agronomy

MDPI

Article

Artificial Intelligence-Based Prediction of Key Textural Properties from LUCAS and ICRAF Spectral Libraries

Mohamed Zakaria Gouda [1], El Mehdi Nagihi [2], Lotfi Khiari [1,3,*], Jacques Gallichand [3] and Mahmoud Ismail [2]

[1] Soil and Fertilizer Research in Africa, Mohammed VI Polytechnic University, Benguerir 43150, Morocco; mohamed.gouda@um6p.ma
[2] Emines School of Industrial Management, Mohammed VI Polytechnic University, Benguerir 43150, Morocco; elmehdi.nagihi@emines.um6p.ma (E.M.N.); mahmoud.ismail@emines.um6p.ma (M.I.)
[3] Department of Soil Science and Agrifood Engineering, Laval University, Quebec, QC G1V 0A6, Canada; Jacques.Gallichand@fsaa.ulaval.ca
* Correspondence: lotfi.khiari@um6p.ma; Tel.: +212-6620-95436

Abstract: Soil texture is a key soil property influencing many agronomic practices including fertilization and liming. Therefore, an accurate estimation of soil texture is essential for adopting sustainable soil management practices. In this study, we used different machine learning algorithms trained on vis–NIR spectra from existing soil spectral libraries (ICRAF and LUCAS) to predict soil textural fractions (sand–silt–clay %). In addition, we predicted the soil textural groups (G1: Fine, G2: Medium, and G3: Coarse) using routine chemical characteristics as auxiliary. With the ICRAF dataset, multilayer perceptron resulted in good predictions for sand and clay (R^2 = 0.78 and 0.85, respectively) and categorical boosting outperformed the other algorithms (random forest, extreme gradient boosting, linear regression) for silt prediction (R^2 = 0.81). For the LUCAS dataset, categorical boosting consistently showed a high performance for sand, silt, and clay predictions (R^2 = 0.79, 0.76, and 0.85, respectively). Furthermore, the soil texture groups (G1, G2, and G3) were classified using the light gradient boosted machine algorithm with a high accuracy (83% and 84% for ICRAF and LUCAS, respectively). These results, using spectral data, are very promising for rapid diagnosis of soil texture and group in order to adjust agricultural practices.

Keywords: textural group; fine, medium and coarse texture; vis-NIR spectrum; dry chemistry; chemometrics; machine learning

Citation: Gouda, M.Z.; Nagihi, E.M.; Khiari, L.; Gallichand, J.; Ismail, M. Artificial Intelligence-Based Prediction of Key Textural Properties from LUCAS and ICRAF Spectral Libraries. *Agronomy* **2021**, *11*, 1550. https://doi.org/10.3390/agronomy11081550

Academic Editor: Baohua Zhang

Received: 12 June 2021
Accepted: 22 July 2021
Published: 3 August 2021

1. Introduction

Soils provide several ecosystem services to support human needs; however, due to the rapid growth of the human population, soils are facing unprecedented pressure through the intensification of agricultural production [1]. Therefore, monitoring the soil status is crucial to support sustainable and high agricultural productivity. Soil analysis is an important key factor for sustainable soil management, as it provides a valuable information about the soil condition prior to making recommendations regarding fertilization, and irrigation, and subsequently improving soil productivity. Gathering precise and detailed information on the soil is essential for crop yield increase [2]. There are many chemical, physical, and biological properties that can affect the soil quality and its function within an agricultural production system [3]. Soil texture, which represents the relative composition of particle sizes (sand: 0.05 to 2 mm, silt: 0.002 to 0.05 mm, and clay: $\leq$0.002 mm [4]), is one the fundamental physical properties. Soil texture usually shows significant spatial variation within a land area. There are many processes affected by soil texture, including plant growth and yield, water retention and infiltration, availability and absorption of plant nutrients, soil organisms and plant root growth, soil quality and productivity, soil temperature, soil compaction, tillage, and the efficiency of fertilizer use and irrigation water [5]. Considering variations in soil texture within a field is now possible with global positioning

system (GPS) and precision agriculture (PA). The aim of these technologies is to divide the field into homogenous management zones suitable for optimal site-specific management of fertilization [6], liming [7], and drainage and irrigation [8]. Consequently, recommending the same management practices irrespective to the textural group can lead to unwanted high crop yield variability. Considering the soil textural group is also necessary to determine the field sampling frequency [9] and sample preparation techniques [10,11]. Precision agriculture, which involves modifying field interventions based on texture, requires a large number of samples to be tested in laboratories.

Laboratories use conventional methods of soil texture testing, such as sieving, hydrometers, and Robinson pipette [12], but such methods are time-consuming, expensive, destructive, require hazardous chemical extractants, and are not suitable for the large number of sample analyses required by precision agriculture. To overcome this obstacle, visible (Vis) and near infrared (NIR) soil analysis techniques have been used as quick, non-destructive, and less expensive soil analysis methods. These methods are efficient for analyzing the large number of samples required for high resolution soil mapping [13]. Over the last few decades, there has been a growing focus on spectral methods, especially with the increasing demand for precision agriculture tools to assist characterization of soils effectively [14,15]. Lately, the application of machine learning (ML) algorithms in soil science has increased. Recent studies used several ML approaches (e.g., support vector machine, artificial neural networks, random forest, partial least squares, and cubist regressions) for predicting soil texture [16,17]. Moreover, ML approaches have paved the way to handle large datasets and produce more accurate predictions of soil properties [18]. Over the last decade, several soil health and safety monitoring research efforts have led to the development of large soil libraries, such as the Land Use/Land Cover Area Frame Survey (LUCAS) [19] and the World Agroforestry Centre (ICRAF) that was established by the International Soil Reference and Information Centre (ISRIC) [20]. These two libraries contain the results of both physical and chemical soil analyses and their vis-NIR spectral signatures. The objective of our study is to use these two libraries, covering a wide range of soils in several countries, to develop a procedure to predict soil textural fractions and groups.

2. Materials and Methods

2.1. Description of Soil Datasets

Both datasets (LUCAS and ICRAF) contain vis-NIR spectral and chemical data from different locations in many countries. LUCAS is a large-scale vis-NIR soil dataset with 19,967 soil samples collected from 23 member states of the European Union under the Land Use/Land Cover Area Frame Survey (LUCAS) that started in 2009 [19]. The standard sampling procedure used around 0.5 kg of topsoil (0–20 cm), which is a composite sample of five subsamples. The first subsample represents the central georeferenced sampling location characterized by the LUCAS point, and the other four subsamples were collected in a cross-shape at 2 m from the central point. These soil samples were dried, sieved, and subjected to spectral measurements using a FOSS XDS Rapid Content Analyzer (NIRSystems, Inc., Laurel, MD, USA), operating with wavelengths ranging from 400 to 2500 nm and a spectral resolution of 0.5 nm. The LUCAS dataset includes various soil properties, such as: coarse fragment percentage, particle size distribution (% of sand, silt, and clay), organic carbon, pH, carbonate content, cation exchange capacity, N, P, extractable K, and metals. The standard analytical methods of soil properties used the ISO methods [21]. The vis-NIR analyses followed the procedure described by the FOSS spectroscope [22].

The ICRAF dataset is composed of 4438 soil samples selected from the soil information system (ISIS) of the International Soil Reference and Information Centre [20]. This library represents soil samples from 58 countries around the world (Africa, Asia, South America, North America, and Europe) [20]. These soil samples were air-dried, clod crushed, and sieved at 2 mm. Soil properties of pH, OC, CEC, exchangeable (K, Ca, Mg), sand, silt, and clay content were determined according to the IRSIC procedures for soil analysis [21]. For the spectral analysis, the soil samples were scanned using a FieldSpec FR spectroradiometer

(Boulder, CO) at spectral wavelengths ranging from 350 to 2500 nm, with a spectral resolution of 1 nm. Figure 1 shows the raw spectra for soil samples of the LUCAS and ICRAF datasets.

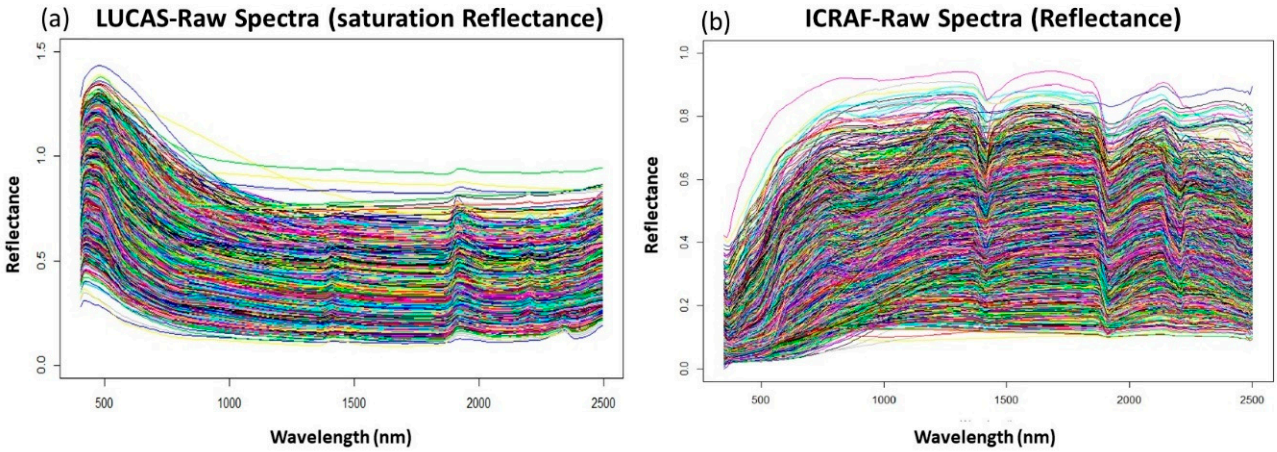

Figure 1. Raw soil spectra for both datasets showing the reflectance against wavelength (nm): (**a**) LUCAS-ESDAC (19,967 samples); (**b**) ICRAF-ISRIC (4438 samples).

2.2. Pre-Processing of Raw Spectra

Pre-processing of spectral data is usually a required step to improve data quality before modeling. This pre-processing step aims at removing undesirable physical phenomena and increasing the signal resolution, as well as maximizing the quality of subsequent data processing. In this study, five spectral pre-processing methods (first derivative, second derivative, continuum removal, detrend, and standard normal variate) were applied on both datasets as shown in Figure 2. The spectral data pre-processing methods were implemented in R using the "prospectr" package [23]. The estimation of the three textural fractions of soil were performed with the h2o AutoML algorithm [24] using the R software (Version 3.0.2) [25]. The best performing pre-processing technique was used for the following steps on the study. The first (1st Der) and second (2nd Der) derivatives can be used to remove the vertical offsets and linearly sloping baselines to produce low error calibration. Converting the raw spectra to 1st Der or 2nd Der aims at accentuating the reflectance features. The 1st Der is effective for removing baseline offset; the 2nd Der is effective for both the baseline offset and linear trend from a spectrum [26,27]. Derivatives also remove both additive and multiplicative effects on the spectra. Continuum removal (CR) was used as a normalization procedure prior to features quantification [28]. Detrend correction (DT) was applied to remove the baseline shifts and curvilinearity. Standard normal variate (SNV) was used to remove the scatter interferences and differences and correct the global intensity effect [29]. A detailed description of these pre-processing methods can be found elsewhere [26].

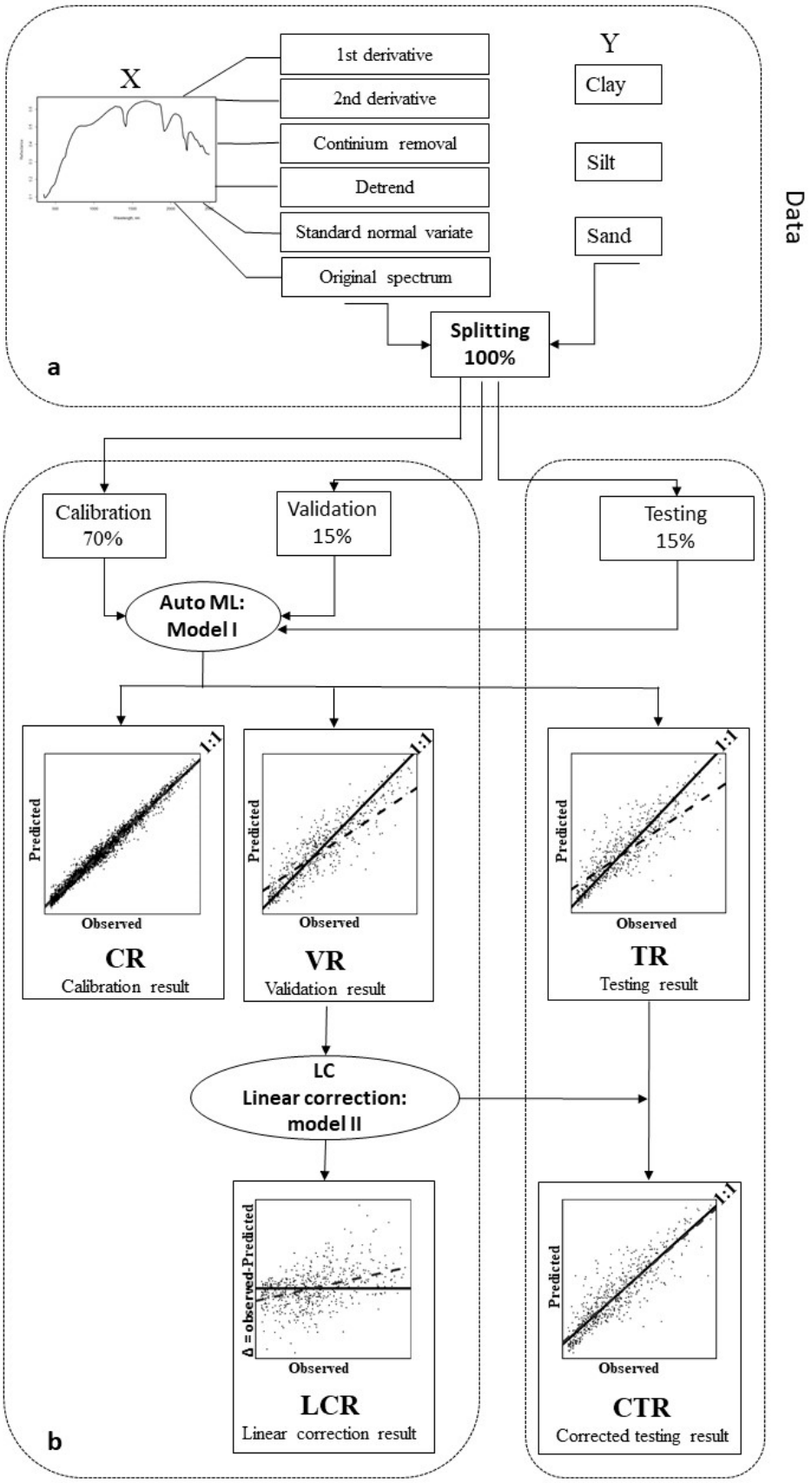

Figure 2. Methodology flowchart: (**a**) the five pre-processing techniques; (**b**) machine learning

models calibration including data splitting into three sets: calibration (70% of data), validation (15% of data) and testing (15% of data) and linear correction of biased residuals to predict the three soil texture fractions (sand, silt, and clay %).

2.3. Machine Learning Algorithms for the Prediction of Clay, Silt and Sand Fractions

Using spectral libraries with machine learning is important because it can analyze more data while delivering faster and more accurate results and help agricultural, environmental and ecosystemic services to strengthen their capacity to estimate soil properties and intensify soil characterization.

This modeling process is aimed at obtaining the particle size fractions by analyzing the spectral features contained in the spectrum acquired using the spectroradiometer scanning of a sample. As a soil spectral signature produces a very large number of spectral data (1250 pairs of wavelength and corresponding reflectance) and reflectances are often autocorrelated, machine learning procedures are the most appropriate option to correlate these 1250 input parameters with a single output factor: a particle size fraction (clay, silt, or sand).

Three steps are required to build a machine learning model: (i) selection of a machine learning algorithm, (ii) training the algorithm on large datasets to create a model that converges to an optimal accuracy, and (iii) validating the performance of the resulting model on a testing dataset.

These machine learning models must be constantly calibrated as the size of the database increases to maintain the most powerful and accurate estimates.

In this study, five machine learning algorithms were used, i.e., random forest (RF), categorical boosting (CatBoost), extreme gradient boosting (XGBoost), linear regression (LR) and multilayer perceptron (MLP). A brief description of these machine learning models follows.

Random forest (RF) is an ensemble learning technique that uses a combination of classification tree predictors to give predictions for the response variable [30]. Each tree is constructed from a bootstrap sample of the dataset and uses a randomly selected subset of the variables from the candidate set of input variables that is generated at each split of data. Therefore, for tree building, random forest uses both bagging (bootstrap aggregation) [31], a method for combining unstable learners, and random variables selection. RF algorithm can handle both categorical and continuous variables. RF has a good capability for high-dimensional data processing while avoiding overfitting, and it can be used for both classification and regression tasks.

Categorical boosting (CatBoost) is a gradient boosting decision tree (GBDT), which uses binary decision trees as base predictors [32]. CatBoost uses a symmetrical adopted tree model to reduce over-fitting and enhance the model's generalizability. Moreover, it avoids the problems of gradient bias and prediction shift.

Extreme gradient boosting (XGBoost) [33] is a scalable tree boosting system that implements a generalized gradient boosting method including an additional regularization term. The regularization term aims to smooth the final learned weights to avoid over-fitting, and to yield accurate models. In addition, parallel calculations are automatically executed for the functions during the training step [34]. This algorithm comes with several improvements in multithreaded processing and optimization function to improve calculation speed and reduce over-fitting events as well.

Linear regression (LR) is a machine learning algorithm used to represent the relationship between independent variables and one dependent variables (simple linear regression) or more than independent variables (multiple linear regression). This linear regression establishes a mathematical model to study and describe a given real-world phenomenon [35].

Multilayer perceptron neural network (MLP) is a class of feedforward artificial neural network (ANN) with a backpropagation algorithm for training. MLP consists of one or more hidden layers between the input and output layers. The hidden layers contain multiple neurons mimicking the biological nervous system [36]. The input signal starts with input nodes and propagates to the output node in the forward pass. The output

calculation in each node is performed layer-by-layer by multiplying it by a weight factor called a synaptic weight. After performing the forward pass calculations, an error is calculated in the backward pass using a loss function [37]. The advantages of MLP is that it has a strong nonlinear fitting ability and can map arbitrarily complex nonlinear relationships [38].

2.4. Development of Predictive Modeling Process

Figure 2 presents the workflow of the applied methodology. All data were randomly split into three parts: 70% for calibration or training, 15% for validation, and 15% for testing using GroupShuffleSplit from scikit-learn [39]. The calibration set is used to learn the parameters of the model: it consists in estimating one granulometric fraction (clay, silt, or sand) according to the reflectances of the vis-NIR spectrum. This step is used to train the chosen algorithm. The validation set is used to avoid introducing any artifact in the prediction results when using the training dataset. This validation adjusts the model hyperparameters to avoid overfitting, a very common phenomenon in prediction model training. The test dataset acts as independent data and is used to verify the performance of the prediction results. In order to avoid the effect of a random seed, the random split was performed 30 times with different seeds. Therefore, for illustrative purposes, the seed chosen in this study is the one giving results closer to the average values of R^2 (Equation (1)) and RMSE (Equation (2)) obtained by the 30 prediction models.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (\overline{y} - y_i)^2} \quad (1)$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2 / n} \quad (2)$$

where R^2 is the coefficient of determination, RMSE is the root mean squared error, $\hat{y}_i$ is the predicted value of the ith observation, $\overline{y}$ is the mean measured value, y_i is the measured value of the *i*th observation, and n is the number of samples.

To correct the residual error (bias adjustment) and ensure unbiased predictions, the observed residuals from the validation model were used to linearly correct the bias [40]. This correction step was only used for prediction of the soil textural fractions.

2.5. Machine Learning Models for Predicting Fine (G1), Medium (G2), and Coarse (G3) Textural Groups

Since both soil libraries ICRAF and LUCAS contain the spectral and chemical measurements of soils, we used the seven available chemical measurements (N_{Total}, $P_{available}$, $K_{exchangeable}$, $Ca_{exchangeable}$, $Mg_{exchangeable}$, $CaCO_3$, pH_{water}, OC) for LUCAS and the five available measurements ($K_{exchangeable}$, $Ca_{exchangeable}$, $Mg_{exchangeable}$, pH_{water}, OC) for ICRAF. The inclusion of these chemical measurements in the spectral measurements in the calibration model is intended to enhance the performance of the textural group prediction. In contrast to the predictions of the size fractions (clay, silt, and sand), which are quantitative features and have shown a better affinity with regression algorithms such as MLP and CatBoost, the textural groups (G1, G2, and G3) are categorical variables, which require classification algorithms, such as the light gradient boosting machine (LightGBM). The LightGBM is a model applying gradient-based one-side sampling (GOSS) and exclusive feature bundling technologies [4]. Unlike the conventional GBM tree splitting method, LightGBM uses the GOSS method to identify the observations that can be used for computing the split. LightGBM has the advantages of reducing memory usage while increasing the training speed. In addition, it produces much more complex trees by following a leaf-wise split approach rather than a level-wise approach, which is the main factor in achieving higher accuracy, and it is capable of handling large-scale data. LightGBM contains two novel techniques, namely gradient-based one-side sampling and exclusive feature bundling, to deal with large numbers of data instances and large numbers of features,

respectively [41]. We used this algorithm to train predictive models for classification of the three textural groups of soil. Many research studies have used auxiliary information derived from different data sources to improve the prediction of soil properties [42,43]. Textural groups are based on the Canadian system of soil classification (Figure 3).

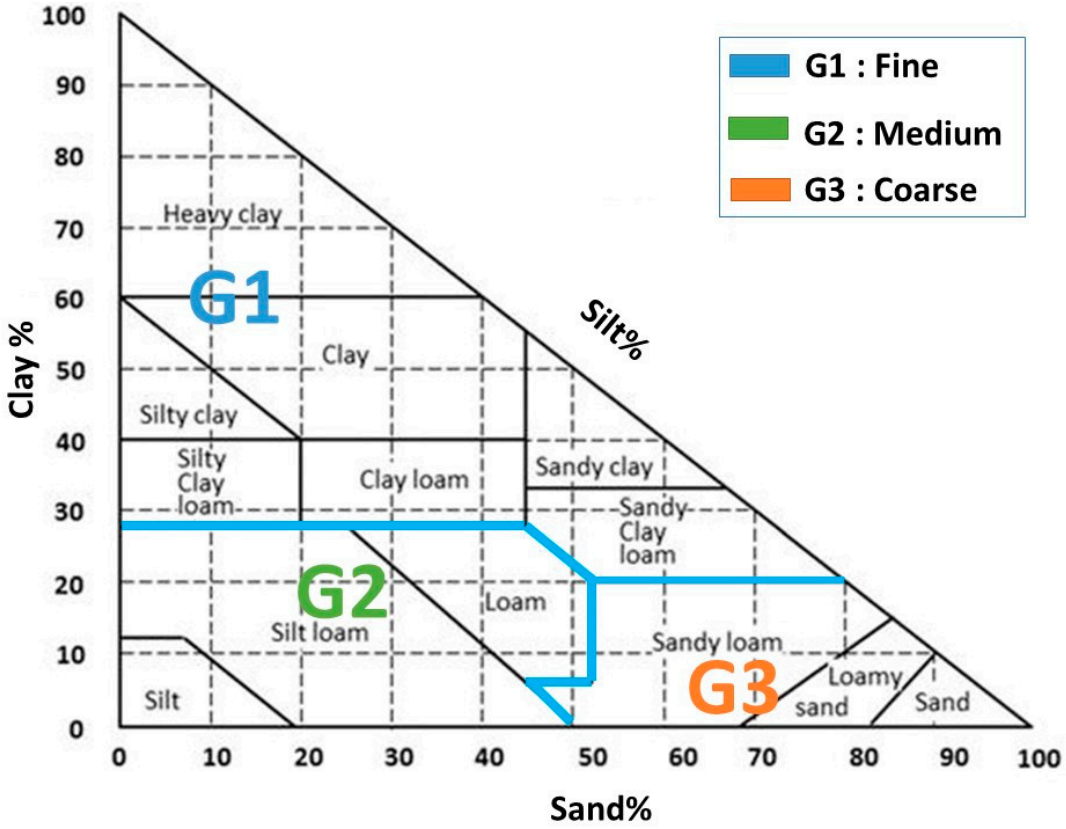

Figure 3. Canadian triangle for soil texture classification for textural classes and groups of mineral soils.

2.6. Data Analysis Platform

All of the models were implemented in python 3.9 using the packages "sklearn" [39], "xgboost" [33], "CatBoost" [32], and "LightGBM" [41]. In addition, "NumPy" [44] was used for math calculations and "pandas" for dataset manipulation [45]. The "plotly" package was used to generate graphs.

2.7. Model Evaluation

To evaluate the performance and accuracy of prediction models for the three soil textural fractions, four parameters were used; the coefficient of determination (R^2), the root mean squared error (RMSE), and the slope and intercept of predicted versus observed values [46]. Li et al. [47] proposed a classification criterion for R^2 values: $R^2 < 0.50$ (unacceptable prediction), $0.50 \leq R^2 < 0.75$ (acceptable prediction) and $R^2 \geq 0.75$ (good prediction). The same criterion was applied in the current study.

For the three textural groups (G1, G2, G3) we used a normalized confusion matrix and visualized it using the "seaborn" package [48] in python 3.9. The performance index in classifying the soil textural groups was evaluated by calculating the accuracy, ranging from 0 to 100% (Equation (3)), which represent the ratio between the number of correctly classified cases over the total number of properly and wrongly classified cases.

$$\text{Accuracy} = \frac{\text{correctly classified}}{\text{Total : correctly and incorrectly classified}} \quad (3)$$

3. Results and Discussion

3.1. Performance of Pre-Processing Techniques

Table 1 shows the five pre-processing techniques applied to the raw vis-NIR spectra. The results show that the first derivative is the best for ICRAF and LUCAS. Therefore, the first derivative spectrum was used for the prediction models of sand, silt, and clay. This choice was based on the four evaluation criteria that reflect the consistency between predicted and measured values when using the testing data. The model performed better when the slope (m) and the coefficient of determination (R^2) were closer to 1.0, and the intercept (a) and the RMSE were closer to 0.

Ben-Dor et al. [49] reported that the first derivative of the spectra improved spectral information and showed spectral changes better. Moreover, our results agree with other research studies. Gerighausen et al. [50] pointed out that the first derivative of the spectra gave the best prediction results for the soil clay and organic carbon content.

In general, there are several baseline removal methods available, such as spectral derivative transformations (first and second derivative), which is one of the best methods for removing baseline effects [27]. Rinnan et al. [26] reviewed the most applied spectral pre-processing techniques for near-infrared spectrometry. They concluded that the first derivative is very effective for removing the baseline offset. Therefore, the first derivative technique was selected in this study as the pre-processing method for the following steps. Compared to the raw spectra that give unacceptable prediction results according to the classification of Li et al. [47] with R^2 mostly around 0.47 for sand and silt and 0.65 for clay, the first derivative transformation of these spectra contributed to increase the robustness of the prediction models by over 20% (Table 1). Even though the two databases are very different in size and origin, the prediction results were comparable for both the raw and first derivative transformed spectra (Table 2). The prediction from the first derivative spectra was significantly higher for clay than for the other two particle size fractions with the highest R^2 and slopes and the lowest intercepts and RMSE. Therefore, it may be concluded that spectrometry is more adapted to clay fraction than to silt and sand fractions.

Table 1. The performance of five pre-processing techniques for prediction of sand, silt, and clay on the ICRAF and LUCAS datasets.

Soil Property	Pre-Processing	ICRAF				LUCAS			
		Testing Set				Testing Set			
		R^2 *	RMSE *	Intercept	Slope	R^2 *	RMSE *	Intercept	Slope
Sand	Spectra	0.47	21.00	20.06	0.469	0.52	17.83	20.28	0.516
	1st Der	0.73	14.77	10.03	0.734	0.66	14.83	14.01	0.669
	2nd Der	0.71	15.38	10.79	0.717	0.67	14.61	13.56	0.676
	CR	0.61	17.89	14.91	0.609	0.54	17.41	19.31	0.534
	DT	0.61	17.87	14.84	0.617	0.58	16.70	18.10	0.574
	SNV	0.62	17.61	14.60	0.626	0.59	16.38	16.95	0.599
Silt	Spectra	0.47	14.49	15.51	0.468	0.47	13.13	20.39	0.475
	1st Der	0.71	10.62	8.37	0.712	0.60	11.44	15.61	0.599
	2nd Der	0.71	10.65	8.42	0.712	0.62	11.13	14.56	0.625
	CR	0.62	12.16	11.11	0.622	0.51	12.77	19.03	0.505
	DT	0.62	12.10	10.79	0.633	0.51	12.69	19.02	0.512
	SNV	0.58	12.78	12.24	0.587	0.51	12.67	18.50	0.518
Clay	Spectra	0.65	13.03	11.42	0.651	0.65	7.610	6.643	0.656
	1st Der	0.81	9.63	6.29	0.807	0.80	5.760	3.990	0.795
	2nd Der	0.77	10.41	7.24	0.773	0.73	6.672	5.316	0.729
	CR	0.71	11.83	9.57	0.711	0.57	8.444	8.320	0.571
	DT	0.70	12.16	10.10	0.691	0.67	7.376	6.259	0.673
	SNV	0.72	11.41	8.92	0.720	0.74	6.566	5.065	0.750

* R^2, coefficient of determination; RMSE, root mean square error.

Table 2. Best prediction accuracy of sand, silt and clay for both datasets ICRAF and LUCAS.

Soil Textural Fraction	SSL *	Model	R^2 *	RMSE *	Slope (m)	Intercept (a)
Sand	ICRAF	MLP	0.78	13.55	0.99	0.29
Silt		CatBoost	0.81	8.45	0.99	0
Clay		MLP	0.85	8.77	0.99	0.62
Sand	LUCAS	CatBoost	0.78	11.77	0.99	0.33
Silt		CatBoost	0.76	8.68	0.98	0.57
Clay		CatBoost	0.85	4.99	1.01	−0.3

* R^2, Coefficient of determination; RSME, root-mean-square error; SSL, soil spectral library.

3.2. Soil Fractions Prediction and Bias Correction

Table 2 shows the results that yielded the best validity (measured vs. predicted) among the five models used with both datasets, ICRAF and LUCAS, compared to the models without bias corrections (Supplementary Tables S1 and S2). This validity assesses the degree of association between the three particle size fractions (clay, silt, and sand) measured by standard techniques (hydrometer or Robinson pipette) and those derived by examining the signal spectrum obtained by spectral scanning of the soil samples by artificial intelligence. In a two-dimensional graph representing the real values on the abscissa (e.g., measured percentage of clay) and the predicted values on the ordinate (predicted percentage of clay), we fitted a linear regression to the testing dataset. This linear regression model shows the four performance criteria (Table 2): robustness (R^2), accuracy (slope, m), sensitivity (intercept, a) and precision (RMSE). Using these four performance criteria, CatBoost was found to be the best in predicting the three particle size components for the LUCAS dataset. Its R^2 values ranged from 0.76 to 0.85, its m-slopes were close to 1 (high accuracy), its a-intercepts were low at 0–0.57% (low sensitivities), and it had low RMSEs of 4.99–11.77. These results show that MLP appears to have performed better than CatBoost for the prediction of sand and clay with the ICRAF testing dataset. MLP showed a R^2 robustness of 0.78–0.85, m accuracies of 0.99; for the intercept, a, it showed a low sensitivity of 0.29–0.62%, and it had a RMSE of 8.77–13.55. It appears that CatBoost is best suited for LUCAS sand. MLP works best with ICRAF databases. According to the R^2 classification of Li et al. [48], it can be stated that all predictions of the three particle size components (clay, silt, and sand) in the two databases LUCAS and ICRAF are considered good, and thus satisfactory to adopt as alternative methods of soil scanning to deduce their particle size composition.

Figure 4 shows all clay prediction results for the ICRAF dataset. These are the four criteria of prediction performance through the linear regression between the predicted and measured values in the training (Figure 4a), validation (Figure 4b), and testing (Figure 4d) datasets. In the same figure, the linear bias on the residuals of the validation dataset (Figure 4c) was used to correct the results in the testing dataset (Figure 4e). Accounting for the linear bias of the prediction residuals (predicted vs. measured residuals) as a function of the percentage of clay according to the equation Y = 0.235X–4.54 (Figure 4c), the predictions were improved significantly. Compared to the validation results (Figure 4b), the linear bias correction helped to increase the R^2 by 0.07, the slope m by 0.24, and sensitivity or intercept a by 4%. In Figure 4e, prediction in the corrected testing can be seen to reach a robustness R^2 of 0.85, an m value close to 1, and sensitivity close to zero. It is therefore clear that the linear bias correction proposed by Song [40] improves the performance criteria for clay prediction from vis-NIR spectra. Curiously, the prediction results were almost the same for the LUCAS database (Figure 5) and the ICRAF database (Figure 4). Similarly, the prediction results for LUCAS in the corrected testing reached the same levels of R^2 robustness of 0.85, m of 0.99, and sensitivity, a, of 0.62% (Figure 5e).

Generally, the bias correction step significantly improved the model accuracy for the slope, as shown in Figures 4 and 5, and furthermore, this step (bias correction) was carried out for the three soil textural fractions in both datasets (ICRAF and LUCAS), as shown in Table 2. For both datasets, the clay content prediction result was the highest between the

soil textural fractions for LUCAS and ICRAF (R^2 = 0.85 for both datasets), followed by the silt (R^2 = 0.81 and 0.76 for ICRAF and LUCAS, respectively), and sand (R^2 = 0.78 for both datasets) (Supplementary Figures S1–S4).

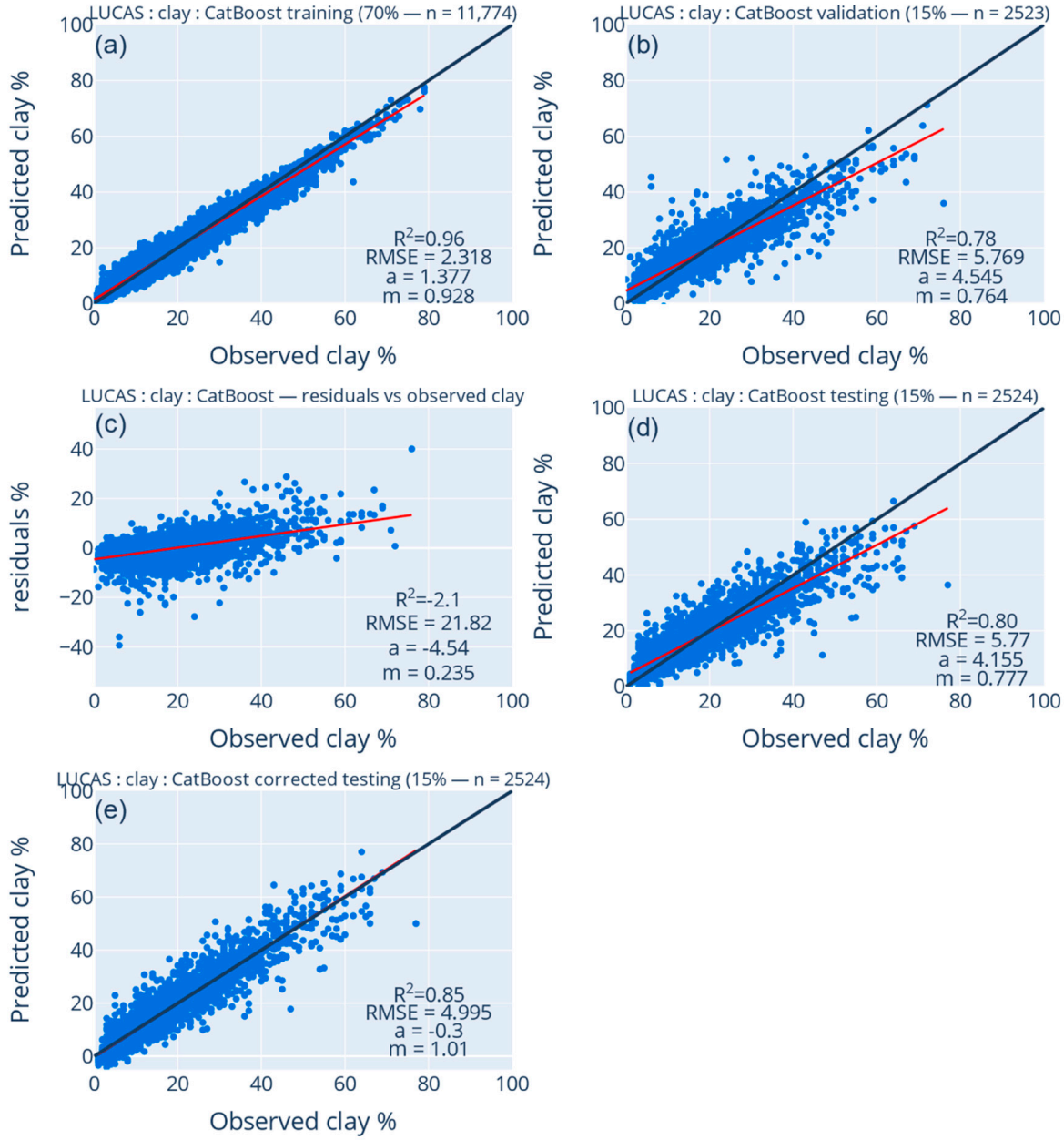

Figure 4. Prediction accuracy of clay content from ICRAF dataset: (**a**) training; (**b**) validation; (**c**) the residual bias correction to be used in the testing model; (**d**) testing without bias correction; (**e**) testing with bias correction. The black line refers to the 1:1 line, and the red line refers to the regression line.

Cozzolino and Moron [51] achieved a good R^2 (0.67, 0.80, and 0.90 respectively) for sand, silt, and clay content from vis-NIR spectra using principal components regression (PCR) models (n = 332 soil samples). Chang et al. [52] obtained an R^2 of 0.67 for clay content prediction, beside other soil properties based on PCR using the first derivatives pre-processing (n = 743 soil samples). Ahmadi et al. [53] concluded the arithmetic mean of R^2 for sand, silt, and clay content prediction (0.76, 0.68, and 0.70, respectively) using vis-NIR data from several studies. These studies used different prediction algorithms, soils, and methodologies that can explain the variation of the prediction results obtained from several studies.

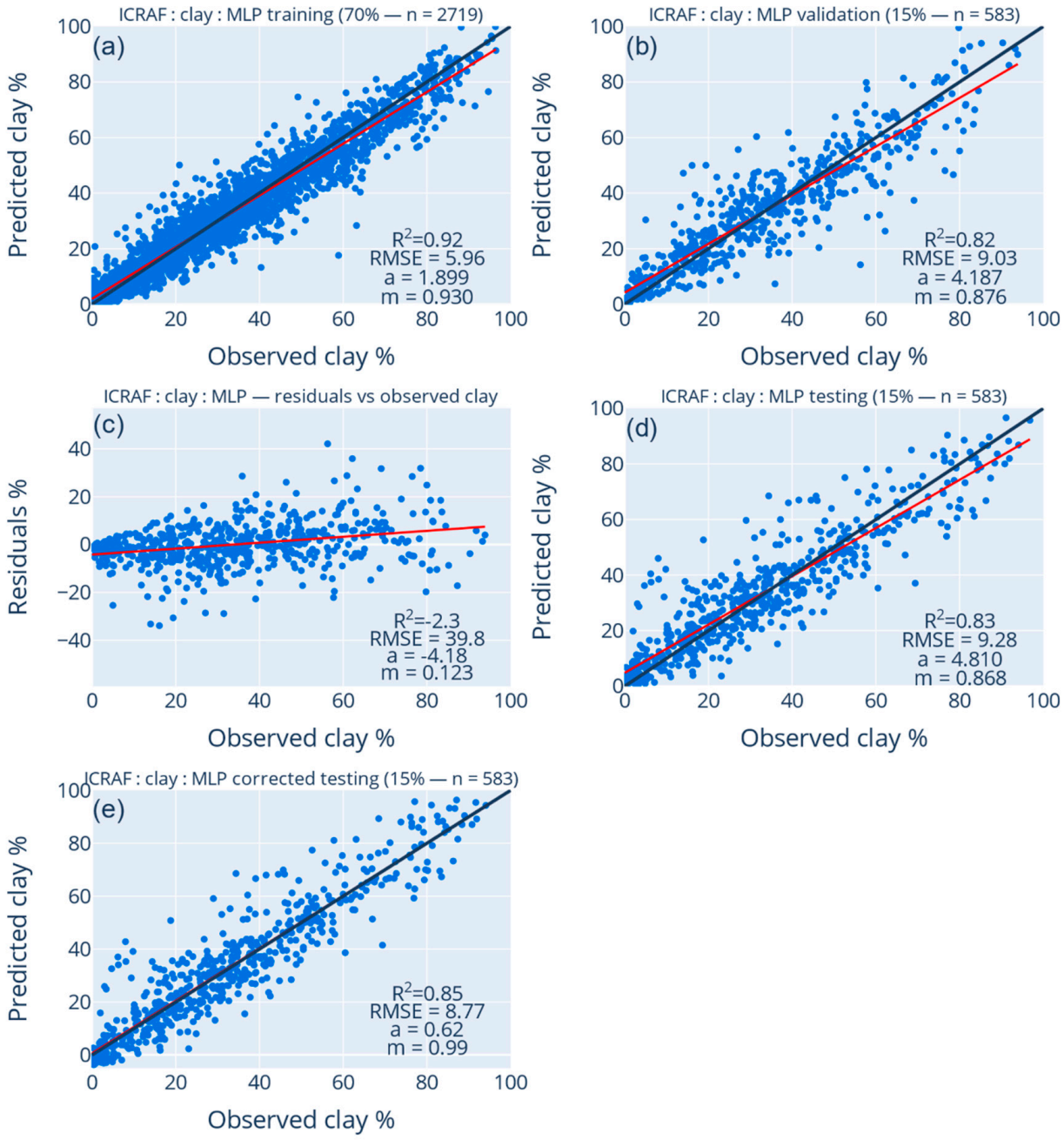

Figure 5. Prediction accuracy of clay content from LUCAS and ESDAC datasets: (**a**) training; (**b**) validation; (**c**) the residual bias correction bias to be used in the testing model; (**d**) testing without bias correction; (**e**) testing with bias correction. The black line refers to the 1:1 line, and the red line refers to the regression line.

The conventional methods used in the laboratory (e.g., pipette and hydrometer methods) for determining soil texture are laborious, costly, and not suitable for large numbers of soil samples. Moreover, the texture hand test method is very subjective and often gives unreliable results. Therefore, spectral soil analysis can be a valuable alternative to the traditional wet chemistry soil analyses once robust predictive models are built. This spectral method is attractive because: (a) the intercepts (sensitivities) do not exceed the value of 0.62% (Table 2); (b) the slope m (accuracy) is 0.99–1; (c) the R^2 (robustness) of the prediction models is 0.76–0.85 on the testing dataset (Table 2); (d) the range of the two datasets cover a wide variety of soils from several countries; (e) the reproducibility of the spectral calibration performance criteria was satisfactory for both independent databases of ICRAF and LUCAS; and (f) the suitability to precision agriculture, which generates a

large number of soil tests. With these high prediction performances, we can expect the spectral method to be a candidate replacement for conventional methods of soil texture diagnosis. If a sufficient number of samples is available on a regional scale, the prediction performance could be further refined, since the training is done only on soils in the region where the analysis results will be used. A laser diffraction granulometry method is proposed as another alternative [54] to the standard techniques (hydrometer, Robinson pipette); however, it remains to be seen if it is competitive in terms of processing time and quality of the results [55].

3.3. Prediction of Textural Groups

Textural soil group information is very valuable for planning agricultural management practices [10]. Figure 6 shows the predicted textural groups for the LUCAS testing dataset using LightGBM. The overall accuracy (Equation (3)) of predicting the textural groups for the LUCAS testing dataset was 75% when considering only the spectra, 58% when considering only the seven available chemical measurements (N_{Total}, $P_{available}$, $K_{exchangeable}$, $Ca_{exchangeable}$, $Mg_{exchangeable}$, $CaCO_3$, pH_{water}, OC), and 84% when combining the spectral and chemical predictors. The overall accuracy of the textural group prediction for the LUCAS training dataset was 100% regardless of whether one considers only spectra as predictors, or chemical measurements or combined measurements (chemical and spectral). This 100% accuracy shows that the training used to estimate the G1, G2, and G3 textural groups resulted in a perfect prediction with a 0% prediction error. However, this same "perfect" training model resulted in a 16–38% decrease in accuracy when applied to independent testing data. This shows that the LightGBM model tends to overtrain, causing a drop in prediction performance. Despite this decrease, the probability of correctly classifying the textural group from spectral analysis and routine chemical testing of the soil is still quite high, at almost 84%. Moreover, a normalized confusion matrix (observation ratios in percentage terms) was used to visualize the performance of the LightGBM model by textural groups G1, G2, and G3, as defined in Figure 3. Their accuracy was 100% for training and 80.8, 80.6, and 89.4% for G1, G2, and G3 respectively. The overall accuracy of predicting the textural groups for the ICRAF testing dataset was 75% when considering only the spectra, 62% when considering only the five available chemical measurements ($K_{exchangeable}$, $Ca_{exchangeable}$, $Mg_{exchangeable}$, pH_{water}, OC), and 83% when combining the spectral and chemical predictors. Figure 7 present the results of testing data from ICRAF dataset with a total or overall prediction accuracy of 100% for training and 83% for testing. Distributed over the three textural groups, this prediction accuracy for testing was 88.12%, 56.52%, and 80.26%, respectively for groups G1, G2, and G3 and 100% for training.

The accuracy of both models shows a good ability to classify soil textural groups. As expected, most of the misclassified data points were located on the borders between neighboring textural groups. Even the conventional methods (e.g., pipette and hydrometer methods) can generate such measurement errors when it comes to neighboring of textural groups [56]. Beretta et al. [57] confirmed that even conventional wet methods can show differences between them in textural class assignments. Therefore, this misclassification between different neighboring soil classes determined by the conventional methods may explain the misclassified data points generated by the LightGBM model. For a tolerance of about 10% on the sand and clay fractions, we can visualize a border effect of 20 $\pm$ 2% when considering the horizontal boundary of 20% clay between G1 and G3, of 30 $\pm$ 3% when considering the horizontal boundary of 30% clay between G1 and G2, and of 45 $\pm$ 4.5% to 55 $\pm$ 5.5% when considering the vertical boundaries of 45–55% sand between G2 and G3 (Figures 6 and 7). The remaining points that are misclassified, and are far from these border effects, are attributed to true prediction errors by LightGBM. At first glance, misclassified points due to border effects are more frequent than those due to prediction errors (Figures 6 and 7).

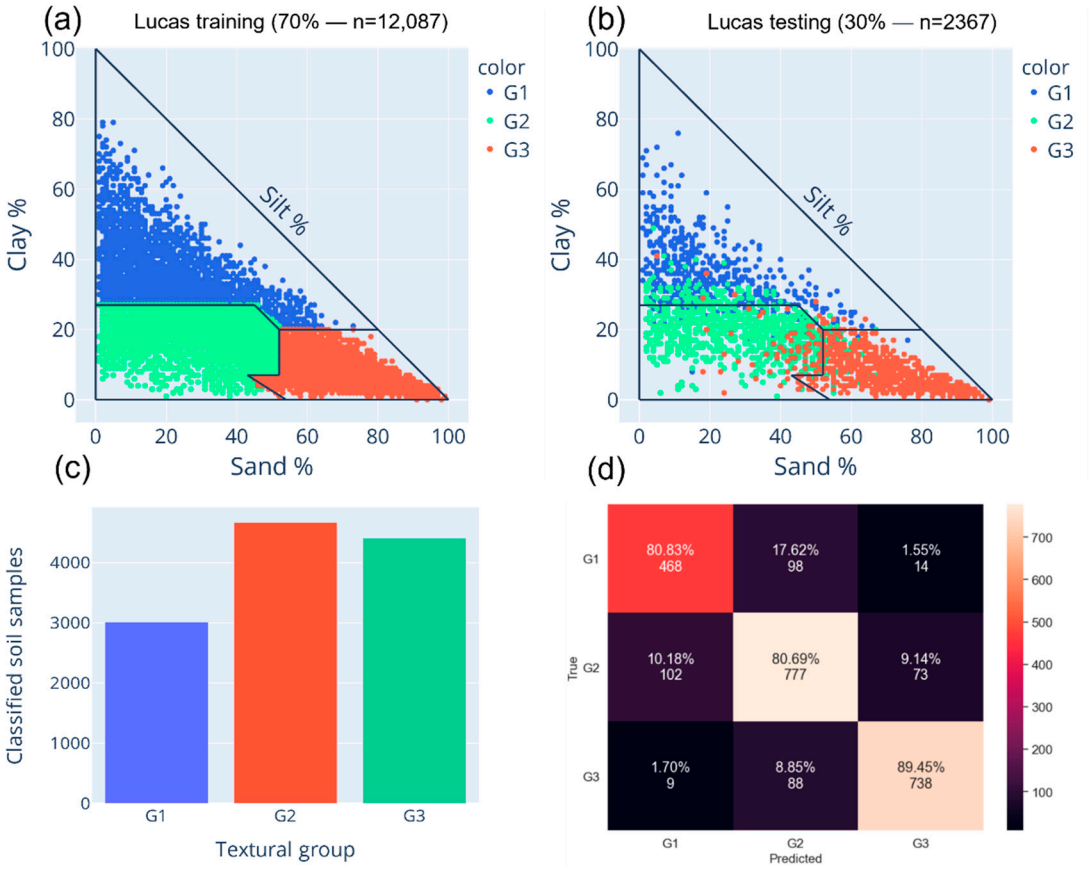

Figure 6. Prediction texture triangles (following the Canadian system of soil classification) obtained from LightGBM algorithm for the LUCAS dataset: (**a**) calibration data (*n* = 12,087); (**b**) testing data (*n* = 2367); (**c**) normalized confusion matrix of the training dataset; (**d**) normalized confusion matrix of testing dataset obtained for the 3 textural groups.

As shown in Figures 6 and 7, the total prediction accuracy for soil textural groups was similar for ICRAF and LUCAS, respectively at 83% and 84% for testing and 100% for training. These high classification accuracies were achieved when spectral information and auxiliary chemical variables are combined in the prediction model. Previous studies showed that the inclusion of auxiliary predictor variables improves the prediction accuracy of a model [8,18,58–60]. Moreover, Cozzolino and Moron [51] obtained a high positive correlation between Ca and clay content (0.80), Cu and Mg (0.71), K and clay (0.60), and Mg and clay (0.51), which may consequently affect the soil texture prediction. They concluded that these high correlations between physical properties and chemical parameters could explain some of the high accuracy obtained. Our results suggest that the LightGBM model can be successfully used to predict the soil textural groups.

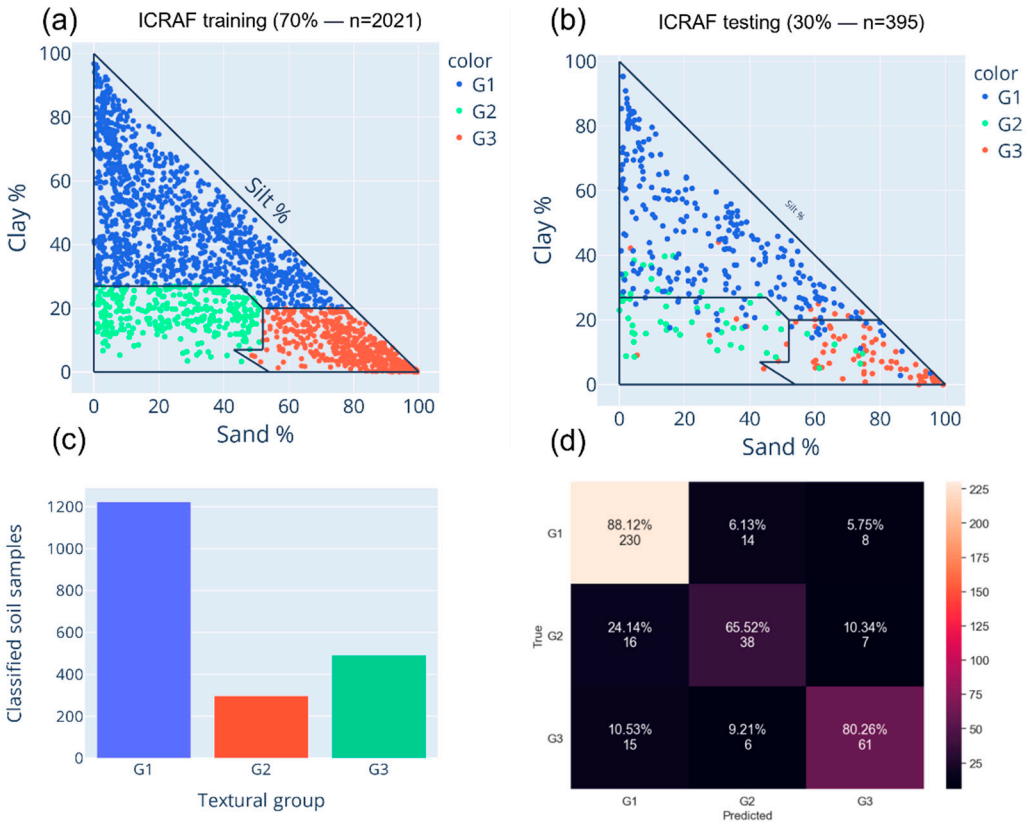

Figure 7. Prediction texture triangles (following Canadian system of soil classification) for obtained from the LightGBM algorithm for the ICRAF dataset: (**a**) training data (n = 2021); (**b**) testing data (n = 395); (**c**) normalized confusion matrix of the training dataset; (**d**) normalized confusion matrix of testing dataset obtained for the 3 textural groups.

4. Conclusions

In this paper, the machine learning algorithms MLP, CatBoost, and LightGBM were used to predict soil textural fractions (clay, silt, and sand) and groups (G1: Fine, G2: Medium, and G3: Coarse) for two large-scale vis-NIR soil spectral libraries (ICRAF and LUCAS). The models' performance on the testing datasets reached R^2 values up to 0.85. The results showed a good performance for the LUCAS testing dataset with R^2 values of 0.78 (sand), 0.81 (silt). and 0.85 (clay). For the ICRAF testing dataset, the R^2 values for sand, silt, and clay were 0.78, 0.76, and 0.85, respectively. This predictive capacity of soil texture properties using soil spectral information is a very promising alternative to the traditional soil laboratory analysis due to its sensitivity, accuracy, reliability, versatility, reproducibility, and adaptability to precision agriculture.

Furthermore, the soil textural groups (G1, G2, and G3) were classified with LightGBM using spectra and chemical auxiliary variables with a high overall accuracy of 100% for training and close to 84% for testing. These findings support the hypothesis that chemical tests are powerful auxiliary variables for improving the prediction of these three textural groups. The algorithms CatBoost, MLP, and LightGBM are promising for soil texture prediction, and they can be used when a soil spectral library is available with enough samples. Moreover, in light of these good prediction accuracies produced for soil texture prediction with the ICRAF and LUCAS spectral libraries and the CatBoost, MLP, and LightGBM algorithms, we can consider further refining this method by using regional soil sample databases with spectral data obtained using in situ spectral methods.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/agronomy11081550/s1, Table S1: Five machine and deep learning algorithms used for sand,

silt, and clay prediction on LUCAS dataset before bias correction, Table S2: Five machine and deep learning algorithms used for sand, silt, and clay prediction on ICRAF dataset before bias correction, Figure S1: Prediction accuracy of Sand content from ICRAF dataset, Figure S2: Prediction accuracy of Silt content from ICRAF dataset, Figure S3: Prediction accuracy of Sand content from LUCAS dataset, Figure S4: Prediction accuracy of Silt content from ESDAC dataset.

Author Contributions: Writing, original draft, M.Z.G. and L.K.; Conceptualization, L.K., J.G., E.M.N. and M.Z.G.; Methodology, M.Z.G., L.K. and M.I.; Modelling and software, J.G. and E.M.N. Data Deployment, E.M.N. and M.I.; Visualization, L.K., E.M.N. and M.I.; Project administration, L.K. and M.I., Editing and revision, L.K., J.G., M.Z.G. and E.M.N. and supervision, L.K. and J.G. All authors have read and agreed to the published version of the manuscript.

Funding: Funding was provided by OCP Fondation, Mohammed VI Polytechnic University (UM6P), MItacs Globalink Canada and the Natural Sciences and Engineering Research Council of Canada (NSERC) (No. RDCPJ528053-18).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used for this study can be found at the following locations: **ICRAF:** http://africasoils.net/services/data/soil-databases/soil-spectral-libraries/. **LUCAS:** https://esdac.jrc.ec.europa.eu/content/lucas2015-topsoil-data.

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

References

1. Kopittke, P.M.; Menzies, N.W.; Wang, P.; McKenna, B.A.; Lombi, E. Soil and the intensification of agriculture for global food security. *Environ. Int.* **2019**, *132*, 105078. [CrossRef] [PubMed]
2. Saiz-Rubio, V.; Rovira-Más, F. From smart farming towards agriculture 5.0: A review on crop data management. *Agronomy* **2020**, *10*, 207. [CrossRef]
3. Delgado, A.; Gómez, J.A. The Soil. Physical, Chemical and Biological Properties. In *Principles of Agronomy for Sustainable Agriculture*; Springer: Cham, Germany, 2016; pp. 15–26.
4. Gee, G.W.; Or, D. 2.4 Particle-Size Analysis. In *Methods of Soil Analysis: Part 4 Physical Methods, 5.4*; Soil Science Society of America, Inc.: Madison, WI, USA, 2002; pp. 255–293.
5. Tümsavaş, Z.; Tekin, Y.; Ulusoy, Y.; Mouazen, A.M. Prediction and mapping of soil clay and sand contents using visible and near-infrared spectroscopy. *Biosyst. Eng.* **2019**, *177*, 90–100. [CrossRef]
6. Shahandeh, H.; Wright, A.L.; Hons, F.M. Use of soil nitrogen parameters and texture for spatially-variable nitrogen fertilization. *Precis. Agric.* **2011**, *12*, 146–163. [CrossRef]
7. Bönecke, E.; Meyer, S.; Vogel, S.; Schröter, I.; Gebbers, R.; Kling, C.; Kramer, E.; Lück, K.; Nagel, A.; Philipp, G.; et al. Guidelines for precise lime management based on high-resolution soil pH, texture and SOM maps generated from proximal soil sensing data. *Precis. Agric.* **2020**, *22*, 493–523. [CrossRef]
8. Vories, E.; O'Shaughnessy, S.; Sudduth, K.; Evett, S.; Andrade, M.; Drummond, S. Comparison of precision and conventional irrigation management of cotton and impact of soil texture. *Precis. Agric.* **2020**, *22*, 414–431. [CrossRef]
9. Khiari, L. *Échantillonnage Conventionnel des Sols Agricoles au Québec*; Centre de Référence en Agriculture et Agroalimentaire du Québec: Québec, QC, Canada, 2014; p. 20.
10. Chelabi, H.; Khiari, L.; Gallichand, J.; Joseph, C.A. Soil sample preparation techniques on routine analyses in Quebec affect lime and fertilizer recommendations. *Can. J. Soil Sci.* **2016**, *96*, 244–255. [CrossRef]
11. Metwally, M.S.; Shaddad, S.M.; Liu, M.; Yao, R.J.; Abdo, A.I.; Li, P.; Jiao, J.; Chen, X. Soil properties spatial variability and delineation of site-specific management zones based on soil fertility using fuzzy clustering in a hilly field in Jianyang, Sichuan, China. *Sustainability* **2019**, *11*, 7084. [CrossRef]
12. Blake, G.R. Physical and mineralogical properties, including statistics of measurement and sampling. In *Methods of Soil Analysis: Part 1 Physical and Mineralogical Properties, Including Statistics of Measurement and Sampling, 9.1*; Soil Science Society of America, Inc.: Madison, WI, USA, 1965.
13. Viscarra Rossel, R.A.; Lark, R.M. Improved analysis and modelling of soil diffuse reflectance spectra using wavelets. *Eur. J. Soil Sci.* **2009**, *60*, 453–464. [CrossRef]
14. Allo, M.; Todoroff, P.; Jameux, M.; Stern, M.; Paulin, L.; Albrecht, A. Prediction of tropical volcanic soil organic carbon stocks by visible-near- and mid-infrared spectroscopy. *Catena* **2020**, *189*, 104452. [CrossRef]
15. Davari, M.; Karimi, S.A.; Bahrami, H.A.; Hossaini, S.M.T.; Fahmideh, S. Simultaneous prediction of several soil properties related to engineering uses based on laboratory Vis-NIR reflectance spectroscopy. *Catena* **2021**, *197*, 104987. [CrossRef]
16. Coblinski, J.A.; Giasson, É.; Demattê, J.A.M.; Dotto, A.C.; Costa, J.J.F.; Vašát, R. Prediction of soil texture classes through different wavelength regions of reflectance spectroscopy at various soil depths. *Catena* **2020**, *189*, 104485. [CrossRef]

17. Katuwal, S.; Knadel, M.; Norgaard, T.; Moldrup, P.; Greve, M.H.; de Jonge, L.W. Predicting the dry bulk density of soils across Denmark: Comparison of single-parameter, multi-parameter, and vis–NIR based models. *Geoderma* **2020**, *361*, 114080. [CrossRef]
18. Stevens, A.; Nocita, M.; Tóth, G.; Montanarella, L.; van Wesemael, B. Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. *PLoS ONE* **2013**, *8*, e66409. [CrossRef]
19. 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]
20. Garrity, D.; Bindraban, P. *A Globally Distributed Soil Spectral Library Visible Near Infrared Diffuse Reflectance Spectra*; ICRAF (International Center for Research in Agroforestry): Nairobi, Kenya; ISRIC (World Soil Information): Wageningen, The Netherlands; Spectral Library: Nairobi, Kenya, 2004.
21. Reeuwijk, L. *Procedures for Soil Analysis*; Tech. Pap.; ISRIC: Wageningen, The Netherlands, 2002.
22. FOSS. *NIR Spectroscopy: A Guide to Near-Infrared Spectroscopic Analysis of Industrial Manufacturing Processes*; Metrohm: Herisau, Switzerland, 2009.
23. Stenvens, A.; Ramirez-López, L. Miscellaneous Functions for Processing and Sample Selection of vis-NIR Diffuse Reflectance Data. 2014. CRAN 'Prospectr' R; An Introduction to the Prospectr Package. R Package Vignette R Version 0.1. Available online: https://cran.r-project.org/web/packages/prospectr/vignettes/prospectr.html (accessed on 28 July 2021).
24. LeDell, E.; Gill, N.; Aiello, S.; Fu, A.; Candel, A.; Click, C.; Kraljevic, T.; Nykodym, T.; Aboyoun, P.; Kurka, M.; et al. R Interface for "H2O". 2019. Available online: https://cran.r-project.org/web/packages/h2o/index.html (accessed on 28 July 2021).
25. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation Statistical Computing: Vienna, Austria, 2019; p. 201.
26. Rinnan, Å.; van den Berg, F.; Engelsen, S.B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends Anal. Chem.* **2009**, *28*, 1201–1222. [CrossRef]
27. Duckworth, J. Mathematical Data Preprocessing. In *Near-Infrared Spectroscopy in Agriculture*; Soil Science Society of America, Inc.: Madison, WI, USA, 2015; Volume 44, pp. 113–132.
28. Clark, R.N.; Roush, T.L. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *J. Geophys. Res.* **1984**, *89*, 6329–6340. [CrossRef]
29. Dhanoa, M.S.; Lister, S.J.; Sanderson, R.; Barnes, R.J. The Link between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) Transformations of NIR Spectra. *J. Near Infrared Spectrosc.* **1994**, *2*, 43–47. [CrossRef]
30. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
31. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
32. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. Catboost: Unbiased boosting with categorical features. In Proceedings of the Advances in Neural Information Processing Systems. *arXiv* **2017**, arXiv:1706.09516.
33. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. *San Fr. Calif.* **2016**, *1*, 1–4.
34. Fan, J.; Wang, X.; Wu, L.; Zhou, H.; Zhang, F.; Yu, X.; Lu, X.; Xiang, Y. Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy Convers. Manag.* **2018**, *164*, 102–111. [CrossRef]
35. Barbur, V.A.; Montgomery, D.C.; Peck, E.A. Introduction to Linear Regression Analysis. *J. R. Stat. Soc.* **1994**, *43*. [CrossRef]
36. Murtagh, F. Multilayer perceptrons for classification and regression. *Neurocomputing* **1991**, *2*, 183–197. [CrossRef]
37. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
38. Liu, B.; Zhao, Q.; Jin, Y.; Shen, J.; Li, C. Application of combined model of stepwise regression analysis and artificial neural network in data calibration of miniature air quality detector. *Sci. Rep.* **2021**, *11*, 1–12. [CrossRef]
39. 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.
40. Song, J. Bias corrections for Random Forest in regression using residual rotation. *J. Korean Stat. Soc.* **2015**, *44*, 321–326. [CrossRef]
41. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. LightGBM: A highly efficient gradient boosting decision tree. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4 December 2017; Volume 30, pp. 3146–3154.
42. Zhao, Z.; Chow, T.L.; Rees, H.W.; Yang, Q.; Xing, Z.; Meng, F.R. Predict soil texture distributions using an artificial neural network model. *Comput. Electron. Agric.* **2009**, *65*, 36–48. [CrossRef]
43. Zhang, S.-W.; Shen, C.-Y.; Chen, X.-Y.; Ye, H.-C.; Huang, Y.-F.; Lai, S. Spatial Interpolation of Soil Texture Using Compositional Kriging and Regression Kriging with Consideration of the Characteristics of Compositional Data and Environment Variables. *J. Integr. Agric.* **2013**, *12*, 1673–1683. [CrossRef]
44. McKinney, W. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*; O'Reilly Media Inc.: Sebastopol, CA, USA, 2018; Volume 71.
45. McKinney, W. pandas: A Foundational Python Library for Data Analysis and Statistics. *Python High Perform. Sci. Comput.* **2011**, *14*, 1–9.
46. Stevens, R.J.; Poppe, K.K. Validation of clinical prediction models: What does the "calibration slope" really measure? *J. Clin. Epidemiol.* **2020**, *118*, 93–99. [CrossRef] [PubMed]

47. Li, L.; Lu, J.; Wang, S.; Ma, Y.; Wei, Q.; Li, X.; Cong, R.; Ren, T. Methods for estimating leaf nitrogen concentration of winter oilseed rape (*Brassica napus* L.) using in situ leaf spectroscopy. *Ind. Crops Prod.* **2016**, *91*, 194–204. [CrossRef]
48. Waskom, M. Seaborn: Statistical data visualization. *J. Open Source Softw.* **2021**, *6*, 3021. [CrossRef]
49. Ben-Dor, E.; Inbar, Y.; Chen, Y. The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400-2500 nm) during a controlled decomposition process. *Remote Sens. Environ.* **1997**, *61*, 1–15. [CrossRef]
50. Gerighausen, H.; Menz, G.; Kaufmann, H. Spatially explicit estimation of clay and organic carbon content in agricultural soils using multi-annual imaging spectroscopy data. *Appl. Environ. Soil Sci.* **2012**, *2012*, 868090. [CrossRef]
51. Cozzolino, D.; Morón, A. The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. *J. Agric. Sci.* **2003**, *140*, 65–71. [CrossRef]
52. Chang, C.-W.; Laird, D.A.; Mausbach, M.J.; Hurburgh, C.R. Near-Infrared Reflectance Spectroscopy-Principal Components Regression Analyses of Soil Properties. *Soil Sci. Soc. Am. J.* **2001**, *65*, 480. [CrossRef]
53. Ahmadi, A.; Emami, M.; Daccache, A.; He, L. Soil Properties Prediction for Precision Agriculture Using Visible and Near-Infrared Spectroscopy: A Systematic Review and Meta-Analysis. *Agronomy* **2021**, *11*, 433. [CrossRef]
54. Yang, X.; Zhang, Q.; Li, X.; Jia, X.; Wei, X.; Shao, M. Determination of Soil Texture by Laser Diffraction Method. *Soil Sci. Soc. Am. J.* **2015**, *79*, 1556–1566. [CrossRef]
55. UNIL Préparation Pour la Granulométrie Laser. l'Institut des Dynamiques de la Surface Terrestre. 2020. Available online: https://www.unil.ch/idyst/files/live/sites/idyst/files/shared/Analytical%20platform/PDF_protocole/french/sols%20et%20s%C3%A9diments/Preparation%20pour%20la%20Granulometrie%20Laser%202.0.pdf (accessed on 28 July 2021).
56. Dharumarajan, S.; Hegde, R. Digital mapping of soil texture classes using Random Forest classification algorithm. *Soil Use Manag.* **2020**. [CrossRef]
57. Beretta, A.N.; Silbermann, A.V.; Paladino, L.; Torres, D.; Bassahun, D.; Musselli, R.; García-Lamohte, A. Soil texture analyses using a hydrometer: Modification of the Bouyoucos method. *Cienc. Investig. Agrar.* **2014**, *41*, 263–271. [CrossRef]
58. Pirie, A.; Singh, B.; Islam, K. Ultra-violet, visible, near-infrared, and mid-infrared diffuse reflectance spectroscopic techniques to predict several soil properties. *Aust. J. Soil Res.* **2005**, *43*, 713–721. [CrossRef]
59. Ramirez-Lopez, L.; Wadoux, A.M.J.C.; Franceschini, M.H.D.; Terra, F.S.; Marques, K.P.P.; Sayão, V.M.; Demattê, J.A.M. Robust soil mapping at the farm scale with vis–NIR spectroscopy. *Eur. J. Soil Sci.* **2019**, *70*, 378–393. [CrossRef]
60. Pinheiro, É.F.M.; Ceddia, M.B.; Clingensmith, C.M.; Grunwald, S.; Vasques, G.M. Prediction of soil physical and chemical properties by visible and near-infrared diffuse reflectance spectroscopy in the Central Amazon. *Remote Sens.* **2017**, *9*, 293. [CrossRef]