JYX
JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Karila, Kirsi; Alves Oliveira, Raquel; Ek, Johannes; Kaivosoja, Jere; Koivumäki, Niko; Korhonen, Panu; Niemeläinen, Oiva; Nyholm, Laura; Näsi, Roope; Pölönen, Ilkka; Honkavaara, Eija

Title: Estimating Grass Sward Quality and Quantity Parameters Using Drone Remote Sensing with Deep Neural Networks

Year: 2022

Version: Published version

Copyright: © 2022 The Author(s).

Rights: CC BY 4.0

Rights url: https://creativecommons.org/licenses/by/4.0/

Please cite the original version:

Karila, K., Alves Oliveira, R., Ek, J., Kaivosoja, J., Koivumäki, N., Korhonen, P., Niemeläinen, O., Nyholm, L., Näsi, R., Pölönen, I., & Honkavaara, E. (2022). Estimating Grass Sward Quality and Quantity Parameters Using Drone Remote Sensing with Deep Neural Networks. Remote Sensing, 14(11), Article 2692. https://doi.org/10.3390/rs14112692

remote sensing

MDPI

Article

Estimating Grass Sward Quality and Quantity Parameters Using Drone Remote Sensing with Deep Neural Networks

Kirsi Karila [1,*], Raquel Alves Oliveira [1], Johannes Ek [2], Jere Kaivosoja [3], Niko Koivumäki [1], Panu Korhonen [3], Oiva Niemeläinen [3], Laura Nyholm [4], Roope Näsi [1], Ilkka Pölönen [5] and Eija Honkavaara [1]

1 Finnish Geospatial Research Institute (FGI), National Land Survey of Finland, 02150 Espoo, Finland; raquel.alvesdeoliveira@nls.fi (R.A.O.); niko.koivumaki@nls.fi (N.K.); roope.nasi@nls.fi (R.N.); eija.honkavaara@nls.fi (E.H.)
2 Department of Applied Physics, School of Science, Aalto University, 02150 Espoo, Finland; johannes.ek@aalto.fi
3 Natural Resources Institute Finland (Luke), 00790 Helsinki, Finland; jere.kaivosoja@luke.fi (J.K.); panu.korhonen@luke.fi (P.K.); oiva.niemelainen@luke.fi (O.N.)
4 Farm Services, Valio Ltd., 00370 Helsinki, Finland; laura.nyholm@valio.fi
5 Faculty of Information Technology, University of Jyväskylä, 40014 Jyväskylä, Finland; ilkka.polonen@jyu.fi
* Correspondence: kirsi.karila@nls.fi; Tel.: +358-50-409-3895

Abstract: The objective of this study is to investigate the potential of novel neural network architectures for measuring the quality and quantity parameters of silage grass swards, using drone RGB and hyperspectral images (HSI), and compare the results with the random forest (RF) method and handcrafted features. The parameters included fresh and dry biomass (FY, DMY), the digestibility of organic matter in dry matter (D-value), neutral detergent fiber (NDF), indigestible neutral detergent fiber (iNDF), water-soluble carbohydrates (WSC), nitrogen concentration (Ncont) and nitrogen uptake (NU); datasets from spring and summer growth were used. Deep pre-trained neural network architectures, the VGG16 and the Vision Transformer (ViT), and simple 2D and 3D convolutional neural networks (CNN) were studied. In most cases, the neural networks outperformed RF. The normalized root-mean-square errors (NRMSE) of the best models were for FY 19% (2104 kg/ha), DMY 21% (512 kg DM/ha), D-value 1.2% (8.6 g/kg DM), iNDF 12% (5.1 g/kg DM), NDF 1.1% (6.2 g/kg DM), WSC 10% (10.5 g/kg DM), Ncont 9% (2 g N/kg DM), and NU 22% (11.9 N kg/ha) using independent test dataset. The RGB data provided good results, particularly for the FY, DMY, WSC and NU. The HSI datasets provided advantages for some parameters. The ViT and VGG provided the best results with the RGB data, whereas the simple 3D-CNN was the most consistent with the HSI data.

Keywords: drone; remote sensing; hyperspectral; RGB; CNN; image transformer; silage production; grass sward

check for updates

Citation: Karila, K.; Alves Oliveira, R.; Ek, J.; Kaivosoja, J.; Koivumäki, N.; Korhonen, P.; Niemeläinen, O.; Nyholm, L.; Näsi, R.; Pölönen, I.; et al. Estimating Grass Sward Quality and Quantity Parameters Using Drone Remote Sensing with Deep Neural Networks. *Remote Sens.* **2022**, *14*, 2692. https://doi.org/10.3390/rs14112692

Academic Editor: Pablo Rodríguez-Gonzálvez

Received: 14 April 2022
Accepted: 30 May 2022
Published: 3 June 2022

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

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Timely estimates of grass sward biomass quantity and quality parameters are needed to produce high-quality feed for animals in dairy and beef production systems. Grasses accumulate biomass rapidly, especially in spring and summer, when approaching the optimal harvest time, which is accompanied by a rapid decrease in the digestibility of biomass [1,2]. Services based on weather data-based grass growth models with or without additional sampling of grass quality are provided for farmers to help in silage sward harvest time optimization (example of services: http://www.vallprognos.se/, http://www.karpe.fi/darvoennuste.php, accessed on 29 May 2022). Equations of one D-value prediction model are described, and its shortages are discussed by Hyrkäs et al. [3]. One main shortage is that the service models do not consider many factors affecting grass sward

yield quantity and quality, such as fertilizer application, age of stand and variability of edaphic factors within a field. Site-specific information of the grass sward biomass quantity and quality within a field would facilitate site-specific management of silage production. Currently, farmers optimize the timing of harvest by using visual approximation of sward growth stage (laborious on larger scale, prone to subjectivity), laboratory analyses of samples (delayed results) and D-value prediction models available online (inaccurate). Remote sensing of grass yield quantity and quality with drones and satellites opens up new possibilities to improve efficiency and accuracy of making timely decisions on optimal harvest times, and new envisioned online harvest optimization tools could utilize the best features of these approaches and scale those cost-effectively to cover larger geographical areas and different production systems.

Interest in utilizing satellite and drone remote sensing to support automated precision farming is growing rapidly. In drone applications, spectral imaging using hyper- and multispectral cameras equipped with specially designed spectral filters and RGB imaging using regular cameras [4] are potential technologies for vegetation remote sensing. Compared to hyper- and multispectral cameras, the advantages of regular RGB cameras include their high spatial resolution and signal-to-noise ratio; their potential disadvantage is the reduced spectral resolution. Furthermore, low-cost RGB cameras enable the implementation of cost-efficient monitoring solutions. Most grass-related studies have used multispectral or RGB cameras.

Machine learning analytics are the foundation for utilizing drones in precision agriculture. Various supervised techniques have been used to estimate grass quality and quantity parameters from drone imagery. Early studies used various indices and linear estimation techniques [5,6]. Machine learning-based approaches have evolved afterwards [7–11]; they extract various structural and/or spectral features from the remote sensing datasets to optimize the models so that they provide the best fit between the features and references.

Recently, deep learning technologies have revolutionized the performance of machine learning analytics. Conventional methods, such as Random Forest (RF) and support vector machines (SVM), use predefined handcrafted features for classification or regression tasks. Deep learning methods learn useful features from the data [12]. Deep neural networks (DNNs) are neural networks with several layers of trainable weights. Convolutional neural networks (CNNs) are neural networks with convolutional layers that are often used in image recognition and classification tasks [13]. The use of DNNs for regression from images has been reviewed by Lathuilière et al. [14]. Deep learning for aerial remote sensing data classification and segmentation has been reviewed by [15–18]. Usually, in image recognition tasks, 2D convolutional filters that operate in the spatial dimension are used in CNN. Often, remote sensing data is acquired with spectral sensors using more than three channels; for them, 3D-CNNs containing 3D convolutional filters take advantage of both spatial and spectral dimensions in hyperspectral data [19]. Vision transformers are new types of DNNs without convolutional layers [20]. Recently, transformers have been successfully used in remote sensing data classification [21,22]. Deep learning requires large training datasets to make the models reliable. Pretrained deep learning models are often designed for image classification and the pretraining datasets consist of millions of hand-annotated RGB images of objects. In transfer learning, the knowledge (i.e., model parameters) gained from the classification of big datasets is used or fine-tuned by additional training for other tasks, obtaining improved results [23,24].

Deep learning in vegetation remote sensing has been reviewed by Kattenborn et al. [25], in agriculture by Kamilaris and Prenafeta-Boldú [26] and in UAV remote sensing by Osco et al. [27]. RGB UAV images and DNNs for regression have been used for forage grass biomass [28], winter wheat biomass [29], and Guineagrass dry matter yield [30] estimation. RGB and 3D information were used in plant species cover fraction estimation in [31]. Yang et al. [32] used RGB and multispectral images to estimate rice grain yield.

Previous studies have shown that it is possible to estimate plants' quality parameters along with biomass or yield. Dvorak et al. [33] estimated the yield and nutritive value

of alfalfa throughout its growth cycle using photogrammetric point clouds; parameters included acid detergent fiber (ADF), neutral detergent fiber (NDF), and crude protein (CP). Gruner et al. [34] developed models for aboveground biomass and NFix estimation for two legume-grass mixtures through a whole vegetation period based on UAV multispectral data using partial least squares regression (PLSR) and RF regression. Wijesingha et al. [10] developed models for CP and ADF for multiple grassland types using UAV-borne imaging spectroscopy. Askari et al. [35] obtained good results for the prediction of biomass and CP using multi-spectral UAV images and PLSR and multilinear regression (MLR). Oliveira et al. [11] used RF and MLR to estimate the fresh and dry biomass, i.e., fresh yield (FY) and dry matter yield (DMY) and five different quality parameters: digestibility of organic matter in dry matter (the D-value), NDF, indigestible neutral detergent fiber (iNDF), water-soluble carbohydrates (WSC), and the nitrogen concentration (Ncont) in dry matter (CP = 6.25 × Ncont [36]) and nitrogen uptake (NU); this dataset will be further investigated in this study.

Recent studies have shown that DNNs outperform classical machine learning algorithms in quantity parameter estimation [26,29]. To investigate the performance of DNNs in grass quality parameter estimation, we studied various neural network architectures for estimating both the quantity and quality parameters of silage grass swards using drone RGB and hyperspectral images (HSI). Simple 2D- and 3D-CNNs trained from scratch and two different types of deep pretrained models, the VGGNet [37] and the Vision Transformer [20], were studied. Our primary objective was to compare their performance with the results of classical machine learning with handcrafted features. Our further objective was to compare the performance of the low-cost RGB camera and hyperspectral camera, as well as to study the advantages of utilizing dense photogrammetric point clouds. To efficiently compare DNN and classical remote sensing methods, we used the same dataset as in an earlier study [11].

2. Materials and Methods

2.1. Test Area and References

The study areas were located in the municipality of Jokioinen in southwest Finland (approximately 60°48′N, 23°30′E) (Figure 1). The datasets were captured in the spring growth (referred to as primary growth, PG) and summer growth (referred to as regrowth, RG) phases in summer 2017. The grass field for the primary growth training was a second-year timothy meadow fescue (*Phleum pratense* and *Lolium pratense*) ley, which was managed as a silage production sward in 2016. The grass fields for testing the primary growth estimation models and for training and testing the regrowth models were located approximately 1.2 km from the primary growth training area. The sward was a second-year sward, established on the 2nd of June in 2015, with a seed mixture composed of 67% timothy fescue and 33% tall fescue. However, the stand composition was nearly pure timothy in 2017.

The primary growth training dataset included a total of 96 sample plots, which consisted of four replicates, six nitrogen fertilizer levels (0, 50, 75, 100, 125 and 150 kg ha^{-1}) and four harvesting/measuring dates (6th, 15th, 19th and 28th of June). The regrowth datasets were obtained from a field in which primary growth was harvested for silage on 19 June. There were 108 sample plots with nine different nitrogen application rates or sources (0–150 kg N ha^{-1}), 4 replicates and three harvest dates (25th of July and 1st and 15th of August). Independent areas were used for testing. The sizes of harvested samples varied in different areas, as discussed in Section 2.2. Details of the materials are given in [9,11].

Various parameters were measured from the harvested datasets, including FY, DMY, D-value, NDF, iNDF, WSC, Ncont and NU (see [11]). The reference data statistics are presented in Table 1.

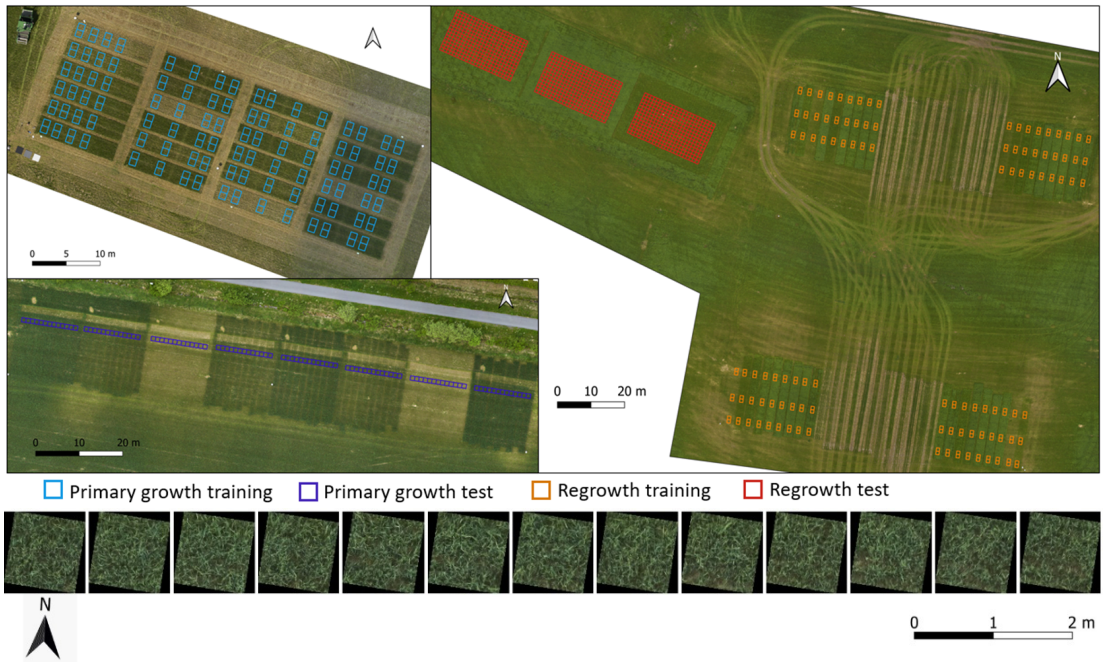


Figure 1. Field plots used as training and testing data and RGB image patches (sub-plots) of 1 m × 1 m for a primary growth test plot.

Table 1. Reference data statistics for the primary growth (PG) and regrowth (RG) training and testing datasets: fresh and dry biomass (FY, DMY), the digestibility of organic matter in dry matter (D-value), neutral detergent fiber (NDF), indigestible neutral detergent fiber (iNDF), water-soluble carbohydrates (WSC), and the nitrogen concentration (Ncont) in dry matter and nitrogen uptake (NU).

		FY (kg/ha)	DMY (kg DM/ha)	D-Value (g/kg DM)	iNDF (g/kg DM)	NDF (g/kg DM)	WSC (g/kg DM)	Ncont (g N/kg DM)	NU (N kg/ha)
PG Train	Mean	11,046.0	2652.6	710.1	62.0	537.8	138.8	22.1	53.6
	Min	1022.2	336.0	632.0	11.0	435.0	76.0	11.8	7.6
	Max	25,459.3	6135.1	770.0	128.0	614.0	246.0	40.8	105.7
	Std	6955.0	1578.8	39.7	31.8	56.1	42.2	7.6	26.7
PG Test	Mean	11,205.7	2408.4	720.8	41.4	591.0	108.4	21.7	54.4
	Min	4796.2	1255.9	708.0	29.0	573.0	79.0	16.0	21.1
	Max	18,783.5	3822.4	741.0	49.0	605.0	160.0	27.8	104.0
	Std	4440.3	816.3	11.8	5.6	10.5	31.4	4.3	26.0
RG train	Mean	16,658.3	3542.9	689.6	65.9	546.0	137.1	20.2	70.5
	Min	1379.3	368.0	618.0	13.0	426.0	50.0	13.0	7.5
	Max	32,390.6	7228.7	756.0	137.0	612.0	294.0	33.1	132.9
	Std	8043.4	1611.1	34.4	30.5	44.4	64.4	4.9	32.3
RG test	Mean	17,144.0	3898.6	677.4	81.0	530.5	141.6	20.3	76.5
	Min	7843.1	2073.3	631.0	57.0	464.0	89.0	14.7	37.8
	Max	27,911.1	7346.0	714.0	111.0	575.0	207.0	28.5	104.9
	Std	7259.5	1621.7	28.6	18.6	36.6	50.4	5.0	27.4

2.2. Drone Remote Sensing Datasets

A Gryphon Dynamics quadcopter drone was used to collect remote sensing datasets. Two cameras were operated onboard, the 36.4 megapixel Sony A7R (RGB) camera (Sony Corporation, Minato, Tokyo, Japan) with a Sony FE 35 mm f/2.8 ZA Carl Zeiss Sonnar T* lens (Sony Corporation, Minato, Tokyo, Japan) and the Fabry Pérot (FPI) hyperspectral 2D frame camera (the prototype FPI2012b) with an image size of 1024 × 648 pixels and 36 bands in the range of 500–900 nm. Datasets were collected under varying conditions using flight heights of 30 and 50 m. Photogrammetric processing was carried out using Agisoft PhotoScan Professional software (version 1.3.4, Agisoft LLC, St. Petersburg, Russia). For the FPI images, the in-house georeferencing, radiometric and orthomosaic processing

pipeline was used [38,39]. Orthophoto mosaics with ground sampling distances (GSDs) of 0.8 cm and 4 cm for the RGB and HSI, respectively, were created. Reflectance calibration was carried out for the HSI, and the RGB datasets were used with and without reflectance calibration. Point clouds were extracted from the RGB images and used to generate digital surface models (DSM) and digital terrain models (DTM), which were employed to calculate the canopy height models (CHM) of the study areas. The details of the processing are given by Viljanen et al. [9] and Oliveira et al. [11].

For machine learning, the following remote sensing data combinations were used: RGB, RGB + CHM, RGB_Refl, HSI and HSI + CHM. The reflectance calibrated RGB mosaic (RGB_Refl) is an additional dataset to the previous paper [11]. Images of 1 m × 1 m subplots of the harvested sample plots were cropped to create the training/validation and testing datasets. For the primary growth, the training/validation set included 192 images of 1 m × 1 m (96 harvested samples of size of 1.5 m × 2.6 m) and the test set included 104 images of 1 m × 1 m (8 harvested samples of size of 1.5 m × 12.6 to 13.4 m). For the regrowth, the training/validation set included 216 images (108 harvested samples of size of 1.5 m × 3 m) and the test set included 2392 sub-images (covering an area of 23 m × 13 m from 8 plots of size of approximately 30 m × 30 m). The same reference value was given as the reference value to each sub-image. The input image size of 1 m × 1 m was used to normalize the input data and to avoid potential problems of varying sizes of the reference plots; this is also a representative resolution to be used in the real farm application. The diversity of the training sample was increased by data augmentation during the training phase. The numbers of training and testing samples are summarized in Table 2 and examples of harvested samples of training and testing datasets and their splitting into sub-images are shown in Figure 1.

Table 2. Summary of the training and testing data.

	Training		**Testing**	
	Sample Plots	**1 m × 1 m Plots**	**Sample Plots**	**1 m × 1 m Plots**
Primary growth	96	192	8	104
Regrowth	108	216	8	2392

2.3. Neural Network Architectures for Estimating Grass Quality and Quantity Parameters

Methods for tackling limited training datasets in deep learning include transfer learning using pretrained models, data augmentation and using simpler models. Strategies for transfer learning are either using the pretrained network as a feature extractor and only changing the last classification layer or finetuning the parameters of the entire network. In this study, finetuning was selected instead of freezing as the transfer learning method, since the grass image data is very different from the image recognition problem. By changing the last model layer to have only one output, the model predicts a single value and can be used for regression instead of classification. Four different models were selected for this study: VGGNet and Vision Transformer (ViT), which represent very deep neural networks as well as in-house designed 2D- and 3D-CNN architectures. We conducted a preliminary study with various DNN architectures available in PyTorch or timm and concluded that the VGG and ViT represented well the CNN and transformer-based approaches. The simpler 2D- and 3D-CNN architectures were used to provide reference to the pretrained models.

VGGNet is a very deep convolutional neural network developed for large-scale image recognition [37]. It has been used in numerous studies since its publication in 2014. It is available with different numbers of layers. In this study, we used VGG16 (referred to as VGG in this paper) with batch normalization pretrained using the ImageNet 21k set [40]. VGG16 has 16 weight layers (13 convolution layers and 3 fully connected layers). The number of trainable parameters is very large, over 134 million. VGG16 for regression was thoroughly analyzed by Lathuilière et al. [14]. In a previous study, a pretrained VGGNet11 showed good performance [28].

ViT [20] is a recent deep learning architecture for image recognition tasks that does not include convolutional filters; rather, it is based on self-attention-based Transformers used in natural language processing. Images are split into patches and a sequence of linear embeddings from flattened patches is fed into a Transformer Encoder. They require even bigger training datasets than deep CNNs, but outstanding results have been reported in image recognition tasks [20] and many pretrained models are available. In this study, we used a small ViT, ViT-S, with 16 patches and an image size of 224 [41]. It consists of 12 transformer encoder blocks with 6 attention heads and has an embedding dimension of 384. The number of parameters was 22 million. The model was also pretrained on the ImageNet dataset.

In addition, simpler 2D-CNN and 3D-CNN models with a lower number of trainable parameters were trained from scratch. These models consist of three 2D or 3D convolutional layers and two fully connected layers (Table 3). The 2D-CNN had ~200,000 parameters and 3D-CNN ~300,000 parameters. The 3D-CNN model was used for the HSI datasets only. In the additional test for the RGB reflectance data, only 2D-CNN and VGG were trained.

Table 3. Examples of simple 2D and 3D convolutional neural network (CNN) layers and layer output sizes used in this study. The inputs for this 3D-CNN are hyperspectral image and canopy height model (HSI + CHM) data cubes. K = Kernel, S = Stride, F = Filters, BN = Batch Normalization.

2D-CNN					**3D-CNN (37 Input Channels)**				
Layer	**K**	**S**	**F**	**Output**	**Layer**	**K**	**S**	**F**	**Output**
Conv2D	3 × 3	1	32	32, 125, 125	Conv3D	3 × 3 × 3	1	32	32, 37, 125, 125
BN2D				32, 125, 125	BN3D				32, 37, 125, 125
ReLU				32, 125, 125	ReLU				32, 37, 125, 125
MaxPool2D	2 × 2	2		32, 62, 62	MaxPool3D	2 × 2 × 2	2		32, 18, 62, 62
Conv2D	3 × 3	1	64	64, 62, 62	Conv3D	3 × 3 × 3	1	64	64, 18, 62, 62
BN2D				64, 62, 62	BN3D				64, 18, 62, 62
ReLU				64, 62, 62	ReLU				64, 18, 62, 62
MaxPool2D	2 × 2	2		64, 31, 31	MaxPool3D	2 × 2 × 2	2		64, 9, 31, 31
Conv2D	3 × 3	1	64	64, 31, 31	Conv3D	3 × 3 × 3	1	64	64, 9, 31, 31
BN2D				64, 31, 31	BN3D				64, 9, 31, 31
ReLU				64, 31, 31	ReLU				64, 9, 31, 31
AdaptiveAvgPool2D				64, 6, 6	AdaptiveAvgPool3D				64, 3, 3, 3
Linear				64	Linear				64
ReLU				64	ReLU				64
Dropout				64	Dropout				64
Linear				1	Linear				1

The models were implemented using PyTorch and timm (PyTorch image models) Python libraries. The number of input channels was changed to match the number of channels in input data: 3 (RGB, RGB_Refl), 4 (RGB + CHM), 36 (HSI), and 37 (HSI + CHM). Pre-trained weights were only available for 3 channels, and for models with more input channels, the weights from channels 1–3 were repeatedly copied over the rest of the channels. In the timm library, the last layers of the VGG and ViT models are linear layers.

The image sizes of the georeferenced 1 m × 1 m subplots were approximately 25 × 25 pixels (HSI) and 125 × 125 (RGB, CHM) pixels. All images were scaled to the ImageNet size of 224 × 224 pixels for the pretrained VGG and ViT networks and to 125 × 125 pixels for the 2D- and 3D-CNN networks using the nearest neighbor resampling. Data cubes for the four input data combinations were stacked from the input data. Labels (reference measurements) were min–max normalized to values between 0 and 1. The range from 0 to 1 was also used for the remote sensing datasets; the RGB images were scaled from the range 0–255 to 0–1, the initial metric scale was used for the CHM, the initial reflectance scale was used for the HSI images, the RGB_Refl values were very small (mostly less than 0.1) and were therefore scaled between 0 and 1. The same scaling factor was used for all bands of the

dataset, keeping the initial band ratios. The training dataset was split into training (80%) and validation (20%) sets during the training phase. The test set was used only in the performance evaluation.

We selected the final hyperparameters for the experiment based on the validation results after several tests. AdamW optimizer [42] with a learning rate of 0.0001 (all except ViT, 0.00001) and a weight decay of 0.01 was used in the training. In AdamW, weight decay is a regularization term, which reduces the risk of overfitting. For the training, we used the MSE cost function. The training batch size was 32. The deep pretrained models were trained for 100 epochs and the simple models for 300 epochs. For each epoch, the diversity of the training data was increased by using data augmentations, which included random horizontal flip and random rotation between −180 and 180 degrees. The experiments were conducted on the NVIDIA Quadro RTX 6000 GPU.

2.4. Performance Assessment

The training was repeated 5 times with different fixed random seeds to study the stability of the models. In each run, the weights for the epoch that yielded the best validation accuracy were used for testing the final models. The final predictions for the test plots were determined as the mean of the predictions for 1 m × 1 m subplots of each test plot. The median, minimum and maximum root-mean-square error (RMSE) and normalized RMSE (NRMSE) were calculated from the estimation errors:

$$\text{NRMSE} = 100\% \times \left(\sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}}\right)/\overline{y} = 100\% \times \text{RMSE}/\overline{y}, \quad (1)$$

where x_i is the predicted and y_i is the reference value of test plot i, $\overline{y}$ is the average of reference values, and n is the number of samples. In addition, the Scott–Knott test for means [43,44] was used for the NRMSEs to rank the models and to find the best model or models when a single model was not statistically superior.

The coefficient of variation (CV) for the NRMSEs of the five runs was used as an indicator of the stability of the models. Modeling quality was also evaluated using the Pearson correlation coefficient (PCC). The results from the DNN and CNN models were compared to the RF results (Table 4) from a previous study by Oliveira et al. [11]. The RF models for different parameters were trained using different combinations of RGB, HSI, multispectral (MS) and CHM 3D features. RGB and HSI features included spectral bands and several handcrafted vegetation and spectral indexes. The MS features were vegetation indices using two spectral bands of the hyperspectral camera with the RGB dataset.

Table 4. Normalized root-mean-square errors (NRMSEs) for the Random Forest models for different parameters trained using 3D structural features from the CHM (3D) as well as spectral bands (_b) and vegetation indices (_i) extracted from hyperspectral (HS) and RGB images [11].

Random Forest	FY	DMY	D-Value	iNDF	NDF	WSC	Ncont	NU
				Primary Growth NRMSE%				
3D + RGB_b + RGB_i	40.8	50.6	5.3	99.8	6.7	35.8	17.2	25.5
3D + RGB_b+ RGB_i + MS_i	24.1	35.7	4.0	78.4	6.7	37.9	12.5	19.0
HS_b + HS_i	36.8	31.6	1.4	37.8	12.9	39.9	19.7	31.3
3D + HS_b + HS_i	20.8	23.0	2.8	67.0	9.3	41.8	17.8	31.2
				Regrowth NRMSE%				
3D + RGB_b + RGB_i	36.4	27.8	2.5	24.2	4.2	32.6	24.6	34.5
3D + RGB_b+ RGB_i + MS_i	26.6	25.9	2.8	27.0	3.6	17.0	19.0	26.3
HS_b + HS_i	29.2	30.7	4.9	41.4	4.0	28.4	14.0	29.3
3D + HS_b + HS_i	30.6	30.8	4.9	41.2	2.5	26.8	13.5	30.4

3. Results

All results are presented in Figures 2–10. NRMSE (%) for each data combination and DNN and comparison to RF is presented in Figures 2–7. The PCC, RMSE and CV values are presented in Figure 8. Observed vs. predicted values for five runs for the models with the lowest median RMSE are plotted for each parameter in Figures 9 and 10. The best models according to the Scott–Knott test are listed in Table 5. In the following, the results are described for each parameter separately.

3.1. Fresh and Dry Matter Yield Estimation

In general, the pretrained RGB VGG and ViT models provided very good estimation accuracies for FY and DMY. The best models according to the Scott–Knott test (Table 5) were the RGB VGG and RGB_refl VGG for the primary growth FY (NRMSE 18.8% and 19.6%, PCC 0.89 and 0.87), RGB_refl VGG for the primary growth DMY (NRMSE 21.2%, PCC 0.79), HSI + CHM 3D-CNN for the regrowth FY (NRMSE 24.7%, PCC 0.94), and RGB + CHM ViT for the regrowth DMY (NRMSE 24.6%, PCC 0.90). However, for regrowth, the results of the HSI 3D-CNN and RGB + CHM VGG and ViT models were at a similar level (Figure 2).

Considering the impact of CHM, the HSI models improved when the CHM was included in the estimation. In the case of the RGB image-based models, the use of CHM was advantageous with the regrowth data.

Most models outperformed the RGB + CHM RF, but only a few outperformed the best RF results that were obtained with 3D + HSI features for primary growth and 3D + RGB + MS features for regrowth [11]. It is noteworthy that the results with the RGB dataset with VGG and ViT were close to or even better than the best results of the RF estimators with the HSI or MS data (Figure 2).

It is also notable that the simple 2D-CNN produced very unstable results for the primary growth RGB data. The variation of the NRMSEs was generally less than 10%; however, some higher values were obtained with the primary growth FY (Figure 8, NRMSE CV). It can be observed that each model produced very similar FY and DMY estimates, which is an indication of the consistency of the models. For the best-performing models, the FY estimates were slightly more accurate than the DMY estimates, which is consistent with the expectations.

3.2. D-Value

For the D-value, the best models (Table 5) were RGB_Refl VGG and RGB + CHM ViT for primary growth (NRMSE 1.2% and 1.2%, PCC 0.82 and 0.82) and RGB + CHM ViT for regrowth (NRMSE 2.5%, PCC 0.91). In general, the CHM improved the results.

For primary growth, all DNNs provided median NRMSE around 2% or better, which is close to the best results with the RF with the HSI bands and indexes. For regrowth, the D-value NRMSE was 2.5–6%, which was worse than the results of the best RF models (Figure 3). The NRMSE CVs were in most cases 10–20%, indicating relatively good repeatability (Figure 8, NRMSE CV).

3.3. NDF and iNDF

The PCCs were low for NDF and iNDF, especially for the primary growth data (Figure 8). For regrowth, however, acceptable results were obtained with some models. For primary growth, the best models (Table 5) were the RGB_Refl VGG (NRMSE 12.3%, PCC 0.40) for iNDF and the HSI + CHM VGG for NDF (NRMSE 1.1%, PCC 0.81). For the regrowth iNDF, all models with RGB data provided good results and the best model was the RGB + CHM ViT (NRMSE 14.4%, PCC 0.91). For the regrowth NDF, the models with HSI data provided the best results; the best was the HSI + CHM 2D-CNN (NRMSE 4.2%, PCC 0.88). The CHM improved the results for the primary growth NDF and the regrowth iNDF.

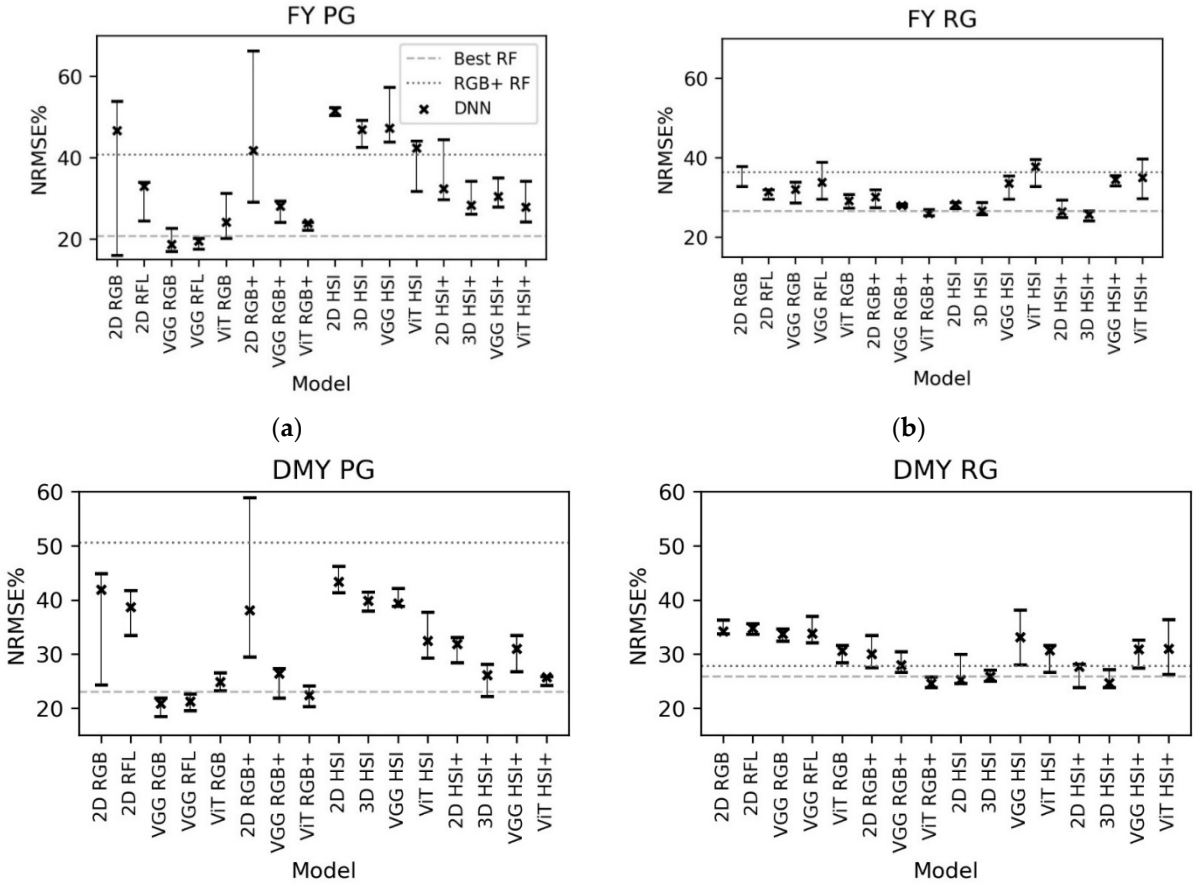


Figure 2. The DNN results for the 5 runs (median (x), min and max NRMSE (%)) and comparison to Random Forest (RF) results [11] for FY and DMY. RGB+ is the RGB + CHM and HSI+ is the HSI + CHM model. RFL is the reflectance-calibrated RGB data. The best RF model is the result with the best feature combination, including any of RGB, CHM, HS and MS features [11].

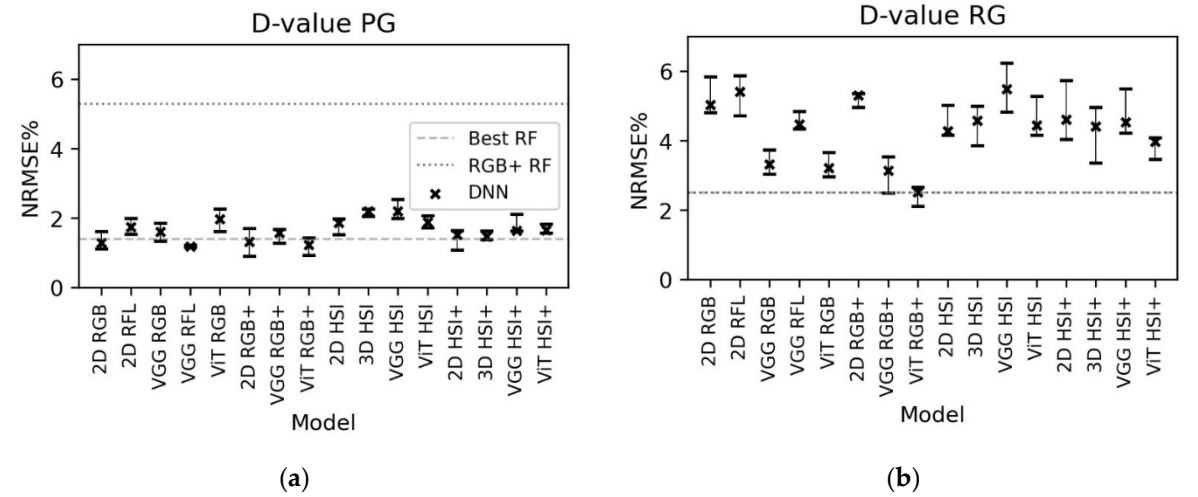


Figure 3. The DNN results for the 5 runs (median (x), min and max NRMSE (%)) and comparison to RF results [11] for D-value. RGB+ is the RGB + CHM and HSI+ is the HSI + CHM model. The best RF model is the result with the best feature combination, including any of RGB, CHM, HS and MS features [11]. In (**b**), the best RF is RGB + RF.

The DNNs outperformed the RF NRMSE, except for the regrowth NDF (Figure 4). There was some variability in the NRMSEs, especially for primary growth; in the above-mentioned best cases, they varied between 5–25% (Figure 8, NRMSE CV).

3.4. WSC

The PCC values indicated good performance of the estimators (Figure 8). Particularly, the 3D-CNNs with the HSI data were consistent in all cases; the best models (Table 5) were HSI + CHM 3D-CNN for primary growth (NRMSE 9.7%, PCC 0.98) and the HSI VGG for regrowth (NRMSE 20.2%, PCC 0.90). In general, the HSI data performed well, but good results were also obtained with RGB data using VGG and ViT. The CHM improved the results for only a few models.

All DNN models outperformed RF models for the primary growth data and the primary growth results were better than the regrowth results. In the case of the regrowth data, the best RF models (RGB + CHM + MS features) outperformed the DNN models, but the DNNs outperformed the RF models with RGB data (Figure 5).

The variability of the NRMSEs was mostly 10–20% for the primary growth and 5–10% for the regrowth (Figure 8, NRMSE CV).

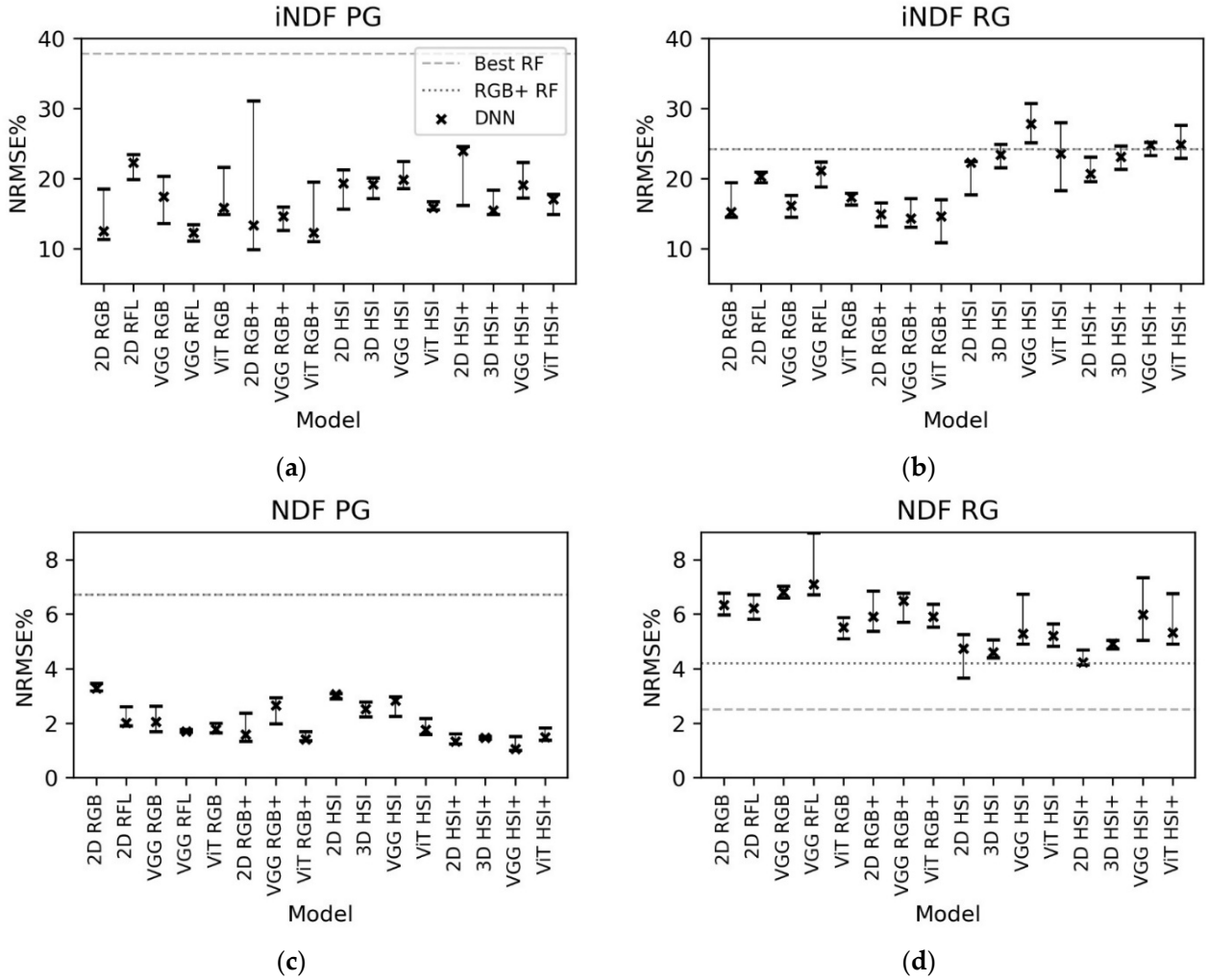
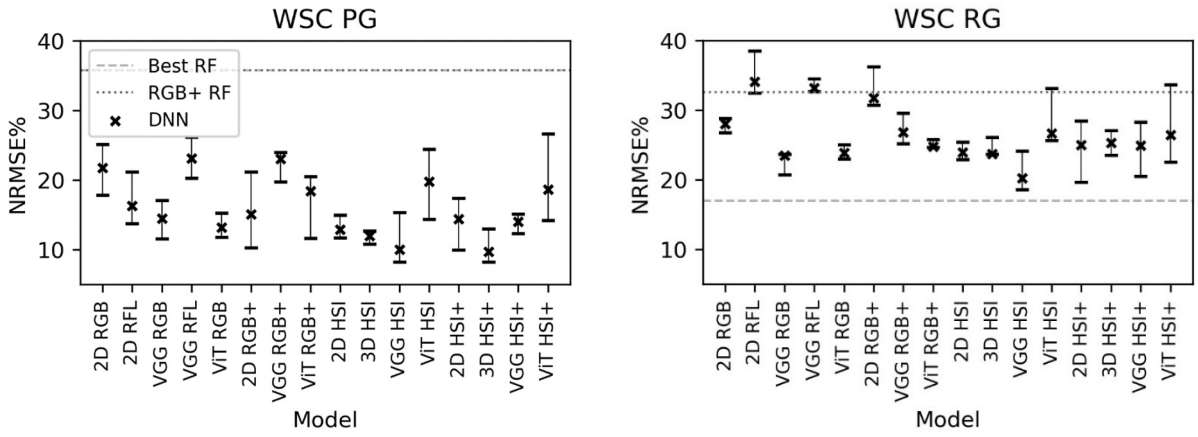


Figure 4. The DNN results for the 5 runs (median (x), min and max NRMSE (%)) and comparison to RF results [11] for iNDF and NDF. RGB+ is the RGB + CHM and HSI+ is the HSI + CHM model. The best RF model is the result with the best feature combination, including any of RGB, CHM, HS and MS features [11]. Note that in (**b**,**c**), the best RF is RGB + RF. In (**a**), RGB+ RF is too high for the range.



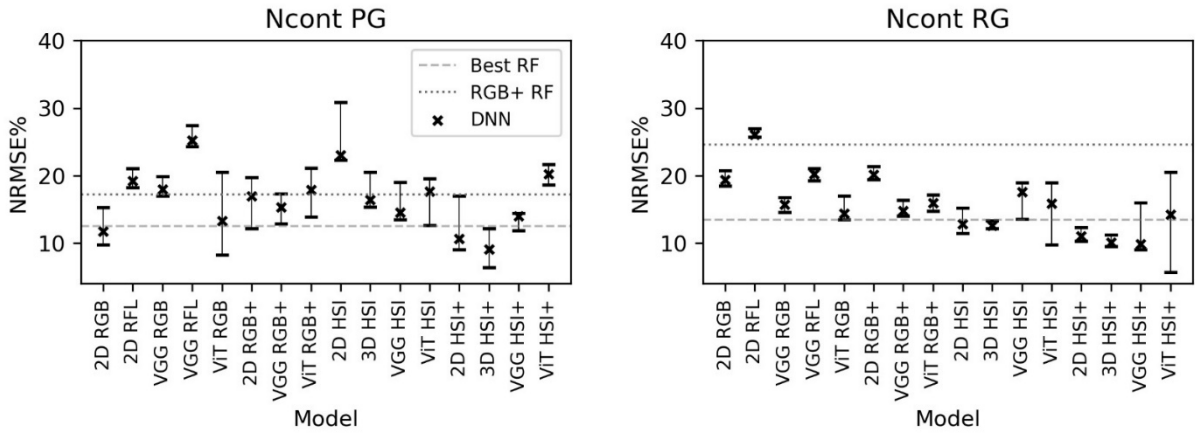
(**a**) (**b**)

Figure 5. The DNN results for the 5 runs (median (x), min and max NRMSE (%)) and comparison to RF results [11] for WSC. RGB+ is the RGB + CHM and HSI+ is the HSI + CHM model. The best RF model is the result with the best feature combination, including any of RGB, CHM, HS and MS features [11]. In (**a**), the best RF is RGB + RF.

3.5. Ncont

One of the biggest improvements of all parameters in comparison to RF was obtained for the Ncont parameter but only for some models. In general, the HSI + CHM 2D- and 3D-CNN models performed well. The best models (Table 5) were HSI + CHM 3D-CNN for primary growth (NRMSE 9.1%, PCC 0.96) and HSI + CHM 3D-CNN for regrowth (NRMSE 10.1%, PCC 0.95). The VGG and ViT RGB models provided better results with the regrowth data than with primary growth data. The CHM clearly improved the estimation results for the HSI data. RGB-only models were worse than HSI, especially for the regrowth data (Figure 6).

The NRMSEs of the best models outperformed or were at the same level as the best RF results (Figure 6), but for them considerable variation of approximately 20–30% can be observed in the NRMSE (Figure 8, NRMSE CV). However, there are also good models with a CV value below 10%, e.g., 2D- and 3D-CNN models for HSI + CHM for the regrowth data.



(**a**) (**b**)

Figure 6. The DNN results for the 5 runs (median (x), min and max NRMSE (%)) and comparison to RF results [11] for Ncont. RGB+ is the RGB + CHM and HSI+ is the HSI + CHM model. The best RF model is the result with the best feature combination, including any of RGB, CHM, HS and MS features [11].

3.6. NU

The best models (Table 5) were HSI 3D-CNN for primary growth (NRMSE 21.9%, PCC 0.89) and HSI + CHM 2D- and 3D-CNN for regrowth data (NRMSE 23.5% and 24.1%, PCC 0.90 and 0.97). For primary growth data, RGB DNNs also produced good results, and for regrowth, practically all models provided satisfactory results, excluding HSI ViT and HSI + CHM ViT. In general, the 2D- and 3D-CNNs were good with the HSI. It also seems that the CHM slightly improved the results for regrowth data but reduced the accuracy with the primary growth data.

The RF appeared to provide better NRMSEs than the evaluated neural networks in the primary growth NU estimation, whereas the DNNs achieved results comparable to the best RF results with the regrowth data (Figure 7). The variation in NRMSEs for the best models was less than 10% (Figure 8, NRMSE CV).

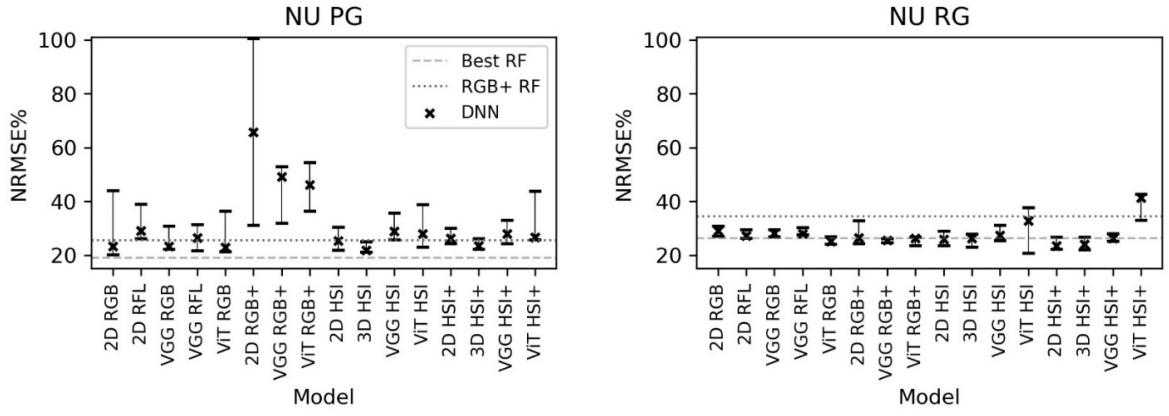
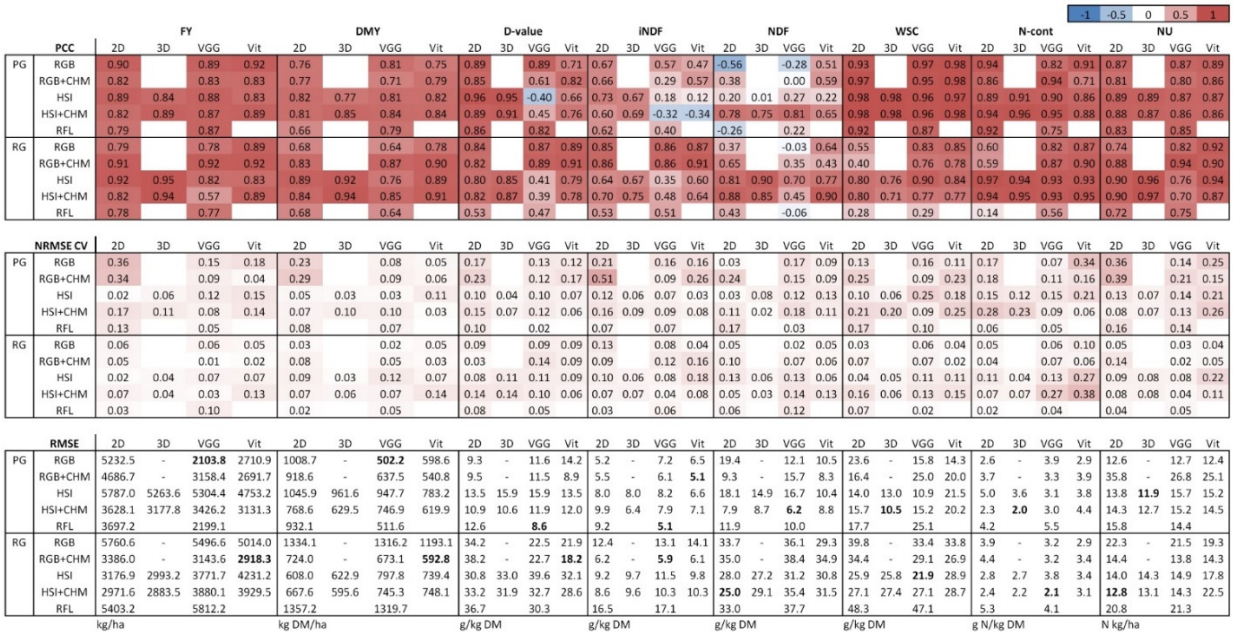


Figure 7. The DNN results for the 5 runs (median (x), min and max NRMSE (%)) and comparison to RF results [11] for NU. RGB+ is the RGB + CHM and HSI+ is the HSI + CHM model. The best RF model is the result with the best feature combination, including any of RGB, CHM, HS and MS features [11].



	PCC	FY 2D	FY 3D	FY VGG	FY Vit	DMY 2D	DMY 3D	DMY VGG	DMY Vit	D-value 2D	D-value 3D	D-value VGG	D-value Vit	iNDF 2D	iNDF 3D	iNDF VGG	iNDF Vit	NDF 2D	NDF 3D	NDF VGG	NDF Vit	WSC 2D	WSC 3D	WSC VGG	WSC Vit	N-cont 2D	N-cont 3D	N-cont VGG	N-cont Vit	NU 2D	NU 3D	NU VGG	NU Vit
PG	RGB	0.90		0.89	0.92	0.76		0.81	0.75	0.89		0.89	0.71	0.67		0.57	0.47	-0.56		-0.28	0.51	0.93		0.97	0.98	0.94		0.82	0.91	0.87		0.87	0.89
	RGB+CHM	0.82		0.83	0.83	0.77		0.71	0.79	0.85		0.61	0.82	0.66		0.29	0.57	0.38		0.00	0.59	0.97		0.95	0.98	0.86		0.94	0.71	0.81		0.80	0.86
	HSI	0.89	0.84	0.88	0.83	0.82	0.77	0.81	0.82	0.96	0.95	-0.40	0.66	0.73	0.67	0.18	0.12	0.20	0.01	0.27	0.22	0.98	0.98	0.96	0.97	0.89	0.91	0.90	0.86	0.89	0.89	0.87	0.87
	HSI+CHM	0.82	0.89	0.87	0.89	0.81	0.85	0.84	0.84	0.89	0.91	0.45	0.76	0.60	0.69	-0.32	-0.34	0.78	0.75	0.81	0.65	0.98	0.98	0.96	0.98	0.94	0.96	0.95	0.88	0.88	0.87	0.86	0.86
	RFL	0.79		0.87		0.66		0.79		0.86		0.82		0.62		0.40		-0.26		0.22		0.92		0.87		0.92		0.75		0.83		0.85	
RG	RGB	0.79		0.78	0.89	0.68		0.64	0.78	0.84		0.87	0.89	0.85		0.86	0.87	0.37		-0.03	0.64	0.55		0.83	0.85	0.60		0.82	0.87	0.74		0.82	0.92
	RGB+CHM	0.91		0.92	0.92	0.83		0.87	0.90	0.82		0.89	0.91	0.86		0.86	0.91	0.65		0.35	0.43	0.40		0.76	0.78	0.59		0.87	0.90	0.88		0.94	0.90
	HSI	0.92	0.95	0.82	0.83	0.89	0.92	0.76	0.89	0.80	0.85	0.41	0.79	0.64	0.67	0.35	0.60	0.81	0.90	0.70	0.77	0.80	0.76	0.90	0.84	0.97	0.94	0.93	0.93	0.90	0.96	0.76	0.94
	HSI+CHM	0.82	0.94	0.57	0.89	0.84	0.94	0.85	0.91	0.82	0.87	0.39	0.78	0.70	0.75	0.48	0.64	0.88	0.85	0.45	0.90	0.80	0.71	0.77	0.77	0.94	0.95	0.93	0.95	0.90	0.97	0.70	0.87
	RFL	0.78		0.77		0.68		0.64		0.53		0.47		0.53		0.51		0.43		-0.06		0.28		0.29		0.14		0.56		0.72		0.75	

	NRMSE CV	FY 2D	FY 3D	FY VGG	FY Vit	DMY 2D	DMY 3D	DMY VGG	DMY Vit	D-value 2D	D-value 3D	D-value VGG	D-value Vit	iNDF 2D	iNDF 3D	iNDF VGG	iNDF Vit	NDF 2D	NDF 3D	NDF VGG	NDF Vit	WSC 2D	WSC 3D	WSC VGG	WSC Vit	N-cont 2D	N-cont 3D	N-cont VGG	N-cont Vit	NU 2D	NU 3D	NU VGG	NU Vit
PG	RGB	0.36		0.15	0.18	0.23		0.08	0.05	0.17		0.13	0.12	0.21		0.16	0.16	0.03		0.17	0.09	0.13		0.16	0.11	0.17		0.07	0.34	0.36		0.14	0.25
	RGB+CHM	0.34		0.09	0.04	0.29		0.09	0.06	0.23		0.12	0.17	0.51		0.09	0.26	0.24		0.15	0.09	0.25		0.09	0.23	0.18		0.11	0.16	0.39		0.21	0.15
	HSI	0.02	0.06	0.12	0.15	0.05	0.03	0.03	0.11	0.10	0.04	0.10	0.07	0.12	0.06	0.07	0.03	0.03	0.08	0.12	0.13	0.10	0.06	0.25	0.18	0.15	0.12	0.15	0.21	0.13	0.07	0.14	0.21
	HSI+CHM	0.17	0.11	0.08	0.14	0.07	0.10	0.10	0.03	0.15	0.07	0.12	0.06	0.16	0.09	0.09	0.08	0.11	0.02	0.18	0.11	0.21	0.20	0.09	0.25	0.28	0.23	0.09	0.06	0.08	0.07	0.13	0.26
	RFL	0.13		0.05		0.08		0.07		0.10		0.02		0.07		0.07		0.17		0.03		0.17		0.10		0.06		0.05		0.16		0.14	
RG	RGB	0.06		0.06	0.05	0.03		0.02	0.05	0.09		0.09	0.09	0.13		0.08	0.04	0.05		0.02	0.05	0.03		0.06	0.04	0.05		0.06	0.10	0.05		0.03	0.04
	RGB+CHM	0.05		0.01	0.02	0.08		0.05	0.03	0.03		0.14	0.09	0.09		0.12	0.16	0.10		0.07	0.06	0.07		0.07	0.02	0.04		0.07	0.06	0.14		0.02	0.05
	HSI	0.02	0.04	0.07	0.07	0.09	0.03	0.12	0.07	0.08	0.11	0.11	0.09	0.10	0.06	0.08	0.18	0.13	0.06	0.13	0.06	0.04	0.05	0.11	0.11	0.11	0.04	0.13	0.27	0.09	0.08	0.08	0.22
	HSI+CHM	0.07	0.04	0.03	0.13	0.07	0.06	0.07	0.14	0.14	0.14	0.10	0.06	0.07	0.07	0.04	0.08	0.05	0.03	0.14	0.13	0.16	0.06	0.13	0.15	0.07	0.07	0.27	0.38	0.08	0.08	0.04	0.11
	RFL	0.03		0.10		0.02		0.05		0.08		0.05		0.03		0.06		0.06		0.12		0.07		0.02		0.02		0.04		0.04		0.05	

	RMSE	FY 2D	FY 3D	FY VGG	FY Vit	DMY 2D	DMY 3D	DMY VGG	DMY Vit	D-value 2D	D-value 3D	D-value VGG	D-value Vit	iNDF 2D	iNDF 3D	iNDF VGG	iNDF Vit	NDF 2D	NDF 3D	NDF VGG	NDF Vit	WSC 2D	WSC 3D	WSC VGG	WSC Vit	N-cont 2D	N-cont 3D	N-cont VGG	N-cont Vit	NU 2D	NU 3D	NU VGG	NU Vit
PG	RGB	5232.5	-	**2103.8**	2710.9	1008.7	-	**502.2**	598.6	9.3	-	11.6	14.2	5.2	-	7.2	6.5	19.4	-	12.1	10.5	23.6	-	15.8	14.3	2.6	-	3.9	2.9	12.6	-	12.7	12.4
	RGB+CHM	4686.7	-	3158.4	2691.7	918.6	-	637.5	540.8	9.5	-	11.5	8.9	5.5	-	6.1	**5.1**	9.3	-	15.7	8.3	16.4	-	25.0	20.0	3.7	-	3.3	3.9	35.8	-	26.8	25.1
	HSI	5787.0	5263.6	5304.4	4753.2	1045.9	961.6	947.7	783.2	13.5	15.9	15.9	13.5	8.0	8.0	8.2	6.6	18.1	14.9	16.7	10.4	14.0	13.0	10.9	21.5	5.0	3.6	3.1	3.8	13.8	**11.9**	15.7	15.2
	HSI+CHM	3628.1	3177.8	3426.2	3131.3	768.6	629.5	746.9	619.9	10.9	10.6	11.9	12.0	9.9	6.4	7.9	7.1	7.9	8.7	**6.2**	8.8	15.7	**10.5**	15.2	20.2	2.3	**2.0**	3.0	4.4	14.3	12.7	15.2	14.5
	RFL	3697.2		2199.1		932.1		511.6		12.6		**8.6**		9.2		**5.1**		11.9		10.0		17.7		25.1		4.2		5.5		15.8		14.4	
RG	RGB	5760.6	-	5496.6	5014.0	1334.1	-	1316.2	1193.1	34.2	-	22.5	21.9	12.4	-	13.1	14.1	33.7	-	36.1	29.3	39.8	-	33.4	33.8	3.9	-	3.2	2.9	22.3	-	21.5	19.3
	RGB+CHM	3386.0	-	3143.6	**2918.3**	724.0	-	673.1	**592.8**	38.2	-	22.7	**18.2**	6.2	-	**5.9**	6.1	35.0	-	38.4	34.9	34.4	-	29.1	26.9	4.4	-	3.2	3.4	14.4	-	13.8	14.3
	HSI	3176.9	2993.2	3771.7	4231.2	608.0	622.9	797.8	739.4	30.8	33.0	39.6	32.1	9.2	9.7	11.5	9.8	28.0	27.2	31.2	30.8	25.9	25.8	**21.9**	28.9	2.8	2.7	3.8	3.4	14.0	14.3	14.9	17.8
	HSI+CHM	2971.6	2883.5	3880.1	3929.5	667.6	595.6	745.3	748.1	33.2	31.9	32.7	28.6	8.6	9.6	10.3	10.3	**25.0**	29.1	35.4	31.5	27.1	27.4	27.1	28.7	2.4	2.2	**2.1**	3.1	**12.8**	13.1	14.3	22.5
	RFL	5403.2		5812.2		1357.2		1319.7		36.7		30.3		16.5		17.1		33.0		37.7		48.3		47.1		5.3		4.1		20.8		21.3	
		kg/ha				kg DM/ha				g/kg DM				g/kg DM				g/kg DM				g/kg DM				g N/kg DM				N kg/ha			

Figure 8. Median PCC, RMSE and NRMSE coefficient of variation (CV) for the 5 runs. For PCC and NRMSE, CV colors from low (blue) to high (red) have been used to make the interpretation of results easier. For the RMSE, the best results are written in bold.

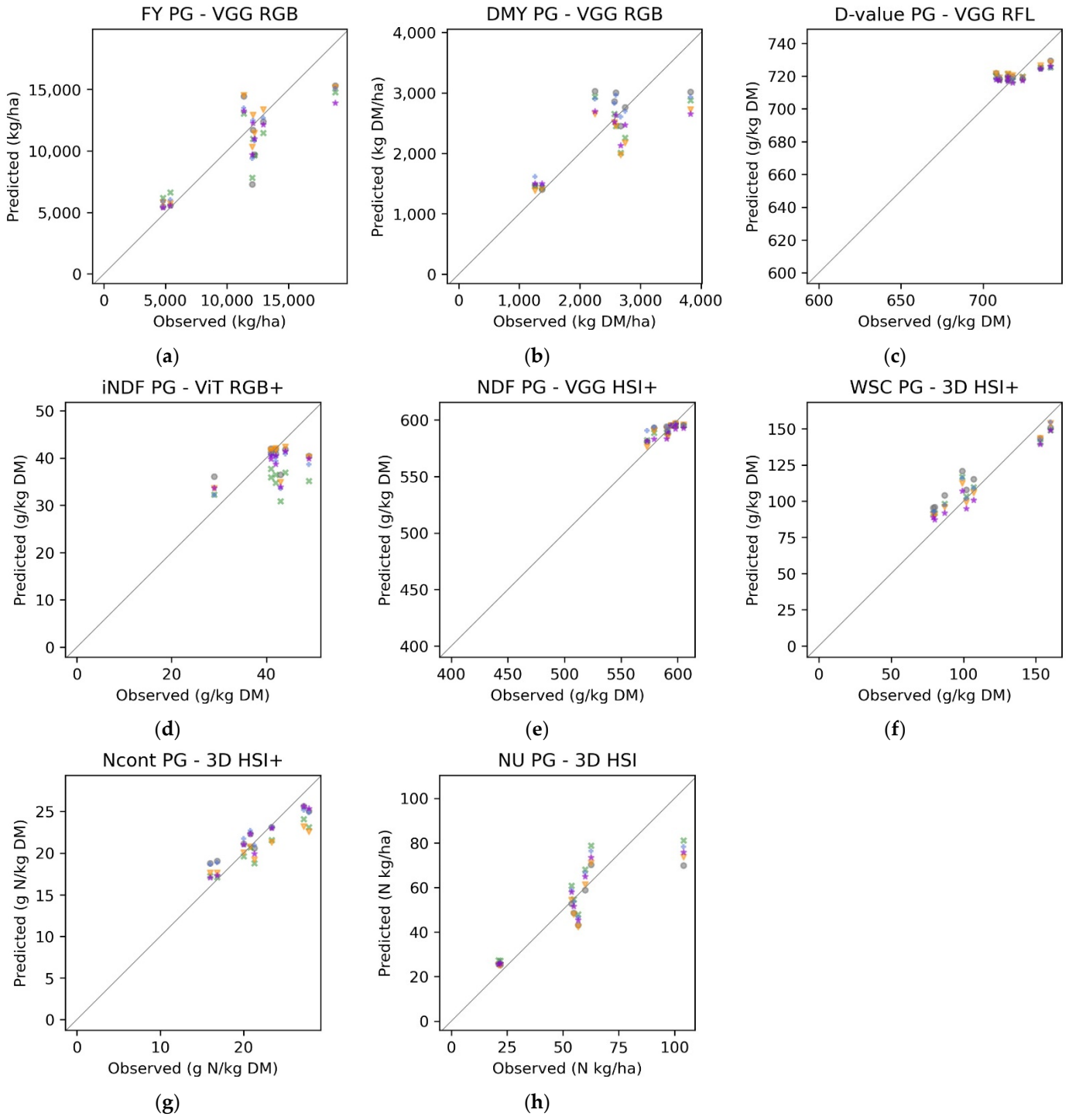


Figure 9. Primary growth observed (measured) vs. predicted (estimated) parameters for the best median RMSE models for (**a**) FY; (**b**) DMY; (**c**) D-value; (**d**) iNDF; (**e**) NDF; (**f**) WSC; (**g**) Ncont; (**h**) NU; each for 5 runs of the models.

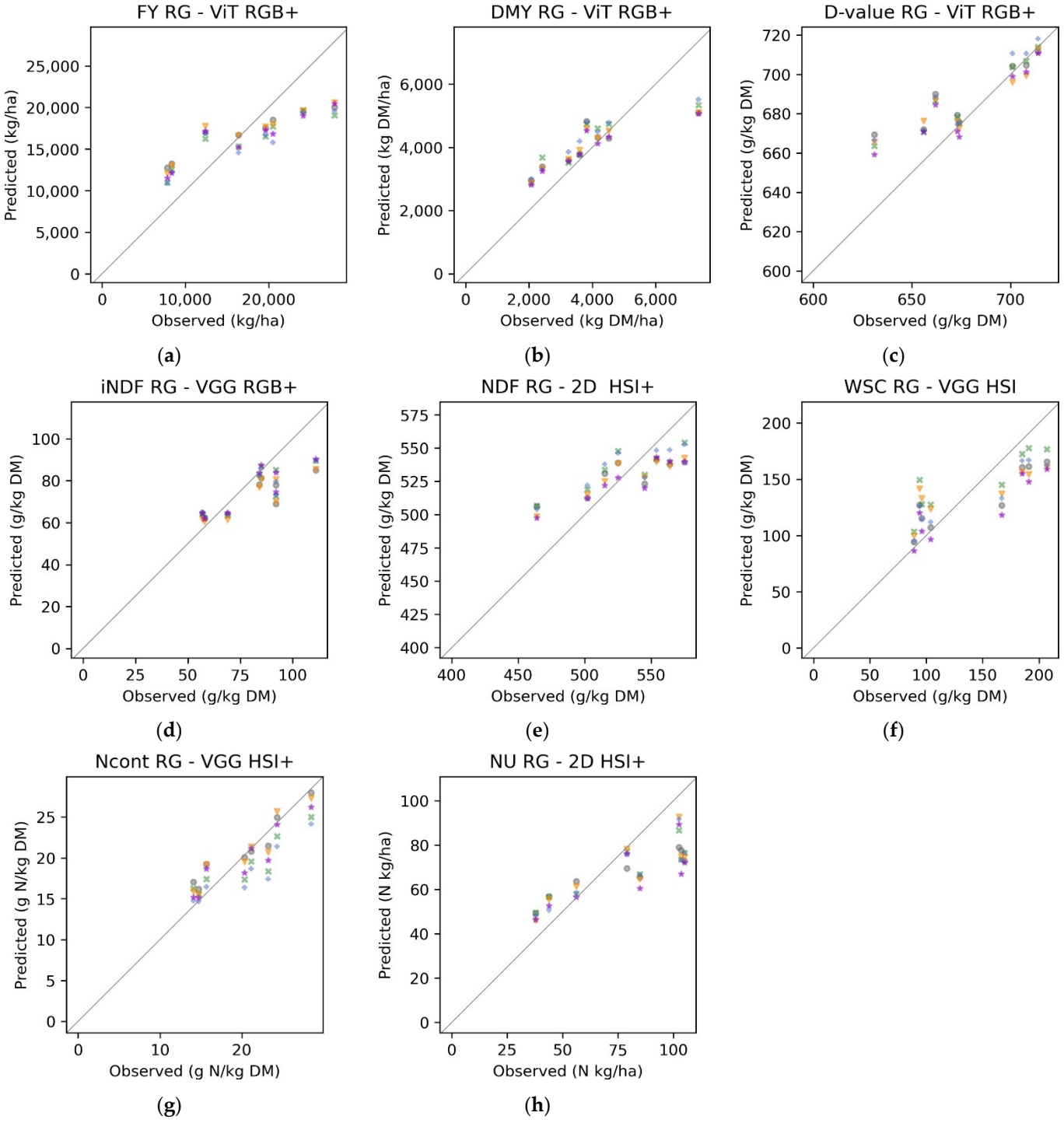


Figure 10. Regrowth observed (measured) vs. predicted (estimated) parameters for the best median RMSE models for (**a**) FY; (**b**) DMY; (**c**) D-value; (**d**) iNDF; (**e**) NDF; (**f**) WSC; (**g**) Ncont; (**h**) NU; each for 5 runs (presented by colors) of the models.

Table 5. The best DNN model(s) for each dataset according to the Scott–Knott test.

	Primary Growth	Regrowth
FY	VGG RGB, VGG RGB_Refl	3D HSI + CHM
DMY	VGG RGB	ViT RGB + CHM
D-value	ViT RGB + CHM, VGG RGB_Refl	ViT RGB + CHM
iNDF	VGG RGB_Refl	ViT RGB + CHM
NDF	VGG HSI + CHM	2D HSI + CHM
WSC	3D HSI + CHM	VGG HSI
Ncont	3D HSI + CHM	3D HSI + CHM
NU	3D HSI	2D HSI + CHM, 3D HSI + CHM

4. Discussion

We studied the performance of two very deep DNN models (VGG and ViT) and simple 3D- and 2D-CNN models for the yield quality and quantity estimation of silage production grass sward using RGB, RGB + CHM, HSI and HSI + CHM remote sensing datasets. The results were compared to our previous study with classical ML techniques RF using handcrafted features [11]. We used NRMSE, PCC and NRMSE CV to evaluate the general success of machine learning (Figure 8).

4.1. Performance of Different Remote Sensing Technologies

In general, good results were obtained for the FY, DMY, WSC, Ncont (or CP) and NU. The lowest PCCs appeared for the iNDF and NDF, particularly in the case of primary growth data. Repeating the estimation with different random number settings showed that most of the models were relatively stable; however, poorer repeatability appeared more systematically for the 2D-CNNs with the primary growth RGB and RGB + CHM data (Figure 8, NRMSE CV).

In most cases, the RGB datasets were comparable to and, in several cases, outperformed the HSI datasets. The RGB data provided good results, particularly for the FY, DMY, WSC and NU; the RGB image-based models outperformed HSI-based models for the FY, DMY, D-value and iNDF (Table 5). Considering the advantages of the HSI datasets, they outperformed the RGB datasets in the case of NDF, WSC, Ncont and NU (Table 5). Comparing the results with RF and handcrafted features, the best RGB DNN models (without CHM) outperformed the RGB RF models in all cases, except for the regrowth DMY, D-value and NDF. The best DNN models outperformed the best RF models, excluding the cases with regrowth D-value, NDF and WSC, and primary growth NU.

The CHM data provided advantages, particularly when combined with the HSI datasets in the case of the primary growth estimation. For the RGB models, the CHM did not provide significant advantages in most cases, indicating that the RGB images provided sufficient information for several parameters. The approach based on images only is attractive, particularly for fast response real-time applications because CHM generation requires extensive data processing. In particular, the DNNs using only RGB images seem useful from the processing and practical data collection point-of-view. When compared with RF models for HSI and multiple data sources, RGB-only DNNs showed promising performance. In this study, the results for CHM are not clear. Viljanen et al. [9] showed that CHM quality varied during four harvesting dates in primary growth when comparing the height measured with manual measuring tape and the height of CHM. The R^2 of the estimations were greater than 90% on 15 and 19 June but less than 80% in the first and last cutting dates 6 and 28 June, respectively. The authors concluded that the reason was that on the first cutting date, the grass was still very sparse and by the last date it was already overgrown and started to bend and lodge. This might have had an effect on the results of this study.

The reflectance calibration of the RGB datasets improved the results for the primary growth data in comparison to the uncalibrated RGB data for some parameters; this may indicate that there is a need to emphasize the importance of calibration. In some cases,

however, the reflectance calibration did not have an impact on or even deteriorated the accuracy. The relatively small impact of reflectance calibration might be explained by the fact that in the case of uncalibrated data, the image values were scaled to the range of 0–1. Moreover, the transformation of consumer RGB images with spectral response optimized for the human visual range to the reflectance scale might not be ideal because the transformation might even reduce the radiometric quality.

The DNNs improved the results of RGB image-based models more than the estimators based on the HSI data when compared to the results obtained with classical RF based on handcrafted features [11]. Potentially, the richer spatial information of the high spatial resolution RGB images, with 0.8 cm GSD, in comparison to the HSI images, with 4 cm GSD, could have contributed to this performance. ViT and VGG appeared to be the best networks with RGB data and provided equally good results in most cases. The lack of pretrained models for HSI data due to different sensor types and channels also complicates the possibilities of applying transfer learning. HSI data are likely to benefit from more sophisticated pretrained DNN models, which can handle joint coupling of spatial and spectral features [45]. It was thus a logical result that the deep RGB-pretrained models did not work as well for HSI as they did for RGB. Additionally, these models focus on 2D information and due to the lower spatial resolution, the geometric features in the HSI image are not so clear. Simple 3D-CNNs, which can consider both spatial and spectral features, appeared to be the best option for HSI data in this study. Methods in which HSI data are regressed to a particular parameter have generally been relatively little studied compared to classification methods. One reason for this is the availability of openly available materials. The hyperspectral datasets used in this study were captured with a relatively old camera model, the FPI prototype 2012b, whose characteristics are similar to commercially available Senop Rikola cameras [46]. Further improvements could be expected for HSI analyses if using the latest versions of the Senop cameras or other novel HSI sensors with improved spectral, radiometric and spatial performance [47].

4.2. Data and Application Area-Related Aspects

Many DNNs were initially designed for image classification using shapes, textures and colors. In this study, the prominent characteristics are color/spectra and texture. For the quantity parameters (FY, DMY), the texture may be more relevant and for some parameters (WSC, NU) the color/spectral information could provide added value. On the other hand, the structure of the grass, such as the stem–leaf ratio, is also linked to quality parameters such as to Ncont and D-value [48]. Such structural patterns could be visible in the texture of grass sward images if the resolution is high enough. We studied timothy-dominated grass sward. It is the most common in Nordic agriculture, but in organic production, mixed grass-clover swards are common. Their management, growth and feeding quality development differ from grass swards [49]. As the clover component in the mixture fix nitrogen biologically for the sward and for the subsequent crop, the site-specific information of clover and grass composition in the sward within the field would be very valuable for management measures, such as fertilizer application [50,51]. Sun et al. [52] studied the estimation of botanical composition in mixed clover-grass fields using machine learning-based image analysis. CNN or ViT might also be useful in drone image analysis with complex mixed swards and providing a very useful approach in organic production. More studies are needed on these aspects in different test setups.

Based on expectations, the biomass-related traits FY and DMY are generally easier to estimate than the traits based on leaf pigments and molecular structures that are visible in spectra. In this study, the lowest NRMSE values were obtained for the D-value, iNDF, NDF, WSC and Ncont. One factor impacting the level of NRMSE could be related to the characteristics of the reference data. Table 1 and Figures 9 and 10 show that the data variability was lower for these quantities; the lower data value range could lead to easier estimation tasks and result in lower NRMSE values. Furthermore, the NRMSEs in this study were consistent with the results from the previous study with RF [11].

There were significant differences in the primary growth and regrowth results. However, also in the previous study [11], similar differences were present (for different growth phases and input data combinations). The first aspect is related to the characteristics of the sward. In primary growth, the share of stems increases as the season proceeds and the plants start heading. Heading was abundant at the 4th harvesting date. In regrowth, stem formation remains small, and the stand is formed from leaf sheaths and leaf lamina [53]. By stem formation, the stand grows higher in primary growth than in regrowth. In the regrowth, the growing time of the stand prior to the second and third harvest times was longer than would be advised for common silage production. Therefore, the primary and regrowth were different due to the physical characteristics of the vegetation. Secondly, the training and testing areas were located in the same field in the regrowth phase and the datasets were collected during the same flight, whereas the training and testing areas were located in different fields with 1.2 km distance in the primary growth phase and were collected in different flights [11]; this resulted in greater differences between the training and testing datasets due to different environments and different conditions during the data capture, which could have deteriorated results. In general, the CV was lower for the models based on regrowth data. Third, in the primary growth, the weak initial growth and the lodging in the final phases caused some challenges for the models [9,11].

The small dataset will also have an influence on the results. Repeated experiments and averaging over the plots were carried out to make the results more reliable. Due to the small dataset, some models were not able to learn adequately, resulting in a high CV. In this study, we had separate training and testing areas as opposed to the sample shuffling method used in many DNN studies, making it more challenging to estimate the parameters than with shuffled plots but also giving more reliable evidence of the performance of the model. The instability in some of the DNN results confirms the need for repeat training, as also stated by Lathuilière et al. [14]. In general, RF does not require as large a training dataset as DNNs for good results. Especially for HSI data, the handcrafted features seem to produce good RF results. In [54], shallow CNN for feature extraction with an RF classification head was used for plant disease identification, and it outperformed deep pretrained models. This kind of hybrid approach could be tested in the future.

4.3. Assessment of the Results and Future Research

The results are comparable to previous studies using DNNs to estimate biomass and yield in different agricultural applications. Ma et al. [29] estimated winter wheat biomass using a DNN inspired by the VGG. The reported NRMSE was 25% (10% better than with RF). Castro et al. [28] estimated forage biomass (corresponding FY) using AlexNet, ResNet18 and VGGNet. The best mean absolute percentage error (MAPE) produced with AlexNet was 13 $\pm$ 2.2% (mean absolute error is usually lower than RMSE, as single bad predictions affect less). Yang et al. [32] estimated rice grain yield with two-branch CNN using RGB and MS imagery; MAPE of 20.4% was reported. In a comparison of RGB and RGB + MS, it was found that spatial information from very high resolution RGB images dominates the prediction. In [30], Guineagrass dry matter yield was estimated using the RGB sensor embedded in the DJi Phantom 4 Pro. The ResNeXt50 pretrained model was the best among several pretrained models from the literature with an RMSE of 413 $\pm$ 118 kg·ha^{-1}.

A few studies have estimated quality parameters using classical remote sensing techniques, but to the author's knowledge, there do not yet exist studies with deep learning models. Michez et al. [55] reported R^2 values of 0.33–0.85 for different quality parameters using RGB and multispectral sensors and MLR models. A study by Wijesingha et al. [10] indicated that the support vector regression provided the highest precision and accuracy when estimating CP (median NMRSE 10.6%) and the cubist regression model proved for the ADF estimation NRMSE of 13.4%. Askari et al. [35] obtained promising results for CP estimation with multispectral UAV imagery with PLSAR and MLR techniques. Dvorak et al. [33] used photogrammetric point clouds for predicting yield, ADF, NDF and CP, with R^2 values of 0.81, 0.81, 0.78, and 0.79, respectively. Feng et al. [56] developed a

multitask learning (MTL) approach to simultaneously predict multiple quality traits of alfalfa. The algorithm first extracts shared information through a long, short-term memory (LSTM)-based common hidden layer. To enhance the model flexibility, it is then divided into multiple branches, each containing the same or a different number of task-specific fully connected hidden layers. The approach outperformed various single-task machine-learning models (e.g., RF and support vector machine).

Our results showed that deep learning had the potential to outperform classical machine learning in grass quality and quantity parameter estimation for most of the parameters. This was validated by repeating training with several random seeds and by using completely independent test data. There are still many questions that still remain to be covered in future studies. First, there are questions concerning the reference datasets. Particularly, the amount and representativeness of training and testing data should be increased, and studies should be repeated with different datasets. The second fundamental question is the ability of remote sensing datasets to model the properties of interest. Our study showed good results for some parameters, but more extensive testing with different datasets is needed to answer this question comprehensively. Third, the quality of the remote sensing dataset is always a question and is expected to influence performance. More specifically, future studies should elaborate which remote sensing data is required and what is the required quality, e.g., with respect to spectral characteristics, spatial resolution, and calibration. Finally, this study still worked with samples cropped from a controlled trial site. The implementation of efficient pipelines for practical estimation tasks over heterogeneous objects will be an important need for future development.

5. Conclusions

This study evaluated the performance of novel pretrained deep neural network (DNN) architectures and simple 2D- and 3D-CNNs in estimating various silage grass stands quality and quantity parameters using drone-based remote sensing images. The models were trained and tested on primary growth and regrowth datasets for four remote sensing data combinations: RGB images, RGB + canopy height model (CHM), hyperspectral images (HSI), and HSI + CHM. The results were validated by repeating the training with several random seeds and by using completely independent test data. Both simple and very deep models produced promising results. Using only RGB data, many DNN models outperformed the random forest models trained with RGB images and CHM or with HSI and CHM. The HSI outperformed the RGB images for some parameters, particularly the nitrogen concentration and nitrogen uptake. The RGB models outperformed the HSI models in fresh yield and dry matter yield. D-value and indigestible neutral detergent fiber (iNDF). The results are promising but still can be considered preliminary, because the training and testing datasets were relatively limited, improved pretrained networks could support hyperspectral data modeling and the networks could potentially be further optimized. To build operational models and to compare the performance of different DNNs comprehensively, more data is needed.

This study was the first to compare novel deep learning architectures with classical random forest with handcrafted features using both RGB and HSI data in various grass quality parameter estimation tasks. The results indicated promising performance of DNNs in silage grass quality and quantity estimation and also indicated several topics for future research.

The approach has the potential for fast estimation of grass characteristics. Increased possibility for rapid results would be highly needed and appreciated by farmers when deciding on harvest time and planning the harvesting process. Simplest equipment (RGB cameras) and data processing would be appreciated by service providers in their farm advisory work. Different parameters are most important depending on the time of season or production type. In organic production, monitoring the nitrogen status of the stands is important, and for grazing, the uniformity of stands and amount of biomass are important for management decisions. Therefore, targeted models for working well on a single specific

parameter could also be very valuable. These new methods could be added to the current harvest time services, e.g., available in Finland and Sweden.

Author Contributions: Conceptualization, E.H., O.N. and J.K.; methodology, K.K., J.E. and E.H.; software, J.E. and K.K.; validation, K.K.; formal analysis, K.K., R.A.O., J.E., J.K., N.K., P.K., O.N., L.N., R.N., I.P. and E.H.; investigation, K.K.; resources, K.K., R.A.O., R.N., N.K., O.N. and L.N.; data curation, K.K., R.A.O., O.N., R.N. and E.H.; writing—original draft preparation, K.K. and E.H.; writing—review and editing, K.K., R.A.O., J.E., J.K., N.K., P.K., O.N., L.N., R.N., I.P. and E.H.; visualization, K.K. and R.A.O.; supervision, I.P., L.N., O.N. and E.H.; project administration, E.H.; funding acquisition, J.K., P.K. and E.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Academy of Finland ICT 2023 Smart-HSI—"Smart hyperspectral imaging solutions for new era in Earth and planetary observations" (Decision no. 335612), by the European Agricultural Fund for Rural Development: Europe investing in rural areas, Pohjois-Savon Ely-keskus (Grant no. 145346) and by the European Regional Development Fund for "CyberGrass I—Introduction to remote sensing and artificial intelligence assisted silage production" project (ID 20302863) in European Union Interreg Botnia-Atlantica programme. This research was carried out in affiliation with the Academy of Finland Flagship "Forest-Human-Machine Interplay—Building Resilience, Redefining Value Networks and Enabling Meaningful Experiences (UNITE)" (Decision no. 337127) ecosystem.

Data Availability Statement: No new data were analyzed in this study.

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

References

1. Pulli, S. Growth factors and management technique used in relation to the developmental rhythm and yield formation pattern of a pure grass stand. *Agric. Food Sci.* **1980**, *52*, 281–330. [CrossRef]
2. Rinne, M. Influence of the Timing of the Harvest of Primary Grass Growth on Herbage Quality and Subsequent Digestion and Performance in the Ruminant Animal. Ph.D. Dissertation, University of Helsinki, Helsinki, Finland, 2000.
3. Hyrkäs, M.; Korhonen, P.; Pitkänen, T.; Rinne, M.; Kaseva, J. Grass growth models for estimating digestibility and dry matter yield of forage grasses in Finland. In *Sustainable Meat and Milk Production from Grasslands*; Wageningen Academic Publishers: Wageningen, The Netherlands, 2018; pp. 252–254.
4. Aasen, H.; Honkavaara, E.; Lucieer, A.; Zarco-Tejada, P.J. Quantitative Remote Sensing at Ultra-High Resolution with UAV Spectroscopy: A Review of Sensor Technology, Measurement Procedures, and Data Correction Workflows. *Remote Sens.* **2018**, *10*, 1091. [CrossRef]
5. Swain, K.C.; Thomson, S.J.; Jayasuriya, H.P. Adoption of an unmanned helicopter for low-altitude remote sensing to estimate yield and total biomass of a rice crop. *Trans. ASABE* **2010**, *53*, 21–27. [CrossRef]
6. Bendig, J.; Yu, K.; Aasen, H.; Bolten, A.; Bennertz, S.; Broscheit, J.; Gnyp, M.; Bareth, G. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *39*, 79–87. [CrossRef]
7. Lary, D.J.; Alavi, A.H.; Gandomi, A.H.; Walker, A.L. Machine learning in geosciences and remote sensing. *Geosci. Front.* **2016**, *7*, 3–10. [CrossRef]
8. Näsi, R.; Viljanen, N.; Kaivosoja, J.; Alhonoja, K.; Hakala, T.; Markelin, L.; Honkavaara, E. Estimating Biomass and Nitrogen Amount of Barley and Grass Using UAV and Aircraft Based Spectral and Photogrammetric 3D Features. *Remote Sens.* **2018**, *10*, 1082. [CrossRef]
9. Viljanen, N.; Honkavaara, E.; Näsi, R.; Hakala, T.; Niemeläinen, O.; Kaivosoja, J. A novel machine learning method for estimating biomass of grass swards using a photogrammetric canopy height model, images and vegetation indices captured by a drone. *Agriculture* **2018**, *8*, 70. [CrossRef]
10. Wijesingha, J.; Astor, T.; Schulze-Brüninghoff, D.; Wengert, M.; Wachendorf, M. Predicting Forage Quality of Grasslands Using UAV-Borne Imaging Spectroscopy. *Remote Sens.* **2020**, *12*, 126. [CrossRef]
11. Oliveira, R.A.; Näsi, R.; Niemeläinen, O.; Nyholm, L.; Alhonoja, K.; Kaivosoja, J.; Jauhiainen, L.; Viljanen, N.; Nezami, S.; Markelin, L.; et al. Machine learning estimators for the quantity and quality of grass swards used for silage production using drone-based imaging spectrometry and photogrammetry. *Remote Sens. Environ.* **2020**, *246*, 111830. [CrossRef]
12. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
13. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1090–1098. [CrossRef]
14. Lathuilière, S.; Mesejo, P.; Alameda-Pineda, X.; Horaud, R. A Comprehensive Analysis of Deep Regression. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2065–2081. [CrossRef]
15. Zhang, L.; Zhang, L.; Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [CrossRef]

16. Ball, J.E.; Anderson, D.T.; Chan Sr, C.S. Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community. *J. Appl. Remote Sens.* **2017**, *11*, 042609. [CrossRef]
17. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]
18. Yuan, X.; Shi, J.; Gu, L. A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Syst. Appl.* **2021**, *169*, 114417. [CrossRef]
19. Chen, Y.; Jiang, H.; Li, C.; Jia, X.; Ghamisi, P. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6232–6251. [CrossRef]
20. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.
21. Bazi, Y.; Bashmal, L.; Rahhal, M.M.A.; Dayil, R.A.; Ajlan, N.A. Vision Transformers for Remote Sensing Image Classification. *Remote Sens.* **2021**, *13*, 516. [CrossRef]
22. Chen, H.; Qi, Z.; Shi, Z. Remote Sensing Image Change Detection with Transformers. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5607514. [CrossRef]
23. Donahue, J.; Jia, Y.; Vinyals, O.; Hoffman, J.; Zhang, N.; Tzeng, E.; Darrell, T. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. In Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 21–26 June 2014; Volume 32, pp. 647–655.
24. Castelluccio, M.; Poggi, G.; Sansone, C.; Verdoliva, L. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. *arXiv* **2015**, arXiv:1508.00092.
25. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [CrossRef]
26. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
27. Osco, L.P.; Junior, J.M.; Ramos, A.P.M.; de Castro Jorge, L.A.; Fatholahi, S.N.; de Andrade Silva, J.; Matsubara, E.T.; Pistori, H.; Gonçalves, W.N.; Li, J. A Review on Deep Learning in UAV Remote Sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102456. [CrossRef]
28. Castro, W.; Marcato Junior, J.; Polidoro, C.; Osco, L.P.; Gonçalves, W.; Rodrigues, L.; Santos, M.; Jank, L.; Barrios, S.; Valle, C.; et al. Deep Learning Applied to Phenotyping of Biomass in Forages with UAV-Based RGB Imagery. *Sensors* **2020**, *20*, 4802. [CrossRef]
29. Ma, J.; Li, Y.; Chen, Y.; Du, K.; Zheng, F.; Zhang, L.; Sun, Z. Estimating above ground biomass of winter wheat at early growth stages using digital images and deep convolutional neural network. *Eur. J. Agron.* **2019**, *103*, 117–129. [CrossRef]
30. de Oliveira, G.S.; Marcato Junior, J.; Polidoro, C.; Osco, L.P.; Siqueira, H.; Rodrigues, L.; Jank, L.; Barrios, S.; Valle, C.; Simeão, R.; et al. Convolutional Neural Networks to Estimate Dry Matter Yield in a Guineagrass Breeding Program Using UAV Remote Sensing. *Sensors* **2021**, *21*, 3971. [CrossRef]
31. Kattenborn, T.; Eichel, J.; Wiser, S.; Burrows, L.; Fassnacht, F.E.; Schmidtlein, S. Convolutional Neural Networks accurately predict cover fractions of plant species and communities in Unmanned Aerial Vehicle imagery. *Remote Sens. Ecol.* **2020**, *6*, 472–486. [CrossRef]
32. Yang, Q.; Shi, L.; Han, J.; Zha, Y.; Zhu, P. Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. *Field Crops Res.* **2019**, *235*, 142–153. [CrossRef]
33. Dvorak, J.S.; Pampolini, L.F.; Jackson, J.J.; Seyyedhasani, H.; Sama, M.P.; Goff, B. Predicting Quality and Yield of Growing Alfalfa from a UAV. *Trans. ASABE* **2021**, *64*, 63–72. [CrossRef]
34. Grüner, E.; Astor, T.; Wachendorf, M. Prediction of Biomass and N Fixation of Legume–Grass Mixtures Using Sensor Fusion. *Front. Plant Sci.* **2021**, *11*, 603921. [CrossRef] [PubMed]
35. Askari, M.S.; McCarthy, T.; Magee, A.; Murphy, D.J. Evaluation of Grass Quality under Different Soil Management Scenarios Using Remote Sensing Techniques. *Remote Sens.* **2019**, *11*, 1835. [CrossRef]
36. Jones, D.B. *Factors for Converting Percentages of Nitrogen in Foods and Feeds into Percentages of Protein*; US Department of Agriculture: Washington, DC, USA, 1931; Volume 183, pp. 1–21.
37. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
38. Honkavaara, E.; Rosnell, T.; Oliveira, R.; Tommaselli, A. Band registration of tuneable frame format hyperspectral UAV imagers in complex scenes. *ISPRS J. Photogramm. Remote Sens.* **2017**, *134*, 96–109. [CrossRef]
39. Honkavaara, E.; Khoramshahi, E. Radiometric correction of close-range spectral image blocks captured using an unmanned aerial vehicle with a radiometric block adjustment. *Remote Sens.* **2018**, *10*, 256. [CrossRef]
40. Ridnik, T.; Ben-Baruch, E.; Noy, A.; Zelnik-Manor, L. Imagenet-21k pretraining for the masses. *arXiv* **2021**, arXiv:2104.10972.
41. Touvron, H.; Cord, M.; Douze, M.; Massa, F.; Sablayrolles, A.; Jegou, H. Training data-efficient image transformers & distillation through attention. *arXiv* **2021**, arXiv:2012.12877.
42. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. *arXiv* **2019**, arXiv:1711.05101.
43. Scott, A.; Knott, M. Cluster-Analysis Method for Grouping Means in Analysis of Variance. *Biometrics* **1974**, *30*, 507–512. [CrossRef]
44. Tantithamthavorn, C.; McIntosh, S.; Hassan, A.E.; Matsumoto, K. The Impact of Automated Parameter Optimization on Defect Prediction Models. *IEEE Trans. Softw. Eng.* **2019**, *45*, 683–711. [CrossRef]
45. Ahmad, M.; Shabbir, S.; Roy, S.K.; Hong, D.; Wu, X.; Yao, J.; Khan, A.; Mazzara, M.; Distefano, S.; Chanussot, J. Hyperspectral Image Classification-Traditional to Deep Models: A Survey for Future Prospects. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *15*, 968–999. [CrossRef]

46. Senop.fi, Hyperspecral Imaging | High Performance Devices—Senop. Available online: https://senop.fi/industry-research/hyperspectral-imaging/ (accessed on 20 May 2022).
47. Nex, F.; Armenakis, C.; Cramer, M.; Cucci, D.A.; Gerke, M.; Honkavaara, E.; Kukko, A.; Persello, C.; Skaloud, J. UAV in the advent of the twenties: Where we stand and what is next. *ISPRS J. Photogramm. Remote Sens.* **2022**, *184*, 215–242. [CrossRef]
48. Van Soest, P.J. *Nutritional Ecology of the Ruminant*; Cornell University Press: Ithaca, NY, USA, 1994.
49. Kuoppala, K. *Influence of Harvesting Strategy on Nutrient Supply and Production of Dairy Cows Consuming Diets Based on Grass and Red Clover Silage*; MTT Science 11; MTT Agrifood Research Finland: Jokioinen, Finland, 2010. Available online: http://urn.fi/URN:ISBN:978-952-487-286-7 (accessed on 29 May 2022).
50. Nykänen, A.; Jauhiainen, L.; Kemppainen, J.; Lindström, K. Field-scale spatial variation in soil nutrients and in yields and nitrogen fixation of clover-grass leys. *Agric. Food Sci.* **2008**, *17*, 376–393. Available online: https://journal.fi/afs/article/view/5927/67185 (accessed on 29 May 2022). [CrossRef]
51. Nykänen, A. *Nitrogen Dynamics of Organic Farming in a Crop Rotation Based on Red Clover (Trifolium pratense) Leys*; Agrifood Research Reports 121; MTT Agrifood Research Finland: Jokioinen, Finland, 2008.
52. Sun, S.; Liang, N.; Zuo, Z.; Parsons, D.; Morel, J.; Shi, J.; Wang, Z.; Luo, L.; Zhao, L.; Fang, H.; et al. Estimation of Botanical Composition in Mixed Clover–Grass Fields Using Machine Learning-Based Image Analysis. *Front. Plant Sci.* **2021**, *12*, 622429. [CrossRef]
53. Virkajärvi, P.; Järvenranta, K. Leaf dynamics of timothy and meadow fescue under Nordic conditions. *Grass Forage Sci.* **2001**, *56*, 294–304. [CrossRef]
54. Li, Y.; Nie, J.; Chao, X. Do we really need deep CNN for plant diseases identification? *Comput. Electron. Agric.* **2020**, *178*, 105803. [CrossRef]
55. Michez, A.; Philippe, L.; David, K.; Sébastien, D.; Christian, D.; Bindelle, J. Can Low-Cost Unmanned Aerial Systems Describe the Forage Quality Heterogeneity? Insight from a Timothy Pasture Case Study in Southern Belgium. *Remote Sens.* **2020**, *12*, 1650. [CrossRef]
56. Feng, L.; Zhang, Z.; Ma, Y.; Sun, Y.; Du, Q.; Williams, P.; Drewry, J.; Luck, B. Multitask Learning of Alfalfa Nutritive Value From UAV-Based Hyperspectral Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 5506305. [CrossRef]