



Review

Remote-Sensing Data and Deep-Learning Techniques in Crop Mapping and Yield Prediction: A Systematic Review

Abhasha Joshi ¹, Biswajeet Pradhan ^{1,*}, Shilpa Gite ^{2,3} and Subrata Chakraborty ^{1,4}

¹ Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), School of Civil and Environmental Engineering, University of Technology Sydney, Ultimo, NSW 2007, Australia

² AIML Department, Symbiosis Institute of Technology, Symbiosis International (Deemed) University, Pune 412115, India

³ Symbiosis Centre of Applied AI (SCAAI), Symbiosis International (Deemed) University, Pune 412115, India

⁴ School of Science and Technology, Faculty of Science, Agriculture, Business and Law, University of New England, Armidale, NSW 2351, Australia

* Correspondence: biswajeet.pradhan@uts.edu.au

Abstract: Reliable and timely crop-yield prediction and crop mapping are crucial for food security and decision making in the food industry and in agro-environmental management. The global coverage, rich spectral and spatial information and repetitive nature of remote sensing (RS) data have made them effective tools for mapping crop extent and predicting yield before harvesting. Advanced machine-learning methods, particularly deep learning (DL), can accurately represent the complex features essential for crop mapping and yield predictions by accounting for the nonlinear relationships between variables. The DL algorithm has attained remarkable success in different fields of RS and its use in crop monitoring is also increasing. Although a few reviews cover the use of DL techniques in broader RS and agricultural applications, only a small number of references are made to RS-based crop-mapping and yield-prediction studies. A few recently conducted reviews attempted to provide overviews of the applications of DL in crop-yield prediction. However, they did not cover crop mapping and did not consider some of the critical attributes that reveal the essential issues in the field. This study is one of the first in the literature to provide a thorough systematic review of the important scientific works related to state-of-the-art DL techniques and RS in crop mapping and yield estimation. This review systematically identified 90 papers from databases of peer-reviewed scientific publications and comprehensively reviewed the aspects related to the employed platforms, sensors, input features, architectures, frameworks, training data, spatial distributions of study sites, output scales, evaluation metrics and performances. The review suggests that multiple DL-based solutions using different RS data and DL architectures have been developed in recent years, thereby providing reliable solutions for crop mapping and yield prediction. However, challenges related to scarce training data, the development of effective, efficient and generalisable models and the transparency of predictions should be addressed to implement these solutions at scale for diverse locations and crops.

Keywords: crop monitoring; crop classification; yield prediction; deep learning; convolutional neural network (CNN); recurrent neural network (RNN)



Citation: Joshi, A.; Pradhan, B.; Gite, S.; Chakraborty, S. Remote-Sensing Data and Deep-Learning Techniques in Crop Mapping and Yield Prediction: A Systematic Review. *Remote Sens.* **2023**, *15*, 2014.

<https://doi.org/10.3390/rs15082014>

Academic Editor: Hirohiko Nagano

Received: 2 March 2023

Revised: 6 April 2023

Accepted: 7 April 2023

Published: 11 April 2023



Copyright: © 2023 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

The global population is proliferating and the demand for food is also increasing. Food production needs to be increased by 50% compared with 2013 to meet the need of approximately 9.1 billion people in 2050 [1]. Sustainable agricultural management has become an essential issue to address this challenge. Reliable and timely information on crop type and yield, along with their spatial components, play an important role in sustainable agricultural-resource management [2]. The accurate mapping of heterogeneous agricultural landscapes is an integral part of agricultural management. Crop maps are useful for

precision agriculture, the monitoring of farming activities, the preparation of crop statistics and the study of the impact of environmental factors on crops [3,4]. Regional-scale crop-yield-prediction models also use crop maps to select data. Furthermore, the accuracy of crop production depends on the accuracy of crop maps. The prediction of crop yield before harvest provides an early warning for food security and supports decision making related to the import and export of food [5,6]. This information is also helpful for the agricultural industry to determine crop pricing, insurance pricing and stock planning [7–9]. Identifying areas with low productivity and applying interventions, such as site-specific fertilisation or insecticides to increase productivity, can also be achieved with yield prediction [10,11]. Together, crop mapping and crop-yield prediction provide a complete picture of crop production and its spatial distribution. Thus, crop mapping and crop-yield prediction are essential in ensuring food security, for supply-chain management in agribusiness and in adapting crop-management practices.

Remote-sensing data captured by satellites, aeroplanes or unmanned aerial vehicles (UAVs) provide a comprehensive snapshot of our environment. The global coverage, repetitive nature, multispectral information and increasing spatial, spectral and temporal resolution of RS data make them tools for monitoring crops at local, regional and global scales. Many researchers have used RS for crop monitoring, including crop-type classification and yield prediction [12–14].

Classically, process-based crop and statistical models have been used for crop-yield prediction. The process-based crop model [15–17] attempts to simulate physical processes in crop growth and yield formation. Remote sensing (RS) data have been assimilated into crop models to provide accurate information about crop-state variables and input parameters, so that a model can be applied for yield prediction at the regional scale [18–22]. Statistical models determine the empirical relations between the yield predictor and yield using available data. One of the initial applications of RS for crop-yield prediction was the use of regression analysis [23]. For crop mapping, pixel-based [24,25] or object-based [26] supervised and unsupervised classification methods have been primarily used over the years. The process-based crop models require extensive local data and are time-consuming to run at large scales [20,27,28]. Meanwhile, traditional statistical models cannot capture complex nonlinear interactions amongst variables. Recently, machine learning (ML) algorithms, such as random forest (RF) [29], decision tree (DT) [30] and support vector machines (SVM) [31], have also been successfully applied for crop mapping [32–34] and crop-yield estimation [35–38]. The ML methods can account for the nonlinear relation between variables, but their performance relies significantly on the appropriate representation of the input data [39]. Thus, the method requires domain knowledge to handcraft optimal features [40–42]. Handcrafted features also often disregard a large amount of valuable data [43].

Deep learning is a ML method that can not only map features onto outputs but also learns appropriate features itself, thereby avoiding the need for feature engineering [39]. Deep learning is a reapplication of neural networks, in which multiple layers of neural networks are used for predictions based on available data [44]. Recently, DL has gained immense popularity in many research and business sectors, including in agricultural applications and RS-image processing [45–47]. The main contributors to this surge in DL applications include increasing computational resources and available data and the development of techniques to train deeper networks [39]. The ability of DL models to extract features automatically at different levels of abstraction from available RS and other data and to make predictions without the need to simulate complex relationships makes them valuable tools for crop monitoring. Evidently, a recent increase in interest in crop mapping and yield prediction using DL and RS has been observed.

Several studies have reviewed DL applications in RS in general [40,48–50]. However, none of these studies has comprehensive coverage in their applications in agriculture monitoring. Kamilaris and Prenafeta-Boldú [47] reviewed 40 studies regarding the use of DL in different areas of agriculture, including weed detection, disease detection, crop-type classification and erosion assessment. Nevertheless, the study only makes a few

references to RS-based crop-mapping and yield-prediction studies. Van Klompenburg et al. [51] conducted a systematic review of publications to identify features, architecture and evaluation measures regarding the application of ML in yield prediction. Furthermore, recently published reviews by Oikonomidis et al. [52] and Muruganantham et al. [53] systematically evaluated DL applications for yield prediction. However, these studies either did not focus on the RS component sufficiently or did not consider some critical attributes, such as the performances of the models, target data used, crop type and study location. Furthermore, to the best of our knowledge, no study has yet reviewed the application of DL in crop mapping.

This paper comprehensively reviews studies that used RS data and DL techniques for two critical crop-monitoring applications, namely, crop mapping and crop-yield prediction. Crop mapping mainly refers to the classification problem, whereas yield prediction is primarily a regression problem. This paper is one of the first in the literature to provide a thorough systematic review of the important scientific works related to state-of-the-art DL techniques and RS in crop mapping and yield estimation. This review aims to explore the following attributes of the studies considered:

- Platform, sensor and input features of models;
- Training data used;
- Architecture used;
- Framework used to implement the architecture;
- Crop type, site, area and scale of the studies;
- Assessment criteria and performance achieved in the studies.

The aim of this study is to summarise the existing research related to crop mapping and yield prediction using DL and RS and to highlight the gaps between them. This review minimises the barriers to implementing the solution and provides recommendations for future research directions.

2. Overview of Deep Learning (DL)

Deep learning is a ML method inspired by the structure of the human brain [39]. It involves the training of neural networks with many layers. Machine learning is a branch of artificial intelligence (AI) that allows a computer to perform tasks by learning from a large amount of data without explicitly programming the computer. Machine learning is valuable when the relationships between variables cannot be efficiently described using the traditional linear, deterministic model-building approach [54]. In DL, multiple layers learn data representation at various abstraction levels. Complex functions can be learned using DL with sufficient data and many layers that represent features at various abstractions [45]. The mainstream DL models include multilayer perceptron (MLP) [55], convolutional neural network (CNN) [55], recurrent neural network (RNN) [56] and autoencoders (AEs) [39]. Recently, a self-attention-based DL model architecture called Transformer [57] has also emerged as a key advancement in DL architecture. Figure 1 summarises the architectures of some of the leading DL models.

A MLP is the simplest form of a feedforward ANN. Each MLP layer has a set of nonlinear functions of the weighted sum of all the inputs from the previous layer [55]. They are also called deep neural networks (DNN). We use both terms interchangeably in this review. Autoencoders are neural networks used for unsupervised learning [58]. They have an encoder–decoder structure, in which the encoder represents the input data in a compressed form and the decoder decodes the representation to the original data [39]. The encoding and decoding are performed automatically, without any feature engineering. One of the purposes of AEs in RS is to reduce the dimensionality of the data.

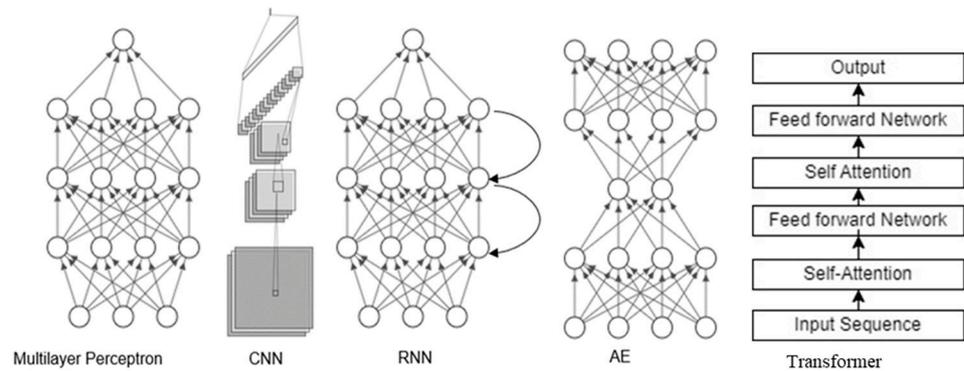


Figure 1. Different DL models: All the models have an input layer, output layer and hidden layers for encoding and decoding. Hidden layers learn representation and perform mapping based on learned features.

A CNN is also a type of feedforward neural network that is commonly used in computer vision [45]. It primarily consists of filters (convolution and pooling) that are used to extract features from the original image [39]. Different features are highlighted in various layers, thereby providing hierarchical representations of data [59]. The convolution layer acts as a feature extractor and dimensionality is reduced by the pooling layer [39]. The pooling layer also prevents the network from overfitting. Often, fully connected layers, which act as classifiers by using the high-level features learned, are found at the end of the network. One of the key advantages of CNNs is parameter sharing. The parameter size remains fixed irrespective of the size of the input grid. The successful application of CNNs dates back to the late 1990s, in LeNet [60]. However, they did not achieve momentum until the development of core-computing systems. A significant landmark in the development of CNNs was the AlexNet [61], which won the ImageNet competition by a large margin. Convolutional neural networks have been highly successful in computer-vision problems, such as image classification [62], object detection [63], the neural-style transfer of images [59] and image segmentation [64].

Recurrent neural networks can take inputs and generate outputs of different lengths. This type of network is suitable for modelling sequential data and is widely used in speech recognition [65], natural language processing [56] and time-series analysis [66]. In RNNs, each layer comprises a set of nonlinear functions of the weighted sum of all the inputs from the previous layer and ‘a state vector’ [45], which contains information about the history of all the past elements in the sequence. The problem with RNNs is the vanishing and exploding gradient during backpropagation [67]. Long short-term memory (LSTM) was developed to address the vanishing-gradient problem in RNNs [68]. The hidden layers of LSTM have memory cells that model temporal sequences and their long-range dependencies more accurately. A gated recurrent neural network is another variant of RNNs developed to solve vanishing- or exploding-gradient problems [69]. They have an update gate and a reset gate, which decides which information should be passed to the output.

Transformer architectures are also used to process sequential data. Unlike RNNs, they receive the entire input data at once and use attention mechanisms to capture the relation between the input and output [57]. Transformer allows the parallel processing of entire sequences, thereby enabling very large networks to be trained significantly faster. The use of Transformer, which was initially developed for natural language processing and is the current state-of-the-art method in the field, has also gained recognition in computer vision [70,71]. The Architecture section further describes the DL architectures used for crop mapping and yield prediction.

3. Literature Identification

A systematic search of the literature on DL- and RS-based crop mapping and yield estimation was conducted on ‘Scopus’ and ‘Web of Science’ databases, respectively. These databases are for peer-reviewed scientific publications. The title, abstract and keyword were searched using the following search string:

((‘Deep Learning’) AND (‘Remote Sensing’ OR ‘Satellite Imag*’) AND (‘Agri*’ OR ‘Crop’) AND (‘Yield’ OR ‘Production’ OR ‘Mapping’ OR ‘Classification’)).

The list was then filtered to exclude review papers and book chapters, works written in languages other than English, papers published in 2023, papers with fewer than five citations and duplicate papers.

Only English-language articles were considered due to the linguistic abilities of the authors and the lack of translation resources. Some of the important works in the field are presented at conferences. Hence, we also included conference proceedings in the review. Citation-based exclusion criteria were used to reduce the number of articles whilst ensuring that the most important and effectual studies were included in the review. The citation constraint might have excluded some high-quality and effectual research. Therefore, we read the abstracts and scanned through all the articles that were removed due to these criteria to determine any significant studies and added them to the list.

The initial search yielded 800 articles from the Scopus database and 268 articles from the Web of Science database. After applying all the exclusion criteria and removing duplicates, 267 articles were considered for further analysis, out of which 81 papers were selected by reading the abstracts and skimming through the contents. Nine relevant articles, which were identified from the bibliography or from the exclusion list, were added to the list. In this way, 90 publications were selected for review. Figure 2 graphically describes the step involving the identification of the publications for this study.

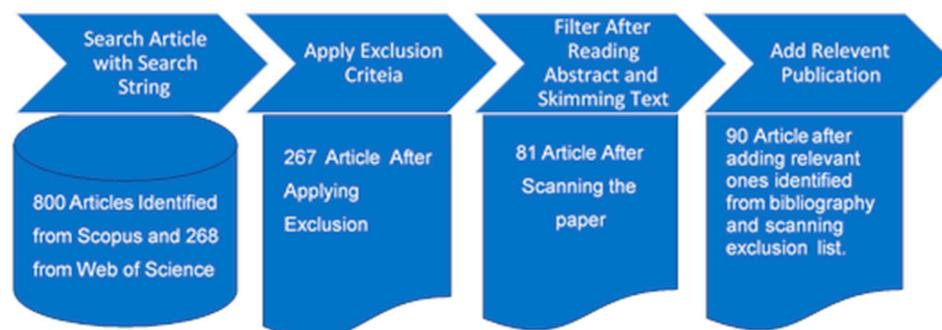


Figure 2. Research identification: research was systematically identified from the scientific database.

Figure 3 presents the five most important sources of the reviewed research. Four sources were peer-reviewed journals and the fifth was the IGARSS proceedings. The IGARSS proceedings also provided significant contributions related to methods and applications in RS. Figure 4 shows the most frequently used terms in the titles, keywords and abstracts in the reviewed studies. Font size corresponds to frequency. The cloud tag provides a comprehensive overview of the topics covered in these papers. The cloud tag shows that Sentinel, Landsat and SAR were the most frequently mentioned input data. The use of temporal data seems to have been popular in these studies. Regarding the crop types, soybeans, wheat, corn and rice were the most popular. The figure also shows that CNN was more popular than other DL methods for crop mapping and yield prediction. Furthermore, attention methods elicited a degree of research focus.

MODIS a preferred choice for regional-level yield-prediction studies. Sentinel-1, Landsat and Sentinel-2 were the most commonly used sensors in the crop-mapping studies. The common merit of all the aforementioned RS data is that they are freely available. These data are also available through the Google Earth engine, so data management and pre-processing are accessible. In fact, some crop-monitoring studies have used the Google Earth engine as a data-management and -processing platform [78–80]. The high temporal resolution of the MODIS and Sentinel sensors also allows the study of crop phenology at a finer level. Radar sensors, such as Sentinel-1 and Radarsat-2, can also work in cloudy weather. This could be the reason for the broader use of these sensors in the study of the phenological characteristics of crops, including rice. The number of UAV-based optical and multispectral sensors is also significant. Notably, hyperspectral sensors were less explored, despite their ability to provide better spectral range and precision, which are required for crop monitoring [81].

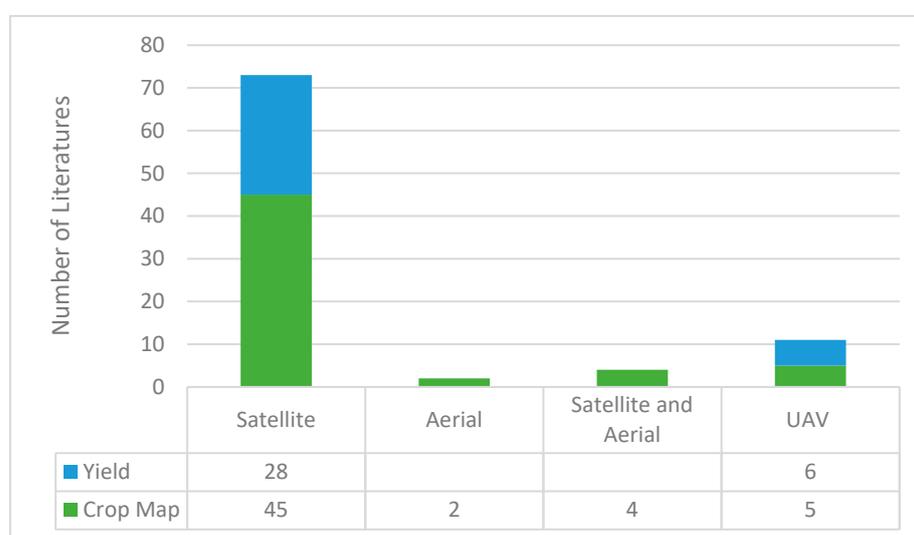


Figure 5. Common RS-based platforms used: most studies used satellite-sensor data.

Table 1. Major RS-based sensors used.

Sensor	Attributes	Number of Usages	
		Yield	Crop Map
MODIS	Multispectral sensor Spatial resolution: 250–1000 m Temporal resolution: up to a day High temporal and moderate spatial resolutions make it suitable for crop mapping and yield prediction at regional and global scales.	22	0
Sentinel-1	Radar sensor Spatial resolution: 10 m Temporal resolution: ~6 days Suitable for monitor crops in cloudy weather	0	14
Landsat	Multispectral sensor Spatial resolution: 30 m Temporal resolution: 16 days Suitable for regional/continental-scale crop mapping and regional-to-field-scale yield-prediction studies; can also be used in historical studies	3	10

Table 1. Cont.

Sensor	Attributes	Number of Usages	
		Yield	Crop Map
Sentinel-2	Multispectral sensor Spatial resolution: 10–60 m Temporal resolution: ~10 days Suitable for crop mapping in smaller fields and field-scale yield-prediction studies.	2	13
UAV(RGB)	Optical sensor Spatial resolution: up to centimetres Flexibility in data capture Suitable for crop mapping with precise field boundaries and field-level yield-prediction studies; limited spectral information for crop monitoring.	2	4
WV-3	Multispectral sensor Spatial resolution: 1.2 m Temporal resolution: ~1 day High cost of data, suitable for mapping crops with precise field-boundary and field-scale yield-prediction studies.	1	2
UAV (Multispectral)	Multispectral sensor Spatial resolution: up to centimetres Flexibility in data capture Crop mapping with precise field-boundary and field-level yield-prediction studies.	1	2

4.2. Input Features

As features input to the DL architecture, crop-mapping studies typically used optical data (RGB), multispectral data, radar data, thermal data, or a combination of these data. Some of the reviewed studies used the time-series enhanced vegetation index (EVI) [82,83] and normalised difference vegetation index [84] derived from the RS data as inputs in their crop-mapping models. Bhosle and Musande [85] reduced the dimensionality of the hyperspectral image using principal component analysis before feeding it to the CNN model. Traditionally, computer-vision CNN models were designed for three-channel red, green and blue (RGB) images. When transferring models developed for computer vision are used in RS applications, the data should be prepared in a three-channel RGB format, so additional multispectral bands cannot be used [86]. For instance, Li et al. [87] used this approach and only used the RGB channel of the multispectral Quickbird image to feed the LeNet model.

For crop-yield studies, environmental data, such as climate and soil data, are increasingly integrated with RS data [88–96]. Optical, multispectral, radar or thermal data, or their combination, were commonly used RS input features in the yield-prediction studies. Vegetation indices were used more often in the yield-prediction studies than in the mapping studies. Approximately 40% of the yield-prediction studies used the vegetation index as an input for their model. However, Nevavuori, Narra and Lipping [73] and Yang et al. [74] found that optical or/and multispectral images performed better than vegetation indices for yield prediction in the CNN model. The authors of [90] attempted to predict yield using satellite-derived climate and soil data without using spectral or VI information, but the model only achieved a coefficient of determination of 0.55.

In crop-yield-prediction studies at the administrative unit (county/district) scale, satellite imagery has a higher resolution than target data. A typical approach in such a scenario is to aggregate each target area's values (county/district) using mean or weighted means. You et al. [42] proposed a histogram method to reduce the dimensionality of RS

data whilst preserving key information for yield prediction. The approach was adopted in several subsequent studies [5,78,97].

Multi-temporal data, which are necessary to distinguish between crop types and estimate yield reliably, capture information on different crop-growth stages [98–100]. Interestingly, in 59% of the studies, the input features had multitemporal dimensions. Although many studies used multitemporal data, none of them intuitively modelled temporal dependencies. A few studies used ‘Explainable AI (XAI)’ techniques to understand the significance of different input features in model prediction. For instance, Wolanin et al. [101] visualised and interpreted features and yield drivers using regression-activation mapping to determine the impact of different drivers in the yield-prediction study. They found that downward shortwave radiation flux is the most influential meteorological variable in yield prediction. The most important variables that influence yields were also identified using attention mechanisms [102]. From an interpretability analysis, Xu et al. [103] identified that the increment in time-series length increased the classification confidence in an in-season-classification scenario.

4.3. Architecture

The deep-learning crop-mapping and yield-prediction applications were typically built using CNN, RNN, DNN, AEs, Transformer and hybrid architectures (Table 2). The CNN was the most popular architecture, with usage in approximately 58% of the reviewed studies. The CNN architecture is more suited to array data, such as RS data. Kuwata and Shibasaki [104] were among the pioneers in the field of crop-yield estimation using DL. They used a CNN network with a single and two fully connected layers (inner product layer) to extract features that affected the crop yield and estimated the yield index. They used satellite, climate and environmental data as inputs for the model. Similarly, for crop mapping, an early approach was applied by Nogueira et al. [105], who used a CNN for feature extraction from RS scenes and classified the scenes into coffee and non-coffee. Kussul et al. [106], in a pioneering work, proposed a CNN-based model to classify multitemporal, multisource RS data for crop mapping.

Table 2. DL architectures: different DL architectures used in the literature.

Architecture	Number of Times Used
CNN	52
RNN	20
MLP	6
Transformer	2
AE	4
Hybrid of CNN and RNN	14
Hybrid of CNN and ML	2
Bayesian NN	1
Domain adversarial NN	1

The use of CNN-based methods for semantic segmentation can be broadly categorised into patch-based approaches and fully convolutional networks (FCN) [107]. In patch-based approaches, the imagery is sliced into different patches. Each patch is fed to the CNN model, which assigns the central pixel or the whole patch to one target value. Some of the examples of patch-based CNN for crop mapping include the 2D CNN classification model by Kussul et al. [106], oil-palm-tree and citrus-tree detection studies [87,108,109], the classification of PolSAR data by Chen and Tao [110] and the study by Nogueira, Miranda and Santos [105], who classified the SPOT scene as coffee and non-coffee scene.

In yield prediction, Tri et al. [75] used LeNet and the inception version to extract features from the image patch. In the inception module, rather than selecting a filter size, multiple convolution filters of different sizes were selected (to learn features at a different scale) and all the outputs were concatenated [111]. To reduce the number of parameters,

one-by-one convolution was used along the depth. Nevavuori et al. [80] designed a CNN architecture similar to that presented by Krizhevsky et al. [46] to predict yield from UAV-based RGB images. The model predicted wheat and barley yield at the field scale with satisfactory accuracy. Jiang, Liu and Wu [83] used a CNN based on LeNet-5 to classify a time-series EVI curve. The authors fine-tuned the model, which was trained to detect handwritten numbers on the MNIST database with parameter-based transfer learning using the curve of the time-series EVI. One of the major disadvantages of the patch-based crop-classification approach is that small features may be smoothed and misclassified in the final classification.

In 2015, a novel CNN approach, called FCN, which used convolutional layers to process the input image and generate an output image of the same size, was introduced. It used the entire image as the input, extracted feature at different levels of abstraction and upsampled features to restore the input resolution in the next part of the network using techniques such as bilinear interpolation, a deconvolution layer and features from an earlier, more spatially accurate layer. The U-net [112] is a widely used FCN architecture with skip connections. It was also used quite often for crop mapping. Typical examples of FCN usage for crop mapping are those of Du et al. [113] and Saralioglu and Gungor [114]. Adrian, Sagan and Maimaitijiang [80] used 3D U-net to extract features from the temporal and spatial dimensions. Notably, FCN is mainly used with higher-resolution imagery. Mullissa et al. [115], La Rosa et al. [116], Chamorro Martinez et al. [117] and Wei et al. [118] implemented FCN in the classification of crops in synthetic-aperture radar (SAR) images. Chew et al. [119] used a VGG16 architecture and the publicly available ImageNet dataset to pretrain their model before feeding the UAV image.

The application of a 1D CNN along the temporal or spectral dimension is also used for crop mapping and yield prediction. Zhong, Hu and Zhou [82] demonstrated that 1D CNN can be effectively and efficiently used to classify multitemporal imagery. The authors compared the output of a 1D CNN with a RNN to classify summer crops using multitemporal Landsat EVI data. In the experiment, the 1D CNN exhibited a higher accuracy and FI score than CNN-, RF- and SVM-based methods. This experiment demonstrated the ability of 1D CNN to represent temporal features to classify crops. However, the limitation of this approach is that it fails to consider the spectral and spatial information of satellite imagery. Zhou et al. [120] used object-based image analysis, which is well-recognised as a classification method for high-resolution images for crop mapping. The authors used a segmentation algorithm to make segments from Sentinel-2 imagery and used a 1D CNN to classify the mean spectral vector of the segments. This approach used the mean of the segments to capture spatial information but still did not model the temporal relation, which would have improved the accuracy.

The RNN models were the second most widely used models, since they were applied in more than 22% of the reviewed studies. In fact, RNNs were preferred for yield prediction. More than 40% of the reviewed yield-prediction studies used RNNs. The RNN is the preferred method for agricultural monitoring when temporal dimensions are involved [47]. The LSTM, a type of RNN, was effective in learning temporal characteristics from multitemporal images for crop mapping [121,122] and yield estimation [5,79,88]. Rußwurm and Korner [123] proposed a LSTM model for temporal-feature extraction to classify multiclass crop types. Xu et al. [122] used the LSTM model to learn time-series spectral features for crop mapping. The authors also studied the spatial transfer of the model amongst six sites within the US corn-maturity zone and found that the approach can learn generalisable feature representation across regions. However, the RNN model typically cannot be used to learn spatial-feature representation.

Although MLPs are not particularly suitable for array data, such as RS and environmental data, a few studies also used MLPs. For instance, Maimaitijiang et al. [72] used a fully connected feedforward neural network and data fusion for yield prediction. They studied the impact of fusing data, such as spectral, crop-height, crop-density, temperature and texture data, at the input and intermediate stages for soybean yields. Chamorro Martinez

et al. [117] used a Bayesian neural network to predict the yield and find the uncertainty associated with the prediction. In the Bayesian neural network, the weight of the neural network is not fixed, but represented by a probability distribution. The use of AE, which is an unsupervised DL technique, featured in some of the crop-mapping applications [116,124,125]. In these studies, the AEs were used to learn the compressed and improved the representation of satellite data, which were then classified using other methods.

Hybrid modules that contain more than one architecture were used to learn spatial, spectral and temporal features for improved decision making. In the hybrid models, different architectures were used to learn features in various domains. These models either merged higher-level features obtained from two networks or used the output feature of one architecture as an input for another. To model the spatial context and temporal information from the multitemporal images jointly, the combinations of RNN and CNN were used [42,117]. Ghazaryan et al. [126] found that a hybrid model provided the highest accuracy out of 3D CNN, LSTM and a combination of CNN and LSTM whilst predicting the yield from multitemporal, multispectral and multisource images. Zhao et al. [127] used a LeNet-5 [128] model and a transfer-learning approach to classify red, green and infrared images in a first step and then improved the classification results using the DT model with phenological information in a second step. Although this hybrid approach improved the accuracy of the rice mapping, the approach could be challenging to implement in a larger scale because the decision rules are local and have to be determined from the field survey.

Attention mechanisms have also become popular in recent years in DL crop-mapping and yield-prediction models. An attention LSTM model with an attention mechanism was used to improve the generalisability of a yield-prediction model and identify the contribution of different variables to the yield [102]. The attention mechanism was also used to identify important features for crop mapping [103]. Wang et al. [129] claimed that crop mapping can be improved by integrating an attention mechanism and geographic information because it reduces the effects of geographic heterogeneity and prevents irrelevant information from being considered. Seydi, Amani and Ghorbanian [84] implemented spatial- and spectral-attention mechanisms to extract hidden features relevant to crop mapping. Self-attention-based transformers, which have been found to be effective for processing sequential data, were also applied for crop mapping. Rußwurm and Körner [130] concluded that transformers were more robust in handling the noise present in raw time-series RS data and were more effective for their classification. Reedha et al. [131] employed the Visual Transformer (ViT) model to classify aerial images captured by UAVs and achieved a similar degree of accuracy to a CNN. The authors also claimed that when the labelled training dataset is small, ViT models can be better than state-of-the-art CNN classification. To evaluate their performance, DL models are typically benchmarked against ML methods, such as SVM, RF and DT.

4.4. Frameworks

Deep-learning frameworks are software libraries with pre-built structures made for implementing DL models. The implementation of DL architectures has been made easier and more accessible in this way. The most popular DL frameworks are convolutional architectures for fast feature embedding (Caffe) [132], Theano [133], TensorFlow [134], PyTorch [135], CNTK [136] and MatConvNet [137]. These frameworks have a robust GPU backend that allows the training of networks with billions of parameters.

TensorFlow was the most widely used framework for crop mapping and yield prediction with DL (Figure 6). TensorFlow is written in Python and interfaces in R and JavaScript are also available. TensorFlow was developed by researchers who work on the Google Brain Team as a ML and DNN framework. It supports multiple GPUs and CPUs. Keras was also frequently used, with a total of 19 mentions. Keras is a high-level neural-network API written in Python and runs on top of TensorFlow or Theano. Of the Keras-based implementations, 11 used TensorFlow as the backend, one used Theano, and the remainder did not report a backend. Keras APIs are intuitive and straightforward, resulting in their

rapid growth. TensorFlow version 2 completely integrates Keras, thus providing a versatile library with a simple interface.

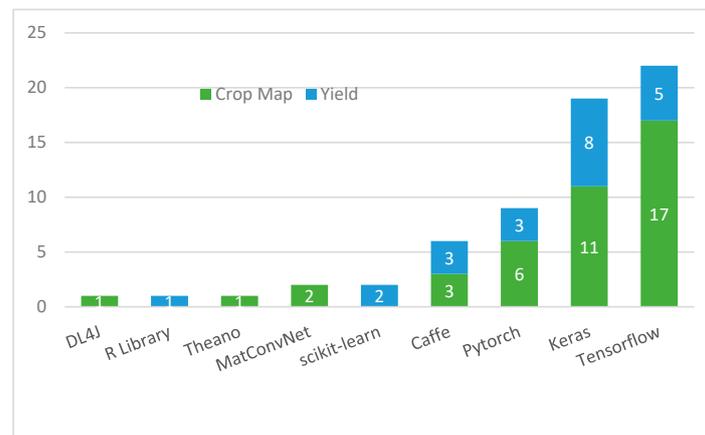


Figure 6. Various types of DL framework used in the literature.

Furthermore, Pytorch was used relatively frequently, with nine mentions. Facebook’s AI-research laboratory developed PyTorch. Providing flexibility, speed and deeper integration with Python, PyTorch has gained a user community in recent years. Caffe is written in C++ with a Python interface and is also popular in computer vision because it incorporates various CNN frameworks and datasets.

Deep neural networks are also built in Scikit-learn [138], a ML library. Mu et al. [139] used Scikit-learn to develop a DNN for yield prediction in their study, and Ma et al. [91] developed a Bayesian neural network. Scikit-learn does not support GPU implementation. Furthermore, DL4J [140], which is suitable for distributed computation, was also used in a study.

4.5. Crop Type

In the crop-yield-prediction studies, DL was most frequently applied to corn and soybeans (Table 3). Although most of the yield-prediction studies used a single crop, some also approached the prediction of the yield of more than one crop without distinguishing between the crops [73,76,96]. For crop mapping, most of the studies detected multiple crops. Rice was the most commonly mapped single crop. The wide use of rice as a staple food crop and the distinct phenological characteristics of rice fields reflected in the sensor data could be the primary reasons for its high rate of detection.

Table 3. Crop types used in the various studies. Most yield-prediction studies used either corn or soybean. Most of the crop-mapping studies mapped multiple crops. In the table, references to more than two crops are denoted as multiple.

Crop	Number of Studies	
	Yield Prediction	Crop Classification
Corn	9	-
Soybean	9	-
Corn and soybean	2	3
Wheat	5	-
Wheat and barley	1	-
Wheat and corn	1	-
Wheat and rapeseed	-	1
Rice	4	6
Multiple	2	41
Coffee	-	2
Oil-palm tree	-	3
Tobacco	-	1

4.6. Training Data

A DL model's accuracy and generalisation ability is determined by the quality and quantity of the training data [39]. Insufficient training data cause models to overfit and affect their prediction accuracy. Most of the training on crop mapping was performed by collecting the crop-type labels of the area of concern through field visits (Table 4). A field visit is a labour- and time-intensive process. After the field survey, the cropland-data layer (CDL) was the primary source of training for the crop-classification models. The CDL is a georeferenced, crop-specific land-cover map of the United States [141]. It is prepared using ground-truth data and moderate-resolution imagery. The CDL has a resolution of 30 m. It is published annually by the United States Department of Agriculture (USDA). It can be inferred that conducting such a study in other parts of the world is challenging, since such standard data are unavailable. Only three studies used government-supplied data other than CDL. Another method for training the data was the visual-image interpretation of higher-resolution images.

Table 4. Source of training data in crop-mapping studies.

Data Source (For Crop Mapping)	Number of Studies
Field survey	17
CDL	13
Visual interpretation (VisI)	6
Benchmark data	5
Government data (excluding CDL)	3
Crowdsourcing	2
CDL and field survey	1
Field survey and VisI	2
Data from Agricultural Company	2

Benchmark data, such as the UC Merced land-use dataset, the NWPU-RESISC45 dataset, the Campo Verde dataset and Breizhcrops, are also available for RS analysis and were used to test models in some of the crop-mapping studies. Crowdsourcing is another source of training data. Wang et al. [142] used crowdsourced crop-type data from farmers to train a network. Saralioglu and Gungor [114] created a web interface to collect training data for crop mapping. However, the challenge presented by the crowdsourcing method is the creation of an incentive or motivation for the contributor. Furthermore, the validation of these data is another challenge. Google Street View Images can also provide an efficient, cost-effective way to deliver ground referencing to train a DNN for crop-type mapping [143].

The county-level yield statistics provided by the USDA National Agricultural Statistics Service and the yield data collected from fields were the most commonly used data for training DL-yield prediction (Table 5). The USDA yield statistics are separated from other government data sources in the table to highlight the frequency of their use. The target data for yield prediction at the field scale are collected during harvesting, either by the harvester [73] or by the weighted grain from each yield plot [72]. Field data are essential for field-level predictions. The data prepared by local governance bodies may not provide confidence inaccuracy. The USDA county-level yields are available for the USA, while CISA data are available for Canada. However, such data are not available for other parts of the world.

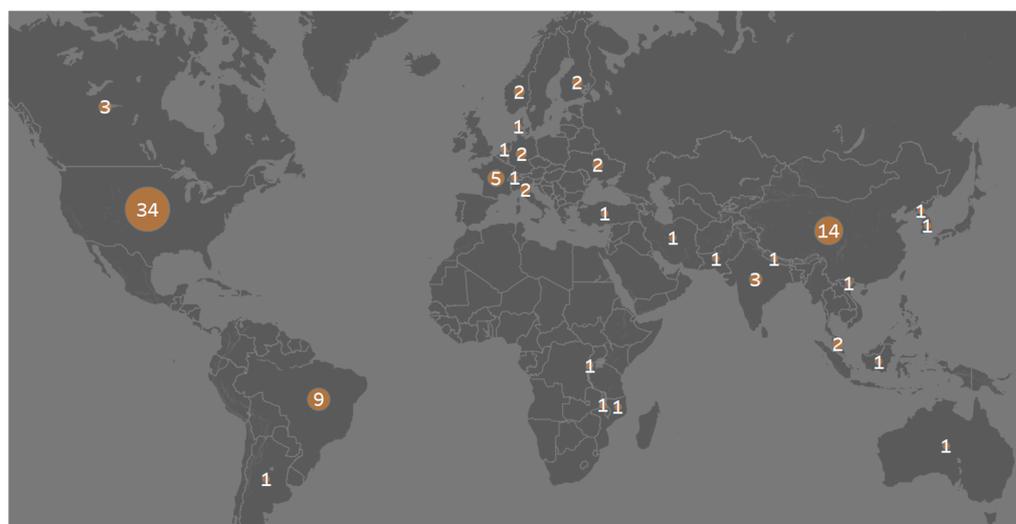
Table 5. Source of training data for crop-yield-prediction studies.

Data Source (For Crop-Yield Prediction)	Number of Studies
Field data	11
Government data (excluding USDA)	9
Government data (USDA)	11
Government data (USDA) and field data	1

Data-augmentation techniques, such as rotation and flips, were also used in the crop-mapping [113] and yield-prediction studies [75] to further enlarge the data and ensure that the model was independent of rotation and flips. Some of the crop-mapping studies used the domain-adaptation technique [144] and weakly supervised learning [145] to address the scarcity of training data. Wang et al. [145] concluded that CNN can perform better than other ML methods for crop mapping, even when the training data are scarce, if the training labels are used efficiently. The authors used two types of training data, a single geotagged point (pixel) and an image-level label, to train a U-net. This training approach gave satisfactory results that demonstrated the applicability of weak supervision. This model should be further validated for different areas and crop types and can be improved using multitemporal features and a DL model with temporal-learning capabilities. The scarcity of crop-type labels and historical yield data are major barriers to the development of the DL model for reliable and accurate crop mapping and yield prediction.

4.7. Location of Study and Area

Figure 7 shows the spatial distribution of the study sites of the reviewed studies. Some of the studies were conducted in more than one area, such as those in which the experiments conducted in one area, while a transfer-learning technique was used to perform estimates at another location [5]. In such cases, both are included on the list. Evidently, the map shows that the studies were concentrated only in some parts of the world. In total, 37% of all the studies were conducted in the USA and 15% were conducted in China. Only 3% of the studies were conducted in Africa, despite the fact that Africa holds 60% of the world's arable land [146]. Agriculture accounted for 55% of Australian land use and 11% of goods-and-services exports in 2019–2020 [147], but only one reviewed study used Australia as its study site.

**Figure 7.** Geographic distribution of RS- and DL-based crop mapping and yield-prediction-study sites.

One of the reasons for the skewed distribution of the sites could be the unavailability of target data. Another reason could be the locations of the research institutes. The areas of

the study sites varied from 65 hectares to as large as the Indian wheat belt and the entire USA. In all the UAV-based studies that reported the area, fewer than 200 hectares were covered. The reason for this could be the high cost of data capture with UAVs. Studies of larger areas provide confidence in model's applicability to diverse landscapes.

4.8. Scale of the Output

The crop-mapping and yield-prediction studies were implemented at different scales. The application of crop-monitoring studies depends on the output's scale. Although regional studies help to monitor crop production at a national and regional scale, within-field variability is necessary to inform field-specific decision making [147,148]. The scale of the output depends on the resolution of the input and target data. In most of the crop-mapping studies, each pixel or pixel group was assigned a crop class. The precision of the field boundary and generalisation depends on the spatial resolution of the RS data. We categorised the yield-prediction studies into two classes, namely, field-level and county- or district-level. Almost 70% of the yield-prediction studies were county-level and the remainder were field-level. The county/district-crop-yield statistics were typically used in the county-scale yield-prediction studies. In contrast, the field data collected from farmers and harvesters were used for field-scale studies. Precise yield data can be used to make predictions at the best possible scale. Notably, the platforms used and the scales of the studies were correlated. The field-level yield was estimated in all the UAV-based yield-prediction studies. The county-level yield-prediction studies were predominantly conducted in the USA. The reason for this could be the availability of USDA yield data.

4.9. Evaluation Metrics and Performance

The most commonly used evaluation metrics in the reviewed crop-mapping studies were the overall accuracy, kappa statistics, precision, recall and F1 score. The majority of the studies used more than one metric to evaluate performance. Approximately 87% of all the crop-mapping studies used overall accuracy to assess their model's performance. Overall accuracy is the most intuitive evaluation measure. It is the ratio of the correct predictions to the total number of predictions made. It is proportional to the area that is correctly mapped. Along with overall accuracy, kappa statistics were often computed in the crop-mapping studies. Precision refers to the ratio of correctly predicted positive observations to the total positive predictions made by the model. The recall is the number of correct positive results divided by the number of all the samples that should have been identified as positive. The F1 score is the harmonic mean of the precision and the recall [149]. The studies reported either the F1 score of individual classes or the mean (weighted or unweighted) of the F1 scores. Notably, the overall accuracy of a dataset can be misleading when the class distribution is uneven. The precision, recall and F1 score might be more useful in such studies.

The mean squared error (MSE), root MSE (RMSE), coefficient of determination (R^2), mean absolute error (MAE) and mean absolute percentage error (MAPE) were the commonly used metrics in assessing the reviewed yield-prediction models. The MSE is the average of the square of the difference between the original values and the predicted values. The MSE penalises larger errors because each value is squared. The RMSE is similar to the MSE, but takes the square root of the output. The problem with using MAE, MSE and RMSE is that the value depends on the units and the scale of the residuals. The mean absolute percentage error (MAPE) attempts to solve this issue. It transforms the errors into percentages; ideally, the MAPE should be as close to 0 as possible. The R^2 is the degree of agreement between the true value and the predicted value. It measures the proportion of variance in the dependent variables explained by the independent variable. The R^2 was the most frequently used evaluation metric in the yield-prediction studies, since it was used in 65% of the studies. Most of the yield-prediction studies also computed multiple metrics. After R^2 , RMSE and MAPE were the most commonly used. In some studies [75,77], the yield prediction was approached as a classification task. Each image segment was assigned

to a yield class and classified within a particular yield category. These studies used the overall accuracy and F1 score as evaluation metrics.

Comparing model performance is not easy when models use different evaluation metrics. We prepared box plots that show the distribution of the achieved R^2 and the overall accuracy percentage in the yield-prediction and crop-mapping studies, respectively (Figure 8). The data for these plots were obtained from the studies that reported the performances of the yield-prediction and crop-mapping models using R^2 and overall accuracy, respectively. The value of the best-performing model was selected to make the graph. Most of the crop-mapping studies reported very high classification accuracies, up to 99.7%, with a mean value of 90.0%. The R^2 of the yield prediction was distributed in the range of 0.5 to 0.96, with a mean value of 0.77.

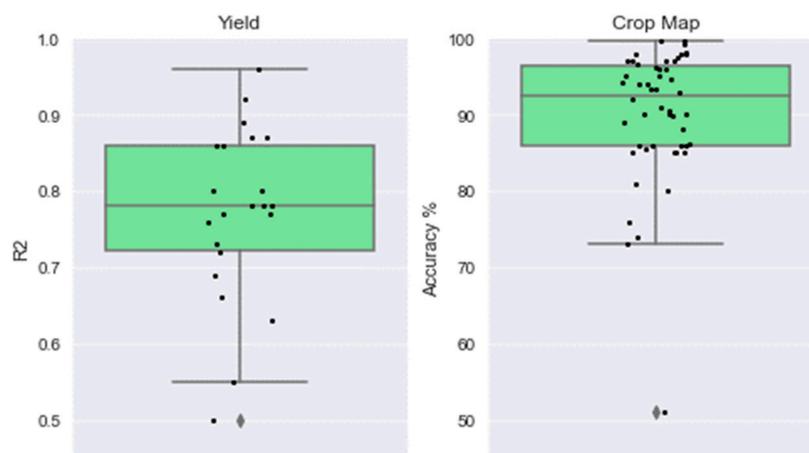


Figure 8. Box plot showing the distribution of reported R^2 and overall accuracy in yield-prediction and crop-mapping studies, respectively. The R^2 ranged from 0.5 to 0.96, with the median at 0.78. The overall accuracy ranged from 51% to 99.7%, with a median value of 92%.

5. Discussion

Deep learning and RS have emerged as promising techniques for crop mapping and yield prediction in recent years. A typical approach to DL- and RS-based crop mapping and yield prediction is summarised in Figure 9. The main platforms for capturing data are UAVs, satellites and aeroplanes. The input data can be raw spectral values from multispectral, optical, hyperspectral, radar or thermal sensors or derived features, such as the vegetation index, histograms of pixel intensities or even graphs of phenological characteristics. In yield-prediction models, RS data are often integrated with environmental data, such as climate and soil data. Although multitemporal data are popular, data from a single date can also be used. The models are built using CNN, RNN, MLP, Transformer or hybrid architectures. The models are usually implemented using a standard framework, such as TensorFlow, PyTorch or Caffe. Target labels (crop labels/yield values) are the most important components in the process and are often the limiting factors in the development of models. Trained models are evaluated using one of the evaluation metrics to assess their performance and fitness for use. In the section below, these aspects of the study are discussed.

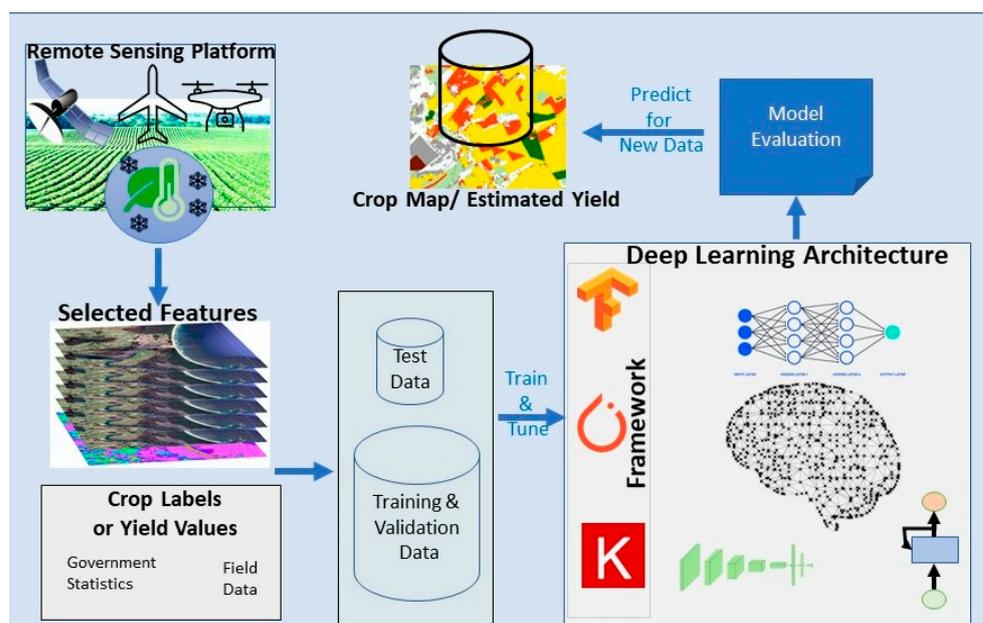


Figure 9. Typical approach for RS- and DL-based crop mapping or/and yield-estimation studies.

This review shows that satellite-based sensors are the most commonly used RS-data sources for crop mapping and yield prediction. This preference could be due to the ease of access to data, the availability of multiple spectral and spatial resolutions, the availability of historical data, or the global coverage of and fewer pre-processing steps in RS data. Additionally, platforms such as Google Earth Engine make the handling of large amounts of satellite data easier. Unmanned aerial vehicles are less often used as RS-data sources, although they provide flexibility in terms of the choice of sensors, spatial resolution and data-capture time. Unmanned aerial vehicles are the preferred platforms when greater precision is needed (e.g., in precision agriculture). With regard to sensors, MODIS is the most commonly used sensor for yield prediction, whereas Sentinel-1,2 and Landsat 2 are frequently used for crop mapping. Although hyperspectral sensors can provide better spectral ranges and precision for crop monitoring, their application is yet to be fully explored. Researchers and practitioners currently have access to RS data at varied resolutions (e.g., spatial, spectral and temporal), obtained from various platforms and sensors, thereby allowing the selection of the most suitable sensor based on the specific needs of the study. This study also summarised the attributes of commonly used RS data for crop mapping and yield prediction (Table 1).

The choice of input features can significantly affect how well a DL model represents underlying phenomena. The input features determine the model's architecture. Commonly used RS-derived input features in crop mapping and yield predictions are optical, multispectral radar or thermal data. The integration of environmental data and RS data for yield prediction is becoming increasingly popular. Environmental data provide additional valuable information for yield prediction beyond RS data [89,150,151]. Vegetation indices, which are compact summaries of vegetation crafted from spectral values, were also commonly used as inputs in the crop-mapping and yield-prediction studies. Some of the yield-prediction studies used the histogram method and summarisation techniques, such as the mean or weighted mean, to simplify input variables while retaining the most important information. Although these methods can allow the training of DL models with limited labelled data, they may feature the drawback of generalising input information. Furthermore, although the use of multitemporal and multispectral data for crop mapping and yield prediction is increasing, most existing approaches do not account for temporal, spatial or spectral dependencies concurrently.

The use of DL architectures for crop mapping and yield prediction has shown significant progress due to the development of new architectures in the DL field. Several DL methods, including MLP, CNN, RNN, Transformer and AEs, have been used for crop mapping and yield prediction. The CNNs are the most widely used architectures for crop mapping and yield prediction. They were specifically designed for processing data with grid-like topologies. Furthermore, 2D CNNs, 3D CNNs and FCNs can effectively capture and analyse the spatial features of satellite imagery. A 1D CNN can be used along temporal or spectral dimensions to capture respective dependencies in crop-mapping and yield-prediction tasks. Furthermore, RNNs can be used to model temporal dependencies in crop mapping and yield prediction, but they suffer from the vanishing- and exploding-gradient problem, are not particularly effective at capturing long-term dependencies and cannot be parallelised. Transformer is known for its ability to capture long-range dependencies in input data and its ability to train very large models efficiently. The reviewed studies also suggested that for crop mapping, Transformer is more robust in handling noise in raw time series and classification when the labelled training dataset is small. The MLPs are not particularly efficient when used to process high-dimensional-array data such as RS and environmental data and they have limited utility for crop mapping and yield prediction. Considering that different architectures have varying strengths and limitations, there is no single best architecture for crop mapping and yield prediction. Rather, the best option depends on the amount, nature and quality of the data, the complexity of their representation and the available computational resources, amongst other factors.

This review suggests that crop-mapping and yield-prediction research are currently skewed towards certain crops and locations. The results of research conducted on a certain area or crop type may not be generalisable to other regions or crops. Hence, studies must be extended to other widely consumed crops and other regions to ensure the usability of models in varied conditions. One of the reasons for the skewed study distribution could be the limited availability of training labels. Scarce historical yield data and crop-type labels are significant limiting factors in the development of DL models for crop mapping and yield prediction. The data prepared by government bodies are amongst the most widely used for training yield-prediction and crop-mapping models. Such data are not available in most of the world, especially in developing countries.

With regard to the performances of the models, the evaluation metrics used in different crop-mapping and yield-prediction studies are not uniform, which makes the comparison of the models challenging. The main evaluation metrics used for the crop-mapping models were the overall accuracy, kappa statistics and F1 score, whereas those in the crop-yield prediction studies were the R², RMSE and MAPE. The reviewed yield-prediction studies reported R² values that ranged from 0.5 to 0.96, with a median of 0.78, which suggests moderate-to-high accuracy. In the crop-mapping studies, generally, a higher accuracy was achieved, with a range of 51% to 99.7% and a median of 92%. However, a higher performance metric does not necessarily mean that the research problem relating to accuracy is solved. Accuracy is affected by various factors, including the choice of evaluation metrics and the training- and test-data selection. For instance, using an overall performance metric in imbalanced class-distribution scenarios can be misleading because the model will perform poorly in identifying minority classes, even though the OA is high. Similarly, a model's performance can be overestimated due to data leakage [152]. Moreover, a CNN model can overfit due to the involuntary overlapping of test data in the receptive fields of the training data, biasing the evaluation [153]. Another consideration could be the spatial distribution of the model performance. Although developed models may exhibit satisfactory performances overall, they may not perform well at specific locations. Moreover, the accuracy of predictions tends to improve as the season progresses, with early-season predictions generally less accurate than later predictions.

6. Future Work

In this review, several important aspects of crop mapping and yield prediction based on DL and RS, which provide a foundation for future research, were discussed. The following section highlights a few avenues for future research in this field.

The most prominent issue is the availability of target data related to different crops in various parts of the world at diverse times. Although DL can learn nonlinear patterns between input and output data, it requires a large amount of training data. However, the availability of target data for crop mapping and yield prediction is limited. In yield prediction, the data are even more scarce. Further research shall be carried out to make data available, learning from limited data [154,155], transfer-learning approaches [156], unsupervised learning [157] and the quantification of the uncertainties in predictions [158,159]. Alternative methods, such as crowdsourcing, using closed-range oblique images, including those obtained from Google Street View, geotagged social media images and interviews with farmers, should be further explored to collect training data. More benchmark datasets must be developed and used to produce a standardised measure of comparison and to allow researchers to evaluate their proposed architectures fairly and consistently. These standard data should be available for varied times, locations, resolutions and sensor types. Data availability can make the development of a global model that encompasses varying times and locations possible. The domain-adaptation technique [144,160] and weakly supervised learning [145] are useful for training DL models in scarce-target-data scenarios. These techniques need to be further validated and explored for multi-temporal scenarios and a large spatial extents with differences exists in environmental conditions and cropping practices. Curriculum learning, the multi-stage transfer-learning approach and few-shot learning also have the potential to improve our existing of crop-mapping and yield-prediction model. Unsupervised-learning techniques can reveal hidden patterns and structures within data without pre-existing labelled data [157]. Furthermore, unsupervised learning can be used to study the applicability of abundant unlabelled RS imagery to crop mapping and yield prediction. Furthermore, the modelling predictive uncertainty can be performed by combining DL models with Bayesian statistics, which can be beneficial when training data are scarce.

Another potential avenue for future research in this domain is to investigate how to develop a more efficient, effective and generalisable method whilst making the best use of available spatial, spectral and temporal richness. Although target data are scarce, satellite imagery is abundant, but it is not utilised properly. Most of the existing crop-mapping and yield-prediction applications fail to model temporal, spatial and spectral dependencies simultaneously. Additionally, DL models require high computing resources and training times, which may not always be accessible to or affordable for every institution. Thus, designing DL models that can optimally use available RS resources to improve generalisability and accuracy whilst considering the computational constraints of implementation is one promising area of research.

Deep-learning algorithms are often considered complex black-box models that pose challenges in terms of interpretability [161]. Interpreting the opaque decision-making processes of DL approaches in crop mapping and yield prediction is crucial to improve transparency, accountability and trust in predictions. Further, the interpretability of crop-yield-prediction models could provide valuable insights into the factors that contribute to crop-yield variability and could help to improve it. Explainable AI [162] should be further explored in crop mapping and yield prediction to address the challenges of model interpretability. A further avenue of research could be the integration of crop models with DL. Deep-learning models can learn complex patterns and relationships from large datasets, whereas crop models provide structured representations of the growth and development of crops. The combination of these two approaches can improve the accuracy, efficiency and interpretability of yield prediction.

7. Conclusions

In this systematic review, 90 papers related to DL- and RS-based crop-mapping and crop-yield-prediction studies were reviewed. The review provided an overview of the approaches used in these studies and presented important observations regarding the employed platforms, sensors, input features, architectures, frameworks, training data, spatial distributions of study sites, output scales, assessment criteria and performances. This review suggests that DL provides a promising solution for crop mapping and yield estimation at different scales. However, the mapping of crops and prediction of yields in new locations for new crops at the desired scale are still challenging. The issues include scarce target data, optimal model designs, generalisability across different domains and transparency. The resolution of these issues will better prepare us to realise the application at scale and thus address the problems of food security and decision making in the food industry and agro-environmental management.

Author Contributions: Conceptualisation, B.P.; methodology, A.J.; software, A.J.; validation, B.P. and S.G.; formal analysis, A.J.; investigation, A.J.; resources, B.P.; data curation, A.J.; writing—original draft preparation, A.J.; writing—review and editing, B.P. and S.G.; visualisation, B.P. and S.G.; supervision, B.P. and S.C.; project administration, B.P.; funding acquisition, B.P. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), University of Technology Sydney. The research was supported by an Australian Government Research Training Program Scholarship, granted to A.J.

Data Availability Statement: Not applicable.

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

References

1. FAO. The Future of Food and Agriculture—Trends and Challenges. *Annu. Rep.* 2017. Available online: <https://www.fao.org/global-perspectives-studies/resources/detail/en/c/458158/> (accessed on 28 February 2023).
2. Foley, J.A.; DeFries, R.; Asner, G.P.; Barford, C.; Bonan, G.; Carpenter, S.R.; Chapin, F.S.; Coe, M.T.; Daily, G.C.; Gibbs, H.K. Global Consequences of Land Use. *Science* **2005**, *309*, 570–574. [[CrossRef](#)] [[PubMed](#)]
3. Joseph, G. *Fundamentals of Remote Sensing*; Universities Press: Hyderabad, India, 2005.
4. Lillesand, T.; Kiefer, R.W.; Chipman, J. *Remote Sensing and Image Interpretation*; John Wiley & Sons: New York, NY, USA, 2015.
5. Wang, A.X.; Tran, C.; Desai, N.; Lobell, D.; Ermon, S. Deep transfer learning for crop yield prediction with remote sensing data. In Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, Menlo Park and San Jose, CA, USA, 20–22 June 2018.
6. FAO. Fao’s Director-General on How to Feed the World in 2050. *Popul. Dev. Rev.* **2009**, *35*, 837–839. [[CrossRef](#)]
7. Hoffman, L.A.; Etienne, X.L.; Irwin, S.H.; Colino, E.V.; Toasa, J.I. Forecast Performance of Wasde Price Projections for Us Corn. *Agric. Econ.* **2015**, *46*, 157–171. [[CrossRef](#)]
8. Sherrick, B.J.; Lanoue, C.A.; Woodard, J.; Schnitkey, G.D.; Paulson, N.D. Crop Yield Distributions: Fit, Efficiency, and Performance. *Agric. Financ. Rev.* **2014**, *74*, 348–363. [[CrossRef](#)]
9. Isengildina-Massa, O.; Irwin, S.H.; Good, D.L.; Gomez, J.K. The Impact of Situation and Outlook Information in Corn and Soybean Futures Markets: Evidence from Wasde Reports. *J. Agric. Appl. Econ.* **2008**, *40*, 89–103. [[CrossRef](#)]
10. Yang, C.; Everitt, J.H.; Du, Q.; Luo, B.; Chanussot, J. Using High-Resolution Airborne and Satellite Imagery to Assess Crop Growth and Yield Variability for Precision Agriculture. *Proc. IEEE* **2012**, *101*, 582–592. [[CrossRef](#)]
11. Basso, B.; Cammarano, D.; Carfagna, E. Review of crop yield forecasting methods and early warning systems. In Proceedings of the First Meeting of the Scientific Advisory Committee of the Global Strategy to Improve Agricultural and Rural Statistics, FAO Headquarters, Rome, Italy, 18 July 2013; p. 19.
12. Razmjooy, N.; Estrela, V.V. *Applications of Image Processing and Soft Computing Systems in Agriculture*; IGI Global: Hershey, PA, USA, 2019.
13. Moran, M.S.; Inoue, Y.; Barnes, E. Opportunities and Limitations for Image-Based Remote Sensing in Precision Crop Management. *Remote Sens. Environ.* **1997**, *61*, 319–346. [[CrossRef](#)]
14. Karthikeyan, L.; Chawla, I.; Mishra, A.K. A Review of Remote Sensing Applications in Agriculture for Food Security: Crop Growth and Yield, Irrigation, and Crop Losses. *J. Hydrol.* **2020**, *586*, 124905. [[CrossRef](#)]
15. Keating, B.A.; Carberry, P.S.; Hammer, G.L.; Probert, M.E.; Robertson, M.J.; Holzworth, D.; Huth, N.I.; Hargreaves, J.N.; Meinke, H.; Hochman, Z. An Overview of Apsim, a Model Designed for Farming Systems Simulation. *Eur. J. Agron.* **2003**, *18*, 267–288. [[CrossRef](#)]

16. Timsina, J.; Humphreys, E. Performance of Ceres-Rice and Ceres-Wheat Models in Rice–Wheat Systems: A Review. *Agric. Syst.* **2006**, *90*, 5–31. [[CrossRef](#)]
17. Brisson, N.; Gary, C.; Justes, E.; Roche, R.; Mary, B.; Ripoche, D.; Zimmer, D.; Sierra, J.; Bertuzzi, P.; Burger, P. An Overview of the Crop Model Stics. *Eur. J. Agron.* **2003**, *18*, 309–332. [[CrossRef](#)]
18. Launay, M.; Guerif, M. Assimilating Remote Sensing Data into a Crop Model to Improve Predictive Performance for Spatial Applications. *Agric. Ecosyst. Environ.* **2005**, *111*, 321–339. [[CrossRef](#)]
19. Jin, X.; Kumar, L.; Li, Z.; Feng, H.; Xu, X.; Yang, G.; Wang, J. A Review of Data Assimilation of Remote Sensing and Crop Models. *Eur. J. Agron.* **2018**, *92*, 141–152. [[CrossRef](#)]
20. Kang, Y.; Özdoğan, M. Field-Level Crop Yield Mapping with Landsat Using a Hierarchical Data Assimilation Approach. *Remote Sens. Environ.* **2019**, *228*, 144–163. [[CrossRef](#)]
21. Yao, F.; Tang, Y.; Wang, P.; Zhang, J. Estimation of Maize Yield by Using a Process-Based Model and Remote Sensing Data in the Northeast China Plain. *Phys. Chem. Earth Pt. A/B/C* **2015**, *87*, 142–152. [[CrossRef](#)]
22. Huang, J.; Gómez-Dans, J.L.; Huang, H.; Ma, H.; Wu, Q.; Lewis, P.E.; Liang, S.; Chen, Z.; Xue, J.-H.; Wu, Y. Assimilation of Remote Sensing into Crop Growth Models: Current Status and Perspectives. *Agric. For. Meteorol.* **2019**, *276*, 107609. [[CrossRef](#)]
23. Delécolle, R.; Maas, S.; Guerif, M.; Baret, F. Remote Sensing and Crop Production Models: Present Trends. *ISPRS J. Photogramm. Remote Sens.* **1992**, *47*, 145–161. [[CrossRef](#)]
24. Arıkan, M. Parcel Based Crop Mapping through Multi-Temporal Masking Classification of Landsat 7 Images in Karacabey, Turkey. In Proceedings of the ISPRS Symposium, Istanbul International Archives of Photogrammetry, Remote Sensing and Spatial Information Science, Istanbul, Turkey, 12–23 July 2004.
25. Beltran, C.M.; Belmonte, A.C. Irrigated Crop Area Estimation Using Landsat Tm Imagery in La Mancha, Spain. *Photogramm. Eng. Remote Sens.* **2001**, *67*, 1177–1184.
26. Castillejo-González, I.L.; López-Granados, F.; García-Ferrer, A.; Peña-Barragán, J.M.; Jurado-Expósito, M.; de la Orden, M.S.; González-Audicana, M. Object-and Pixel-Based Analysis for Mapping Crops and Their Agro-Environmental Associated Measures Using Quickbird Imagery. *Comput. Electron. Agric.* **2009**, *68*, 207–215. [[CrossRef](#)]
27. Huang, J.; Tian, L.; Liang, S.; Ma, H.; Becker-Reshef, I.; Huang, Y.; Su, W.; Zhang, X.; Zhu, D.; Wu, W. Improving Winter Wheat Yield Estimation by Assimilation of the Leaf Area Index from Landsat Tm and Modis Data into the Wofost Model. *Agric. For. Meteorol.* **2015**, *204*, 106–121. [[CrossRef](#)]
28. Lobell, D.B.; Thau, D.; Seifert, C.; Engle, E.; Little, B. A Scalable Satellite-Based Crop Yield Mapper. *Remote Sens. Environ.* **2015**, *164*, 324–333. [[CrossRef](#)]
29. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
30. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Routledge: New York, NY, USA, 2017.
31. Cortes, C.; Vapnik, V. Support Vector Machine. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
32. Ok, A.O.; Akar, O.; Gungor, O. Evaluation of Random Forest Method for Agricultural Crop Classification. *Eur. J. Remote Sens.* **2012**, *45*, 421–432. [[CrossRef](#)]
33. Peña-Barragán, J.M.; Ngugi, M.K.; Plant, R.E.; Six, J. Object-Based Crop Identification Using Multiple Vegetation Indices, Textural Features and Crop Phenology. *Remote Sens. Environ.* **2011**, *115*, 1301–1316. [[CrossRef](#)]
34. Feng, S.; Zhao, J.; Liu, T.; Zhang, H.; Zhang, Z.; Guo, X. Crop Type Identification and Mapping Using Machine Learning Algorithms and Sentinel-2 Time Series Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3295–3306. [[CrossRef](#)]
35. Jeong, J.H.; Resop, J.P.; Mueller, N.D.; Fleisher, D.H.; Yun, K.; Butler, E.E.; Timlin, D.J.; Shim, K.-M.; Gerber, J.S.; Reddy, V.R. Random Forests for Global and Regional Crop Yield Predictions. *PLoS ONE* **2016**, *11*, e0156571. [[CrossRef](#)]
36. Shekoofa, A.; Emam, Y.; Shekoufa, N.; Ebrahimi, M.; Ebrahimie, E. Determining the Most Important Physiological and Agronomic Traits Contributing to Maize Grain Yield through Machine Learning Algorithms: A New Avenue in Intelligent Agriculture. *PLoS ONE* **2014**, *9*, e97288. [[CrossRef](#)] [[PubMed](#)]
37. Gandhi, N.; Armstrong, L.J.; Petkar, O.; Tripathy, A.K. Rice Crop Yield Prediction in India Using Support Vector Machines. In Proceedings of the 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), Khon Kaen, Thailand, 13–15 July 2016; pp. 1–5.
38. Foody, G.M.; Mathur, A. Toward Intelligent Training of Supervised Image Classifications: Directing Training Data Acquisition for Svm Classification. *Remote Sens. Environ.* **2004**, *93*, 107–117. [[CrossRef](#)]
39. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
40. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J. Deep Learning in Environmental Remote Sensing: Achievements and Challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [[CrossRef](#)]
41. Liang, S. *Quantitative Remote Sensing of Land Surfaces*; John Wiley & Sons: Hoboken, NJ, USA, 2005.
42. You, J.; Li, X.; Low, M.; Lobell, D.; Ermon, S. Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. In Proceedings of the Thirty-First AAAI conference on artificial intelligence, San Francisco, CA, USA, 4–9 February 2017.
43. 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](#)]
44. Hinton, G.E.; Osindero, S.; Teh, Y.-W. A Fast Learning Algorithm for Deep Belief Nets. *Neural Comput.* **2006**, *18*, 1527–1554. [[CrossRef](#)]
45. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]

46. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep Learning in Remote Sensing Applications: A Meta-Analysis and Review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [[CrossRef](#)]
47. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep Learning in Agriculture: A Survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [[CrossRef](#)]
48. Weiss, M.; Jacob, F.; Duveiller, G. Remote Sensing for Agricultural Applications: A Meta-Review. *Remote Sens. Environ.* **2020**, *236*, 111402. [[CrossRef](#)]
49. Khan, M.J.; Khan, H.S.; Yousaf, A.; Khurshid, K.; Abbas, A. Modern Trends in Hyperspectral Image Analysis: A Review. *IEEE Access* **2018**, *6*, 14118–14129. [[CrossRef](#)]
50. Cheng, G.; Xie, X.; Han, J.; Guo, L.; Xia, G.-S. Remote Sensing Image Scene Classification Meets Deep Learning: Challenges, Methods, Benchmarks, and Opportunities. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 3735–3756. [[CrossRef](#)]
51. Van Klompenburg, T.; Kassahun, A.; Catal, C. Crop Yield Prediction Using Machine Learning: A Systematic Literature Review. *Comput. Electron. Agric.* **2020**, *177*, 105709. [[CrossRef](#)]
52. Oikonomidis, A.; Catal, C.; Kassahun, A. Deep Learning for Crop Yield Prediction: A Systematic Literature Review. *N. Z. J. Crop Hortic. Sci.* **2022**, 1–26. [[CrossRef](#)]
53. Murugantham, P.; Wibowo, S.; Grandhi, S.; Samrat, N.H.; Islam, N. A Systematic Literature Review on Crop Yield Prediction with Deep Learning and Remote Sensing. *Remote Sens.* **2022**, *14*, 1990. [[CrossRef](#)]
54. Mohri, M.; Rostamizadeh, A.; Talwalkar, A. *Foundations of Machine Learning*; MIT Press: Cambridge, MA, USA, 2018.
55. Hornik, K.; Stinchcombe, M.; White, H. Multilayer Feedforward Networks Are Universal Approximators. *Neural Netw.* **1989**, *2*, 359–366. [[CrossRef](#)]
56. Mesnil, G.; Dauphin, Y.; Yao, K.; Bengio, Y.; Deng, L.; Hakkani-Tur, D.; He, X.; Heck, L.; Tur, G.; Yu, D. Using Recurrent Neural Networks for Slot Filling in Spoken Language Understanding. *IEEE T. Audio Speech* **2014**, *23*, 530–539. [[CrossRef](#)]
57. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5999–6009.
58. Kramer, M.A. Nonlinear Principal Component Analysis Using Autoassociative Neural Networks. *AICHE J.* **1991**, *37*, 233–243. [[CrossRef](#)]
59. Gatys, L.A.; Ecker, A.S.; Bethge, M. A Neural Algorithm of Artistic Style. *arXiv* **2015**, arXiv:1508.06576. [[CrossRef](#)]
60. LeCun, Y.; Boser, B.; Denker, J.; Henderson, D.; Howard, R.; Hubbard, W.; Jackel, L. Handwritten Digit Recognition with a Back-Propagation Network. *Adv. Neural Inf. Process. Syst.* **1989**, *2*, 396–404.
61. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [[CrossRef](#)]
62. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
63. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
64. Novikov, A.A.; Lenis, D.; Major, D.; Hladůvka, J.; Wimmer, M.; Bühler, K. Fully Convolutional Architectures for Multiclass Segmentation in Chest Radiographs. *IEEE Trans. Med. Imaging* **2018**, *37*, 1865–1876. [[CrossRef](#)]
65. Graves, A.; Mohamed, A.-R.; Hinton, G. Speech Recognition with Deep Recurrent Neural Networks. In Proceedings of the 2013 IEEE international conference on acoustics, speech and signal processing, Vancouver, Canada, 26–31 May 2013; pp. 6645–6649.
66. Selvin, S.; Vinayakumar, R.; Gopalakrishnan, E.; Menon, V.K.; Soman, K. Stock Price Prediction Using Lstm, Rnn and Cnn-Sliding Window Model. In Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, 13–16 September 2017; pp. 1643–1647.
67. Bengio, Y.; Simard, P.; Frasconi, P. Learning Long-Term Dependencies with Gradient Descent Is Difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166. [[CrossRef](#)]
68. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
69. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations Using Rnn Encoder-Decoder for Statistical Machine Translation. *arXiv* **2014**, arXiv:1406.1078.
70. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
71. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.
72. Maimaitijiang, M.; Sagan, V.; Sidike, P.; Hartling, S.; Esposito, F.; Fritsch, F.B. Soybean Yield Prediction from Uav Using Multimodal Data Fusion and Deep Learning. *Remote Sens. Environ.* **2020**, *237*, 111599. [[CrossRef](#)]
73. Nevavuori, P.; Narra, N.; Lipping, T. Crop Yield Prediction with Deep Convolutional Neural Networks. *Comput. Electron. Agric.* **2019**, *163*, 104859. [[CrossRef](#)]
74. 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](#)]
75. Tri, N.C.; Duong, H.N.; Van Hoai, T.; Van Hoa, T.; Nguyen, V.H.; Toan, N.T.; Snasel, V. A Novel Approach Based on Deep Learning Techniques and Uavs to Yield Assessment of Paddy Fields. In Proceedings of the 2017 9th International Conference on Knowledge and Systems Engineering, Hue, Vietnam, 19–21 October 2017; pp. 257–262.

76. Nevavuori, P.; Narra, N.; Linna, P.; Lipping, T. Crop Yield Prediction Using Multitemporal Uav Data and Spatio-Temporal Deep Learning Models. *Remote Sens.* **2020**, *12*, 4000. [[CrossRef](#)]
77. Yang, W.; Nigon, T.; Hao, Z.; Dias Paiao, G.; Fernández, F.G.; Mulla, D.; Yang, C. Estimation of Corn Yield Based on Hyperspectral Imagery and Convolutional Neural Network. *Comput. Electron. Agric.* **2021**, *184*. [[CrossRef](#)]
78. Sun, J.; Di, L.; Sun, Z.; Shen, Y.; Lai, Z. County-Level Soybean Yield Prediction Using Deep Cnn-Lstm Model. *Sensors* **2019**, *19*, 4363. [[CrossRef](#)]
79. Schwalbert, R.A.; Amado, T.; Corassa, G.; Pott, L.P.; Prasad, P.V.V.; Ciampitti, I.A. Satellite-Based Soybean Yield Forecast: Integrating Machine Learning and Weather Data for Improving Crop Yield Prediction in Southern Brazil. *Agric. For. Meteorol.* **2020**, *284*, 107886. [[CrossRef](#)]
80. Adrian, J.; Sagan, V.; Maimaitijiang, M. Sentinel Sar-Optical Fusion for Crop Type Mapping Using Deep Learning and Google Earth Engine. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 215–235. [[CrossRef](#)]
81. Adão, T.; Hruška, J.; Pádua, L.; Bessa, J.; Peres, E.; Morais, R.; Sousa, J.J. Hyperspectral Imaging: A Review on Uav-Based Sensors, Data Processing and Applications for Agriculture and Forestry. *Remote Sens.* **2017**, *9*, 1110. [[CrossRef](#)]
82. Zhong, L.; Hu, L.; Zhou, H. Deep Learning Based Multi-Temporal Crop Classification. *Remote Sens. Environ.* **2019**, *221*, 430–443. [[CrossRef](#)]
83. Jiang, T.; Liu, X.; Wu, L. Method for Mapping Rice Fields in Complex Landscape Areas Based on Pre-Trained Convolutional Neural Network from HJ-1 a/B Data. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 418. [[CrossRef](#)]
84. Seydi, S.T.; Amani, M.; Ghorbanian, A. A Dual Attention Convolutional Neural Network for Crop Classification Using Time-Series Sentinel-2 Imagery. *Remote Sens.* **2022**, *14*, 498. [[CrossRef](#)]
85. Bhosle, K.; Musande, V. Evaluation of Deep Learning Cnn Model for Land Use Land Cover Classification and Crop Identification Using Hyperspectral Remote Sensing Images. *J. Ind. Soc. Remote Sens.* **2019**, *47*, 1949–1958. [[CrossRef](#)]
86. Hoerer, T.; Kuenzer, C. Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends. *Remote Sens.* **2020**, *12*, 1667. [[CrossRef](#)]
87. Li, W.; Fu, H.; Yu, L.; Cracknell, A. Deep Learning Based Oil Palm Tree Detection and Counting for High-Resolution Remote Sensing Images. *Remote Sens.* **2017**, *9*, 22. [[CrossRef](#)]
88. Jiang, H.; Hu, H.; Zhong, R.; Xu, J.; Xu, J.; Huang, J.; Wang, S.; Ying, Y.; Lin, T. A Deep Learning Approach to Conflating Heterogeneous Geospatial Data for Corn Yield Estimation: A Case Study of the Us Corn Belt at the County Level. *Glob. Chang. Biol.* **2020**, *26*, 1754–1766. [[CrossRef](#)]
89. Zhang, L.; Zhang, Z.; Luo, Y.; Cao, J.; Tao, F. Combining Optical, Fluorescence, Thermal Satellite, and Environmental Data to Predict County-Level Maize Yield in China Using Machine Learning Approaches. *Remote Sens.* **2020**, *12*, 21. [[CrossRef](#)]
90. Cunha, R.L.F.; Silva, B.; Netto, M.A.S. A Scalable Machine Learning System for Pre-Season Agriculture Yield Forecast. In Proceedings of the IEEE 14th International Conference on eScience, Amsterdam, The Netherlands, 9 October 2018; pp. 423–430.
91. Ma, Y.; Zhang, Z.; Kang, Y.; Özdoğan, M. Corn Yield Prediction and Uncertainty Analysis Based on Remotely Sensed Variables Using a Bayesian Neural Network Approach. *Remote Sens. Environ.* **2021**, *259*, 112408. [[CrossRef](#)]
92. Fernandez-Beltran, R.; Baidar, T.; Kang, J.; Pla, F. Rice-Yield Prediction with Multi-Temporal Sentinel-2 Data and 3d Cnn: A Case Study in Nepal. *Remote Sens.* **2021**, *13*, 1391. [[CrossRef](#)]
93. Wang, X.; Huang, J.; Feng, Q.; Yin, D. Winter Wheat Yield Prediction at County Level and Uncertainty Analysis in Main Wheat-Producing Regions of China with Deep Learning Approaches. *Remote Sens.* **2020**, *12*, 1744. [[CrossRef](#)]
94. Jeong, S.; Ko, J.; Yeom, J.M. Predicting Rice Yield at Pixel Scale through Synthetic Use of Crop and Deep Learning Models with Satellite Data in South and North Korea. *Sci. Total Environ.* **2022**, *802*, 149726. [[CrossRef](#)]
95. Zhang, L.; Zhang, Z.; Luo, Y.; Cao, J.; Xie, R.; Li, S. Integrating Satellite-Derived Climatic and Vegetation Indices to Predict Smallholder Maize Yield Using Deep Learning. *Agric. For. Meteorol.* **2021**, *311*, 108666. [[CrossRef](#)]
96. Engen, M.; Sando, E.; Sjolander, B.L.O.; Arenberg, S.; Gupta, R.; Goodwin, M. Farm-Scale Crop Yield Prediction from Multi-Temporal Data Using Deep Hybrid Neural Networks. *Agronomy* **2021**, *11*, 2576. [[CrossRef](#)]
97. Khaki, S.; Pham, H.; Wang, L. Simultaneous Corn and Soybean Yield Prediction from Remote Sensing Data Using Deep Transfer Learning. *Sci. Rep.* **2021**, *11*, 11132. [[CrossRef](#)] [[PubMed](#)]
98. Reed, B.C.; Brown, J.F.; VanderZee, D.; Loveland, T.R.; Merchant, J.W.; Ohlen, D.O. Measuring Phenological Variability from Satellite Imagery. *J. Veg. Sci.* **1994**, *5*, 703–714. [[CrossRef](#)]
99. Xiao, X.; Boles, S.; Liu, J.; Zhuang, D.; Frolicking, S.; Li, C.; Salas, W.; Moore III, B. Mapping Paddy Rice Agriculture in Southern China Using Multi-Temporal Modis Images. *Remote Sens. Environ.* **2005**, *95*, 480–492. [[CrossRef](#)]
100. Vega, F.A.; Ramirez, F.C.; Saiz, M.P.; Rosua, F.O. Multi-Temporal Imaging Using an Unmanned Aerial Vehicle for Monitoring a Sunflower Crop. *Biosys. Eng.* **2015**, *132*, 19–27. [[CrossRef](#)]
101. Wolanin, A.; Mateo-García, G.; Camps-Valls, G.; Gómez-Chova, L.; Meroni, M.; Duveiller, G.; Liangzhi, Y.; Guanter, L. Estimating and Understanding Crop Yields with Explainable Deep Learning in the Indian Wheat Belt. *Environ. Res. Lett.* **2020**, *15*, 024019. [[CrossRef](#)]
102. Tian, H.; Wang, P.; Tansey, K.; Han, D.; Zhang, J.; Zhang, S.; Li, H. A Deep Learning Framework under Attention Mechanism for Wheat Yield Estimation Using Remotely Sensed Indices in the Guanzhong Plain, Pr China. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102375. [[CrossRef](#)]

103. Xu, J.; Yang, J.; Xiong, X.; Li, H.; Huang, J.; Ting, K.; Ying, Y.; Lin, T. Towards Interpreting Multi-Temporal Deep Learning Models in Crop Mapping. *Remote Sens. Environ.* **2021**, *264*, 112599. [[CrossRef](#)]
104. Kuwata, K.; Shibasaki, R. Estimating Crop Yields with Deep Learning and Remotely Sensed Data. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 858–861.
105. Nogueira, K.; Miranda, W.O.; Santos, J.A.D. Improving Spatial Feature Representation from Aerial Scenes by Using Convolutional Networks. In Proceedings of the Brazilian Symposium of Computer Graphic and Image Processing, Salvador, Brazil, 26–29 August 2015; pp. 289–296.
106. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782. [[CrossRef](#)]
107. Hoerer, T.; Bachofer, F.; Kuenzer, C. Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review—Part II: Applications. *Remote Sens.* **2020**, *12*, 3053. [[CrossRef](#)]
108. Csillik, O.; Cherbini, J.; Johnson, R.; Lyons, A.; Kelly, M. Identification of Citrus Trees from Unmanned Aerial Vehicle Imagery Using Convolutional Neural Networks. *Drones* **2018**, *2*, 39. [[CrossRef](#)]
109. Li, W.; Dong, R.; Fu, H.; Yu, L. Large-Scale Oil Palm Tree Detection from High-Resolution Satellite Images Using Two-Stage Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 11. [[CrossRef](#)]
110. Chen, S.-W.; Tao, C.-S. PolSAR Image Classification Using Polarimetric-Feature-Driven Deep Convolutional Neural Network. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 627–631. [[CrossRef](#)]
111. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
112. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
113. Du, Z.; Yang, J.; Ou, C.; Zhang, T. Smallholder Crop Area Mapped with a Semantic Segmentation Deep Learning Method. *Remote Sens.* **2019**, *11*, 888. [[CrossRef](#)]
114. Saralioglu, E.; Gungor, O. Semantic Segmentation of Land Cover from High Resolution Multispectral Satellite Images by Spectral-Spatial Convolutional Neural Network. *Geocarto Int.* **2020**, *37*, 657–677. [[CrossRef](#)]
115. Mullissa, A.G.; Persello, C.; Tolpekin, V. Fully Convolutional Networks for Multi-Temporal Sar Image Classification. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Valencia, Spain, 22–27 July 2018; pp. 6635–6638.
116. La Rosa, L.E.C.; Feitosa, R.Q.; Happ, P.N.; Sanches, I.D.; da Costa, G.A.O.P. Combining Deep Learning and Prior Knowledge for Crop Mapping in Tropical Regions from Multitemporal Sar Image Sequences. *Remote Sens.* **2019**, *11*, 2029. [[CrossRef](#)]
117. Chamorro Martinez, J.A.; Cué La Rosa, L.E.; Feitosa, R.Q.; Sanches, I.D.; Happ, P.N. Fully Convolutional Recurrent Networks for Multidate Crop Recognition from Multitemporal Image Sequences. *ISPRS J. Photogramm. Remote Sens.* **2021**, *171*, 188–201. [[CrossRef](#)]
118. Wei, P.; Chai, D.; Lin, T.; Tang, C.; Du, M.; Huang, J. Large-Scale Rice Mapping under Different Years Based on Time-Series Sentinel-1 Images Using Deep Semantic Segmentation Model. *ISPRS J. Photogramm. Remote Sens.* **2021**, *174*, 198–214. [[CrossRef](#)]
119. Chew, R.; Rineer, J.; Beach, R.; O’neil, M.; Ujeneza, N.; Lapidus, D.; Miano, T.; Hegarty-Craver, M.; Polly, J.; Temple, D.S. Deep Neural Networks and Transfer Learning for Food Crop Identification in Uav Images. *Drones* **2020**, *4*, 7. [[CrossRef](#)]
120. Zhou, Z.; Li, S.; Shao, Y. Object-Oriented Crops Classification for Remote Sensing Images Based on Convolutional Neural Network. In Proceedings of the SPIE—The International Society for Optical Engineering, Berlin, Germany, 10–12 September 2018.
121. Sun, Z.; Di, L.; Fang, H. Using Long Short-Term Memory Recurrent Neural Network in Land Cover Classification on Landsat and Cropland Data Layer Time Series. *Int. J. Remote Sens.* **2019**, *40*, 593–614. [[CrossRef](#)]
122. Xu, J.; Zhu, Y.; Zhong, R.; Lin, Z.; Xu, J.; Jiang, H.; Huang, J.; Li, H.; Lin, T. Deepcropmapping: A Multi-Temporal Deep Learning Approach with Improved Spatial Generalizability for Dynamic Corn and Soybean Mapping. *Remote Sens. Environ.* **2020**, *247*, 111946. [[CrossRef](#)]
123. Rußwurm, M.; Korner, M. Temporal Vegetation Modelling Using Long Short-Term Memory Networks for Crop Identification from Medium-Resolution Multi-Spectral Satellite Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 11–19.
124. Castro, J.B.; Feitosa, R.Q.; La Rosa, L.C.; Diaz, P.A.; Sanches, I. A Comparative Analysis of Deep Learning Techniques for Sub-Tropical Crop Types Recognition from Multitemporal Optical/Sar Image Sequences. In Proceedings of the 2017 30th Sibgrapi Conference on Graphics, Patterns and Images (SIBGRAPI), Niteroi, Brazil, 17–20 October 2017; pp. 382–389.
125. Lavreniuk, M.; Kussul, N.; Novikov, A. Deep Learning Crop Classification Approach Based on Sparse Coding of Time Series of Satellite Data. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Valencia, Spain, 22–27 July 2018; pp. 4812–4815.
126. Ghazaryan, G.; Skakun, S.; König, S.; Rezaei, E.E.; Siebert, S.; Dubovyk, O. Crop Yield Estimation Using Multi-Source Satellite Image Series and Deep Learning. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October; pp. 5163–5166.

127. Zhao, S.; Liu, X.; Ding, C.; Liu, S.; Wu, C.; Wu, L. Mapping Rice Paddies in Complex Landscapes with Convolutional Neural Networks and Phenological Metrics. *GISci. Remote Sens.* **2020**, *57*, 37–48. [[CrossRef](#)]
128. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
129. Wang, Y.; Zhang, Z.; Feng, L.; Ma, Y.; Du, Q. A New Attention-Based Cnn Approach for Crop Mapping Using Time Series Sentinel-2 Images. *Comput. Electron. Agric.* **2021**, *184*, 106090. [[CrossRef](#)]
130. Rußwurm, M.; Körner, M. Self-Attention for Raw Optical Satellite Time Series Classification. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 421–435. [[CrossRef](#)]
131. Reedha, R.; Dericquebourg, E.; Canals, R.; Hafiane, A. Transformer Neural Network for Weed and Crop Classification of High Resolution Uav Images. *Remote Sens.* **2022**, *14*, 592. [[CrossRef](#)]
132. Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; Darrell, T. Caffe: Convolutional Architecture for Fast Feature Embedding. In Proceedings of the 22nd ACM international conference on Multimedia, Orlando, FL, USA, 3–7 November 2014; pp. 675–678.
133. Bergstra, J.; Breuleux, O.; Bastien, F.; Lamblin, P.; Pascanu, R.; Desjardins, G.; Turian, J.; Warde-Farley, D.; Bengio, Y. Theano: A Cpu and Gpu Math Expression Compiler. In Proceedings of the Python for scientific computing conference (SciPy), Austin TX, USA, 28 June–3 July 2010; pp. 1–7.
134. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M. Tensorflow: A System for Large-Scale Machine Learning. In Proceedings of the 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
135. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L. Pytorch: An Imperative Style, High-Performance Deep Learning Library. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 8026–8037.
136. Seide, F.; Agarwal, A. Cntk: Microsoft’s Open-Source Deep-Learning Toolkit. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 13 August 2016; p. 2135.
137. Vedaldi, A.; Lenc, K. Matconvnet: Convolutional Neural Networks for Matlab. In Proceedings of the 23rd ACM international conference on Multimedia, Brisbane, Australia, 26–30 October 2015; pp. 689–692.
138. Scikit-Learn. Available online: <https://scikit-learn.org/stable/index.html> (accessed on 1 March 2022).
139. Mu, H.; Zhou, L.; Dang, X.; Yuan, B. Winter Wheat Yield Estimation from Multitemporal Remote Sensing Images Based on Convolutional Neural Networks. In Proceedings of the 2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images, MultiTemp 2019, Shanghai, China, 5–7 August 2019.
140. Eclipse. DeepLearning4j. Available online: <https://github.com/deeplearning4j> (accessed on 1 February 2023).
141. NASS. Usda National Agricultural Statistics Service Cropland Data Layer. Available online: <https://nassgeodata.gmu.edu/CropScape> (accessed on 1 November 2022).
142. Wang, S.; Di Tommaso, S.; Faulkner, J.; Friedel, T.; Kennepohl, A.; Strey, R.; Lobell, D.B. Mapping Crop Types in Southeast India with Smartphone Crowdsourcing and Deep Learning. *Remote Sens.* **2020**, *12*, 2957. [[CrossRef](#)]
143. Yan, Y.; Ryu, Y. Exploring Google Street View with Deep Learning for Crop Type Mapping. *ISPRS J. Photogramm. Remote Sens.* **2021**, *171*, 278–296. [[CrossRef](#)]
144. Elshamli, A.; Taylor, G.W.; Berg, A.; Areibi, S. Domain Adaptation Using Representation Learning for the Classification of Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4198–4209. [[CrossRef](#)]
145. Wang, S.; Chen, W.; Xie, S.M.; Azzari, G.; Lobell, D.B. Weakly Supervised Deep Learning for Segmentation of Remote Sensing Imagery. *Remote Sens.* **2020**, *12*, 207. [[CrossRef](#)]
146. Bruzzone, B. Agriculture in Africa 2021: Focus Report. Available online: <https://oxfordbusinessgroup.com/blog/bernardo-bruzzone/focus-reports/agriculture-africa-2021-focus-report> (accessed on 15 March 2022).
147. Becker-Reshef, I.; Justice, C.; Sullivan, M.; Vermote, E.; Tucker, C.; Anyamba, A.; Small, J.; Pak, E.; Masuoka, E.; Schmaltz, J. Monitoring Global Croplands with Coarse Resolution Earth Observations: The Global Agriculture Monitoring (Glam) Project. *Remote Sens.* **2010**, *2*, 1589–1609. [[CrossRef](#)]
148. Lobell, D.B. The Use of Satellite Data for Crop Yield Gap Analysis. *Field Crops Res.* **2013**, *143*, 56–64. [[CrossRef](#)]
149. Scikit-Learn. Sklearn.Metrics.F1_Score. Available online: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html (accessed on 1 March 2022).
150. Wang, Y.; Zhang, Z.; Feng, L.; Du, Q.; Runge, T. Combining Multi-Source Data and Machine Learning Approaches to Predict Winter Wheat Yield in the Conterminous United States. *Remote Sens.* **2020**, *12*, 1232. [[CrossRef](#)]
151. Cai, Y.; Guan, K.; Lobell, D.; Potgieter, A.B.; Wang, S.; Peng, J.; Xu, T.; Asseng, S.; Zhang, Y.; You, L. Integrating Satellite and Climate Data to Predict Wheat Yield in Australia Using Machine Learning Approaches. *Agric. For. Meteorol.* **2019**, *274*, 144–159. [[CrossRef](#)]
152. Kaufman, S.; Rosset, S.; Perlich, C.; Stitelman, O. Leakage in Data Mining: Formulation, Detection, and Avoidance. *ACM Trans. Knowl. Discov. Data (TKDD)* **2012**, *6*, 1–21. [[CrossRef](#)]
153. Audebert, N.; Le Saux, B.; Lefèvre, S. Deep Learning for Classification of Hyperspectral Data: A Comparative Review. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 159–173. [[CrossRef](#)]
154. Zhou, Z.-H. A Brief Introduction to Weakly Supervised Learning. *Natl Sci. Rev.* **2018**, *5*, 44–53. [[CrossRef](#)]
155. Shorten, C.; Khoshgoftaar, T.M. A Survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [[CrossRef](#)]

156. Torrey, L.; Shavlik, J. Transfer Learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*; IGI Global: Hershey, PA, USA, 2010; pp. 242–264.
157. Hinton, G.; Sejnowski, T.J. *Unsupervised Learning: Foundations of Neural Computation*; MIT Press: Cambridge, MA, USA, 1999.
158. Abdar, M.; Pourpanah, F.; Hussain, S.; Rezazadegan, D.; Liu, L.; Ghavamzadeh, M.; Fieguth, P.; Cao, X.; Khosravi, A.; Acharya, U.R. A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges. *Inform. Fusion* **2021**, *76*, 243–297. [[CrossRef](#)]
159. Kendall, A.; Gal, Y. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5580–5590.
160. Ma, Y.; Zhang, Z.; Yang, H.L.; Yang, Z. An Adaptive Adversarial Domain Adaptation Approach for Corn Yield Prediction. *Comput. Electron. Agric.* **2021**, *187*, 106314. [[CrossRef](#)]
161. Castelvechi, D. Can We Open the Black Box of Ai? *Nat. News* **2016**, *538*, 20. [[CrossRef](#)] [[PubMed](#)]
162. Gunning, D.; Stefik, M.; Choi, J.; Miller, T.; Stumpf, S.; Yang, G.-Z. Xai—Explainable Artificial Intelligence. *Sci. Robot.* **2019**, *4*, eaay7120. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.