Prediction of pasture yield using machine learning-based optical sensing: a systematic review

Christoph Stumpe[1] · Joerg Leukel[2] · Tobias Zimpel[2]

Abstract

Accurate and reliable predictions of biomass yield are important for decision-making in pasture management including fertilization, pest control, irrigation, grazing, and mowing. The possibilities for monitoring pasture growth and developing prediction models have greatly been expanded by advances in machine learning (ML) using optical sensing data. To facilitate the development of prediction models, an understanding of how ML techniques affect performance is needed. Therefore, this review examines the adoption of ML-based optical sensing for predicting the biomass yield of managed grasslands. We carried out a systematic search for English-language journal articles published between 2015-01-01 and 2022-10-26. Three coders screened 593 unique records of which 91 were forwarded to the full-text assessment. Forty-three studies were eligible for inclusion. We determined the adoption of techniques for collecting input data, preprocessing, and training prediction models, and evaluating their performance. The results show (1) a broad array of vegetation indices and spectral bands obtained from various optical sensors, (2) an emphasis focus on feature selection to cope with high-dimensional sensor data, (3) a low reporting rate of unitless performance metrics other than R^2, (4) higher variability of R^2 for models trained on sensor data of larger distance from the pasture sward, and (5) the need for greater comparability of study designs and results. We submit recommendations for future research and enhanced reporting that can help reduce barriers to the integration of evidence from studies.

Introduction

Pastures account for about 70% of the world's agricultural land (Squires et al., 2018) and provide essential sources of high-quality forage for ruminants (Bouwman et al., 2005). Thus, pastures assume a key role in nourishing a growing global population with dairy and meat products (Henchion et al., 2017; Tripathi et al., 2018). Moreover, grasslands fulfill eco-

---

Extended author information available on the last page of the article

🌲 Springer

system services such as carbon storage and habitat conservation; hence, they help mitigate climate change and preserve biodiversity (O'Mara, 2012; Zhao et al., 2020).

Pasture management is being challenged by increasing competition between forage and energy crops (Donnison & Fraser, 2016), land sealing due to infrastructure and housing, greater yield volatility due to climate change (Hopkins & Del Prado, 2007), and stronger constraints on fertilization (Buckley et al., 2016). Against this backdrop, accurate and reliable information about future pasture yields gains importance for agricultural management. These yield predictions can support the decision-making processes regarding fertilization, pest control, irrigation, stocking rates, and mowing. Overall, accurate yield predictions allow more efficient use of all inputs, resulting in less environmental impact and greater profits for farmers (Hedley, 2015; Kent Shannon et al., 2018).

Machine learning-based optical sensing has become the prevailing approach to predictive modeling for pasture yields. In this approach, past observational data is analyzed to learn a mapping function between pasture characteristics and biomass at harvest. This function is used to predict the biomass for pasture characteristics obtained via sensors at a future time. Predictive modeling using machine learning (ML) takes advantage of significant improvements in technology for optical sensing (Adão et al., 2017; Zeng et al., 2020), enhanced availability of field data at different levels of granularity (Murphy et al., 2021), and greater performance of the underlying ML algorithms. The importance of MLbased optical sensing for pasture yield prediction is reflected in the high number of studies in recent years.

Evidence for the effectiveness of ML-based predictive modeling has increased. The evidence concerns different grass species, such as perennial ryegrass (*Lolium perenne*) (Nguyen et al., 2022), signalgrass (*Brachiaria)* (Bretas et al., 2021), and clover (*Trifolium pratense*) (Li et al., 2021), and different types of optical sensors, including portable spectroradiometers (Murphy et al., 2022), sensors mounted on unmanned aerial vehicles (UAVs) (van der Merwe et al., 2020), and carried by satellites (Bretas et al., 2021). This variety converges with a broad set of ML techniques that developers can adopt. Developers must select techniques for the transformation of input data into features, the training of prediction models from past observations of input data and biomass, and the evaluation of the trained models on new observations. To inform these decisions on the development of prediction models, insights into the effectiveness of specific ML techniques are required.

Notwithstanding the increased evidence base, the understanding of the effectiveness of specific ML techniques is still limited. Regarding the types of optical sensor data, a previous review found better performance for prediction models that were trained on in-field imagery compared to models that processed satellite data (Morais et al., 2021). However, about half of the included studies examined non-managed grasslands, such as steppe, semiarid grassland, bunchgrass, and shrub on drylands; hence, the finding cannot necessarily be generalized to prediction models for pasture yield. Two literature reviews focused on UAVs so that the results do not extend to models trained on data from field spectroradiometers and satellites (Bazzo et al., 2023; Lyu et al., 2022). Another review had a broader scope by including studies that collected non-optical sensor data (Murphy et al., 2021). One related review only provided aggregated information but no results at the study level (Subhashree et al., 2023).

Collectively, the burgeoning field of pasture yield prediction using ML-based optical sensing calls for the assessment of current evidence to facilitate the development of prediction models. To address this need, we conducted a systematic review that is conceptually guided by the ML process. Specifically, the objectives are to: (1) determine the adoption of

ML-based optical sensing in previous research examining yield prediction for pasture management, (2) collate the performance results, and (3) propose recommendations for future research and the reporting of studies.

Method

We conducted a systematic review of studies and report the results based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, where applicable (Moher et al., 2009). To ensure the reliability of the coding, three authors independently screened the identified records, assessed the full-text articles for eligibility, and extracted data from the selected studies.

Eligibility criteria

We included studies that applied machine learning on data obtained from optical sensing for predicting the yield of pastures, with yield defined as the current or future biomass of a specific pasture area, such as a plot, paddock, or field. The studies were required to report empirical results from the processing of real-world data. We focused on studies in refereed journals and written in English. The time interval of the past eight years (2015-01-01 through 2022-10-26) allowed us to assess studies that benefit from advances in ML and optical sensors in recent years, and thus have high relevance for research and practice. Studies were excluded if any of the following criteria were met: (1) dependent variable not related to a managed pasture but a different crop, nature conservation, biodiversity, or grassland coverage; (2) no prediction of yield but a different variable (e.g., nutrients, sward composition); (3) no predictive modeling but explanatory modeling or conceptual research; (4) no use of machine learning; (5) no processing of real-world data; and (6) no use of data from optical sensing (e.g., exclusively weather data).

Information sources and search

We identified articles through an automated search of journal articles published between January 1, 2015, and October 26, 2022. The search used the electronic database Scopus, which is the largest database of scientific literature and has larger coverage of peer-reviewed literature than the Web of Science (Mongeon & Paul-Hus, 2016; Singh et al., 2021; Thelwall & Sud, 2022). We designed the search query to cover the wide variety of terminology found in the literature. The search query had four concatenated components for pasture, yield, prediction, and machine learning. The pasture was represented as follows: (pasture* OR forage* OR grassland* OR *grass OR herbage* OR meadow*). Yield was covered by the following term: (yield OR biomass OR agb OR "herbage mass" OR "pasture mass" OR "grassland production" OR "forage production" OR quantity). The prediction component included different words as follows: (predict* OR assess* OR estimat* OR forecast*). Machine learning was represented by abstract terms and specific algorithms as follows: ("machine learning" OR "deep learning" OR "support vector" OR "random forest*" OR "neural network" OR "partial least square*" OR "predict* model*" OR "regression model").

Study selection

We ensured the reliability of the study selection through the following procedure. We defined a codebook that provided the eligibility and exclusion criteria. The codebook was used in the screening phase by three authors who independently coded the first nine articles based on the title, abstract, and keywords. The coders met to compare their codes and resolve any conflicting codes through discussion. The coding commenced with the remaining articles. Once all articles were coded and discussed, we downloaded the full texts of the articles that passed the screening. The assessment of the full texts employed the same codebook and was organized in two rounds of coding and resolving disagreements.

Data collection process

The data collection for the included studies was carried out by the same three authors, who independently filled in an Excel spreadsheet form for 96 data items per article. The data items operationalize the conceptual framework described in the following section. All individual codes were compared in two rounds in which disagreements were resolved through discussion and consensus.

Data items

Figure 1 illustrates the conceptual model of the review based on the process of predictive modeling using machine learning. This process begins with the prediction problem and ends with the evaluated prediction model. The figure also denotes the principal data items that we collected during the review.

Forecasting the yield of pastures based on optical sensor data recorded during the vegetation period represents the prediction problem. *Plant species* are pasture plants that are cultivated and constitute the sward composition for which the prediction is made. *Grazing* indicates whether the pasture is grazed by animals or managed by machinery for forage conservation. *Country* denotes the location where the study was conducted. The process shown in Fig. 1 defines four phases, which we discuss in the following paragraphs.

Data collection includes the creation of a data set by recording prediction-relevant data of the pasture vegetation and the observed yield at the time of harvest. *Study conditions* describe the number of fields, sample plots, and seasons for which these data were recorded. A sample plot is defined as the smallest partial area from which an independent sample of biomass is collected by cutting. Studies vary in the number of plots per field as well as in the size of plots, which usually range from 0.25 m^2 to a handful of square meters. The *input data* can be classified into the following groups: vegetation indices calculated from spectral
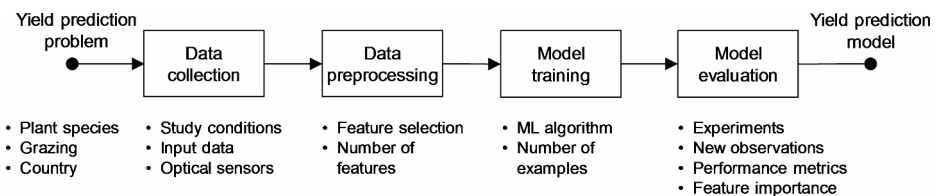

Fig. 1 Process and principal data items for pasture yield prediction

measurements (Xue & Su, 2017); spectral bands taken from imagery; textural features as properties of the surface calculated based on the Grey Level Co-occurrence Matrix (GLCM) method (Haralick et al., 1973); sward height above the ground; weather data (e.g., precipitation, temperature) (Yao et al., 2022); site data (e.g., soil type, angle); and agronomic data (e.g., fertilizer input, irrigation, grazing rotation, stocking rates, species selection, and pest and weed control) (Smit et al., 2008). Another classification of input data is based on the dichotomy of biotic and abiotic factors affecting plant growth (Lange et al., 2014). Biotic factors refer to living organisms, such as grazing animals, insects, microorganisms, and other plants that influence pasture production (Kallenbach, 2015; Klaus et al., 2013). Abiotic factors encompass non-living elements, such as soil composition, temperature, water supply, and global radiation that determine plant growth (Baldocchi et al., 2004; Sorkau et al., 2018). A sound knowledge of biotic factors, abiotic factors, and their complex interactions helps to develop effective prediction models, although the conceptual differences between explanatory and predictive modeling need to be considered (Shmueli, 2010).

Optical sensors for gathering input data can be categorized as follows. In-field sensors operate near the ground and foremost include field spectroradiometers for obtaining reflectance data, such as vegetation indices and chlorophyll content of plants, but also laser scanners, such as LiDAR (Light Detection and Ranging), to create a 3D map of the pasture. Aerial sensors are mounted on an aircraft or unmanned aerial vehicle (UAV) to collect high-resolution imagery from a low flying height (Feng et al., 2021); they include hyperspectral, multispectral, and RGB cameras as well as thermal sensors and LiDAR sensors. Satellite remote sensing enables to record vegetation reflectance from the orbit.

Data preprocessing is the second phase, which produces so-called features from the input data. Features represent characteristics of the empirical phenomenon on which a prediction model can be learned. Including all input data as features in the prediction model incurs the risk of learning from noise in the data and lacking in prediction performance. For this reason, *feature selection* provides different techniques for identifying smaller sets of features. The techniques are usually grouped into three categories (Chandrashekar & Sahin, 2014): (1) Filter-based techniques select features based on a metric calculated for each feature. For instance, correlation analysis can identify pairs of highly correlated features from which only one feature will be retained. Another technique is principal component analysis (PCA), which transforms a set of strongly correlated features into a smaller set. (2) Wrapper-based techniques remove one or more features from the initial set by iteratively training and evaluating alternative prediction models. For instance, backward elimination starts with the full set of features and removes features based on *p*values passing a specific threshold (in case of multiple linear regression). (3) Embedded techniques are specific to an ML algorithm. One example is Random Forests feature selection, which calculates the so-called feature importance metric and then removes features that do not pass a threshold for the metric.

The third phase is model training in which example observations are used to learn a function that best maps a set of feature values to the corresponding observed yield; these examples are also referred to as input-output pairs. For estimating the mapping function, the field of supervised machine learning provides a large variety of *ML algorithms*. Frequently used algorithms for predicting pasture yields include Random Forests (RF) (Ho, 1995), Artificial Neural Networks (ANN) (Bishop, 2006), and Support Vector Regression (SVR) (Drucker et al., 1996) but also different types of linear regression, such as ordinary least squares (OLS)

and partial least squares (PLS) regression. All these algorithms require a sufficiently large training set that includes a *number of examples*.

Model evaluation is the final phase, which assesses the prediction performance of a trained model. Because of the many design alternatives to choose from in the preceding phases, developers usually evaluate alternative prediction models through experiments in which one or more factors are manipulated. By conducting factorial experiments, developers can devise a variety of experimental conditions, gain insights into how the factors affect performance, and eventually identify the best-performing model. Irrespective of the experimental design, evaluation calls for testing the prediction model on *new observations*, thus observations that were not included in the training phase. The evaluation can be accomplished using cross-validation, a test set, or both techniques. In cross-validation (CV), the data set is iteratively divided into subsets for training and testing. For instance, *k*-fold CV divides the data set into *k* subsets (folds) of equal size, trains a model for each combination of *k-1* folds, evaluates the model on the left-out fold, and reports the mean performance for all *k* models. In other words, in each iteration, one fold is left out of the training set. Another type of CV is leave-one-out, which trains the model using all but one observation and tests the model on the left-out observation. The training and testing must be repeated *n* times, with *n* standing for the total number of observations. Different from CV, a test set indicates a technique that uses a separate data set of new observations. Either technique can apply *performance metrics* to quantify the accuracy of predicted vis-á-vis observed yields. Metrics for yield predictions include, for instance, the coefficient of determination (R^2), the root mean square error (RMSE), the normalized RMSE (NRMSE), and the mean absolute error (MAE). The report on prediction performance can be supplemented by information on so-called *feature importance*, i.e., a quantitative assessment of the extent to which individual features have contributed to the prediction performance. Various techniques are available for measuring importance, for example, by indicating how a specific performance metric would change in absolute or relative terms if the feature in question were removed. Other techniques specify importance as a percentage value, summing to 100% for all features. Such information is typically presented in column charts or placed in tabular appendices.

Results

Study selection

Figure 2 shows the PRISMA flow chart of study selection. A total of 591 records were identified through database searching. We considered two additional studies that reported prediction models for pasture yield using optical sensing; the studies were listed in Scopus, but their records were not automatically retrieved. Of the 91 articles that were forwarded to the full-text assessment, 43 articles fulfilled the eligibility criteria, and these studies were included in the review.

Table 1 shows an overview of the included studies. The most frequently studied plants were perennial ryegrass (14 studies, *Lolium perenne*), clover species (8, *Trifolium*), signalgrass (5, *Brachiaria*), and timothy (4, *Phleum pratense*). Twenty-five of the pastures were mechanically harvested, and the remaining pastures were grazed by animals. Twenty-two studies were conducted in Europe, nine in Australia, six in South America, and four in North
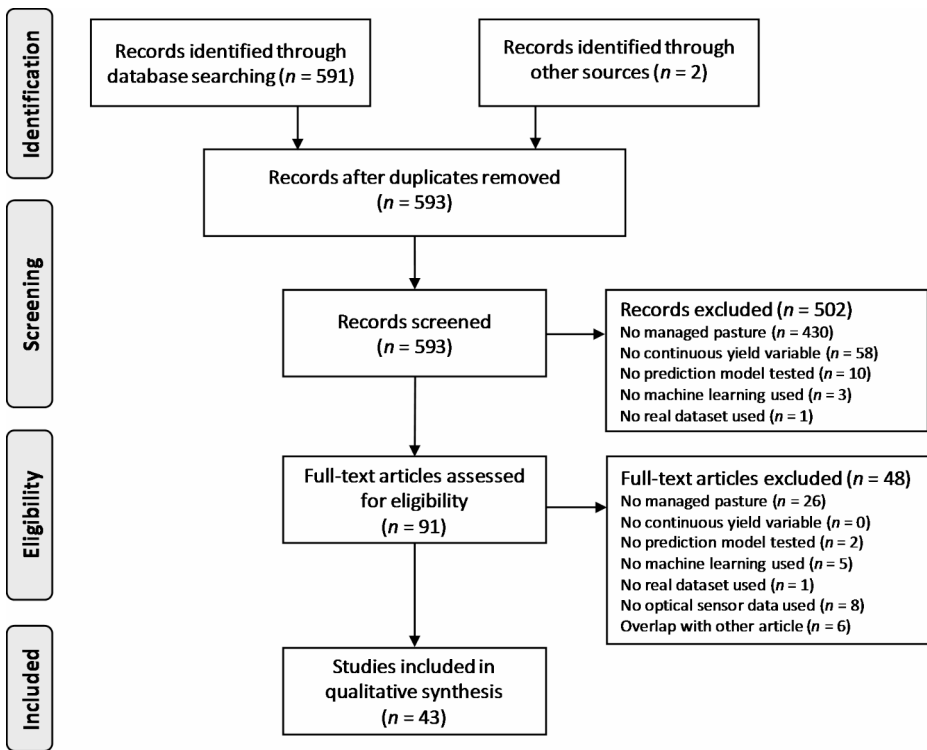
Fig. 2 PRISMA flow chart

America, whereas only each one study was carried out in Africa and Asia. Predictions were always made right before harvest or very near to that day, except for three studies that used prediction horizons of 13 days (Schwieder et al., 2020), 38 days (Li et al., 2021), and 152 days (Hamada et al., 2021), respectively.

Data collection

Table 2 provides the number of fields, sample plots, and seasons, followed by the different types of input data. Almost half of the studies were limited to data from a single field. The number of sample plots ranged from only two to more than one thousand (mean: 114; median: 54; $n=35$). Two-thirds of the studies collected data in one growing season, and every fifth study covered two seasons. Two studies even covered eight (Jaberalansar et al., 2017) and twelve seasons (Ali et al., 2017), respectively. Twothirds of the studies were conducted at research facilities and one-third on fields operated by farmers (not tabulated).

Vegetation indices were processed as input data in 29 studies. The number of VIs spanned from one index in four studies to more than 20 indices in eight studies. Sward height was used in 19 studies and it was either measured by UAV (13 studies), rising plate meter (3), LiDAR (1), meter ruler (1), or satellite (1). Nineteen studies processed spectral bands, which exhibited large variability between 2 and 2150 different bands. In three studies, vegetation indices were complemented with textural features. All other types of input data played a

Table 1 Overview of the studies (N=43)

Study	Plant species	Grazing	Country
Ali et al. (2017)	NR	Yes	Ireland
Askari et al. (2019)	*Trifolium repens, Lolium perenne*	Yes	Ireland
Barnetson et al. (2021)	Various	Yes	Australia
Borra-Serrano et al. (2019)	*Lolium perenne*	–	Belgium
Bretas et al. (2021)	*Brachiaria* species	Yes	Brazil
Chen et al. (2021)	*Lolium perenne, Trifolium repens*	Yes	Australia
Chiarito et al. (2021)	NR	Yes	Italy
Da Silva et al. (2022)	*Brachiaria ruziziensis, Pennisetum glaucum*	Yes	Brazil
Rosa et al. (2021)	*Lolium perenne*	Yes	Australia
Dos Reis et al. (2020)	*Brachiaria ruziziensis, Pennisetum glaucum*	Yes	Brazil
Franceschini et al. (2022)	*Lolium perenne*	–	Germany
Freitas et al. (2022)	*Brachiaria ruziziensis, Pennisetum glaucum*	Yes	Brazil
Geipel et al. (2021)	*Phleum pratense, Festuca pratensis, Trifolium pratense*	–	Norway
Grüner et al. (2019)	*Lolium multiflorum, Festuca pratensis, Trifolium* species, *Medicago sativa*	–	Germany
Grüner et al. (2020)	*Lolium multiflorum, Festuca pratensis, Trifolium* species, *Medicago sativa*	–	Germany
Hamada et al. (2021)	*Panicum virgatum*	–	USA
Jaberalansar et al. (2017)	NR	–	Iran
Jackman et al. (2021)	NR	–	Ireland
Karunaratne et al. (2020)	*Lolium perenne*	Yes	Australia
Li et al. (2021)	*Trifolium pratense, Festuca pratensis*	–	Estonia
Lussem et al. (2020)	Various	–	Germany
Lussem et al. (2022)	*Lolium perenne, Alopecurus* species, *Festuca* species	–	Germany
Mundava et al. (2015)	Various	Yes	Australia
Murphy et al. (2022)	*Lolium perenne*	–	Ireland
Näsi et al. (2018)	*Phleum pratense, Festuca pratensis*	–	Finland
Nguyen et al. (2022)	*Lolium perenne*	–	Australia
de Oliveira et al. (2020)	*Phleum pratense, Festuca pratensis*	–	Finland
de Oliveira et al. (2021)	*Megathyrsus maximus*	–	Brazil

Table 1 (continued)

Study	Plant species	Grazing	Country
Pereira et al. (2022)	*Brachiaria ruziziensis, Pennisetum glaucum*	Yes	Brazil
Pranga et al. (2021)	*Lolium perenne*	–	Belgium
Raab et al. (2020)	Various	–	Germany
Schucknecht et al. (2022)	*Lolium multiflorum, Trifolium repens, Lolium perenne, Poa pratensis*	Yes	Germany
Schulze-Brüninghoff et al. (2021)	*Nardus stricta, Trisetum flavescens*	Yes	Germany
Schwieder et al. (2020)	Various	–	Germany
Sibanda et al. (2017)	Various	–	South Africa
Sun et al. (2022)	*Phleum pratense, Trifolium pratense*	–	Sweden
Théau et al. (2021)	Various	Yes	Canada
Thomson et al. (2020)	*Lolium perenne*	Yes	Australia
Togeiro de Alckmin et al. (2021)	*Lolium perenne*	–	Australia
Togeiro de Alckmin et al. (2022)	*Lolium perenne*	–	Australia
van der Merwe et al. (2020)	*Andropogon gerardi, Panicum virgatum, Sorghastrum nutans*	–	USA
Zeng and Chen (2018)	*Thinopyrum species, Medicago sativa*	–	USA
Zhou et al. (2019)	Various grass and legume species	–	Sweden

Note. NR=not reported

Table 2 Data collection in the included studies (*N*=43)

Study	Fields	Sample plots	Seasons	Input data VI	Spectral bands	Texture	Sward height	Other
Ali et al. (2017)	2	2	12	5	2	–	–	–
Askari et al. (2019)	1	117	2	21	–	–	–	–
Barnetson et al. (2021)	2	12	2	5	–	–	UAV	–
Borra-Serrano et al. (2019)	1	39	1	2	–	–	UAV	W
Bretas et al. (2021)	1	40	5	3	–	–	–	W
Chen et al. (2021)	334	NR	2	1	10	–	–	W
Chiarito et al. (2021)	2	NR	3	–	1024	–	–	–
Da Silva et al. (2022)	4	100	1	1	–	–	–	–
Rosa et al. (2021)	2	18	3	–	–	–	–	S, W
Dos Reis et al. (2020)	4	100	1	24	–	24	–	–
Franceschini et al. (2022)	1	60	2	–	–	–	–	A
Freitas et al. (2022)	4	100	1	25	5	12	–	–
Geipel et al. (2021)	2	264	2	2	–	–	UAV	–
Grüner et al. (2019)	1	24	1	–	–	–	UAV	–
Grüner et al. (2020)	1	48	1	13	4	32	–	–
Hamada et al. (2021)	5	79	1	20	–	–	–	–
Jaberalansari et al. (2017)	16	NR	8	2	–	–	–	S, W
Jackman et al. (2021)	NR	NR	1	–	12	–	–	–
Karunaratne et al. (2020)	3	9	1	15	4	–	UAV	–
Li et al. (2021)	2	144	1	6	–	–	–	–
Lussem et al. (2020)	1	25	3	–	–	–	UAV	–
Lussem et al. (2022)	1	156	1	15	4	–	UAV	–
Mundava et al. (2015)	1	19	2	1	–	–	RPM	–
Murphy et al. (2022)	NR	NR	3	–	1200	–	–	–
Näsi et al. (2018)	1	32	1	13	36	–	UAV	–
Nguyen et al. (2022)	1	160	1	1	–	–	LiDAR	–
de Oliveira et al. (2020)	4	96	1	22	38	–	UAV	–

Table 2 (continued)

Study	Fields	Sample plots	Seasons	Input data				
				VI	Spectral bands	Texture	Sward height	Other
de Oliveira et al. (2021)	1	330	1	–	–	–	–	–
Pereira et al. (2022)	4	43	1	51	–	–	–	–
Pranga et al. (2021)	1	468	1	20	–	–	UAV	–
Raab et al. (2020)	5	15	3	77	10	–	–	–
Schucknecht et al. (2022)	3	NR	1	26	–	–	RPM	–
Schulze-Brüninghoff et al. (2021)	4	60	1	–	118	–	Laser	–
Schwieder et al. (2020)	1	21	1	6	3	–	–	–
Sibanda et al. (2017)	1	1080	1	12	8	–	Satellite	–
Sun et al. (2022)	NR	NR	2	–	296*	–	MR	–
Théau et al. (2021)	1	99	1	9	–	–	UAV	–
Thomson et al. (2020)	2	54	1	–	NR	–	RPM	–
Togeiro de Alckmin et al. (2021)	1	30	1	97	–	–	–	–
Togeiro de Alckmin et al. (2022)	1	30	2	6	–	–	–	–
van der Merwe et al. (2020)	13	65	1	–	–	–	UAV	–
Zeng and Chen (2018)	1	54	1	–	2150	–	–	–
Zhou et al. (2019)	2	NR	2	–	60*	–	–	–

Note. * = no exact figure reported. A = agronomic data. MR = meter ruler. NR = not reported. RPM = rising plate meter. S = site data. UAV = unmanned aerial vehicle. VI = vegetation index. W = weather data

minor role. Specifically, five studies used weather data, two studies considered site data (e.g., soil type, elevation, slope, and aspect), and only one study integrated fertilizer input as agronomic data (Franceschini et al., 2022). No study learned a model based on biotic data.

Table 3 summarizes the adoption of the different types of optical sensors. Fourteen studies obtained data from in-field sensors, which included spectroradiometers (12 studies). Twenty-four studies collected input data from cameras mounted on aerial vehicles; the cameras recorded RGB images (11 studies), multispectral images (13), and hyperspectral images (5). Thirteen studies retrieved image data from satellites, including Sentinel (9 studies), MODIS (3), PlanetScope (2), Landsat (1), PlanetDove (1), and WorldView (1).

Data preprocessing

Table 4 reports the adoption of feature selection techniques and provides the number of features per study. Feature selection was present in 24 studies, and the most frequent techniques were correlation analysis (8 studies), PCA (8), and stepwise regression (4). The number of features ranged between 1 and 101, although eight studies did not report this information. On one hand, four studies spared out feature selection but collected sensor data for a single feature. On the other hand, 18 studies trained models from at least 10 different features.

Model training

Table 5 provides information about the adoption of 16 different ML algorithms. The most frequent algorithms were Random Forests (20 studies), PLS regression (13), OLS regression using a single predictor (10) or multiple predictors (8), and Support Vector Regression (8). The size of the training set was stated in 41 studies, which either reported the number of examples, a percentage value of the examples used, or both types of information (not tabulated). In 10 studies, the prediction models were trained on less than 100 examples.

Model evaluation

Experimental manipulation

Table 6 shows the frequency of each manipulated factor. The most frequent factors were feature set (19 studies) and ML algorithm (17). The former studies compared the performance of prediction models using different combinations of features, such as vegetation indices, spectral bands, sward height, and weather features. Nine studies investigated alternative sensors (e.g., UAV versus satellite). Six further factors were only examined in one study each. The number of manipulated factors per study was either one (16 studies), two (15), or three (2), whereas no manipulation was present in ten studies.

Model assessment

Table 7 shows the adoption of cross-validation and test set as techniques for model assessment. Twelve studies were limited to cross-validation, and another 13 studies only used a test set including new observations. The remaining 18 studies applied both techniques.

Table 3 Optical sensors used in the included studies ($N=43$)

Study	In-field	Aerial	Satellite
Ali et al. (2017)	–	–	MODIS
Askari et al. (2019)	Spectroradiometer	Multispectral	Sentinel-2
Barnetson et al. (2021)	–	RGB	PlanetDove
Borra-Serrano et al. (2019)	–	RGB	–
Bretas et al. (2021)	–	–	Landsat-8, Sentinel-2
Chen et al. (2021)	–	–	Sentinel-2
Chiarito et al. (2021)	Spectroradiometer	–	Sentinel-1
Da Silva et al. (2022)	–	–	MODIS, Sentinel-2
Rosa et al. (2021)	–	Multispectral	–
Dos Reis et al. (2020)	–	–	PlanetScope
Franceschini et al. (2022)	–	Hyperspectral, RGB	–
Freitas et al. (2022)	–	Multispectral	–
Geipel et al. (2021)	–	Hyperspectral	–
Grüner et al. (2019)	–	RGB	–
Grüner et al. (2020)	–	Multispectral	–
Hamada et al. (2021)	Spectroradiometer	–	Sentinel-2
Jaberalansar et al. (2017)	–	–	MODIS
Jackman et al. (2021)	Spectroradiometer	–	–
Karunaratne et al. (2020)	–	Multispectral	–
Li et al. (2021)	–	Multispectral	–
Lussem et al. (2020)	–	RGB	–
Lussem et al. (2022)	–	Multispectral, RGB	–
Mundava et al. (2015)	Spectroradiometer	–	–
Murphy et al. (2022)	Spectroradiometer	–	–
Näsi et al. (2018)	–	Hyperspectral, RGB	–
Nguyen et al. (2022)	LiDAR vehicle	Multispectral	–
de Oliveira et al. (2020)	–	Hyperspectral, RGB	–
de Oliveira et al. (2021)	–	RGB	–
Pereira et al. (2022)	–	Multispectral	PlanetScope, Sentinel-2
Pranga et al. (2021)	–	Multispectral, RGB	–
Raab et al. (2020)	–	–	Sentinel-1/2
Schucknecht et al. (2022)	–	Multispectral	–
Schulze-Brüninghoff et al. (2021)	Laser scanner	Hyperspectral	–
Schwieder et al. (2020)	Spectroradiometer	–	Sentinel-2
Sibanda et al. (2017)	–	–	WorldView-3
Sun et al. (2022)	Spectroradiometer	–	–
Théau et al. (2021)	–	Multispectral	–
Thomson et al. (2020)	Spectroradiometer	–	–
Togeiro de Alckmin et al. (2021)	Spectroradiometer	–	–
Togeiro de Alckmin et al. (2022)	Spectroradiometer	Multispectral	–
van der Merwe et al. (2020)	–	RGB	–
Zeng and Chen (2018)	Spectroradiometer	–	–
Zhou et al. (2019)	Spectroradiometer	–	–

Note. RGB=red, green, and blue

Table 4 Feature selection techniques and number of features in the included studies ($N=43$)

Study	Technique	Number of features
Ali et al. (2017)	PCA	5
Askari et al. (2019)	OLS: correlations, stepwise regression; PLS: Marten's test	OLS: 2–9; PLS: 4–8
Barnetson et al. (2021)	–	NR
Borra-Serrano et al. (2019)	OLS: correlation; PLS: separate PCA	5
Bretas et al. (2021)	Stepwise regression (backward elimination)	OLS: 6; RF: 9
Chen et al. (2021)	–	19
Chiarito et al. (2021)	Genetic algorithm	13
Da Silva et al. (2022)	–	1
Rosa et al. (2021)	–	4–8
Dos Reis et al. (2020)	RReliefF	24
Franceschini et al. (2022)	Variable Importance in Projection	101
Freitas et al. (2022)	Correlations; RReliefF	10
Geipel et al. (2021)	Built-in PCA	NR
Grüner et al. (2019)	–	1
Grüner et al. (2020)	–	49
Hamada et al. (2021)	Correlations	1
Jaberalansar et al. (2017)	Correlations	16
Jackman et al. (2021)	PCA, genetic algorithm	12
Karunaratne et al. (2020)	–	30
Li et al. (2021)	–	NR
Lussem et al. (2020)	–	1
Lussem et al. (2022)	–	35
Mundava et al. (2015)	Stepwise regression (forward selection)	3
Murphy et al. (2022)	PCA	NR
Näsi et al. (2018)	–	49
Nguyen et al. (2022)	Correlations; branch-and-bound algorithm	2; 3; 4
de Oliveira et al. (2020)	–	3–38
de Oliveira et al. (2021)	–	NR
Pereira et al. (2022)	–	3–60
Pranga et al. (2021)	–	27
Raab et al. (2020)	Permutation-based selection	13
Schucknecht et al. (2022)	–	26
Schulze-Brüninghoff et al. (2021)	Variable Selection using Random Forests	29
Schwieder et al. (2020)	–	9
Sibanda et al. (2017)	Sparse PLS regression	19
Sun et al. (2022)	PLS: Variable Importance in Projection; SVR: –	PLS: 10; SVR: NR
Théau et al. (2021)	Coefficient of variation; simple linear regression	NR
Thomson et al. (2020)	PCA, spectral analysis	11
Togeiro de Alckmin et al. (2021)	Correlations; recursive elimination	20
Togeiro de Alckmin et al. (2022)	PCA	4
van der Merwe et al. (2020)	–	1
Zeng and Chen (2018)	PLS: correlations; MLR: stepwise (backward elimination)	PLS: 2150, MLR: 7
Zhou et al. (2019)	PLS: built-in PCA; SVR: –	NR

Note. NR=not reported. PCA=principal component analysis. OLS=ordinal least squares. PLS=partial least squares. RF=Random Forests. RRreliefF=regressional ReliefF algorithm. SVR=Support Vector Regression

Table 5 Machine-learning algorithms in the included studies ($N=43$)

Algorithm	No.	Studies
Random Forests	20	Borra-Serrano et al. (2019); Bretas et al. (2021); Da Silva et al. (2022); Rosa et al. (2021); Dos Reis et al. (2020); Freitas et al. (2022); Grüner et al. (2019); Karunaratne et al. (2020); Li et al. (2021); Lussem et al. (2022); Näsi et al. (2018); de Oliveira et al. (2020); Pereira et al. (2022); Pranga et al. (2021); Raab et al. (2020); Schucknecht et al. (2022); Schulze-Brüninghoff et al. (2021); Schwieder et al. (2020); Togeiro de Alckmin et al. (2021); Togeiro de Alckmin et al. (2022)
Partial least squares regression	13	Askari et al. (2019); Borra-Serrano et al. (2019); Franceschini et al. (2022); Geipel et al. (2021); Jackman et al. (2021); Lussem et al. (2022); Murphy et al. (2022); Pranga et al. (2021); Sibanda et al. (2017); Sun et al. (2022); Thomson et al. (2020); Zeng and Chen (2018); Zhou et al. (2019)
Simple linear regression	10	Borra-Serrano et al. (2019); Bretas et al. (2021); Geipel et al. (2021); Grüner et al. (2019); Hamada et al. (2021); Lussem et al. (2020); Lussem et al. (2022); Pranga et al. (2021); Togeiro de Alckmin et al. (2021); van der Merwe et al. (2020)
Multiple linear regression	8	Ali et al. (2017); Askari et al. (2019); Borra-Serrano et al. (2019); Bretas et al. (2021); Grüner et al. (2020); Mundava et al. (2015); Nguyen et al. (2022); de Oliveira et al. (2020)
Support Vector Regression	8	Chiarito et al. (2021); Li et al. (2021); Lussem et al. (2022); Pranga et al. (2021); Sun et al. (2022); Thomson et al. (2020); Togeiro de Alckmin et al. (2021); Zhou et al. (2019)
Artificial Neural Networks	4	Ali et al. (2017); Barnetson et al. (2021); Chen et al. (2021); Li et al. (2021)
Gradient Boosting	3	Dos Reis et al. (2020); Togeiro de Alckmin et al. (2022); Schucknecht et al. (2022)
Ensemble learner	2	Thomson et al. (2020); Togeiro de Alckmin et al. (2021)
Bagged Trees	1	Togeiro de Alckmin et al. (2022)
Classification and Regression Trees	1	Togeiro de Alckmin et al. (2022)
Convolutional Neural Networks	1	de Oliveira et al. (2021)
Cubist	1	Togeiro de Alckmin et al. (2022)
General Additive Modeling	1	Rosa et al. (2021)
Multivariate adaptive regression spline	1	Togeiro de Alckmin et al. (2021)
Nonlinear regression	1	Théau et al. (2021)
Nonparametric multiplicative regression	1	Jaberalansar et al. (2017)

We note that the application and reporting of cross-validation exhibit large variation by leaving out one fold (19 studies), one example (7), or one site (1) in the training phase, whereas four studies provided no information in that respect. Regarding the test set, 29 of the 31 studies reported its size as a percentage of the whole data set. The number of examples ranged between 10 and 433. Six studies had very small test sets with at most 24 examples, but ten studies had much larger test sets with more than one hundred examples.

Table 6 Manipulated factors in the included studies ($N=43$)

Factor	No.	Studies
Feature set	19	Borra-Serrano et al. (2019); Chen et al. (2021); Rosa et al. (2021); Dos Reis et al. (2020); Freitas et al. (2022); Grüner et al. (2020); Jaberalansar et al. (2017); Karunaratne et al. (2020); Lussem et al. (2020); Lussem et al. (2022); Näsi et al. (2018); Nguyen et al. (2022); de Oliveira et al. (2020); Pereira et al. (2022); Pranga et al. (2021); Schucknecht et al. (2022); Sibanda et al. (2017); Sun et al. (2022); Togeiro de Alckmin et al. (2021)
ML algorithm	17	Ali et al. (2017); Askari et al. (2019); Borra-Serrano et al. (2019); Bretas et al. (2021); Rosa et al. (2021); Dos Reis et al. (2020); Grüner et al. (2020); Li et al. (2021); Lussem et al. (2022); de Oliveira et al. (2020); Pranga et al. (2021); Schucknecht et al. (2022); Sun et al. (2022); Thomson et al. (2020); Togeiro de Alckmin et al. (2021); Togeiro de Alckmin et al. (2022); Zhou et al. (2019)
Sensor	9	Askari et al. (2019); Barnetson et al. (2021); Da Silva et al. (2022); Grüner et al. (2019); Pereira et al. (2022); Raab et al. (2020); Schucknecht et al. (2022); Schulze-Brüninghoff et al. (2021); Togeiro de Alckmin et al. (2022)
Flying altitude	2	Karunaratne et al. (2020); Näsi et al. (2018)
CNN architecture	1	de Oliveira et al. (2021)
Feature selection	1	Freitas et al. (2022)
Prediction horizon	1	Li et al. (2021)
Test set sampling	1	Franceschini et al. (2022)
Training set ratio	1	Rosa et al. (2021)
None	10	Chiarito et al. (2021); Geipel et al. (2021); Hamada et al. (2021); Jackman et al. (2021); Mundava et al. (2015); Murphy et al. (2022); Schwieder et al. (2020); Théau et al. (2021); van der Merwe et al. (2020); Zeng and Chen (2018)

Note. CNN=Convolutional Neural Networks

Performance metrics

Table 8 reveals that R^2 was reported in all but five studies. Thirty studies provided the root mean square error in kg per ha. Unitless normalizations of the RMSE were present in 25 studies. This normalization was either based on the mean (11 studies), range (4), standard deviation (2), and interquartile range (1) of the observed yield, or its specification was missing (7).

Performance by types of optical sensors

The high adoption rate of the unitless R^2 metric allowed us to collate performance results as shown in Fig. 3, which groups 41 prediction models by the types of optical sensors used. The R^2 ranged between 0.42 and 0.90 in the satellite group, between 0.50 and 0.94 in the aerial sensors group, and between 0.62 and 0.92 in the in-field sensors group. The range was even smaller for the few prediction models that complemented in-field data by satellite data (0.71 to 0.90) and UAV data (0.81 to 0.92), respectively. Similarly, the mean performance in the in-field (0.79) and aerial groups (0.77) was higher than for the satellite (0.67) group. Two studies that processed data from aerial sensors and satellites at the same time reported R^2 values of 0.70 (Pereira et al., 2022) and 0.72 (Barnetson et al., 2021), respectively (not shown in Fig. 3).

Table 7 Model assessment in the included studies (*N*=43)

Study	Cross-validation	Test set: Percentage	Test set: Examples
Ali et al. (2017)	–	30	>180*
Askari et al. (2019)	Not specified	30, 33	28, 37, 53
Barnetson et al. (2021)	–	30	19, 30
Borra-Serrano et al. (2019)	–	30	46*
Bretas et al. (2021)	5-fold	20	24
Chen et al. (2021)	–	25	433
Chiarito et al. (2021)	5-fold	20	12*
Da Silva et al. (2022)	5-fold	30	99
Rosa et al. (2021)	–	30, 50, 70	151, 252, 420
Dos Reis et al. (2020)	5-fold	30	104*
Franceschini et al. (2022)	3-fold (100 iterations)	30	58
Freitas et al. (2022)	10-fold	30	54
Geipel et al. (2021)	–	50	353
Grüner et al. (2019)	–	25	45
Grüner et al. (2020)	4-fold (100 iterations)	–	–
Hamada et al. (2021)	–	25 (31 runs)	21
Jaberalansar et al. (2017)	Not specified	–	–
Jackman et al. (2021)	10-fold (5 iterations)	–	–
Karunaratne et al. (2020)	10-fold	20	20
Li et al. (2021)	Leave-one-out (36)	50	36
Lussem et al. (2020)	Leave-one-out (140)	–	–
Lussem et al. (2022)	10-fold (5 iterations)	–	–
Mundava et al. (2015)	Leave-one-out (NR); Leave-one-site-out	–	–
Murphy et al. (2022)	4-fold	11	197
Näsi et al. (2018)	Leave-one-out (32)	–	–
Nguyen et al. (2022)	3 to 30 folds; Leave-one-out (368)	40	147*
de Oliveira et al. (2020)	Leave-one-out (NR)	NR	10
de Oliveira et al. (2021)	10-fold	–	–
Pereira et al. (2022)	10-fold	30	34*
Pranga et al. (2021)	10-fold	–	–
Raab et al. (2020)	10-fold (1000 iterations)	–	–
Schucknecht et al. (2022)	6-fold (10 iterations)	–	–
Schulze-Brüninghoff et al. (2021)	–	15 (100 runs)	33
Schwieder et al. (2020)	–	30 (100 runs)	13
Sibanda et al. (2017)	Leave-one-out (1080)	–	–
Sun et al. (2022)	10-fold	33	78
Théau et al. (2021)	–	50	50
Thomson et al. (2020)	20-fold	25	51
Togeiro de Alckmin et al. (2021)	10-fold (5 iterations)	30	270
Togeiro de Alckmin et al. (2022)	Not specified	38	300
van der Merwe et al. (2020)	–	26	19
Zeng and Chen (2018)	Not specified	35	35
Zhou et al. (2019)	–	33	126

Note. NR=not reported. * = no exact figure reported

Table 8 Performance metrics reported in the included studies (N=43)

Study	R^2	RMSE	NRMSE	MAE	Other
Ali et al. (2017)	Yes	–	–	–	–
Askari et al. (2019)	Yes	Yes	–	Yes	Yes
Barnetson et al. (2021)	Yes	–	–	Yes	–
Borra-Serrano et al. (2019)	Yes	Yes	Yes[5]	–	Yes
Bretas et al. (2021)	Yes	Yes	–	–	–
Chen et al. (2021)	Yes	Yes	–	Yes	–
Chiarito et al. (2021)	Yes	Yes	–	–	–
Da Silva et al. (2022)	Yes	–	Yes[5]	–	–
Rosa et al. (2021)	Yes	–	Yes[1]	–	Yes
Dos Reis et al. (2020)	Yes	Yes	Yes[5]	–	–
Franceschini et al. (2022)	Yes	Yes	Yes[4]	–	–
Freitas et al. (2022)	Yes	Yes	Yes[5]	–	–
Geipel et al. (2021)	Yes	Yes	Yes[5]	–	–
Grüner et al. (2019)	Yes	Yes	Yes[1]	–	Yes
Grüner et al. (2020)	Yes	Yes	Yes[2]	–	–
Hamada et al. (2021)	Yes	Yes	–	Yes	–
Jaberalansar et al. (2017)	Yes	Yes	Yes[3]	Yes	–
Jackman et al. (2021)	Yes	–	Yes[1]	–	–
Karunaratne et al. (2020)	Yes	Yes	Yes[1]	–	Yes
Li et al. (2021)	Yes	–	Yes[2]	–	–
Lussem et al. (2020)	Yes	Yes	–	–	–
Lussem et al. (2022)	Yes	Yes	Yes[5]	–	–
Mundava et al. (2015)	–	–	–	–	Yes
Murphy et al. (2022)	Yes	–	–	–	Yes
Näsi et al. (2018)	–	Yes	Yes[5]	–	Yes
Nguyen et al. (2022)	Yes	Yes	–	Yes	Yes
de Oliveira et al. (2020)	–	Yes	Yes[1]	–	–
de Oliveira et al. (2021)	–	Yes	–	Yes	Yes
Pereira et al. (2022)	Yes	Yes	Yes[3]	Yes	Yes
Pranga et al. (2021)	–	Yes	Yes[1]	–	–
Raab et al. (2020)	Yes	Yes	–	–	–
Schucknecht et al. (2022)	Yes	Yes	Yes[2]	–	–
Schulze-Brüninghoff et al. (2021)	Yes	–	Yes[2]	–	–
Schwieder et al. (2020)	Yes	–	Yes[1]	–	–
Sibanda et al. (2017)	Yes	Yes	Yes[5]	–	–
Sun et al. (2022)	Yes	–	Yes[1]	–	–
Théau et al. (2021)	Yes	Yes	Yes[1]	–	Yes
Thomson et al. (2020)	Yes	Yes	Yes[1]	–	Yes
Togeiro de Alckmin et al. (2021)	Yes	Yes	–	Yes	–
Togeiro de Alckmin et al. (2022)	Yes	Yes	–	–	Yes
van der Merwe et al. (2020)	Yes	–	–	–	Yes
Zeng and Chen (2018)	Yes	–	–	–	–
Zhou et al. (2019)	Yes	Yes	–	Yes	–

Note. MAE=mean absolute error. NRMSE=normalized RMSE by (1) mean, (2) range, (3) standard deviation, (4) interquartile range, or (5) not specified. R^2=coefficient of determination. RMSE=root mean square error
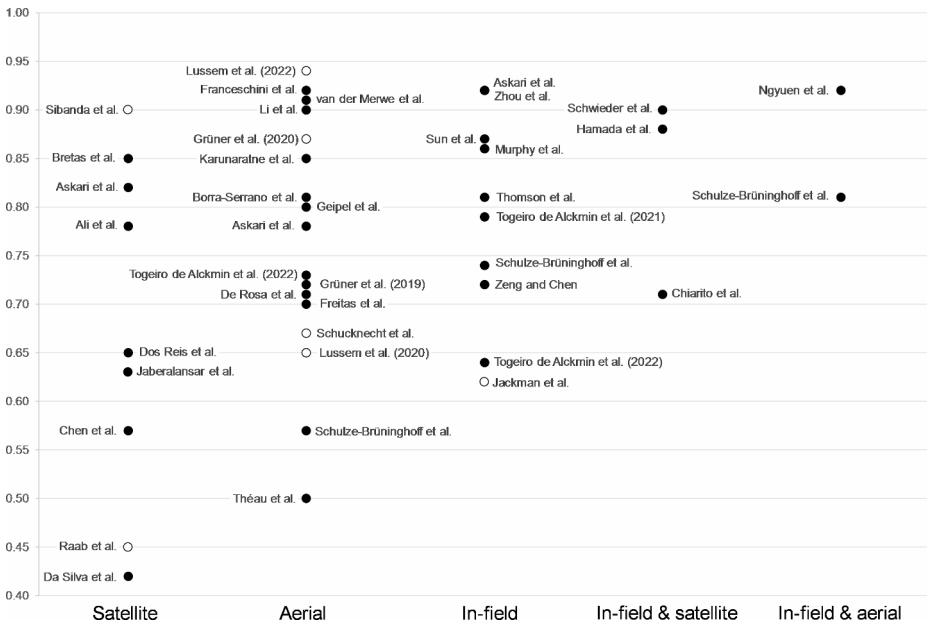
Fig. 3 R^2 of 41 prediction models by the types of optical sensors used (dots with filling indicate R^2 on the test set, dots without filling stand for cross-validation)

Table 9 Techniques for feature importance reported in the included studies ($N=43$)

Technique	No.	Studies
Not specified	6	Bretas et al. (2021); Freitas et al. (2022); Grüner et al. (2020); Karunaratne et al. (2020); Näsi et al. (2018); Sibanda et al. (2017)
Increase in mean square error (IncMSE)	3	Schucknecht et al. (2022); Schulze-Brüninghoff et al. (2021); Schwieder et al. (2020)
Variable importance in the projection (VIP)	2	Franceschini et al. (2022); Sun et al. (2022)
Decrease in R^2	1	Raab et al. (2020)
Feature importances (skitlearn)	1	Li et al. (2021)
Gini importance	1	Togeiro de Alckmin et al. (2021)
Increase in RMSE	1	Lussem et al. (2022)
Information gain (XGBoost)	1	Dos Reis et al. (2020)
Variable importance measure (permimp)	1	Pranga et al. (2021)

Feature importance

Seventeen studies provided information on the importance of features. As shown in Table 9, many different techniques have been adopted, including variable importance in projection (Chong & Jun, 2005) and increase in RMSE (Kuhn, 2008). Six studies reported on the

importance but lacked a specification of the technique used. In fifteen studies, only features based on spectral data were assessed (which is consistent with the focus on spectral variables in the data collection). In one study, the highest feature importance was assigned to canopy height (Schucknecht et al., 2022), and another study found that the relative importance of three weather features was one third, while three vegetation indices contributed two thirds (Bretas et al., 2021).

Discussion

This review examined the adoption of machine-learning techniques for pasture yield prediction using optical sensor data. We analyzed forty-three studies that have been published in journals between 2015-01-01 and 2022-10-26. This section discusses the principal findings of the review and draws implications for future research and the reporting of studies. We also discuss the limitations of our review.

Data collection

For assessing the reliability of a trained prediction model, the number of fields, plots, and seasons are important factors that determine the size of the training set. These numbers can provide indications of how far the temporal and spatial variability of pasture yields have been considered in the data collection. Yet, more than half of the studies were conducted in one season, which restricts the training data to specific weather and growing conditions. One-third of the studies were limited to data from one field in one season. Even in the latter group of studies, the number of sample plots exhibited an enormous span from 21 to 1080. The chances that a model will perform similarly in future seasons can be enhanced by training the model on multi-seasonal data of different fields in which many plots capture the in-field variability. However, this approach puts a burden on researchers and farmers because the required effort for data collection increases significantly.

The restriction of many studies to a single field limits the applicability of the trained models to a local level; hence, conclusions about their prediction performance beyond the specific field cannot be drawn. Three approaches are feasible to develop global models. First, data from a larger number of fields from different regions can be collected to develop models from data of greater heterogeneity. Second, data representing biotic and abiotic factors can be integrated to represent a larger set of growing conditions, thereby incorporating these factors into the development. A third possibility is to train a global prediction model by integrating multiple local models, i.e., the reuse of training data from different local sources (Liu et al., 2022).

The included studies do not inform developers about the minimum size of the training set to achieve a reasonable level of prediction performance. One study manipulated the size but it found only marginal effects on performance (Rosa et al., 2021). The results of our review highlight that little is known about how to specify the size of the training set. This finding points to the need for further examination of the relationship between the training set and prediction performance.

A common theme in the studies is the application of features derived from optical data, namely VIs (29 studies), spectral bands (19), textures (19), and sward height (15). Most

studies collected data from at least two different types and considered multiple input variables, either alternatively or supplementary. The focus is on exploiting the potential of optical sensing and techniques for transforming image data into variables that are associated with plant growth. Therefore, it is not surprising that the number of VIs (1 to 97), spectral bands (2 to 2150), and textures (8 to 32) varied considerably across studies. It is noteworthy that the most frequent approach for measuring the sward height was UAV, which has substituted the manual measurement using hand-held devices. Consistent with the large array of VIs, spectral bands, textures, and sward height variables collected in combination, the types of optical sensors used indicate comprehensive coverage of current sensor technologies.

All other types of input data, such as weather data (5 studies), site data (2), and agronomic data (1) played a minor role. This finding is surprising in view of the fact that such data can be obtained with relatively little effort or are already available. For instance, the retrieval of weather data is facilitated by online portal and programming interfaces (Jaffrés, 2019). Historic and current weather data specifically for agricultural purposes are offered by companies, national meteorological agencies, and agricultural departments (Farmers Guide, 2022). Site and agronomic data are increasingly recorded by farmers and processed in digital farm management information systems. These data represent a so far hardly exploited potential for supplementing the spectral data and thus for training even more accurate prediction models.

Data preprocessing

Given the often-high dimensionality of the optical input data used, the objective of data processing is to reduce the number of features derived from the input data. This dimensionality is foremost due to the large number of VIs and spectral bands in many studies. The results of our review provide evidence for the relevance of feature selection, which was present in more than half of the studies (it was not relevant for four studies that collected data for a single feature). Overall, the techniques used span across filter-based, wrapper-based, and embedded techniques, and reflect the variety of techniques available from the ML literature. In many studies, the number of features was effectively reduced without increasing the error of prediction. For instance, one study started with a set of 20 VIs and eventually selected one VI based on correlation analysis (Hamada et al., 2021). Another study considered 2150 different spectral bands obtained from a hand-held spectroradiometer and then performed stepwise regression using backward elimination to arrive at a linear model with only seven features (Zeng & Chen, 2018). Similarly, a study by Chiarito et al. (2021) applied a genetic algorithm to an initial set of 1024 spectral bands to choose a 13-feature model.

Concerning the final set of features processed in the training phase, our review identified six studies in which the lack of feature selection led to models trained from at least 26 features (Grüner et al., 2020; Karunaratne et al., 2020; Lussem et al., 2022; Näsi et al., 2018; Pranga et al., 2021; Schucknecht et al., 2022). For these studies, it might be possible that a model with fewer features exists that would perform similarly. Therefore, we recommend exploring the impact of feature selection on prediction performance when the model was trained on a large number of features derived from optical data. For studies that adopt a hypothetico-deductive approach to the model development, we suggest determining the relative importance of each feature and relating the results back to the model development. We also note that almost one-fifth of the studies provided no information on the number of fea-

tures. Presenting complete information on the features included in the trained model would help assess the input data that necessarily must be collected and data that can be omitted. This information can effectively be reported in a table showing each feature along its unit of measurement and definition (Chen et al., 2021). In case the number of features exceeds the possibilities of a table, the information can be summarized by indicating the numbers per type of sensor as well as the initial and final numbers (Schulze-Brüninghoff et al., 2021).

Model training

The highest adoption rates were found for Random Forests, PLS regression, and OLS regression, whereas only four studies used Artificial Neural Networks. It is noteworthy that PLS regression was employed in 13 studies. PLS regression is a form of multiple linear regression in which the number of initially used independent variables can automatically be reduced by an in-built principal component analysis to a smaller set of features. Therefore, PLS regression appears specifically relevant for the training from spectral input data, while it still assumes linear relationships between the features and the pasture yield. However, two studies that compared the performance of PLS and MLR models reported conflicting results concerning the R^2 metric (Askari et al., 2019; Borra-Serrano et al., 2019). Regarding the performance of linear vis-á-vis non-linear models, four studies found better performance for non-linear models, but three other studies came to the opposite conclusion. Overall, the results of our review suggest no evidence for the superiority of any group of learning algorithms.

Forty studies stated the size of the training set, which varied greatly because of the different numbers of fields, sampling plots, and seasons between studies. This size must be seen in the context of the temporal and spatial variability of the specific pasture under study. If a prediction model is learned from too few observations that do not sufficiently represent this variability, the model will perform very differently for other test data. Our review identified two studies that had very small training sets of 32 (Näsi et al., 2018) and 36 observations (Li et al., 2021), respectively. Although no clear guidance is available for determining the minimum size required, it can give readers a hint about the reliability of the prediction model. Therefore, we suggest reporting complete and unequivocal information on the training and test sets used. The reporting should always include the absolute and relative numbers for each data set. For instance, this information can be visualized in a flow diagram that specifies the data processing (Murphy et al., 2022), or provided in a single sentence, such as the following: "the complete dataset was first divided into two parts in a 70:30 ratio (231 observations for the training dataset and 99 observations for the test dataset)" (Da Silva et al., 2022, pp. 6039–6040). Moreover, any removal of observations throughout the preprocessing and learning phases should be justified, instead of generally referring to so-called outlier removal.

Model evaluation

Insights into the role of specific ML techniques for achieving high performance can be obtained from systematic testing of alternative models. The preferred method is the controlled experiment in which one or more factors are manipulated (33 of 43 studies). Our review identified feature sets, ML algorithms, and optical sensors as the most frequently

used factors. The relevance of feature sets and sensors can be explained by the range of input data to capture the temporal and spatial variability of pasture yields.

The evaluation should be conducted in a way that can mitigate the overfitting of a learned prediction model. A model is said to overfit if it fits well to the training set but exhibits much lower performance on new observations. For instance, it might be possible that a model that has been trained on a rich set of sensor data collected from several fields in different seasons will perform much worse for a different field or in a future season. To address the overfitting problem, cross-validation is appropriate for situations in which the total number of observations is insufficiently large to divide them into training and test sets. Different types of cross-validation are available, and they have been adopted in several ways. With respect to *k*-fold CV, the number of folds ranged between 3 and 20, with no study providing a rationale for the number chosen. This deficit also holds for the number of iterations of a CV; iterations were present in seven studies, and they spanned from only 5 to 1000. Leave-one-out CV must be regarded as the weakest type of CV because it likely overstates the prediction performance for extremely small ratios of new observations, such as 1080:1 in one study (Sibanda et al., 2017).

Almost three-fourths of the included studies assessed performance on a separate test set, instead of solely relying on cross-validation. Two studies stand out that trained models in one season (Togeiro de Alckmin et al., 2022) or two seasons (Murphy et al., 2022) to evaluate them in the subsequent season. Another study performed the evaluation on test data from a different field in the same season (de Oliveira et al., 2020). The size of the test set was most often stated explicitly in a table or text in the results section. For some articles, the size could be derived from a percentage value in relation to the entire data set. In seven other studies, the reporting in that respect was incomplete or inconsistent, but we could determine the number of observations through counting of dots shown in scatter plots. In a similar vein as for the training set, we recommend specifying the test procedure including the absolute and relative number of observations, any further data cleansing that led to the removal of observations, and the number of runs (if relevant).

Regarding the reporting of performance metrics, our review provides three major findings that have implications for the interpretation of study results. The first finding is the relatively high adoption of the R^2 (91%), followed by the RMSE (71%). Because R^2 is a unitless metric, it can in principle help compare the performance of different models that predict the same type of pasture yield. Its interpretation is less dependent on the study context compared to the RMSE. The latter metric is more informative for farmers of the specific fields from which the observations had been collected and it can help assess how useful the prediction model is compared to current means of yield prediction. However, the RMSE cannot be used to compare models that have been trained on different observations. In other words, any claim about a proposed model that its RMSE would be smaller than that of a model proposed in previous research must be taken with great caution. In view of the duality of metrics that serve very different purposes, we contend that there is no reason *not* to report the R^2 and RMSE.

The second finding concerns the heterogeneity and ambiguity of the normalized RMSE. Given that the NRMSE was reported in more than half of the studies, this unitless metric might be used for comparing the results of similar studies. Unfortunately, we identified at least four different definitions, which render the comparison of results impossible. Only eleven studies reported the RMSE divided by the mean observed yield. This definition

might be regarded as the 'common' meaning of the NRMSE, but seven studies provided no further details. This practice is questionable because readers might assume a definition that is different from the calculation done in the study. To make matters worse, the abstracts of 14 studies only indicated the NRMSE but not its exact definition. Moreover, different designations were used including NRMSE, nRMSE, relative RMSE (rRMSE), RRMSE, RMSE%, and RMSE with a percentage value. Taken together, our finding highlights the need for greater clarity in the reporting to avoid misinterpretations of the NRMSE.

The third finding is the lack of consensus regarding the reporting of metrics to comprehensively describe prediction performance. The three most commonly used metrics (R^2, RMSE, and NRMSE) were only reported in one-third of the studies. The frequencies of metrics such as MAE (10), Lin's concordance correlation coefficient (2), Willmott's d (2), and mean average percentage error (1) were negligible, although other literature has highlighted the usefulness of these supplementary metrics (Chai & Draxler, 2014; Willmott & Matsuura, 2005). Irrespective of the advantages and disadvantages of specific metrics, we recommend reporting a larger set of metrics, including but not limited to the R^2, RMSE, and NRMSE.

Because of the large heterogeneity in the reporting of metrics discussed above, our analysis of prediction performance by the types of optical sensors used was limited to the R^2 metric reported for 41 prediction models from 36 studies. We found that the variability of R^2 decreased for smaller distance from the pasture sward; thus, the largest variability was observed for models trained from satellite data and it decreased considerably for aerial sensors and in-field sensors. These results suggest that the higher spatial and temporal resolution of in-field imagery can make a difference in the training of effective models.

Although the individual contribution of specific features to prediction performance can be determined using feature importance, this was the case in only 17 studies. All but two of these studies focused on features derived from spectral data. Therefore, there is yet little evidence for the roles of biotic versus abiotic features, including weather, site, and agronomic features. The results of previous studies using optical sensing do not inform us about the influence of such features on the accuracy of models. Opportunities exist to examine the supplementing roles of non-spectral features, especially of those features for which data collection is relatively easy or the data is already available from the farmers.

Collectively, the results of our review also highlight challenges for prediction models to become less local and increasingly global. Especially the large differences of pastures regarding soil properties, weather conditions, and plant species make the development of generalizable prediction model challenging. The collection of data reflecting biotic and abiotic factors is not possible by remote sensing alone. To develop more globally applicable models, it is necessary to include data from complementary sources (e.g., weather stations, soil analysis, and farm management information systems).

Limitations

The results of this review should be understood in light of its limitations. The included studies varied greatly in the processing of input data and how prediction models were trained. Therefore, it was not possible to conduct in-depth comparisons of performance results, except for the types of optical sensors used. Another limitation of the review is due to the large heterogeneity of the reporting of performance metrics. This heterogeneity limited

the number of studies for which performance results could be synthesized; this synthesis was only possible for the R^2 but not for the NRMSE. Third, although our data collection involved three independent coders, all results presented in this review were bound to the information reported in the original studies (in a few cases, we contacted the authors to clarify the meaning of specific statements though).

Conclusion

This systematic review provides a comprehensive account of the application of ML for the prediction of pasture yield using optical sensor data. The results highlight the richness of techniques used for the collection and preprocessing of input data as well as the training and evaluation of prediction models. Our review also revealed some shortcomings in the assessment of prediction performance and the presentation of study designs and results. Specifically, we identified deficits in the reporting of feature sets and complete information on the training and test data, and a lack of consensus in the reporting of performance metrics. We suggest specific recommendations to enhance the uniformity and comparability of study results, which can then facilitate the integration of evidence on the role of specific ML techniques for accurate and reliable yield predictions of managed pastures.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Adão, T., Hruška, J., Pádua, L., Bessa, J., Peres, E., Morais, R., & Sousa, J. (2017). Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry. *Remote Sensing*, *9*(11), 1110. https://doi.org/10.3390/rs9111110.

Ali, I., Cawkwell, F., Dwyer, E., & Green, S. (2017). Modeling managed grassland biomass estimation by using multitemporal remote sensing data—a machine learning approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(7), 3254–3264. https://doi.org/10.1109/JSTARS.2016.2561618.

Askari, M. S., McCarthy, T., Magee, A., & Murphy, D. J. (2019). Evaluation of grass quality under different soil management scenarios using remote sensing techniques. *Remote Sensing*, *11*(15), 1835. https://doi.org/10.3390/rs11151835.

Baldocchi, D. D., Xu, L., & Kiang, N. (2004). How plant functional-type, weather, seasonal drought, and soil physical properties alter water and energy fluxes of an oak–grass savanna and an annual grassland. *Agricultural and Forest Meteorology*, *123*(1–2), 13–39. https://doi.org/10.1016/j.agrformet.2003.11.006.

Barnetson, J., Phinn, S., & Scarth, P. (2021). Climate-resilient grazing in the pastures of Queensland: An integrated remotely piloted aircraft system and satellite-based deep-learning method for estimating pasture yield. *AgriEngineering*, *3*(3), 681–703. https://doi.org/10.3390/agriengineering3030044.

Bazzo, C. O. G., Kamali, B., Hütt, C., Bareth, G., & Gaiser, T. (2023). A review of estimation methods for aboveground biomass in grasslands using UAV. *Remote Sensing*, *15*(3), 639. https://doi.org/10.3390/rs15030639.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Borra-Serrano, I., de Swaef, T., Muylle, H., Nuyttens, D., Vangeyte, J., Mertens, K., Saeys, W., Somers, B., Roldán-Ruiz, I., & Lootens, P. (2019). Canopy height measurements and non-destructive biomass estimation of Lolium perenne swards using UAV imagery. *Grass and Forage Science*, *74*(3), 356–369. https://doi.org/10.1111/gfs.12439.

Bouwman, A. F., van der Hoek, K. W., Eickhout, B., & Soenario, I. (2005). Exploring changes in world ruminant production systems. *Agricultural Systems*, *84*(2), 121–153. https://doi.org/10.1016/j.agsy.2004.05.006.

Bretas, I. L., Valente, D. S. M., Silva, F. F., Chizzotti, M. L., Paulino, M. F., D'Áurea, A. P., Paciullo, D. S., Pedreira, B. C., & Chizzotti, F. H. M (2021). Prediction of aboveground biomass and dry-matter content in Brachiaria pastures by combining meteorological data and satellite imagery. *Grass and Forage Science*, *76*(3), 340–352. https://doi.org/10.1111/gfs.12517.

Buckley, C., Wall, D. P., Moran, B., O'Neill, S., & Murphy, P. N. C. (2016). Farm gate level nitrogen balance and use efficiency changes post implementation of the EU Nitrates Directive. *Nutrient Cycling in Agroecosystems*, *104*(1), 1–13. https://doi.org/10.1007/s10705-015-9753-y.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014.

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, *40*(1), 16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024.

Chen, Y., Guerschman, J., Shendryk, Y., Henry, D., & Harrison, M. T. (2021). Estimating pasture biomass using Sentinel-2 imagery and machine learning. *Remote Sensing*, *13*(4), https://doi.org/10.3390/rs13040603.

Chiarito, E., Cigna, F., Cuozzo, G., Fontanelli, G., Mejia Aguilar, A., Paloscia, S., Rossi, M., Santi, E., Tapete, D., & Notarnicola, C. (2021). Biomass retrieval based on genetic algorithm feature selection and support vector regression in Alpine grassland using ground-based hyperspectral and Sentinel-1 SAR data. *European Journal of Remote Sensing*, *54*(1), 209–225. https://doi.org/10.1080/22797254.2021.1901063.

Chong, I. G., & Jun, C. H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, *78*(1–2), 103–112. https://doi.org/10.1016/j.chemolab.2004.12.011.

Da Silva, Y. F., Dos Reis, A., Sampaio Werner, A., Valadares, J. P. V., Campbell, R., Camargo Lamparelli, E. E. A., Magalhães, R., P. S. G., & Figueiredo, G. K. D. A (2022). Assessing the capability of MODIS to monitor mixed pastures with high-intensity grazing at a fine-scale. *Geocarto International*, *37*(20), 6033–6051. https://doi.org/10.1080/10106049.2021.1926559.

de Oliveira, R. A., Näsi, R., Niemeläinen, O., Nyholm, L., Alhonoja, K., Kaivosoja, J., Jauhiainen, L., Viljanen, N., Nezami, S., Markelin, L., Hakala, T., & Honkavaara, E. (2020). Machine learning estimators for the quantity and quality of grass swards used for silage production using drone-based imaging spectrometry and photogrammetry. *Remote Sensing of Environment*, *246*, 111830. https://doi.org/10.1016/j.rse.2020.111830.

de Oliveira, G. S., Marcato Junior, J., Polidoro, C., Osco, L. P., Siqueira, H., Rodrigues, L., Jank, L., Barrios, S., Valle, C., Simeão, R., Carromeu, C., Silveira, E., de Castro Jorge, A., Gonçalves, L., Santos, W., M., & Matsubara, E. (2021). Convolutional neural networks to estimate dry matter yield in a Guineagrass breeding program using UAV remote sensing. *Sensors (Basel, Switzerland)*, *21*(12), https://doi.org/10.3390/s21123971.

De Rosa, D., Basso, B., Fasiolo, M., Friedl, J., Fulkerson, B., Grace, P. R., & Rowlings, D. W. (2021). Predicting pasture biomass using a statistical model and machine learning algorithm implemented with remotely sensed imagery. *Computers and Electronics in Agriculture*, *180*, 105880. https://doi.org/10.1016/j.compag.2020.105880.

Togeiro de Alckmin G., Kooistra, L., Rawnsley, R., & Lucieer, A. (2021). Comparing methods to estimate perennial ryegrass biomass: Canopy height and spectral vegetation indices. *Precision Agriculture*, *22*(1), 205–225. https://doi.org/10.1007/s11119-020-09737-z.

Togeiro de Alckmin G., Lucieer, A., Rawnsley, R., & Kooistra, L. (2022). Perennial ryegrass biomass retrieval through multispectral UAV data. *Computers and Electronics in Agriculture*, *193*, 106574. https://doi.org/10.1016/j.compag.2021.106574.

Donnison, I. S., & Fraser, M. D. (2016). Diversification and use of bioenergy to maintain future grasslands. *Food and Energy Security*, *5*(2), 67–75. https://doi.org/10.1002/fes3.75.

Dos Reis, A. A., Werner, J. P. S., Silva, B. C., Figueiredo, G. K. D. A., Antunes, J. F. G., Esquerdo, J. C. D. M., Coutinho, A. C., Lamparelli, R. A. C., Rocha, J. V., & Magalhães, P. S. G. (2020). Monitoring pasture aboveground biomass and canopy height in an integrated crop–livestock system using textural information from PlanetScope imagery. *Remote Sensing*, *12*(16), 2534. https://doi.org/10.3390/RS12162534.

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1996). Support Vector Regression Machines. In M. C. Mozer, M. Jordan, & T. Petsche (Eds.), *Advances in neural information processing systems* (Vol. 9). MIT Press. https://proceedings.neurips.cc/paper/1996/file/d38901788c533e8286cb6400b40b386d-Paper.pdf.

Farmers Guide (2022). *UK-wide weather station network provides data to aid agronomic decision making*. https://www.farmersguide.co.uk/local-weather-data-aids-agronomic-decision-making/.

Feng, L., Chen, S., Zhang, C., Zhang, Y., & He, Y. (2021). A comprehensive review on recent applications of unmanned aerial vehicle remote sensing with various sensors for high-throughput plant phenotyping. *Computers and Electronics in Agriculture*, *182*, 106033. https://doi.org/10.1016/j.compag.2021.106033.

Franceschini, M. H. D., Becker, R., Wichern, F., & Kooistra, L. (2022). Quantification of grassland biomass and nitrogen content through UAV hyperspectral imagery—active sample selection for model transfer. *Drones*, *6*(3), 73. https://doi.org/10.3390/drones6030073.

Freitas, R. G., Pereira, F. R., Reis, D., Magalhães, A. A., Figueiredo, P. S. G., D., G. K. A.,do, & Amaral, L. R. (2022). Estimating pasture aboveground biomass under an integrated crop-livestock system based on spectral and texture measures derived from UAV images. *Computers and Electronics in Agriculture*, *198*, Article 107122. https://doi.org/10.1016/j.compag.2022.107122.

Geipel, J., Bakken, A. K., Jørgensen, M., & Korsaeth, A. (2021). Forage yield and quality estimation by means of UAV and hyperspectral imaging. *Precision Agriculture*, *22*(5), 1437–1463. https://doi.org/10.1007/s11119-021-09790-2.

Grüner, E., Astor, T., & Wachendorf, M. (2019). Biomass prediction of heterogeneous temperate grasslands using an SfM approach based on UAV imaging. *Agronomy*, *9*(2), 54. https://doi.org/10.3390/agronomy9020054.

Grüner, E., Wachendorf, M., & Astor, T. (2020). The potential of UAV-borne spectral and textural information for predicting aboveground biomass and N fixation in legume-grass mixtures. *PLOS ONE*, *15*(6), e0234703. https://doi.org/10.1371/journal.pone.0234703.

Hamada, Y., Zumpf, C. R., Cacho, J. F., Lee, D., Lin, C. H., Boe, A., Heaton, E., Mitchell, R., & Negri, M. C. (2021). Remote sensing-based estimation of advanced perennial grass biomass yields for bioenergy. *Land*, *10*(11), 1221. https://doi.org/10.3390/land10111221.

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-3*(6), 610–621. https://doi.org/10.1109/TSMC.1973.4309314.

Hedley, C. (2015). The role of precision agriculture for improved nutrient management on farms. *Journal of the Science of Food and Agriculture*, *95*(1), 12–19. https://doi.org/10.1002/jsfa.6734.

Henchion, M., Hayes, M., Mullen, A. M., Fenelon, M., & Tiwari, B. (2017). Future protein supply and demand: Strategies and factors influencing a sustainable equilibrium. *Foods*, *6*(7), https://doi.org/10.3390/foods6070053.

Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition (ICDAR)* (pp. 278–282). IEEE. https://doi.org/10.1109/ICDAR.1995.598994.

Hopkins, A., & Del Prado, A. (2007). Implications of climate change for grassland in Europe: Impacts, adaptations and mitigation options: A review. *Grass and Forage Science*, *62*(2), 118–126. https://doi.org/10.1111/j.1365-2494.2007.00575.x.

Jaberalansar, Z., Tarkesh, M., & Bassiri, M. (2017). Soil moisture index improves models of forage production in central Iran. *Archives of Agronomy and Soil Science*, *63*(12), 1763–1775. https://doi.org/10.1080/03650340.2017.1296138.

Jackman, P., Lee, T., French, M., Sasikumar, J., O'Byrne, P., Berry, D., Lacey, A., & Ross, R. (2021). Predicting key grassland characteristics from hyperspectral data. *AgriEngineering*, *3*(2), 313–322. https://doi.org/10.3390/agriengineering3020021.

Jaffrés, J. B. (2019). GHCN-Daily: A treasure trove of climate data awaiting discovery. *Computers & Geosciences*, *122*, 35–44. https://doi.org/10.1016/j.cageo.2018.07.003.

Kallenbach, R. L. (2015). Describing the dynamic: Measuring and assessing the value of plants in the pasture. *Crop Science*, *55*(6), 2531–2539. https://doi.org/10.2135/cropsci2015.01.0065.

Karunaratne, S., Thomson, A., Morse-McNabb, E., Wijesingha, J., Stayches, D., Copland, A., & Jacobs, J. (2020). The fusion of spectral and structural datasets derived from an airborne multispectral sensor for estimation of pasture dry matter yield at paddock scale with time. *Remote Sensing*, *12*(12), 2017. https://doi.org/10.3390/rs12122017.

Kent Shannon, D., Clay, D. E., & Sudduth, K. A. (2018). An introduction to precision agriculture. In D. K. Shannon, D. E. Clay, & N. R. Kitchen (Eds.), *Precision agriculture basics* (pp. 1–12). John Wiley & Sons. https://doi.org/10.2134/precisionagbasics.2016.0084.

Klaus, V. H., Kleinebecker, T., Prati, D., Gossner, M. M., Alt, F., Boch, S., Gockel, S., Hemp, A., Lange, M., Müller, J., Oelmann, Y., Pašalić, E., Renner, S. C., Socher, S. A., Türke, M., Weisser, W. W., Fischer, M., & Hölzel, N. (2013). Does organic grassland farming benefit plant and arthropod diversity at the expense of yield and soil fertility? *Agriculture Ecosystems & Environment*, *177*, 1–9. https://doi.org/10.1016/j.agee.2013.05.019.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5), https://doi.org/10.18637/jss.v028.i05.

Lange, M., Habekost, M., Eisenhauer, N., Roscher, C., Bessler, H., Engels, C., Oelmann, Y., Scheu, S., Wilcke, W., Schulze, E. D., & Gleixner, G. (2014). Biotic and abiotic properties mediating plant diversity effects on soil microbial communities in an experimental grassland. *PLOS ONE*, *9*(5), e96182. https://doi.org/10.1371/journal.pone.0096182.

Li, K. Y., Burnside, N. G., Sampaio de Lima, R., Villoslada Peciña, M., Sepp, K., Yang, M. D., Raet, J., Vain, A., Selge, A., & Sepp, K. (2021). The application of an unmanned aerial system and machine learning techniques for red clover-grass mixture yield estimation under variety performance trials. *Remote Sensing*, *13*(10), 1994. https://doi.org/10.3390/rs13101994.

Liu, J., [Ji], Huang, J., Zhou, Y., Li, X., [Xuhong], Ji, S., Xiong, H., & Dou, D. (2022). From distributed machine learning to federated learning: A survey. *Knowledge and Information Systems*, *64*(4), 885–917. https://doi.org/10.1007/s10115-022-01664-x.

Lussem, U., Schellberg, J., & Bareth, G. (2020). Monitoring forage mass with low-cost UAV data: Case study at the Rengen Grassland Experiment. *PFG – Journal of Photogrammetry Remote Sensing and Geoinformation Science*, *88*(5), 407–422. https://doi.org/10.1007/s41064-020-00117-w.

Lussem, U., Bolten, A., Kleppert, I., Jasper, J., Gnyp, M. L., Schellberg, J., & Bareth, G. (2022). Herbage mass, N concentration, and N uptake of temperate grasslands can adequately be estimated from UAV-based image data using machine learning. *Remote Sensing*, *14*(13), 3066. https://doi.org/10.3390/rs14133066.

Lyu, X., Li, X., [Xiaobing], Dang, D., Dou, H., Wang, K., & Lou, A. (2022). Unmanned aerial vehicle (UAV) remote sensing in grassland ecosystem monitoring: A systematic review. *Remote Sensing*, *14*(5), 1096. https://doi.org/10.3390/rs14051096.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, *151*(4), 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135.

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of web of Science and Scopus: A comparative analysis. *Scientometrics*, *106*(1), 213–228. https://doi.org/10.1007/s11192-015-1765-5.

Morais, T. G., Teixeira, R. F., Figueiredo, M., & Domingos, T. (2021). The use of machine learning methods to estimate aboveground biomass of grasslands: A review. *Ecological Indicators*, *130*, Article 108081. https://doi.org/10.1016/j.ecolind.2021.108081.

Mundava, C., Schut, A. G. T., Helmholz, P., Stovold, R., Donald, G., & Lamb, D. W. (2015). A novel protocol for assessment of aboveground biomass in rangeland environments. *The Rangeland Journal*, *37*(2), 157. https://doi.org/10.1071/RJ14072.

Murphy, D. J., Murphy, M. D., O'Brien, B., & O'Donovan, M. (2021). A review of precision technologies for optimising pasture measurement on irish grassland. *Agriculture*, *11*(7), 600. https://doi.org/10.3390/agriculture11070600.

Murphy, D. J., O' Brien, B., O' Donovan, M., Condon, T., & Murphy, M. D. (2022). A near infrared spectroscopy calibration for the prediction of fresh grass quality on irish pastures. *Information Processing in Agriculture*, *9*(2), 243–253. https://doi.org/10.1016/j.inpa.2021.04.012.

Näsi, R., Viljanen, N., Kaivosoja, J., Alhonoja, K., Hakala, T., Markelin, L., & Honkavaara, E. (2018). Estimating biomass and nitrogen amount of barley and grass using UAV and aircraft based spectral and photogrammetric 3D features. *Remote Sensing*, *10*(7), 1082. https://doi.org/10.3390/rs10071082.

Nguyen, P. T., Shi, F., Wang, J., Badenhorst, P. E., Spangenberg, G. C., Smith, K. F., & Daetwyler, H. D. (2022). Within and combined season prediction models for perennial ryegrass biomass yield using ground- and air-based sensor data. *Frontiers in Plant Science*, *13*, 950720. https://doi.org/10.3389/fpls.2022.950720.

O'Mara, F. P. (2012). The role of grasslands in food security and climate change. *Annals of Botany*, *110*(6), 1263–1270. https://doi.org/10.1093/aob/mcs209.

Pereira, F. S., de Lima, J. P., Freitas, R. G., Reis, D., Amaral, A. A., Figueiredo, L., Lamparelli, G. K. D. A., R., & Magalhães, P. (2022). Nitrogen variability assessment of pasture fields under an integrated crop-livestock system using UAV, PlanetScope, and Sentinel-2 data. *Computers and Electronics in Agriculture*, *193*, Article 106645. https://doi.org/10.1016/j.compag.2021.106645.

Pranga, J., Borra-Serrano, I., Aper, J., de Swaef, T., an, Ghesquiere, Quataert, P., Roldán-Ruiz, I., Janssens, I. A., Ruysschaert, G., & Lootens, P. (2021). Improving accuracy of herbage yield predictions in perennial ryegrass with UAV-based structural and spectral data fusion and machine learning. *Remote Sensing*, *13*(17), 3459. https://doi.org/10.3390/rs13173459.

Raab, C., Riesch, F., Tonn, B., Barrett, B., Meißner, M., Balkenhol, N., & Isselstein, J. (2020). Target-oriented habitat and wildlife management: Estimating forage quantity and quality of semi-natural grasslands with Sentinel-1 and Sentinel-2 data. *Remote Sensing in Ecology and Conservation*, *6*(3), 381–398. https://doi.org/10.1002/rse2.149.

Schucknecht, A., Seo, B., Krämer, A., Asam, S., Atzberger, C., & Kiese, R. (2022). Estimating dry biomass and plant nitrogen concentration in pre-alpine grasslands with low-cost UAS-borne multispectral data – a comparison of sensors, algorithms, and predictor sets. *Biogeosciences*, *19*(10), 2699–2727. https://doi.org/10.5194/bg-19-2699-2022.

Schulze-Brüninghoff, D., Wachendorf, M., & Astor, T. (2021). Remote sensing data fusion as a tool for biomass prediction in extensive grasslands invaded by L. polyphyllus. *Remote Sensing in Ecology and Conservation*, *7*(2), 198–213. https://doi.org/10.1002/rse2.182.

Schwieder, M., Buddeberg, M., Kowalski, K., Pfoch, K., Bartsch, J., Bach, H., Pickert, J., & Hostert, P. (2020). Estimating grassland parameters from Sentinel-2: A model comparison study. *PFG – Journal of Photogrammetry Remote Sensing and Geoinformation Science*, *88*(5), 379–390. https://doi.org/10.1007/s41064-020-00120-1.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), https://doi.org/10.1214/10-STS330.

Sibanda, M., Mutanga, O., Rouget, M., & Kumar, L. (2017). Estimating biomass of native grass grown under complex management treatments using WorldView-3 spectral derivatives. *Remote Sensing*, *9*(1), 55. https://doi.org/10.3390/rs9010055.

Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, *126*(6), 5113–5142. https://doi.org/10.1007/s11192-021-03948-5.

Smit, H. J., Metzger, M. J., & Ewert, F. (2008). Spatial distribution of grassland productivity and land use in Europe. *Agricultural Systems*, *98*(3), 208–219. https://doi.org/10.1016/j.agsy.2008.07.004.

Sorkau, E., Boch, S., Boeddinghaus, R. S., Bonkowski, M., Fischer, M., Kandeler, E., Klaus, V. H., Kleinebecker, T., Marhan, S., Müller, J., Prati, D., Schöning, I., Schrumpf, M., Weinert, J., & Oelmann, Y. (2018). The role of soil chemical properties, land use and plant diversity for microbial phosphorus in forest and grassland soils. *Journal of Plant Nutrition and Soil Science*, *181*(2), 185–197. https://doi.org/10.1002/jpln.201700082.

Squires, V. R., Dengler, J., Feng, H., & Hua, L. (Eds.). (2018). *Grasslands of the world: Diversity, management and conservation*. CRC PRESS. https://www.taylorfrancis.com/books/9781315156125.

Subhashree, S. N., Igathinathane, C., Akyuz, A., Borhan, M., Hendrickson, J., Archer, D., Liebig, M., Toledo, D., Sedivec, K., Kronberg, S., & Halvorson, J. (2023). Tools for predicting forage growth in rangelands and economic analyses—a systematic review. *Agriculture*, *13*(2), 455. https://doi.org/10.3390/agriculture13020455.

Sun, S., Zuo, Z., Yue, W., Morel, J., Parsons, D., Liu, J., [Jian], Peng, J., Cen, H., He, Y., Shi, J., Li, X., [Xiaolong], & Zhou, Z. (2022). Estimation of biomass and nutritive value of grass and clover mixtures by analyzing spectral and crop height data using chemometric methods. *Computers and Electronics in Agriculture*, *192*, 106571. https://doi.org/10.1016/j.compag.2021.106571.

Théau, J., Lauzier-Hudon, É., Aubé, L., & Devillers, N. (2021). Estimation of forage biomass and vegetation cover in grasslands using UAV imagery. *PLOS ONE*, *16*(1), e0245784. https://doi.org/10.1371/journal.pone.0245784.

Thelwall, M., & Sud, P. (2022). Scopus 1900–2020: Growth in articles, abstracts, countries, fields, and journals. *Quantitative Science Studies*, *3*(1), 37–50. https://doi.org/10.1162/qss_a_00177.

Thomson, A. L., Karunaratne, S. B., Copland, A., Stayches, D., McNabb, E. M., & Jacobs, J. (2020). Use of traditional, modern, and hybrid modelling approaches for in situ prediction of dry matter yield and nutritive characteristics of pasture using hyperspectral datasets. *Animal Feed Science and Technology*, *269*, 114670. https://doi.org/10.1016/j.anifeedsci.2020.114670.

Tripathi, A. D., Mishra, R., Maurya, K. K., Singh, R. B., & Wilson, D. W. (2018). Estimates for world population and global food availability for global health. In R. R. Watson, R. B. Singh, & T. Takahashi (Eds.), *The role of functional food security in global health* (pp. 3–24). Elsevier Science & Technology. https://doi.org/10.1016/B978-0-12-813148-0.00001-3.

van der Merwe, D., Baldwin, C. E., & Boyer, W. (2020). An efficient method for estimating dormant season grass biomass in tallgrass prairie from ultra-high spatial resolution aerial imaging produced with small unmanned aircraft systems. *International Journal of Wildland Fire*, *29*(8), 696. https://doi.org/10.1071/WF19026.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*, 79–82. https://doi.org/10.3354/cr030079.

Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: a review of developments and applications. *Journal of Sensors*, *2017*, 1–17. https://doi.org/10.1155/2017/1353691.

Yao, Z., Xin, Y., Yang, L., Zhao, L., & Ali, A. (2022). Precipitation and temperature regulate species diversity, plant coverage and aboveground biomass through opposing mechanisms in large-scale grasslands. *Frontiers in Plant Science*, *13*, 999636. https://doi.org/10.3389/fpls.2022.999636.

Zeng, L., & Chen, C. (2018). Using remote sensing to estimate forage biomass and nutrient contents at different growth stages. *Biomass and Bioenergy*, *115*, 74–81. https://doi.org/10.1016/j.biombioe.2018.04.016.

Zeng, L., Wardlow, B. D., Xiang, D., Hu, S., & Li, D. (2020). A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sensing of Environment*, *237*, 111511. https://doi.org/10.1016/j.rse.2019.111511.

Zhao, Y., Liu, Z., & Wu, J. (2020). Grassland ecosystem services: A systematic review of research advances and future directions. *Landscape Ecology*, *35*(4), 793–814. https://doi.org/10.1007/s10980-020-00980-3.

Zhou, Z., Morel, J., Parsons, D., Kucheryavskiy, S. V., & Gustavsson, A. M. (2019). Estimation of yield and quality of legume and grass mixtures using partial least squares and support vector machine analysis of spectral data. *Computers and Electronics in Agriculture*, *162*, 246–253. https://doi.org/10.1016/j.compag.2019.03.038.

Authors and Affiliations

Christoph Stumpe[1] · Joerg Leukel[2] · Tobias Zimpel[2]

✉ Christoph Stumpe
christoph.stumpe@uni-hohenheim.de

Joerg Leukel
joerg.leukel@uni-hohenheim.de

Tobias Zimpel
tobias.zimpel@uni-hohenheim.de

[1] Department of Fundamentals of Agricultural Engineering, University of Hohenheim, Garbenstr. 9, 70599 Stuttgart, Germany

[2] Department of Information Systems 2, University of Hohenheim, Schwerzstr. 35, 70599 Stuttgart, Germany