussion of each learning-based model that has been considered in this study is given below:

3.2.1. Random Forest

RF regression is an ML algorithm used for predicting continuous output variables. It works by building multiple decision trees on different subsets of the training data and combining their predictions. This algorithm selects the best feature at each split by randomly selecting a subset of features, which helps reduce the correlation between trees and improve model accuracy. The RF regression is robust, can handle large and complex datasets, and is resistant to overfitting issues [83].

3.2.2. Decision Tree

DT regression shows a similar characteristic to RF regression. It partitions the feature space into smaller regions based on the values of input features and assigns a constant value to each region as the prediction. It selects the feature that maximizes the reduction in variance at each split, and the process continues until a stopping criterion is met. It is a simple and interpretable algorithm that can handle numerical and categorical features but tends to overfit the training data [84].

3.2.3. Gradient Boosting

GB regression is also an ML algorithm used for regression tasks. It works by iteratively adding weak decision tree models to the ensemble, where each new tree corrects the errors of the previous tree. The algorithm minimizes a loss function by gradient descent,

where the gradient is computed concerning the predictions of the current ensemble. It can handle complex nonlinear relationships between the input and output variables and resist overfitting [85].

3.2.4. Self Organizing Map

SOM also known as Kohonen map, is an unsupervised machine learning algorithm used for dimensionality reduction, data visualization, and clustering [86]. It maps high-dimensional input data onto a lower-dimensional grid of nodes or neurons, where each neuron represents a cluster of similar data points. During training, the algorithm adjusts the position of each neuron and its weights to match the input data better. SOM is a powerful visualization tool that can reveal the underlying structure and patterns in high-dimensional data and can be used for unsupervised clustering or to initialize the weights of other ML algorithms. However, it requires careful tuning of hyperparameters and may be sensitive to the initialization of the neurons. However, in this study, we have used the SOM packages proposed in article [27]. This SOM framework involves two SOM architecture that combines unsupervised and supervised learning.

3.2.5. K Nearest Neighbors

KNN is an ML algorithm used for both classification and regression tasks. It finds the k-closest training examples in the feature space to a new test example and assigns the label or value of most of the k-nearest neighbors to the test example. KNN is simple to implement and can handle both numerical and categorical features. Still, it can be sensitive to the choice of k and the distance metric used to measure the similarity between examples. It is also non-parametric, meaning it makes no assumptions about the data distribution [87].

3.2.6. Artificial Neural Network

An ANN is an ML algorithm inspired by the structure and function of the human brain. It consists of a network of interconnected nodes, called neurons, organized into layers. During training, the algorithm adjusts the weights of the connections between neurons to minimize a loss function. ANNs can handle complex nonlinear relationships between input and output variables and are widely used in many applications. However, they can be computationally expensive to train and may require careful tuning of the architecture and hyperparameters [88].

3.2.7. Support Vector Regression

SVR is an ML algorithm that finds a hyperplane that maximizes the margin between the predicted output values and the actual output values while still being able to predict new examples accurately. SVR uses kernel functions to handle non-linear relationships between input and output variables, and it can handle large datasets and high-dimensional input spaces. However, it can be sensitive to the choice of hyperparameters and computationally expensive for large datasets [81].

3.2.8. One-Dimensional Convolution Neural Network

A 1dCNN is a neural network commonly used for processing data sequences, such as time series data. However, in this study, we have used 1dCNN with spectral series data to understand the prediction performance. The network applies a set of learnable filters to the input sequence using a convolutional layer and then downsamples the feature maps using a pooling layer to extract the essential features. One-dimensional CNNs are well-suited for time series data and can handle variable-length sequences [89].

All ML and DL regressor models are tuned to obtain a good prediction performance, and the models tuning parameters are given in Table 1. However, GB and DT provide excellent overall performance on the basic packages of the scikit-study library [90]. After optimizing the hyperparameter of all ML and DL models, training and testing started in Jupyter Notebook Python environments.

Table 1. Hyperparameter optimization for different regression models.

Model	Library	Hyperparameter
RF	scikit-learn	n_estimators = 100, n_jobs = -1
DT	scikit-learn	–
GB	scikit-learn	–
SOM	susi	n_rows = 35, n_columns = 35, n_iter_unsupervised = 10,000, n_iter_supervised = 10,000
KNN	scikit-learn	n_neighbors = 5, algorithm = auto,
ANN	scikit-learn	hidden_layer_sizes = (20, 20, 20)
SVR	scikit-learn	C = np.logspace(-8, 8, 17), γ = np.logspace(-8, 8, 17)
1dCNN	PyTorch	activation = LeakyRelu(), optimiser = Adam, learning rate = 0.01 batch size = 1000, epochs = 500

3.3. Evaluation Parameter

The performance of each model was evaluated by calculating the coefficient of determination (R^2), mean absolute error (MAE), and root means squared error (RMSE).

R^2 describes how well the data fits compared to the ground truth data. R^2 value ranging from 0 to 1 and expressed as a percentage. MAE defines the absolute difference between the predicted value and the original values calculated from the average difference from the whole dataset. Mean squared error (MSE) is the difference between the predicted and the actual value calculated by squaring the difference from the entire dataset. RMSE is the error rate calculated by the square root of MSE. The closer the value of R^2 is 1, the more excellent correlation between the actual and predicted value. The lower the value of MAE and RMSE, the better the prediction accuracy [91].

The mathematical formulation of these terms is given as follows:

$$R^2 = 1 - \frac{\sum (y_{true} - y_{\hat{pred}})^2}{\sum (y_{true} - y_{\bar{avg}})^2}, \quad (3)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i^{true} - y_i^{\hat{pred}}|, \quad (4)$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i^{true} - y_i^{\hat{pred}})^2}, \quad (5)$$

where, y_{true} and $y_{\hat{pred}}$ is the original and predicted value, $y_{\bar{avg}}$ is the average value, N is the number of samples.

4. Prediction Results

This section presents a comparative study of seven different ML models and one DL model to predict SM, SOC, and NC from the three different datasets. The main objective is to develop a model to perform satisfactorily on unseen data and eliminate under-fitting and overfitting problems. Thus, to make the model more robust, a quite limited number of cross-validations has been considered to understand the average prediction performance. We divided the entire dataset into ten groups, where nine groups were used for training the model, and the remaining group was used for testing, as depicted in Figure 5. We repeated this process ten times, and the model's performance was recorded for each testing data set. We then calculated each ML and DL model's mean prediction accuracy (Equation (6)). We experimented on the three HS datasets to predict SM, SOC, and NC.

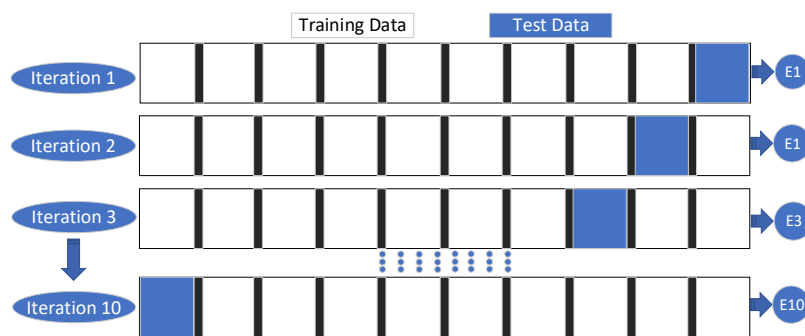


Figure 5. Schematic diagram of 10-fold cross-validation for model evaluation.

$$R_{average} = \frac{1}{10} \sum_{iteration=1}^{10} R_{iteration} \tag{6}$$

4.1. Soil Moisture Prediction from Visual Band

The prediction performance of SM with five different feature combinations is given in Table 2. A comparative study is also included when all the HS band (454 nm to 950 nm) was considered to predict SM [26]. This table shows that when only three RGB channels have been considered an input feature to predict SM, RF provides the best prediction accuracy with R^2 , MAE, and RMSE 84.83%, 0.86, and 1.37, respectively. After that, we consider the whole range of VB (454 nm to 700 nm) to observe the SM prediction performance. The prediction accuracy is improved for all the ML and DL models, and the best results are obtained for SVR with 93.89% prediction accuracy. The prediction performance is further enhanced when PCA is applied to VB, and the best result was obtained for RF ($R^2 = 95.51\%$, MAE = 0.53, RMSE = 0.88). This study performs EMD on VB, taking the maximum variance as an input feature with the VB reflectance data. Finally, the PCA of EMD with VB has been studied to understand the SM prediction performance. However, EMD does not have an excellent contribution to improving the results.

Table 2. Soil moisture prediction of different ML and DL models regression results (R^2 , MAE, RMSE). The best result is in bold format considering other models.

Model	AHSB ¹ [26]	RGB ²	VB ³	PCA (VB)	EMD + VB	PCA (EMD + VB)
RF	92.35%, 0.61, 1.02	84.83%, 0.86, 1.37	91.47%, 0.63, 0.98	95.51% , 0.53, 0.88	90.73%, 0.70, 1.04	92.68%, 0.60, 1.04
DT	88.48%, 0.66, 1.34	78.75%, 0.87, 1.65	81.32%, 0.81, 1.64	94.44%, 0.57, 1.17	84.15%, 0.72, 1.41	87.32%, 0.72, 1.44
GB	92.27%, 0.64, 1.09	81.19%, 0.95, 1.51	91.71%, 0.63, 1.03	95.42%, 0.52, 0.81	91.96%, 0.66, 1.04	93.56%, 0.58, 0.87
SOM	89.81%, 0.72, 1.10	86.85%, 0.85, 1.27	90.67%, 0.70, 1.06	92.36%, 0.61, 0.92	81.81%, 1.00, 1.47	82.39%, 0.92, 1.38
KNN	90.95%, 0.68, 1.08	82.29%, 0.93, 1.46	90.72%, 0.70, 1.11	91.14%, 0.62, 1.06	76.11%, 1.08, 1.63	79.68%, 0.96, 1.57
ANN	64.93%, 1.42, 2.07	46.12%, 1.91, 2.58	49.60%, 1.74, 2.38	90.94%, 0.70, 0.95	53.55%, 1.71, 2.34	88.15%, 0.91, 1.21
SVR	95.43%, 0.49, 0.80	73.43%, 1.29, 1.85	93.89%, 0.57, 0.94	91.29%, 0.62, 1.01	84.81%, 0.83, 1.33	89.13%, 0.77, 1.15
1dCNN	-	72.87%, 0.12, 0.20	79.95%, 0.07 , 0.09	-	79.96%, 0.07, 0.09	-

¹ AHSB = All HS bands (454 nm to 950 nm), ² RGB = Red green blue (3 channel), ³ VB = Visual bands.

Figure 6 presents a box plot comparison for seven different ML models based on three criteria: i. VB, ii. PCA of VB, and iii. PCA of EMD and VB. This box plot reflects the outcome of the ten-fold cross-validation utilized for model validation. A circle, cross-line, and diamond on the box represent the mean, median, and outliers. This figure shows that PCA of VB is the most significant feature combination, and RF exhibits the highest average prediction accuracy with minimal fluctuation in the prediction range.

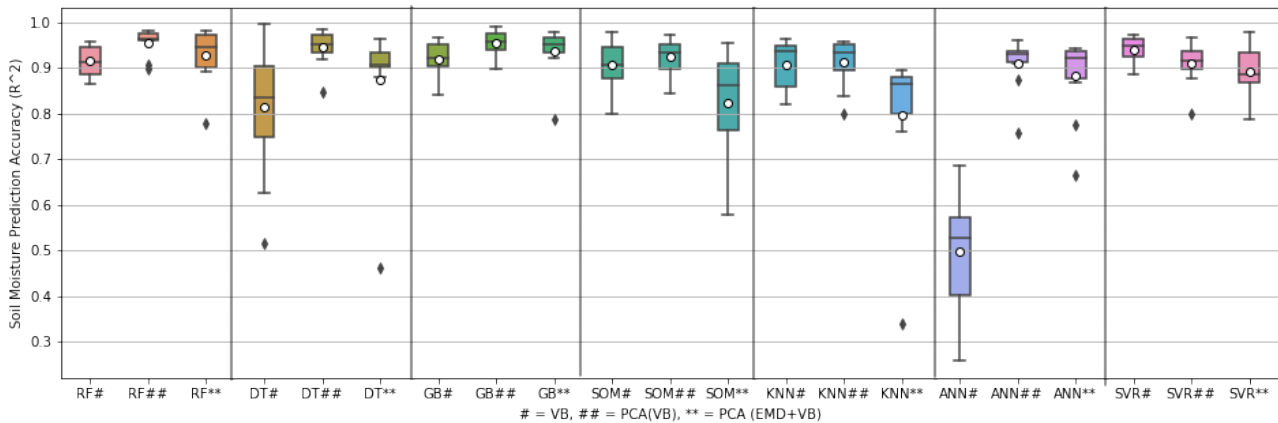


Figure 6. Soil moisture prediction range in box-plot, where circle and diamond represent mean and outliers, respectively.

4.2. Soil Organic Carbon Prediction from Visual Band

We investigated the potential of using our proposed methodology to predict SOC from Sweden’s LUCAS dataset; a comparative study was listed from paper [26] when all HS band (400 nm to 2500 nm) was considered, shown in Table 3. Here, we also studied five feature combinations to predict SOC. When only RGB 3-channel band was considered, the SVR model provides the best prediction accuracy ($R^2 = 79.48\%$, $MAE = 39.75$, $RMSE = 65.72$). The prediction accuracy for all the ML and DL models improved when the full range of VB (400 nm to 700 nm) was considered. When PCA was done on VB, the best performance was noted for SVR with 85.77% prediction accuracy. The contribution of EMD is evaluated on SOC prediction, and the overall best performance is obtained for SVR (85.97%) when PVA was done on EMD with VB.

Figure 7 shows the comparison box plot for the seven different ML models with three feature combinations. This figure shows that SVR performed the best prediction when PCA was done on EMD with VB.

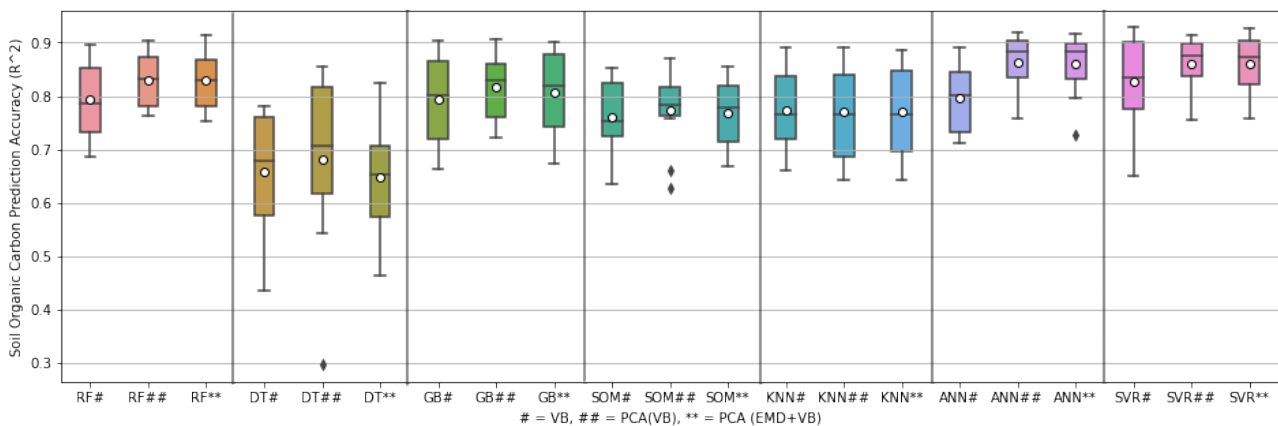


Figure 7. Soil organic carbon prediction range in box-plot, where circle and diamond represent mean and outliers, respectively.

Table 3. Soil organic carbon prediction of different ML and DL models regression results (R^2 , MAE, RMSE). The best result is in bold format considering other models.

Model	AHSB ¹ [26]	RGB ²	VB ³	PCA (VB)	EMD + VB	PCA (EMD + VB)
RF	83.93%, 35.13, 62.46	73.47%, 43.72, 76.79	79.34%, 38.34, 67.80	82.91%, 36.45, 62.71	79.63%, 38.59, 68.04	82.97%, 36.34, 62.46
DT	67.43%, 42.94, 77.95	52.30%, 55.82, 103.11	65.73%, 47.97, 86.04	68.23%, 49.24, 88.01	66.09%, 47.52, 86.54	64.77%, 52.14, 91.95
GB	81.56%, 36.08, 64.15	73.77%, 43.22, 76.08	79.33%, 37.99, 67.59	81.61%, 37.89, 65.82	79.38%, 38.45, 67.69	80.70%, 37.64, 64.44
SOM	78.97%, 41.02, 67.07	75.97%, 43.64, 74.78	76.02%, 40.64, 69.76	77.23%, 42.23, 71.96	77.10%, 41.13, 71.42	76.56%, 42.89, 72.32
KNN	83.08%, 35.92, 60.92	73.96%, 43.02, 75.82	77.44%, 39.30, 69.60	77.04%, 39.67, 70.05	77.44%, 39.30, 69.60	83.56%, 35.76, 61.50
ANN	80.33%, 38.58, 62.58	71.12%, 48.49, 81.99	79.58%, 42.95, 71.64	85.31%, 33.17, 55.42	74.88%, 40.21, 68.38	85.91%, 32.19, 53.19
SVR	78.94%, 48.06, 67.39	79.48%, 39.75, 65.72	82.74%, 33.50, 64.14	85.77%, 33.12, 55.38	81.76%, 36.58, 64.27	85.97% , 32.93, 56.80
1dCNN	-	76.84%, 0.10, 0.14	78.22%, 0.07 , 0.12	-	77.80%, 0.07, 0.13	-

¹ AHSB = All HS bands (400 nm to 2500 nm), ² RGB = Red green blue (3 channel), ³ VB = Visual bands.

4.3. Soil Nitrogen Prediction from Visual Band

A similar study has been done to predict NC from LUCAS (Sweden) dataset, and a comparative analysis is included from paper [26] in Table 4. When three bands of RGB were considered, SVR provided the best prediction accuracy ($R^2 = 72.43%$, MAE = 1.90, RMSE = 3.23). The best prediction results were obtained for SVR (79.61%) when the VB range was considered. The PCA of VB and EMD was also studied to predict NC, and most of the ML models showed improved performance when EMD was considered. Figure 8 shows the comparison box plot of predicting NC. From this figure, SVR performed best when PCA was done on EMD with VB.

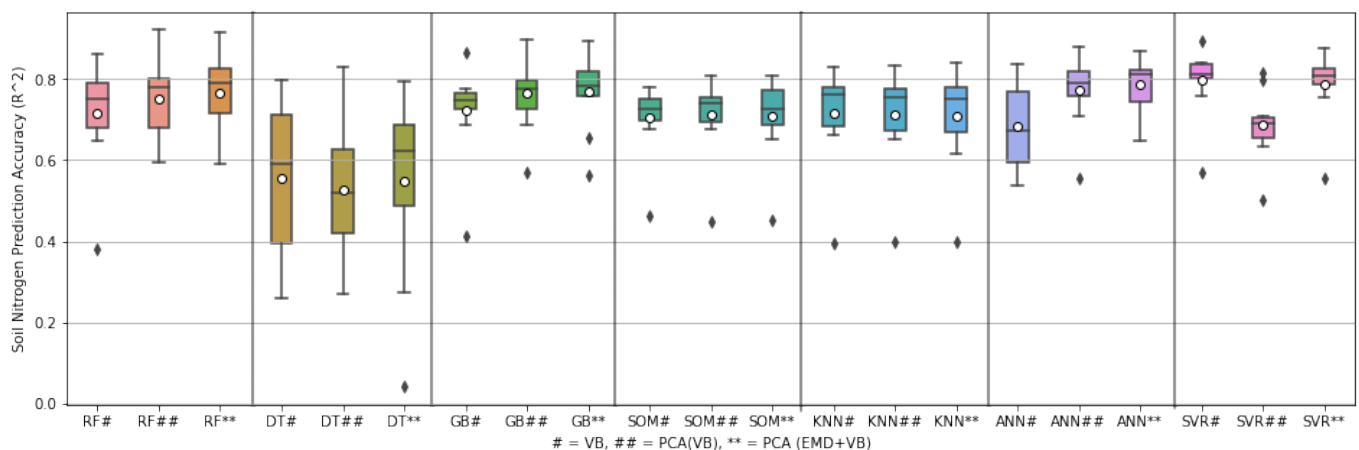


Figure 8. Soil nitrogen prediction range in box-plot, where circle and diamond represent mean and outliers, respectively.

Table 4. Soil nitrogen prediction of different ML and DL models regression results (R^2 , MAE, RMSE). The best result is in bold format considering other models.

Model	AHSB ¹ [26]	RGB ²	VB ³	PCA (VB)	EMD + VB	PCA (EMD + VB)
RF	73.98%, 1.80, 3.03	67.84%, 2.03, 3.37	71.53%, 1.93, 3.29	75.09%, 1.75, 2.90	70.86%, 1.98, 3.29	76.69%, 1.72, 2.84
DT	56.60%, 2.29, 3.89	49.61%, 2.72, 4.57	55.49%, 2.46, 4.17	52.87%, 2.26, 3.95	54.66%, 2.48, 4.24	53.98%, 2.29, 3.80
GB	71.91%, 1.80, 3.01	68.15%, 2.12, 3.52	72.21%, 1.96, 3.26	76.43%, 1.80, 2.91	72.77%, 1.95, 3.28	76.88%, 1.78, 2.85
SOM	74.71%, 1.87, 3.01	70.54%, 1.99, 3.40	70.41%, 1.93, 3.24	71.07%, 1.96, 3.33	70.61%, 1.88, 3.20	70.89%, 1.94, 3.26
KNN	73.96%, 1.83, 2.95	67.39%, 2.02, 3.38	71.61%, 1.96, 3.27	71.16%, 1.96, 3.26	71.61%, 1.96, 3.27	70.82%, 1.94, 3.22
ANN	74.30%, 1.80, 2.80	70.64%, 2.04, 3.41	68.33%, 1.99, 3.18	77.20%, 1.76, 2.79	68.72%, 2.17, 3.25	78.54%, 1.78, 2.89
SVR	71.64%, 2.09, 2.99	72.43%, 1.90, 3.23	79.61% , 1.59, 2.75	68.75%, 2.23, 3.40	75.86%, 1.76, 2.99	78.74%, 1.64, 2.82
1dCNN	-	70.23%, 0.05, 0.10	74.74%, 0.04 , 0.08	-	73.59%, 0.05, 0.08	-

¹ AHSB = All HS bands (400 nm to 2500 nm), ² RGB = Red green blue (3 channel), ³ VB = Visual bands.

5. Discussion

The primary purpose of this study is to understand how accurately we can predict SM, SOC, and NC from VB data. Previous studies [26] have explored the potential of using HS bands for soil property prediction using a number of ML models. However, HS sensors are expensive and require specialized training, limiting their widespread use for soil analysis. In contrast, VB data can be acquired using readily available devices, such as smartphones or digital cameras, making it an attractive option for rapid soil testing. The following observations can be summarised from the results discussed in Section 4.

- Soil properties (SM, SOC, and NC) can be predicted only using 3-channel RGB band data.
- Improved prediction performance for most of the ML and DL models can be obtained when the full range of VB has been considered.
- For SM prediction, the best prediction accuracy is obtained for RF when PCA of VB is considered an input feature.
- EMD plays a good role in SOC prediction; the best result is obtained for SVR when PCA is performed on EMD transformation on VB.
- The good NC prediction is obtained for SVR when only VB is taken as an input feature.
- More stable NC prediction performance (Figure 8) is noted for SVR when PCA is done on EMD with VB.
- The performance of 1dCNN is satisfactory for all of those soil properties predictions in terms of error rate.

Table 5 shows the summary and provides a clear decision about the learning-based models and corresponding prediction accuracy. If we have only the 3-channel RGB bands, we can also predict SM, SOC, and NC with 86.85%, 79.48%, and 72.43% prediction accuracy, respectively. However, accuracy was lost when we used only the 3-channel RGB band and a full spectrum of VB compared with the AHSB. For SM prediction, the best performance was obtained for SVR (95.43%) [26], whereas when only the RGB band was used, the best prediction was obtained for SOM (86.85%), indicating a 9% accuracy loss. However, with the availability of the VB band with some mathematical preprocessing, RF noted a satisfactory improvement in SM prediction accuracy (95.51%).

Table 5. Soil properties prediction results with specific feature combinations.

Soil Property	Best Model (RGB)	Accuracy (RGB) (R^2)	Best Feature & Model	Best Result (R^2)
Soil Moisture	SOM	86.85%	PCA (VB), RF	95.51%
Organic Carbon	SVR	79.48%	PCA (EMD + VB), SVR	85.97%
Nitrogen Content	SVR	72.43%	PCA (EMD + VB), SVR	78.74%

A similar pattern was also observed for SOC and NC prediction. When AHSB was available, RF recorded 83.93% prediction accuracy [26]. However, with only the RGB band, the prediction accuracy was reduced by 5%, and SVR obtained the best result (79.48%). The prediction accuracy was improved with the full VB range, and the SVR model provided the best prediction accuracy. ANN recorded 74.30% prediction accuracy when AHSB was considered for NC prediction [26]. When RGB was used, only 2.5% prediction accuracy was dropped (SVR), and the best result was recorded for SVR (78.74%) with some mathematical preprocessing when VB was considered.

In summary, the results of this study suggest that VB data can be used to predict SM, SOC, and NC, even using only the 3-channel RGB bands. In this study, we have used three popular datasets, and we hope that our proposed approach is also suitable for other datasets and regions. This methodology can provide a rapid and cost-effective alternative to traditional soil analysis methods, enabling more frequent and widespread soil testing. Additionally, more advanced ML and DL models and preprocessing techniques can improve prediction accuracy, opening up new opportunities for precision agriculture and environmental monitoring.

6. Conclusions

This study explores the possibility of predicting SM, SOC, and NC from VB data. The results showed that the 3-channel RGB bands provide comparable results to predict those soil properties, so traditional imaging systems (RGB band) can be used. Improved prediction performance can be achieved when the full range of VB data is considered as an input feature. The EMD and PCA have some roles in improving the performance of ML and DL models. The best-performing models and feature combinations were identified for each soil property, and their corresponding prediction accuracies were presented. This study provides valuable insights into the potential of using low-cost and easily accessible VB data for rapid soil testing.

The results of this study have practical implications in precision agriculture and environmental management. With the increasing demand for sustainable agricultural practices, accurate and rapid soil testing is crucial to optimize crop yields and reduce environmental impacts. The findings of this study suggest that VB data can be used as an alternative to expensive HS data for soil testing. The learning-based prediction models developed in this study can be used in smartphone applications or low-cost soil sensors for on-site soil testing, enabling farmers to make informed decisions about crop management practices.

Overall, this study highlights the potential of VB data for soil testing and provides a roadmap for future research in this area. Future studies could focus on generalizing the developed models across different soil types and locations and integrating other data sources, such as weather, on improving the accuracy of soil property predictions.

Author Contributions: Conceptualization, D.D.; methodology, D.D.; software, D.D.; validation, D.D., M.P., M.M., S.W.T. and L.S.; formal analysis, D.D.; investigation, D.D.; resources, D.D., M.P., M.M., S.W.T. and L.S.; data curation, D.D.; writing—original draft preparation, D.D.; writing—review and editing, D.D., M.P., M.M., S.W.T. and L.S.; visualization, D.D., M.P., M.M., S.W.T. and L.S.; supervision, D.D., M.P., M.M., S.W.T. and L.S.; project administration, M.P.; funding acquisition, M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Soil CRC Australia (No. 2.S.006 PhD Scholarship), with contribution from Charles Sturt University.

Data Availability Statement: Not applicable.

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

References

1. Yadav, A.N.; Singh, J.; Singh, C.; Yadav, N. *Current Trends in Microbial Biotechnology for Sustainable Agriculture*; Springer: Berlin/Heidelberg, Germany, 2021.
2. Herrick, J.E.; Wander, M.M. Relationships between soil organic carbon and soil quality in cropped and rangeland soils: The importance of distribution, composition, and soil biological activity. In *Soil Processes and the Carbon Cycle*; CRC Press: Boca Raton, FL, USA, 2018; pp. 405–425.
3. Fageria, N.; Moreira, A. The role of mineral nutrition on root growth of crop plants. *Adv. Agron.* **2011**, *110*, 251–331.
4. Denmead, O.T.; Shaw, R.H. Availability of soil water to plants as affected by soil moisture content and meteorological conditions 1. *Agron. J.* **1962**, *54*, 385–390. [[CrossRef](#)]
5. Wang, W.; Akhtar, K.; Ren, G.; Yang, G.; Feng, Y.; Yuan, L. Impact of straw management on seasonal soil carbon dioxide emissions, soil water content, and temperature in a semi-arid region of China. *Sci. Total Environ.* **2019**, *652*, 471–482. [[CrossRef](#)] [[PubMed](#)]
6. Leghari, S.J.; Wahocho, N.A.; Laghari, G.M.; HafeezLaghari, A.; MustafaBhabhan, G.; HussainTalpur, K.; Bhutto, T.A.; Wahocho, S.A.; Lashari, A.A. Role of nitrogen for plant growth and development: A review. *Adv. Environ. Biol.* **2016**, *10*, 209–219.
7. Njira, K.O.; Nabwami, J. A review of effects of nutrient elements on crop quality. *Afr. J. Food Agric. Nutr. Dev.* **2015**, *15*, 9777–9793. [[CrossRef](#)]
8. Verma, B.C.; Datta, S.P.; Rattan, R.K.; Singh, A.K. Monitoring changes in soil organic carbon pools, nitrogen, phosphorus, and sulfur under different agricultural management practices in the tropics. *Environ. Monit. Assess.* **2010**, *171*, 579–593. [[CrossRef](#)]
9. Xu, S.; Liu, Y.; Wang, X.; Zhang, G. Scale effect on spatial patterns of ecosystem services and associations among them in semi-arid area: A case study in Ningxia Hui Autonomous Region, China. *Sci. Total Environ.* **2017**, *598*, 297–306. [[CrossRef](#)]
10. Kumar, S.V.; Dirmeyer, P.A.; Peters-Lidard, C.D.; Bindlish, R.; Bolten, J. Information theoretic evaluation of satellite soil moisture retrievals. *Remote Sens. Environ.* **2018**, *204*, 392–400. [[CrossRef](#)]
11. Visconti, F.; Jiménez, M.G.; de Paz, J.M. How do the chemical characteristics of organic matter explain differences among its determinations in calcareous soils? *Geoderma* **2022**, *406*, 115454. [[CrossRef](#)]
12. McGill, W.; Figueiredo, C. Total nitrogen. In *Soil Sampling and Methods of Analysis*; CRC: Boca Raton, FL, USA, 1993; pp. 201–211.
13. Govender, M.; Chetty, K.; Bulcock, H. A review of hyperspectral remote sensing and its application in vegetation and water resource studies. *Water SA* **2007**, *33*, 145–151. [[CrossRef](#)]
14. Zhang, Y.; Migliavacca, M.; Penuelas, J.; Ju, W. Advances in hyperspectral remote sensing of vegetation traits and functions. *Remote Sens. Environ.* **2021**, *252*, 112121. [[CrossRef](#)]
15. Stuart, M.B.; McGonigle, A.J.; Willmott, J.R. Hyperspectral imaging in environmental monitoring: A review of recent developments and technological advances in compact field deployable systems. *Sensors* **2019**, *19*, 3071. [[CrossRef](#)] [[PubMed](#)]
16. Liu, Y.; Pu, H.; Sun, D.W. Hyperspectral imaging technique for evaluating food quality and safety during various processes: A review of recent applications. *Trends Food Sci. Technol.* **2017**, *69*, 25–35. [[CrossRef](#)]
17. Krishna, G.; Sahoo, R.N.; Singh, P.; Bajpai, V.; Patra, H.; Kumar, S.; Dandapani, R.; Gupta, V.K.; Viswanathan, C.; Ahmad, T.; et al. Comparison of various modelling approaches for water deficit stress monitoring in rice crop through hyperspectral remote sensing. *Agric. Water Manag.* **2019**, *213*, 231–244. [[CrossRef](#)]
18. Peyghambari, S.; Zhang, Y. Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: An updated review. *J. Appl. Remote Sens.* **2021**, *15*, 031501. [[CrossRef](#)]
19. Edelman, G.J.; Gaston, E.; Van Leeuwen, T.G.; Cullen, P.; Aalders, M.C. Hyperspectral imaging for non-contact analysis of forensic traces. *Forensic Sci. Int.* **2012**, *223*, 28–39. [[CrossRef](#)]
20. Suzuki, S.; Matsui, T. Remote sensing for medical and health care applications. In *Remote Sensing—Applications*; Boris, E., Ed.; BoD—Books on Demand: Norderstedt, Germany, 2012; pp. 479–492.
21. Fei, B. Hyperspectral imaging in medical applications. In *Data Handling in Science and Technology*; Elsevier: Amsterdam, The Netherlands, 2020; Volume 32, pp. 523–565.
22. Weber, C.; Aguejdad, R.; Briottet, X.; Avala, J.; Fabre, S.; Demuynek, J.; Zenou, E.; Deville, Y.; Karoui, M.S.; Benhalouche, F.Z.; et al. Hyperspectral imagery for environmental urban planning. In *Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018*; pp. 1628–1631.
23. Shimoni, M.; Haelterman, R.; Perneel, C. Hypersectral imaging for military and security applications: Combining myriad processing and sensing techniques. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 101–117. [[CrossRef](#)]
24. Boubanga-Tombet, S.; Huot, A.; Vitins, I.; Heuberger, S.; Veuve, C.; Eisele, A.; Hewson, R.; Guyot, E.; Marcotte, F.; Chamberland, M. Thermal infrared hyperspectral imaging for mineralogy mapping of a mine face. *Remote Sens.* **2018**, *10*, 1518. [[CrossRef](#)]
25. Majda, A.; Wietecha-Posluszny, R.; Mendys, A.; Wójtowicz, A.; Łydźba-Kopczyńska, B. Hyperspectral imaging and multivariate analysis in the dried blood spots investigations. *Appl. Phys. A* **2018**, *124*, 1–8. [[CrossRef](#)]

26. Datta, D.; Paul, M.; Murshed, M.; Teng, S.W.; Schmidtke, L. Soil Moisture, Organic Carbon, and Nitrogen Content Prediction with Hyperspectral Data Using Regression Models. *Sensors* **2022**, *22*, 7998. [[CrossRef](#)]
27. Riese, F.M.; Keller, S. Introducing a framework of self-organizing maps for regression of soil moisture with hyperspectral data. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 6151–6154.
28. Ge, X.; Ding, J.; Jin, X.; Wang, J.; Chen, X.; Li, X.; Liu, J.; Xie, B. Estimating agricultural soil moisture content through UAV-based hyperspectral images in the arid region. *Remote Sens.* **2021**, *13*, 1562. [[CrossRef](#)]
29. Xu, C.; Zeng, W.; Huang, J.; Wu, J.; Van Leeuwen, W.J. Prediction of soil moisture content and soil salt concentration from hyperspectral laboratory and field data. *Remote Sens.* **2016**, *8*, 42. [[CrossRef](#)]
30. Castaldi, F.; Palombo, A.; Pascucci, S.; Pignatti, S.; Santini, F.; Casa, R. Reducing the influence of soil moisture on the estimation of clay from hyperspectral data: A case study using simulated PRISMA data. *Remote Sens.* **2015**, *7*, 15561–15582. [[CrossRef](#)]
31. Zhang, F.; Wu, S.; Liu, J.; Wang, C.; Guo, Z.; Xu, A.; Pan, K.; Pan, X. Predicting soil moisture content over partially vegetation covered surfaces from hyperspectral data with deep learning. *Soil Sci. Soc. Am. J.* **2021**, *85*, 989–1001. [[CrossRef](#)]
32. Haijun, Q.; Xiu, J.; Liu, Z.; Maxime, D.I.; Shaowen, L. Predicting sandy soil moisture content with hyperspectral imaging. *Int. J. Agric. Biol. Eng.* **2017**, *10*, 175–183. [[CrossRef](#)]
33. Wu, T.; Yu, J.; Lu, J.; Zou, X.; Zhang, W. Research on inversion model of cultivated soil moisture content based on hyperspectral imaging analysis. *Agriculture* **2020**, *10*, 292. [[CrossRef](#)]
34. Yuan, J.; Wang, X.; Yan, C.x.; Wang, S.r.; Ju, X.p.; Li, Y. Soil moisture retrieval model for remote sensing using reflected hyperspectral information. *Remote Sens.* **2019**, *11*, 366. [[CrossRef](#)]
35. Nocita, M.; Stevens, A.; Toth, G.; Panagos, P.; van Wesemael, B.; Montanarella, L. Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression approach. *Soil Biol. Biochem.* **2014**, *68*, 337–347. [[CrossRef](#)]
36. Steinberg, A.; Chabrilat, S.; Stevens, A.; Segl, K.; Foerster, S. Prediction of common surface soil properties based on Vis-NIR airborne and simulated EnMAP imaging spectroscopy data: Prediction accuracy and influence of spatial resolution. *Remote Sens.* **2016**, *8*, 613. [[CrossRef](#)]
37. Hou, L.; Zheng, Y.; Liu, M.; Li, X.; Lin, X.; Yin, G.; Gao, J.; Deng, F.; Chen, F.; Jiang, X. Anaerobic ammonium oxidation and its contribution to nitrogen removal in China’s coastal wetlands. *Sci. Rep.* **2015**, *5*, 15621. [[CrossRef](#)]
38. Xu, S.; Wang, M.; Shi, X.; Yu, Q.; Zhang, Z. Integrating hyperspectral imaging with machine learning techniques for the high-resolution mapping of soil nitrogen fractions in soil profiles. *Sci. Total Environ.* **2021**, *754*, 142135. [[CrossRef](#)] [[PubMed](#)]
39. Pechanec, V.; Mráz, A.; Rozkošný, L.; Vyvlečka, P. Usage of airborne hyperspectral imaging data for identifying spatial variability of soil nitrogen content. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 355. [[CrossRef](#)]
40. Coops, N.C.; Smith, M.L.; Martin, M.E.; Ollinger, S.V. Prediction of eucalypt foliage nitrogen content from satellite-derived hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1338–1346. [[CrossRef](#)]
41. Vohland, M.; Ludwig, M.; Thiele-Bruhn, S.; Ludwig, B. Quantification of soil properties with hyperspectral data: Selecting spectral variables with different methods to improve accuracies and analyze prediction mechanisms. *Remote Sens.* **2017**, *9*, 1103. [[CrossRef](#)]
42. Pudełko, A.; Chodak, M.; Roemer, J.; Uhl, T. Application of FT-NIR spectroscopy and NIR hyperspectral imaging to predict nitrogen and organic carbon contents in mine soils. *Measurement* **2020**, *164*, 108117. [[CrossRef](#)]
43. Pan, T.; Wu, Z.-T.; Chen, H. Waveband optimization for near-infrared spectroscopic analysis of total nitrogen in soil. *Chin. J. Anal. Chem.* **2012**, *40*, 920–924. [[CrossRef](#)]
44. Kuang, B.; Mouazen, A.M. Non-biased prediction of soil organic carbon and total nitrogen with vis-NIR spectroscopy, as affected by soil moisture content and texture. *Biosyst. Eng.* **2013**, *114*, 249–258. [[CrossRef](#)]
45. Fatholouloumi, S.; Vaezi, A.R.; Alavipanah, S.K.; Ghorbani, A.; Saurette, D.; Biswas, A. Effect of multi-temporal satellite images on soil moisture prediction using a digital soil mapping approach. *Geoderma* **2021**, *385*, 114901. [[CrossRef](#)]
46. Drusch, M. Initializing numerical weather prediction models with satellite-derived surface soil moisture: Data assimilation experiments with ECMWF’s Integrated Forecast System and the TMI soil moisture data set. *J. Geophys. Res. Atmos.* **2007**, *112*, 3102. [[CrossRef](#)]
47. Baldwin, D.; Manfreda, S.; Keller, K.; Smithwick, E. Predicting root zone soil moisture with soil properties and satellite near-surface moisture data across the conterminous United States. *J. Hydrol.* **2017**, *546*, 393–404. [[CrossRef](#)]
48. Laiolo, P.; Gabellani, S.; Campo, L.; Silvestro, F.; Delogu, F.; Rudari, R.; Pulvirenti, L.; Boni, G.; Fascetti, F.; Pierdicca, N.; et al. Impact of different satellite soil moisture products on the predictions of a continuous distributed hydrological model. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *48*, 131–145. [[CrossRef](#)]
49. Pinnington, E.; Amezcua, J.; Cooper, E.; Dadson, S.; Ellis, R.; Peng, J.; Robinson, E.; Morrison, R.; Osborne, S.; Quaife, T. Improving soil moisture prediction of a high-resolution land surface model by parameterising pedotransfer functions through assimilation of SMAP satellite data. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 1617–1641. [[CrossRef](#)]
50. Tiwari, S.K.; Saha, S.K.; Kumar, S. Prediction modeling and mapping of soil carbon content using artificial neural network, hyperspectral satellite data and field spectroscopy. *Adv. Remote Sens.* **2015**, *4*, 63. [[CrossRef](#)]
51. Meng, X.; Bao, Y.; Liu, J.; Liu, H.; Zhang, X.; Zhang, Y.; Wang, P.; Tang, H.; Kong, F. Regional soil organic carbon prediction model based on a discrete wavelet analysis of hyperspectral satellite data. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *89*, 102111. [[CrossRef](#)]

52. Dou, X.; Wang, X.; Liu, H.; Zhang, X.; Meng, L.; Pan, Y.; Yu, Z.; Cui, Y. Prediction of soil organic matter using multi-temporal satellite images in the Songnen Plain, China. *Geoderma* **2019**, *356*, 113896. [[CrossRef](#)]
53. Zhou, T.; Geng, Y.; Ji, C.; Xu, X.; Wang, H.; Pan, J.; Bumberger, J.; Haase, D.; Lausch, A. 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. *Sci. Total Environ.* **2021**, *755*, 142661. [[CrossRef](#)]
54. Wang, S.; Zhuang, Q.; Jin, X.; Yang, Z.; Liu, H. Predicting soil organic carbon and soil nitrogen stocks in topsoil of forest ecosystems in northeastern china using remote sensing data. *Remote Sens.* **2020**, *12*, 1115. [[CrossRef](#)]
55. Pande, C.B.; Kadam, S.A.; Jayaraman, R.; Gorantiwar, S.; Shinde, M. Prediction of soil chemical properties using multispectral satellite images and wavelet transforms methods. *J. Saudi Soc. Agric. Sci.* **2022**, *21*, 21–28. [[CrossRef](#)]
56. Hajjar, C.S.; Hajjar, C.; Esta, M.; Chamoun, Y.G. Machine learning methods for soil moisture prediction in vineyards using digital images. *E3s Web Conf.* **2020**, *167*, 02004. [[CrossRef](#)]
57. Gorthi, S.; Swetha, R.; Chakraborty, S.; Li, B.; Weindorf, D.C.; Dutta, S.; Banerjee, H.; Das, K.; Majumdar, K. Soil organic matter prediction using smartphone-captured digital images: Use of reflectance image and image perturbation. *Biosyst. Eng.* **2021**, *209*, 154–169. [[CrossRef](#)]
58. Taneja, P.; Vasava, H.K.; Daggupati, P.; Biswas, A. Multi-algorithm comparison to predict soil organic matter and soil moisture content from cell phone images. *Geoderma* **2021**, *385*, 114863. [[CrossRef](#)]
59. Swetha, R.; Bende, P.; Singh, K.; Gorthi, S.; Biswas, A.; Li, B.; Weindorf, D.C.; Chakraborty, S. Predicting soil texture from smartphone-captured digital images and an application. *Geoderma* **2020**, *376*, 114562. [[CrossRef](#)]
60. de Oliveira Morais, P.A.; de Souza, D.M.; de Melo Carvalho, M.T.; Madari, B.E.; de Oliveira, A.E. Predicting soil texture using image analysis. *Microchem. J.* **2019**, *146*, 455–463. [[CrossRef](#)]
61. Pendleton, R.L.; Nickerson, D. Soil colors and special Munsell soil color charts. *Soil Sci.* **1951**, *71*, 35–44. [[CrossRef](#)]
62. Wills, S.A.; Burras, C.L.; Sandor, J.A. Prediction of soil organic carbon content using field and laboratory measurements of soil color. *Soil Sci. Soc. Am. J.* **2007**, *71*, 380–388. [[CrossRef](#)]
63. Liles, G.C.; Beaudette, D.E.; O’Geen, A.T.; Horwath, W.R. Developing predictive soil C models for soils using quantitative color measurements. *Soil Sci. Soc. Am. J.* **2013**, *77*, 2173–2181. [[CrossRef](#)]
64. Minasny, B.; McBratney, A.B.; Mendonça-Santos, M.; Odeh, I.; Guyon, B. Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. *Soil Res.* **2006**, *44*, 233–244. [[CrossRef](#)]
65. Li, F.; Frosio, I.; Timofte, R.; Zhu, C. Spectral reflectance reconstruction from RGB images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 23–28 June 2013; pp. 2626–2633.
66. Bailoni, J.; Waske, B.; Benediktsson, J.A.; Ressel, R. Spectral imaging with a consumer camera: A comparison to a scientific camera and opportunities for citizen science. *J. Appl. Remote Sens.* **2017**, *11*, 026015.
67. Tan, R.T.; Kong, H.K.; Quan, L. Visibility in bad weather from a single image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
68. Islam, M.R.; Paul, M.; Antolovich, M.; Kabir, A. Sports Highlights Generation using Decomposed Audio Information. In Proceedings of the 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Shanghai, China, 8–12 July 2019; pp. 579–584.
69. Riese, F.M.; Keller, S. Hyperspectral benchmark dataset on soil moisture. In Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Valencia, Spain, 22–27 July 2018; p. 1227837. [[CrossRef](#)]
70. Orgiazzi, A.; Ballabio, C.; Panagos, P.; Jones, A.; Fernández-Ugalde, O. LUCAS Soil, the largest expandable soil dataset for Europe: A review. *Eur. J. Soil Sci.* **2018**, *69*, 140–153. [[CrossRef](#)]
71. Panagos, P.; Van Liedekerke, M.; Jones, A.; Montanarella, L. European Soil Data Centre: Response to European policy support and public data requirements. *Land Use Policy* **2012**, *29*, 329–338. [[CrossRef](#)]
72. Tóth, G.; Jones, A.; Montanarella, L. The LUCAS topsoil database and derived information on the regional variability of cropland topsoil properties in the European Union. *Environ. Monit. Assess.* **2013**, *185*, 7409–7425. [[CrossRef](#)] [[PubMed](#)]
73. Ballabio, C.; Panagos, P.; Monatanarella, L. Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma* **2016**, *261*, 110–123. [[CrossRef](#)]
74. Apaydin, H.; Sibtain, M. A multivariate streamflow forecasting model by integrating improved complete ensemble empirical mode decomposition with additive noise, sample entropy, Gini index and sequence-to-sequence approaches. *J. Hydrol.* **2021**, *603*, 126831. [[CrossRef](#)]
75. Lawrence, N.; Hyvärinen, A. Probabilistic non-linear principal component analysis with Gaussian process latent variable models. *J. Mach. Learn. Res.* **2005**, *6*, 1783–1816.
76. Liu, Y.; Wang, Y.; Zhang, J. New machine learning algorithm: Random forest. In Proceedings of the International Conference on Information Computing and Applications, Chengde, China, 14–16 September 2012; pp. 246–252.
77. Myles, A.J.; Feudale, R.N.; Liu, Y.; Woody, N.A.; Brown, S.D. An introduction to decision tree modeling. *J. Chemom. J. Chemom. Soc.* **2004**, *18*, 275–285. [[CrossRef](#)]
78. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [[CrossRef](#)]
79. Zhang, Z. Introduction to machine learning: K-nearest neighbors. *Ann. Transl. Med.* **2016**, *4*, 218. [[CrossRef](#)]

80. Kukreja, H.; Bharath, N.; Siddesh, C.; Kuldeep, S. An introduction to artificial neural network. *Int. J. Adv. Res. Innov. Ideas Educ.* **2016**, *1*, 27–30.
81. Awad, M.; Khanna, R. Support vector regression. In *Efficient Learning Machines*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 67–80.
82. Wu, C.; Jiang, P.; Ding, C.; Feng, F.; Chen, T. Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional neural network. *Comput. Ind.* **2019**, *108*, 53–61. [[CrossRef](#)]
83. Sauer, P.; Cootes, T.F.; Taylor, C.J. Accurate Regression Procedures for Active Appearance Models. In Proceedings of the BMVC, Dundee, UK, 29 August–2 September 2011; pp. 1–11.
84. Xu, M.; Watanachaturaporn, P.; Varshney, P.K.; Arora, M.K. Decision tree regression for soft classification of remote sensing data. *Remote Sens. Environ.* **2005**, *97*, 322–336. [[CrossRef](#)]
85. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
86. Purbasari, I.; Puspaningrum, E.; Putra, A. Using self-organizing map (SOM) for clustering and visualization of new students based on grades. *J. Phys. Conf. Ser.* **2020**, *1569*, 022037. [[CrossRef](#)]
87. Larose, D.T.; Larose, C.D. k-Nearest Neighbor Algorithm. Wiley Data and Cybersecurity. 2014. Available online: <https://ieeexplore.ieee.org/abstract/document/10066854> (accessed on 24 March 2023).
88. Agatonovic-Kustrin, S.; Beresford, R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J. Pharm. Biomed. Anal.* **2000**, *22*, 717–727. [[CrossRef](#)] [[PubMed](#)]
89. Datta, D.; Sarkar, N.I. Deep Learning Frameworks for Internet of Things. In *Artificial Intelligence-Based Internet of Things Systems*; Springer: Cham, Switzerland, 2022; pp. 137–161.
90. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
91. Alexander, D.L.; Tropsha, A.; Winkler, D.A. Beware of R²: Simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. *J. Chem. Inf. Model.* **2015**, *55*, 1316–1322. [[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.