

Review

Deep Learning Models for the Classification of Crops in Aerial Imagery: A Review

Igor Teixeira ^{1,*} , Raul Morais ^{1,2} , Joaquim J. Sousa ^{1,3}  and António Cunha ^{1,3} 

¹ Engineering Department, School of Science and Technology, UTAD—University of Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal; rmorais@utad.pt (R.M.); jjsousa@utad.pt (J.J.S.); acunha@utad.pt (A.C.)

² Centre for the Research and Technology of Agro-Environmental and Biological Sciences, University of Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal

³ Institute for Systems and Computer Engineering, Technology and Science (INESC-TEC), 4200-465 Porto, Portugal

* Correspondence: igor_teixeira@hotmail.com

Abstract: In recent years, the use of remote sensing data obtained from satellite or unmanned aerial vehicle (UAV) imagery has grown in popularity for crop classification tasks such as yield prediction, soil classification or crop mapping. The ready availability of information, with improved temporal, radiometric, and spatial resolution, has resulted in the accumulation of vast amounts of data. Meeting the demands of analysing this data requires innovative solutions, and artificial intelligence techniques offer the necessary support. This systematic review aims to evaluate the effectiveness of deep learning techniques for crop classification using remote sensing data from aerial imagery. The reviewed papers focus on a variety of deep learning architectures, including convolutional neural networks (CNNs), long short-term memory networks, transformers, and hybrid CNN-recurrent neural network models, and incorporate techniques such as data augmentation, transfer learning, and multimodal fusion to improve model performance. The review analyses the use of these techniques to boost crop classification accuracy by developing new deep learning architectures or by combining various types of remote sensing data. Additionally, it assesses the impact of factors like spatial and spectral resolution, image annotation, and sample quality on crop classification. Ensembling models or integrating multiple data sources tends to enhance the classification accuracy of deep learning models. Satellite imagery is the most commonly used data source due to its accessibility and typically free availability. The study highlights the requirement for large amounts of training data and the incorporation of non-crop classes to enhance accuracy and provide valuable insights into the current state of deep learning models and datasets for crop classification tasks.

Keywords: deep learning; crop; image classification; aerial imagery



Citation: Teixeira, I.; Morais, R.; Sousa, J.J.; Cunha, A. Deep Learning Models for the Classification of Crops in Aerial Imagery: A Review. *Agriculture* **2023**, *13*, 965. <https://doi.org/10.3390/agriculture13050965>

Academic Editors: Wen-Hao Su and Zhou Zhang

Received: 27 March 2023

Revised: 24 April 2023

Accepted: 25 April 2023

Published: 27 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

Aerial imagery refers to the process of obtaining visual data of Earth using various systems, including manned aircraft, unmanned aerial vehicles (UAVs), and satellites [1], as well as other vehicles such as helicopters, balloons, and rockets [2], that are mounted with sensors for capturing images. Aerial imagery, also known as aerial photography [3], enables the examination of a broad spectrum of land areas, from a small plot to entire countries. It has been utilized for various purposes for many years, including generating precise maps and 3D models for urban planning and land management, monitoring changes in land use, and conducting environmental analysis. It is a valuable tool for assessing and monitoring land cover, including forests and agricultural fields, and for understanding the dynamics of different types of land use, including commercial, residential, transportation, and cadastral (property) areas [3]. Moreover, it is also employed in surveillance and military operations [4].

In the agricultural context, aerial imagery offers numerous advantages, such as identifying the sown area, predicting production and regulating produce distribution. Incorporating remote sensing technology to obtain data during key stages of a crop's phenological cycle and conducting multitemporal image analysis enables the measurement of specific agricultural variables and provides valuable insights into the underlying processes that affect crop development. It is especially useful for government agencies and organizations that provide financial assistance to make informed decisions regarding crop management interventions [5,6]. Aerial acquisition systems can supply either multispectral and multitemporal images and synthetic aperture radar (SAR) data.

The scientific community has leveraged artificial intelligence, particularly deep learning (DL) techniques, to automatically identify patterns in data [7,8], often achieving excellent classification results. DL for classification works by using a neural network to learn a mapping from inputs to outputs, in which the input is typically an image, text, or other types of data, and the output is a label or category [9]. However, in image classification, low resolution, subpixel sample quality, and insufficient image annotation can impair accuracy. Researchers have concentrated on investigating temporal, spatial, and spectral data. To enhance metric results, some have employed fusion techniques to combine different types of images and applied them to diverse datasets. Many studies have utilized datasets containing images from satellites or UAVs since they are readily accessible. Overall, the datasets employed vary in size and complexity, offering a wide range of data for constructing and assessing crop classification models [10–12].

This study examines the utilization of DL models for crop classification via aerial images. It scrutinizes the quantity and kind of classes employed, data sources, and model architectures. The review is organized as follows: Section 2 outlines the contributions of relevant review studies; Section 3 presents the review's scope and potential; Section 4 describes the selection process of studies included in the review; Section 5 provides a summary of the selected papers, grouped by source data, while Sections 6 and 7 analyse and summarize the research's findings.

2. Related Work

Numerous research studies have concentrated on DL classification models for aerial crops since 2016, but only a small subset of them can be classified as review studies.

In the paper [13], a review was conducted on the use of DL techniques for crop classification in SAR images. The authors conducted a thorough search of relevant papers from 2016 to 2020 in publication databases to identify research gaps and challenges in previous studies. The paper primarily focuses on two categories of classification techniques: conventional machine learning (ML) techniques and DL techniques, with convolutional neural networks (CNNs) being the most commonly used DL algorithm. The commonly used evaluation parameters include user's accuracy (UA), producer's accuracy (PA), and overall accuracy (OA). The authors observe that single-date imagery leads to inaccurate crop maps and food estimates, and they suggest that multitemporal data should be used and spatial autocorrelation should be considered to improve the classification performance of SAR data. Despite these limitations, SAR has shown significant potential in crop classification due to its ability to obtain structural information about ground targets and operate in all weather and light conditions.

In the paper [14], a survey was conducted to examine research conducted between 2017 and 2019 using DL techniques to identify or categorize weeds in different crops, including sunflower, carrot, soybean, sugar beet, and maize (i.e., corn), using CNN and deep convolutional neural network (DCNN) models. The study identified a research gap in autonomous weeding applications, in which DL has shown promising results for achieving high accuracy in weed identification and autonomous spray application. However, a scarcity of extensive datasets was revealed, highlighting the necessity for additional investigation of DL methods. The highest reported accuracy achieved was 94.74% using a dataset of 10,000 images with sugar beet as the target crop, and other

studies reported accuracies ranging from 90% to 93.64% using datasets of varying sizes to classify different crops. Furthermore, there is a lack of research on the application of DL for important crops such as sugarcane, rice, wheat, and cotton, as well as a gap in research exploring various crop and weed combinations.

The review article [15] comprehensively surveyed the existing literature on the use of DL methods for crop classification using UAV imagery. The authors discussed the importance of crop classification and the potential of UAVs and DL methods in this domain. They reviewed a wide range of studies that used various DL models, such as CNNs, recurrent neural networks (RNNs), and transfer learning, for crop classification from UAV imagery. They discussed the advantages and limitations of these models and highlighted the importance of data augmentation, feature extraction, and interpretability in crop classification. They also discussed the impact of the number and quality of the training datasets on classification accuracy.

Although the studies have analysed the use of DL and found that CNNs and DCNNs are widely used, they intend to study different aspects. The first focused only on identifying weeds. It does not allow obtaining a global view of the studies published in the research period because it excludes all publications with less than five pages and some that have already been cited. The second reviewed papers that used only SAR data and DL techniques, and the third focused only on CNN architectures applied to UAV imagery.

Thus, this work intends to answer the following research questions:

1. Which deep learning architectures are commonly employed for crop classification?

Motivation: Identify the models that can achieve higher performance.

2. How does the performance of deep learning models compare to that of machine learning?

Motivation: Evaluate the ability of deep learning to recognize and categorize images of crops.

3. What type of aerial imagery and data sources are used for training models?

Motivation: Assess the availability of the datasets and scrutinize the crops that are classified.

4. What is the number of classes employed in the classification process?

Motivation: Examine whether the number of categories utilized has an impact on the model's performance.

3. Materials and Methods

The objective of this study is to conduct a literature review on the application of DL models for the identification and classification of various types of crops using aerial images. As highlighted in Section 2, there is a paucity of research in this area. This review seeks to bridge the gap by examining recent studies, encompassing all crop types, and emphasizing the data sources, model architectures, and the number of classes involved.

The exploration of DL models and the kind of data utilized for crop classification was carried out by utilizing the keywords “image”, “crop classification”, and “deep learning”. In addition, some synonyms were defined to expand the scope of published articles in the title. Studies published between 2020 and 2022, written in English and peer-reviewed, were established as inclusion criteria. The search was limited to this time frame to obtain the most recent and relevant studies and to fill the gap identified in the previous section. Only peer-reviewed studies were included to ensure the reliability and quality of the information gathered and to prevent the inclusion of publications from predatory journals. Studies that featured the terms “leaves”, “trunks” or “disease” in the title were excluded, since they generally employ images not acquired by aerial systems and are not relevant to crop identification.

The search was conducted using Harzing's Publish or Perish software to search for papers from Google Scholar and Scopus databases, utilizing different combinations of the keywords as search terms. The search outcomes were exported to CSV format and merged

into a single Excel file to eliminate duplicates, which was done using both automated and manual methods.

4. Results

Figure 1 displays the PRISMA flow diagram [16], which encompasses a search period ranging from 29 November to 7 December 2022. Initially, 262 records were identified, and after eliminating duplicates, 166 studies remained. After reviewing the titles and abstracts, 111 records were excluded because they were not relevant to crop classification using DL models. One article was not accessible for free, despite efforts to request access from the authors, and therefore was not included in the review. Additionally, four articles were in the preprint stage, and four more had not undergone peer review, so they were also excluded. Furthermore, two articles written in languages other than English were excluded. After a comprehensive evaluation of the full text of 44 papers, 36 studies were included in the systematic review, while the excluded studies were primarily focused on segmentation, on nonaerial image classification, or on non-DL methods.

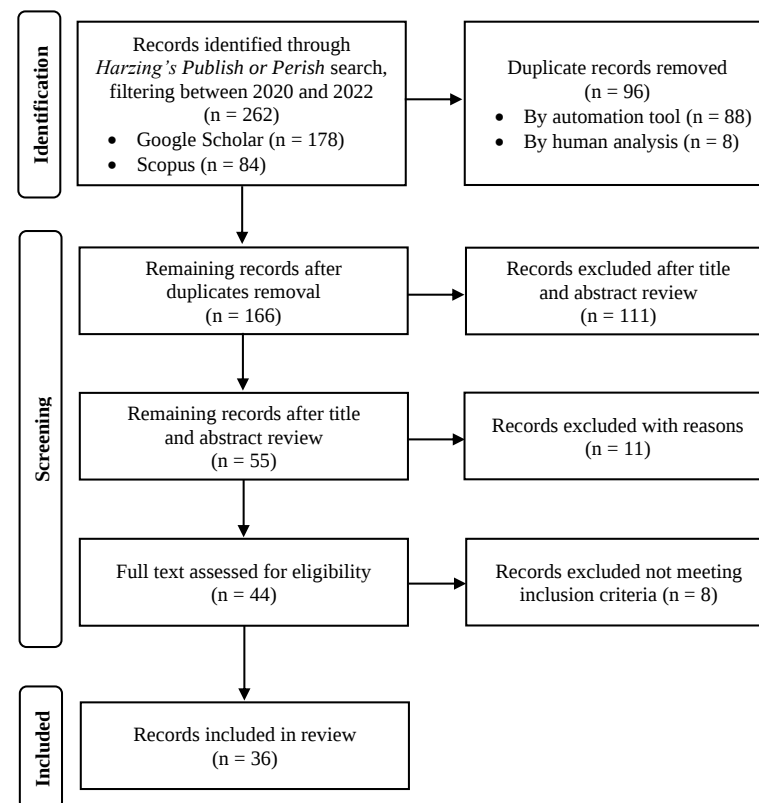


Figure 1. Flowchart depicting the identification of studies in accordance with PRISMA guidelines.

The majority of the studies considered in this review utilized DL techniques for crop classification employing remote sensing (RS) data, obtained from satellite or UAV imagery. Two studies, namely [17,18], employed aerial orthoimages of extremely high resolution obtained from aeroplanes, with the former utilizing Sentinel-2 and the latter moderate resolution imaging spectroradiometer (MODIS) satellite imagery. Additionally, reference [19,20] relied on spectral data derived from AVIRIS and ROSIS spectral sensors, respectively. These were the only exceptions to the utilization of RS data from satellite or UAV imagery across the studies included in this review. A substantial portion of the reviewed papers employed CNNs as the primary method for crop classification, while others incorporated different types of DL architectures, such as long short-term memory networks (LSTMs), transformers, and hybrid CNN-RNN models. To improve their model's performance, some papers also integrated techniques such as data augmentation, transfer

learning, and multimodal fusion. Furthermore, a significant number of papers aimed to enhance crop classification accuracy by developing novel DL architectures or by combining various RS data types, such as multispectral and multitemporal images or optical and SAR data. In addition, other papers analysed the impact of various factors on crop classification accuracy, such as spatial and spectral resolution, sample quality, and image annotation.

5. Crop Classification

The effectiveness of DL methods in crop classification is influenced by the quality and quantity of available images. Figure 2 shows the analysis of papers included in this review that utilized images and data acquired from three types of aerial systems for crop classification.

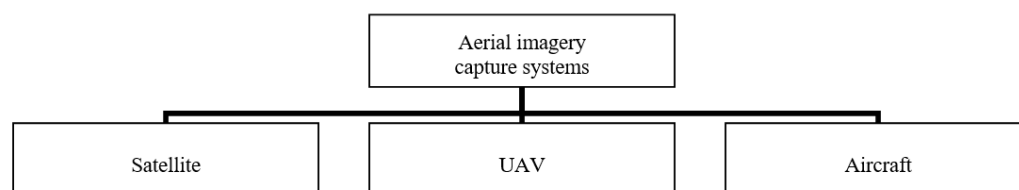


Figure 2. Aerial capture systems for images used in crop classification research.

5.1. Crop Classification Using Satellite Data

Satellite aerial data are acquired by satellites orbiting the Earth and the acquisition process involves the use of sensors that capture electromagnetic radiation reflected or emitted by the Earth's surface. The sensors used in satellites can vary in their specifications, such as spatial resolution, spectral resolution, and temporal resolution, which affect the quality and types of data that can be acquired. In addition, environmental factors such as cloud cover, atmospheric conditions, and time of day can also impact the quality of the acquired data [1]. The recent advancements in remote sensing technology have led to the development of sensors that are capable of acquiring high-quality data with improved spatial and spectral resolution. One such example is the Pleiades Neo sensor, which provides a very high spatial resolution of 0.30 m for panchromatic data acquisition. In contrast, the MODIS/Terra sensor enables access to multispectral data with resolutions ranging from 250 m to 1 km and 36 bands, covering an imaging width of 2330 km [21].

Based on the findings presented in Table 1, it can be concluded that the Sentinel satellites, belonging to the Copernicus Programme of the European Space Agency (ESA), are the most frequently utilized data source, which is in line with the goals of the AgriSAR project to evaluate the effects of the Sentinel sensor and mission characteristics on land applications, to assess quantitative trade-offs, such as spatial and radiometric resolution, and to revisit time [22]. Optical Sentinel-2 is often preferred due to its capability to provide access to multispectral and multitemporal data. In some instances, Sentinel-2 data is combined with SAR data [23–27] to address issues related to inadequate resolution, substandard image quality, as well as limitations caused by cloud cover or the inability to collect data under low-light conditions. Among the various DL models utilized in the reviewed studies, LSTM networks and their variations, as well as CNNs, were the most commonly employed.

In order to improve crop classification, the authors of [28] utilized a hybrid CNN-transformer approach to model subtle differences in crop phenology. The CNN-transformer architecture takes normalized feature maps from various sensors as input for classification. To compare performance with other classification models, the authors tested the proposed hybrid approach against a CNN-LSTM, a CNN, a support vector machine (SVM), and a random forest (RF) model. The experiment dataset contained 39,560 samples of ten different crop types, with 1% used for training and 99% for verification. Results indicate that the hybrid approach achieved an OA of 98.97%, an average accuracy (AA) of 98.92%, and a kappa coefficient of 0.9884, outperforming the other classification models, with particular success at classifying rice, corn, and grapes. However, the hybrid approach struggled with

classifying tomatoes and almonds, leading the authors to suggest further research for these specific crop types. The authors utilized a cropland data layer (CDL) from the United States Department of Agriculture (USDA) as the ground truth for crop types and removed clouds and cloud shadows from the dataset using an algorithm. In addition, the authors utilized self-organizing Kohonen maps (SOMs) to reconstruct missing data due to cloudy holes.

Table 1. Model architectures for satellite data analysis.

Paper	Year	Models	Data Source *
[28]	2020	CNN-Transformer, CNN; CNN-LSTM	Sentinel-2 (3), Landsat-8 (9)
[29]	2020	2D-CNN, 3D-CNN, LSTM	Sentinel-2, Landsat-8 (10)
[30]	2020	Conv1D-RF, VGG-RF, Conv1D, VGG	Sentinel-2 (4)
[31]	2020	1D-CNN, 2D-CNN, RNN-LSTM, RNN-GRU	Sentinel-2 (10)
[25]	2020	LSTM, MLP, U-net	Sentinel-1, Sentinel-2 (14)
[32]	2021	ANN	Sentinel-2 (4)
[33]	2021	PSE + LTAE	Sentinel-2 (20)
[34]	2021	Bi-LSTM, LSTM	Sentinel-2 (16)
[35]	2021	CNN	Sentinel-2 (11)
[36]	2021	CNN-CRF, CNN	Sentinel-1 (9)
[37]	2021	MSFCN, CNN,	Sentinel-1 (14)
[38]	2021	LSTM, CNN, GAN	Landsat-8 (3)
[39]	2021	CNN	AgriSAR (6)
[40]	2022	CNN	Sentinel2-Agri (20)
[41]	2022	CNNDAM, R-CNN, 2D-CNN, 3D-CNN	Sentinel-2 (10)
[24]	2022	2D-CNN, 3D-CNN, MLP	Sentinel-1, Sentinel-2 (7)
[42]	2022	LSTM	Sentinel-1, Sentinel-2 (6)
[43]	2022	LSTM	Sentinel-1, Sentinel-2 (2)
[44]	2022	Conv1D, LSTM	MODIS (5)
[45]	2022	ConvLSTM-RFC	Sentinel-1 (2)

* The number of classes used by the authors for each data source is indicated in parentheses.

The work presented in [29] introduced CropNet, a method that utilizes time-series multispectral images to classify crops. It employs spatial, temporal, and spectral information to improve the accuracy of classification. The method includes two primary components: 3D CNNs for deep spatial-spectral feature learning and LSTM networks for deep temporal-spectral feature learning. The outputs of these two components are merged and fed into a softmax classifier for the final crop classification. The study presented results of the method's effectiveness on two different datasets. The first dataset achieved an OA of 83.57% and a kappa coefficient of 0.7920, while the second dataset obtained an OA of 85.19% and a kappa coefficient of 0.7778, demonstrating that it outperforms other DL and ML methods.

In a particular research paper [30], a method is introduced that utilizes RS imagery acquired at three different dates. The method involves the selection of features using an optimal feature selection method (OFSM) and uses a hybrid classifier that combines a CNN with random forest (CNN-RF). The authors propose two hybrid CNN-RF networks that combine the advantages of Conv1D and Visual Geometry Group (VGG) with RF for crop classification. The study evaluates the method on a dataset consisting of rice, corn, soybean crops, and urban areas, using Sentinel-2 imagery. According to the results, the proposed method performs well in terms of OA (94.97%) and kappa coefficient (0.917).

The focus of study [31] was to classify crops in India using various techniques such as SVMs, RFs, CNNs, and RNNs, with the aid of temporal multispectral images from Sentinel-2. The authors utilized the normalized difference vegetation index (NDVI) as a feature and assessed the performance of the different models through stratified 10-fold cross-validation. Results indicated that the SVM model showed the strongest correlation with the crop areas surveyed on the ground, achieving an agreement rate of 95.9% and

the highest classification accuracy with an F1 score of 0.994, followed by the RNN with a single layer of LSTM, which yielded an F1 score of 0.783. The traditional ML models outperformed the DL models in general, and the authors speculate that this could be attributed to the limited size of the training dataset.

In order to overcome the limitations of Sentinel-2 data, the fusion of SAR and optical data can be employed. A study presented in [25] proposes a method to combine data from Sentinel-1 and Sentinel-2. The authors used three datasets from 2018, where one dataset contained all the available optical data, another had less than 10% cloud coverage, and the third contained radar images from Sentinel-1. They employed three different DL models, namely, multilayer perceptron, U-net, and a deep recurrent neural network with LSTM cells, to classify 14 types of crops in the Kyiv region of Ukraine. The LSTM network achieved the best performance, with an OA of 93.7% when using all available Sentinel-2 data and 97.5% when using the fused data. The authors concluded that their method outperforms other ML techniques and is resilient to gaps and noise in the data.

In paper [32], a technique for crop identification using artificial neural networks (ANNs) is presented. The method entails utilizing all satellite bands and information from images captured throughout the year, treating each pixel as an independent element, and striving to achieve patterns that are less reliant on specific meteorological conditions in a given year. The method is composed of three main phases: downloading and clipping Sentinel data for each polygon, preparing the input pixels for the ANN, and training the ANN model. Figure 3 demonstrates an example of the clipping process. The research centres on tobacco detection, with other crops having similar phenological patterns serving as negative examples. The results reveal that utilizing data from multiple years in the training phase enhances accuracy, with a precision of 0.9921 obtained when data from 2017, 2018, and 2019 are utilized. However, this method's primary drawback is the need for substantial storage capacity and processing time due to the usage of millions of training pixels with year-round information and data from multiple years. The technique was put into practice in a study region in Spain to manage subsidies under the European Union (EU) Common Agricultural Policy (CAP).



Figure 3. The process of image clipping using a GIS polygon. The polygon shape is created to encompass the desired region of the image, and the software then clips the image to fit within the boundary of the polygon.

In the study described in [33], a framework is presented that models crop rotation explicitly, both at the intra-annual and inter-annual scales, using a combination of Pixel Set Encoder (PSE) and Lightweight Temporal Attention Encoder (LTAE) to process Sentinel-2 satellite data. The PSE + LTAE network is also modified to model crop rotation. The authors evaluate their method on a dataset consisting of 103,602 parcels with three image time sequences and three crop annotations. In terms of crop rotation, the model achieved a mean intersection over union (mIoU) of 97.3% for permanent cultures, 77.77% for structured cultures, and 66.6% for other crops, demonstrating that their mixed-year model outperforms single-year models and leads to better performance than models that only consider a single year. However, they note that further evaluation will be possible as more Sentinel-2 data becomes available.

Satellite data has also been used in detecting crop abandonment [46–50]. For instance, the authors of [34] sought to evaluate the capability of Sentinel-2 satellite data in detecting land abandonment in the Valencia province of Spain. They utilized five spectral indices to detect temporal differences between abandoned and active land parcels and trained two DL models—an LSTM network and its bi-directional counterpart (Bi-LSTM)—on the spectral index data. The results, illustrated in Figure 4, indicated that the Bi-LSTM model achieved a maximum OA of 98.2% in identifying abandoned parcels, outperforming both LSTM and other ML models. The study also revealed that the selection of dates for calculating distances between active and abandoned parcels affected the distinguishability between classes and that the Bi-LSTM model exhibited more resilience to changes in the selected dates. Moreover, the authors demonstrated that incorporating additional metadata, such as parcel area, can enhance the model’s performance.

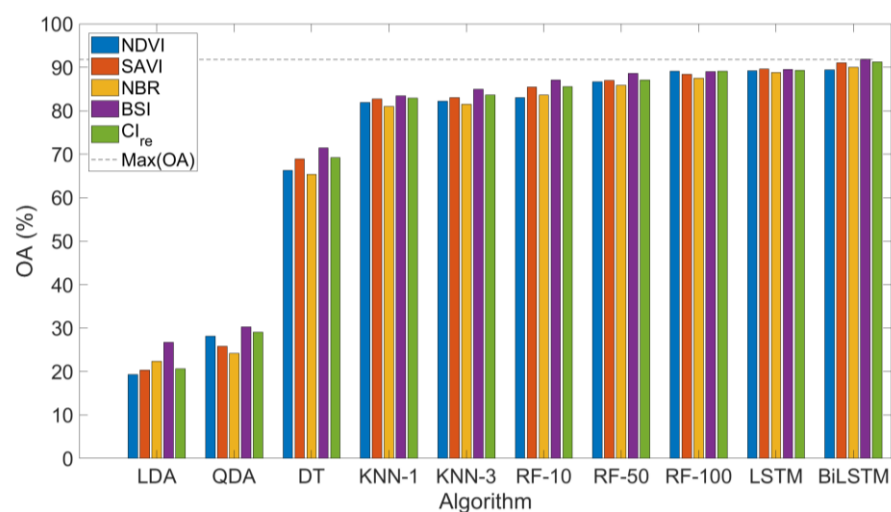


Figure 4. Comparison results of models and spectral index used.

In Spain, a novel tool was developed to aid in the implementation of the EU’s CAP. The method, as described in study [35], serves as a support tool for the Spanish government and utilizes synthetic images. These images are generated by extracting information from crop pixels and their changes over time, which are then used as input for a CNN. The method employs 13 spectral bands from Sentinel-2 images and a cloud detection system to generate the synthetic images. Data is organized in a matrix format, with one dimension representing satellite bands and the other representing temporal variation. The CNN utilized in the study employs techniques such as dropout and batch-normalization to facilitate learning and prevent overtraining. The proposed method was tested on seven crop groups in 2020 using models trained on data from 2017, 2018, and 2019. The results demonstrate that the method achieved a high level of accuracy, with an average F-score of 96% and an average accuracy of 96.23%.

The approach presented in [36] involves utilizing multitemporal RS image sequences from tropical regions for crop type classification. The method, known as CNN-CRF, combines a CNN module with a conditional random field (CRF) module to leverage both spatial and temporal context. The authors evaluated the method using a dataset of SAR Sentinel-1 images from the Campo Verde region in Brazil. The results indicate that the CNN-CRFA variant of the CRF module consistently outperformed the CNN-CRFG variant in terms of OA and average user’s accuracy (avgUA). Additionally, the CNN-CRFA variant exhibited a higher F1 score in 5 out of 9 months and was more robust, achieving better results for 7 out of 9 classes when compared to the baseline method. Specifically, the baseline method yielded an average F1-score of 69.6%, while the CNN-CRFG and CNN-CRFA models obtained 67.3% and 72.9%, respectively.

The paper [37] introduces a multiscale CNN model that utilizes time-series fully polarized synthetic aperture radar (PolSAR) images. The authors employ a sparse auto-encoder network with non-negativity constraint (NC-SAE) to reduce feature dimensions and a classifier based on a multiscale feature classification network (MSFCN). The proposed approach was tested on a site with 14 crop classes and a “non-crop” class, using simulated Sentinel-1 data and an established ground truth map for evaluation. The dual attention CNN, which is a new method, utilizes a two-stream CNN architecture with attention blocks and multiscale residual blocks to extract deep spectral and spatial features for crop mapping. The results indicate that the classification accuracy using NC-SAE compressed features improved by over 6% compared to conventional methods and that the MSFCN classifier outperformed other classifiers. The dual attention CNN model outperformed other CNN frameworks and achieved the highest accuracy for crop type mapping, with an OA of 99.33% and a kappa coefficient of 0.9919.

The authors of [38] proposed a novel approach that combines LSTM, CNN, and generative adversarial network (GAN) for crop classification. The proposed method was evaluated on three crop types (corn, soybeans, and other crops) using Landsat 8 satellite images and compared against several other methods, including SVM, SegNet, CNN, LSTM, and various combinations of these methods. The results show that the proposed model achieved the highest OA (86%) and kappa coefficient (0.7933) compared to the other methods. The authors found that the model performed best when using bands 5-6-4 of Landsat 8 for classification. Moreover, the model was tested on new data from Fayette and Pickaway County and demonstrated good generalization ability, achieving an OA of 81%.

In [39], the authors proposed a deep CNN approach that utilizes dual-polarization SAR images and H/α decomposition for feature extraction to improve classification accuracy. The study evaluated the performance of different feature combinations as input to the CNN classifier using data from the AgriSAR project of the ESA. The best classification results for the six major crops in the Indian Head dataset were achieved by using the combination of H , α , θ , and intensity as input to the CNN, with an OA of 99.30% and a kappa coefficient of 0.9903. The CNN method demonstrated significant improvement in the classification accuracy of flax.

The authors of reference [40] introduced a technique for identifying different types of crops using satellite images and a two-stream network with temporal self-attention. This approach aims to track agricultural crops at the national and international levels on a continuous basis. To achieve this, the authors employed a temporal attention module (AM) that utilizes a temporal convolutional network (TCN) to create a self-attention query. This query summarizes the satellite image sequences over multiple time steps into a one-dimensional output for the entire time series. They also generated a pseudo modality by computing the temporal differences of the original time series and integrating it with the original input data. The proposed method was tested on the publicly available Sentinel2-Agri dataset, which covers an area of 12,100 km² in southern France and consists of 13 spectral bands. The authors discovered that the TCN was the most appropriate method for summarizing the temporal information of the satellite images, and that the proposed use of the TCN output was superior to the “query as parameter” approach. Additionally, the authors conducted an ablation study to assess the significance of each stream in the crop classification task and determined that both the original data stream and the pseudo-modality stream played a critical role in improving the model’s performance. The study compared the two-stream network approach with state-of-the-art architectures and demonstrated its superiority, achieving an OA of 94.31% and an mIoU of 53.66%, outperforming the existing techniques.

The limitations of previous methods used for crop mapping using spectral-temporal information are addressed in [41] with the introduction of a new algorithm. The proposed two-stream CNN with AMs aims to overcome the shortcomings of previous methods, such as limited use of 2D/3D convolution blocks for feature extraction and a lack of spatial information. The algorithm extracts deep features for crop mapping using multiscale and

residual blocks. The first stream uses multiscale residual convolution blocks and spectral attention blocks to explore deep features, while the second stream investigates deep spatial features using spatial attention blocks. The dataset used in the study consisted of time-series images and NDVI data from Sentinel-2 images. These images were taken from the visible to shortwave infrared domains of the electromagnetic spectrum, with a total of 13 bands available in three different spatial resolutions. Thirteen images were used in the study, with some weeks being excluded due to cloud cover. The reference samples were divided using random sampling, with 3% for training, 0.1% for validation, and 96.9% for testing. Preprocessing steps, including cloud masking and atmospheric correction, were applied to the dataset. The proposed algorithm was evaluated and compared with other commonly used ML and DL methods, and it achieved the highest OA (98.54%) and kappa coefficient (0.981). The results demonstrate that the proposed algorithm is effective in mapping crop types during the growing season.

In study [24], the authors devised a novel fusion technique to combine data sources using a 3D-CNN, as depicted in Figure 5. They employed this approach to extract features from both data sources and fuse them at the feature level. The performance of this method was compared to that of other methods, such as a multilayer perceptron (MLP) and a 2D-CNN. The study used a dataset that included seven different crop types in a specific study area. The findings indicated that the fusion approach outperformed other methods, particularly when using time series data. The 3SI-3D-CNN model produced the highest OA of 88.6% and achieved the highest kappa coefficient of 86.7%. Furthermore, the results showed that the kernel depth in the 3D-convolution operator had a significant impact on the performance of the 3D-CNN.

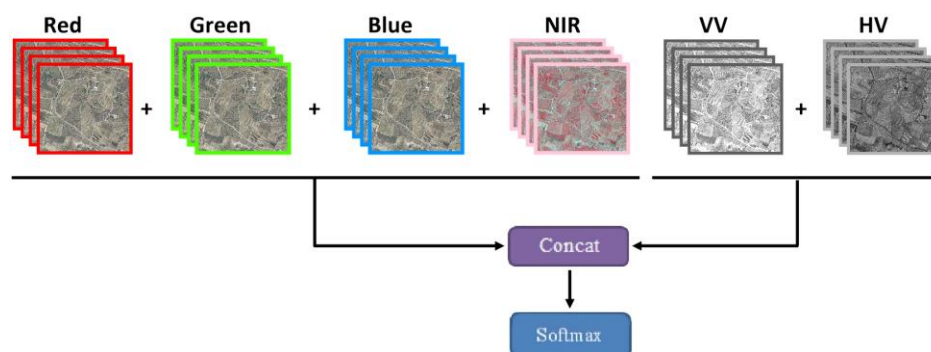


Figure 5. Network structure of optical and radar data fusion.

LSTM models, Sentinel-1 and Sentinel-2 data were utilized in studies [42,43]. The first study proposes an approach for improving classification accuracy in mountainous areas with rain and clouds. An enhanced SAR-optical (ESO) network was developed by the authors, which merges data to enhance classification performance. The network was tested in a region of Taiwan, and the precision, recall, and overall accuracy metrics were used to evaluate the results. The study found that the ESO network outperformed the use of optical data alone, achieving an OA of 62.2%, and accurately classified various crops, such as rice, corn, tea, and orchards. However, the model did not perform well for abandoned farmland because of the lack of change in SAR data for this land cover type. Additionally, the model performed better in plains than in mountainous areas. In the second study, the objective was to classify sugarcane crops utilizing time series data from both optical (NDVI) and SAR (VH) images. The authors found that using data together improved the model accuracy compared to using either NDVI or VH data alone. The model achieved an accuracy of 98%. Furthermore, the study determined that the use of both NDVI and VH data helped to prevent misidentification of algae-filled water bodies as sugarcane crops. Finally, the authors acknowledge that although the results of the study are satisfactory, there is always room for improvement, particularly since it only concentrated on a small region.

The objective of the investigation described in reference [44] is to compare the efficacy of ML and DL models for crop classification using time-series data across a large region in China. The authors employ MODIS 16-day composite surface reflectance products as input and utilize six algorithms (stacking, SVM, k-nearest neighbors, RF, Conv1D, and LSTM) on the data. The performance of the models is assessed using metrics such as accuracy, F1-score, and kappa coefficient. The findings indicate that the stacking model outperforms the other models, followed by the SVM and k-nearest neighbors models, with accuracies of 77.12%, 76.75%, and 76.19%, respectively. Additionally, the authors investigate three different sets of input features and six training strategies. Based on the results, they suggest that combining multiple classifiers and input features can lead to improved accuracy.

The study [45] evaluated various classification methods to distinguish rice fields in China by utilizing a combination of remote sensing and ground-based data. The results indicated that integrating data from different sources, such as satellite imagery, ground-based data, and weather data, substantially enhanced the precision of rice field classification. The ConvLSTM-RFC model was particularly successful, achieving an accuracy of 98.08% and a false-positive rate of just 15.08%. Moreover, it obtained the highest area under the curve (AUC) value of 88% among all the models tested. However, the study revealed that ground-based data was crucial for achieving high accuracy. The findings underscore the significance of utilizing multisource data for crop classification and the potential of these methods to facilitate precision agriculture.

The success of crop area identification through the classification of satellite images relies on the number of classes considered. Researchers have utilized various architectures for this purpose, ranging from 2 to 20 classes. The performance of these models is dependent on the data source, which determines the spatial resolution of the images. For instance, very high-resolution satellite data, with a spatial resolution of 0.50 cm/px, is only accessible through paid services. On the other hand, freely available data from Sentinel-2 has a maximum spatial resolution of ten meters, which makes manual analysis challenging and poses a difficulty for DL models to accurately classify small crops. To overcome this obstacle, some authors have chosen to incorporate hyperspectral data with five meters of spatial resolution obtained from the Sentinel-1 satellite in their research. For example, the study [51] recently focused on crop type classification using data from the complete growing season and analysed eight different crop types. Authors found that a DL approach called pixel-set encoder-temporal-attention encoder (PSE-TAE) algorithm outperformed classical approaches like RF. They also found that their method for data fusion enabled the training of models that performed better than using only Sentinel-1 or Sentinel-2 data, which is in line with previous studies using data fusion [23,24,43].

5.2. Crop Classification Using UAV Data

The choice of cameras to be mounted on UAVs plays a crucial role in the success of the UAV system. Parameters such as focal point, resolution, and quality of CCD/CMOS chips are common factors to be considered when selecting a camera. Multispectral cameras are often used alongside RGB cameras in the UAV sensors family, and they can provide high-resolution data that is not typically attainable with conventional multispectral cameras. Hyperspectral sensors are highly useful for many applications, as they capture images with hundreds of spectral bands, and their resolution can reach levels as low as 2–5 cm. Thermal infrared sensors are capable of measuring temperature in real-time, and LIDAR sensors, although being highly accurate in acquiring geometric data, the resolution of the data obtained is generally low, even with well-calibrated lightweight sensors [52].

Approximately half of the articles that employ satellite images in their research use images acquired from UAV systems, as revealed by the survey. As demonstrated in Table 2, the majority of these studies use CNN and DCNN models. Researchers usually develop their own systems to capture images and use the Wuhan University (WHU)-Hi dataset to train and validate their models.

Table 2. Model architectures for UAV data analysis.

Paper	Year	Models	Data Source *
[53]	2020	Bi-LSTM, LSTM	Custom (4)
[54]	2021	CNN, LSTM, Convolutional LSTM	Custom (4)
[55]	2021	DNN-CRF, DNN	WHU-Hi-HongHu (18), Xiong'an (20)
[56]	2021	GoogLeNet; VGG-16	Custom (6); Custom (4)
[57]	2022	DCNN, AlexNet, VGG-16, VGG-19, ResNet-50	Custom (5)
[58]	2022	ViT, EfficientNet, ResNet	Custom (5)
[59]	2022	ResNet50, VGG16, VGG19	Custom (6)
[60]	2022	Inception V3 + MFMS-DCNN, Inception V3, MFMS-DCNN, MR and PR Ensemble	Plant Seedling (12), Custom (12)
[61]	2022	1D-CNN, 3D-CNN	WHU-Hi-HongHu (22), WHU-Hi-HanChuan (16), WHU-Hi-LongKou (9)
[62]	2022	CNNCRF, SSFCN-CRF	WHU-Hi-HongHu (22), WHU-Hi-HanChuan (16), WHU-Hi-JiaYu (12)

* The number of classes used by the authors for each data source is indicated in parentheses.

Reference [53] investigated the application of the Bi-LSTM network for crop classification utilizing multitemporal data. The research was conducted in a small agricultural area in Korea, where images with a spatial resolution of 50 cm were used as inputs. The overall and class-wise accuracy of the Bi-LSTM network was compared to those of the forward and backward unidirectional LSTM networks for various combinations of input images. Three combination cases were tested, and the classification performance of the Bi-LSTM with the first and second cases (C1 and C2) achieved overall accuracies of 96.8% and 97.6%, respectively. In contrast, a significant decrease in overall accuracy was observed when the third combination case (C3) was used for classification. The outcomes indicated that the Bi-LSTM version performed well when certain images that did not provide adequate visual discrimination between crops were incorporated. Furthermore, the performance of the unidirectional LSTM was significantly impacted by the classification results of the beginning date.

Paper [54] investigated the use of the convolutional variant of the LSTM network. The study presents a two-stage DL model, which includes an LSTM-based autoencoder (LAE) and a CNN. The LAE is utilized to extract latent features from the input images, which are then fed into the CNN for classification. The effectiveness of the proposed model is evaluated using images from Anbandegi, Korea, and compared with various other DL models, such as CNN and LSTM. The findings reveal that the proposed model surpasses the other models in terms of accuracy and noise reduction, achieving the highest overall accuracy of 90.97%. Furthermore, the efficacy of the LAE in extracting informative features for crop classification is demonstrated through quantitative evaluation using the Mahalanobis distance and Jeffries–Matusita distance.

The method proposed in [55] aims to perform fine-grained crop classification using DL and multifeature fusion. The authors extract three spatial features from the images, namely morphological profile, GLCM texture, and endmember abundance features, which are then combined with the original spectral information to create fused features. A deep neural network (DNN) model is trained on these fused features using two datasets, WHU-Hi-HongHu and Xiong'an, both having a spatial resolution of around 0.4–0.5 m. The performance of the proposed method is evaluated using various metrics, including OA and kappa coefficient, and is compared with single-feature classification and different fusion strategies. The results indicate that the proposed method outperforms single-feature classification and other fusion strategies, achieving overall accuracies of up to 98.71% and 99.71%, and a kappa coefficient of 0.985 and 0.995 on the Honghu and Xiong'an

datasets, respectively. The authors also find that the classification accuracy improves with an increase in the number of training samples, and the choice of classifier significantly affects the results.

The research [56] put forward a technique for crop classification using transfer learning and ML. According to the authors, employing transfer learning can enhance the accuracy of crop classification by initially training on a substantial dataset with similar characteristics. They also found that utilizing diverse ML algorithms can further enhance classification accuracy. However, the study has limitations due to the small dataset used for testing and the need for more data to validate the proposed technique. Nevertheless, the research indicates that the proposed approach can be a valuable tool for crop classification, particularly in situations where there is a lack of training data, as it demonstrated a high level of accuracy in two specific cases, reaching nearly 83% for the Malawi dataset and up to 90% for the Mozambique dataset.

The system described in [57] utilizes a conjugated dense convolutional neural network (CD-CNN) architecture and a novel activation function known as SL-ReLU. The study was conducted on images of five distinct crops (rice, sugarcane, wheat, beans, and cumbu napier grass) obtained from a Quadcopter UAV in India, with a spatial resolution of 2.73 cm. The CD-CNN model achieved an accuracy of 96.2%, sensitivity of 96.2%, specificity of 99.05%, F1-score of 96.20%, and false positive rate of 0.95%. It was compared to other ML algorithms and standard CNN architectures and was found to perform better in terms of accuracy and other evaluation metrics.

The utilization of a transformer neural network for the classification of weeds and crops is proposed in [58]. The study was carried out in the Centre-Val de Loire region of France, where images of diverse crops, such as red-leaf beet, green-leaf beet, parsley, and spinach, were captured using a high-resolution camera-equipped drone. The images were annotated with bounding boxes and image patches of weeds and crops were extracted. The off-type beet class was upsampled 4 times to 3265 samples using data augmentation techniques to address the class imbalance. The authors employed various models, including ViT-B32, ViT-B16, EfficientNet, and ResNet and found that the transformer models outperformed the CNN models in terms of overall accuracy and precision. The best F1-score achieved was 94.4%, and the minimum loss recorded was 0.656. They also found that increasing the size of the training dataset enhanced the performance, as shown in Figure 6.

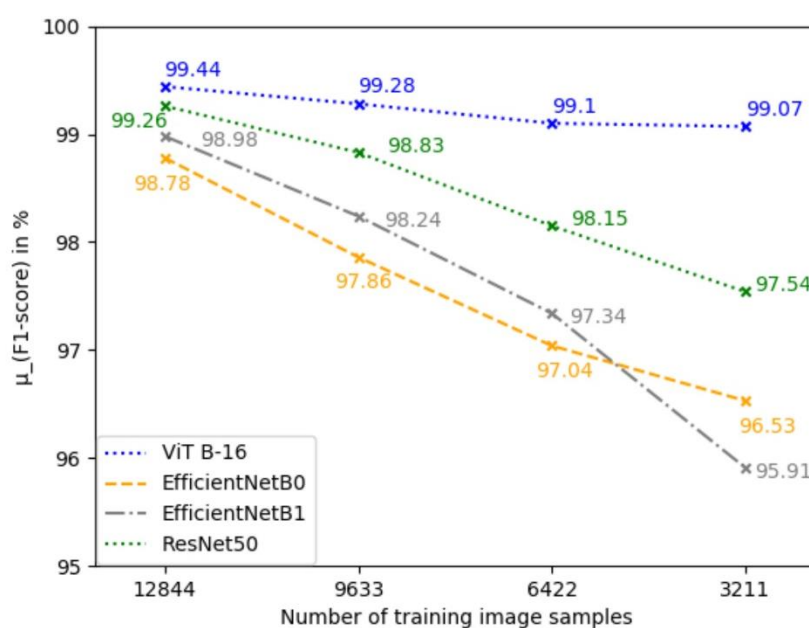


Figure 6. Comparison results of trained models.

In article [59], the authors explore the impact of sample quality on the accuracy of DL classification models for identifying multiple crop types. Two aspects of sample quality were investigated: the ratio of training and validation samples and the spatial resolution of the images. The study was conducted in North China and used three DL models (VGG16, VGG19, and ResNet50) to classify six types of crops. Different spatial resolutions and training to validation sample ratios were tested. With accuracies of up to 97.04%, the results indicate that the accuracy of classification models improves when the training samples have an equal or higher ratio than the validation samples and when higher spatial resolution images are used. The classification accuracy decreases significantly as the spatial resolution decreases. Overall, the study suggests that a flight altitude of approximately 40 m, which corresponds to a spatial resolution of around 0.28 cm, strikes a good balance between recognition accuracy and operating cost.

The authors of [60] suggest a new approach that combines DCNNs with ensemble learning to enhance classification performance. They also introduce a novel multifilter multiscale DCNN (MFMS-DCNN) architecture and two fresh datasets comprising images captured in the agricultural areas of Kolkata and Assam in India. The performance of the proposed method is assessed on these datasets, as well as on a publicly available plant seedling dataset, and it is shown to outperform current state-of-the-art techniques, achieving average accuracy levels ranging from approximately 96% to 99%.

Studies [61,62] used the WHU-Hi dataset to evaluate their methods using high spectral and very high spatial resolution imagery. In the first study, the authors examined the impact of spatial and spectral resolution on classification precision. They observed that reducing spatial resolution had a positive impact, especially when it was below 0.4 m. However, reducing the number of spectral bands had a negative impact on accuracy, with the 1D-CNN method being the most affected. The spectral-spatial residual network (SSRN) was found to be less affected, as it could make use of the spatial information in the image. Overall, SSRN was the most stable and achieved the highest accuracy under different conditions. In the second study, the authors proposed a method called S³ANet, which used a spectral-spatial-scale network with spectral, spatial, and scale attention modules and an additive angular margin loss function in an end-to-end classification framework. They also tested the method on a new dataset containing 12 rice varieties. The proposed method outperformed seven other crop classification methods in terms of classification accuracy, visualization performance, and achieved a kappa coefficient of 0.9823. The OA of the proposed method exceeded 96% for all experiments.

In the study [63], a crop classification method using UAV imagery and DL with an emphasis on explainability is proposed. The findings reveal that the method can attain high accuracy in classification and provide transparent explanations for the results. The significance of explainability in crop classification to establish trust and comprehension among users is highlighted. Nevertheless, the study's limitations include the small dataset size used for testing and the need for additional validation with larger datasets. Consequently, the research suggests that the proposed method can serve as a beneficial tool for crop classification, particularly when interpretability and transparency are necessary.

Precision agriculture has seen promising research in crop classification using UAV data, but it is essential to consider the spatial resolution of the data, as it provides detailed crop information. However, higher resolution data requires more processing, which can affect model performance. Although the use of drones for image acquisition limits the analysis to smaller areas, the high resolution of the data obtained enables DL models to classify crops with small distances between planted rows. The number of classes used in crop classification also affects the model's complexity and varies depending on the region and crop being studied, with custom datasets typically using four to twelve classes and publicly available datasets including twelve to twenty-two classes, such as corn, soybeans, wheat, and cotton. The use of CNN models is prevalent in crop classification using UAV systems due to the difficulty in capturing temporal images of the same cultivated area, presenting a challenge for other models, caused by operational constraints and the data

acquisition process associated with drones. Despite this, in the study [64], the CNN model has been proven to be highly effective in weed detection amongst commercial crops, such as Chinese cabbage. The CNN-based classifier was integrated with UAV imagery, achieving an average sensitivity of 80.29%, average specificity of 93.88%, average precision of 75.45%, and average accuracy of 92.41%, outperforming the RF-based classifier. The CNN's superiority over the RF model can be attributed to its use of convolution layers to magnify and refine features, filter out irrelevant information, and learn diverse gradient pathways with dropout, making it less likely to get trapped in a local optimum caused by an imbalanced dataset.

5.3. Crop Classification Using Multisource Data

Table 3 presents a list of papers where researchers have employed models for crop classification using data from two sources, as opposed to just one. The papers [17–20] utilized data acquired from satellites and aircraft, while [65–68] utilized data from satellite and UAV systems. Among the different network versions implemented, convolutional-based models are the most common.

Table 3. Model architectures for Satellite and UAV / Aircraft data analysis.

Paper	Year	Models	Data Source *
[17]	2020	VGG-19	Sentinel-2 (6), Single-date VHR orthoimages (2)
[18]	2020	VGG16, ResNet50, DenseNet201, LSTM, CNN, MLP	VHR USDA NAIP, MODIS (6)
[19]	2022	2D-CNN, 3D-CNN, VGG16/19, ResNet, DenseNet	Indian Pines (16), Pavia (9), Salinas (16)
[20]	2022	CNN, Convolutional Autoencoder, DNN	EO-1 Hyperion (16), Indian Pines (4)
[65]	2021	SS-OCNN, PCNN, OBIA, OCNN, MOCNN	Single-date UAVSAR (9), RapidEye (10)
[66]	2022	TS-OCNN, OCNN, OBIA, PCNN	UAVSAR (10), RapidEye (9)
[67]	2022	2D-CNN, CNN-MFL	Indian Pines (16), Salinas (16), UAV WHU-Hi-HongHu (22)
[68]	2022	DMLP, DMLPFFN, CNN, ResNet, MLP	Salinas (16), KSC (13), WHU-Hi-LongKou (9), WHU-Hi-HanChuan (16)

* The number of classes used by the authors for each data source is indicated in parentheses.

The aim of the study described in paper [17] was to develop a CNN architecture that could accurately classify various crop types in agricultural parcels using both very high resolution (VHR) aerial imagery and Sentinel-2 time-series data. The CNN was trained on a combination of VHR and multispectral images and achieved an overall classification accuracy of over 93% for different crop types, such as cereals, fruit trees, olive trees, vineyards, grasslands, and arable land. Furthermore, as shown in Figure 7, the study used the trained CNN to automatically detect the condition of permanent crops (fruit trees, olive trees, and vineyards) in the area and found that it could detect abandonment with an overall accuracy of 99%.

In paper [18] the authors proposed a method for crop classification using a multimodal DL approach that leverages both high spatial and high temporal resolution satellite imagery. The authors used National Agricultural Imagery Program (NAIP) data, which has a resolution of 1 m but low temporal resolution, and MODIS data, which has a resolution of 250 m but high temporal resolution. They extracted NDVI information from the MODIS data and used a multimodal network with two streams: a spatial stream to extract relevant spatial information from the VHR imagery and a temporal stream to extract phenological information from the high temporal resolution imagery. The authors evaluated their approach using data from various locations across the USA for six different crops: corn, cotton, soy, spring wheat, winter wheat, and barley. They compared their approach's performance with

various single-modality approaches and found that their multimodal method outperforms other methods in terms of classification accuracy (98.41%), kappa coefficient (98.08%) and average F1-score (98.44%).

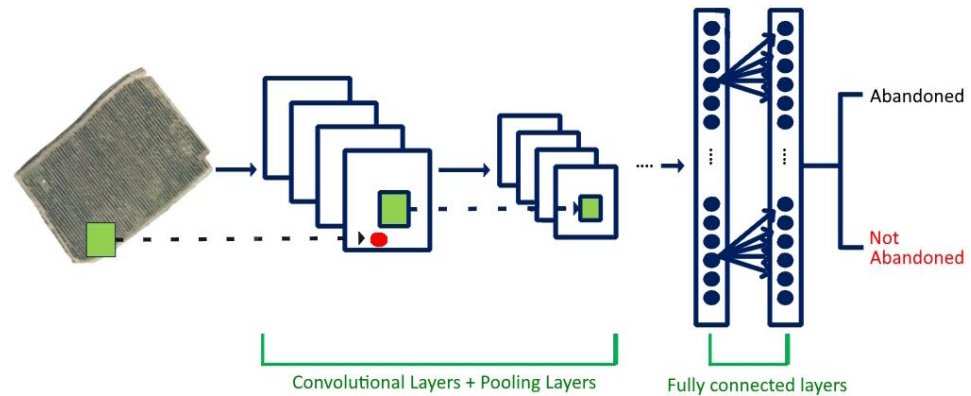


Figure 7. CNN architecture for classification of abandoned and non-abandoned crop images. The input image is analyzed and generates feature maps (green squares) that are connected to the next layer (red circle). Before the classification decision, all input values (blue circles) from one layer are connected to every next layer. (Adapted from [edrawmax.com/templates/1021526](https://www.edrawmax.com/templates/1021526) accessed on 9 January 2023).

The use of DL techniques with hyperspectral images is discussed in paper [19]. The authors suggest utilizing transfer learning to address the issue of limited training samples. They compared the performance of various DL models, including VGG16 and ResNet, as well as two customized models (2D-CNN and 3D-CNN) on three image datasets. The authors discovered that customized models outperformed other models when using homogeneous transfer learning, where the source and target datasets have similar features, with the proposed method achieving accuracies of 99% for both the Indian Pines and Pavia University datasets. To enhance efficiency and performance, the authors employed dimensionality reduction and batch normalization techniques.

The objective of the research presented in [20] was to assess the effectiveness of a CNN model using data from the Indian Pines standard dataset and the EO-1 spaceborne Hyperion sensor. The researchers compared the CNN model's performance to that of a convolutional autoencoder and a DNN. The study results demonstrated that the CNN model outperformed the other models, achieving an overall accuracy of 97% on the Indian Pines dataset and 78% on the study area dataset. Moreover, the CNN model delivered impressive performance even with a limited number of training samples. As per the authors' conclusion, the optimized model can be applied to other regions.

The reference [65] presents a novel approach, called scale sequence object-based convolutional neural network (SS-OCNN), for land cover classification using fine spatial resolution (FSR) imagery. The SS-OCNN model is an extension of the object-based convolutional neural network (OCNN) model, and it aims to address the challenge of determining the optimal input window size (scale) for the OCNN model. The performance of the SS-OCNN was evaluated on two different agricultural fields using optical and radar FSR remotely sensed imagery and compared against four other conventional methods, including the standard pixel-wise convolutional neural network (PCNN), the OCNN, the multiscale object-based convolutional neural network (MOCNN), and the object-based image analysis (OBIA). The experimental results demonstrated that the SS-OCNN model consistently outperformed the other methods in terms of OA and kappa coefficient for both the first and second study areas, achieving an OA of 87.79% and 89.46%, and a kappa coefficient of 0.86 and 0.87, respectively. The SS-OCNN model classifies the entire remotely sensed image by predicting the label of each segmented object, which can significantly

enhance computational efficiency and accuracy. Furthermore, the SS-OCNN was found to have the highest overall accuracy when tested on the SAR and optical imagery separately.

The reference [66] proposes the use of the “Temporal Sequence Object-based CNN (TS-OCNN)” method to develop a crop classification model using fine-resolution RS time series data. The method employs a CNN to process multitemporal imagery and feeds single-date images into the model in a forward temporal sequence from early to late acquisition. Additionally, an OCNN is utilized to classify crops at the object level to maintain accurate crop parcel boundaries. The study evaluates the TS-OCNN approach in two agricultural sites (S1 and S2) in California, USA using L-band radar and optical images, and compares it to traditional object-based image analysis, standard pixel-wise CNN, and standard object-based CNN methods. The findings demonstrate that the TS-OCNN approach outperforms the other methods in terms of OA (82.68% for S1 and 87.40% for S2) and kappa coefficient (0.80 for S1 and 0.85 for S2). The crop classification maps can be seen in Figure 8.

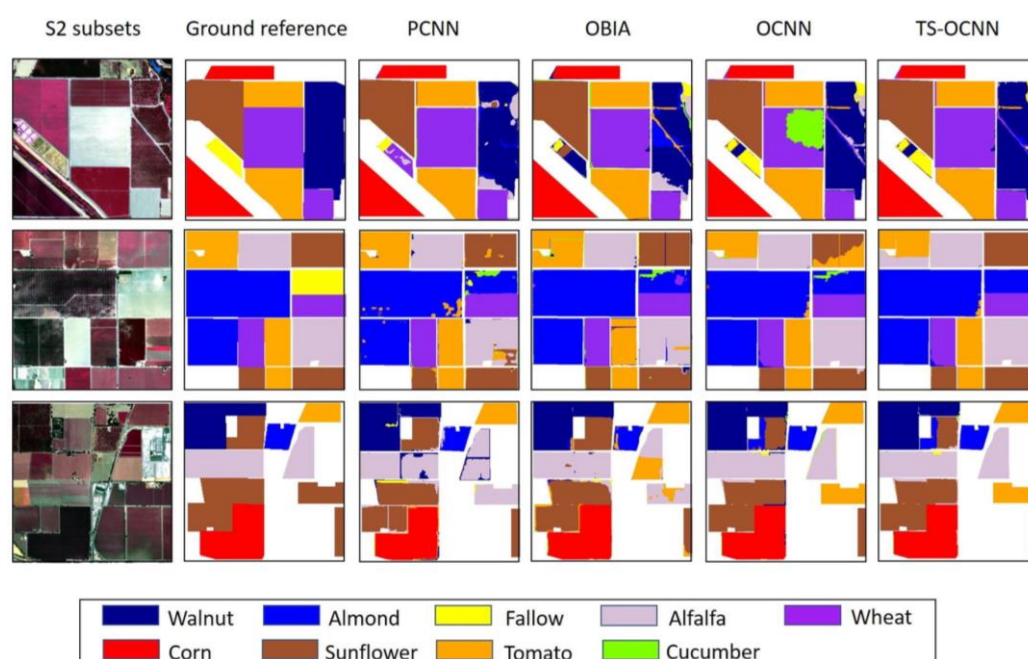


Figure 8. Crop classification maps of agricultural site S2 generated using various methods.

The approach presented in [67] consists of two stages, namely band selection and classification. During the band selection step, the original image is partitioned into three regions based on the physical and biological characteristics of the plants, and bands are chosen from each area using metrics like entropy, NDVI, and modified normalized difference water index (MNDWI). In the classification stage, the selected bands are fed into a 2D-CNN to obtain precise crop classification. The effectiveness of this method is evaluated using two datasets comprising satellite images. The method achieved an AA of 95.84% and an OA of 97.62% when classifying 16 classes in the Indian Pines dataset. Similarly, for the Salinas dataset, the method obtained an AA of 97.24% and an OA of 96.08%. Furthermore, the authors classify 15 crop categories from UAV data with an AA of 98.56%.

The use of two methods for fine crop classification, namely dilation-based multilayer perceptrons (DMLP) and DMLP with a feature fusion network (DMLPFFN), is presented in [68]. The DMLP method modifies the convolution operation with a dilated convolution to capture contextual information without losing resolution. On the other hand, DMLPFFN extracts multiscale features from the image using different branches at various stages of the network and then fuses the features using element summation to create a feature map with comprehensive information. These two methods are compared with other methods like RBF-SVM, EMP-SVM, CNN, ResNet, MLP-Mixer, RepMLP, and DFFN using

four datasets, showing their superior performance, outperforming CNN by 6.81%, 12.45%, 4.38%, and 8.84% and ResNet by 4.48%, 7.74%, 3.53%, and 6.39% on the Salinas, KSC, WHU-Hi-LongKou, and WHU-Hi-HanChuan datasets, respectively.

Multisource data is used for crop classification, which involves using different data sources such as satellite imagery, UAV data, and aircraft-captured images to identify and classify crops. This approach has the potential to improve the accuracy of crop classification compared to using single-source data. Its main objective is to combine data with high spatial resolution but low temporal resolution with data of low spatial resolution but a high temporal resolution to achieve a more accurate understanding of crops.

6. Discussion

Crop classification using DL models has become more and more popular with the increasing availability of aerial imagery, satellite data, UAV data, and other multisource data. In aerial imagery, DL models, particularly CNNs [35,37,39,41,54], have been widely used for crop classification. These models automatically learn features from images and can detect complex patterns and features that may not be visible to the naked eye. However, the accuracy of the model depends on the quality of the images used, and obtaining labelled data can be time-consuming and expensive [58].

Aerial images differ from maps in that they result from a perspective or central projection, whereas maps are created through an orthographic projection [4]. This distinction is important to consider, as aerial images can experience distortion from motion or lens, and displacement from curvature of the Earth, tilt, or topography. Because of these factors, geometric and radiometric calibration of aerial images is essential for accurate and reliable interpretation of remote sensing data [69], as it ensures alignment with their corresponding geographic locations on the ground and consistent image intensities across the entire image, accurately reflecting the true reflectance values of objects [1,3].

In satellite data, DL models can handle large datasets and can detect subtle changes in crops that are not easily detectable by human eyes. Nevertheless, the accuracy of the classification depends on the spectral resolution of the satellite data, and the cost of obtaining high-resolution satellite data can be expensive. The use of DL models for crop classification using UAV data has several advantages, including real-time information and high-resolution images. However, the cost of acquiring UAVs and the necessary equipment can be high, and the operation of UAVs is highly regulated.

Multisource data integration has become a trend in crop classification, with the combination of data from different sources improving the accuracy of crop classification. The fusion of data from different sources remains a subject of study in various fields. The recent study [70] presents a new technique for crop classification using a combination of optical and SAR image time series. The method is evaluated on a large and imbalanced dataset of 18 crop types, and the results show that it outperforms state-of-the-art methods with up to 0.42% better overall accuracy and 0.53% better mIoU. The fusion of different types of data were also studied in [71]. The authors explored the potential of DL for fusing ground-based terrestrial LiDAR point cloud and satellite (WorldView-III)-based multispectral imagery to identify three horticulture crops at different nitrogen (N) levels. The study faces challenges such as the relatively lower height and lack of sturdy geometric profiles of horticulture crops, and the need for discernible self-derived features in the DL-based model. The results show that the DCNN performs substantially better on the fused dataset when sensitivity to N level is not considered. The study concludes that the inclusion of structural crop information along with spatial data enhances performance in the fused dataset classification, and LiDAR data exhibits better results in comparison to the WorldView-III classification.

The use of transfer learning has shown promising results in crop classification by reducing the data requirements for training the DL model [19,58,72]. The newly released literature [73] presents transfer learning based on hyperspectral images and CNNs, with initial testing done on standard datasets such as Indian Pines and Salinas for validation. Additionally, a novel dataset from crops in the Kota region of Rajasthan (India) is also

experimented with. The results show that the transfer learning approach is effective with limited training samples, and the proposed model is simple, fast, and promising, even with minimal training samples.

In agricultural context, besides having been used for crop classification and identification [74–76], DL techniques have played an important role in areas such as detecting diseases [77–81], yield prediction [82–86] and weed detection [14,87–89] and have also shown great potential in detecting agricultural abandonment using remote sensing data [17,34].

As presented in this study, several studies have been conducted on crop classification using DL models. In [45], authors integrated satellite imagery, ground-based data, and weather data to classify different crops in China, reporting a significant improvement in accuracy compared to using single-source data. Authors of [56] used transfer learning with pretrained CNN models to classify different crops in UAV imagery, achieving an accuracy of over 90%. Paper [90] presents the development of an explainable DL model for crop classification using UAV imagery, using the Grad-CAM method to provide insights into how the model made decisions.

Given the swift and recent advancements in this area, it is anticipated that DL models for crop classification could prove to be a useful instrument in assisting farmers with decisionmaking. DL models can offer valuable insights into crop growth, health, potential yield, and areas of suboptimal performance, enabling farmers to refine their crop management practices and enhance their productivity and profitability. Nevertheless, it is crucial to acknowledge the models' limitations and provide farmers with adequate training and assistance in utilizing the models' findings effectively.

Thus, it was possible to answer the raised questions:

1. Which deep learning architectures are commonly employed for crop classification?

R: Researchers mostly used convolutional neural networks, long short-term memory networks, transformers, and hybrid CNN-recurrent neural network models.

2. How does the performance of deep learning models compare to that of machine learning?

R: Of the 36 articles analysed, the machine learning approach only outperformed the deep learning methods in one study. The authors suggested that it was due to the small size of the dataset.

3. What type of aerial imagery and data sources are used for training models?

R: The models were trained with data obtained from satellites, UAVs and aircraft. The most commonly used datasets in the studies reviewed are based on multitemporal and multispectral data, UAV images, and multiannual satellite imagery. Hyperspectral and dual-polarization SAR imagery were also used in some studies. Aircraft VHR images are less used due to not being freely available. The most commonly used crops include corn, rice, grapes, almonds, and walnuts. Other crops that have been studied include grass, cherries, safflower, wheat, barley, canola, and garden crops. Non-crop classes that have been included in some studies include water, built-up areas, barren land, and areas of natural vegetation. It is also worth noting that many studies have focused on classifying specific crop groups or types, such as cereal crops, fruit trees, olive trees, grassland, and arable land.

4. What is the number of classes employed in the classification process?

R: Researchers studied the classification of only two classes up to twenty-two. The number of classes affects the model's performance, mainly when the classes have similar phenological characteristics. However, the use of non-crop classes improves the overall accuracy.

7. Conclusions

After conducting a systematic review of thirty-six articles, this study aimed to answer four research questions regarding deep learning models for crop classification using remote

sensing data from aerial imagery. The reviewed papers highlighted that deep learning techniques, particularly those based on CNNs and LSTM networks, are commonly used for crop classification and tend to outperform machine learning models when sufficient data is available. The appropriate selection of the model depends on various factors such as the type of data and crop being analysed, as well as the available sample size.

Researchers analysed data from different capture systems, including satellite, UAV, and aircraft and utilized techniques such as data augmentation, transfer learning, and multimodal fusion to improve classification accuracy. Factors such as spatial and spectral resolution, sample quality, and image annotation were found to have a significant impact on the accuracy of crop classification.

In summary, this systematic review provides insights into the current state of deep learning models for crop classification and highlights important factors that affect their performance, including the requirement for large amounts of training data and the incorporation of non-crop classes to enhance accuracy. These findings can assist researchers and practitioners in selecting appropriate models and datasets for crop classification tasks. In the future, potential areas of exploration could involve integrating various data sources and creating hybrid models that incorporate both deep learning methods and conventional machine learning algorithms.

Author Contributions: Conceptualisation, I.T. and A.C.; methodology, I.T. and A.C.; formal analysis, I.T., J.J.S. and A.C.; resources, I.T., J.J.S. and A.C.; data curation, I.T., J.J.S. and A.C.; writing—original draft preparation, I.T.; writing—review and editing, I.T., J.J.S. and A.C.; supervision, R.M., J.J.S. and A.C.; project administration, R.M., J.J.S. and A.C.; funding acquisition, R.M., J.J.S. and A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by National Funds by FCT (Portuguese Foundation for Science and Technology), under the projects UIDB/04033/2020, UIDP/04033/2020, and LA/P/0063/2020 and by project “DATI—Digital Agriculture Technologies for Irrigation efficiency”, PRIMA—Partnership for Research and Innovation in the Mediterranean Area, (Research and Innovation activities), financed by the states participating in the PRIMA partnership and by the European Union through Horizon 2020.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Lillesand, T.M.; Kiefer, R.W.; Chipman, J.W. *Remote Sensing and Image Interpretation*, 7th ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2015.
2. Warf, B. Aerial Imagery: Data. In *Encyclopedia of Geography*; SAGE Publications, Inc.: Thousand Oaks, CA, USA, 2010.
3. Jensen, J.R. *Introductory Digital Image Processing: A Remote Sensing Perspective*, 4th ed.; Pearson Education, Inc.: Glenview, IL, USA, 2015.
4. Paine, D.P.; Kiser, J.D. *Aerial Photography and Image Interpretation*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2012.
5. European Court of Auditors. *Using New Imaging Technologies to Monitor the Common Agricultural Policy: Steady Progress Overall, but Slower for Climate and Environment Monitoring. Special Report No 04, 2020*; Publications Office: Luxembourg, 2021.
6. Joint Research Centre (European Commission); Milenov, P.; Lemoine, G.; Devos, W.; Fasbender, D. *Technical Guidance on the Decision to Go for Substitution of OTSC by Monitoring*; Publications Office: Ispra, Italy, 2018; ISBN 978-92-79-94173-3.
7. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
8. Chollet, F. *Deep Learning with Python*, 2nd ed.; Manning Publications Co.: Shelter Island, NY, USA, 2021; ISBN 978-1-61729-686-4.
9. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
10. Sykas, D.; Sdraka, M.; Zografakis, D.; Papoutsis, I. A Sentinel-2 Multiyear, Multicountry Benchmark Dataset for Crop Classification and Segmentation With Deep Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 3323–3339. [[CrossRef](#)]
11. Selea, T.; Pslaru, M.-F. AgriSen—A Dataset for Crop Classification. In Proceedings of the 2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), Timisoara, Romania, 1–4 September 2020; pp. 259–263.

12. Choumos, G.; Koukos, A.; Sitokonstantinou, V.; Kontoes, C. Towards Space-to-Ground Data Availability for Agriculture Monitoring. In Proceedings of the 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), Nafplio, Greece, 26–29 June 2022; pp. 1–5.
13. Desai, G.; Gaikwad, A. Deep Learning Techniques for Crop Classification Applied to SAR Imagery: A Survey. In Proceedings of the 2021 Asian Conference on Innovation in Technology (ASIANCON), Pune, India, 27–29 August 2021; pp. 1–6.
14. Moazzam, S.I.; Khan, U.S.; Tiwana, M.I.; Iqbal, J.; Qureshi, W.S.; Shah, S.I. A Review of Application of Deep Learning for Weeds and Crops Classification in Agriculture. In Proceedings of the 2019 International Conference on Robotics and Automation in Industry (ICRAI), Rawalpindi, Pakistan, 21–22 October 2019; pp. 1–6.
15. Bouguettaya, A.; Zarzour, H.; Kechida, A.; Taberkit, A.M. Deep Learning Techniques to Classify Agricultural Crops through UAV Imagery: A Review. *Neural Comput Applic* **2022**, *34*, 9511–9536. [[CrossRef](#)] [[PubMed](#)]
16. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ* **2021**, *372*, n71. [[CrossRef](#)]
17. Ruiz, L.A.; Almonacid-Caballer, J.; Crespo-Peremarch, P.; Recio, J.A.; Pardo-Pascual, J.E.; Sánchez-García, E. Automated Classification of Crop Types and Condition in a Mediterranean Area Using a Fine-Tuned Convolutional Neural Network. In *Proceedings of the The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*; Copernicus Publications: Nice, France, 2020; Volume XLIII-B3-2020, pp. 1061–1068.
18. Gadiraju, K.K.; Ramachandra, B.; Chen, Z.; Vatsavai, R.R. Multimodal Deep Learning Based Crop Classification Using Multispectral and Multitemporal Satellite Imagery. In *Proceedings of the Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*; Association for Computing Machinery: New York, NY, USA, 2020; pp. 3234–3242.
19. Patel, U.; Pathan, M.; Kathiria, P.; Patel, V. Crop Type Classification with Hyperspectral Images Using Deep Learning: A Transfer Learning Approach. *Model. Earth Syst. Environ.* **2022**. [[CrossRef](#)]
20. Bhosle, K.; Musande, V. Evaluation of CNN Model by Comparing with Convolutional Autoencoder and Deep Neural Network for Crop Classification on Hyperspectral Imagery. *Geocarto Int.* **2022**, *37*, 813–827. [[CrossRef](#)]
21. Pluto-Kossakowska, J. Review on Multitemporal Classification Methods of Satellite Images for Crop and Arable Land Recognition. *Agriculture* **2021**, *11*, 999. [[CrossRef](#)]
22. Hajnsek, I.; Bianchi, R.; Davidson, M.; D’Urso, G.; Gomez-Sanchez, J.A.; Hausold, A.; Horn, R.; Howse, J.; Loew, A.; Lopez-Sanchez, J.; et al. AGRISAR 2006—Airborne SAR and Optics Campaigns for an Improved Monitoring of Agricultural Processes and Practices. *Geophys. Res. Abstr.* **2007**, *9*, 04085.
23. Orynbaikyzy, A.; Gessner, U.; Mack, B.; Conrad, C. Crop Type Classification Using Fusion of Sentinel-1 and Sentinel-2 Data: Assessing the Impact of Feature Selection, Optical Data Availability, and Parcel Sizes on the Accuracies. *Remote Sens.* **2020**, *12*, 2779. [[CrossRef](#)]
24. Teimouri, M.; Mokhtarzade, M.; Baghdadi, N.; Heipke, C. Fusion of Time-Series Optical and SAR Images Using 3D Convolutional Neural Networks for Crop Classification. *Geocarto Int.* **2022**, *37*, 1–18. [[CrossRef](#)]
25. Kussul, N.; Lavreniuk, M.; Shumilo, L. Deep Recurrent Neural Network for Crop Classification Task Based on Sentinel-1 and Sentinel-2 Imagery. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 6914–6917.
26. 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](#)]
27. Jiao, X.; McNairn, H.; Yekkehkhany, B.; Dingle Robertson, L.; Ihuoma, S. Integrating Sentinel-1 SAR and Sentinel-2 Optical Imagery with a Crop Structure Dynamics Model to Track Crop Condition. *Int. J. Remote Sens.* **2022**, *43*, 6509–6537. [[CrossRef](#)]
28. Li, Z.; Chen, G.; Zhang, T. A CNN-Transformer Hybrid Approach for Crop Classification Using Multitemporal Multisensor Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 847–858. [[CrossRef](#)]
29. Luo, C.; Meng, S.; Hu, X.; Wang, X.; Zhong, Y. Cropnet: Deep Spatial-Temporal-Spectral Feature Learning Network for Crop Classification from Time-Series Multi-Spectral Images. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 4187–4190.
30. Yang, S.; Gu, L.; Li, X.; Jiang, T.; Ren, R. Crop Classification Method Based on Optimal Feature Selection and Hybrid CNN-RF Networks for Multi-Temporal Remote Sensing Imagery. *Remote Sens.* **2020**, *12*, 3119. [[CrossRef](#)]
31. Koppaka, R.; Moh, T.-S. Machine Learning in Indian Crop Classification of Temporal Multi-Spectral Satellite Image. In Proceedings of the 2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM), Taichung, Taiwan, 3–5 January 2020; pp. 1–8.
32. Lozano-Tello, A.; Fernández-Sellers, M.; Quirós, E.; Frago-Campón, L.; García-Martín, A.; Gutiérrez Gallego, J.A.; Mateos, C.; Trenado, R.; Muñoz, P. Crop Identification by Massive Processing of Multiannual Satellite Imagery for EU Common Agriculture Policy Subsidy Control. *Eur. J. Remote Sens.* **2021**, *54*, 1–12. [[CrossRef](#)]
33. Quinton, F.; Landrieu, L. Crop Rotation Modeling for Deep Learning-Based Parcel Classification from Satellite Time Series. *Remote Sens.* **2021**, *13*, 4599. [[CrossRef](#)]
34. Portalés-Julià, E.; Campos-Taberner, M.; García-Haro, F.J.; Gilabert, M.A. Assessing the Sentinel-2 Capabilities to Identify Abandoned Crops Using Deep Learning. *Agronomy* **2021**, *11*, 654. [[CrossRef](#)]

35. Siesto, G.; Fernández-Sellers, M.; Lozano-Tello, A. Crop Classification of Satellite Imagery Using Synthetic Multitemporal and Multispectral Images in Convolutional Neural Networks. *Remote Sens.* **2021**, *13*, 3378. [\[CrossRef\]](#)
36. Rosa, L.E.C.L.; Oliveira, D.A.B.; Feitosa, R.Q. End-to-End CNN-CRFs for Multi-Date Crop Classification Using Multitemporal Remote Sensing Image Sequences. In Proceedings of the Proceedings of the CIKM 2021 Workshops Co-Located with 30th ACM International Conference on Information and Knowledge Management (CIKM 2021), Gold Coast, QLD, Australia, 1–5 November 2021.
37. Zhang, W.-T.; Wang, M.; Guo, J. A Novel Multi-Scale CNN Model for Crop Classification with Time-Series Fully Polarization SAR Images. In Proceedings of the 2021 2nd China International SAR Symposium (CISS), Shanghai, China, 3–5 November 2021; pp. 1–5.
38. Li, J.; Shen, Y.; Yang, C. An Adversarial Generative Network for Crop Classification from Remote Sensing Timeseries Images. *Remote Sens.* **2021**, *13*, 65. [\[CrossRef\]](#)
39. Guo, J.; Bai, Q.-Y.; Li, H.-H. Crop Classification Using Differential-Scattering-Feature Driven CNN for Dual-Pol SAR Images. In Proceedings of the 2021 2nd China International SAR Symposium (CISS), Shanghai, China, 3–5 November 2021; pp. 1–5.
40. Stergioulas, A.; Dimitropoulos, K.; Grammalidis, N. Crop Classification from Satellite Image Sequences Using a Two-Stream Network with Temporal Self-Attention. In Proceedings of the 2022 IEEE International Conference on Imaging Systems and Techniques (IST), Kaohsiung, Taiwan, 21–23 June 2022; pp. 1–6.
41. 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. [\[CrossRef\]](#)
42. Sun, Y.; Li, Z.-L.; Luo, J.; Wu, T.; Liu, N. Farmland Parcel-Based Crop Classification in Cloudy/Rainy Mountains Using Sentinel-1 and Sentinel-2 Based Deep Learning. *Int. J. Remote Sens.* **2022**, *43*, 1054–1073. [\[CrossRef\]](#)
43. Sreedhar, R.; Varshney, A.; Dhanya, M. Sugarcane Crop Classification Using Time Series Analysis of Optical and SAR Sentinel Images: A Deep Learning Approach. *Remote Sens. Lett.* **2022**, *13*, 812–821. [\[CrossRef\]](#)
44. Wang, X.; Zhang, J.; Xun, L.; Wang, J.; Wu, Z.; Hanchiri, M.; Zhang, S.; Zhang, S.; Bai, Y.; Yang, S.; et al. Evaluating the Effectiveness of Machine Learning and Deep Learning Models Combined Time-Series Satellite Data for Multiple Crop Types Classification over a Large-Scale Region. *Remote Sens.* **2022**, *14*, 2341. [\[CrossRef\]](#)
45. Chang, Y.-L.; Tan, T.-H.; Chen, T.-H.; Chuah, J.H.; Chang, L.; Wu, M.-C.; Tatini, N.B.; Ma, S.-C.; Alkhaleefah, M. Spatial-Temporal Neural Network for Rice Field Classification from SAR Images. *Remote Sens.* **2022**, *14*, 1929. [\[CrossRef\]](#)
46. Yusoff, N.M.; Muharam, F.M. The Use of Multi-Temporal Landsat Imageries in Detecting Seasonal Crop Abandonment. *Remote Sens.* **2015**, *7*, 11974–11991. [\[CrossRef\]](#)
47. Szatmári, D.; Feranec, J.; Goga, T.; Rusnák, M.; Kopecká, M.; O’ahel’, J. The Role of Field Survey in the Identification of Farmland Abandonment in Slovakia Using Sentinel-2 Data. *Can. J. Remote Sens.* **2021**, *47*, 569–587. [\[CrossRef\]](#)
48. López-Andreu, F.J.; Erena, M.; Dominguez-Gómez, J.A.; López-Morales, J.A. Sentinel-2 Images and Machine Learning as Tool for Monitoring of the Common Agricultural Policy: Calasparra Rice as a Case Study. *Agronomy* **2021**, *11*, 621. [\[CrossRef\]](#)
49. Löw, F.; Fliemann, E.; Abdullaev, I.; Conrad, C.; Lamers, J.P.A. Mapping Abandoned Agricultural Land in Kyzyl-Orda, Kazakhstan Using Satellite Remote Sensing. *Appl. Geogr.* **2015**, *62*, 377–390. [\[CrossRef\]](#)
50. Volpi, I.; Marchi, S.; Petacchi, R.; Hoxha, K.; Guidotti, D. Detecting Olive Grove Abandonment with Sentinel-2 and Machine Learning: The Development of a Web-Based Tool for Land Management. *Smart Agric. Technol.* **2023**, *3*, 100068. [\[CrossRef\]](#)
51. Weilandt, F.; Behling, R.; Goncalves, R.; Madadi, A.; Richter, L.; Sanona, T.; Spengler, D.; Welsch, J. Early Crop Classification via Multi-Modal Satellite Data Fusion and Temporal Attention. *Remote Sens.* **2023**, *15*, 799. [\[CrossRef\]](#)
52. Yao, H.; Qin, R.; Chen, X. Unmanned Aerial Vehicle for Remote Sensing Applications—A Review. *Remote Sens.* **2019**, *11*, 1443. [\[CrossRef\]](#)
53. Kwak, G.-H.; Park, C.-W.; Ahn, H.-Y.; Na, S.-I.; Lee, K.-D.; Park, N.-W. Potential of Bidirectional Long Short-Term Memory Networks for Crop Classification with Multitemporal Remote Sensing Images. *Korean J. Remote Sens.* **2020**, *36*, 515–525. [\[CrossRef\]](#)
54. Kwak, G.-H.; Park, N.-W. Two-Stage Deep Learning Model with LSTM-Based Autoencoder and CNN for Crop Classification Using Multi-Temporal Remote Sensing Images. *Korean J. Remote Sens.* **2021**, *37*, 719–731. [\[CrossRef\]](#)
55. Wei, L.; Wang, K.; Lu, Q.; Liang, Y.; Li, H.; Wang, Z.; Wang, R.; Cao, L. Crops Fine Classification in Airborne Hyperspectral Imagery Based on Multi-Feature Fusion and Deep Learning. *Remote Sens.* **2021**, *13*, 2917. [\[CrossRef\]](#)
56. Nowakowski, A.; Mrziglod, J.; Spiller, D.; Bonifacio, R.; Ferrari, I.; Mathieu, P.P.; Garcia-Herranz, M.; Kim, D.-H. Crop Type Mapping by Using Transfer Learning. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *98*, 102313. [\[CrossRef\]](#)
57. Pandey, A.; Jain, K. An Intelligent System for Crop Identification and Classification from UAV Images Using Conjugated Dense Convolutional Neural Network. *Comput. Electron. Agric.* **2022**, *192*, 106543. [\[CrossRef\]](#)
58. 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. [\[CrossRef\]](#)
59. Li, X.; Cui, Y.; Zhou, X.; Liu, S. Impact of Sample Quality to Deep Learning Classification Model of Multiple Crop Types on UAV Remotely Sensed Images. *Appl. Math. Model. Comput. Simul.* **2022**, *20*, 475–488. [\[CrossRef\]](#)
60. Kalita, I.; Singh, G.P.; Roy, M. Explainable Crop Classification from by Analyzing an Ensemble of DCNNs under Multi-Filter & Multi-Scale Framework. *Multimed. Tools Appl.* **2022**, *82*, 18409–18433. [\[CrossRef\]](#)

61. Yang, B.; Hu, S. What Kind Of Spatial And Spectral Resolution Of Uav-Borne Hyperspectral Image Is Required For Precise Crop Classification When Using Deep Learning. In Proceedings of the 2022 12th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Rome, Italy, 13–16 September 2022; pp. 1–5.
62. Hu, X.; Wang, X.; Zhong, Y.; Zhang, L. S3ANet: Spectral-Spatial-Scale Attention Network for End-to-End Precise Crop Classification Based on UAV-Borne H2 Imagery. *ISPRS J. Photogramm. Remote Sens.* **2022**, *183*, 147–163. [\[CrossRef\]](#)
63. Wang, Y.; Li, Z.; Li, P.; Liu, X.; Li, Y.; Wang, Y. Explainable Crop Classification from UAV Imagery Based on Deep Learning. *Remote Sens.* **2020**, *12*, 1842. [\[CrossRef\]](#)
64. Ong, P.; Teo, K.S.; Sia, C.K. UAV-Based Weed Detection in Chinese Cabbage Using Deep Learning. *Smart Agric. Technol.* **2023**, *4*, 100181. [\[CrossRef\]](#)
65. Li, H.; Zhang, C.; Zhang, Y.; Zhang, S.; Ding, X.; Atkinson, P.M. A Scale Sequence Object-Based Convolutional Neural Network (SS-OCNN) for Crop Classification from Fine Spatial Resolution Remotely Sensed Imagery. *Int. J. Digit. Earth* **2021**, *14*, 1528–1546. [\[CrossRef\]](#)
66. Li, H.; Tian, Y.; Zhang, C.; Zhang, S.; Atkinson, P.M. Temporal Sequence Object-Based CNN (TS-OCNN) for Crop Classification from Fine Resolution Remote Sensing Image Time-Series. *Crop J.* **2022**, *10*, 1507–1516. [\[CrossRef\]](#)
67. Agilandeewari, L.; Prabukumar, M.; Radhesyam, V.; Phaneendra, K.L.N.B.; Farhan, A. Crop Classification for Agricultural Applications in Hyperspectral Remote Sensing Images. *Appl. Sci.* **2022**, *12*, 1670. [\[CrossRef\]](#)
68. Wu, H.; Zhou, H.; Wang, A.; Iwahori, Y. Precise Crop Classification of Hyperspectral Images Using Multi-Branch Feature Fusion and Dilation-Based MLP. *Remote Sens.* **2022**, *14*, 2713. [\[CrossRef\]](#)
69. Shafique, A.; Cao, G.; Khan, Z.; Asad, M.; Aslam, M. Deep Learning-Based Change Detection in Remote Sensing Images: A Review. *Remote Sens.* **2022**, *14*, 871. [\[CrossRef\]](#)
70. Yuan, Y.; Lin, L.; Zhou, Z.-G.; Jiang, H.; Liu, Q. Bridging Optical and SAR Satellite Image Time Series via Contrastive Feature Extraction for Crop Classification. *ISPRS J. Photogramm. Remote Sens.* **2023**, *195*, 222–232. [\[CrossRef\]](#)
71. Reji, J.; Nidamanuri, R.R. Deep Learning Based Fusion of LiDAR Point Cloud and Multispectral Imagery for Crop Classification Sensitive to Nitrogen Level. In Proceedings of the 2023 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS), Hyderabad, India, 27–29 January 2023; Volume 1, pp. 1–4.
72. Divyanth, L.G.; Guru, D.S.; Soni, P.; Machavaram, R.; Nadimi, M.; Paliwal, J. Image-to-Image Translation-Based Data Augmentation for Improving Crop/Weed Classification Models for Precision Agriculture Applications. *Algorithms* **2022**, *15*, 401. [\[CrossRef\]](#)
73. Munipalle, V.K.; Nelakuditi, U.R.; Nidamanuri, R.R. Agricultural Crop Hyperspectral Image Classification Using Transfer Learning. In Proceedings of the 2023 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS), Hyderabad, India, 27–29 January 2023; Volume 1, pp. 1–4.
74. 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. Indian Soc. Remote Sens.* **2019**, *47*, 1949–1958. [\[CrossRef\]](#)
75. Patel, H.; Bhagia, N.; Vyas, T.; Bhattacharya, B.; Dave, K. Crop Identification and Discrimination Using AVIRIS-NG Hyperspectral Data Based on Deep Learning Techniques. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 3728–3731.
76. Yi, Z.; Jia, L.; Chen, Q.; Jiang, M.; Zhou, D.; Zeng, Y. Early-Season Crop Identification in the Shiyang River Basin Using a Deep Learning Algorithm and Time-Series Sentinel-2 Data. *Remote Sens.* **2022**, *14*, 5625. [\[CrossRef\]](#)
77. Hruška, J.; Adão, T.; Pádua, L.; Marques, P.; Peres, E.; Sousa, A.; Morais, R.; Sousa, J.J. Deep Learning-Based Methodological Approach for Vineyard Early Disease Detection Using Hyperspectral Data. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 9063–9066.
78. Kulkarni, O. Crop Disease Detection Using Deep Learning. In Proceedings of the 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 16–18 August 2018; pp. 1–4.
79. Park, H.; JeeSook, E.; Kim, S.-H. Crops Disease Diagnosing Using Image-Based Deep Learning Mechanism. In Proceedings of the 2018 International Conference on Computing and Network Communications (CoCoNet), Astana, Kazakhstan, 15–17 August 2018; pp. 23–26.
80. Long, Y.; Liu, C. Research on Deep Learning Method of Crop Disease Identification. In *Proceedings of the Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1–6.
81. Rangarajan, A.K.; Purushothaman, R.; Ramesh, A. Tomato Crop Disease Classification Using Pre-Trained Deep Learning Algorithm. *Procedia Comput. Sci.* **2018**, *133*, 1040–1047. [\[CrossRef\]](#)
82. 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 Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies; Association for Computing Machinery: New York, NY, USA, 2018; pp. 1–5.
83. Muruganatham, 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. [\[CrossRef\]](#)
84. Tsouli Fathi, M.; Ezziyyani, M.; Ezziyyani, M.; El Mamoune, S. Crop Yield Prediction Using Deep Learning in Mediterranean Region. In Proceedings of the Advanced Intelligent Systems for Sustainable Development (AI2SD'2019); Ezziyyani, M., Ed.; Springer International Publishing: Cham, Switzerland, 2020; pp. 106–114.

85. Renju, R.S.; Deepthi, P.S.; Chitra, M.T. A Review of Crop Yield Prediction Strategies Based on Machine Learning and Deep Learning. In Proceedings of the 2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS), Kochi, India, 23–25 June 2022; pp. 1–6.
86. 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](#)]
87. Bah, M.D.; Hafiane, A.; Canals, R. Deep Learning with Unsupervised Data Labeling for Weed Detection in Line Crops in UAV Images. *Remote Sens.* **2018**, *10*, 1690. [[CrossRef](#)]
88. Osorio, K.; Puerto, A.; Pedraza, C.; Jamaica, D.; Rodríguez, L. A Deep Learning Approach for Weed Detection in Lettuce Crops Using Multispectral Images. *AgriEngineering* **2020**, *2*, 471–488. [[CrossRef](#)]
89. Khan, S.; Tufail, M.; Khan, M.T.; Khan, Z.A.; Anwar, S. Deep Learning-Based Identification System of Weeds and Crops in Strawberry and Pea Fields for a Precision Agriculture Sprayer. *Precis. Agric* **2021**, *22*, 1711–1727. [[CrossRef](#)]
90. Bai, X.; Wang, X.; Liu, X.; Liu, Q.; Song, J.; Sebe, N.; Kim, B. Explainable Deep Learning for Efficient and Robust Pattern Recognition: A Survey of Recent Developments. *Pattern Recognit.* **2021**, *120*, 108102. [[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.