

## Article

# Assessment of Soil Fertility Using Induced Fluorescence and Machine Learning

Louis Longchamps <sup>1,\*</sup> , Dipankar Mandal <sup>2</sup>  and Raj Khosla <sup>2,3</sup><sup>1</sup> School of Integrative Plant Sciences, Cornell University, Ithaca, NY 14853, USA<sup>2</sup> Department of Agronomy, Kansas State University, Manhattan, KS 66506, USA; dmandal@ksu.edu (D.M.); rkhosla@ksu.edu (R.K.)<sup>3</sup> Department of Soil & Crop Sciences, Colorado State University, Fort Collins, CO 80523, USA

\* Correspondence: ll928@cornell.edu

**Abstract:** Techniques such as proximal soil sampling are investigated to increase the sampling density and hence the resolution at which nutrient prescription maps are developed. With the advent of a commercial mobile fluorescence sensor, this study assessed the potential of fluorescence to estimate soil chemical properties and fertilizer recommendations. This experiment was conducted over two years at nine sites on 168 soil samples and used random forest regression to estimate soil properties, fertility classes, and recommended N rates for maize production based on induced fluorescence of air-dried soil samples. Results showed that important soil properties such as soil organic matter, pH, and CEC can be estimated with a correlation of 0.74, 0.75, and 0.75, respectively. When attempting to predict fertility classes, this approach yielded an overall accuracy of 0.54, 0.78, and 0.69 for NO<sub>3</sub>-N, SOM, and Zn, respectively. The N rate recommendation for maize can be directly estimated by fluorescence readings of the soil with an overall accuracy of 0.78. These results suggest that induced fluorescence is a viable approach for assessing soil fertility. More research is required to transpose these laboratory-acquired soil analysis results to in situ readings successfully.

**Keywords:** soil fertility; induced fluorescence; precision agriculture; proximal soil sensing



**Citation:** Longchamps, L.; Mandal, D.; Khosla, R. Assessment of Soil Fertility Using Induced Fluorescence and Machine Learning. *Sensors* **2022**, *22*, 4644. <https://doi.org/10.3390/s22124644>

Academic Editor: Ernest W. Tollner

Received: 10 May 2022

Accepted: 14 June 2022

Published: 20 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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

To achieve a higher level of precision in nutrient management, farmers require information about soil fertility at every location of their fields. However, the acquisition and processing of soil samples is time consuming, labor intensive, and the associated cost remains unaffordable to farmers [1]. Moreover, soil samples acquired at a sampling density of one sample per ha (i.e., common commercial practice for field mapping) often fail to characterize the spatial variability of soil properties that happens at a much smaller scale [2]. Proximal soil sensing (PSS) can be used to estimate soil properties (<2 m below soil surface) rapidly at a high sampling density [3]. It can be in situ or mobile. For decades, agronomists and soil scientists have used color as a soil classification parameter [4]. Findings reported in more recent research studies suggest that soil's optical properties are influenced by its physical and chemical properties [5]. Optical techniques were investigated using visible, near-infrared, and mid-infrared reflectance spectroscopy and were reported to be very good ( $R^2 > 0.81$ ) predictors of soil carbon and organic matter and good ( $R^2$  0.61–0.81) predictors of soil texture, calcium (Ca), cation exchange capacity (CEC), and magnesium (Mg), among others [6]. Near-infrared reflectance can be used to detect soil organic matter content ( $r > 0.9$ ) of dried soil samples [7]. This concept is widely used in the commercial optical soil sensors (OpticMapper; Veris Technologies Inc., Salina, KS, USA) that enable on-the-go detection of organic matter using visible and near-infrared wavebands. More recent sensing platforms can be used attached to another implement to characterize the soil while implementing another task. For instance, the Smart Firmer (Precision Planting, Tremont, TN, USA) is attached to the seeder's firmer and uses optical sensors to estimate

soil organic matter and moisture on-the-go while seeding the crop, and the Veris iScan (Veris Technologies Inc., Salina, KS, USA) is attached to the tillage tool, fertilizer bar, or planter to map soil organic matter, electrical conductivity, moisture, and temperature on-the-go. A new generation of soil sensors also uses soil optical properties to try to detect soil properties such as nitrogen (N), phosphorus (P), potassium (K), organic matter (OM), pH, moisture, CEC, and minor nutrients in situ and in real time (ChrysaLabs Inc., Montreal, QC, Canada; AgroCares Scanner, Wageningen, The Netherlands).

One aspect of optical soil sensing that has received low attention for the characterization of soil properties is ultra-violet (UV)-visible (Vis)-induced fluorescence [8]. Active fluorescence measurement has long been implemented in laboratory conditions, and mobile platforms were limited by both the power of the excitation energy source and the weakness of the fluorescence signal itself [9]. With recent technology developments, notably the advent of powerful UV light emitting diodes and increasingly sensitive optical sensors, portable fluorescence sensors can now be brought to the field. A review of PSS techniques did not mention the use of fluorescence to predict soil physical and chemical properties, most likely due to the scarcity of literature on this topic [3]. In a recent series of research experiments on the topic of soil characterization using portable fluorimeters, it is reported that several chemical and physical soil properties such as texture, pH, and CEC can be detected using X-ray fluorescence [8,10–12]. More recently, a study assessed the potential of UV-induced fluorescence in the visible spectrum to assess soil properties and found that the predictive ability of UV-induced fluorescence was very significant for topsoil [13]. They suggested that this approach could be useful for farmers wanting to update information on soil fertility for fertilization purposes.

Induced fluorescence emitted by a material is influenced by both the material's waveband-selective light absorption properties and its capacity and specificity to emit fluorescence [14]. Fluorophores are chemical compounds that have the property to emit fluorescence upon light absorption. Soil fluorophores can either be organic (present in humus as coumarins, quinones, 2-aminobenzoic acid, salicylic acid, among others) or inorganic [15]. Soil is a mixture of several organic and inorganic compounds that can potentially emit fluorescence. Martins et al. [16] reported the fluorescence spectrum of a whole soil sample as showing a broad (from 480 to 630 nm) spectrum with a peak around 520 nm. Consistently, [17] observed the laser induced fluorescence spectrum of a whole soil sample to be broad in the visible wavebands, and they also observed a much lower fluorescence intensity once the organic fraction of the sample was removed. Laser-induced fluorescence can be used to determine the humification of soil organic matter [16]. Portable X-ray fluorescence was used in situ to measure soil Ca, potassium (K), manganese (Mn), iron (Fe), copper (Cu), and other micro-elements to predict sand, silt, and clay content with success [8,18].

An optical sensor based on fluorescence allows for in situ measurement of UV-Vis-induced fluorescence, the Multiplex3 (Force-A, Orsay, France). This sensor was developed to measure anthocyanins and flavonoids in grapes and provide precise evaluation of fruit maturation [19]. This sensor was also shown to provide early estimates of N status in maize (*Zea mays* L.), rice (*Oryza sativa* L.), and potatoes (*Solanum tuberosum* L.), among other crops [20–23]. Studies also demonstrated the potential to detect biotic stresses such as powdery mildew in sugar beets (*Beta vulgaris* L.) and various fungal infections in wheat (*Triticum aestivum* L.) using this sensor [24,25]. This indicates that handheld UV-Vis-induced sensors can be versatile in their uses, providing an interesting tool to characterize crops for precision agriculture. The study conducted by [13] found encouraging results using UV-Vis-induced fluorescence to predict soil chemical properties of air-dried samples.

Fluorescence spectroscopy consists of measuring the photoluminescence of molecules that emit light after having absorbed ultraviolet, visible, or infrared light. For plants, pigments (chlorophyll and anthocyanin) act as key responsive parameters for fluorescence emission at a specific wavelength of excitation. Differential sensitivity of the emission spectrum (at far-red and red) at specific excitation wavelengths (green and red) can be

observed in wheat leaves [26]. These differences in the emission spectrum help to rationalize the chlorophyll indices (SFR\_G and SFR\_R) and anthocyanin indices (ANTH\_RG and ANTH\_RB) for plants that are generated by the Multiplex3 sensor. The soil fluorescence excitation-emission matrix also demonstrates such differential sensitivities at the FRF and RF emission spectrum [27]. Among soil components, organic matter components such as humic and fulvic acids have fluorescent properties [28,29]. Such fluorescent behavior relies on the aromaticity, aliphatic character, degree of polycondensation, content of carboxylic groups or organic free radicals, or presence of amide groups or polysaccharidic structures' pH value [7,13,30,31]. Such differences in the emission spectrum induced by excitation wavelengths can be potentially traced using similar indices such as SFR\_G and SFR\_R. Hence, it was hypothesized that soil parameters might be predicted via a fluorescence spectroscopy instrument designed for plants.

The hypothesis of this project was that UV-Vis-induced fluorescence sensors can be used for instantaneous assessment of soil fertility. The specific objectives were to assess if induced fluorescence of soil can be used to (1) estimate soil chemical properties, (2) classify soil samples within fertility classes of soil properties, and (3) predict the N fertilizer rate recommendation.

## 2. Materials and Methods

### 2.1. Sites and Soil Sampling

Field data were acquired during the crop growing season over two years (2013 and 2016) from nine sites located in Colorado (Table 1). Fields were selected with the information from the farmers about contrasting soils within their farm. The objective was to acquire soil samples having a broad range of values for each soil property. Soil was sampled from the top 20 cm at random locations within fields with a 2.5 cm diameter soil sampling probe. At each location, samples were composites collected to total about 500 g of wet soil (6 to 10 cores collected within a 1 m radius circle around the geolocated sampling point). Soil samples were air dried, evenly spread in a shallow container, and scanned with a fluorescence sensor (described below). After fluorescence reading acquisition, soil samples were sent to Servi-Tech Laboratories (Dodge City, KS, USA) where soil analysis was performed. Soil pH and soluble salts were determined by measurement in a 1:1 soil:water slurry. Soil organic matter content (SOM) was determined using the weight-loss-on-ignition method [32]. Soil NO<sub>3</sub>-N was determined by the Cd reduction method, and P was extracted by the Mehlich-3 method [33,34]. Soil concentration in K, S, Ca, Mg, and Na was measured with the ammonium acetate method [35]. Soil concentration in Zn, Fe, Mn, and Cu was measured with the diethylenetriaminepentaacetic acid (DTPA) method [36]. Particle size analysis (soil texture) was performed using the hydrometer method [37].

**Table 1.** Location and sample size acquired along with soil classification for each site.

Site	Location (Lat. Lon.)	Sample Size	Soil Series †
Site 1	Wellington, CO (40°40' N, 104°59' W)	60	Kim loam (Fine-loamy, mixed, active, calcareous, mesic Ustic Torriorthents)
		22	Nunn clay loam (Fine, smectitic, mesic Aridic Argiustolls)
Site 2	Atwood, CO (40°33' N, 103°16' W)	10	Nunn clay loam (Fine, smectitic, mesic Aridic Argiustolls)
		2	Haverson loam (Fine-loamy, mixed, superactive, calcareous, mesic Aridic Ustifluvents)
Site 3	Ault, CO (40°34' N, 104°43' W)	12	Kim loam (Fine-loamy, mixed, active, calcareous, mesic Ustic Torriorthents)
Site 4	Iliff, CO (40°46' N, 103°02' W)	8	Loveland clay loam (Fine-loamy over sandy or sandy-skeletal, mixed, superactive, calcareous, mesic Fluvaquentic Endoaquolls)
		6	Nunn clay loam (Fine, smectitic, mesic Aridic Argiustolls)
Site 5	Fort Collins, CO (40°36' N, 104°59' W)	6	Nunn clay loam (Fine, smectitic, mesic Aridic Argiustolls)
		4	Santana loam (Loamy, mixed, superactive, mesic Aridic Lithic Haplustolls)
Site 6	Severance, CO (40°31' N, 104°52' W)	10	Kim loam (Fine-loamy, mixed, active, calcareous, mesic Ustic Torriorthents)
Site 7	Lucerne, CO (40°28' N, 104°41' W)	5	Colby loam (Fine-silty, mixed, superactive, calcareous, mesic Aridic Ustorthent)
		4	Weld loam (Fine-silty, mixed, mesic, Aridic Argiustoll)
Site 8	LaSalle, CO (40°17' N, 104°39' W)	1	Ascalon loam (Fine-loamy, mixed, mesic, Aridic Argiustoll)
		5	Olney fine sandy loam (Fine-loamy, mixed Ustolic Haplargids)
		6	Otero sandy loam (Coarse-loamy, mixed (calcareous), mesic Ustic Torriorthents)

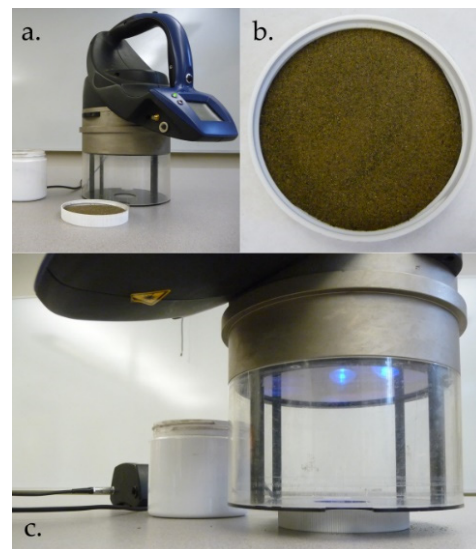
**Table 1.** *Cont.*

Site	Location (Lat. Lon.)	Sample Size	Soil Series †
Site 9	Pierce, CO (40°36' N, 104°42' W)	5	Docono clay loam (Clayey over sandy or sandy-skeletal, smectitic, mesic Aridic Argiustolls)
		2	Nunn clay loam (Fine, smectitic, mesic Aridic Argiustolls)

† [38,39].

## 2.2. Fluorescence Sensor

The sensor used for this study was the portable Multiplex MX3 multi-parameter fluorescence sensor (FORCE-A, Orsay, France; Figure 1a). The four excitation channels are UV (around 375 nm), blue (around 470 nm), green (around 515 nm), and red (around 625 nm; Table 2). Excitation light pulses (20  $\mu$ s per flash) are delivered by high-power light emitting diode arrays located around the detectors and pointing in the direction of the sensed area. The three detection channels are yellow (590 nm  $\pm$  40 nm; YF), red (678 nm  $\pm$  22 nm; RF), and far-red (750 nm  $\pm$  65 nm; FRF). The detectors consist of three silicon photodiodes (20 mm  $\times$  20 mm), each having an optical bandpass filter allowing only YF, RF, or FRF light to reach the photodiode. The flash induces the emission of fluorescence, and the filters allow the selection of the wavebands of interest. A firmware synchronizes the light pulses and the detectors in order to acquire each combination (12 in total) of excitation wavebands and detection channels for about 70 readings per second. Fluorescence based indices are generated by the sensor's firmware (Table 3). More details about the sensor hardware can be found in Cerovic et al. [40].



**Figure 1.** Multiplex MX3 sensor (a), soil disposed in a plate (container lid) ready for sensing (b), and fluorescence acquisition of the soil sample with the Multiplex MX3 (c).

**Table 2.** Presentation of the nine fluorescence and two reflectance (underlined) signals acquired by the Multiplex MX3 at each reading. Subscripts indicate the induction channel.

Detection Channel	Induction Channel			
	UV	Red (R)	Green (G)	Blue (B)
Yellow (YF)	$\underline{YF_{UV}}$	$\underline{YF_R}$	$\underline{YF_G}$	$\underline{YF_B}$
Red (RF)	$\underline{RF_{UV}}$	$\underline{RF_R}$	$\underline{RF_G}$	$\underline{RF_B}$
Far-red (FRF)	$\underline{FRF_{UV}}$	$\underline{FRF_R}$	$\underline{FRF_G}$	$\underline{FRF_B}$

**Table 3.** Fluorescence indices used for this study along with their description and formula.

Parameter	Description	Formula *
SFR_G	Chlorophyll index with green induction	$\frac{1}{250} \sum_{i=1}^{250} \frac{FRF_{G_i}}{RF_{G_i}}$
SFR_R	Chlorophyll index with red induction	$\frac{1}{250} \sum_{i=1}^{250} \frac{FRF_{R_i}}{RF_{R_i}}$
FLAV	Index of compounds which absorbs at 375 nm, often flavonoids	$\log \left( \frac{1}{250} \sum_{i=1}^{250} \frac{FRF_{R_i}}{FRF_{UV_i}} \right)$
FER_RG	Chlorophyll ratio originally designed for fruit anthocyanin content	$\frac{1}{250} \sum_{i=1}^{250} \frac{FRF_{R_i}}{FRF_{G_i}}$
FERARI	Index of anthocyanins on grapes	$\log \left( \frac{1}{250} \sum_{i=1}^{250} \frac{5000}{FRF_{R_i}} \right)$
ANTH_RG	Index of anthocyanin with green induced denominator	$\log \left( \frac{1}{250} \sum_{i=1}^{250} \frac{FRF_{R_i}}{FRF_{G_i}} \right)$
ANTH_RB	Index of anthocyanin with blue induced denominator	$\log \left( \frac{1}{250} \sum_{i=1}^{250} \frac{FRF_{R_i}}{FRF_{B_i}} \right)$
NBI_R	Nitrogen balance index (red)	$\frac{1}{250} \sum_{i=1}^{250} \frac{FRF_{UV_i}}{RF_{R_i}}$
NBI_G	Nitrogen balance index (green)	$\frac{1}{250} \sum_{i=1}^{250} \frac{FRF_{UV_i}}{RF_{G_i}}$
NBI_Rm **	Ratio of UV induced far-red fluorescence on red light induced red fluorescence	$\frac{\frac{1}{250} \sum_{i=1}^{250} FRF_{UV_i}}{\frac{1}{250} \sum_{i=1}^{250} RF_{R_i}}$
NBI_Gm **	Ratio of UV induced far-red fluorescence on green light induced red fluorescence	$\frac{\frac{1}{250} \sum_{i=1}^{250} FRF_{UV_i}}{\frac{1}{250} \sum_{i=1}^{250} RF_{G_i}}$
NBI_Bm **	Ratio of UV induced far-red fluorescence on blue light induced red fluorescence	$\frac{\frac{1}{250} \sum_{i=1}^{250} FRF_{UV_i}}{\frac{1}{250} \sum_{i=1}^{250} RF_{B_i}}$
NBI_UVm **	Ratio of UV induced far-red fluorescence on UV induced red fluorescence	$\frac{\frac{1}{250} \sum_{i=1}^{250} FRF_{UV_i}}{\frac{1}{250} \sum_{i=1}^{250} RF_{UV_i}}$

\* Induction waveband is in subscript. UV = Ultra-violet; G = Green; R = Red; B = Blue. \*\* This parameter was not automatically computed by the sensor but calculated afterward.

### 2.3. Data Acquisition

Fluorescence readings were acquired in laboratory conditions on air dried soil samples. Dried soil samples were placed in a plate (container lid) to an even 1 cm thickness of soil (Figure 1b). A black metal mask was installed on the sensor to restrain the field-of-view of the sensor to a 40 mm diameter circle (Figure 1a). The Multiplex MX3 was placed on the soil plate and triggered to acquire the fluorescence measurement (Figure 1c). The sensor was set to acquire an average of over 250 induction/detection cycles for each sample. Soil measurements were acquired over the 168 soil samples collected from the field.

### 2.4. Statistical Analysis

Descriptive statistics including average, minimum, maximum, standard deviation, skewness, and kurtosis were calculated for soil properties over the 168 soil samples in this database. The R software [41] package stats was used to produce the descriptive statistics. Density plots were generated for the soil target variables to realize the underlying distribution of the data. Subsequently, the Anderson–Darling test was performed to determine whether the data followed a normal distribution. This type of test is useful for testing normality, which is a common assumption used in many statistical approaches including linear regression and ANOVA. The test statistics (TS) value of the Anderson–Darling test is generally compared to each critical value that corresponds to significance level  $\alpha = 0.01$  and  $0.05$  to see if the test results are significant. The density plots indicated non-normal distribution for all soil target variables (data not shown). The Anderson–Darling test statistics results were significant at both significance levels, which means the null hypothesis can be rejected. It was apparent from the test statistics that the sample data sets were not normally distributed.

#### 2.4.1. Estimation of Soil Properties with Random Forest Regression (RFR)

Linear regression techniques are often used for estimation of dependent parameters (soil properties) from independent parameters (optical measurements). However, the linear regression is a parametric method that needs explicit modeling of nonlinearities in the data and interactions between the parameters [42]. Moreover, the inferential procedures for linear regression are typically based on a normality assumption for the residuals. For estimation of soil properties, these conditions rarely apply since data distributions are usually not known a priori. Moreover, the optical measurements in soil fluorescence spectroscopy are not always independent of each other. In such conditions, machine learning-based methods such as Random Forest Regression (RFR) are more robust, as they do not require that the underlying distribution of the data be known a priori and do not assume independence amongst the predictors. The RFR is a nonparametric method, and unlike linear regression, it does not require nonlinearities and parameter interactions to be explicitly modeled since these can be learned from the data themselves.

In recent years, RFR has been used widely in applications related to soil-crop sensing methods [43–45]. The RFR is an ensemble learning technique developed by Breiman [46] that involves combining a large set of decision trees generated independently so that no two trees are the same. The independence between the trees is achieved by randomly selecting one third of the predictors at each node for node splitting and by using a random bootstrap sample comprising about 67% of the training samples to build each tree of the random forest. The remaining 33% of samples are called out-of-bag samples that are used to obtain an error estimate based on the bootstrap subset. At each node, the best split is chosen to form child nodes [47]. The value of each child node is the average of the sample values in that node. The node splitting is based on minimization of the Mean Squared Errors for a tree.

The RFR also identifies a subset of important independent parameters from the total parameter set which are relevant for the regression model for each independent parameter. In RFR, the most used approach is the Mean Decrease in Impurity (MDI) score-based feature importance. Features with high scores are further used in the regression model for each soil parameter. The Pearson's  $r$  coefficient of correlation between observed and estimated values were calculated for the training and test datasets. The Random Forest regression model training and validation were implemented using the open-source Python Scikit-learn packages.

#### 2.4.2. Predicting Fertility Classes

To assess the potential of the Multiplex MX3 as a maize fertilization management tool, soil properties were classified as per the maize fertilization guide of the Colorado State University Extension [48,49] (Table 4). Three, four, or five classes were created depending on the soil property (Table 4). Soil properties that most often require intervention in Colorado soils (i.e.,  $\text{NO}_3\text{-N}$ , SOM, P, K, Zn, and S) were reported as well as salts' and minor elements' (i.e., Mn and Cu) content [48,50]. From these soil properties, only the ones for which there was enough representation in each fertility class were used for classification, notably  $\text{NO}_3\text{-N}$ , SOM, and Zn.

The random forest classifier was used to predict the soil fertility classes based on fluorescence readings. Similar to the RFR approach, the random forest classifier uses an ensemble learning technique which is constructed by several decision trees that are trained, and their results are combined through a voting process by the majority of the individual decision trees [46]. The multiple decision trees of the random forest are trained on a bootstrapped sample of the original training data. In general, the random forest increases the diversity among the decision trees by randomly resampling the data with replacement and by randomly changing the parameter subsets for node splitting at each node of every decision tree. The R package "random Forest" was used to conduct the random forest classification [41,51].

**Table 4.** Classification of selected soil properties values for maize fertilization in Colorado.

Soil Property	Soil Fertility Level				
	Very Low	Low	Medium	High	Very High
NO <sub>3</sub> -N (ppm)	0–6	7–12	13–18	19–24	>24
SOM (%)	-	0–1.0	1.1–2.0	>2.0	-
P (ppm) †	-	0–10	11–31	31–56	>56
K (ppm)	-	0–60	61–120	>120	-
Zn (ppm)	-	0–0.9	1.0–1.5	>1.5	-
S (ppm)	-	0–6	6–8	>8	-
Fe (ppm) ‡	-	0–3	3–5	>5	-
Salts ‡	-	0–2	2–4	4–8	>8
Mn (ppm) ‡	-	0–0.5	>0.5	-	-
Cu (ppm) ‡	-	0–0.2	>0.2	-	-

† P was not reported based on Melich-3 method in Davis and Westfall, 2014, but was reported by Bauder et al., 2003 [49], ‡ Classes for these properties come from Self, 2010 [50].

From the random forest classification analysis output (i.e., soil fertility classes estimated from fluorescence readings), a confusion matrix was built between observed and estimated values, which enabled quantification of true and false positives and calculation of prediction performance statistics. For each class, the area under the curve (*AUC*) was calculated as follows:

$$AUC = \frac{TP \times FP}{2} + \frac{(1 - FP) \times (1 + TP)}{2} \quad (1)$$

where *TP* is the true positive rate, and *FP* is the false positive rate. The *AUC* was computed for each class of each selected soil property. The *AUC* ranges from 0 to 1 with an *AUC* below 0.5 indicating estimation worse than randomness, an *AUC* of 0.5 being equivalent to random prediction, and an *AUC* of 0.6–0.7, 0.7–0.8, 0.8–0.9, or 0.9–1 indicating a poor, average, good, or excellent prediction potential, respectively. The overall accuracy (*OA*) and the balanced accuracy (*BA*) were calculated for both the training and the test dataset of each soil property.

The *OA* is calculated as follows:

$$OA = \frac{\sum_{i=1}^c TP_i}{N} \quad (2)$$

where *OA* is the overall accuracy; *i* is the *i*th class (e.g., low, medium, high); *c* is the number of classes in the confusion matrix; *TP<sub>i</sub>* is the number of true positives for class *i*; and *N* is the total number of observations in the dataset.

The *BA* is calculated as follows:

$$BA = \frac{\sum_{i=1}^c Recall_i}{c} \quad (3)$$

where *BA* is the balanced accuracy; *i* is the *i*th class (e.g., low, medium, high); *c* is the number of classes in the confusion matrix; and *Recall* is the number of correctly estimated observations in class *i* out of the number of actual observations in class *i* [52]. The *BA* cannot be calculated with classes containing zero observations, and thus classes with zero observation are eliminated before calculation of *BA*. The *BA* enables a better prediction for under-represented classes than the traditional overall accuracy, which provides a prediction assessment for the entire dataset. A simple measure of accuracy may be misleading and is better interpreted when accompanied by a baseline calculation. The baseline accuracy (*BASE*) is calculated by classifying all samples as the most common class in the dataset and by using equation 2 to measure the *OA* of this modified confusion matrix.

In order to compare the prediction of continuous data (Obj. 1) to the prediction of fertility classes (Obj. 2), the continuous data were converted to classes using Table 4. This enabled the comparison of both approaches based on overall accuracy rather than comparing Pearson's *r* of the regression method to the overall accuracy of the classification method.

The *AUC*, *BASE*, *OA*, and *BA* were reported for each selected soil properties. A custom R code was written to compute the confusion matrix, the baseline, the *OA*, the *BA*, and the *AUC* [41].

### 2.4.3. Predicting Nitrogen Rate Recommendation

The N rate recommendation for maize can be calculated based on the soil  $\text{NO}_3\text{-N}$  and SOM content [53]. The decision algorithm is shown in Table 5. For each soil sample, the recommended N rate was calculated based on this decision algorithm. Subsequently, for each sampling point, the N rate is tabulated against corresponding fluorescence readings. A random forest classification model was generated using fluorescence readings as predictors and the N rate as the target variable. This random forest based trained model allowed us to estimate the N rate directly for a given fluorescence reading.

**Table 5.** Suggested nitrogen rates ( $\text{kg N ha}^{-1}$ ) for irrigated maize, as related to  $\text{NO}_3\text{-N}$  in the soil and soil organic matter content, calculated from the algorithm. Target yield for this algorithm is 11 Mg grain per ha, and recommended N rate does not account for other N credits. Adapted from Davis and Westfall (2014) [48].

$\text{NO}_3\text{-N}$ (mg/kg) *	Soil Organic Matter (%)		
	0–1.0	1.1–2.0	>2.0
0–6	235	207	185
7–12	179	151	129
13–18	123	95	73
19–24	67	39	17
>24	11	0	0

\* Average weighted concentration ( $\text{mg kg}^{-1}$ ) in the tillage and subsoil layers.

## 3. Results and Discussion

### 3.1. Statistical Description of Soil Properties

As anticipated, the collected soil samples from nine experimental sites displayed a wide range of values for most soil properties (Table 6). As per the Kurtosis of the distributions, most soil properties showed a fair level of variability, except for Zn and Cu, for which the standard deviation was low, and Kurtosis was high. The range of values (i.e., minimum to maximum) of  $\text{NO}_3\text{-N}$ , SOM, P, and Zn included all levels of soil fertility (Table 4), and the range of values of S and Salts included at least two levels of soil fertility from Table 4. The soil properties  $\text{NO}_3\text{-N}$ , SOM, and Zn had at least 10% of observations in each fertility class, while the soil properties P and Fe had less than 10% of observations in the low classes. Other soil properties that could be divided into fertility classes as per Table 4 (i.e., K, S, Salt, Mn, and Cu) had classes with 1% or fewer observations. This indicates that  $\text{NO}_3\text{-N}$ , SOM, and Zn properties did provide enough data to meet the objective of assessing classification of fertility based on fluorescence data, while P and Fe may provide only partial insights, and the other properties are not suitable to meet this objective.

**Table 6.** Descriptive statistics of soil properties for the entire dataset.

	Mean	Min.	Max.	Standard Deviation	Kurtosis	Skewness
pH	8.09	7.10	8.40	0.24	3.31	−1.63
Salts	0.66	0.23	2.28	0.39	3.18	1.90
SOM † (%)	1.43	0.40	2.60	0.45	−0.05	0.43
$\text{NO}_3\text{-N}$ (mg/kg)	15	1	100	16	9.46	2.76
P (mg/kg)	59	3	284	49	6.13	2.13
K (mg/kg)	317	147	815	129	2.13	1.39
S (mg/kg)	97	7	709	124	7.64	2.64
Ca (mg/kg)	4113	1188	5322	950	1.68	−1.58
Mg (mg/kg)	594	167	931	125	2.34	−1.24
Na (mg/kg)	125	25	819	136	8.36	2.96
Zn (mg/kg)	2.2	0.3	20.8	2.4	21.84	3.99
Fe (mg/kg)	7.5	2.0	42.0	6.3	11.45	3.27
Mn (mg/kg)	7.1	3.0	23.0	2.5	10.85	2.30
Cu (mg/kg)	1.2	0.4	8.5	0.9	28.49	4.43



Table 6. Cont.

	Mean	Min.	Max.	Standard Deviation	Kurtosis	Skewness
CEC †	26.9	8.0	33.0	5.51	2.55	−1.75
Sand (%)	49.3	30.0	80.0	9.34	1.09	0.83
Silt (%)	22.0	10.0	37.5	5.15	0.16	0.38
Clay (%)	28.8	10.0	37.5	6.11	1.01	−1.07

† Soil organic matter content., ‡ Cation exchange capacity.

### 3.2. Fluorescence Features to Estimate Soil Parameter Using RFR

The details of the RFR feature importance are presented in Table 7. Results showed a strong disparity in terms of which optical feature was the most important for each soil parameter based on MDI scores (Table 7). When considering only the twelve signals, and not the optical ratios and parameters, YF\_R, YF\_UV, and YF\_G were the most frequent important predictors. This indicates that the yellow filter seems important for soil properties' prediction. As mentioned in the previous section, YF\_R and YF\_G are reflectance rather than fluorescence signals (Table 2) because induction wavebands for these signals cannot trigger fluorescence in the yellow part of the spectrum. Nevertheless, because YF\_UV is also one of the best predictors, it does not lead to concluding that reflectance is better than fluorescence to estimate soil properties. The yellow filter (~590 nm) signals being better predictors opens the question of the performance of filters for shorter wavebands such as the blue-green filter, which is an option with the Multiplex MX3 system. The UV-induced fluorescence (351 nm) of whole soil samples shows the highest intensity of fluorescence from 475 to 525 nm, which may be an optimal zone for detecting subtle differences in soil samples' fluorescence [17]. Better performances with filters in shorter wavebands are consistent with [54] who observed greater disparity in fluorescence intensity from soils' humic acid component at shorter (~400 to 600 nm) wavebands. In their study, the greatest differences in fluorescence from soils coming from different sampling locations appeared around 475 nm. In the current study, the yellow filters performed better as independent signals, and the literature seems to suggest that shorter wavebands' filter (e.g., blue) may perform even better.

In terms of induction channels, the fluorescence measurements with UV and red induction were more important in predicting soil properties as per MDI scores (Table 7), followed by green and blue induction bands. Even if fluorophores react differently to different induction bands, it is unclear why this sequence was observed because few studies have tested the effect of using different induction bands on the discrimination potential of whole soil samples. Although conducted on forest soil from the Amazon Forest with high organic matter content, a study using two different induction wavebands (i.e., 378 nm and 445 nm) observed resulting fluorescence spectra with marked differences, notably with two peaks for the shorter waveband and a single peak for the longer waveband [55]. This indicates that different fluorophores may be triggered with different induction wavebands, thus resulting in different fluorescence spectra. In this study, both red and UV induction bands performed somewhat similarly in terms of prediction and may be preferred to other bands for designing a soil properties sensor that contains fewer elements (e.g., number of induction LEDs).

In general, ratios and fluorescence indices provided better prediction power than individual signals (Table 7). For instance, when cumulating relative importance (i.e., MDI scores) of each signal over all soil properties in Table 7, the two best indices, NBI\_UVm and FLAV, were 5.4 and 2.6 times more important, respectively, than the best predictor signal (YF\_R) for estimating soil properties. Both indices use signals with UV and/or red induction, which are the induction bands that provided the best prediction potential as per results shown above. However, no indices from Table 3 used the YF filter signals, which were the best raw signals to predict soil properties as per results shown above. This is related to the Multiplex MX3 being originally designed to sense fruits and vegetation rather than soil, and all proposed indices target vegetation fluorophores such as chlorophyll, anthocyanins, and flavonoids [40]. For instance, the NBI\_UVm is a fluorescence index

that is not automatically generated by the sensor, and yet, it is the one that yielded the best results for the prediction of soil properties. As suggested by Longchamps and Khosla (2014), it is possible that the calculation of the index using averaged signal values over 250 readings, rather than calculating the index 250 times and then averaging it, may have helped to stabilize the values and improved the estimation power. Findings from the current study suggest that research should be conducted on full spectrum fluorescence of whole soil samples induced by different wavelengths to identify new indices dedicated to soil analysis.

**Table 7.** Random Forest Regression feature or optical measurement importance for each soil parameter. The Mean Decrease in Impurity (MDI) scores are indicated along with a grayscale gradient showing higher values in darker gray tones. The sum of each column is equal to one.

Fluorescence Measurements	Soil Properties																	
	pH	Salt	OM	N	P	K	S	Ca	Mg	Na	Zn	Fe	Mn	Cu	CEC	Sand	Silt	Clay
YF_UV	0.00	0.01	0.00	0.05	0.13	0.03	0.01	0.17	0.01	0.01	0.01	0.01	0.26	0.08	0.02	0.01	0.02	0.02
RF_UV	0.00	0.00	0.01	0.02	0.04	0.01	0.03	0.04	0.01	0.00	0.01	0.01	0.05	0.00	0.09	0.01	0.02	0.00
FRF_UV	0.00	0.00	0.01	0.01	0.04	0.02	0.01	0.01	0.01	0.02	0.01	0.01	0.02	0.00	0.05	0.01	0.02	0.01
YF_B	0.01	0.01	0.02	0.01	0.02	0.01	0.05	0.20	0.01	0.00	0.01	0.01	0.03	0.02	0.04	0.02	0.02	0.02
RF_B	0.00	0.00	0.01	0.02	0.01	0.00	0.02	0.01	0.00	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.01	0.01
FRF_B	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.00	0.01	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01
YF_G	0.01	0.01	0.30	0.02	0.19	0.06	0.02	0.02	0.01	0.00	0.02	0.01	0.06	0.04	0.02	0.03	0.01	0.01
RF_G	0.01	0.00	0.00	0.01	0.01	0.00	0.03	0.01	0.01	0.00	0.00	0.01	0.01	0.02	0.01	0.01	0.01	0.01
FRF_G	0.00	0.00	0.02	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.01
YF_R	0.10	0.00	0.41	0.01	0.08	0.03	0.02	0.04	0.01	0.00	0.03	0.03	0.04	0.04	0.03	0.03	0.02	0.01
RF_R	0.00	0.01	0.03	0.04	0.04	0.04	0.01	0.01	0.01	0.00	0.00	0.02	0.02	0.01	0.00	0.03	0.01	0.01
FRF_R	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.01	0.02	0.00
SFR_G	0.14	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.01	0.01	0.00	0.07	0.06	0.00	0.00	0.01	0.01	0.00
SFR_R	0.00	0.14	0.01	0.10	0.04	0.02	0.04	0.02	0.02	0.00	0.05	0.01	0.05	0.00	0.01	0.17	0.46	0.09
FLAV	0.08	0.59	0.00	0.20	0.01	0.45	0.39	0.06	0.10	0.26	0.01	0.01	0.02	0.01	0.04	0.04	0.02	0.14
FER_RG	0.01	0.00	0.00	0.04	0.02	0.05	0.00	0.01	0.02	0.00	0.01	0.03	0.02	0.00	0.01	0.02	0.02	0.03
ANTH_RG	0.01	0.00	0.00	0.03	0.02	0.05	0.01	0.01	0.02	0.01	0.01	0.03	0.04	0.00	0.00	0.02	0.02	0.06
ANTH_RB	0.02	0.00	0.01	0.16	0.01	0.02	0.01	0.01	0.03	0.01	0.01	0.01	0.04	0.01	0.02	0.03	0.03	0.10
NBL_G	0.04	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.00	0.03	0.04	0.01	0.01	0.01	0.01	0.00
NBL_R	0.00	0.09	0.01	0.12	0.06	0.08	0.10	0.02	0.02	0.22	0.01	0.01	0.02	0.01	0.02	0.08	0.02	0.02
FERARI	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.02	0.00
NBL_Rm	0.00	0.10	0.01	0.09	0.04	0.05	0.10	0.01	0.02	0.29	0.01	0.00	0.02	0.02	0.02	0.14	0.02	0.02
NBL_Gm	0.05	0.00	0.00	0.02	0.02	0.01	0.01	0.02	0.01	0.00	0.00	0.04	0.01	0.02	0.01	0.02	0.01	0.01
NBL_Bm	0.02	0.03	0.12	0.01	0.02	0.01	0.08	0.02	0.02	0.12	0.01	0.01	0.04	0.08	0.03	0.04	0.02	0.02
NBL_UVm	0.47	0.01	0.01	0.02	0.15	0.03	0.02	0.30	0.58	0.01	0.79	0.63	0.06	0.59	0.55	0.24	0.16	0.38

### 3.3. Estimating Soil Parameters Using Induced Fluorescence

The Figure 2a,b show scatter plots between the observed and estimated values of the training and test datasets for each soil properties and report the Pearson's  $r$  coefficient of correlation. In general, there was a large difference (i.e., average of 0.15-point difference) between the  $r$  values of the training versus the test datasets, which seems to indicate a slight underfitting of the training data associated with a small number of observations covering a limited range of values. Nevertheless, RFR is an artificial intelligence technique that is less prone to under/overfitting, and thus due diligence was conducted to avoid overfitting [46].

When considering the Pearson's  $r$  coefficient of correlation for the test dataset, all soil properties displayed an  $r$  value of 0.57 or above between the estimated and observed values (Figure 2a,b). This demonstrates that there was a significant positive correlation when trying to predict soil properties using optical features, which was one of the main objectives of this study. Nevertheless, the  $r$  values ranged from 0.57 for Mg to 0.81 for Na, indicating that this approach did not perform similarly across all soil properties. There was no discernible commonality among the soil properties with the highest  $r$  values, and neither among the soil properties with the lowest  $r$  values. Nevertheless, soil organic matter (SOM) is a soil fraction containing several fluorophores [31,56], and it did appear among the top five soil properties with an  $r$  value of 0.74. Ref. [57] observed a drastic difference between the UV-induced fluorescence spectrum of the whole soil sample with and without (i.e., calcinated soil) the organic matter. This indicates that the large majority of whole soil sample fluorescence is emitted from organic matter, which explains the relatively high  $r$  value (i.e., 0.74 for the test dataset) between observed and estimated SOM

in this study. Other soil properties, notably pH, CEC, Fe, Ca, and Na, showed a better correlation between observed and estimated values than SOM with Pearson's  $r$  of 0.75, 0.75, 0.76, 0.76, and 0.81, respectively, for the test dataset (Figure 2a,b). Ref. [13], who used the Multiplex MX3 equipped with a blue-green filter in place of the yellow filter to estimate chemical properties of whole soil samples, also found pH, Ca, CEC, and Fe on top of the list in terms of predictability. In their study, the Na prediction using the same system showed lower accuracy. Interestingly, pH and CEC are soil properties rather than elemental concentrations, and yet both showed high potential for predictability using UV-induced fluorescence. Because of their influence on soil fertility and elemental concentration, other soil properties combined can act as a proxy for determining soil pH (e.g., Al, Fe, and Mn) and CEC (e.g., Ca, Mg, K, Cu, Zn, and Fe) using induced fluorescence [11,58]. This may explain the higher accuracy of induced fluorescence to predict pH and CEC.

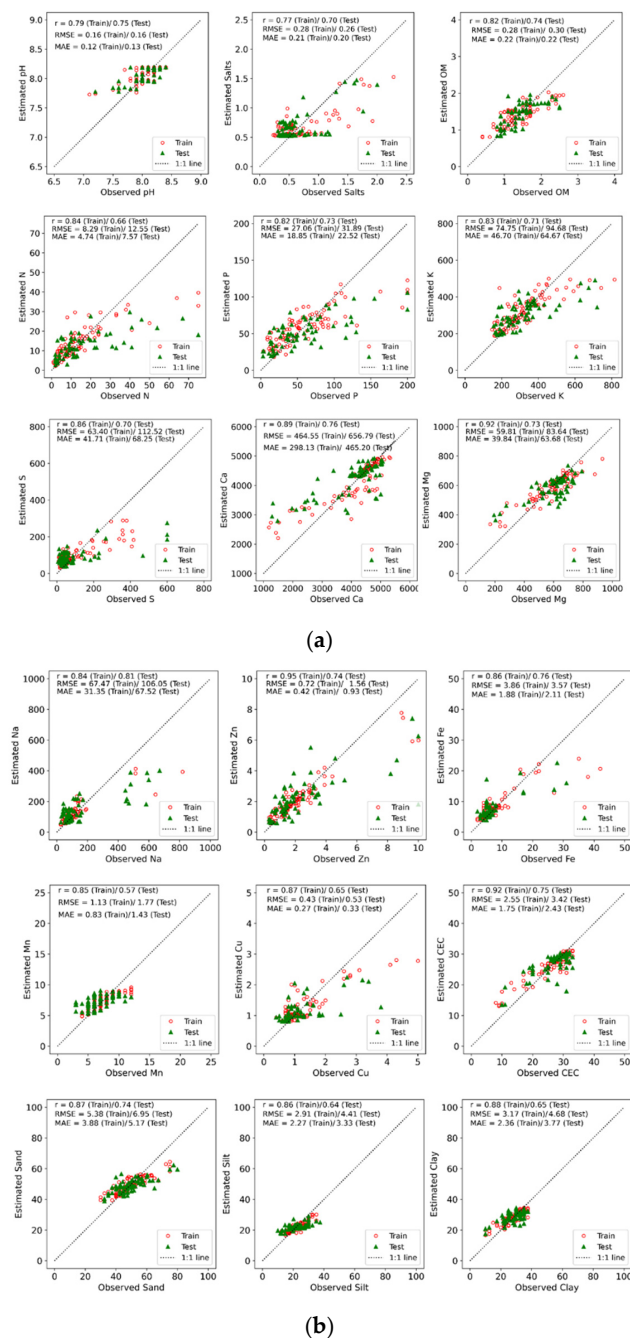


Figure 2. (a) Scatter plot of the estimated to observed values of each soil property as per random

forest regression analysis. The Pearson's  $r$  coefficient of correlations and two error estimates, i.e., Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), are indicated for both training and test dataset for each soil property. (b) Scatter plot of the estimated to observed values of each soil property as per random forest regression analysis. The Pearson's  $r$  coefficient of correlations and two error estimates, i.e., Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), are indicated for both training and test dataset for each soil property.

Overall, the simple use of the portable Multiplex MX3 sensor enabled high prediction of a number of soil properties (e.g., SOM, pH, and CEC) that are important for assessing soil fertility. This indicates that, as also observed by [13] and by [11,12], fluorescence sensing has good potential for the rapid assessment of soil fertility. As indicated by the observed performance and statistics of the current dataset, a larger dataset showing a wider range of each soil property is required to improve the prediction algorithms further. While obtaining a precise estimate of the soil properties is desirable from an agronomic standpoint, categorical estimates may be more convenient for decision making while requiring less accurate sensing technologies.

### 3.4. Estimating Fertility Classes of Selected Soil Properties Using UV-Induced Fluorescence

The soil properties  $\text{NO}_3\text{-N}$ , SOM, and Zn had enough observations ( $n \geq 10\%$ ) in each fertility class to test the second objective (i.e., predicting fertility classes with fluorescence) of this study. Results show that the OA was 0.54, 0.78, and 0.69 for the test dataset of  $\text{NO}_3\text{-N}$ , SOM, and Zn, respectively (Table 8). This represents an improvement of 0.23, 0.12, and 0.25 points over the baseline accuracy for  $\text{NO}_3\text{-N}$ , SOM, and Zn, respectively. In general, the BA was slightly inferior than the OA, indicating an unbalanced distribution of observations across the classes. The hypothesis for this procedure was that trying to predict broad classes would yield better results than trying to predict exact values, with the rationale that, in the end, exact values are often converted to fertility classes for decision making. These results seem to suggest that this approach did not yield substantial improvement over a direct regression approach. Because of the inability directly to compare regression performance coefficients such as Pearson's  $r$  to classification performance coefficients such as OA, continuous data output from the RFR was converted to classes. The OA of the continuous data output from RFR converted to classes using Table 4 was 0.50, 0.75, and 0.60 for the test dataset of  $\text{NO}_3\text{-N}$ , SOM, and Zn, respectively (data not shown). The results from the class prediction method thus suggest a slight improvement (i.e., 0.05-, 0.03-, and 0.09-point improvement for  $\text{NO}_3\text{-N}$ , SOM, and Zn, respectively) as compared to the prediction of continuous values. Since predicting classes is not an additional burden for computation as compared to predicting continuous values and that continuous values are often converted to classes, these results suggest that predicting classes should be the preferred option. Trontelj ml. and Chambers [59] observed an improvement in prediction results with a decrease in the number of classes when using artificial intelligence to predict soil properties out of the optical spectral signature. Although hardly comparable across different properties, the results of the current study seem to indicate that a larger number of classes (e.g., five classes for  $\text{NO}_3\text{-N}$ ) yielded lower OA as compared to fewer classes (e.g., three classes for SOM and Zn). Overall, the class prediction approach seems to be preferable over the continuous prediction approach, and there is a trend towards a fewer number of classes yielding better prediction accuracy.

**Table 8.** Area under the curve (AUC) values for each class of each soil property that can be separated into fertility classes (see Table 4). The percentage of N (number of observation) in each class within each dataset is indicated within parenthesis. The baseline (BASE), overall accuracy (OA), and balanced accuracy (BA) calculated with the confusion matrix of each soil property are indicated.

Soil Parameter	Training Dataset (n = 100)								Test Dataset (n = 68)							
	Fertility Classes					BASE	OA	BA	Fertility Classes					BASE	OA	BA
	Very Low	Low	Medium	High	Very High				Very Low	Low	Medium	High	Very High			
NO <sub>3</sub> -N	0.82 (31)	0.71 (39)	0.66 (10)	0.75 (7)	0.92 (13)	0.39	0.65	0.68	0.81 (31)	0.72 (25)	0.66 (16)	0.54 (13)	0.74 (15)	0.31	0.54	0.50
SOM		0.87 (23)	0.79 (66)	0.68 (11)		0.66	0.84	0.81		0.86 (24)	0.72 (66)	0.50 (10)		0.66	0.78	0.57
P		0.83 (8)	0.84 (31)	0.77 (31)	0.90 (30)	0.43	0.80	0.66		0.50 (2)	0.79 (40)	0.63 (4)	0.78 (54)	0.40	0.66	0.48
K				1.00 (100)		1.00	1.00	1.00				1.00 (100)		1.00	1.00	1.00
Zn		0.82 (22)	0.60 (25)	0.81 (53)		0.53	0.74	0.70		0.80 (26)	0.60 (29)	0.79 (54)		0.44	0.69	0.64
S			0.50 (98)	0.50 (2)		1.00	0.98	1.00				1.00 (100)		1.00	1.00	1.00
Fe		0.64 (7)	0.54 (36)	0.65 (57)		0.50	0.65	0.90		0.50 (3)	0.55 (41)	0.57 (56)		0.56	0.60	0.37
Salt		0.50 (99)	0.50 (1)			0.99	0.99	1.00		1.00 (100)				1.00	1.00	1.00
Mn			1.00 (100)			1.00	1.00	1.00			1.00 (100)			1.00	1.00	1.00
Cu			1.00 (100)			1.00	1.00	1.00			1.00 (100)			1.00	1.00	1.00

### 3.5. Estimating N fertilization Recommendation Directly Using UV-induced Fluorescence

The fluorescence features were used to measure the N rate recommendation of each sample directly as it was calculated using the algorithm in Table 5. Among all possible cases in the algorithms, only the N rates 0, 39, 95, 151, 179, 207, and 235 kg N Ha<sup>-1</sup> were represented. The confusion matrix was thus generated only with those seven classes. The baseline accuracy (BASE) was 0.41 and 0.28 for the training and the test datasets, respectively. Using machine learning to predict the N recommendation rates, the OA was 0.91 and 0.78 for the training and test datasets, respectively (Table 9). This represents an improvement of 0.5 points for both the training and test datasets. The BA yielded similar results, suggesting a good distribution of cases in each class. These results seem to indicate that directly using UV-induced fluorescence measurements to predict the N fertilizer recommendation is a viable approach. There was no significant bias in terms of over- and under-prediction for both the training and the testing datasets. However, in some instances, the error of prediction was large (data not shown). For example, in the test dataset, one sample for which the N recommendation should have been 207 kg N per Ha was estimated at 0 kg N per Ha based on fluorescence measurements. Inversely, one sample for which the N recommendation should have been 39 kg N per Ha was estimated at 207 kg N per Ha. This indicates that despite a good overall accuracy (i.e., OA = 0.78 for the test dataset), important errors may occur, which could cause N loadings in the environment or yield loss if not addressed. Interestingly, there was no example in the literature of studies attempting to estimate fertilizer rates directly using soil optical properties. In the current study, this approach was attempted to make soil fluorescence sensing more practical for decision making. This was motivated by the availability of machine learning analytics enabling to account for a large number of factors for accurate prediction, and by the availability of the decision algorithm for the N rate based on soil properties (i.e., NO<sub>3</sub>-N and SOM). If this approach was to be implemented commercially, a different algorithm would be required for each different crop and possibly for different yield goals. Similar to other conclusions in this study, a larger number of samples would help confirm the potential of the approach, develop more robust algorithms, and provide a better assessment of the bias along with its magnitude. Nevertheless, this study clearly demonstrates the potential of this approach, and further development would inevitably require training the machine learning algorithm on a much larger dataset.

**Table 9.** Accuracy of N rate prediction using fluorescence features. The baseline accuracy (BASE), overall accuracy (OA), and balanced accuracy (BA) were calculated from a multi-class confusion matrix. The percentage of cases when estimated N rate was below or above the actual recommended rate is indicated in the under- and over-estimated columns.

	N	BASE	OA	BA	Under-Estimated	Over-Estimated
Training	100	0.41	0.91	0.91	5%	4%
Test	68	0.28	0.78	0.77	10%	12%

### 3.6. General Discussion

The dataset used in this study enabled the assessment of the potential of UV- and Vis-induced fluorescence to predict soil properties. Despite using a dataset of 168 samples, results showed that more data may be required to (1) train the machine learning algorithms and (2) provide a broader range of values for each soil property. The observations-to-features ratio of the training dataset was 4:1 (i.e., 100 observations to 25 features), which may be considered low as compared to the general rule of thumb of using a 10:1 ratio as a starting point [60]. However, random forest models are usually well suited for classification algorithms of small datasets, and previous studies show that prediction accuracy may not be impacted by the observations-to-features ratio as much as other machine learning approaches [61,62]. Nevertheless, in the case of the current study, despite using an adapted machine learning approach, there may not have been enough variability in the dataset for certain soil properties. For instance, the range of K, Mn, and Cu was all enclosed in one fertility class of Table 4. The results of this study thus do not allow to conclude for those soil properties and suggest that more samples may be needed to test the approach on a broader range of values.

As per the results of this study, the prediction of soil properties using induced fluorescence on whole dried soil samples presents a good potential for a rapid assessment of soil fertility, and possibly for fertilizer recommendations. Such an approach has multiple advantages over a traditional laboratory analysis and may even be transformational in countries where soil analysis laboratories are not available [63]. Yet, certain aspects are important to consider, one of which is the condition of the soil samples. It has been demonstrated that the presence of organic material (e.g., crop residues) in the sample may have a significant impact on the UV- and Vis-induced fluorescence readings [22]. It is thus important that samples be cleaned of such crop residues before the fluorescence readings' acquisition. The moisture content of the sample may also impact the outcome of the readings. The effect of water on optical properties of soil is visible to the naked eye with wet samples appearing darker than dry samples, and this alteration of optical properties may also apply to fluorescence. Ref. [64] found a significant effect of soil moisture on trace element prediction by x-ray fluorescence, and the impact depended on the trace element measured and on the level of moisture content. This indicates that a soil wetness calibration curve may be required for each element independently. The absorption of light by liquid water in the UV to visible spectrum decreases down to about 420 nm and increases again after this minimum [65]. This indicates that UV induction may be more impacted by moisture than the visible light induction. In situ readings would present a significant advantage in terms of efficiency and practicality. However, the presence of crop residues and organic tissues on the surface as well as the variable levels of moisture in the soil can complicate the use of this technology. For instance, rather than a simple reading such as what would be used on the crop canopy, acquiring a reliable soil reading may require a cleaning of the surface from residues and the acquisition of separate soil moisture readings.

## 4. Conclusions

This study aimed at assessing the potential of a multi-parameter fluorescence sensor (Multiplex MX3) to predict soil properties of whole air-dried samples. Results showed that this approach can predict NO<sub>3</sub>-N, SOM, and Zn fertility classes with an overall accuracy of 0.54, 0.78, and 0.69, respectively. Certain fluorescence signals such as the ones using a yellow

filter generated better results which seem to indicate that fluorescence at shorter wavebands may perform better than the longer ones for soil properties' prediction. Moreover, the index NBI\_UVm, which was calculated a posteriori, was the best index for soil fertility classification. This study also demonstrated that induced fluorescence can be used to predict the N rate directly (OA of 0.78), which indicates that this approach can be practical for farmers. Further development of this approach is required to explore its full potential and its expansion to in situ measurements. Notably, a larger training dataset composed of observations covering a broader range of soil properties' values are required to confirm the potential for soil properties beyond NO<sub>3</sub>-N, SOM, and Zn, and to develop the random forest models further. Moreover, the transposition of this approach to in situ readings will require dedicated studies on the effects of organic components in the sensor field of view and on the effect of soil water content on the predictability of each soil property.

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

**Funding:** This research received support from the Colorado State University Agricultural Experiment Station and the Cornell University Agricultural Experiment Station. In-kind support was provided by Force-A, Orsay, France.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** We want to acknowledge the generous contribution of Force-A, Orsay, France for lending the prototype of the fluorescence sensor at no cost for the duration of this study.

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

## References

1. Ferguson, R.B.; Gotway, C.A.; Hergert, G.W.; Peterson, T.A. Soil Sampling for Site-Specific Nitrogen Management. In *Proceedings of the Third International Conference on Precision Agriculture*; American Society of Agronomy, Crop Science Society of America, Soil Science Society of America: Madison, WI, USA, 1996; pp. 13–22. [[CrossRef](#)]
2. Nolin, M.C.; Guertin, S.P.; Wang, C. Within-Field Spatial Variability of Soil Nutrients and Corn Yield in a Montreal Lowlands Clay Soil. In *Proceedings of the Third International Conference on Precision Agriculture*; American Society of Agronomy, Crop Science Society of America, Soil Science Society of America: Madison, WI, USA, 1996; pp. 257–270.
3. Adamchuk, V.I.; Rossel, R.A.V. Development of On-the-Go Proximal Soil Sensor Systems. In *Proximal Soil Sensing. Progress in Soil Science*; Viscarra Rossel, R., McBratney, A., Minasny, B., Eds.; Springer: Dordrecht, The Netherlands, 2010; pp. 15–28. [[CrossRef](#)]
4. Munsell, C. *Munsell Soil Color Charts, Munsell Color*; Macbeth Division of Kollmorgen Corporation: Baltimore, MD, USA, 1950.
5. Adamchuk, V.I.; Hummel, J.W.; Morgan, M.T.; Upadhyaya, S.K. On-the-Go Soil Sensors for Precision Agriculture. *Comput. Electron. Agric.* **2004**, *44*, 71–91. [[CrossRef](#)]
6. Stenberg, B.; Viscarra Rossel, R.A. Diffuse Reflectance Spectroscopy for High-Resolution Soil Sensing. In *Proximal Soil Sensing*; Viscarra Rossel, R.A., McBratney, A.B., Minasny, B., Eds.; Springer Netherlands: Dordrecht, The Netherlands, 2010; pp. 29–47. ISBN 978-90-481-8858-1.
7. Rinnan, R.; Rinnan, Å. Application of near Infrared Reflectance (NIR) and Fluorescence Spectroscopy to Analysis of Microbiological and Chemical Properties of Arctic Soil. *Soil Biol. Biochem.* **2007**, *39*, 1664–1673. [[CrossRef](#)]
8. Zhu, Y.; Weindorf, D.C.; Zhang, W. Characterizing Soils Using a Portable X-ray Fluorescence Spectrometer: 1. Soil Texture. *Geoderma* **2011**, *167*, 167–177. [[CrossRef](#)]
9. Barnes, E.M.; Sudduth, K.A.; Hummel, J.W.; Lesch, S.M.; Corwin, D.L.; Yang, G.; Dougherty, C.S.T.; Bausch, W.C. Remote-and Ground-Based Sensor Techniques to Map Soil Properties. *Photogramm. Eng. Remote Sens.* **2003**, *69*, 619–630. [[CrossRef](#)]
10. Weindorf, D.C.; Zhu, Y.; McDaniel, P.; Valerio, M.; Lynn, L.; Michaelson, G.; Clark, M.; Ping, C.L. Characterizing Soils via Portable X-Ray Fluorescence Spectrometer: 2. Spodic and Albic Horizons. *Geoderma* **2012**, *189*, 268–277. [[CrossRef](#)]
11. Sharma, A.; Weindorf, D.C.; Man, T.; Aldabaa, A.A.A.; Chakraborty, S. Characterizing Soils via Portable X-Ray Fluorescence Spectrometer: 3. Soil Reaction (PH). *Geoderma* **2014**, *232*, 141–147. [[CrossRef](#)]

12. Sharma, A.; Weindorf, D.C.; Wang, D.; Chakraborty, S. Characterizing Soils via Portable X-Ray Fluorescence Spectrometer: 4. Cation Exchange Capacity (CEC). *Geoderma* **2015**, *239*, 130–134. [[CrossRef](#)]
13. Vaudour, E.; Cerovic, Z.G.; Ebengo, D.M.; Latouche, G. Predicting Key Agronomic Soil Properties with UV-Vis Fluorescence Measurements Combined with Vis-NIR-SWIR Reflectance Spectroscopy: A Farm-Scale Study in a Mediterranean Viticultural Agroecosystem. *Sensors* **2018**, *18*, 1157. [[CrossRef](#)]
14. Herman, B.; Frohlich, V.C.; Lakowicz, J.R.; Murphy, D.B.; Spring, K.R.; Davidson, M.W. Basic Concepts in Fluorescence. *Microscopy Resource Center Olympus*. 2003. Available online: <https://www.olympus-lifescience.com/en/microscope-resource/primer/techniques/fluorescence/fluorescenceintro/> (accessed on 17 June 2022).
15. Ma, X.; Green, S.A. Fractionation and Spectroscopic Properties of Fulvic Acid and Its Extract. *Chemosphere* **2008**, *72*, 1425–1434. [[CrossRef](#)]
16. Martins, T.; Saab, S.C.; Milori, D.M.B.P.; Brinatti, A.M.; Rosa, J.A.; Cassaro, F.A.M.; Pires, L.F. Soil Organic Matter Humification under Different Tillage Managements Evaluated by Laser Induced Fluorescence (LIF) and C/N Ratio. *Soil Tillage Res.* **2011**, *111*, 231–235. [[CrossRef](#)]
17. Milori, D.M.B.P.; Galeti, H.V.A.; Martin-Neto, L.; Dieckow, J.; González-Pérez, M.; Bayer, C.; Salton, J. Organic Matter Study of Whole Soil Samples Using Laser-Induced Fluorescence Spectroscopy. *Soil Sci. Soc. Am. J.* **2006**, *70*, 57–63. [[CrossRef](#)]
18. Zhu, Y.; Weindorf, D.C. Determination of Soil Calcium Using Field Portable X-Ray Fluorescence. *Soil Sci.* **2009**, *174*, 151–155. [[CrossRef](#)]
19. Ghozlen, N.B.; Cerovic, Z.G.; Germain, C.; Toutain, S.; Latouche, G. Non-Destructive Optical Monitoring of Grape Maturation by Proximal Sensing. *Sensors* **2010**, *10*, 10040–10068. [[CrossRef](#)]
20. Ben Abdallah, F.; Philippe, W.; Goffart, J.P. Comparison of Optical Indicators for Potato Crop Nitrogen Status Assessment Including Novel Approaches Based on Leaf Fluorescence and Flavonoid Content. *J. Plant Nutr.* **2018**, *41*, 2705–2728. [[CrossRef](#)]
21. Huang, S.; Miao, Y.; Yuan, F.; Cao, Q.; Ye, H.; Lenz-Wiedemann, V.I.S.; Bareth, G. In-Season Diagnosis of Rice Nitrogen Status Using Proximal Fluorescence Canopy Sensor at Different Growth Stages. *Remote Sens.* **2019**, *11*, 1847. [[CrossRef](#)]
22. Longchamps, L.; Khosla, R. Early Detection of Nitrogen Variability in Maize Using Fluorescence. *Agron. J.* **2014**, *106*, 511–518. [[CrossRef](#)]
23. Siqueira, R.; Longchamps, L.; Dahal, S.; Khosla, R. Use of Fluorescence Sensing to Detect Nitrogen and Potassium Variability in Maize. *Remote Sens.* **2020**, *12*, 1752. [[CrossRef](#)]
24. Leufen, G.; Noga, G.; Hunsche, M. Fluorescence Indices for the Proximal Sensing of Powdery Mildew, Nitrogen Supply and Water Deficit in Sugar Beet Leaves. *Agriculture* **2014**, *4*, 58–78. [[CrossRef](#)]
25. Wang, Y.; Zia, S.; Owusu-Adu, S.; Gerhards, R.; Müller, J. Early Detection of Fungal Diseases in Winter Wheat by Multi-Optical Sensors. *APCBEE Procedia* **2014**, *8*, 199–203. [[CrossRef](#)]
26. Noble, E.; Kumar, S.; Görlitz, F.G.; Stain, C.; Dunsby, C.; French, P.M.W. In Vivo Label-Free Mapping of the Effect of a Photosystem II Inhibiting Herbicide in Plants Using Chlorophyll Fluorescence Lifetime. *Plant Methods* **2017**, *13*, 48. [[CrossRef](#)]
27. Nakaya, Y.; Nakashima, S.; Moriizumi, M.; Oguchi, M.; Kashiwagi, S.; Naka, N. Three Dimensional Excitation-Emission Matrix Fluorescence Spectroscopy of Typical Japanese Soil Powders. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2020**, *233*, 118188. [[CrossRef](#)]
28. Senesi, N.; Miano, T.M.; Provenzano, M.R.; Brunetti, G. Characterization, Differentiation, and Classification of Humic Substances by Fluorescence Spectroscopy. *Soil Sci.* **1991**, *152*, 259–271. [[CrossRef](#)]
29. Miano, T.M.; Senesi, N. Synchronous Excitation Fluorescence Spectroscopy Applied to Soil Humic Substances Chemistry. *Sci. Total Environ.* **1992**, *117–118*, 41–51. [[CrossRef](#)]
30. Del Vecchio, R.; Blough, N.V. On the Origin of the Optical Properties of Humic Substances. *Environ. Sci. Technol.* **2004**, *38*, 3885–3891. [[CrossRef](#)] [[PubMed](#)]
31. Fuentes, M.; González-Gaitano, G.; García-Mina, J.M. The Usefulness of UV-Visible and Fluorescence Spectroscopies to Study the Chemical Nature of Humic Substances from Soils and Composts. *Org. Geochem.* **2006**, *37*, 1949–1959. [[CrossRef](#)]
32. Magdoff, F.; Tabatabai, M.A.; Hanlon, E.A. Soil Organic Matter: Analysis and Interpretation. In Proceedings of the Symposium Sponsored by Divisions S-4 and S-8 of the Soil Science Society of America in Seattle, Washington, DC, USA, 14 November 1994.
33. Keeney, D.R.; Nelson, D.W. Nitrogen—Inorganic Forms. In *Methods of soil analysis. Part 2*, 2nd ed.; Page, A.L., Miller, R.H., Keeney, D.R., Eds.; American Society of Agronomy and Soil Science Society of America Publisher: Madison, WI, USA, 1983; pp. 643–698.
34. Mehlich, A. Mehlich 3 Soil Test Extractant: A Modification of Mehlich 2 Extractant. *Commun. Soil Sci. Plant Anal.* **1984**, *15*, 1409–1416. [[CrossRef](#)]
35. Warncke, D.; Brown, J.R. Potassium and Other Basic Cations. In *Recommended Chemical Soil Test Procedures for the North Central Region*; Brown, J.R., Ed.; NCR Publication No. 221; Missouri Agricultural Experiment Station: Columbia, MO, USA, 1998; pp. 31–33.
36. Lindsay, W.L.; Norvell, W.A. Development of a DTPA Soil Test for Zinc, Iron, Manganese, and Copper. *Soil Sci. Soc. Am. J.* **1978**, *42*, 421–428. [[CrossRef](#)]
37. Gee, G.W.; Bauder, J.W. Particle-Size Analysis. In *Methods of Soil Analysis. Part 1. Physical and Mineralogical Methods*; Klute, A., Ed.; Soil Science Society of America: Madison, WI, USA, 1986; pp. 383–411.



38. United States Department of Agriculture Soil Conservation Service; Colorado Agricultural Experiment Station. *Soil Survey Staff Soil Survey of Logan County, Colorado*; USDA Cooperative Soil Survey U.S. Government Printing Office: Washington, DC, USA, 1977.
39. United States Department of Agriculture Soil Conservation Service; Colorado Agricultural Experiment Station. *Soil Survey Staff Soil Survey of Larimer County, Colorado*; USDA Cooperative Soil Survey U.S. Government Printing Office: Washington, DC, USA, 1980.
40. Cerovic, Z.G.; Goutouly, J.P.; Hilbert, G.; Destrac-Irvine, A.; Martinon, V.; Moise, N. Mapping Winegrape Quality Attributes Using Portable Fluorescence-Based Sensors. In *FRUTIC 09*; Best, S., Ed.; Concepcion, Chile; Propag INIA: Chilian, Chile, 2009; pp. 301–310.
41. R Development Core Team. *R: A Language and Environment for Statistical Computing*; Version R 3.3.0; The R Foundation for Statistical Computing: Vienna, Austria, 2016.
42. Grömping, U. Variable Importance Assessment in Regression: Linear Regression versus Random Forest. *Am. Stat.* **2009**, *63*, 308–319. [[CrossRef](#)]
43. Liu, X.; Guanter, L.; Liu, L.; Damm, A.; Malenovsky, Z.; Rascher, U.; Peng, D.; Du, S.; Gastellu-Etchegorry, J.-P. Downscaling of Solar-Induced Chlorophyll Fluorescence from Canopy Level to Photosystem Level Using a Random Forest Model. *Remote Sens. Environ.* **2019**, *231*, 110772. [[CrossRef](#)]
44. Peng, J.; Manevski, K.; Kørup, K.; Larsen, R.; Andersen, M.N. Random Forest Regression Results in Accurate Assessment of Potato Nitrogen Status Based on Multispectral Data from Different Platforms and the Critical Concentration Approach. *Field Crop. Res.* **2021**, *268*, 108158. [[CrossRef](#)]
45. Zhang, Y.; Sui, B.; Shen, H.; Ouyang, L. Mapping Stocks of Soil Total Nitrogen Using Remote Sensing Data: A Comparison of Random Forest Models with Different Predictors. *Comput. Electron. Agric.* **2019**, *160*, 23–30. [[CrossRef](#)]
46. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
47. Liaw, A.; Wiener, M. Classification and Regression by RandomForest. *R News* **2002**, *2*, 18–22.
48. Davis, J.G.; Westfall, D.G. Fertilizing Corn-0.538. Available online: <https://extension.colostate.edu/topic-areas/agriculture/fertilizing-corn-0-538/> (accessed on 13 December 2021).
49. Bauder, T.A.; Waskom, R.M.T.; Schneekloth, J.P.; Alldredge, J. *Best Management Practices for Colorado Corn*; Colorado State University Cooperative Extension: Fort Collins, CO, USA, 2003.
50. Self, J.R. Soil Test Explanation—0.502. 2010. Available online: <https://extension.colostate.edu/topic-areas/agriculture/soil-test-explanation-0-502/> (accessed on 20 December 2021).
51. Ripley, B.D. *R Package Tree, Version 1.0–2*; The R Foundation for Statistical Computing: Vienna, Austria, 2006; Available online: <http://CRAN.R-project.org/package=tree> (accessed on 13 December 2021).
52. Grandini, M.; Bagli, E.; Visani, G. Metrics for Multi-Class Classification: An Overview. *arXiv* **2020**, arXiv:2008.05756.
53. Davis, J.G.; Westfall, D.G. *Fertilizing Corn*; Colorado State University Extension: Fort Collins, CO, USA, 2009.
54. Velasco, M.I.; Campitelli, P.A.; Ceppi, S.B.; Havel, J. Analysis of Humic Acid from Compost of Urban Wastes and Soil by Fluorescence Spectroscopy. *Agriscientia* **2004**, *21*, 31–38.
55. Nicolodelli, G.; Tadini, A.M.; Nogueira, M.S.; Pratavieira, S.; Mounier, S.; Huaman, J.L.C.; dos Santos, C.H.; Montes, C.R.; Milori, D.M.B.P. Fluorescence Lifetime Evaluation of Whole Soils from the Amazon Rainforest. *Appl. Opt. AO* **2017**, *56*, 6936–6941. [[CrossRef](#)]
56. Ammari, F.; Bendoula, R.; Bouveresse, D.J.-R.; Rutledge, D.N.; Roger, J.-M. 3D Front Face Solid-Phase Fluorescence Spectroscopy Combined with Independent Components Analysis to Characterize Organic Matter in Model Soils. *Talanta* **2014**, *125*, 146–152. [[CrossRef](#)]
57. González-Pérez, M.; Milori, D.M.B.P.; Colnago, L.A.; Martin-Neto, L.; Melo, W.J. A Laser-Induced Fluorescence Spectroscopic Study of Organic Matter in a Brazilian Oxisol under Different Tillage Systems. *Geoderma* **2007**, *138*, 20–24. [[CrossRef](#)]
58. Ross, D.S.; Ketterings, Q. Recommended Methods for Determining Soil Cation Exchange Capacity. *Recomm. Soil Test. Proced. Northeast. U.S.* **2011**, *493*, 62.
59. ML, J.T.; Chambers, O. Machine Learning Strategy for Soil Nutrients Prediction Using Spectroscopic Method. *Sensors* **2021**, *21*, 4208. [[CrossRef](#)]
60. Feeny, A.K.; Chung, M.K.; Madabhushi, A.; Attia, Z.I.; Cikes, M.; Firouznia, M.; Friedman, P.A.; Kalscheur, M.M.; Kapa, S.; Narayan, S.M.; et al. Artificial Intelligence and Machine Learning in Arrhythmias and Cardiac Electrophysiology. *Circ. Arrhythmia Electrophysiol.* **2020**, *13*, e007952. [[CrossRef](#)]
61. Merali, Z.G.; Witiw, C.D.; Badhiwala, J.H.; Wilson, J.R.; Fehlings, M.G. Using a Machine Learning Approach to Predict Outcome after Surgery for Degenerative Cervical Myelopathy. *PLoS ONE* **2019**, *14*, e0215133. [[CrossRef](#)]
62. Pham, H.T. Generalized Weighting for Bagged Ensembles. Ph.D. Thesis, Iowa State University, Ames, IA, USA, 2018.
63. Andrade, R.; Faria, W.M.; Silva, S.H.G.; Chakraborty, S.; Weindorf, D.C.; Mesquita, L.F.; Guilherme, L.R.G.; Curi, N. Prediction of Soil Fertility via Portable X-Ray Fluorescence (PXRF) Spectrometry and Soil Texture in the Brazilian Coastal Plains. *Geoderma* **2020**, *357*, 113960. [[CrossRef](#)]
64. Sahraoui, H.; Hachicha, M. Effect of Soil Moisture on Trace Elements Concentrations Using Portable X-Ray Fluorescence Spectrometer. *J. Fundam. Appl. Sci.* **2017**, *9*, 468–484. [[CrossRef](#)]
65. Fry, E.S. Visible and Near-Ultraviolet Absorption Spectrum of Liquid Water: Comment. *Appl. Opt. AO* **2000**, *39*, 2743–2744. [[CrossRef](#)] [[PubMed](#)]