



Article

# Multispectral Cameras and Machine Learning Integrated into Portable Devices as Clay Prediction Technology

Gilson Augusto Helfer<sup>1,2,\*</sup>, Jorge Luis Victória Barbosa<sup>1</sup>, Douglas Alves<sup>2</sup>, Adilson Ben da Costa<sup>3</sup>, Marko Beko<sup>4,5</sup> and Valderi Reis Quietinho Leithardt<sup>5,6</sup>

<sup>1</sup> Applied Computing Graduate Program, University of Vale do Rio dos Sinos, Av. Unisinos 950, São Leopoldo 93022-750, RS, Brazil; jbarbosa@unisinos.br

<sup>2</sup> Department of Engineering, Architecture and Computing, University of Santa Cruz do Sul, Av. Independencia 2293, Santa Cruz do Sul 96815-900, RS, Brazil; alves\_543@hotmail.com

<sup>3</sup> Industrial Systems and Processes Graduate Program, University of Santa Cruz do Sul, Av. Independencia 2293, Santa Cruz do Sul 96815-900, RS, Brazil; adilson@unisc.br

<sup>4</sup> Instituto de Telecomunicações, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisbon, Portugal; marko.beko@tecnico.ulisboa.pt

<sup>5</sup> COPELABS, University Lusófona—ULHT, 1749-024 Lisbon, Portugal; valderi@ippportalegre.pt

<sup>6</sup> VALORIZA—Research Centre for Endogenous Resource Valorization, Polytechnic Institute of Portalegre, 7300-555 Portalegre, Portugal

\* Correspondence: ghelfer@gmail.com



**Citation:** Helfer, G.A.; Barbosa, J.L.V.; Alves, D.; da Costa, A.B.; Beko, M.; Leithardt, V.R.Q. Multispectral Cameras and Machine Learning Integrated into Portable Devices as Clay Prediction Technology. *J. Sens. Actuator Netw.* **2021**, *10*, 40. <https://doi.org/10.3390/jsan10030040>

Academic Editors: Jameela Al-Jaroodi, Sanja Lazarova-Molnar and Nader Mohamed

Received: 25 May 2021  
Accepted: 23 June 2021  
Published: 25 June 2021

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



**Copyright:** © 2021 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/>).

**Abstract:** The present work proposed a low-cost portable device as an enabling technology for agriculture using multispectral imaging and machine learning in soil texture. Clay is an important factor for the verification and monitoring of soil use due to its fast reaction to chemical and surface changes. The system developed uses the analysis of reflectance in wavebands for clay prediction. The selection of each wavelength is performed through an LED lamp panel. A NoIR microcamera controlled by a Raspberry Pi device is employed to acquire the image and unfold it in RGB histograms. Results showed a good prediction performance with  $R^2$  of 0.96, RMSEC of 3.66% and RMSECV of 16.87%. The high portability allows the equipment to be used in a field providing strategic information related to soil sciences.

**Keywords:** machine learning; multispectral image; soil; clay; agriculture

## 1. Introduction

Smart farming represents the use of information and communication technology systems applied in agriculture with the objective of obtaining better results, greater performance and higher quality production with safety and precision while optimizing human work [1,2].

From these new technologies, a cultivation area can be divided into as many plots as it has internal differences supported by soil analysis and each plot can receive a customized treatment to obtain the maximum benefit from it. This is also known as precision agriculture [3–5].

However, it is necessary to characterize the variability of the chemical and physical attributes of the soil through a representative sample of such variations. So, soil analysis is the only method that allows, before planting, to recommend adequate amounts of correctives and fertilizers to increase crop productivity and, as a consequence, crop production and profitability. Soil science is considered a strategic research topic for precision agriculture and smart farms [6,7].

The clay content defines the texture of the soil. It interferes with several factors including plant growth and productivity, water infiltration into the soil and its storage, retention and transport, availability and absorption of plant nutrients, living organisms, soil quality, productivity and temperature, levels of structure and compaction, soil preparation,

irrigation and fertilizer efficiency. Therefore, the clay content plays a fundamental role in crop productivity [8].

The traditional way of collecting soil in the fields and analyzing it in the laboratory is the most accurate, but it takes time and uses an alkaline solution that needs to be neutralized before wasting [9]. New research has been proposed to optimize this, but with limitations. Satellite images are important for obtaining quick information on the surface of soils in large areas. However, mapping large areas of soil presents difficulties as most areas are usually covered by vegetation [10].

The use of spectral images expands the capacity of studies in several areas and their application has been growing in agriculture in order to recognize patterns [11–13]. For these types of analyses, two well-known scientific methodologies are used: Spectroscopy and imaging. Optical spectroscopy is a term used to describe the phenomena involving a spectrum of light intensities at different wavelengths. Imaging can be conceptualized as the science of image acquisition of the spatial shape of objects. Currently, the most advanced way to capture images is digitally [14].

Multispectral and hyperspectral imaging systems are image analysis techniques that are also based on capturing the same image at different wavelengths. The difference consists in the number of captured wavelengths: While multispectral systems use up to 10, hyperspectral systems can exceed 100 wavelengths, with the latter generating larger amounts of data [15].

Since single-board computers have become more accessible to the general public, the Raspberry Pi has become one of the more popular systems, mainly in the scientific community, promoting research in IoT and all features involved [16,17]. Leithardt et al. [18] and Felipe Viel et al. [19] developed works that exemplified the application of the Raspberry Pi in IoT.

In addition, machine learning tools as Partial Least Regression (PLSR) have been applied for multivariate calibration in soil spectroscopy [20], images [21] and sensor data [22]. These algorithms eliminate variables that do not correlate with the property of interest, such as those that add noise, non-linearities or irrelevant information [23].

Considering the importance of research in areas involving soils (agriculture, geochemistry, geology), the ability to use devices such as the Raspberry Pi and the use of computer vision techniques such as spectral imaging, the following research problem was defined: “Is it possible to use multispectral imaging techniques to predict clays?”.

The main objective is to develop a computer vision system to predict the amount of clay in the soil using multispectral imaging techniques on a Raspberry Pi device. The relevance of this work is in the absence of a fast, mobile, cheap and non-destructive method to measure clay content in soil.

This article is structured in six sections. Section 2 presents an approach to the soil texture and colors, multispectral images, machine learning and OpenCV libraries. Section 3 describes the related works, while Section 4 presents materials and methods employed. Section 5 shows the results of the implementation and its discussions. Finally Section 6 is intended for conclusions and future works.

## 2. Background

The process of building a clay prediction system based on multispectral images covers several areas of knowledge such as optics, soil science, computer vision and artificial intelligence.

### 2.1. Optics

The extraction of characteristics from the objects can be performed through the reflected energy, depending on some factors such as the positioning of the object, the composition of the material, the roughness of the material and the type of surface that this material displays at the time of capture. Each material can have a specific spectral behavior where characteristics such as humidity, deterioration and decomposition are agents that influence

the performance of its identification. Thus, the amount of bands in the spectrum required for the identification of a given material depends on the amount of material discriminated and also on its variations [24].

Light is a special band of electromagnetic radiation within the spectrum that can be perceived by the human eye, this band is divided into six regions, which are violet, blue, green, yellow, orange and red; the perception of these colors is determined by light reflected by an object. For example, green colored objects mainly reflect wavelengths between 500 and 570 nm (green color in the electromagnetic spectrum), and absorb most other wavelengths [25].

A spectral imaging system is, essentially, composed of four components: Lighting, focus lens, a detector and a wavelength selection system. The first spectral imaging systems were designed to filter an object's light and use a monochrome digital camera to record the reflected light. More modern systems illuminated the sampling object with a monochromatic light [26]. In recent years, LED lamps have been adopted because they present the advantage of less variation in brightness when compared to ordinary white lamps [27].

## 2.2. Soil Science

Smart farms employ different technologies and sciences, among which soil science stands out. Soil science strategies allowed implementation of technologies related to precision agriculture that can enable several smart services on a farm [28].

Soil classification is important for correct soil management in a sustainable manner; in view of the different physico-chemical compositions of each type of soil, color is an indicator of chemical composition. The Munsell table is the most common method for this type of classification. It consists of a simple device for comparing the color of the soil divided into three color patterns, which are:

1. Hue: This is usually red or yellow;
2. Value: This is light or dark; the darker, the closer the value is zero;
3. Chroma: This corresponds to the brightness, with zero corresponding to gray [29].

The color of the soil is directly influenced by three factors: Organic matter, water concentration and the oxidation state of iron and manganese oxides. Since soils with higher water contents are darker than when dry, water also interferes with the amount of microorganisms present, which also makes the soil darker. Oxidation in high quantity leaves the soil more grayish or bluish; otherwise, it will be more reddish [30].

So, soil color is one of the most useful characteristics for soil evaluation, informing about redox, aeration, organic material and soil fertility. Some colors and characteristics of good quality soils are:

1. Superficial dark brown: This offers a wealth of organic matter, good aggregation and a good amount of nutrients;
2. Light yellow and red in the subsoil: This indicates high concentrations of iron oxide and good drainage; iron oxides also contribute to the aggregation of the soil, containing air and water for root development.

Some colors and characteristics of poor quality soils are:

1. Spotted or stained with opaque yellow and orange, bluish gray or olive green: This indicates permanent flooding of the soil and lack of oxygenation and aeration of the soil;
2. Rusted colors (ferrihydrite): Indicates constant flooding;
3. Whitish and pale colors: This indicates the presence of a water layer above the clay [31].

## 2.3. Computer Vision

The RGB model is the most famous color model; as the name suggests, the primary colors are red (red), green (green) and blue (blue). Any shade of color can be obtained

by mixing different amounts of primary colors, this color system can be represented by three perpendicular axes, such that each axis represents a color, resulting in a cube where the entire spectrum of colors meets inside. Each point inside the cube describes the primary color components that result in a certain color; each point has values defined in a triple, each reference a primary RGB color, where the values range from 0 to 255 [32].

A histogram is a graph that describes the scale of luminosity values that a camera can register; ranging from pure black to pure white, the scale itself varies between 0 and 255, respectively. The image histogram is also known as frequency distribution, it is the graphical representation that describes the gray levels of the pixels present in the scene. It is determined by simply counting the number of times that a given level of gray occurs in the image. Histograms are constructed band by band, separately; each band has a unique histogram. The histogram only specifies the number of pixels in each level of gray, without informing the spatial distribution between pixels; another important aspect is that the histogram can be interpreted as a distribution of the probability of occurrence of a certain level of gray [33].

OpenCV is a widely used computer vision and machine learning library and provides algorithms for image processing, resource detection, object detection and video analysis. In addition, it is open source and built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in commercial products. It has more than 2500 optimized algorithms, which include a comprehensive set of classic and state-of-the-art machine vision and machine learning algorithms [34].

#### 2.4. Machine Learning

Partial Least Squares Regression (PLSR) is a machine learning algorithm method for building predictive models when there are several factors and these are highly collinear. The algorithm's emphasis is on predicting responses and not necessarily finding understanding of relationships between variables. For example, PLS is generally not suitable for tracking factors that have an insignificant effect on the response. PLS has as its central idea the extraction of latent factors that represent both the variation of the factors, while modeling the responses. The general objective is to use fewer factors to predict responses in the population [35].

The number of factors or latent variables (LVs) in the model is chosen based on the value of Root Mean Squares Errors of Cross Validation (RMSECV). When working with a larger set of samples, validation can be performed by continuous blocks or random subsets in which a larger number of samples is taken and the model is built with the remaining ones, estimating their concentrations. The prediction errors are averaged for each number of LVs and the one with the lowest error will be the number of LVs in the model.

Other figures of merit also used to evaluate a prediction model are:

1. Linearity, defined by the coefficient of determination ( $R^2$ ), is aware of the model's ability to provide results directly proportional to the amount of analyte present in the sample, as shown in the Equation (1) [36].

$$R^2 = \left( \frac{n(\sum(x * y)) - (\sum x) * (\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2] * [n \sum y^2 - (\sum y)^2]}} \right)^2 \quad (1)$$

2. Veracity: This is the degree of accuracy between the reference values and the predicted values. In the case of multivariate analysis the Root Means Squares Errors of Calibration (RMSEC) is used, as shown in the Equation (2) [36]. The same equation is applied to evaluate RMSECV.

$$RMSEC = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (2)$$

The Kennard–Stone algorithm is a uniform mapping algorithm that selects samples that best represent the training set. To ensure uniform distribution of the subset of samples represented by the data space based on the instrumental response, the Kennard–Stone algorithm segregates samples in regions of the distant space of the samples already selected. For this purpose, the algorithm uses the Euclidean distance [37].

### 3. Related Work

There are several works in the field of multispectral imaging in the most diverse areas of application. The selected articles’ subjects are related to multispectral data, smart farm and soil prediction.

In 2014, Svensgaard et al. built a mobile and closed multispectral imaging system to estimate crop physiology in field experiments. This system shuts out wind and sunlight to ensure the highest possible precision and accuracy. Multispectral images were acquired in an experiment with four different wheat varieties and two different nitrogen levels, replicated on two different soil types at different dates. The results showed potentials, especially at the early growth stages [38].

In 2015, Hassan-Esfahan developed an Artificial Neural Network (ANN) model to quantify the effectiveness of using satellite spectral images to estimate surface soil moisture. The model produces acceptable estimations of moisture results by combining field measurements with inexpensive and readily available remotely sensed inputs [39].

Treboux and Genoud presented in 2018 the usage of decision tree methodology on segregation of vineyard and agricultural objects using hyperspectral images from a drone. This technique demonstrates that results can be improved to obtain 94.27% of accuracy and opens new perspectives for the future of high precision agriculture [40].

Žížala et al. performed in 2019 an evaluation of the prediction ability of models assessing soil organic carbon (SOC) using real multispectral remote sensing data from different platforms in South Moravia (Czechia). The adopted methods included field sampling and predictive modeling using satellite data. Random forest, support vector machine, and the cubist regression technique were applied in the predictive modeling. The obtained results show similar prediction accuracy for all spaceborne sensors, but some limitations occurred in multispectral data [41].

Lopez-Ruiz et al. presented in 2017 the development of a low-cost system for general purposes that was tested by classifying fruits (putrefied and overripe), plastic materials and determining water characteristics. This work shows the development of a general-purpose portable system for object identification using Raspberry Pi and multispectral imaging, which served as the basis for the present study [42].

Table 1 compares articles regarding the application, data analysis, type of sensors and properties, including the proposed work. The following criteria allowed comparison of the proposed model and the aforementioned studies identifying relevant characteristics for evaluation:

1. Sensors: This shows the sensors used in the related studies;
2. Analysis: This identifies which tool is used for data analysis. In other words, it identifies how results were generated for decision making or information to users;
3. Spectral range: This informs the type of spectral image employed (multispectral or hyperspectral);
4. Application: This describes the object, material or scenery analysed.

**Table 1.** Comparison of related works.

Criterion	[38]	[39]	[40]	[41]	[42]	This Work
Sensors	Camera	Satellite	Camera	Satellite	Camera	Camera
Analysis	MMA	ANN	DT	SVM	N/A	ML
Spectral range	Multi	Multi	Hyper	Multi	Multi	Multi
Application	Wheat	Soil	Vineyard	Soil	Fruit	Soil

No study directly related to clay prediction in precision agriculture was found in the literature based on multispectral analysis using LED lamps. The research that presented the multispectral data for decision making mostly made use of resources from satellite images or non-portable solutions, differently to as suggested in the present article.

#### 4. Materials and Methods

A total of 50 soil samples were selected from different collection points of the Vale do Rio Pardo/RS, Brazil, where the clay concentrations ranged from 4% to 72%. These samples were supplied by the Central Analitica soil laboratory (Santa Cruz do Sul, Brazil), where they were dried in a MA037 oven (Marconi, Piracicaba, Brazil) with air circulation for a period of at least 24 h at a temperature between 45 and 60 °C. Afterwards, the samples were ground in a NI040 hammer mill (Marconi, Piracicaba, Brazil), with 2 mm strainer, and stored in cardboard boxes.

This work proposes a system that predicts the amount of clay contained in soil samples using a panel of LED lamps of various colors. The lamps were arranged around a Raspberry Pi NoIR microcamera with OV5647 sensor (OmniVision Technologies). The OV5647 is a low voltage, high performance, 5 megapixel CMOS image sensor that provides up to 2592 × 1944 video output and multiple resolution raw images via the control of the serial camera control bus or MIPI interface. The sensor has an image array capable of operating up to 15 fps in high resolution with user control of image quality. The camera is connected to the BCM2835/BCM2836 processor on the Pi via the CSI bus, a higher bandwidth link that carries pixel data from the camera back to the processor. This bus travels along the ribbon cable that attaches the camera board to the Pi. So, the OV5647 sensor core generates streaming pixel data at a constant frame rate [43].

The microcamera (Pi NoIR v1.3) was coupled to a Raspberry Pi 3 Model B computer that processes the captured images. This device was launched in February 2016 and it uses a 1.2 GHz 64-bit quad-core Arm Cortex-A53 CPU, has 1GB RAM, integrated with 40 extended GPIO pins and CSI camera port for connecting a Raspberry Pi camera [44]. The analysis consisted in capturing images of the soil samples and each captured image is a result of the exposure of the sample to a certain color emitted by a specific LED.

The system used light by means of multispectral spectroscopy, which analyzes light as a set of waves, using bands of the electromagnetic spectrum between 460 to 630 nm (nanometers), which corresponds to the range of visible light and which corresponds to the set of LEDs, according to Table 2.

**Table 2.** LED set used.

LED	Wavelength (nm)	Size (mm)	Voltage (V)
White	500–620	5	3.0–3.2
Yellow	580–590	5	2.8–3.1
Red	620–630	5	2.8–3.1
Green	570–573	5	3.0–3.4
Blue	460–470	5	3.0–3.4

The use of LEDs of various colors allows the analysis of the object reflectance in various bands of the electromagnetic spectrum, which is captured by the NoIR camera—selected due to its low-cost and reduced dimensions. When compared to conventional lamps, the use of LED lamps results in a reduced variation in brightness on the object as well as in the consumption [45].

The processing of the reflected spectra captured by the microcamera is performed by the Raspberry Pi, which has the capacity to perform this type of task, presenting small dimensions and low-energy consumption and offering low-costs. The entire system was arranged inside a black box to avoid the disturbance caused by natural light. The techniques applied in the processing of the selected images were:

1. The generation of histograms of the image in each light spectrum;

2. The use of the histograms in a machine learning training algorithm;
3. All the results obtained in already existing methods will present the results.

Figure 1 shows the general functioning of the system. The soil samples were placed in front of the LED panel and the camera is centered between a set of LED lamps of five different colors (blue, green, red, yellow and white). When starting the program, it captures images of the same sample on each color of light emitted by the panel.

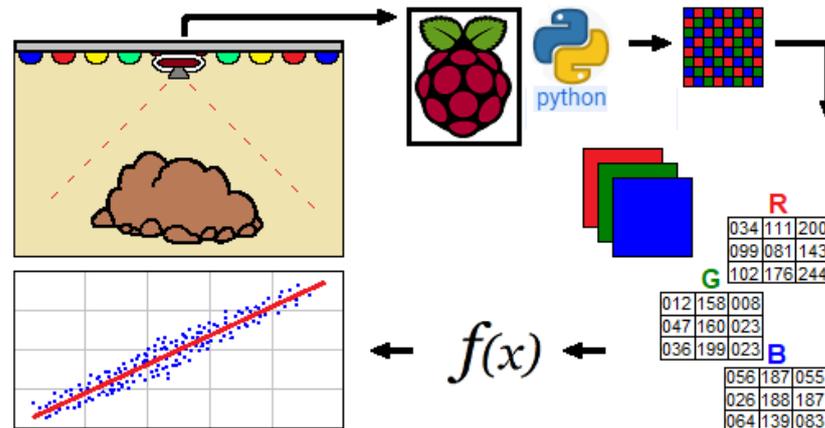


Figure 1. Representation diagram of the instrument.

The initial development of the work was carried out through the construction of the imaging capture structure, which was composed by the LED lights and the NoIR camera inside a box that was duly painted black in order to provide minimal interference in the reflection of the light emitted from the panel. The welding of 5-volt resistors with 330 ohms of resistance was also performed between the LEDs and the GPIO of the Raspberry Pi to prevent the lamps from burning.

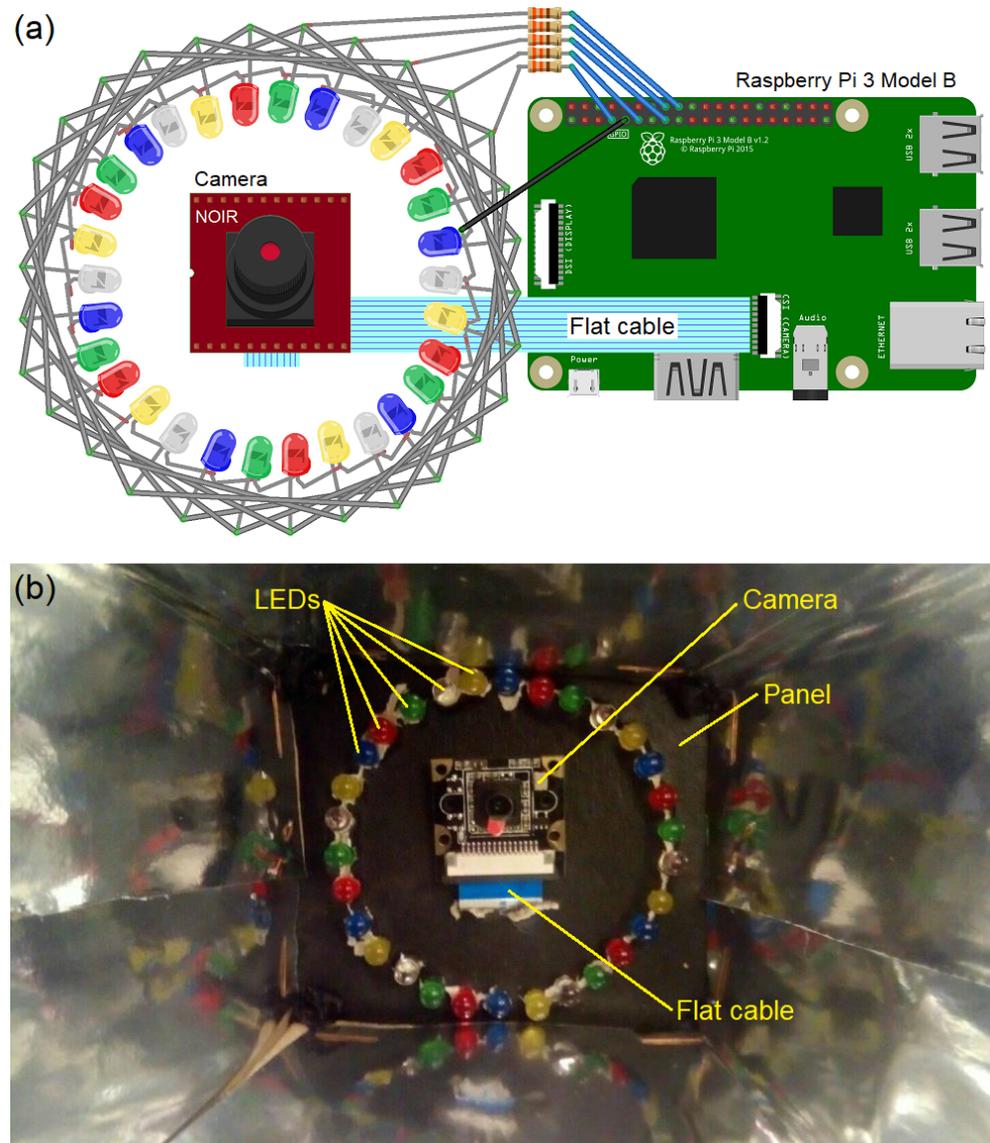
Therefore, the assembly becomes simpler, not requiring the multiplexing of the LED lamps. Figure 2a presents the resulting structure. Regarding the arrangement of the lamps and camera; they were organized in such a way that the incident light was as linear as possible around the camera. So, 30 lamps were installed, six in each wavelength, resulting in a total set of five different wavelengths for analysis.

After welding and assembling the hardware components, tests were carried out to verify the existence of shaded parts on the samples which would affect the performance of the application. Still regarding the cause of shading on the sample, parameters such as disposition and quantity of employed LEDs were essential factors that caused this effect on the images.

Thus, two approaches to solving the problem were possible, the first being an increase in the number of used lamps and the second being the approximation of the lamps, which would result in the modification of its arrangement without changing the quantity.

The first approach—the increase in the number of lamps—was discarded due to the limited availability of resistors and space for the installation of the lamps, as well as the Raspberry Pi GPIO’s ability to support the number of lamps. Therefore, the defined solution was the modification of the arrangement of the LEDs that were already being used, resulting in a distribution in a circular shape.

After changing the arrangement of the lamps there was an improvement in shading; however, this did not fully resolve the issue, so a light deflector was developed in order to avoid the dispersion of the light beam, significantly improving the linearity of the light reflected in the sample, as presented in Figure 2b.



**Figure 2.** LED system and micro-camera disposition scheme (a) and photograph (b).

The application software was developed using Python 2.7 programming language with GPIO libraries to manipulate the Raspberry pins, with the CV2 library referring to the OpenCV to process the images and the PiCamera library to manipulate the camera as well as to capture the images. This software in Python works in the shell terminal of Raspbian GNU/Linux 10. Figure 3a presents an image that was the result of capturing the soil sample in each beam of light.

OpenCV is a library that can be applied to computer vision and machine learning, offering computational power for admission, detection and image processing. Regarding computer vision, it covers the extraction, manipulation and analysis of images in order to obtain useful information from them to perform a specific task [34].

The software structure consists of a main class responsible for defining the area of interest before capturing and defining the activation sequence of the LEDs, in addition to the acquisition of the images, one of each color. Finally, the image is cropped in order to follow the area of interest used, in this case  $128 \times 128$  pixels, as shown in Figure 3b.

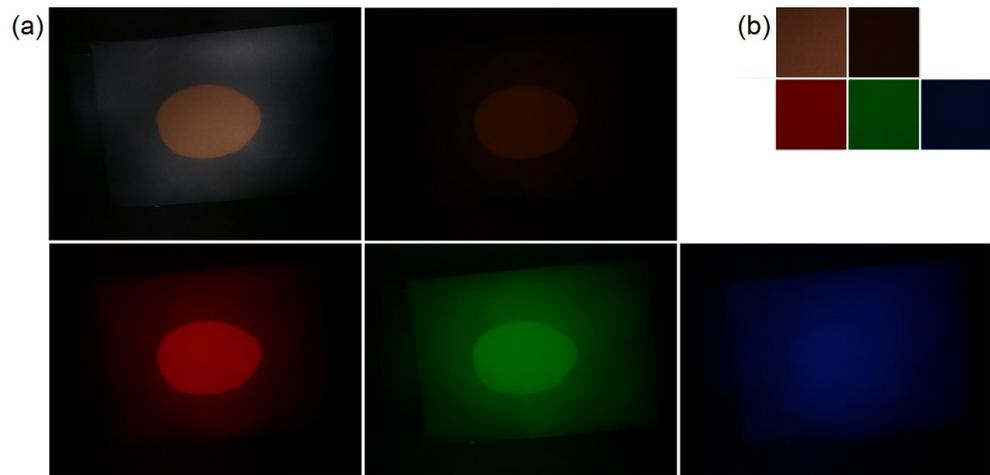


Figure 3. Image of the soil sample captured (a) and processing crop (b).

We employed the Scikit Learn module with machine learning libraries, more specifically the Partial Least Squares Regression (PLSR) technique. For the correct use of the PLSR method there is a need for a linear data profile, which does not occur with luminance values. Thus, only the histograms of the images were used, as follows:

1. Extraction of the image under the effect of a certain LED color;
2. The image is divided into three histograms;
3. The histograms are concatenated, as well as each of the LED colors.
4. As a result, a CSV file (Comma Separated Values) is generated with all histograms in all LED colors, as illustrated in Figure 4.

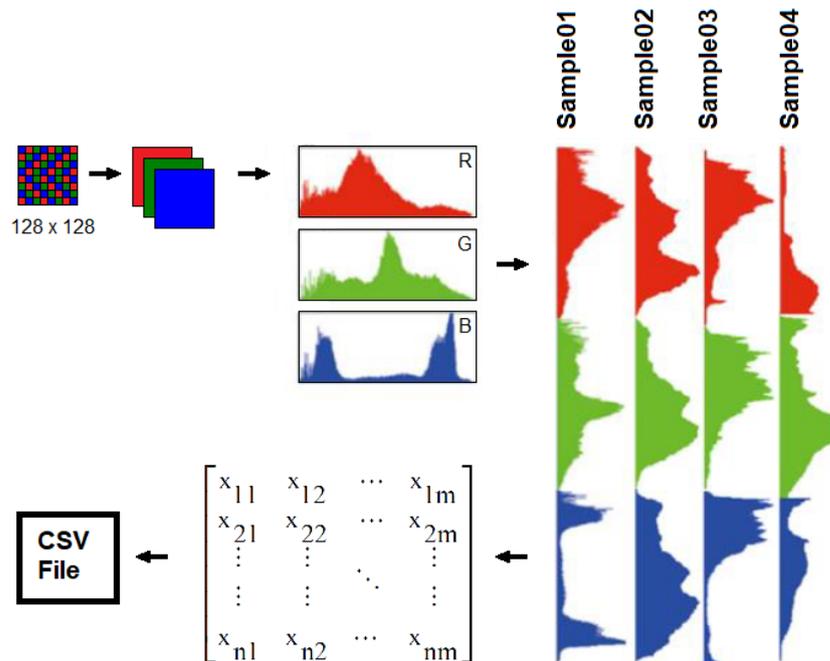


Figure 4. Processing to generate CSV file from matrix of histograms.

Algorithm 1 shows the procedures developed for the acquisition and processing of images and the prediction of clay results.

**Algorithm 1** Procedure of predicting clay through histogram images.**Input:** Image parameters (ROI) and number of PLSR factors**Output:** Prediction charts and reporting data

```

1: roi = 128
2: factors = 6
3: leds = ['green','red','white','yellow','blue']
4: histograms = [ ]
5: for led in leds do
6:   files = acquireImages(led)
7:   for file in files do
8:     histograms.add(processingHistograms(file, roi))
9:   end for
10: end for
11: csv = generateCSV(histograms)
12: ref = loadReferences()
13: predictionModel = computePLSR(csv, ref, factors)
14: reportingData(predictionModel)
15: plottingData(predictionModel)

```

The first step consists in acquiring images controlling all LEDs individually using Raspberry Pi GPIOs. All images are exported in PNG format. The second step creates the red, green and blue histograms from the images obtained in the previous step. The third step transforms the histograms created in the CSV file, joining all 256 color levels of each histogram forming 768 variables per sample. The fourth step computes the prediction model using the PLSR algorithm using the CSV data to correlate with reference data using a specific number of factors. After the model generation reports with predictions and their charts are showed.

All PLSR models were developed based on Daniel Pelliccia's website, which provides a step by step tutorial on how to build a calibration model using partial least squares regression in Python [46].

## 5. Results and Discussions

The calibration models for all sets of histograms were embedded in a Raspberry Pi device. For later comparison of the performance of the models that presented better results, quantifications were made from the figures of merits, both to validate them based on linearity ( $R^2$ ) and on the Root Mean Square Error of Calibration (RMSEC), as well as to evaluate them as calculated by the Root Mean Square Error of Cross Validation (RMSECV).

At first, the addition of three histograms for each LED separately allowed to build the calibration model, thus originating 768 variables. Each model employed a different number of factors according to the best result of RMSECV, as shown in Table 3.

**Table 3.** Results obtained comparing all training models by LED.

LED	Green	Red	White	Yellow	Blue
Variables	768	768	768	768	768
Factors	6	10	8	9	6
$R^2$	0.82	0.607	0.857	0.839	0.806
RMSEC (%)	7.93	11.74	7.06	7.51	8.24
RMSECV (%)	19.36	23.89	13.66	13.59	26.35

The white LED generated the best model with the highest  $R^2$ , estimated at 0.857, and the lowest RMSEC (7.06%). Regarding the RMSECV (13.66%), the value was very close to the yellow LED (13.59%) in this same item.

Then, seeking better results, new modeling was carried out with separate RGB histograms for each LED. In this case, a number of factors equal to 10 were set-up, so that all models could be compared from the same configuration, according to Table 4.

**Table 4.** Results obtained comparing all training models by LED and RGB histogram.

LED	Green	Red	White	Yellow	Blue	Histogram
Variables	256	256	256	256	256	-
Factors	10	10	10	10	10	-
$R^2$	0.507	0.586	0.838	0.614	0.236	Red
RMSEC (%)	13.13	12.03	7.56	11.61	16.34	Red
RMSECV (%)	17.35	24.21	20.86	18.73	20.33	Red
$R^2$	0.705	0.243	0.751	0.590	0.378	Green
RMSEC (%)	12.16	16.27	9.33	11.97	14.74	Green
RMSECV (%)	23.64	25.72	21.61	22.14	22.61	Green
$R^2$	0.552	0.151	0.818	0.221	0.799	Blue
RMSEC (%)	12.51	17.23	7.98	16.50	8.37	Blue
RMSECV (%)	18.06	48.31	22.29	19.67	32.56	Blue

Comparing the generated values, no model obtained better indexes of figures of merit than the white LED, which presented the highest  $R^2$ , estimated at 0.857, and the lowest RMSEC (7.06%). In relation to the RMSECV (13.66%), the value was very close to the yellow LED (13.59%) in this same item.

Lastly, a single model was built with all histograms for all LEDs, resulting in 3840 variables. In this case, the model employed a better number of factors, according to the result of RMSECV, as demonstrated in Table 5. This model generated the best results regarding linearity ( $R^2$  equal to 0.962) and RMSEC (3.66%). About the figure of merit RMSECV, the result was shown to be greater than in the first generated model. Figure 5 shows the linear performance of this model.

**Table 5.** Results obtained with all data.

LED	Joined Histogram
Variables	3840
Factors	5
$R^2$	0.962
RMSEC (%)	3.66
RMSECV (%)	16.87

Therefore, the Kennard–Stone algorithm was applied to segregate the samples in a group for calibration and another for validation. From 50 soil samples, the algorithm selected 34 samples for the calibration model and 16 samples for the validation or test model. Table 6 presents the predicted results from the calibration model and its reference values.

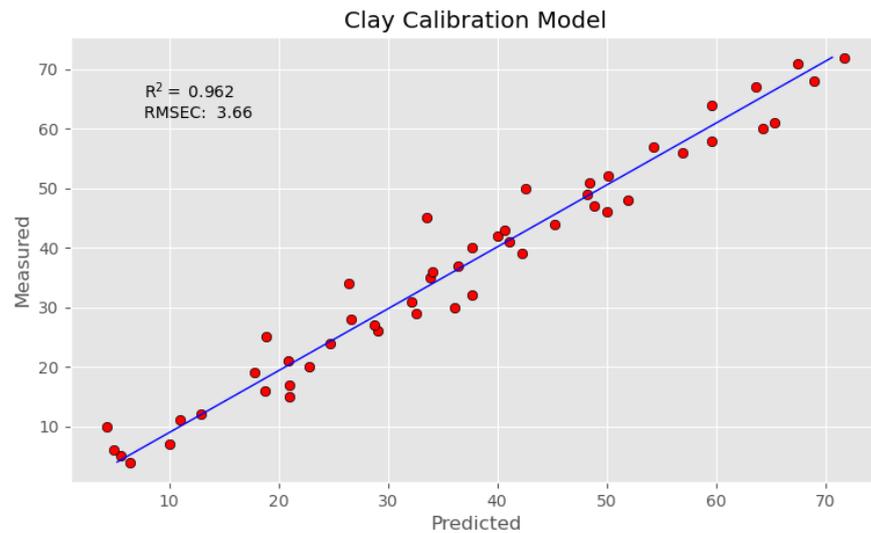


Figure 5. Results of machine learning calibration model.

Table 6. Prediction results using Kennard–Stone algorithm.

#Sample	Clay Ref%	Clay LED%	#Sample	Clay Ref%	Clay LED%
55121	4	6.12	55830	36	35.51
55051	6	5.34	55892	37	37.51
55066	7	12.14	53981	39	40.20
55129	10	1.83	56181	40	35.84
55049	11	11.04	55433	41	49.42
55446	15	22.31	53982	44	40.73
55478	19	19.59	55375	45	37.17
56145	20	22.84	56005	46	46.96
55469	24	27.21	55406	47	41.86
56148	25	23.78	55360	49	50.63
56103	26	24.95	60231	51	50.92
53977	27	28.50	56479	58	54.81
55988	29	33.24	56259	61	64.97
56105	31	28.64	55189	64	62.73
55962	32	35.45	60199	68	70.11
55437	34	29.86	60172	71	67.03
54015	35	34.20	60182	72	70.54

When comparing the results of the calibration model generated in this multispectral LED system with works by different authors, better predictive assessment rates were achieved. Wetterlind et al. in 2015 [47] obtained an  $R^2$  of 0.76 with an RMSECV of 6.4% and Tümsavaş in 2019 [8] found an  $R^2$  of 0.91 with an RMSECV of 3.4%, both using the NIR spectroscopy method. The minor RMSECVs of these authors are due to the low sample representativeness. Both focused their experiments on batches of approximately 0.5 km<sup>2</sup>, while this work represents various points in Vale do Rio Pardo, Brazil, around 13,255.7 km<sup>2</sup>. In addition, other factors such as sensitivity, reproducibility and equipment interference, intrinsic to the method, could also be discussed.

### 6. Conclusions

This work presented the use of multivariate calibration techniques in soil imaging from a multispectral camera in order to predict the amount of clay present in the samples. Concerning calibration results with low RMSEC values, however, the performance during the prediction could present better indexes.

Clay is one of the required parameters, among others, in order to assess soil fertility. As presented in this work, the concentration of this substance in the soil was quantitatively achieved through a multispectral camera. It was also possible to perceive a great potential in the correlation with the official routine analysis.

As advantages, the methodologies that were developed in this work are simple and maintain the integrity of the samples without the need for methods of greater complexity, presenting relatively low cost. The samples are analyzed in less time without the use of reagents and in a non-invasive way.

The combination of OpenCV and machine learning libraries with a low powered device, as a Raspberry Pi, will allow a wide range of research opportunities in agriculture, more precisely smart farms.

As future work, a larger number of samples could make the model more full bodied, predicting more linear effects due to a larger population. Another approach that could be taken is the generation of smaller calibration groups or range calibration. After a global model estimates an initial result, smaller models, with a restricted calibration range, could improve the accuracy of the sample.

Future research will organize the data collected in Context Histories [48,49] to allow pattern recognition [50], context prediction [51] and similarity analysis [52]. These analyses will improve the possibilities to implement intelligent services in the agriculture environments. Finally, the proposed technology can be embedded on equipment used in smart farms such as smart tractors and drones.

**Author Contributions:** Conceptualization, D.A., G.A.H. and A.B.d.C.; Investigation, D.A. and G.A.H.; Methodology, D.A., G.A.H. and A.B.d.C.; Software, D.A. and G.A.H.; Project Administration, G.A.H. and A.B.d.C.; Supervision, G.A.H. and A.B.d.C.; Validation, D.A., G.A.H. and A.B.d.C.; Writing—original draft, G.A.H. and J.L.V.B.; Writing—review and editing, G.A.H., J.L.V.B. and V.R.Q.L.; Financial, V.R.Q.L. and M.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** We would like to thank to Seed Funding ILIND—Instituto Lusófono de Investigação e Desenvolvimento, COPELABS.

**Informed Consent Statement:** This research did not require ethical approval in accordance with the regulations of the University of Vale do Rio dos Sinos (UNISINOS) and University of Santa Cruz do Sul (UNISC).

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to thank the University of Vale do Rio dos Sinos (Unisinos), the Applied Computing Graduate Program (PPGCA), the Mobile Computing Laboratory (Mobilab), the Central Analítica (soil analysis laboratory), the Research Support Foundation of the State of Rio Grande do Sul (FAPERGS), the National Development Council Scientific and Technological (CNPq), the Coordination for the Improvement of Higher Education Personnel-Brazil (CAPES)-Code Funding 001.

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

## Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial Neural Network
CMOS	Complementary Metal-Oxide-Semiconductor
CPU	Central Process Unit
CSI	Camera Serial Interface
CSV	Comma Separated Values
DT	Decision Trees
GPIO	General Purpose Input/Output

IoT	Internet of Things
LED	Light Emitting Diode
LVs	Latent Variables
ML	Machine Learning
MMA	Methods of Multivariate Analysis
MIPI	Mobile Industry Processor Interface
N/A	Not Available
NoIR	No Infra-Red
OpenCV	Open Source Computer Vision Library
PLS	Partial Least Squares
PNG	Portable Network Graphics
R <sup>2</sup>	Coefficient of Determination
RAM	Random Access Memory
Ref	Reference sample
RGB	Red Green Blue
RMSEC	Root Mean Square Error of Calibration
RMSECV	Root Mean Square Error of Cross Validation
SVM	Support Vector Machines
Symbols	
$n$	Number of samples
$x$	Predicted clay concentration
$y$	Reference clay concentration

## References

1. Fiehn, H.B.; Schiebel, L.; Avila, A.F.; Miller, B.; Mickelson, A. Smart Agriculture System Based on Deep Learning. In Proceedings of the 2nd International Conference on Smart Digital Environment (ICSDE'18), Rabat, Morocco, 18–20 October 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 158–165. [\[CrossRef\]](#)
2. Lytos, A.; Lagkas, T.; Sarigiannidis, P.; Zervakis, M.; Livanos, G. Towards smart farming: Systems, frameworks and exploitation of multiple sources. *Comput. Netw.* **2020**, *172*, 107147. [\[CrossRef\]](#)
3. Hochman, Z.; Carberry, P.; Robertson, M.; Gaydon, D.; Bell, L.; McIntosh, P. Prospects for ecological intensification of Australian agriculture. *Eur. J. Agron.* **2013**, *44*, 109–123. [\[CrossRef\]](#)
4. Saiz-Rubio, V.; Rovira-Más, F. From Smart Farming towards Agriculture 5.0: A Review on Crop Data Management. *Agronomy* **2020**, *10*, 207. [\[CrossRef\]](#)
5. Wolfert, S.; Ge, L.; Verdouw, C.; Bogaardt, M.J. Big Data in Smart Farming—A review. *Agric. Syst.* **2017**, *153*, 69–80. [\[CrossRef\]](#)
6. Demattê, J.A.M.; Dotto, A.C.; Bedin, L.G.; Sayão, V.M.; e Souza, A.B. Soil analytical quality control by traditional and spectroscopy techniques: Constructing the future of a hybrid laboratory for low environmental impact. *Geoderma* **2019**, *337*, 111–121. [\[CrossRef\]](#)
7. Bolfe, É.L.; de Castro Jorge, L.A.; Sanches, I.D.; Júnior, A.L.; da Costa, C.C.; de Castro Victoria, D.; Inamasu, R.Y.; Grego, C.R.; Ferreira, V.R.; Ramirez, A.R. Precision and Digital Agriculture: Adoption of Technologies and Perception of Brazilian Farmers. *Agriculture* **2020**, *10*, 653. [\[CrossRef\]](#)
8. 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. [\[CrossRef\]](#)
9. Griebeler, G.; da Silva, L.S.; Cargnelutti Filho, A.; Santos, L.d.S. Avaliação de um programa interlaboratorial de controle de qualidade de resultados de análise de solo. *Rev. Ceres* **2016**, *63*, 371–379. [\[CrossRef\]](#)
10. Demattê, J.A.M.; Alves, M.R.; da Silva Terra, F.; Bosquilia, R.W.D.; Fongaro, C.T.; da Silva Barros, P.P. Is It Possible to Classify Topsoil Texture Using a Sensor Located 800 km Away from the Surface? *Rev. Bras. Ciência Solo* **2016**, *40*. [\[CrossRef\]](#)
11. Nanni, M.R.; Demattê, J.A.M.; Rodrigues, M.; dos Santos, G.L.A.A.; Reis, A.S.; de Oliveira, K.M.; Cezar, E.; Furlanetto, R.H.; Crusiol, L.G.T.; Sun, L. Mapping Particle Size and Soil Organic Matter in Tropical Soil Based on Hyperspectral Imaging and Non-Imaging Sensors. *Remote Sens.* **2021**, *13*, 1782. [\[CrossRef\]](#)
12. Guo, Y.; Chen, S.; Wu, Z.; Wang, S.; Bryant, C.R.; Senthilnath, J.; Cunha, M.; Fu, Y.H. Integrating Spectral and Textural Information for Monitoring the Growth of Pear Trees Using Optical Images from the UAV Platform. *Remote Sens.* **2021**, *13*, 1795. [\[CrossRef\]](#)
13. Crucil, G.; Oost, K.V. Towards Mapping of Soil Crust Using Multispectral Imaging. *Sensors* **2021**, *21*, 1850. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Garini, Y.; Young, I.T.; McNamara, G. Spectral imaging: Principles and applications. *Cytom. Part A* **2006**, *69A*, 735–747. [\[CrossRef\]](#)
15. Amigo, J.M.; Grassi, S. Configuration of hyperspectral and multispectral imaging systems. In *Data Handling in Science and Technology*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 17–34. [\[CrossRef\]](#)
16. Ambrož, M. Raspberry Pi as a low-cost data acquisition system for human powered vehicles. *Measurement* **2017**, *100*, 7–18. [\[CrossRef\]](#)
17. Lucca, A.V.; Sborz, G.M.; Leithardt, V.; Beko, M.; Zeferino, C.A.; Parreira, W. A Review of Techniques for Implementing Elliptic Curve Point Multiplication on Hardware. *J. Sens. Actuator Netw.* **2020**, *10*, 3. [\[CrossRef\]](#)

18. Leithardt, V.; Santos, D.; Silva, L.; Viel, F.; Zeferino, C.; Silva, J. A Solution for Dynamic Management of User Profiles in IoT Environments. *IEEE Lat. Am. Trans.* **2020**, *18*, 1193–1199. [[CrossRef](#)]
19. Viel, F.; Silva, L.A.; Leithardt, V.R.Q.; Santana, J.F.D.P.; Teive, R.C.G.; Zeferino, C.A. An Efficient Interface for the Integration of IoT Devices with Smart Grids. *Sensors* **2020**, *20*, 2849. [[CrossRef](#)]
20. Helfer, G.A.; Barbosa, J.L.V.; dos Santos, R.; da Costa, A.B. A computational model for soil fertility prediction in ubiquitous agriculture. *Comput. Electron. Agric.* **2020**, *175*, 105602. [[CrossRef](#)]
21. Da Costa, A.; Helfer, G.; Barbosa, J.; Teixeira, I.; Santos, R.; dos Santos, R.; Voss, M.; Schlessner, S.; Barin, J. PhotoMetrix UVC: A New Smartphone-Based Device for Digital Image Colorimetric Analysis Using PLS Regression. *J. Braz. Chem. Soc.* **2021**. [[CrossRef](#)]
22. Martini, B.G.; Helfer, G.A.; Barbosa, J.L.V.; Modolo, R.C.E.; da Silva, M.R.; de Figueiredo, R.M.; Mendes, A.S.; Silva, L.A.; Leithardt, V.R.Q. IndoorPlant: A Model for Intelligent Services in Indoor Agriculture Based on Context Histories. *Sensors* **2021**, *21*, 1631. [[CrossRef](#)]
23. Baumann, L.; Librelotto, M.; Pappis, C.; Helfer, G.A.; Santos, R.O.; Santos, R.B.; Costa, A.B. NanoMetrix: An app for chemometric analysis from near infrared spectra. *J. Chemom.* **2020**, *34*. [[CrossRef](#)]
24. Pozo, S.D.; Rodríguez-González, P.; Sánchez-Aparicio, L.J.; Muñoz-Nieto, A.; Hernández-López, D.; Felipe-García, B.; González-Aguilera, D. Multispectral Imaging in Cultural Heritage Conservation. *ISPRS Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2017**, *42*, 155–162. [[CrossRef](#)]
25. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 2nd ed.; Addison-Wesley Longman Publishing Co., Inc.: New York, NY, USA, 2001.
26. Cao, A.; Pang, H.; Zhang, M.; Shi, L.; Deng, Q.; Hu, S. Design and Fabrication of an Artificial Compound Eye for Multi-Spectral Imaging. *Micromachines* **2019**, *10*, 208. [[CrossRef](#)]
27. Carstensen, J.M. LED spectral imaging with food and agricultural applications. *Image Sensing Technologies: Materials, Devices, Systems, and Applications V*; Dhar, N.K., Dutta, A.K., Eds.; SPIE: Orlando, FL, USA, 2018. [[CrossRef](#)]
28. Dagar, R.; Som, S.; Khatri, S.K. Smart Farming—IoT in Agriculture. In Proceedings of the IEEE 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 11–12 July 2018. [[CrossRef](#)]
29. Han, P.; Dong, D.; Zhao, X.; Jiao, L.; Lang, Y. A smartphone-based soil color sensor: For soil type classification. *Comput. Electron. Agric.* **2016**, *123*, 232–241. [[CrossRef](#)]
30. Brady, N.C.; Weil, R.R. *The Nature and Properties of Soils*, 14th ed.; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2008.
31. Peverill, K.I. (Ed.) *Soil Analysis: An Interpretation Manual*, Reprinted ed.; CSIRO Publ: Melbourne, Australia, 2005.
32. Elias, R. *Digital Media: A Problem-Solving Approach for Computer Graphics*, 1st ed.; Springer: Cham, Switzerland, 2014.
33. Gerlach, J.B. *Digital Nature Photography*; 2nd ed.; Routledge: Oxford, UK, 2015.
34. Domínguez, C.; Heras, J.; Pascual, V. IJ-OpenCV: Combining ImageJ and OpenCV for processing images in biomedicine. *Comput. Biol. Med.* **2017**, *84*, 189–194. [[CrossRef](#)] [[PubMed](#)]
35. Mehmood, T.; Liland, K.H.; Snipen, L.; Sæbø, S. A review of variable selection methods in Partial Least Squares Regression. *Chemom. Intell. Lab. Syst.* **2012**, *118*, 62–69. [[CrossRef](#)]
36. Wang, J.; Tiyip, T.; Ding, J.; Zhang, D.; Liu, W.; Wang, F.; Tashpolat, N. Desert soil clay content estimation using reflectance spectroscopy preprocessed by fractional derivative. *PLoS ONE* **2017**, *12*, e0184836. [[CrossRef](#)]
37. Nawar, S.; Mouazen, A.M. Optimal sample selection for measurement of soil organic carbon using on-line vis-NIR spectroscopy. *Comput. Electron. Agric.* **2018**, *151*, 469–477. [[CrossRef](#)]
38. Svensgaard, J.; Roitsch, T.; Christensen, S. Development of a Mobile Multispectral Imaging Platform for Precise Field Phenotyping. *Agronomy* **2014**, *4*, 322–336. [[CrossRef](#)]
39. Hassan-Esfahani, L.; Torres-Rua, A.; Jensen, A.; McKee, M. Assessment of Surface Soil Moisture Using High-Resolution Multi-Spectral Imagery and Artificial Neural Networks. *Remote Sens.* **2015**, *7*, 2627–2646. [[CrossRef](#)]
40. Treboux, J.; Genoud, D. Improved Machine Learning Methodology for High Precision Agriculture. In Proceedings of the IEEE 2018 Global Internet of Things Summit (GIoTS), Bilbao, Spain, 4–7 June 2018. [[CrossRef](#)]
41. Žížala, D.; Minařík, R.; Zádorová, T. Soil Organic Carbon Mapping Using Multispectral Remote Sensing Data: Prediction Ability of Data with Different Spatial and Spectral Resolutions. *Remote Sens.* **2019**, *11*, 2947. [[CrossRef](#)]
42. Lopez-Ruiz, N.; Granados-Ortega, F.; Carvajal, M.A.; Martínez-Olmos, A. Portable multispectral imaging system based on Raspberry Pi. *Sens. Rev.* **2017**, *37*, 322–329. [[CrossRef](#)]
43. OmniVision Technologies Inc. OV5647 Sensor Datasheet. Available online: [https://cdn.sparkfun.com/datasheets/Dev/RaspberryPi/ov5647\\_full.pdf](https://cdn.sparkfun.com/datasheets/Dev/RaspberryPi/ov5647_full.pdf) (accessed on 8 June 2021).
44. Raspberry Pi Foundation. FAQs—Raspberry Pi Documentation. Available online: <https://www.raspberrypi.org/documentation/faqs/> (accessed on 8 June 2021).
45. Park, J.I.; Lee, M.H.; Grossberg, M.D.; Nayar, S.K. Multispectral Imaging Using Multiplexed Illumination. In Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–21 October 2007. [[CrossRef](#)]
46. Pelliccia, D. Partial Least Squares Regression in Python. 2019. Available online: <https://nirpyresearch.com/partial-least-squares-regression-python/> (accessed on 3 October 2020).
47. Wetterlind, J.; Piikki, K.; Stenberg, B.; Söderström, M. Exploring the predictability of soil texture and organic matter content with a commercial integrated soil profiling tool. *Eur. J. Soil Sci.* **2015**, *66*, 631–638. [[CrossRef](#)]

48. Aranda, J.A.S.; Bavaresco, R.S.; de Carvalho, J.V.; Yamin, A.C.; Tavares, M.C.; Barbosa, J.L.V. A computational model for adaptive recording of vital signs through context histories. *J. Ambient. Intell. Humaniz. Comput.* **2021**. [[CrossRef](#)] [[PubMed](#)]
49. Rosa, J.H.; Barbosa, J.L.V.; Kich, M.; Brito, L. A Multi-Temporal Context-aware System for Competences Management. *Int. J. Artif. Intell. Educ.* **2015**, *25*, 455–492. [[CrossRef](#)]
50. Dupont, D.; Barbosa, J.L.V.; Alves, B.M. CHSPAM: a multi-domain model for sequential pattern discovery and monitoring in contexts histories. *Pattern Anal. Appl.* **2019**, *23*, 725–734. [[CrossRef](#)]
51. Da Rosa, J.H.; Barbosa, J.L.; Ribeiro, G.D. ORACON: An adaptive model for context prediction. *Expert Syst. Appl.* **2016**, *45*, 56–70. [[CrossRef](#)]
52. Filippetto, A.S.; Lima, R.; Barbosa, J.L.V. A risk prediction model for software project management based on similarity analysis of context histories. *Inf. Softw. Technol.* **2021**, *131*, 106497. [[CrossRef](#)]