



## Article

# Unbiasing the Estimation of Chlorophyll from Hyperspectral Images: A Benchmark Dataset, Validation Procedure and Baseline Results

Bogdan Ruszczak <sup>1,2</sup> , Agata M. Wijata <sup>2,3</sup> and Jakub Nalepa <sup>2,4,\*</sup>

<sup>1</sup> Faculty of Electrical Engineering, Automatic Control and Informatics, Department of Informatics, Opole University of Technology, Prószkowska 76, 45-758 Opole, Poland

<sup>2</sup> KP Labs, Konarskiego 18C, 44-100 Gliwice, Poland

<sup>3</sup> Faculty of Biomedical Engineering, Silesian University of Technology, Roosevelta 40, 41-800 Zabrze, Poland

<sup>4</sup> Faculty of Automatic Control, Electronics and Computer Science, Department of Algorithmics and Software, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland

\* Correspondence: jnalepa@ieeee.org

**Abstract:** Recent advancements in hyperspectral remote sensing bring exciting opportunities for various domains. Precision agriculture is one of the most widely-researched examples here, as it can benefit from the non-invasiveness and enormous scalability of the Earth observation solutions. In this paper, we focus on estimating the chlorophyll level in leaves using hyperspectral images—capturing this information may help farmers optimize their agricultural practices and is pivotal in planning the plants' treatment procedures. Although there are machine learning algorithms for this task, they are often validated over private datasets; therefore, their performance and generalization capabilities are virtually impossible to compare. We tackle this issue and introduce an open dataset including the hyperspectral and in situ ground-truth data, together with a validation procedure which is suggested to follow while investigating the emerging approaches for chlorophyll analysis with the use of our dataset. The experiments not only provided the solid baseline results obtained using 15 machine learning models over the introduced training-test dataset splits but also showed that it is possible to substantially improve the capabilities of the basic data-driven models. We believe that our work can become an important step toward standardizing the way the community validates algorithms for estimating chlorophyll-related parameters, and may be pivotal in consolidating the state of the art in the field by providing a clear and fair way of comparing new techniques over real data.

**Dataset:** DOI:10.1016/j.dib.2022.108087.

**Dataset License:** The license under which the data set is made available is (CC-BY).

**Keywords:** hyperspectral imaging; machine learning; chlorophyll estimation; benchmark; validation



**Citation:** Ruszczak, B.; Wijata, A.M.; Nalepa, J. Unbiasing the Estimation of Chlorophyll from Hyperspectral Images: A Benchmark Dataset, Validation Procedure and Baseline Results. *Remote Sens.* **2022**, *14*, 5526. <https://doi.org/10.3390/rs14215526>

Academic Editors: Yuhong He and Anita Simic-Milas

Received: 9 September 2022

Accepted: 31 October 2022

Published: 2 November 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

Recent advancements in sensor technology bring new possibilities in hyperspectral image (HSI) analysis—such data effectively captures hundreds of spectral bands in the electromagnetic spectrum. In precision agriculture, acquiring detailed information concerning the chlorophyll saturation lets the plant breeders optimize their operation and plan the plants' treatment. Because the chlorophyll fluorescence, which is induced by solar radiation, is a direct representative of the actual vegetation photosynthesis, it is also the main vegetation performance indicator [1]. Therefore, monitoring the chlorophyll fluorescence parameters could bring important information on plants' stress or help us to detect the moment of the crop photosynthesis termination [2]. Furthermore, such measurements are valuable because they could make us understand the plant response to herbicide treatments and enable us to react quickly to possible plant conditions' changes. Additionally,

the exposition to toxicants can be inferred from HSI [3]. Finally, the decreasing amount of chlorophyll—as a result of combustion—causes changes in the characteristics of the spectral signature of hyperspectral measurements [4]. It allows us to extract important insights concerning the scanned area, e.g., to assess whether it is under an active fire (because there are continuous changes over time), the area is burnt (because the chlorophyll content is low) or if there is a risk of the active area re-developing (partial burnout). It is worth mentioning that the process of monitoring and prevention of fires in Europe is carried out by the European Forest Fire Information System, which is part of the Earth Observation Programme. As part of it, monitoring of fires is carried out using multispectral data from Sentinel-2—the key information source is the red-edge band, which is one of the best descriptors of chlorophyll [4]. Furthermore, the chlorophyll-a index allows for the preparation of the pigmentation map [5], which can be the basis for detecting harmful algae blooms using the remotely sensed data [6,7]. Thus, determining the concentration of chlorophyll is important in monitoring water quality as well. Overall, although we focus on estimating the chlorophyll in leaves, non-invasive determination of its level is of paramount importance in an array of applications that could be potentially targeted at an enormous scale thanks to airborne and satellite imaging [8].

There are several ways to determine the level of chlorophyll in leaves, but most of them require direct and invasive access to leaves. Therefore, due to the high costs of such economically-infeasible, time-consuming and non-scalable procedures, the non-invasive techniques have become an important yet still under-developed research venue. The approaches which are focused on exploiting the multi- and hyperspectral images for this task span across those that use the in-field [9], airborne [10] and satellite [11,12] imaging, with the latter offering immediate scalability over large areas. Although there are works that reported promising results for the multispectral data [13], hyperspectral imaging is the current main focus, as it can allow for precise chlorophyll estimation thanks to the very detailed spectral information available in such data [14].

The state-of-the-art techniques for estimating the chlorophyll level exploit classic and deep supervised machine learning [15], with the latter benefiting from automated representation learning [16]. Such algorithms, however, require representative and large training sets capturing both image data and in situ measurements to generalize over new data. Unfortunately, albeit chlorophyll estimation is an important topic, there are no publicly available and established datasets that could provide an unbiased way of comparing the approaches for this task. Therefore, we are currently facing the reproducibility crisis [17]. Additionally, collecting high-quality ground truth is time- and cost-inefficient, hence such datasets are often synthesized [18]. Haboudane et al. focused on estimating the chlorophyll level from HSI and used a private set containing the images (72 VIS-NIR bands with 2 m GSD) and 12 reference measurements [10]. Similarly, the HSI band selection targeting the chlorophyll estimation was tackled over the in-house data in [19]. Although it is possible to obtain a rough chlorophyll level approximation using the spectral indices [12,20], its quality is questionable [21].

Acquiring precise in situ measurements is a pivotal step in building datasets that could be used to train and validate machine learning chlorophyll estimation algorithms. The majority of approaches that measure the actual level of chlorophyll utilize the soil-plant analysis development (SPAD) parameter [22]. There are, however, techniques exploiting the photosynthesis efficiency parameter and the chemical reflectance index [23]. Interestingly, some works presented the methodology to model the relationship between the chlorophyll content and the plant stress that could be investigated using the maximal photochemical efficiency of PSII ( $F_v/F_m$ ) [24] or the fluctuations in the light intensity. Overall, the algorithms for non-invasive monitoring of chlorophyll parameters from multi/hyperspectral image data have been intensively researched due to their practical applicability and potential scalability (e.g., if deployed on board a satellite [8]) in precision agriculture, but there are no standardized procedures to validate them in an unbiased way. Furthermore, there are no publicly available and adopted datasets that could be used in such validation pipelines.

### 1.1. Contribution

In this paper, we address both research gaps (the lack of standardized procedures to validate the algorithms for chlorophyll estimation from HSI, and the lack of public and adopted datasets that could be utilized to investigate the approaches for this task) and introduce an end-to-end and reproducible validation procedure coupled with the real-life dataset of hyperspectral imagery and in situ measurements. We captured the in situ chlorophyll measurements together with the high-resolution HSI data and introduced a standardized approach for using this dataset to validate the emerging chlorophyll estimation techniques. This validation procedure will help us avoid any experimental flaws—we discussed such flaws concerned with the training/test dataset splits in our previous work in the context of the HSI segmentation [25].

Our contributions are therefore threefold:

- We introduce a publicly available set of (i) chlorophyll content measurements with some complementary information, including the soil moisture, weather parameters collected during the measurements or the relative water content, being the amount of water in a leaf at the time of sampling relative to the maximal water a leaf can hold and (ii) the corresponding high-resolution hyperspectral imagery (2.2 cm GSD). The dataset encompasses the orthophotomaps with the marked plots where the chlorophyll sampling has been completed, as well as the extracted images of separate plots. We performed the on-the-ground chlorophyll measurements, which resulted in four ground-truth parameters (Section 3.1):
  1. The SPAD index [22];
  2. The maximum quantum yield of the PSII photochemistry ( $F_v/F_m$ ) [24];
  3. The performance index for energy conservation from photons absorbed by PSII to the final PSI electron acceptors (PI) [26];
  4. Relative water content (RWC) measurements for the sampled canopy, for capturing additional derivative information on the nutrition of the plants.
- We introduce a procedure for the unbiased validation of machine learning algorithms for estimating the chlorophyll-related parameters from HSI, and we ensure the full reproducibility of the experiments over our dataset (Section 3.2).
- We deliver the baseline results obtained for the introduced dataset (for four ground-truth parameters) using 15 machine learning techniques which can constitute the reference for any future studies emerging from our work (Section 4). Additionally, we show that the performance of a selected model can be further improved through regularization.

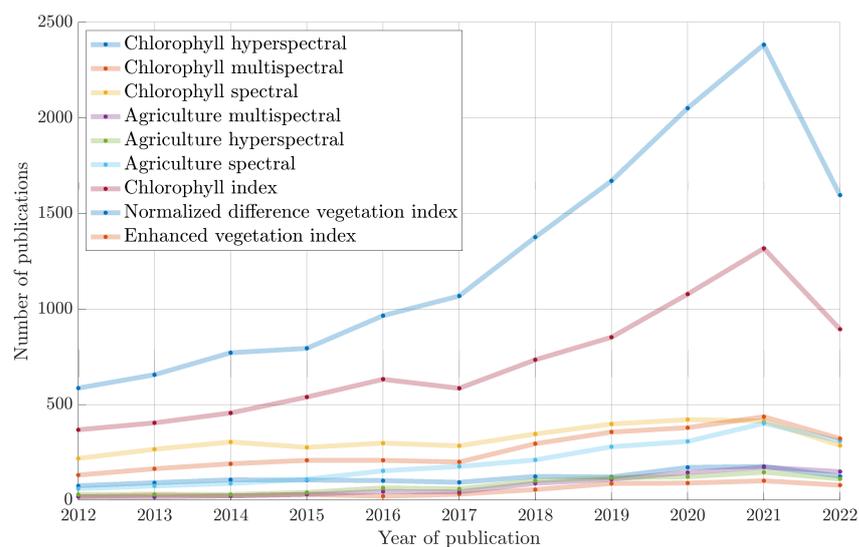
### 1.2. Structure of the Paper

In Section 2, we contextualize our work within state of the art by providing a review of an array of applications which can benefit from machine learning techniques operating on multi- and hyperspectral data, with a special emphasis put on chlorophyll estimation and on the way such approaches are validated. Section 3 introduces our chlorophyll estimation dataset, together with the validation procedure which is suggested to follow while exploiting it for experimentation. Our experimental results, obtained for various machine learning techniques over the suggested training-test dataset split, are gathered and discussed in Section 4. Finally, Section 5 concludes the paper.

## 2. Related Literature

The nature of activities in the agricultural sector has changed over the years as a result of a broadly-understood human activity, which encompasses—among other factors—rapidly growing population, environmental pollution, climate change and depletion of natural resources. The premise of precision agriculture is an effective food production process with a reduced impact on the environment. To achieve this goal, however, it is required to assess the soil quality, its irrigation, fertilizer content and seasonal changes

that occur in the ecosystem. Estimating the yield volume planned for a given region may also constitute the important information related to the effectiveness of the implemented agricultural practices [27]. Remote sensing may easily become a tool enabling the identification of soil and crop parameters due to the possibility of assessing a large area in subsequent time points. For agriculture, it is carried out using both passive and active methods. In the former case, multi- and hyperspectral remote sensing is used. The approaches using multispectral images (MSIs) are mainly based on the content of chlorophyll and its related parameters [28,29]. Nevertheless, the wide bandwidth that characterizes multispectral imaging results in limited accuracy in the early detection of negative symptoms such as nutrient deficiency or plant diseases [30]. The use of hyperspectral imaging, on the other hand, which is characterized by high spectral resolution (the bands are narrow and continuous), allows for the detection of more subtle details in the spectral response of a given area [31]. HSI-based methods can detect potential abnormalities, such as plant diseases, much faster than the MSI-based ones because the spectral signature contains more detailed characteristics derived from significantly narrower bands [32]. Additionally, satellites equipped with multispectral sensors (e.g., WorldView, QuickBird, Sentinel-2, Landsat) are still more popular than those with hyperspectral sensors (see, e.g., the EO-1 Hyperion mission and various emerging missions, including Intuition-1). Furthermore, there are some practical challenges that need to be faced for HSI missions, as such data may be extremely large, hence should be processed on board a satellite to downlink the “information” instead of raw image data. However, hyperspectral analysis in agriculture is popularly carried out by field-point methods using a spectroradiometer. The limitation to a few selected places makes spatial estimation impossible; therefore, research is conducted to determine the correlation between data collected by the field methods and data recorded by satellites [29,33,34] or by manned or unmanned airplanes [27,35–37]. In Figure 1, we show that the popularity of the topics (quantified as the number of papers published yearly) related to the MSI/HSI analysis in agriculture and including chlorophyll estimation, has been steadily growing over the last ten years. It also confirms the importance of introducing standardized validation procedures, which can be easily used to confront the emerging approaches for a given task in an unbiased and reproducible way.



**Figure 1.** The popularity of topics related to the HSI analysis in various agricultural applications, including the chlorophyll estimation, quantified as the number of papers on such topics published between 2012 and 2022 (this analysis is based on <https://app.dimensions.ai/discover/publication> (accessed on 1 November 2022), the analysis was performed on 8 September 2022; in the legend, we present the keyphrase which was used). We can appreciate that the number of articles tackling the automated chlorophyll determination from hyperspectral imagery is increasing at a steady pace.

HSIs provide a tremendous amount of spectral-spatial information. On its basis, target agricultural parameters can be estimated—the narrow infrared and near-infrared bands can be used to accurately calculate the leaf area index (LAI) [38]. Similarly, the analysis of spectral data allowed us to define a vegetation index based on the assessment of chlorophyll content [35,39]. Assessing the content of chlorophyll in crops is one of the most important approaches, as chlorophyll is a reliable indicator of the crop health. The reason for the high usefulness of this biophysical pigment is that it enables us to evaluate the biochemical processes, which reflects the productivity of plants [31]. Vegetation coefficients are the basis for the estimation and monitoring of biomass, as well as for the assessment of soil composition and its moisture. There have been an array of machine learning techniques proposed for such tasks—the biomass estimation was performed using random forests [35,36,40,41], support vector machines [36,40,41] and multivariate regression modeling [28,36,41]. To tackle the problem of the high dimensionality of hyperspectral data, band selection and feature extraction, using, e.g., principal component analysis [42], are commonly deployed [36,43]. They elaborate a subset of the most discriminative bands or features [44]—the experiments using HSI obtained by unmanned aerial vehicles suggest that limiting the spectral range in the context of monitoring the plant growth to 454–950 nm [35] or 454–882 nm [27] is enough to achieve this goal.

In Table 1, we gather a set of the selected works focusing on MSI/HSI analysis in various agricultural applications (the papers tackling the chlorophyll estimation are presented in green), whereas Table 2 presents the corresponding experimental results reported in those publications. We can appreciate that both classic and deep machine learning models have been extensively developed throughout the years, but their direct comparison is virtually impossible, as the authors (i) utilize different datasets (of different cardinality and underlying characteristics, as they may have been captured using various sensors, in different acquisition conditions and following different acquisition procedures) and validation scenarios (e.g., different cross-validation approaches), and (ii) commonly report different metrics quantifying the capabilities of the investigated techniques. In this work, we tackle this issue and introduce a dataset capturing high-resolution hyperspectral imagery coupled with the in situ chlorophyll measurements, alongside the suggested validation procedure (its training-test split and a set of metrics, which should be used to quantify the prediction performance of the algorithms) and our baseline results obtained in this experimental scenario. We believe that it may be an important step toward unbiasing the way the community verifies the emerging approaches for estimating chlorophyll-related parameters from hyperspectral imagery, and it can help us effectively tackle the reproducibility crisis in the machine learning-based HSI analysis [17].

**Table 1.** Selected works focusing on the machine learning-powered analysis in agricultural applications, together with the additional feature extraction step performed in the corresponding method. The papers focusing on the chlorophyll estimation are in green.

Ref.	Year	Goal	Feature Extraction	Algorithm
[31]	2022	prediction of chlorophyll content in maize	selected bands; vegetation indices	random forest
[35]	2021	monitoring of above-ground biomass of maize	vegetation indices	stepwise regression; random forest; extreme gradient boosting
[45]	2021	monitoring of wheat yellow rust	vegetation indices; meteorological features	linear discriminant analysis; support vector machine; artificial neural network
[46]	2020	mapping of live fuel moisture content	NDVI; NDWI; NIRv	recurrent neural network
[28]	2020	estimation of biomass	vegetation indices	deep neural network

Table 1. Cont.

Ref.	Year	Goal	Feature Extraction	Algorithm
[27]	2020	monitoring of crop growth	high-frequency IWD information; continuous wavelet transform (CWT)	multiple linear stepwise regression
[40]	2020	dry biomass estimation of paddy rice	VH; VV	random forest; support vector machine; <i>k</i> -nearest neighbor; gradient boosting decision tree
[47]	2020	LAI detection	LAI	Gaussian process regression (GPR); variational heteroscedastic GPR
[33]	2020	estimation of soil organic carbon	principal component analysis; NDI; RI; DI	discrete wavelet transform at different scales; random forest; support vector machine; back-propagation neural network
[48]	2020	estimation of soil organic carbon	vegetation indices: NDVI, SAVI, NBSI, NDWI, NDBI, FI	random forest
[36]	2019	monitoring of above-ground biomass of maize	recursive-feature elimination	multiple linear regression; support vector machine; artificial neural network; random forest
[41]	2019	pasture conditions, seasonal dynamics of LAI and AGB	NDVI; EVI; LSWI	multiple linear regression; support vector machine; random forest
[38]	2018	canopy green leaf area	GBVI; NDVI; CI	empirical vegetation index regression (NDVIa-b and CIa-b); physically-based inversion, support vector regression
[49]	2016	maize biomass estimation, the seasonal variation	specific leaf area	simple algorithm for yield estimates
[43]	2015	detection of favorable wavelengths	singular value decomposition	stepwise regression

Table 2. The results reported in the selected works on machine learning-powered analysis in various agricultural applications. The papers focusing on the chlorophyll estimation are in green.

Ref.	Date Source	Type	Wavelength	Amount of Data	Measure	Value
[31]	DJI S1000 UAV; MicaSense Altum; Downwelling Light Sensor 2 (DLS-2)	MSI	465, 532, 630 nm, 680–730 nm, 1200–1600 nm	3576	$R^2$ RMSE RRMSE	results reported for the seasons

Table 2. Cont.

Ref.	Date Source	Type	Wavelength	Amount of Data	Measure	Value
[35]	DJI S1000 UAV; Cubert UHD 185	HSI	454–950 nm	1809	$R^2$ RMSE RRMSE	0.85 0.27 $\frac{t}{ha}$ 0.84%
[45]	Sentinel-2; National Meteorological Information Center	MSI	490, 560, 665, 842, 705 nm	58	accuracy	84.20%
[46]	Sentinel-1; Landsat-8	SAR MSI	C-band; 450–510 nm, 530–590 nm, 640–670 nm, 850–880 nm, 1570–1650 nm, 2110–2290 nm	not specified	$R^2$ RMSE bias	0.63 25.00% 1.90%
[28]	Sentinel-2	MSI	443–2190 nm	209	$R^2$ RMSE RRMSE	0.87 1.84 $\frac{t}{ha}$ 24.76%
[27]	DJI S1000 UAV; Cubert UHD 185; Sony DSC QX100	HSI	454–882 nm	144	$R^2$ RMSE MAE	0.85 0.79 $\frac{t}{ha}$ 1.01 $\frac{t}{ha}$
[40]	Sentinel-1A	SAR	C-band	175	$R^2$ RMSE	0.72 362.40 $\frac{g}{m^2}$
[47]	Sentinel-2	MSI	400–2400 nm	114	$R^2$ RMSE $R^2$ RMSE	0.78 (GPR) 0.70 (GPR) 0.80 (VHGPR) 0.63 (VHGPR)
[33]	Gaofen-5	HSI	433–1342 nm, 1460–1763 nm, 1990–2445 nm	14	$R^2$ RMSE	0.83 2.89 $\frac{g}{kg}$
[48]	ALOS PALSAR; Landsat TM	SAR MSI	L-band; 530–590 nm, 640–670 nm, 850–880 nm, 1570–1650 nm, 2110–2290 nm	not specified	$R^2$ RMSE RDP	0.59 9.27 $\frac{g}{kg}$ 1.98
[36]	DJI S1000 UAV; 1.2 megapixel Parrot Sequoia camera	MSI	550–790 nm	120 185	$R^2$ RMSE MAE	0.94 0.50 0.36
[41]	Sentinel-1A; Landsat-8; Sentinel-2	SAR MSI	C-band; 452–512 nm, 636–673 nm, 851–879 nm, 1566–1651 nm	not specified	$R^2$ RMSE	0.78 119.40 $\frac{g}{m^2}$
[38]	HyMap; CHRIS/PROBA	HSI MSI	677–707 nm	118	$R^2$	0.79
[49]	Formosat-2; SPOT4-Take5; Landsat-8; Deimos-1	MSI	specific to the sensor	195	$R^2$ RRMSE	0.96 4.6%
[43]	Landsat ETM+; KONOS; GeoEye-1; WorldView-2; Hyperion	HSI MSI	772, 539, 758, 914, 1130, 1320 nm	9, 23, 23, 24, 10	$R^2$ RMSE	0.12–0.97 1.15–2.47 $\frac{g}{m^2}$

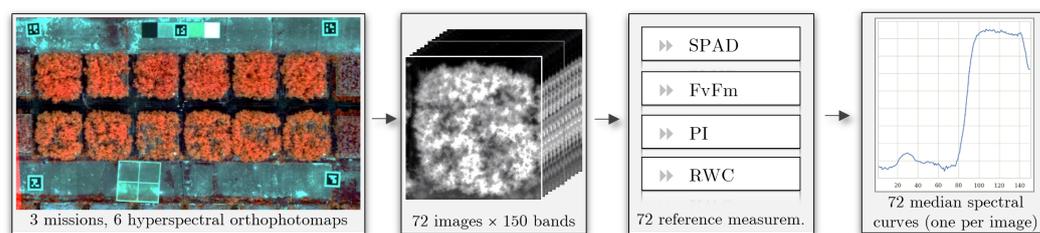
### 3. Materials and Methods

In this section, we present our dataset containing high-resolution hyperspectral imagery coupled with the in situ chlorophyll measurements (Section 3.1). The validation procedure which should be followed while utilizing this dataset to investigate the capabilities of emerging chlorophyll estimation techniques is discussed in detail in Section 3.2.

#### 3.1. Chlorophyll Estimation Dataset (CHESS)

In Figure 2, we visualize the process of building the CHlorophyll ESTimation DataSet (CHESS). The data was collected in 2020 in the Plant Breeding and Acclimatization Institute—National Research Institute (IHAR-PIB) facility located in the Central Poland region (Jadwisin, Masovian Voivodeship). For the selected 24 outdoor plots of two different soil profiles (12 plots for each soil profile, without any repetitions or overlaps), two popular (in the central Europe) potato varieties were planted: *Lady Claire* and *Markies* (split evenly). The acquisition was carried out in June and July 2020 (3 rounds of acquisition, 4 weeks apart: 3 flights over 2 sets of 12 plots resulting in 72 HSIs acquired in total) when the leaves were fully developed. The images were captured using an unmanned aerial vehicle with the push-broom imaging spectrometer that registers 150 continuous spectral bands (460–902 nm with the 2.2 cm GSD). The orthorectification procedure was executed using the collected image material of each spectral band—it was possible thanks to using four location targets (see the far left image in Figure 2) whose geographical positions have been collected with a precise GPS device. The spectral correction of those maps using four calibration targets of different spectral characteristics (selected to cover the spectrum range in which plants are perceived best) was performed. This allowed us to finalize the image acquisition process with low location error (less than 1 cm), high image resolution (2.2 cm GSD), and consistent spectral characteristics.

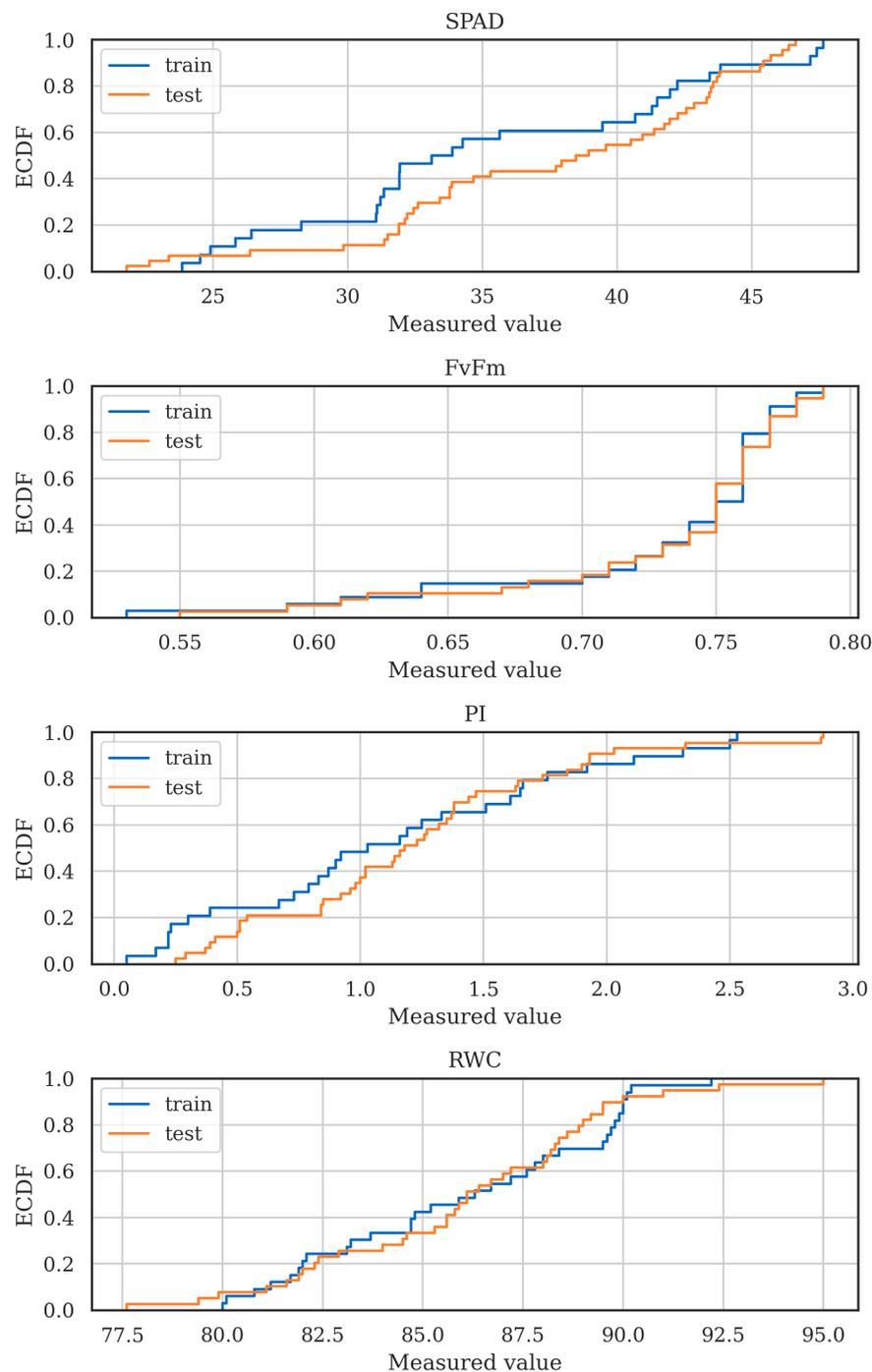
In parallel to the image acquisition campaign, the in situ on-the-ground measurements were performed on each plot (sampling was executed at the same time). To provide precise measurements, we captured (i) the readout of the photosynthesis efficiency quantified as the chlorophyll content using the SPAD index using the Minolta SPAD-502 device, (ii) the measurement of the maximum quantum yield of the PSII photochemistry (Fv/Fm) using the Multifunctional Plant Efficiency Analyzer (Handy-PEA fluorimeter, Hansatech Instruments Ltd. and Pea Plus software), (iii) the performance of the electron flux to the final PSI electron acceptors, as discussed in [50], and (iv) RWC which reflects the lab-measured degree of hydration of the leaf's tissue [51–53].



**Figure 2.** The dataset preparation procedure: 6 hyperspectral orthophotomaps, one for each flight for each set of 12 plots (left) were used to extract 72 hyperspectral 150-band images (middle left). The ground measurements of four parameters were performed for each plot (middle right). We extracted the spectral curves, individually for each pixel and aggregated across all pixels—see, e.g., the median spectral curve in the far right image.

The detailed agronomic setup and the dataset [54] with the training-test split constituting the validation procedure suggested in this paper, hence ensuring full reproducibility of the study, are available at <https://data.mendeley.com/datasets/xn2wy75f8m> (accessed on 1 November 2022). Since the measurements were collected in three independent rounds of data acquisition performed in the outdoor environment, CHESS reflects different plant characteristics and is intrinsically heterogeneous. We believe that exploiting a standardized

and non-biased validation procedure built upon such real-life data is of utmost importance to ensure full reproducibility and to avoid the “illusion of progress” in the field [17].

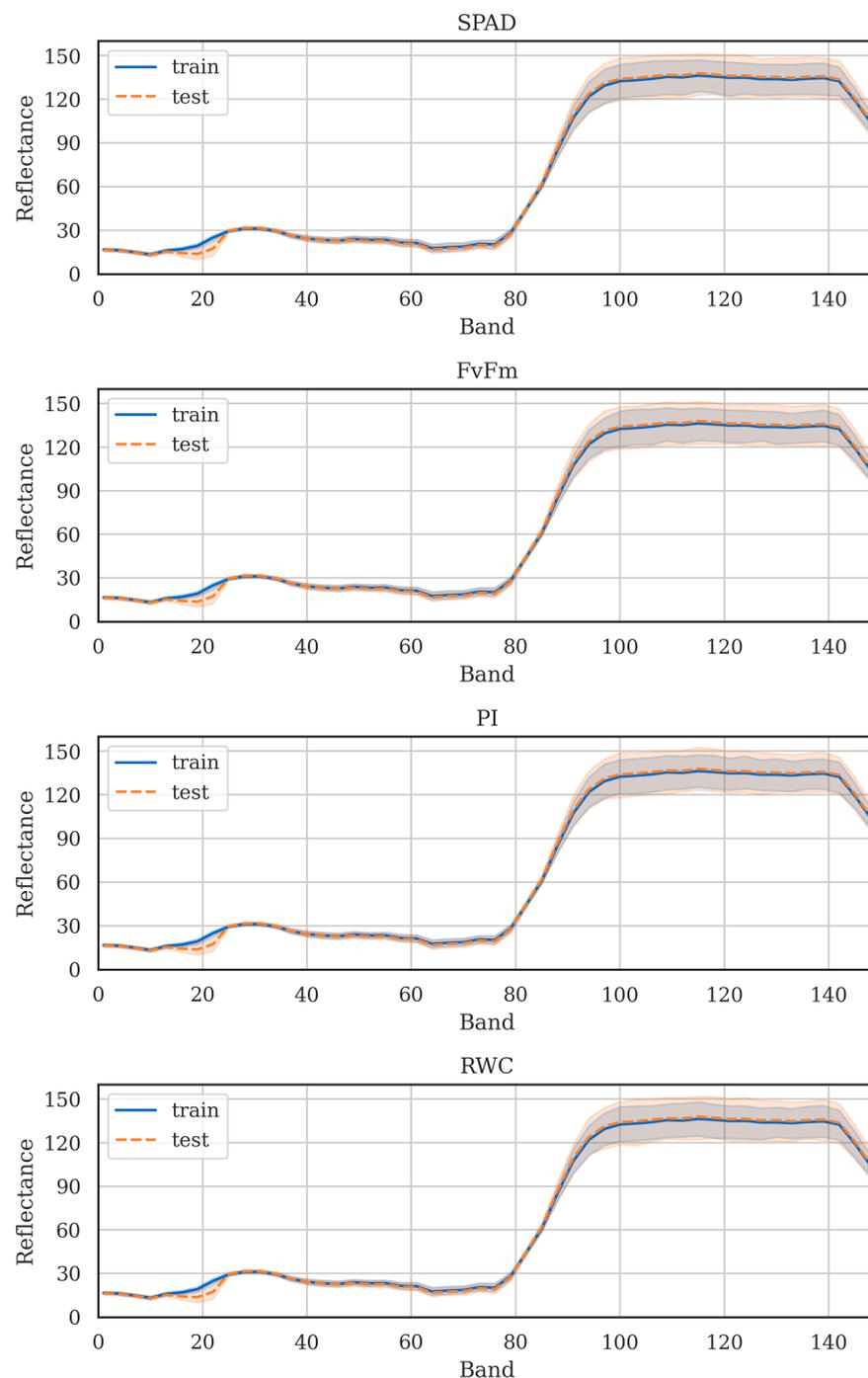


**Figure 3.** Empirical cumulative distribution (ECDF) of all parameters (SPAD, FvFm, PI and RWC) in the training and test sets.

### 3.2. Unbiased Validation of Chlorophyll Estimation

Unbiased and fair validation of the emerging algorithms for the non-invasive estimation of the chlorophyll-related parameters from HSI is critical to allow the community to track the progress in the field, and to accelerate the practical adoption of such approaches. Since there are differences across the measurement methodology followed for each in situ parameter (the SPAD index, Fv/Fm, PI and RWC), we provide four separate training-test dataset splits, independently for each ground-truth chlorophyll-related parameter. Each

split (i.e., for each parameter) is equinumerous, meaning that the number of plots is equal in the training and test subsets (36 HSIs with the ground-truth measurements captured for 36 separate plots of interest in both training and test sets). To be able to effectively quantify the generalization capabilities of the machine learning models trained and validated over such dataset splits, we stratified them according to the corresponding parameter's distribution to maintain similar distributions in both training and test sets (Figure 3). The reflectance characteristics of all HSIs across all folds are rendered in Figure 4. They show that there is a high agreement in the spectral features captured for the training and test images. Hence, the test set indeed resembles the characteristics of the training data and may be used to quantify the generalizability of the data-driven algorithms.



**Figure 4.** Spectral characteristics of all HSIs in the training and test sets. The mean spectral curves are rendered as blue and orange (dashed) lines for the training and test set, respectively.

Although the captured ground-truth parameters are commonly utilized in agronomy to assess the condition of the plants, they are measured differently, hence their characteristics are inherently varying. In Table 3, we gather the correlation coefficients across all of the parameters, indicating that, albeit some of them are indeed correlated (e.g., SPAD and PI, with the Pearson’s and Spearman’s coefficients amounting to 0.675 and 0.674, respectively), RWC is unrelated to other ones.

**Table 3.** Inter-parameter correlation between the measured ground-truth parameters (the Pearson’s correlation coefficient values are reported over the main diagonal, whereas the Spearman’s correlation coefficient values are given below the main diagonal; the corresponding  $p$ -values for the presented correlation coefficients are shown in brackets).

Parameter	SPAD	FvFm	PI	RWC
SPAD	1.000 (1.000)	0.668 (0.000)	0.675 (0.000)	0.109 (0.363)
FvFm	0.574 (0.000)	1.000 (1.000)	0.743 (0.000)	0.352 (0.002)
PI	0.674 (0.000)	0.877 (0.000)	1.000 (1.000)	−0.094 (0.434)
RWC	0.697 (0.550)	0.253 (0.161)	−0.047 (0.692)	1.000 (1.000)

#### 4. Experimental Results

The objectives of our experiments are twofold: (i) to present the baseline results, obtained using a variety of machine learning algorithms (15 in total), over the introduced chlorophyll estimation dataset (CHESS) using the proposed training-test dataset splits (independently for SPAD, Fv/Fm, PI, and RWC), and to (ii) show that the predictive power of a selected model can be improved through additional regularization. To quantify the prediction performance of the algorithms (over the test sets), we exploit the classic metrics, including the coefficient of determination  $R^2$  (upper bounded by the value of one indicating the perfect score, and all negative values of  $R^2$  indicating a worse fit than the average fit [55]), the mean absolute percentage error (MAPE), the mean squared error (MSE) and the mean absolute error (MAE)—all errors should be minimized. For all models, we utilized their default parameters (Table 4), as suggested by Pedregosa et al. [56]—we intentionally have not executed any additional hyperparameter optimization to present the baseline solutions elaborated using the machine learning techniques with default parameterization. The algorithms are fed with the median spectral curves (hence, the feature vectors contain 150 values corresponding to the median value of each band within the image), and each model predicts a single chlorophyll-related parameter (SPAD, Fv/Fm, PI or RWC). Therefore, we do not perform any additional feature extraction or band selection (they may easily improve the performance of data-driven HSI analysis algorithms [44]).

**Table 4.** The most important hyperparameter values, as suggested by Pedregosa et al. [56], of all parameterized machine learning algorithms investigated in this study.

Algorithm	Hyperparameters
Ada Boost	<b>Maximum number of estimators:</b> 50, <b>learning rate:</b> 1, linear loss.
Decision Tree	<b>Loss function:</b> squared error, <b>maximum depth of the tree:</b> not set, <b>minimum number of samples required to split an internal node:</b> 2, <b>minimum number of samples required to be at a leaf node:</b> 1, allowing all features to be considered for the best split with unlimited number of leaf nodes.
Extra Trees	<b>Maximum number of estimators:</b> $10^2$ , <b>loss function:</b> squared error, <b>maximum depth of the tree:</b> not set, <b>minimum number of samples required to split an internal node:</b> 2, <b>minimum number of samples required to be at a leaf node:</b> 1, allowing all features to be considered for the best split with samples’ bootstrapping while building trees, unlimited number of leaf nodes.
Extreme Gradient Boosting	<b>Learning rate:</b> 0.3, <b>maximum depth of an individual estimator:</b> 6, <b>minimum sum of instance weight:</b> 1, <b>maximum delta step:</b> 0, <b>regularization terms <math>\lambda</math>:</b> 1 and $\alpha$ : 0.

Table 4. Cont.

Algorithm	Hyperparameters
Gradient Boosting	<b>Learning rate:</b> 0.1, <b>maximum number of estimators:</b> $10^2$ , <b>maximum depth of an individual estimator:</b> 3, <b>loss function:</b> squared error, <b>percentage of samples required to split an internal node:</b> 10%.
Kernel Ridge	$\alpha$ : 1, <b>the degree of the linear polynomial kernel:</b> 3, <b>zero coefficient for polynomial and sigmoid kernels:</b> 1.
Lasso	$\alpha$ : 1, <b>maximum number of iterations:</b> $10^3$ , <b>tolerance stopping criteria:</b> $10^{-4}$ .
Light Gradient Boosting Machine	<b>Maximum tree leaves for base learners:</b> 31 without any limit for the tree depth for base learners, <b>boosting learning rate:</b> $10^{-1}$ , <b>number of boosted trees to fit:</b> $10^2$ , <b>number of samples for constructing bins:</b> $2 \cdot 10^4$ , no minimum loss reduction required to make a further partition on a leaf node, <b>minimum sum of instance weight in a leaf:</b> $10^{-3}$ , <b>minimum samples in a child:</b> 20.
Linear Support Vector Machine	<b>Regularization parameter C:</b> 1, <b>L1 loss,</b> <b>maximum number of iterations:</b> $10^3$ , <b>the tolerance stopping criteria:</b> $10^{-4}$ .
Nu Support Vector Machine	<b>Kernel:</b> Radial Basis Function, <b>upper bound on the fraction of training errors and a lower bound of the fraction of support vectors:</b> 0.5, <b>regularization parameter C:</b> 1, $\gamma$ : $1 / ( \mathcal{F}  \cdot \sigma^2(T))$ , where $\mathcal{F}$ is the number of features, and $\sigma^2(T)$ is the variance of the training set.
Random Forest	<b>Maximum number of estimators:</b> $10^2$ , <b>function measuring the quality of split:</b> squared error, <b>minimum number of samples required to split an internal node:</b> 2, <b>minimum number of samples required to be at a leaf:</b> 1, allowing all features to be considered for the best split with unlimited number of leaf nodes, no maximum number of samples for bootstrapping.
Ridge	$\alpha$ : 1, <b>tolerance stopping criteria:</b> $10^{-3}$ .
Stochastic Gradient Descent	<b>Loss function:</b> squared error with L2 regularization, $\alpha = 10^{-4}$ , <b>L1 ratio:</b> 0.15, <b>maximum number of passes over the training data:</b> $10^3$ , <b>the stopping criterion for loss:</b> $10^{-3}$ , data shuffling after each epoch, <b>the initial learning rate:</b> $10^{-2}$ , <b>the exponent for inverse scaling learning rate:</b> 0.25, 10% of training data is the validation set for early stopping with 5 iterations with no improvement termination.
Support Vector Machine	<b>Kernel:</b> Radial Basis Function, $\gamma$ : $1 / ( \mathcal{F}  \cdot \sigma^2(T))$ , where $\mathcal{F}$ denotes the number of features, and $\sigma^2(T)$ is the variance of the training set, <b>tolerance for stopping criterion:</b> $10^{-3}$ , <b>regularization parameter C:</b> 1, $\epsilon$ : $10^{-1}$ , where $\epsilon$ specifies the epsilon-tube within which no penalty is associated in the training loss function with points predicted within a distance $\epsilon$ from the actual value.

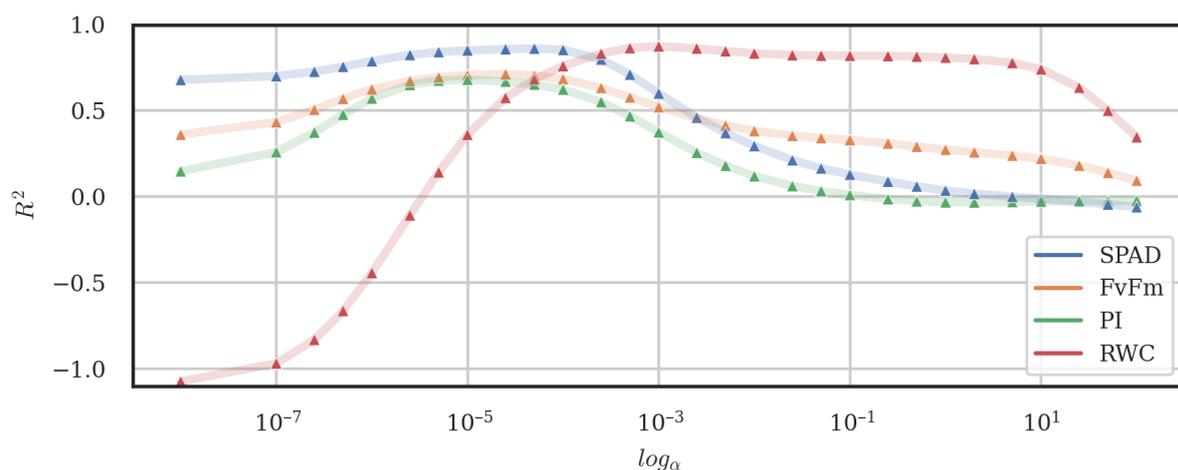
In Table 5, we gather the results obtained for all investigated parameters (SPAD, FvFm, PI and RWC) using top-3 machine learning models (with default parameterization), according to the  $R^2$  metric, being the most widely utilized quality measure in precision agriculture. We can appreciate that Linear Regression allows for obtaining the best coefficients of determination for SPAD, FvFm and PI, which are significantly larger than those achieved by the second-best algorithm ( $R^2$  smaller by 0.099, 0.110, and 0.288 than for Extreme Gradient Boosting, Gradient Boosting and Extra Trees for SPAD, FvFm and PI, respectively). For RWC, this linear model resulted in  $R^2$  of 0.720 (it was ranked fourth), and it was outperformed by the non-linear regression techniques. The results indicate that building heterogeneous regression ensembles, capturing both linear and non-linear models [57], may further improve the overall quality of estimating the parameters of interest.

To show that a selected model can be further improved, we employed a classic L2 regularization to the Ridge regression model, as it outperformed the other techniques for RWC (note that it failed to deliver high-quality prediction for other parameters, as presented later). A similar application of a regularized model was utilized to estimate the chlorophyll concentration by Lin and Lin [58], and it was shown to be effective in enhancing the algorithm's generalization capabilities. In Figure 5, we present the  $R^2$  values over the

test sets (for each parameter) obtained for a range of the  $\alpha$  hyperparameter values which controls the regularization strength (the larger  $\alpha$  becomes, the stronger regularization is). We can observe that for a fine-tuned  $\alpha$  parameter, we can significantly improve the model's performance for all chlorophyll parameters. Ridge regression with L2 regularization (with  $\alpha$  of  $2.5 \times 10^{-5}$ ,  $5 \times 10^{-5}$ ,  $10^{-11}$  and  $10^{-3}$  for SPAD, FvFm, PI and RWC, respectively) not only did provide statistically-significant improvements over the baseline Ridge model for all parameters, but also outperformed the other investigated models in 3/4 chlorophyll parameters (Table 6). Here, only for PI, the results were the same as those obtained by Linear Regression ( $R^2$  of 0.667). Thus, further improvements of the machine learning models, including the optimization of their hyperparameters or selection of appropriate training and/or feature sets, can easily lead to better regressors which may be confronted with the other techniques using our validation procedure in an unbiased and fair way.

**Table 5.** Three best machine learning models (according to  $R^2$ ) with default parameterization for each ground-truth parameter. We indicate if the metric should be minimized ( $\downarrow$ ) or maximized ( $\uparrow$ ).

Param.	Model	$R^2 \uparrow$	MAPE $\downarrow$	MSE $\downarrow$	MAE $\downarrow$
SPAD	Linear Regression	0.818	0.072	9.583	1.569
	Extreme Gradient Boosting	0.719	0.092	14.808	2.784
	AdaBoost	0.698	0.080	15.935	1.814
FvFm	Linear Regression	0.718	0.037	0.001	0.021
	Gradient Boosting	0.608	0.037	0.001	0.017
	AdaBoost	0.600	0.037	0.001	0.016
PI	Linear Regression	0.667	0.532	0.169	0.280
	Extra Trees	0.379	1.251	0.315	0.368
	Random Forest	0.213	1.594	0.400	0.470
RWC	Ridge	0.817	0.014	2.249	1.207
	Extreme Gradient Boosting	0.793	0.014	2.541	1.021
	Support Vector Machine	0.745	0.017	3.127	1.541



**Figure 5.** The Ridge regression results ( $R^2$  over the test sets for each parameter) obtained using a range of the  $\alpha$  hyperparameter values.

**Table 6.** The Ridge regression model with the L2 regularization.

Param.	$\alpha$	$R^2 \uparrow$	MAPE $\downarrow$	MSE $\downarrow$	MAE $\downarrow$
SPAD	$2.5 \times 10^{-5}$	0.827	0.072	9.095	1.625
FvFm	$5 \times 10^{-5}$	0.727	0.036	0.001	0.021
PI	$10^{-11}$	0.667	0.532	0.169	0.280
RWC	$10^{-3}$	0.859	0.013	1.731	0.941

Finally, in Table 7, we present the experimental results obtained for all investigated parameters using all regression models. Here, the best results obtained using the models with default parameterization are **boldfaced in black**, whereas the Ridge regression model with additional L2 regularization is rendered in **green** (we discuss the process of improving a selected baseline model to enhance its capabilities later in this section). Finally, if the Ridge regression model with regularization led to obtaining the globally-best metric value (when compared with other techniques), we **boldfaced and underlined** the corresponding entry. The results indeed show that it is possible to enhance the generalization capabilities of the “default” machine learning models.

**Table 7.** Performance of the investigated machine learning models for each chlorophyll-related parameter of interest. The best results obtained using the baseline machine learning models are **boldfaced in black**, whereas the model with further regularization is indicated in **green** (if it resulted in the best metric value across all models, the values are boldfaced and underlined). We indicate if the metric should be minimized ( $\downarrow$ ) or maximized ( $\uparrow$ ).

Parameter	Model	$R^2 \uparrow$	MAPE $\downarrow$	MSE $\downarrow$	MAE $\downarrow$
SPAD	AdaBoost	0.698	0.080	15.935	1.814
	Decision Tree	0.178	0.132	43.324	2.640
	ExtraTrees	0.649	0.099	18.494	2.307
	Extreme Gradient Boosting	0.719	0.092	14.808	2.784
	Gradient Boosting	0.639	0.095	19.025	2.059
	Kernel Ridge	0.132	0.179	45.768	4.787
	Lasso	−0.091	0.213	57.509	6.045
	Light Gradient Boosting Machine	0.180	0.178	43.255	5.016
	Linear Regression	<b>0.818</b>	<b>0.072</b>	<b>9.583</b>	<b>1.569</b>
	Linear Support Vector Machine	0.216	0.131	41.321	3.242
	Nu Support Vector Machine	−0.142	0.219	60.211	6.065
	Random Forest	0.587	0.123	21.760	3.029
	Ridge	0.037	0.202	50.789	6.153
	<b>Ridge with L2</b>	<b>0.827</b>	<b>0.072</b>	<b>9.095</b>	<b>1.625</b>
FvFm	Stochastic Gradient Descent	0.119	0.187	46.465	5.794
	Support Vector Machine	−0.289	0.231	67.984	6.831
	AdaBoost	0.600	<b>0.037</b>	<b>0.001</b>	0.016
	Decision Tree	−0.005	0.055	0.003	0.020
	ExtraTrees	0.136	0.049	0.003	<b>0.016</b>
	Extreme Gradient Boosting	0.407	0.044	0.002	0.019
	Gradient Boosting	0.608	0.037	<b>0.001</b>	0.017
	Kernel Ridge	−1.609	0.100	0.009	0.054
	Lasso	−0.002	0.061	0.003	0.028
	Light Gradient Boosting Machine	−0.002	0.061	0.003	0.028
	Linear Regression	<b>0.718</b>	0.037	<b>0.001</b>	0.021
	Linear Support Vector Machine	0.363	0.048	0.002	0.022
	Nu Support Vector Machine	0.316	0.047	0.002	0.022
	Random Forest	0.510	0.043	0.002	0.019
Ridge	0.272	0.051	0.003	0.025	
<b>Ridge with L2</b>	<b>0.727</b>	<b>0.036</b>	<b>0.001</b>	<b>0.021</b>	
Stochastic Gradient Descent	−4.424	0.155	0.019	0.102	
Support Vector Machine	−0.112	0.070	0.004	0.034	

Table 7. Cont.

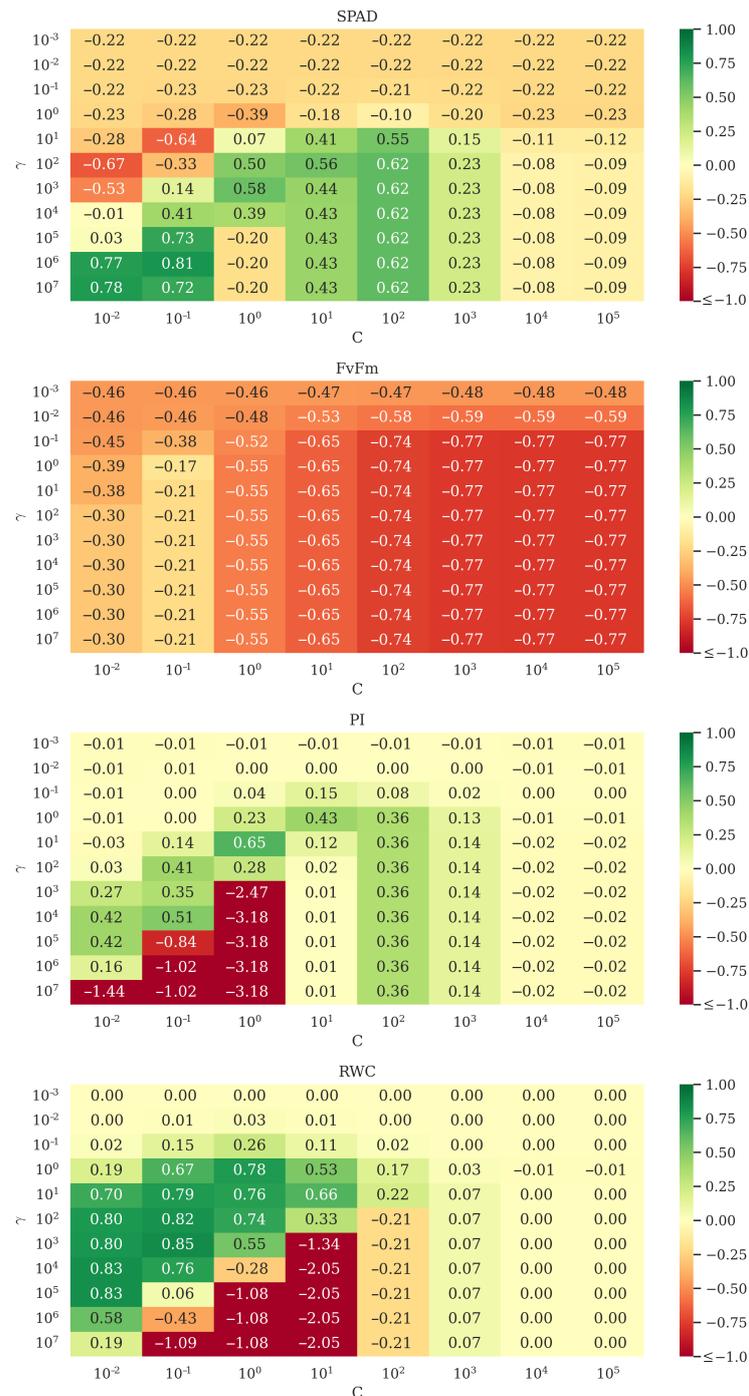
Parameter	Model	$R^2 \uparrow$	MAPE $\downarrow$	MSE $\downarrow$	MAE $\downarrow$
PI	AdaBoost	−0.067	1.825	0.542	<b>0.262</b>
	Decision Tree	−0.150	0.955	0.584	0.350
	Extra Tree	0.379	1.251	0.315	0.368
	Extreme Gradient Boosting	0.168	1.407	0.423	0.388
	Gradient Boosting	0.114	1.508	0.450	0.296
	Kernel Ridge	−0.147	1.834	0.583	0.588
	Lasso	−0.020	1.941	0.518	0.502
	Light Gradient Boosting Machine	−0.010	1.623	0.513	0.456
	Linear Regression	<b>0.667</b>	<b>0.532</b>	<b>0.169</b>	0.280
	Linear Support Vector Machine	−0.133	1.334	0.576	0.465
	Nu Support Vector Machine	−0.010	1.529	0.513	0.634
	Random Forest	0.213	1.594	0.400	0.470
	Ridge	−0.036	1.993	0.526	0.442
	<b>Ridge with L2</b>	<b>0.667</b>	<b>0.532</b>	<b>0.169</b>	<b>0.280</b>
	Stochastic Gradient Descent	−0.108	1.550	0.563	0.601
Support Vector Machine	0.019	1.818	0.499	0.506	
RWC	AdaBoost	0.657	0.018	4.213	1.233
	Decision Tree	0.432	0.025	6.977	1.900
	Extra Tree	0.704	0.018	3.641	1.090
	Extreme Gradient Boosting	0.793	<b>0.014</b>	2.541	<b>1.021</b>
	Gradient Boosting	0.646	0.020	4.350	1.250
	Kernel Ridge	−4.435	0.074	66.776	4.045
	Lasso	0.000	0.036	12.288	3.549
	Light Gradient Boosting Machine	0.000	0.036	12.288	3.549
	Linear Regression	0.720	0.018	3.440	1.385
	Linear Support Vector Machine	−14.497	0.125	190.396	7.235
	Nu Support Vector Machine	0.695	0.019	3.745	1.610
	Random Forest	0.709	0.018	3.581	1.499
	Ridge	<b>0.817</b>	<b>0.014</b>	<b>2.249</b>	1.207
	<b>Ridge with L2</b>	<b>0.859</b>	<b>0.013</b>	<b>1.731</b>	<b>0.941</b>
	Stochastic Gradient Descent	−1.491	0.051	30.598	3.334
Support Vector Machine	0.745	0.017	3.127	1.541	

## 5. Conclusions and Future Work

Capturing the information concerning the chlorophyll level in leaves is an important practical issue in precision agriculture, as it helps practitioners optimize their operation and appropriately plan and monitor the treatment process of various plants. Although there exist in-field methods that allow us to recognize the actual chlorophyll level through elaborating an array of indicators, they are invasive, time-inefficient and lack scalability. Therefore, developing non-invasive approaches benefiting from the detailed information available in hyperspectral imaging has attracted research attention. However, the emerging data-driven algorithms for this task are commonly evaluated using private datasets without any standardized validation procedure, which makes their comparison with the current state of the art virtually impossible. In this paper, we tackled this issue and proposed an open dataset (coupling HSI with in situ ground-truth measurements), together with its training-test splits and quality metrics that can be used to confront the emerging and existing techniques in a fully-unbiased and fair way. Our experimental study not only provided a solid baseline obtained using 15 classic machine learning predictors, but we also showed that it is possible to enhance such models to improve their generalizability. We believe that our work may constitute an important step toward standardizing the way we compare the chlorophyll-analysis algorithms and may help consolidate state of the art in the field by providing a clear way of comparing new approaches over real data.

Our work is an interesting point of departure for further research. In this work, we did not intend to introduce a new, “ground-breaking” algorithm for estimating the chlorophyll level from HSI. There are, however, immediate next steps which should be

performed to improve the performance capabilities of machine learning models, with the hyperparameter optimization being one of them. In Figure 6, we show how two most important hyperparameters of Support Vector Machines (C and  $\gamma$ ) affect their performance. Observing the results of the models optimized for each target parameter separately and gathered in Table 8, we can appreciate that the grid-searched Support Vector Machines significantly outperformed their default parameterization (Table 7). Therefore, optimizing the most important hyperparameters of other techniques would likely lead to their noticeable improvements.



**Figure 6.** The results elaborated using a Support Vector Machine over the test sets for each parameter, and obtained using a range of the C and  $\gamma$  hyperparameter values.

**Table 8.** The results elaborated using a Support Vector Machine with the hyperparameters optimized for each parameter separately. We indicate if the metric should be minimized ( $\downarrow$ ) or maximized ( $\uparrow$ ).

Param.	$\gamma$	$C$	$R^2 \uparrow$	MAPE $\downarrow$	MSE $\downarrow$	MAE $\downarrow$
SPAD	$10^6$	$10^{-1}$	0.808	0.073	10.136	2.401
FvFm	$10^0$	$10^{-1}$	-0.169	0.076	0.004	0.052
PI	$10^1$	$10^0$	0.652	1.157	0.177	0.372
RWC	$10^3$	$10^{-1}$	0.845	0.016	1.904	1.006

We are observing an unprecedented success of deep learning in HSI analysis—such techniques may certainly improve the quality of the estimated chlorophyll-related parameters [15]. To unchain the full scalability potential of HSI, we can perform the analysis process on board a satellite to extract knowledge from raw pixels. However, the algorithms to be deployed in such hardware-constrained execution environments should be resource-frugal and robust against various noise affecting the in-orbit image acquisition [8] and must be thoroughly validated before they can run in space [59]. Developing the (deep) machine learning models for on-board processing is currently widely-explored due to a significant number of emerging Earth observation missions, including our Intuition-1 satellite.

**Author Contributions:** Conceptualization, B.R. and J.N.; Methodology, B.R., J.N. and A.M.W.; Software, B.R.; Validation, B.R. and J.N.; Formal Analysis, B.R.; Literature Review, A.M.W., B.R. and J.N.; Investigation, B.R.; Resources, B.R.; Data Curation, B.R.; Writing—Original Draft Preparation, B.R., J.N. and A.M.W.; Writing—Review and Editing, J.N., B.R. and A.M.W.; Visualization, B.R. and A.M.W.; Supervision, B.R. and J.N.; Funding Acquisition, J.N. All authors have read and agreed to the submitted version of the manuscript.

**Funding:** A.M.W. and J.N. were supported by the Silesian University of Technology grants for maintaining and developing research potential (A.M.W.: 07/010/BKM22/1017). This work was partially supported by The National Centre for Research and Development of Poland under project POIR.04.01.04-00-0009/19.

**Data Availability Statement:** Publicly available dataset was analyzed in this study. This data can be found here: <https://doi.org/10.1016/j.dib.2022.108087> and here: <https://data.mendeley.com/datasets/xn2wy75f8m/> (accessed on 1 November 2022).

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

## Abbreviations

The following abbreviations are used in this manuscript:

AGB	Above-ground biomass
ALOS	Advanced Land Observing Satellite
CHESS	CHlorophyll ESTimation DataSet
CHRIS	Compact High Resolution Imaging Spectrometer
CWT	Continuous wavelet transform
DNN	Deep neural network
EFFIS	The European Forest Fire Information System
EVI	Enhanced vegetation index
FvFm	The maximum quantum yield of the PSII photochemistry
GBDT	Gradient boosting decision tree
GBVI	Green brown vegetation index
GSD	Ground sample distance
GPR	Gaussian process regression
HSI	Hyperspectral image
$k$ -NN	$k$ -nearest neighbor

LAI	Leaf area index
LSWI	Land surface water index
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MLR	Multiple linear regression
MSE	Mean squared error
MSI	Multispectral image
NBSI	Non-binary snow index
NDVI	Normalized difference vegetation index
NIR	Near-infrared spectra
NIRv	Near-infrared reflectance of vegetation
PALSAR	Phased Array L-band Synthetic Aperture Radar
PCA	Principal component analysis
PI	The index for energy conservation from photons absorbed by PSII to PSI electron acceptors
$R^2$	Coefficient of determination
RDP	Ratio of the performance to deviation
RF	Random forest
RFE	Recursive-feature-elimination
RMSE	Root mean squared error
RRMSE	Relative root mean squared error
RWC	Relative water content measurements for the sampled canopy
SAVI	Soil-adjusted vegetation index
SAR	Synthetic-aperture radar
SOC	Soil organic carbon
SPAD	The actual level of chlorophyll utilizing the soil-plant analysis development
SVD	Singular value decomposition
SVM	Support vector machine
VHGPR	Variational heteroscedastic GPR
VIS-NIR	Visible–near-infrared spectra
VH	Vertical transmitted and horizontal received
VV	Vertical transmitted and vertical received polarization
XGB	Extreme gradient boosting

## References

- Shen, Q.; Lin, J.; Yang, J.; Zhao, W.; Wu, J. Exploring the Potential of Spatially Downscaled Solar-Induced Chlorophyll Fluorescence to Monitor Drought Effects on Gross Primary Production in Winter Wheat. *IEEE J-STARS* **2022**, *15*, 2012–2022. [[CrossRef](#)]
- Long, Y.; Ma, M. Recognition of Drought Stress State of Tomato Seedling Based on Chlorophyll Fluorescence Imaging. *IEEE Access* **2022**, *10*, 48633–48642. [[CrossRef](#)]
- Oláh, V.; Hepp, A.; Irfan, M.; Mészáros, I. Chlorophyll Fluorescence Imaging-Based Duckweed Phenotyping to Assess Acute Phytotoxic Effects. *Plants* **2021**, *10*, 2763. [[CrossRef](#)] [[PubMed](#)]
- Lazzeri, G.; Frodella, W.; Rossi, G.; Moretti, S. Multitemporal Mapping of Post-Fire Land Cover Using Multiplatform PRISMA Hyperspectral and Sentinel-UAV Multispectral Data: Insights from Case Studies in Portugal and Italy. *Sensors* **2021**, *21*. [[CrossRef](#)] [[PubMed](#)]
- Pyo, J.; Duan, H.; Baek, S.; Kim, M.S.; Jeon, T.; Kwon, Y.S.; Lee, H.; Cho, K.H. A Convolutional Neural Network Regression for Quantifying Cyanobacteria Using Hyperspectral Imagery. *Remote Sens. Environ.* **2019**, *233*, 111350. [[CrossRef](#)]
- Hill, P.R.; Kumar, A.; Temimi, M.; Bull, D.R. HABNet: Machine Learning, Remote Sensing-Based Detection of Harmful Algal Blooms. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 3229–3239. [[CrossRef](#)]
- Torres Palenzuela, J.M.; Vilas, L.G.; Bellas Aláez, F.M.; Pazos, Y. Potential Application of the New Sentinel Satellites for Monitoring of Harmful Algal Blooms in the Galician Aquaculture. *Thalass. Int. J. Mar. Sci.* **2020**, *36*, 85–93. [[CrossRef](#)]
- Nalepa, J.; Myller, M.; Cwiek, M.; Zak, L.; Lakota, T.; Tulczyjew, L.; Kawulok, M. Towards On-Board Hyperspectral Satellite Image Segmentation: Understanding Robustness of Deep Learning through Simulating Acquisition Conditions. *Remote Sens.* **2021**, *13*, 1532. [[CrossRef](#)]
- Liu, N.; Liu, G.; Sun, H. Real-Time Detection on SPAD Value of Potato Plant Using an In-Field Spectral Imaging Sensor System. *Sensors* **2020**, *20*, 3430. [[CrossRef](#)]
- Haboudane, D.; Miller, J.R.; Tremblay, N.; Zarco-Tejada, P.J.; Dextraze, L. Integrated Narrow-Band Vegetation Indices for Prediction of Crop Chlorophyll Content for Application to Precision Agriculture. *Remote Sens. Environ.* **2002**, *81*, 416–426. [[CrossRef](#)]

11. Hai-ling, J.; Li-fu, Z.; Hang, Y.; Xiao-ping, C.; Shu-dong, W.; Xue-ke, L.; Kai, L. Comparison of Accuracy and Stability of Estimating Winter Wheat Chlorophyll Content Based on Spectral Indices. In Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014; pp. 2985–2988. [[CrossRef](#)]
12. Raya-Sereno, M.D.; Alonso-Ayuso, M.; Pancorbo, J.L.; Gabriel, J.L.; Camino, C.; Zarco-Tejada, P.J.; Quemada, M. Residual Effect and N Fertilizer Rate Detection by High-Resolution VNIR-SWIR Hyperspectral Imagery and Solar-Induced Chlorophyll Fluorescence in Wheat. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–17. [[CrossRef](#)]
13. Wang, J.; Zhou, Q.; Shang, J.; Liu, C.; Zhuang, T.; Ding, J.; Xian, Y.; Zhao, L.; Wang, W.; Zhou, G.; et al. UAV- and Machine Learning-Based Retrieval of Wheat SPAD Values at the Overwintering Stage for Variety Screening. *Remote Sens.* **2021**, *13*, 5166. [[CrossRef](#)]
14. Yuan, Z.; Ye, Y.; Wei, L.; Yang, X.; Huang, C. Study on the Optimization of Hyperspectral Characteristic Bands Combined with Monitoring and Visualization of Pepper Leaf SPAD Value. *Sensors* **2021**, *22*, 183. [[CrossRef](#)] [[PubMed](#)]
15. Ye, H.; Tang, S.; Yang, C. Deep Learning for Chlorophyll-a Concentration Retrieval: A Case Study for the Pearl River Estuary. *Remote Sens.* **2021**, *13*, 3717. [[CrossRef](#)]
16. Tulczyjew, L.; Kawulok, M.; Longépé, N.; Le Saux, B.; Nalepa, J. A Multibranch Convolutional Neural Network for Hyperspectral Unmixing. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [[CrossRef](#)]
17. Kapoor, S.; Narayanan, A. Leakage and the Reproducibility Crisis in ML-based Science. *arXiv* **2022**, arXiv:2207.07048. [[CrossRef](#)].
18. Inoue, Y.; Guérif, M.; Baret, F.; Skidmore, A.; Gitelson, A.; Schlerf, M.; Darvishzadeh, R.; Olioso, A. Simple and Robust Methods for Remote Sensing of Canopy Chlorophyll Content: A Comparative Analysis of Hyperspectral Data for Different Types of Vegetation. *Plant Cell Environ.* **2016**, *39*, 2609–2623. [[CrossRef](#)]
19. Mayranti, F.P.; Saputro, A.H.; Handayani, W. Chlorophyll A and B Content Measurement System of Velvet Apple Leaf in Hyperspectral Imaging. In Proceedings of the ICICOS, Semarang, Indonesia, 29–30 October 2019; pp. 1–5. [[CrossRef](#)]
20. Tomaszewski, M.; Gasz, R.; Smykała, K. Monitoring Vegetation Changes Using Satellite Imaging—NDVI and RVI4S1 Indicators. In Proceedings of the Control, Computer Engineering and Neuroscience, Opole, Poland, 21 September 2021; Paszkiel, S., Ed.; Springer: Berlin/Heidelberg, Germany, 2021; pp. 268–278. [[CrossRef](#)]
21. Bannari, A.; Khurshid, K.S.; Staenz, K.; Schwarz, J.W. A Comparison of Hyperspectral Chlorophyll Indices for Wheat Crop Chlorophyll Content Estimation Using Laboratory Reflectance Measurements. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3063–3074. [[CrossRef](#)]
22. El-Hendawy, S.; Dewir, Y.H.; Elsayed, S.; Schmidhalter, U.; Al-Gaadi, K.; Tola, E.; Refay, Y.; Tahir, M.U.; Hassan, W.M. Combining Hyperspectral Reflectance Indices and Multivariate Analysis to Estimate Different Units of Chlorophyll Content of Spring Wheat under Salinity Conditions. *Plants* **2022**, *11*, 456. [[CrossRef](#)] [[PubMed](#)]
23. Middleton, E.M.; Julitta, T.; Campbell, P.E.; Huemmrich, K.F.; Schickling, A.; Rossini, M.; Cogliati, S.; Landis, D.R.; Alonso, L. Novel Leaf-Level Measurements of Chlorophyll Fluorescence for Photosynthetic Efficiency. In Proceedings of the IGARSS, Milan, Italy, 26–31 July 2015; pp. 3878–3881. [[CrossRef](#)]
24. Jia, M.; Zhou, C.; Cheng, T.; Tian, Y.; Zhu, Y.; Cao, W.; Yao, X. Inversion of Chlorophyll Fluorescence Parameters on Vegetation Indices at Leaf Scale. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 4359–4362. [[CrossRef](#)]
25. Nalepa, J.; Myller, M.; Kawulok, M. Validating Hyperspectral Image Segmentation. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1264–1268. [[CrossRef](#)]
26. Singh, H.; Kumar, D.; Soni, V. Performance of Chlorophyll a Fluorescence Parameters in Lemna Minor under Heavy Metal Stress Induced by Various Concentration of Copper. *Sci. Rep.* **2022**, *12*, 10620. [[CrossRef](#)] [[PubMed](#)]
27. Yue, J.; Zhou, C.; Guo, W.; Feng, H.; Xu, K. Estimation of Winter-Wheat Above-Ground Biomass Using the Wavelet Analysis of Unmanned Aerial Vehicle-Based Digital Images and Hyperspectral Crop Canopy Images. *Int. J. Remote Sens.* **2021**, *42*, 1602–1622. [[CrossRef](#)]
28. Jin, X.; Li, Z.; Feng, H.; Ren, Z.; Li, S. Deep Neural Network Algorithm for Estimating Maize Biomass Based on Simulated Sentinel 2A Vegetation Indices and Leaf Area Index. *Crop J.* **2020**, *8*, 87–97. [[CrossRef](#)]
29. Lu, B.; He, Y. Evaluating Empirical Regression, Machine Learning, and Radiative Transfer Modelling for Estimating Vegetation Chlorophyll Content Using Bi-Seasonal Hyperspectral Images. *Remote Sens.* **2019**, *11*, 1979. [[CrossRef](#)]
30. Adão, T.; Hruška, J.; Pádua, L.; Bessa, J.; Peres, E.; Morais, R.; Sousa, J.J. Hyperspectral Imaging: A Review on UAV-Based Sensors, Data Processing and Applications for Agriculture and Forestry. *Remote Sens.* **2017**, *9*, 1110. [[CrossRef](#)]
31. Brewer, K.; Clulow, A.; Sibanda, M.; Gokool, S.; Naiken, V.; Mabhaudhi, T. Predicting the Chlorophyll Content of Maize over Phenotyping as a Proxy for Crop Health in Smallholder Farming Systems. *Remote Sens.* **2022**, *14*, 518. [[CrossRef](#)]
32. Lu, B.; Dao, P.D.; Liu, J.; He, Y.; Shang, J. Recent Advances of Hyperspectral Imaging Technology and Applications in Agriculture. *Remote Sens.* **2020**, *12*, 2659. [[CrossRef](#)]
33. Meng, X.; Bao, Y.; Liu, J.; Liu, H.; Zhang, X.; Zhang, Y.; Wang, P.; Tang, H.; Kong, F. Regional Soil Organic Carbon Prediction Model Based on a Discrete Wavelet Analysis of Hyperspectral Satellite Data. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *89*, 102111. [[CrossRef](#)]
34. Hong, Y.; Chen, S.; Chen, Y.; Linderman, M.; Mouazen, A.M.; Liu, Y.; Guo, L.; Yu, L.; Liu, Y.; Cheng, H.; et al. Comparing Laboratory and Airborne Hyperspectral Data for the Estimation and Mapping of Topsoil Organic Carbon: Feature Selection Coupled with Random Forest. *Soil Tillage Res.* **2020**, *199*, 104589. [[CrossRef](#)]

35. Zhang, Y.; Xia, C.; Zhang, X.; Cheng, X.; Feng, G.; Wang, Y.; Gao, Q. Estimating the Maize Biomass by Crop Height and Narrowband Vegetation Indices Derived from UAV-based Hyperspectral Images. *Ecol. Indic.* **2021**, *129*, 107985. [[CrossRef](#)]
36. Han, L.; Yang, G.; Dai, H.; Xu, B.; Yang, H.; Feng, H.; Li, Z.; Yang, X. Modeling Maize Above-Ground Biomass Based on Machine Learning Approaches Using UAV Remote-Sensing Data. *Plant Methods* **2019**, *15*, 1746–4811. [[CrossRef](#)] [[PubMed](#)]
37. Ji, S.; Zhang, C.; Xu, A.; Shi, Y.; Duan, Y. 3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images. *Remote Sens.* **2018**, *10*, 75. [[CrossRef](#)]
38. Cui, Z.; Kerekes, J.P. Potential of Red Edge Spectral Bands in Future Landsat Satellites on Agroecosystem Canopy Green Leaf Area Index Retrieval. *Remote Sens.* **2018**, *10*, 1458. [[CrossRef](#)]
39. Zhang, F.; Zhou, G. Estimation of Vegetation Water Content using Hyperspectral Vegetation Indices: A Comparison of Crop Water Indicators in Response to Water Stress Treatments for Summer Maize. *BMC Ecol.* **2019**, *19*, 1–12. [[CrossRef](#)]
40. Mansaray, L.R.; Zhang, K.; Kanu, A.S. Dry Biomass Estimation of Paddy Rice With Sentinel-1A Satellite Data Using Machine Learning Regression Algorithms. *Comput. Electron. Agric.* **2020**, *176*, 105674. [[CrossRef](#)]
41. Wang, J.; Xiao, X.; Bajgain, R.; Starks, P.; Steiner, J.; Doughty, R.B.; Chang, Q. Estimating Leaf Area Index and Aboveground Biomass of Grazing Pastures Using Sentinel-1, Sentinel-2 and Landsat Images. *ISPRS J. Photogramm. Remote Sens.* **2019**, *154*, 189–201. [[CrossRef](#)]
42. Guo, H.; Liu, J.; Xiao, Z.; Xiao, L. Deep CNN-based Hyperspectral Image Classification Using Discriminative Multiple Spectral-Feature Fusion. *Remote Sens. Lett.* **2020**, *11*, 827–836. [[CrossRef](#)]
43. Marshall, M.; Thenkabail, P. Advantage of Hyperspectral EO-1 Hyperion over Multispectral IKOMOS, Geoeye-1, Worldview-2, Landsat ETM+, and MODIS Vegetation Indices in Crop Biomass Estimation. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 205–218. [[CrossRef](#)]
44. Ribalta Lorenzo, P.; Tulczyjew, L.; Marcinkiewicz, M.; Nalepa, J. Hyperspectral Band Selection Using Attention-Based Convolutional Neural Networks. *IEEE Access* **2020**, *8*, 42384–42403. [[CrossRef](#)]
45. Zheng, Q.; Ye, H.; Huang, W.; Dong, Y.; Jiang, H.; Wang, C.; Li, D.; Wang, L.; Chen, S. Integrating Spectral Information and Meteorological Data to Monitor Wheat Yellow Rust at a Regional Scale: A Case Study. *Remote Sens.* **2021**, *13*, 278. [[CrossRef](#)]
46. Rao, K.; Williams, A.P.; Flefil, J.F.; Konings, A.G. SAR-enhanced Mapping of Live Fuel Moisture Content. *Remote Sens. Environ.* **2020**, *245*, 111797. [[CrossRef](#)]
47. Estévez, J.; Vicent, J.; Rivera-Caicedo, J.P.; Morcillo-Pallarés, P.; Vuolo, F.; Sabater, N.; Camps-Valls, G.; Moreno, J.; Verrelst, J. Gaussian Processes Retrieval of LAI from Sentinel-2 Top-of-atmosphere Radiance Data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 289–304. [[CrossRef](#)] [[PubMed](#)]
48. Wang, X.; Zhang, Y.; Atkinson, P.M.; Yao, H. Predicting Soil Organic Carbon Content in Spain by Combining Landsat TM and ALOS PALSAR Images. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *92*, 102182. [[CrossRef](#)]
49. Battude, M.; Al Bitar, A.; Morin, D.; Cros, J.; Huc, M.; Marais Sicre, C.; Le Dantec, V.; Demarez, V. Estimating Maize Biomass and Yield over Large Areas Using High Spatial and Temporal Resolution Sentinel-2 like Remote Sensing Data. *Remote Sens. Environ.* **2016**, *184*, 668–681. [[CrossRef](#)]
50. Kalaji, H.M.; Račková, L.; Paganová, V.; Swoczyna, T.; Rusinowski, S.; Sitko, K. Can Chlorophyll-a Fluorescence Parameters Be Used as Bio-indicators to Distinguish Between Drought and Salinity Stress in *Tilia Cordata* Mill? *Environ. Exp. Bot.* **2018**, *152*, 149–157. [[CrossRef](#)]
51. Li, C.; Chen, P.; Ma, C.; Feng, H.; Wei, F.; Wang, Y.; Shi, J.; Cui, Y. Estimation of potato chlorophyll content using composite hyperspectral index parameters collected by an unmanned aerial vehicle. *Int. J. Remote Sens.* **2020**, *41*, 8176–8197. [[CrossRef](#)]
52. Liu, N.; Xing, Z.; Zhao, R.; Qiao, L.; Li, M.; Liu, G.; Sun, H. Analysis of Chlorophyll Concentration in Potato Crop by Coupling Continuous Wavelet Transform and Spectral Variable Optimization. *Remote Sens.* **2020**, *12*, 2826. [[CrossRef](#)]
53. Yang, H.; Hu, Y.; Zheng, Z.; Qiao, Y.; Zhang, K.; Guo, T.; Chen, J. Estimation of Potato Chlorophyll Content from UAV Multispectral Images with Stacking Ensemble Algorithm. *Agronomy* **2022**, *12*, 2318. [[CrossRef](#)]
54. Ruszczak, B.; Boguszewska-Mańkowska, D. Deep potato—The Hyperspectral Imagery of Potato Cultivation with Reference Agronomic Measurements Dataset: Towards Potato Physiological Features Modeling. *Data Brief* **2022**, *42*, 108087. [[CrossRef](#)]
55. Chicco, D.; Warrens, M.J.; Jurman, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Sci.* **2021**, *7*, e623. [[CrossRef](#)]
56. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
57. Nalepa, J.; Myller, M.; Tulczyjew, L.; Kawulok, M. Deep Ensembles for Hyperspectral Image Data Classification and Unmixing. *Remote Sens.* **2021**, *13*, 4133. [[CrossRef](#)]
58. Lin, C.Y.; Lin, C. Using Ridge Regression Method to Reduce Estimation Uncertainty in Chlorophyll Models Based on Worldview Multispectral Data. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 1777–1780. [[CrossRef](#)]
59. Ziája, M.; Bosowski, P.; Myller, M.; Gajoch, G.; Gumiela, M.; Protich, J.; Borda, K.; Jayaraman, D.; Dividino, R.; Nalepa, J. Benchmarking Deep Learning for On-Board Space Applications. *Remote Sens.* **2021**, *13*, 3981. [[CrossRef](#)]