Data analytics for crop management: a big data view

Nabila Chergui[1] and Mohand Tahar Kechadi[1,2*]

*Correspondence:
tahar.kechadi@ucd.ie

[1] Faculty of Technology, Ferhat
Abbas University, Sétif, Algeria
[2] School of Computer Science,
University College Dublin,
Dublin, Ireland

Abstract

Recent advances in Information and Communication Technologies have a significant impact on all sectors of the economy worldwide. Digital Agriculture appeared as a consequence of the democratisation of digital devices and advances in artificial intelligence and data science. Digital agriculture created new processes for making farming more productive and efficient while respecting the environment. Recent and sophisticated digital devices and data science allowed the collection and analysis of vast amounts of agricultural datasets to help farmers, agronomists, and professionals understand better farming tasks and make better decisions. In this paper, we present a systematic review of the application of data mining techniques to digital agriculture. We introduce the crop yield management process and its components while limiting this study to crop yield and monitoring. After identifying the main categories of data mining techniques for crop yield monitoring, we discuss a panoply of existing works on the use of data analytics. This is followed by a general analysis and discussion on the impact of big data on agriculture.

Keywords: Digital agriculture, Data analytics, Crop management, Big data, Data mining, Machine learning

Introduction

DA, (also called digital farming or smart farming)[1] [78, 105, 130], is a modern approach that uses digital and smart devices [sensors, cameras, satellite, drones, the Global Positioning System (GPS)] in conjunction with Data Mining (or data analytics) to improve productivity and to optimise the use of resources. Digital Agriculture (DA) comes as a response to the increasing demand for improving productivity while reducing farming operational costs. Moreover, the improvement of productivity should not be done at any cost, e.g., overuse of natural resources and chemical products. DA can, for example, manage crop growth by finding appropriate fertilisation program for each farming field and can help farmers to reduce their operational costs and respect the environment by refining their farming operations based on the needs of each part of the farming field.

Since agriculture has a direct and significant impact on the population and therefore its economic environment, DA in its turn should be viewed as the next natural step to

---

[1] European Commission. Brussels. Preparing for Future AKIS in Europe, 2019.
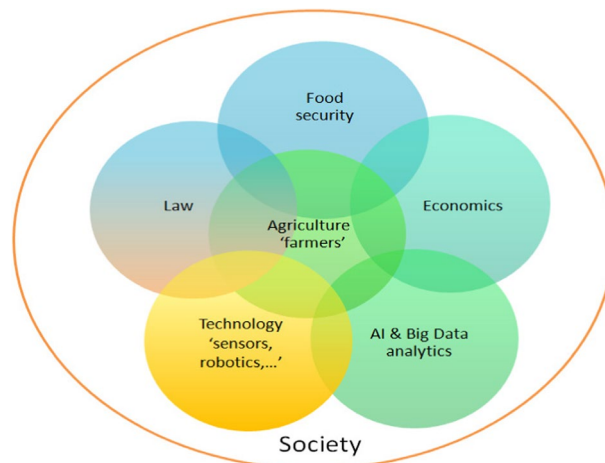
Fig. 1 The interlocked sectors involved in DA

respond to the world population's needs while protecting the environment, by taking advantage of the recent technological advances in digital devices, communications systems, and artificial intelligence. These allow us to construct multidimensional domains, where the farms and farmers are their central subjects. Figure 1 shows the agriculture ecosystem and its direct impact on other sectors of the economy.

Besides, since DA involves the development, adoption and iteration with digital technologies [39], and Artificial Intelligence (data analytics, ...), these developments and interactions should be well-defined (laws, regulations and policies) to guarantee rights and benefits of all the involved actors (farmers, farm holders', data owners', developers and analysts, technology vendors',...) [70, 77, 78, 92, 113, 146].

DA can be regarded as a data driven form of farming, in which decision-making processes are based on explicit information derived from data collected through various sources [148]. DA and Precision Agriculture (PA) seem to refer to the same thing, however, as stated in [148], DA involves the development and adoption of modern technologies in both collecting the data and its analysis in various farming contexts, while PA takes into account only the in-field variability [147]. DA aims to exploit advanced digital devices, ranging from a simple sensor to complex robots, to offer the required farmland treatment with high accuracy. DA can be applied in almost all agricultural fields. For instance, in crop production: DA allows accurate management of crops, which includes fields, wasteland, crop, pest, and irrigation management, soil classification, etc. In Animal production: DA allows monitoring the animal over its whole life cycle, its food quantity, health control and protection from diseases, and so on. Fishery, animal Husbandry, livestock and dairy farming are some examples [14]. In Forestry: We can efficiently manage forests by supporting the environmental and sustainable decision [36]. DA can help in detecting unhealthy trees, air pollution, discriminate different tree species, protect the wildlife, etc. From the economy point of view, the application of DA for forest management enhances the wood quality and its production, which can augment profits; reduce waste and maintain the environment [138].

Addressing DA from all the above mentioned views is a challenging task and cannot be achieved without the participation of specialists from all these sectors. In this study,
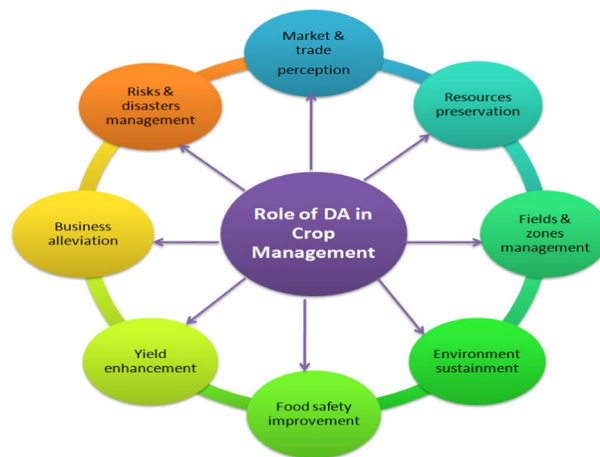
Fig. 2 Role of DA in crop production process

we focus on the use of Big Data in crop management, it is, not only one of the pillars in agriculture but also it can profoundly affect biodiversity. Moreover, crop growth is a very complex process involving various endogenous and exogenous factors. Recent advances in digital technologies allow us to collect data about all these factors. DA has the ability to elucidate the correlations and interactions of these factor to help farmers and agronomists optimise the productivity while reducing the side effects on the environment. DA exhibits several benefits to agriculture as shown in Figure 2. These benefits were discussed in [10, 13, 70, 98, 104, 112, 113, 130, 135, 148] and summarised in the following:

- DA provides a farmer with useful information to support their decision-making processes, such as soil and weather monitoring and prediction, weed and pest monitoring, crop yield dynamic predictions, etc.
- DA can sustain the environment and improve the products' quality, since it provides high quality information and measurements for optimal farming operations on each field.
- DA can provide farmers advanced management methods against climate change and other environmental challenges. The farmer can continuously monitor crop growth and protect them against diseases.
- DA offers valuable feedback to farmers and good assessment of risks, to minimise microbiological or disaster-related risks.
- DA can provide prediction and assistance to farmers against adverse weather incidence, disasters and market instability by assessing the loss at the farm level.
- Farmers/agronomists can benefit from advanced models to understand the market and forecast which products could be more profitable.

The contributions of this study are in the investigation of big data analytics applications to crop production. Crop farming is a complex task, and it depends on many factors that should be taken into account. To optimise the operational cost and reduce the impact on the environment, the big data analytics emerges as one of the most cost effective approaches nowadays. The contributions, therefore, include the following:

- A comprehensive overview of Digital Agriculture big-data with a presentation of the conceptual-layered framework to show the effectiveness of data analytics on Digital Agriculture, when some necessary steps have been implemented. For instance, large-scale data analytics can only be effective if the historical data is available, carefully collected, and it is of high quality.
- A highlight of the different types of data used in the existing studies, and a presentation of the classification of different techniques applied to crop yield monitoring and their effectiveness of the overall results.
- A review and analytical studies of the most widely used data mining techniques to crop farming, with a report of their advantages and shortcomings.
- A discussion on the advantages of big data in agriculture, and how this can be used efficiently for crop farming and its extension to the agricultural field in general.
- A discussion on Digital Agriculture applications for crop management in small and large scale holders.
- A discussion on Digital Agriculture challenges and potential paths for future research.

Methodology

To study the impact of data analytics and big data on DA based on previous works, we conducted a systematic review approach that consists of three steps: (1) collection of related work, (2) selection of relevant work, and (3) examination and analysis of the filtered related work.

In the first step, we performed keyword-based research and We gathered a large number of studies from well-known and popular online sources (Web of Science, Scopus, IEEEXplore, ACM, etc.). We used a combination of keywords from the two sets (Big data, data mining, data analytics, machine learning, Internet Of Things, sensors) and (Digital agriculture, smart farming, precision agriculture, agriculture, farming). We gathered more than 327 articles. In the next steps, We selected a small number of articles, which are considered relevant for further analysis, based on their ideas, methods, data types and sources, addressed problems, proposed solutions, tools used and quality of the results.

Through the literature analysis, the study aims to find responses to the following research questions and discuss findings in the following sections.

- What is the process of DA for crop management?
- What are the various data types generated by farms and used in DA applications for crop management?
- What role does big data analytics play in DA?
- How are big data analytics used for crop management?
- What are the influences of the farm's scale on the application of DA?
- How big is the data used in the proposed DA solutions'?
- What are the challenges facing the DA?
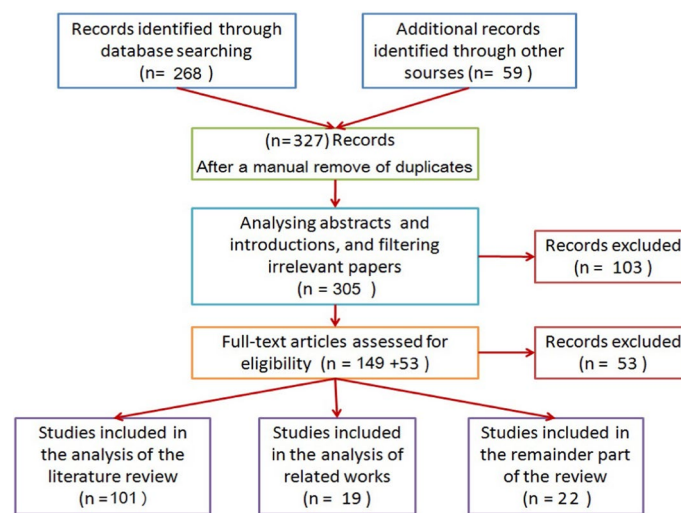
Fig. 3 The research methodology flowchart

Figure 3 summarises the overall approach, adopted from the PRISMA[2] flow diagram.

Related work

Despite that DA and Big Data being relatively recent research fields, their scientific literature is rich and covers several concepts. As DA is at the cross boundaries between agriculture and ICT, three major dimensions have emerged as of a very high importance; technology, social economics and ethics, and decision-making based on Machine Learning. The first dimension focuses on the use of advanced technologies to improve practices and productivity [56, 124]. In Ref. [124], the authors studied the impact of sensor networks in agriculture, including remote sensing technologies, wireless devices, and other IoT devices. Ref. [56] reviewed some developments in remote sensing within Big Data processing and management in agriculture. The second dimension concerns legal, ethics, social and economic factors of DA, to provide insights into the impact of digitised information and its analysis on the farm management; farmer identity, skills, privacy, production, and value chains in food systems [39, 70, 77, 78, 92, 113, 146, 148]. The third dimension focuses on the application of big data analysis and machine learning (ML), to optimise and forecast the production and the use of resources. In this paper, we only consider this dimension.

Various studies have been conducted on the application of data analytics to crop yield management. For instance, [71] presented a systematic review on crop yield prediction using ML techniques, and extracted major ML algorithms, features and evaluation metrics used in those studies. Ref.[35] discussed the yield estimation by integrating agrarian factors in ML techniques. This allowed them to show a strong relationship between crop yield and climatic factors. Ref. [103] Provided a systematic review on the use of computer vision and AI to enhance the grain quality of five crops (maize, rice, wheat, soybean and barley), disease detection and phenotyping. Ref. [64] reviewed the application

of big data analysis in some fields of agriculture. It highlighted solutions to some key well-known problems, used tools and algorithms, along with input datasets. The authors concluded that big data analytics in agriculture is still at its early stage, and many barriers need to be overcome, despite the availability of the data and tools to analyse it. To measure the level of usage of big data in DA, the authors defined big data metrics (low, medium, high) for each of its dimensions (volume, velocity, and variety). However, while it is a very simple model, it is not easy to specify thresholds, as some dimensions, such as volume and velocity depend on technological advances. Ref. [12] presented a review on the use of ML methods to detect biotic stress in crop protection. The authors analysed the potential of these techniques and their suitability to deal with crop protection from weeds, diseases and insects. In addition, they provided very good instructive examples from different fields of DA. An earlier similar study was presented in [89], where the authors studied four very popular learning approaches; Artificial Neural Network (ANN), Support Vector Machine (SVM), K-means, and K-Nearest Neighbour (KNN). Ref. [25] presented a survey on data mining clustering methods applied to food and agricultural domains. It first described major techniques of unsupervised classification, then it examined some existing techniques applied to agriculture products; like fruit classification, wine classification, analysis of remote sensing in forest images and machine vision.

This study is not just an update of previous surveys. The main objective is to examine the effectiveness of big data analytics in crop yield monitoring and discuss the challenges of such paradigm shift in the agriculture domain. Moreover, It is important to understand the sources of datasets, their types, and which ML techniques are more suitable to analyse them.

DA: it's all about data

Digital Agriculture (DA) relies heavily on the data sources and techniques used to collect it. This data is then organised in agricultural data warehouses and analysed [93]. The results of this data analysis provide significant insights to farmers and agronomists about how to improve the production, minimise the farming operational costs, manage risks, and protect the environment. The process of deploying DA is derived from data science.

Digital agriculture process

Figure 4, adopted from the knowledge pyramid DIKW, shows a data-driven process, which is at the heart of DA. This usually shows how data from past experiences and models serve as input to techniques of mining and analysis to help in future decisions and acting accordingly. The newly collected data will be used to further refine the process and adapt it to an ever-evolving agricultural world.

This is a data-driven methodology derived from the overall knowledge discovery process. The first phase, data collection, is crucial to the validity of the whole analysis. One needs to carefully identify the type of data that should be collected and the approach of gathering it and maintain it through its whole life cycle. This is even more complex in DA, as the data is issued from various and heterogeneous sources, and contains a number of factors of uncertainties. The second phase, data representation and analysis, is very sophisticated, as there is no common standards in the way the data should be
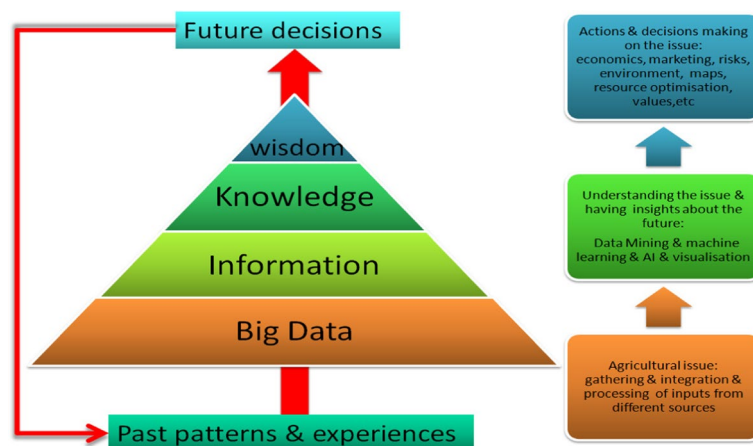
Fig. 4 DA– a data-driven process
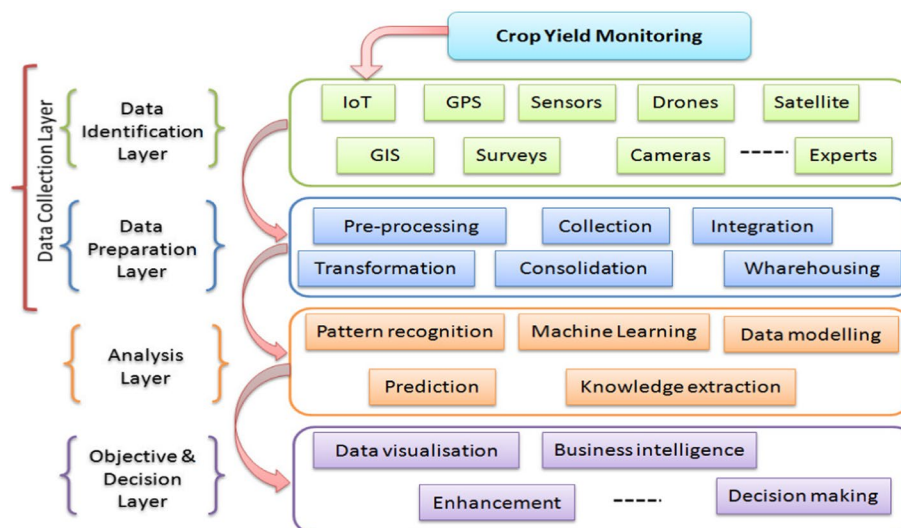
Fig. 5 Big Data Analytics system architecture for crop yield monitoring

integrated, consolidated, to derive a unified representation that is suitable for its analysis, and in the choice of the analysis techniques. Finally, the decision-making is a laborious task, where the extracted knowledge will be associated to the expertise of farmers and agronomists, farming constraints and regulations to derive new management processes with the view to improve productivity and quality of products, reduce and their impact on the environment. Figure 5 depicts a diagram presenting the DA process for crop yield monitoring, as explained below.

- *Data collection and preparation* It is important to identify the data types and attributes based on the problem at hand (e.g., crop management), and the level of granularity of the data. The required data sources should also be identified and assessed for their data quality. As mentioned above, the data is then prepared for analysis. This includes data integration, representation, selection, transformation, etc.

- *Data analysis* the complex nature of the agricultural data requires an elaborative analysis approach, ranging from methods of feature selection or extraction to various learning algorithms to discover models, patterns (or knowledge in general term) for data analysis. These will be evaluated against the expected quality of results and their suitability to a decision-making process.
- *Decision-making* The main goal of the DA process is the decision-making. Any decision should follow the state-of-the-art practice, be justifiable and scientifically sound.

Digital agriculture data

In agriculture, Very large amounts of data can be collected from various sources. These include sensors, weather stations, satellite imagery, drone imagery, and many other instruments. The datasets include weather data, farm records, environmental conditions, soil parameters (nutrients, texture, moisture, and so on. The data is usually rich, large, very complex, and heterogeneous. Therefore, its analysis is not straightforward.

The heterogeneity is not only expressed by the data types and formats, but it can be collected using different equipment of different quality. In addition, historical data may be described with different sets of attributes compared to very recent data. This can present inconsistencies in naming conventions and measures when the data is collected from different locations and times. Moreover, the data can be static and historical, which is considered as offline data, and can be online weather data collected at regular intervals (streams of data values), such as weather data (e.g., every 15 minutes), satellite imagery, which is characterised of being spatio-temporal, such as Geo-spatial data, Moderate-Resolution Imaging Spectroradiometer (MODIS) images, etc.

As mentioned earlier, the data collection is not well tackled in the literature. Most of the studies assume that the data is known already, and the experimental setup was already in place. Therefore, more effort is allocated to the data analysis and interpretation rather than on the complete environmental parameters and conditions. In the following sections, we discuss the data analysis process. This discussion is structured based on the main categories of the data analysis; classification, and clustering [24]. Note that, for high quality results, the data needs to be pre-processed, as discussed in the previous section. The pre-processing includes cleaning (dealing with missing values, redundant data, noise and outliers), data transformation, dimensionality or data reduction, and so on.

Classification for crop monitoring

Big Data analytics system architecture is depicted in Fig. 5. While this system is targeted specifically to crop yield management, it can be adapted to any data-driven application. This architecture implements faithfully what we have highlighted in the previous sections. In this section, we will focus on the data analysis layer of the architecture, moreover, we will pay attention to the data types and their sources, techniques of data acquisition, the learning algorithms. The main objective of the crop management data analysis is to get some insights about the crop monitoring problems and show the potential of DA through big data analytics, also called data mining. Data mining and its techniques are involved in several roles in crop production. Farmers may want to know the

future yield of their crop, specific areas of their farms suffer from the spread of weeds or under-nutrition. Researchers can look for information such as plant growth patterns, optimum growing conditions, best pest and disease control environment and so on. Data mining offers panoply of sophisticated techniques required to meet all of these needs.

There are two major categories of data analysis: Classification and Clustering. In the work of [24], authors studied applications of data mining techniques in crop management and proposed a classification of these applications. They found that the classification and clustering are the main used categories, where the classification includes prediction, detection, protection, and categorisation). The choice between classification or clustering analysis is very simple. If the models or classes we are looking for were known in advance and we have an annotated data to support the training of the learning algorithms, then classification is the right choice. However, the annotated data is not always available and easy to generate, and in many cases we do not know even which models or patterns we are looking for. In these situations, clustering analysis is the right alternative.

In this section, we focus on the studies that use classification methods for their data analysis. Clustering analysis will be covered in the next section. We structure these classification studies based on the application objectives or targets which are *categorisation*, *prediction*, *detection*, and *protection*.

Categorisation

While the classification main objective is to assign a given object into one of the predetermined classes, in the agricultural world, the use of classification process may vary depending on the stakeholders interests. In this study, we report four different applications (or targets) which are widely used in agriculture *categorisation*, *prediction*, *detection*, and *protection*.

Categorisation aims at defining the classes (or class labels) based on the simple recognition of similarities that exist across a set of entities. For example, categorisation can be used to classify small fruit from fruit with normal to big size, to make an estimation of yields; which may have an economic impact if the farmer wants to make different packages or prices for each type of fruit separately. It can also be used to classify damaged crops from good ones in order to estimate losses, or to prepare for the harvest and marketing. Categorisation can also be applied for crop mapping (e.g., poor, average, high yield), which aims to provide information on farmed fields given a specific type of crops, or to identify a type of crops that are more suitable for a particular field. Based on the input data, categorisation can help improve the farming operations based on the meaningful categories (classes) predefined in advance.

Producing accurate crop maps is essential for effective agricultural monitoring [131]. Categorisation approaches can be applied to study regional crop distribution within or post growing season. For this purpose, it can offer:

- A good understanding of how crops are distributed at early stage of their development; allowing for an opportune decision making and management, as well as adjusting crop planting structure, is crucial. Besides, the timely available of (spatial

or maps) distribution of crop types is required for statistical and economic purposes [131].

- The availability of crops maps is critical for the diverse agricultural monitoring activities, such as crop acreage estimation, yield modelling, harvest operation schedules [131, 144], etc.

Moreover, categorisation has been applied for agricultural field mapping [31], to quantify the cropping intensity for small-scale farms [58], to identify and map crops and to retrieve the area of major cultivation [100] and to classify land-cover and crop [76]. Table 1 highlights the major fields, ideas and tools used for crops categorisation. We can see that data issued from satellites and remote sensing, and the features with vegetation indices especially NDVI and EVI, the RGB colours, are the most used.

Crop yield prediction

The estimation of crop yield is crucial in DA, as it enables efficient planning of resources. Economically, an early and accurate prediction of yields can help decision-makers to react to the crops market. Moreover, crop yield prediction permits the study of factors that influence and affect the production, such as climate and weather, natural soil fertility and its physical structure and topography, crop stress, the incidence of pests and diseases, etc.

The prediction of crop yields has been the subject of many studies. Ref. [71] presented a literature review on crop forecasting, where the authors highlighted the most used machine learning algorithms along with the applied metrics and measures. In this section, we examine the learning algorithms that have been used in crop yield prediction from different views: data types, the pre-processing methods, and features or the predictor variables used in each study. Tables 2 and 3 summarise some relevant studies.

The crop yield forecasting approaches follow two major types of sources of data. The first type is related to the sources that have direct impact on the crops. These sources are soil data, weather data, environmental parameter data. These are usually used to predict crop yield [27, 34, 42, 46, 51, 73]. The second type of sources are the use of advanced technologies and tools like satellite multi/hyper spectral images, remote sensing and sensors to collect the data [62, 83, 102, 114, 152]. Some advanced studies use both types of data sources [1, 40, 54, 59, 65, 67, 68, 97, 120, 121].

The forecasting models based on the first type of data sources provides a pre-season estimation of the yield, even before the beginning of the crop season. This allows farmers to decide which strategy to both optimise the farming operations and crop production. These decisions include choosing seeds and crop type, type of fertiliser and its applications. Moreover, This data can also be used for some crop monitoring during the growing season.

The monitoring systems based on the second type of data analysis - data imagery obtained from satellite, cameras, scanner, sensors - allow for on-season estimation (emergence, detect stress conditions of crop, harvest dates, ...). These models are complex since they have to analyse the data that consists of both spatio-temporal and non-spacial. While the spatial data is of high resolution, some images can be of very poor

Table 1 An analytical study on examples of crop categorisation approaches; demonstrates: the type of categorisation application, the used learning algorithm, the data type, data pre-processing and selected features for each algorithm

References	Application	Algo	Target	Data type	Data pre-processing	Extracted features
[31]	Crop fields mapping	RF	/	Satellite DigitalGlobe World-view-2	Hand digitisation	Randomised Quasi-Exhaustive features
[41]	Crop mapping	Decision tree	Soybean	Satellite	Multi resolution segmentation	NDVI NIR (near infrared) SWIR (short wave infrared)
[131]	Crops mapping	RF	/	GF-1 WFV sensor satellite images	Multi-resolution segmentation	temporal, spectral textural features vegetation indexes(NDVI, EVI...)
[22]	Crop fields mapping	RF	Paddy rice	Satellite images	Polarisation for cloud contamination by Google Earth Engine	NDVI, EVI land surface water index LSWI
[115]	Crop mapping	Deep learning: autoencoder CNN, Full CNN	Soybean, maize cotton	Satellite images	Data were pre-processed	Texture pixel's features based on the image patch
[157]	Crop classification	LSTM	/	Satellite & optical images field surveys	Segmentaion pan-sharpening and mosaic of optical imag-esthermal noise removal and radiometric correction	Spatial features
[76]	Crop classification	Deep learning CNN	Wheat, maize sunflower soy-beans, sugar beet	Satellite images	segmentation and data restoration using unsupervised NN self-organising Kohonen maps)	Spectral and spatial features
[33]	Plant classification	Deep learning CNN	22 plants	Camera and cell phone images	Data are not pre-processed	Self-learned features
[26]	Crop classification	Ensemble learning ANN, DT, SVM	Rice, soybean, corn cotton	Remote sensing images	USGS online system, used a cubic convolution 245 re-sampling and a standard terrain correction incorporating ground truth points	NDVI, levels of greenmoisture
[100]	Crop classification	set of classifiers SVM(RBF kernel), RF, Spectral Angle Mapper	Tree crops, sugar beet alfalfa, cereals	Sensor satellite Time series and images	Atmospheric correction and Radiometric calibration and Pan-Sharpening	NDVI

Table 2 Part 1: an analytical study on examples of crop prediction methods; highlights: the applied learning algorithms, the crop type, data type and pre-processing, the other studied and considered parameters in each proposed approach and the predictor variables for each used algorithm

References	Algorithm	Plant	Data type	Data pre-processing	Other parameters	Predictor variables
[46]	KNN, MLP, SVR regression trees	Pepper, bean chickpea, corn potato, tomato Mexican husk tomato	Historical yield climate irrigation plans	/	/	Planting area min, max and avg temperature precipitation, irrigation solor radiation
[51]	RF	Groundnut millet	Historical yield	KNN for data imputation	/	Sunlight, humidity precipitation min, max, avg temperature
[54]	SVM, RF Gaussian process regression	Winter wheat	Remote sensing climate, soil yield, crop map	Google earth engine (GEE)	Regional differences of yield-variable importance	Min, max temperature NDVI, EVI palmer drought severity index, precipitation soil: moisture, physical and chemical properties
[121]	LSTM	Soybean	Satellite images weather and historical yield	GEE	/	NDVI, EVI, land surface and air temperature precipitation
[120]	RF	Corn, soybean	Historical yield satellite remote sensors ancillary and environment	/	/	Dynamic ranged vegetation index (WDRVI), temperature precipitation, soil moisture shortwave radiation statistics related to county-level irrigated harvested cropland
[83]	DNN	Soybean	Multi remote sensing data	Pix4D mapper softwrare: UAI RGB, multi spectral and TIR images conversion of radio-metric value	/	25 features: canopy spectral structure, thermal and texture features NIR, NDVI, WDRVI, EVI
[1]	SVR	Potato	proximal sensing (soil and crop properties) yield data	Effects of data-set size on accuracy	/	Soil electrical conductivity soil moisture, soil slope, soil chemistry NDVI
[65]	SVR	Wheat	Satellite images climate and yield records and maps	KNN	/	NDVI, precipitation max temperature
[40]	RF	Wheat, barley canola	Yield, soil, climate remote sensing Geo-physical data	/	Pre-sowing mid and late season	Soil maps, surveys rainfull, NDVI

Table 2 (continued)

References	Algorithm	Plant	Data type	Data pre-processing	Other parameters	Predictor variables
[42]	RF	Mango	Irrigation, historical yield	/	Different irrigation regimes	/
[34]	ANN	Tomato	Historical data	/	Water monitoring different radiations values	CO_2, day, water radiation, temperature
[97]	RNN	Soybean maize	Multi sources: satellite, soil properties	/	Pre-season yield	Min and max temperature precipitation soil, pH and other 10 features
[59]	RF	Wheat, maize potato	Multi sources: climate, soil photo-period water, yields fertilisation	/	Climate and biophysical variables at global and regional scales	Many features on climate and soil and nitrogen fertiliser
[73]	ELM	Robusta coffee	Soil components	/	Soil fertility	Exchangeable calcium boron, magnesium and nitrogen, PH Zinc potassium, sulphur phosphorus
[102]	ANN supervised kohonen counter propagation XY-fusion	Wheat	Multi-spectral satellite data	Orthorectification in-band reflectance calibration	Physico-chemical soil parameters	NDVI
[114]	non-linear regression	Cabbage	Sensor data	/	Nitrogen variation	NDVI

Table 3 Part 2: an analytical study on examples of crop prediction methods; highlights: the applied learning algorithms, the crop type, data type and pre-processing, the other studied and considered parameters in each proposed approach and the predictor variables for each used algorithm

References	Algorithm	Plant	Data type	Data pre/processing	Other parameters	Predictor variables
[152]	CNN, LSTM Gaussian	Soybean	Sensor MODIS data	Transform multi-spectral images to individual histograms	/	Histogramsgeographic location year
[62]	Regression Rulequest Cubist	Corn, soybean	Satellite data	8-days periods data-points composed then averaged	Pre and within season	NDVI, precipitation day and night land surface temperature
[67]	DT (Extremely randomised) SVM(RBF);DNN	Corn	Satellite images climate, yield	/	Seasonal sensitivities	NDVI and many other features
[27]	Deep learning semi-parametric NN for training; bayesian hyper-parameter optimisation and early stopping	Corn	Historical yield weather	/	Climate change impact and semi-parametric prediction model	Precipitation temperature humidity, wind radiation Latitude and longitude Growing degree, soil County, irrigation
[68]	DNN; ANN; RF; multivariate adaptive regression splines SVM; extremely randomised trees	Corn and soybean	Satellite images MODIS historical yield meteorological and crop landhydrological	/	Effect of phenology	EVI Leaf Area Index Gross Primary Production precipitation; min, max air temperature soil Moisture, NDVI
[50]	SVM	Rice	Climate and geographical data	/	Effect of phenology and climate pre-season	Mean, max, mintemperature Daily Sunshine hours precipitation mean relative humidity min relative humidity mean wind speed maximum wind speed
[60]	LR	Corn	MODIS remote sensing	Savitzky-Golay filter for smoothing NDVI time series	Effect of phenology	Max correlation NDVI crop growth rate crop growth days
[111]	RF	Chickpea	Modis images weather data yield statistics remote sensing	/	Drylands sensitivity to data time	EVI, NDVI, Leaf Area Index precipitation and 5 other features
[4]	Bidirectional LSTM	Tomato, Potato	Climate datairrigation scheduling soil water content	Moving average method for data imputation multi-collinear parametersremoval	Effect of irrigation scheduling	Min, max, mean temperature min, max, average relative humidity average solar radiation min and average wind speed precipitation
[91]	3D-CNN	Wheat, Barley, Oats	Weather data UAV RGB image yield data	Images resizing	Effects of timeefficiency of using time series data vs point-in-time data	RGB Images, cumulative temperature
[19]	LSTM	Winter, wheat	Climate satellite data soil surveys	/	/	Min and max temperatureprecipitation, EVI, soil depth and texture, pH geographic properties

quality, (e.g., images with lot of clouds). Features or predictor variables used in this kind of applications depend on the type of data sources, NDVI and EVI are the most used vegetation indices for satellite and remote/approximate sensors' data source, min/max temperature and precipitation for weather data source, soil moisture and nitrogen fertiliser for soil based data source.

Crop protection

Crop disease is considered as a major menace for food security in many regions of the world since it causes serious crops losses. While the detection of crop diseases correctly and timely when they first appear is crucial in crop monitoring, this remains a difficult task. One of the solutions to deal with this issue is to use data analytics approach. This will reduce yield losses and prevent farmers to take effective reactive actions. Forewarning can be seen as the outputs of data mining process. Usually, this consists of examining the features of a newly presented case and assigning it to a predefined class.

Several interesting efforts have been developed to prevent crops losses due to diseases, Tables 4 and 5 summarise some major studies. Ref. [7] presented an overview of ML techniques for crop disease classification. In addition, it presented to a case study where a deep learning algorithm was successfully used. Ref. [45] provided a review on advanced ANN techniques to process hyper-spectral data for plant disease detection. Recently, deep learning approaches have been emerged and widely used for plant disease detection and classification, with a variety of network architectures (CNN, AlexNet, googLeNet, CaffeNet, DenseNet, Inception, LeNet, VGGNet,...) and training methods (shallow, deep, from scratch) [9, 16, 21, 28, 38, 63, 79, 82, 125, 139, 143, 150, 155]. Moreover, [127] presented an interesting study on the potential of the use of deep learning for plant stress phenotyping.

Crop protection, that consists of disease, stress, and weed detection, aims to offer tools that detect plants disease caused by various biotic (pathogen, insect, pest, and weed) or abiotic (temperature stress, nutrient deficiency, toxicity, herbicide) variables [126]. The earlier the stress, disease or their symptoms are detected, the greater the chance of reducing the disease spread within a field. This has gained significant advantage from the advances in image collection and processing and their analysis using ML algorithm. The state-of-the-art is very rich. The large majority of studies carried out so far were using image processing, consequently image-based data and classification techniques. These are capable of detecting disease at the scale of leaf, canopy or field [126].

Disease detection at a leaf level uses images collected using digital cameras, which are stored in data warehouses. For instance, *PlantVillage* database [6, 9, 21, 28, 63, 79, 88, 106, 125, 129, 150] is created for this purpose. The objective of this repository is to build classifiers with high accuracy. The basic classifiers can simply assign to an unseen image a label *healthy* or *infected*, while more elaborated classifiers can identify the disease - in other words, classify unseen images to disease classes. However, this approach has some limitations. First, it depends on the quality of the images, as when taken in natural environment, these images are subject to different degrees of light, shadow, dust and leaves overlapping and requires sophisticated image processing, which is not an easy task. Second, usually the datasets sizes are small, which affect the learning phase of the classifiers

Table 4 Part 1: an analytical study on examples of crop diseases protection and weeds detection approaches; highlights the applied algorithm, plant and data type, data pre-processing and the extracted features

References	Application	Algorithm	Plant	Data type	Data pre-processing	Extracted features
[21]	Leaf disease recognition	CNN	Tea plant	Digital images	Data augmentation	/
[63]	Plant disease detection	CNN	Plant leaves	Digital images	Data augmentation	/
[9]	Plant disease detection	DNN	Plant leaves	Digital images	Data augmentation	
[143]	Light leaf spot detection	SVM	Soil seed rape	Multi-spectral images	Removal of:background and redundant features	Carter Index 1 light leaf spot index Spectral signature
[82]	Disease detection	SVM	Wheat	Sensing data	/	NDVI, Photochemical reflection index Pigment-specific simple ratio, water index
[155]	Disease detection	CNN	Wheat	Remote sensing data	Segmentaion: sliding-window	3D blocks
[125]	Leaf disease recognition	CNN	Maize	Digital images	/	RGB
[15]	Plant disease detection	Gated recurrent unit CNN	Soybean	Satellite images Crop rotation	Time-series	Spectral bands of: red, green, blue NIR, NDVI
[150]	Leaf disease detection	CNN	Tomato	Digital images	Data augmentation:resolution reducingbicubic method to enlarge images	Patches
[136]	Plant disease detection	RF	Wheat	Aerial multi-spectral images	/	RVI, NDVI, OSAVI NIR, Red
[49]	Plant disease detection	Partial least squares regression	Wheat	Aerial hyper-spectral images	Image fusion and mosaicking	Disease index many vegetation indexes texture features
[153]	Plant detection	CNN	Maize	Aerial RGB images	Segmentation by RF	Morphological caracteristics of maize tassels
[80]	Plant disease detection	ANN	Wheat	Hyper-spectral aerial images	Fusion and stitchingradiometric calibrationatmospheric correction	11 Vegetation indexes spectral bands texture features
[88]	Plant disease detection	deep learning CNN: AlexNet GoogLeNet	Plant leaves	Digital images	Coloured, gray-scaled, segmented	/

Table 4 (continued)

References	Application	Algorithm	Plant	Data type	Data pre-processing	Extracted features
[2]	Crop and weed classification	SVM RBF kernel	Chilli, Pigweed Marsh herb Lamb's quarters Cogongrass, cucumber	Digital images	Segmentation: binarisation technique: -global threshold noise removal morphological opening and morphological closing	14 features: RGB colours, shape features Moment invariant features
[44]	Weeds detection	RF	Maize	Hyper-spectral images	Segmentation	185 features Ratio Vegetation Index,NDVI
[3]	Weeds detection	SVM Gaussian kernel	Corn leaves and broad silver beet leaves	Spectral reflectance images	/	NDVI
[8]	Crop and weed classification	ANN Generalised Softmax Perceptron and the Posterior Probability Model Selection algorithm	Sunflower	Digital images	Special process of segmentation	13 morphological features: Number of boundary pixelsCompactness, Perimeter, Centroid and Elongation,The geometric centre Area, Number of pixels of objects, Major and minor axis of the best fit ellipse

Table 5 Part 2: an analytical study on examples of crop diseases protection and weeds detection approaches; highlights the applied algorithm, plant and data type, data pre-processing and the extracted features

References	Application	Algorithm	Plant	Data type	Data pre-processing	Extracted features
[129]	Leaf disease detection	Deep learning CNN: CaffeNet	13 plants leaves	Images	augmentation by: afine transformation and perspective transformation and rotation manually pre-processing by image cropping and labelling	
[119] [29]	Weeds classification	ANN;PSO and bee for optimisation	Potatoes rice	Stereo video	Segmentation	Color features & vegetation indices
[28]	Mid-late season weed detection	CNN	Soybean	Aerial images	Overlapped images removal dimenssion reduction annotation	Patches
[118]	Weed detection	DNN	Sugar beet and weeds	Multi-spectral UAV	Segmentation	RGB Color-Infrared NDVI
[106]	Leaf disease detection	RF, SVM, KNN	Alfalfa	Digital images	Lesion: artificial cutting-segmentation:12 lesion segmentation with K-median clustering and linear discriminant analysis	129 texture colour and shape
[55]	Seeds disease detection	ANN	Orchids	Digital images	Segmentation: an exponential transform with an adjustable parameter	Texture and colours
[109]	Plant disease detection	RF	Soybean	Satellite images Crop rotation	Geometric distortions removal radiometrically and sensor correction image rotation	Spectral bands of:red, green, blueNDVI, NIR
[28]	Leaves disease detection	Transfer learning CNN: abstraction level fusion	Olive	Digital images	segmentation: automatic cropping: Otsu's algorithm	Edge magnitudes: Gray-scaled-Shape features: area, perimeter
[6]	10 leaves disease detection	Transfer learning CNN	Eggplant, hyacinth beans ladies finger, lime	Digital images	Segmentation data augmentation	/

Table 5 (continued)

References	Application	Algorithm	Plant	Data type	Data pre-processing	Extracted features
[79]	Leaves disease detection	Deep learning: Alex NetGoogLeNet	Apple	Digital images	No pre-processing AlexNet Precursor for features maps max-pooling for GoogLeNet for features extractiondata augmentaion:light disturbance &rotationnoise removal	/
[81]	Plant disease detection	KNN	Wheat	Satellite imagesfield survey	Radiometric calibration atmospheric correction	Red and green bands NIR, vegetation indices:disease water stress index optimised soil adjusted vegetation index shortwave infrared water stress index triangular vegetation index and others
[156]	Plant disease detection	RF	Wheat	Satellite images field canopy hyperspectral	Noise removal image mosaicking Atmospheric correction spatial resolution re-sampling	Disease indexNDVi, EVI and others

and more importantly the potential of some advanced learning algorithms such as deep learning. Data augmentation (rotation, light shade's variation, colour inversion, translation and changes in intensity and so on) is one of the methods used to overcome this problem to artificially increase the number of images [6, 9, 21, 63, 79, 129, 150], but it does not always work. Transfer learning is another solution to scarce/small data-set, where the knowledge obtained from solving a task in a given domain is transferred to the target domain in which the dataset is small [6, 11, 28]. The transfer learning can only be efficient if the source and target domains share some similarities in terms of diseases and their symptoms, for example. Moreover, it is very challenging to transfer knowledge from representations learned using RGB images to a target task using multi-spectral images from UAV or satellite [126].

Third, this approach cannot detect more than a single disease at a time, and the detection of diseases if the symptoms are manifested in another area than leaves. Plant canopy based-image was proposed as a solution to this problem. The idea is to collect data relative to disease in situations where single-leaf phenotypes alone would not provide sufficient information. Such features include the size, the height, the structure, and branching of canopy [126]. The canopy-based detection uses UAV equipped with (multi/ hyper) spectral cameras and sensors to collect the data [32, 49, 80, 82, 136, 143, 153, 155]. Then data needs to be processed to extract features which are usually related to vegetation indices like NDVI and EVI or colours like RGB and NIR. The benefit from UAV images comes with cost on complexity of analysis since images taken by UAV are susceptible to occlusion, overlapping, and atmospheric effects. Also, UAV is not able to fly at higher altitudes, which decreases the quality of the collected images. To cover larger zones and fields, satellite-based remote sensing and images has been proposed as a very good alternative [15, 81, 109, 156]. However, the problem with satellite remote sensing is the revisit time, which is 16 days on average, which makes protection applications difficult, and some diseases can spread rapidly in fields before they are detected. Moreover, passive sensors cannot penetrate clouds [149]. The integration of these data with additional data sources like field surveys, contextual information of field and crop rotation can improve the accuracy [15, 81, 109].

Detecting diseases only from one data source based on digital images or sensor data is not sufficient. Besides, variations in symptoms may lead to false positives due to dynamic nature of plant changes [126]. Consequently, the appearance-based identification of diseases is not reliable enough to accurately detect unhealthy plants, especially in the early growth stages. The use of multi-data sources can improve the accuracy of the detection. For instance, the use of physiological features and morphological characteristics (growth attributes, yield-related features, soil) [66], or the employment of satellite-based images and canopy-based images [156], where the disease can be identified at the plant canopy level and at the field level.

Crop maturity monitoring

Crop maturity is a kind of crop yield prediction, but it is based on image data. This technique has been used in fruit detection, like apples, tomatoes, oranges, etc, and provides an early estimation of yield. It is also used for crop monitoring to provide information to farmers with the view to plan their farming operations, adjust management practices

before harvesting, etc. Such intelligent systems for monitoring crop implement the data mining process incorporating machine vision and image processing methods among with advanced learning algorithms, such as CNN, SVM and ANN. Unlike crop yield prediction process described above, this process is based on a single-data source; digital images [5, 23, 52, 75, 108, 122] or sensor based-images [117, 123, 153]. Table 6 summarises such techniques. The challenges of these systems are more or less the same as those of systems for crop disease detection and protection. For instance, images with different illumination and lighting angles, complex surroundings and backgrounds, noise, the presence of clouds, etc.

Clustering for crop monitoring

Clustering techniques are not widely employed in DA, few efforts have been deployed to investigate the potential of these techniques for zones' delineation within a field. There are several reasons for splitting an agricultural field into zones. Some traditional reasons include crop diversification within a field, crop-rotation, facilitating the management tasks, and more recently we defined the zones based on yield maps. This usually helps to improve the overall crop yield of the field, by managing the zones more effectively. Therefore, delineation of Management zones (DMZ) is a very important task for farming operations since determining zones of low-or-high yields, and understanding the reasons behind low yields, can help come up with specific solution for each zone with the view to increase the yields. In addition, it has other economic benefits, because we can target each zone with the right amount of fertilisers, water, and other nutrients.

According to [69], delineation of management zones is an effective way to manage the variability of soil within a field, such that each zone will receive specific management. In [145], a management zone is defined as a subregion of a field that has a relatively homogeneous combination of yield-limiting factors, for which a single rate of a specific crop input is appropriate to reach maximum efficiency of farm inputs. In [53], it is defined as a subregion of a field that is relatively homogeneous with regard to soil attributes.

DMZ is a complex spatial problem, which is addressed in the literature from several perspectives. This has attracted interest from many researchers [61, 85, 87, 110, 140]. A literature review has been presented in [90], where the authors discussed the delineation of soil management zones from the variable-rate fertilisation point of view. many other studies presented the delineation based on various criteria. Some techniques that have been used include topographic maps, direct soil sampling, non-invasive soil sampling by electrical conductivity equipment, soil organic matter or organic estimated by remote sensing, and yield maps built using data collected over several seasons/years [99].

Figure 6 depicts the general process of delineation of management zones designed according to methodologies followed by the majority of the literature.

The majority of problems that are related to crop management imply the management of fields and zones. Therefore, the collected data is usually characterised by geographic coordinates and time associated with each sample, which leads to the use of data mining techniques that are more suitable for spatial and temporal datasets. It is well recognised that agricultural datasets are typically spatio-temporal, as the data is always associated with location and time. However, these datasets contain a significant amount of noise, outliers, and even missing values. For instance, GPS capture devices introduce

Table 6 An analytical study on examples of crop maturity monitoring (fruits detection and counting) approaches; highlights the applied algorithm, plant and data type, data pre-processing and the extracted features

References	Application	Algorithm	Plant	Data type	Data pre-processing	Extracted features
[123]	Fruit detection	EM	Tomato	High spacial resolution sensor images	Noise and stalks removing spacial segmentation	Shape and size
[23]	Fruit detection	ANN	Apple fruit and tree canopy	Digital images	Segmentation	Area of fruit-sarea of small fruits cross-sectional area of foliage fruit number total cross-section total cross-sectional
[108]	Fruit detection	SVM	Coffee	Digital images	Segmentation: homogeneous information	42 colours features
[5]	Fruit detection	BC Gaussian	Cherry	Digital images	Segmentation: enhancements and specular reflections removing by inward interpolation method	Colours features:RGB
[52]	Fruit detection &classification	CNN	Strawberries	Digital images	Hand marking regions of interests	/
[75]	Immature fruit detection	ANN	Peach	Digital images	Hue-Saturation-Intensity for illumination enhancementpixels' normalisation histogram equalisation reconstruction of images backgoud elimination	Texture features
[117]	Fruit counting	CNN	Sweet pepperrock melon strawberry apple, avocado mango, orange	Multi-spectral images(RGB,NIR)	Pixel-wise segmentaion bounding box annotation	Colour and texture features
[122]	Immature fruitcounting	SVM	Green citrus	Digital images	Images conversion from RGB to graycircular Hough transform	13 texture features
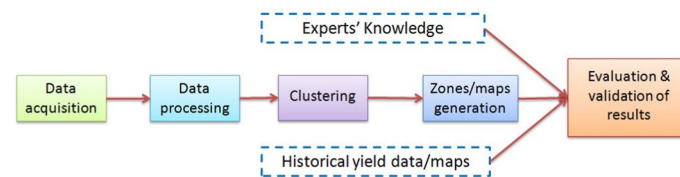
Fig. 6 The delineation management zones process

some noise, imprecisions, and even outliers in the data. Satellite imagery also faces huge imprecision and noise (such as clouds, ...).

Because of the type of the datasets, which is spatio-temporal, it is not surprising to notice that the majority of the clustering algorithms used are of type partitional. K-means and Fuzzy C-Mean (FCM) are considered among the most popular clustering techniques and heavily used to cluster agricultural data [17, 18, 84, 134, 137, 142, 151, 154]. The FCM approach has an advantage over K-means, as it deals better with imprecision and noisy data. Moreover, other types of clustering algorithms have also been proven to be efficient in DA, such as density-based and hierarchical-based clustering techniques applied to DMZ [48, 116].

As mentioned above, besides its huge importance in crop management, delineation of management zone (DMZ) has received much attention, as the data is now available not only from traditional sources but also from refined sources, including advanced data pre-processing techniques. In addition, the recently collected data integrates knowledge of experts and farmers experiences on their fields, which improves significantly the quality of the data [84, 141]. Advanced imaging enhancement techniques improve further the data quality, and they offer the ability to track the development of crops and provide a Geo-referenced data that can describe the spatial and the temporal variability of soil and crops variables at high resolution, covering large areas [17, 84, 101, 132, 133, 141, 151].

Systematic analysis

In the following we will explore the application of data analytics in DA and its extension to big data, and illustrate the practical challenges that hinder the full adoption of DA by farmers.

DA in (small /large) scale farming

Farming can be carried out on a small or large-scale fields depending on several factors like land size, capital, farmer skills, level of use of machinery and technology, etc. According to FAO[3] and Grain[4], over 90% of all farms worldwide are of small-scale holding on average 2.2 hectares (from 0.6 to 10 hectares), except for Northern America where small farms have an average size of 67.7 hectares[5]. Small-scale farms represent 25% of the world's farmland today, where 73.12% are located in developing countries.

---

[3] http://faostat3.fao.org/faostat-gateway/go/to/home/.

[4] https://grain.org.

[5] According to the criterion put forward by Lincoln University in Nebraska, which defines a small farm in the US as one with an annual turnover of less than US$50,000)

In [10] the authors described three categories of smart farming technology, which are complementary:

- Data acquisition technologies: they are used to acquire the data that is related to the farm. These include remote sensing, weather data, etc.;
- Data analysis and evaluation technologies: these technologies usually take as input the data that has been collected so far and deliver insight to the farmer. These include computer-based visualisation and decision models, farm management and information systems;
- Precision application technologies: these are focusing on variable-rate application and guidance technologies.

The application of smart technologies and data analytics for crop management are not restricted to one kind of farm. Nowadays, every farm should adopt smart technologies, as they are needed for variable rates applications (irrigation, pesticides, fertilisers) [72, 102, 154] while protecting the environment.

The size of the farm determines how these technologies will be used. Large farms tend to develop their smart technology to monitor their farming land, or to afford some of the existing sophisticated systems like CropX as they hold the scale and margins. While small farms tend to rent sophisticated machinery and smart applications on demand, especially with the proliferation of cloud technologies that makes these smart applications reasonable, the work of [30] is an example among others, of a smart irrigation system designed for smallholders. Besides, some technologies are more suitable for large-scale farms like drones and aerial vehicles used to monitor crops which are not as profitable or efficient for small scales because they have less difficulty visualising their crops. On the other side, large-scale farms are responsible for 70% of current deforestation[6], the largest share of agriculture-related greenhouse-gazes emissions, agricultural water use and habitat disruption resulting in biodiversity loss. Generally, small-scale farms require considerably fewer external inputs and cause minor damage to the environment.

Table 7 summarises the main differences between small and large-scale farming from several perspectives. However, DA can be applied to any kind of farm without restriction. Yet, we have found that the number of papers that addressed large-scale farms is almost the same as works on large-scale farms.

Technologies for data acquisition Table 7 can be used to all types of farms, such as remote sensing, imagery data systems, and so on. The acquired data, over the years, can lead to the phenomenon of Big Data. If pre-processed and stored properly, this will give a significant competitive advantage to farms that collected them, whether they are small or large. Some of the applications and data analyses that can be performed of the collected are summarised in the Tables 1, 2, 3, 4 and 5, 6.

DA and big data

Big data is not just characterised by the volume, but also by velocity, variety, and others [86]. These are enough to challenge the existing data mining techniques, as trying

---

[6] IPBES, 2019: Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.

Table 7 Comparison between small-scale and large-scale farming

Basis for comparison	Small-scale farming	Large-scale farming	Basis for comparison	Small-scale farming	Large-scale farming
% of all farms	92.3%	25%	Application of modern smart technologies	Yes	Yes
% of world's farmland	7.7%	75%	Mapping technologies	Yes	Yes
Budget (investment)	Low to medium	High	Data acquisition (cameras surveying, sensing and navigation)	Yes	Yes
Machinery vs labour	Labour-based	Machinery-based	Variable rate application	Highly used: Irrigation pesticides fertilisation seeding for some crops	Low to medium use: Irrigation fertilisation pesticides
Target crops and cropping system	Humain Food Backyard Arable crops Forage crops Vineyards Field vegetables Orchards	Commercial crops plants grown for animal feed or biofuels wood products other non-food crops	Navigation systems usability (GPS, INS,..) Inertial Navigation system (INS)+ Unmanned vehicles and drones	Low to medium (73.12% of farms situated on developing countries)	High
Farming method	Extensive	intensive	Applications	All applications	All applications
Contributors to agricultural information and knowledge	Low	Medium to high	Users of agricultural information and knowledge	Low	Medium to high
Effective climate change capacity measures and adaptation	Environmentally friendly	Environmentally not friendly	Sustainability	yes	No
Scope	Production for local communities	Corporate farms Factory farms: profit and business oriented plan exportation	Examples of references	[114, 123, 136, 132, 137]	[26, 43, 62, 67, 100, 102, 154]

to develop techniques to deal with large volumes of data (volume), various types of data attributes (variety or heterogeneity), and be able to analyse the new data as soon as it is collected (velocity) are extremely challenging tasks. Moreover, many other characteristics can be found in some big data-driven applications, these include veracity, value, viscosity, veracity, visualisation, etc. In this study, we added veracity, as the data, collected by various instruments and sensors, is of different quality, which creates a huge challenge to the data pre-processing task, and therefore its analysis. In the following, we discuss the impact of Big Data challenges on DA.

- Velocity: many studies that have been examined do not consider the data velocity during their data collection. In DA, the frequency of collecting data depends on its source and the problem for which the data was collected. Some applications need real-time data and others do not. For instance, crop yield prediction does not need real-time data or data streams. It is performed at ad-hoc, while crop protection and disease detection require high quality sensors and imagery data connected to efficient methods of data analysis, which need continuous control.
- Variety: this is very common in agricultural datasets, as multiple sources were used to collect all the necessary information about the farm and farming operations. The data values can be a simple number such as temperatures to more elaborated such as imagery data, NDVI, soil texture, etc. This makes the definition of distance measures and other parameters of the learning algorithms very difficult.
- Veracity: Agricultural data contains many missing values and collected from various sources of varying quality. The data is very noisy, and more importantly it contains many missing values. Therefore, it is very challenging to clean and prepare it for the analysis. This was the case in the work conducted by [37], and also in [93–96, 107] where data was collected from very large farming areas.

Table 8, summarises a set of representative papers reported in the paper according to their usage of big data. For each paper, we identify the type, the size, the heterogeneity of data used, and the frequency of its collection. Also, we consider the number and type of ML algorithms used, the complexity of the proposed analysis algorithms and devices used to collect data, data analysis applied to a given crop and problem to solve. One can notice while the data analysis algorithms and techniques were heavily used and varied, the rigorous process of knowledge discovery was not followed, usually the data is relatively small either in size (small observations) or the data has few dimensions (for instance, considering only weather data, or fertiliser, without taking into account other factors).

From Table 8, we can extract three classes of applications according to their usage of big data: Full usage (the data contains all the characteristics of big data), light usage (the data contains some characteristics), non-usage (the data does not contain any characteristic of big data).

To examine the degree of use of the big data concept and to figure out which of its dimension is more present, we conducted a statistical study where we classify works according to their employment of the 4Vs of big data.

Table 8 DA applications and their usage of big data concepts

| References	V1	V2	V3	V4	ML	Complexity	Device	Task
[2]	224 images	/	No	Image Data digital images	SVM	$O(n^2 * P + n^3) + O(n_{sv} * P)$	Digital camera	Classification
[18]	3*2 years of data monitoring	1 year	/	Sensor data: soil properties	FCM	time: $O(n * d * c^2 * i)$ space:$O(n * d + n * c)$	Pressure-based AgLeader	Clustering
[26]	/	/	No	Satellite data: Images in GeoTiff	EnsembleLearning (DT+ SVM+ ANN)	$O(n^2 * P) + O(P) + O(n^2 * P + n^3) + O(n_{sv} * P) + O(ep * n(nl_1 * nl_2 + nl_2 * nl_3 + ...) + O(P * nl_1 + nl_1 * nl_2 + nl_2 * nl_3 + ...)$	Satellite	Classification
[38]	87.8K	/	Yes	Image data: Open database images	CNN	$O(T * Q * t * q)$	/	Classification
[40]	/	/	Yes	All data types: yield climate informationsoil Geo-physical NDVI Remote sensed	RF	$O(n^2 * P * n_{trees}) + O(P * n_{trees})$	Yield monitor soil-maps, EM gamma survey MODIS	Prediction
[46]	6217	/	/	Historicalsensor data: Yield climate	SVR, KNN, ANN	$O(n^2 P + n^3) + O(n_{sv} * P); O(n * P); O(ep * n(nl_1 * nl_2 + nl_2 * nl_3 + ...) + O(P * nl_1 + nl_1 * nl_2 + nl_2 * nl_3 + ...)$	Spriter-GIS system	Prediction
[57]	229	1 year	Yes	Historical data: Crop yield	K-means	$O(n * c * d * i)$	/	Clustering
[59]	Precipitation: 47554 min/ max temperature:24542 mean temperature:14835	/	Yes	Sensor data: Crop yield soil Biophysical climate water photo-period, fertilisation	RF	$O(n^2 * P * n_{trees}) + O(P * n_{trees})$	/	Prediction
[33]	10413	/	/	Image data: Digital images	CNN	$O(T * Q * t * q)$	Cell phone	Classifiation
[73]	/	1 year	No	Historical data: Crop yield soil parameters	ELM	$O(L^3 + L^2 * n)$	/	Prediction
[75]	96	1year	Yes	Image data: Digital images	SVM, ANN, NBKNN DT Discriminant analysis	Discriminant analysis: $O(n * P^2)$ NB: $O(n * p) + O(P)$	Camera Nikon CoolpixL22	Classification
[76]	4 Landsat-8 scenes 15 Sentinel-1 scenes	/	Yes	Satellite data: Multi-temporal multi-source images	CNN	$O(T * Q * t * q)$	Landsat-8, Sentinel-1A satellites	Classification

Table 8 (continued)

References	V1	V2	V3	V4	ML	Complexity	Device	Task
[152]	8945	multi-spectral image: 8 days interval for 30 per year	Yes	Satellite, sensor data: surface reflectance land surface temperature land cover	Gaussian CNN	$O(T * Q * t * 2)$	MODIS satellite	Prediction
[62]	/	2006-2011 8 days period 32 times	/	Satellite sensor data: NDVI Precipitation land surface temperature	Rulequest cubist	$O(n^2 * P)$	MODIS satellite	Prediction

No data was clear,Yes data was cleaned and filtered and some samples were not considered because of abnormalities, inconsistencies or duplication and for other reasons,*n* number of data points,n_{sv}: number of support vectors,*P* number of features,n_{trees}: number of trees,*c* number of cluster,*d* number of dimension.*i* number of iterations,*L* number of hidden layers,*T* * *Q* size of input feature map; spatial, two/three-dimensional kernels are of size (*t* * *q*),*nl.*: number of neurons at layer *i,ep*: number of epochs
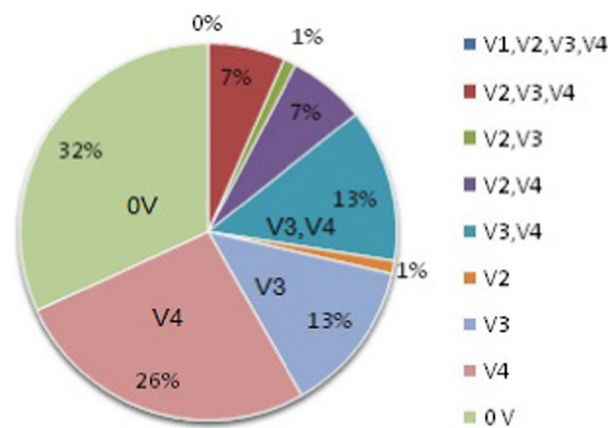
Fig. 7 Distribution of works according to the used Big Data dimensions'

(a) Considering the zero-V (0V) applications.   (b) Without the zero-V (0V) applications.
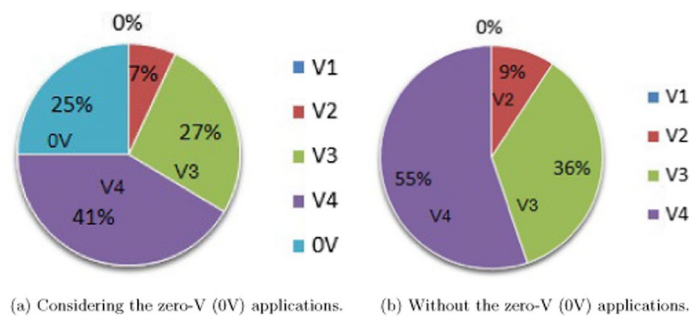
Fig. 8 Percent of employment of big data dimensions

Figures 7 and 8 show that no work has a full employment of big data (4Vs). One can notice that the agricultural data is multidimensional and heterogeneous (variety). Moreover, we have found that the prediction applications display more use of big data, there exist studies that have used three dimensions such as DMZ applications. It is worth noting that these applications, either prediction or delineation of zones, have the potential to use big data to provide stable and accurate results.

If we put aside the volume dimension (V1) (see Figure 7, only 7% of the reviewed studies used (V2, V3 and V4), and 32% of studies just employed data mining techniques for agriculture problems. The most employed data mining techniques are for prediction, including yield prediction, forecasting, prediction of fertiliser applications, etc.

DA practical challenges

There exist a number of challenges and obstacles impeding the potential benefit of DA. In [104], the authors studied the barriers that prevent the adoption of smart farming in their country, Brazil. Some of these barriers include lack of integration and compatibility between different agriculture systems, lack of advanced data manipulation of data obtained from different equipment, poor telecommunications infrastructure on rural areas, and finally, the lack of training in deploying and using new technologies. These barriers are common to the majority of countries in the world.

From the Table 7, we can see that over 73% of crop farms are located in developing countries. So that, the investment in high and sophisticated DA technologies is not there. Most of the main technologies used in DA systems (GPS, UAV, auto-steering and variable rate technology) are designed for relatively large-scale farms located in developed countries [10] or designed by developed countries. Some of these technologies are becoming available recently. For instance, since 2018 African scientists can have access to free and open-source satellite data as a result of a deal signed by the African Union with the European Commission's Copernicus programme.

As DA is relatively new technology, there is a lack of standards and common solutions for data collection, preparation and storage. In addition, there is a lack of data for many reasons, farmers did not record their data and it takes time to build significant historical datasets [20, 39, 77, 78, 92, 146]. Another major barrier is that many farmers are relying more on their expertise and refusing to adopt these new and complex technologies [10]. Moreover, the transition from their traditional practices and farming habits to these technologies comes with a cost and energy (training and learning new skills).

[20] States that the legal and regulatory frameworks around the collection, sharing and use of agricultural data contributes to a range of challenges. Many laws potentially influence the ownership, control of and data access. Ref. [74] presented a set of socio-ethical imperatives associated with the use of data in agriculture, including dependency risks, data concentration, potential lock-in effects, and the peril of transformation of farmers into information tools, in addition to the sustainability challenges.

Finally, according to [47], the real economic value of the use of big data in farming is still unknown, especially for small-scale farming. Consequently, it will be hard to convince them to switch from process-driven towards data and machine learning driven. This is reaffirmed in [20], where the authors stated that on one side, farmers are enticed with promises of increased profits and farming efficiency, on the other hand the proofs are not there yet.

Conclusion

Digital agriculture (DA) is a data-driven approach that exploits the hidden information within the collected data to gain new insights; transforming the farming practices from intuitive-based decision-making to informed-based decision-making. DA relies on efficient data collection practices, efficient data preparation and storage techniques, efficient data analytics, and efficient deployment and exploitation of the gained insights to make optimal farming decisions.

In this study, we presented a systematic review of the potential use of the data mining process in crop production and management and highlighted serious gaps which can be considered in future studies. The majority of the current practices were dominated by statistical analyses and small machine learning systems. However, these can only give some ideas within a very limited view of the overall system. Agricultural data-driven applications collect a significant amount of data from various sources. This constitutes an excellent opportunity to the field to answer numerous research and practical questions that were not possible before. Nevertheless, despite all the advantages that can be gained from DA, there are several other challenges and obstacles that need to be

addressed, among them lack of data, lack of skills, and lack of maturity and standards so that it can be adopted and deployed quickly and easily.

In this study, we cover approaches that deal the entire process of data mining; from data collection to knowledge deployment. We cover this process from big data view, with more focus on crop monitoring and management in an attempt to understand the challenges that DA is currently facing. We defined the research questions addressed by the study and provided a classification of data mining techniques used in the field. For each class, a set of representative existing works have been reviewed, and an analytical study has been provided to highlight the category of machine learning method applied and for which purpose. We discussed the big data concepts and its current impact on DA, and showed that from the data analyst's view, the transition towards DA is ready to embrace big data analytics concepts. This provides new opportunities of investment into these challenges and allows for a efficient ways of managing crops. Besides, it will provide farmers with new insights into how they can grow crops more efficiently, while minimising the impact on the environment. It also promises new levels of scientific discovery and innovative solutions to more complex problems.

Abbreviations

ANN	Artificial Neural Network
BC	Bayesian classifier
CNN	Convolution Neural Network
DT	Decision tree
DMZ	Delineation of management zones
DNN	Deep Neural Network
ELM	Extreme learning machine
EVI	Enhanced Vegetation Index
FCM	Fuzzy C-means
GIS	Geographical information system
GPS	Global positioning system
INS	Inertial navigation system
KNN	K-Nearest Neighbour
LSTM	Long/Short Term Memory Network
KNN	K-Nearest Neighbour
MLP	Multi-layer perceptron
MODIS	Moderate-resolution imaging spectro-radiometer
NDVI	Normalised difference vegetation index
OSAVI	Optimised soil adjusted vegetation index
RF	Random forest
RBF	Radial basis function
RGB	RedGreenBlue
RNN	Recurrent neural network
RVI	Ratio Vegetation Index
SVM	Support vector machine
SVI	Spectral vegetation index
SVR	Support vector regression
UAV	Unmanned aerial vehicle
UGV	Unmanned ground vehicles
WDRVI	Weighted dynamic ranged vegetation index

Author contributions
NC and TK conceived of the presented idea. NC designed the paper and figures, collected and analysed the results extracted from the reviewed papers. TK verified the relevance of the bibliography and the consistency of the results. All authors participated at the writing of the manuscript and provided critical feedback and helped shape the research, analysis and manuscript. All authors read and approved the final manuscript.

References

1. Abbas F, Afzaal H, Farooque A, Tang S. Crop yield prediction through proximal sensing and machine learning algorithms. Agronomy. 2020. https://doi.org/10.3390/agronomy10071046.
2. Ahmed F, Al-Mamun H, Bari H, Hossain E, Kwan P. Classification of crops and weeds from digital images: a support vector machine approach. Crop Prot. 2012;40:98–104. https://doi.org/10.1016/j.cropro.2012.04.024.
3. Akbarzadeh S, Paap A, Ahderom S, Apopei B, Alameh K. Plant discrimination by support vector machine classifier based on spectral reflectance. Comput Electron Agric. 2018;148:250–8. https://doi.org/10.1016/j.compag.2018.03.026.
4. Alibabaei K, Gaspar P, Lima T. Crop yield estimation using deep learning based on climate big data and irrigation scheduling. Energies. 2021;14:3004. https://doi.org/10.3390/en14113004.
5. Amatya S, Karkee M, Gongal A, Zhang Q, Whiting M. Detection of cherry tree branches with full foliage in planar architecture for automated sweet-cherry harvesting. Biosyst Eng. 2015;146:3–15. https://doi.org/10.1016/j.biosystemseng.2015.10.003.
6. Aravind K, Raja P. Automated disease classification in (selected) agricultural crops using transfer learning. Autom J Control Meas Electron Comput Commun. 2020;62:260–72. https://doi.org/10.1080/00051144.2020.1728911.
7. Aravind K, Maheswari P, Raja P, Szczepanski C. Crop disease classification using deep learning approach: an overview and a case study. In: Das H, Pradhan C, Dey N, editors. Deep learning for data analytics foundations, biomedical applications, and challenges. Cambridge: Academic Press; 2020. p. 173–95. https://doi.org/10.1016/b978-0-12-819764-6.00010-7.
8. Arribas J, Sanches-Ferrero G, Ruiz-Ruiz G, Gomez-Gil J. Leaf classification in sunflower crops by computer vision and neural networks. Comput Electron Agric. 2011;78:9–18. https://doi.org/10.1016/j.compag.2011.05.007.
9. Arsenovic M, Karanovic M, Sladojevic S, Anderla A, Stefanovic D. Solving current limitations of deep learning based approaches for plant disease detection. Symmetry. 2019. https://doi.org/10.3390/sym11070939.
10. Balafoutis AT, Beck B, Fountas S, Tsiropoulos Z, Vangeyte J, van der Wal T, Soto-Embodas I, Gomez-Barbero M, Pedersen S,. Smart farming technologies–description taxonomy and economic impact. In: Pedersen SM, Lind K, editors. Precision agriculture: technology and economic perspectives, progress in precision agriculture, chapter 2. Cham: Springer; 2017. p. 21–78. https://doi.org/10.1007/978-3-319-68715-5.
11. Barbedo JA. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. Comput Electron Agric. 2018;153:46–53. https://doi.org/10.1016/j.compag.2018.08.013.
12. Behmann J, Mahlein AK, Rumpf T, Romer C, Plumer L. A review of advanced machine learning methods for the detection of biotic stress in precision crop protection. J Precis Agric. 2014;16:239–60. https://doi.org/10.1007/s11119-014-9372-7.
13. Bendre M, Thool R, Thool V. Big data in precision agriculture through ICT: rainfall prediction using neural network approach. In: Satapathy S, Bhatt Y, Joshi A, Mishra D, editors. Proceedings of the International congress on information and communication technology. Singapore: Springer; 2016. p. 165–75.
14. Berckmans D. Precision livestock farming technologies for welfare management in intensive livestock systems. Rev Sci. 2014;33:189–96.
15. Bi L, Hu G, Raza M, Kandel Y, Leandro L, Mueller D. A gated recurrent units (gru)-based model for early detection of soybean sudden death syndrome through time-series satellite imagery. Remote Sens. 2020. https://doi.org/10.3390/rs12213621.
16. Brahimi M, Arsenovic M, Laraba S, Sladojevic S, Boukhalfa K, Moussaoui A. Deep learning for plant diseases: detection and saliency map visualisation. In: Zhou J, Chen F, editors. Human and machine learning. Cham: Springer; 2018. p. 93–117. https://doi.org/10.1007/978-3-319-90403-0_6.
17. Breunig F, Galvao L, Dalagnol R, Dauve C, Parraga A, Santi A, Flora DD, Chen S. Delineation of management zones in agricultural fields using cover-crop biomass estimates from planetscope data. Int J Appl Earth Obs Geoinf. 2020. https://doi.org/10.1016/j.jag.2019.102004.
18. Brock A, Brouder S, Blumhoff G, Hofmann B. Defining yield-based management zones for corn-soybean rotations. Agron J. 2005;97:1115–28. https://doi.org/10.2134/agronj2004.0220.
19. Cao J, Zhao Z, Luo Y, Zhang L, Zhang J. ZLi, Tao F, Wheat yield predictions at a county and field scale with deep learning, machine learning, and google earth engine. Eur J Agron. 2021;123: 126204. https://doi.org/10.1016/j.eja.2020.126204.

20. Carolan M. Acting like an algorithm: digital farming platforms and the trajectories they (need not) lock-in. Agric Hum Values. 2020;37:1041–53. https://doi.org/10.1007/s10460-020-10032-w.

21. Chen J, Liu Q, Gao L. Visual tea leaf disease recognition using a convolutional neural network model. Symmetry. 2019. https://doi.org/10.3390/sym11030343.

22. Chen N, Yu L, Zhang X, Shen Y, Zeng L, Hu Q, Niyogi D. Mapping paddy rice fields by combining multi-temporal vegetation index and synthetic aperture radar remote sensing data using google earth engine machine learning platform. Remote Sens. 2020;2020. https://doi.org/10.3390/rs12182992.

23. Cheng H, Damerow L, Sun Y, Blanke M. Early yield prediction using image analysis of apple fruit and tree canopy features with neural networks. J Imaging. 2017. https://doi.org/10.3390/jimaging3010006.

24. Chergui N, Kechadi T, McDonnell M, The impact of data analytics in digital agriculture: a review. In: the 2020 IEEE International multi-conference on: organization of knowledge and advanced technologies (OCTA). Isko-Maghreb: 'International society for knowledge organization'. February 6-8, 2020 Tunis (Tunisia). 2020. https://doi.org/10.1109/OCTA49274.2020.9151851

25. Chinchuluun R, Lee W, Bhorania J, Pardalos P. Clustering and classification algorithms in food and agricultural applications: a survey. In: Papajorgji PJ, Pardalos PM, editors. Advances in modelling agricultural systems springer optimisation and its applications. Boston: Springer; 2008. p. 433–54.

26. Contiu S, Groza A. Improving remote sensing crop classification by argumentation-based conflict resolution in ensemble learning. Expert Syst Appl. 2016;64:269–86. https://doi.org/10.1016/j.eswa.2016.07.037.

27. Crane-Droesch A. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. Environ Res Lett. 2018. https://doi.org/10.1088/1748-9326/aae159.

28. Cruz A, Luvisi A, Bellis LD, Ampatzidis Y. X-fido: an effective application for detecting olive quick decline syndrome with deep learning and data fusion. Front Plant Sci. 2017. https://doi.org/10.3389/fpls.2017.01741.

29. Dadashzadeh M, Abbaspour-Gilandeh Y, Mesri-Gundoshmian T, Sabzi S, Hernández-Hernández J, Hernández-Hernández M, Arribas J. Weed classification for site-specific weed management using an automated stereo computer-vision machine-learning system in rice fields. Plants. 2020;5:22–36. https://doi.org/10.3390/plants9050559.

30. Dahane A, Benameur R, Kechar B. An IoT low-cost smart farming for enhancing irrigation efficiency of smallholders farmers. Wirel Pers Commun. 2022. https://doi.org/10.1007/s11277-022-09915-4.

31. Debats S, Luo D, Estes L, Fuchs T, Caylor K. A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes. Remote Sens Environ. 2016;179:210–21. https://doi.org/10.1016/j.rse.2016.03.010.

32. Du CJ, Kechadi M, Zhang YB, Huang BQ. A hybrid HMM-SVM method for online handwriting symbolrecognition. Intell Syst Des Appl. 2006;3:887–91. https://doi.org/10.1109/ISDA.2006.61.

33. Dyrmann M, Karstoft H, Midtiby H. Plant species classification using deep convolutional neural network. Biosyst Eng. 2016;151:72–80. https://doi.org/10.1016/j.biosystemseng.2016.08.024.

34. Ehret D, Hill B, Helmer T, Edwards D. Neural network modeling of greenhouse tomato yield, growth and water use from automated crop monitoring data. Comput Electron Agric. 2011;79:82–9. https://doi.org/10.1016/j.compag.2011.07.013.

35. Elavarasan D, Vincent D, Sharma V, Zomaya A, Srinivasan K. Forecasting yield by integrating agrarian factors and machine learning models: A survey. Comput Electron Agric. 2018;155:257–82. https://doi.org/10.1016/j.compag.2018.10.024.

36. Fardusi MJ, Chianucci F, Barbati A. Concept to practice of geospatial-information tools to assist forest management and planning under precision forestry framework a review. Ann Silvic Res. 2017;41:3–14. https://doi.org/10.12899/asr-1354.

37. Feldman B, Martin E, Skotnes T. Big data in healthcare hype and hope, october 2012.dr. bonnie 2012;360, 2012. Http://www.westinfo.eu/files/big-data-inhealthcare

38. Ferentinos PK. Deep learning models for plant disease detection and diagnosis. Comput Electron Agric. 2018;145:311–8. https://doi.org/10.1016/j.compag.2018.01.009.

39. Fielke S, Taylor B, Jakku E. Digitalisation of agricultural knowledge and advice networks: a state-of-the art. Agric Syst. 2020. https://doi.org/10.1016/j.agsy.2019.102763.

40. Filippi P, Jones E, Bishop T, Acharige N, Dewage S, Johnson L, Ugbaje S, Jephcott T, Paterson S, Whelan B. A big data approach to predicting crop yield. In: Proceedings of the 7th Asian-Australasian Conference on Precision Agriculture 16-18 October 2017. Hamilton; 2017.https://doi.org/10.5281/zenodo.893668

41. Formaggio A, Vieira M, Renno C. Object based image analysis (obia) and data mining (dm) in landsat time series for mapping soybean in intensive agricultural regions. In: Proceedings of IEEE International Geoscience and Remote Sensing Symposium. 22-27 July 2012. Munich; 2012. p. 2257–2260. https://doi.org/10.1109/IGARSS.2012.6351047

42. Fukuda S, Spreer W, Yasunaga E, Yuge K, Sardsud V, Muller J. Random forests modelling for the estimation of mango (Mangifera indica l. cv.chok anan) fruit yields under different irrigation regimes. J Agric Water Manag. 2013;116:142–50. https://doi.org/10.1016/j.agwat.2012.07.003.

43. Galambosova J, Rataj V, Prokeinova R, Presinska J. Determining the management zones with hierarchic and non-hierarchic clustering methods. Res Agric Eng. 2014;60:44–51. https://doi.org/10.17221/34/2013-RAE.

44. Gao J, Nuyttens D, Lootens P, He Y, Pieters J. Recognising weeds in a maize crop using a random forest machine-learning algorithm and near-infrared snapshot mosaic hyperspectral imagery. Biosyst Eng. 2018;170:30–50. https://doi.org/10.1016/j.biosystemseng.2018.03.006.

45. Golhani K. KBalasundram S, Vadamalai G, Pradhan B, A review of neural networks in plant disease detection using hyperspectral data. Inf Proc Agric. 2018;5:354–71. https://doi.org/10.1016/j.inpa.2018.05.002.

46. Gonzalez-Sanchez A, Frausto-Solis J, Ojeda-Bustamante W. Predictive ability of machine learning methods for massive crop yield prediction. Spanish J Agric Res. 2014;12:313–28. https://doi.org/10.5424/sjar/2014122-4439.

47. Griffin T, Mark T, Ferrell S, Janzen T, Ibendahl G, Bennett J, Maurer J, Shanoyan A. Big data considerations for rural property professionals. Am Soc Farm Manage Rural Appraisers. 2016;79:167–80.

48. Guastaferro F, Castrignano A, Benedetto DD, Sollitto D, Troccoli A, Cafarelli B. A comparison of different algorithms for the delineation of management zones. Precis Agric. 2010;11:600–20. https://doi.org/10.1007/s11119-010-9183-4.

49. Guo A, Huang W, Dong Y, Ye H, Ma H, Liu B, Wu W, Ren Y, Ruan C, Geng Y. Wheat yellow rust detection using UAV-based hyperspectral technology. Remote Sensing. 2021. https://doi.org/10.3390/rs13010123.

50. Guo Y, Fu Y, Hao F, Zhang X, Wu W, Jin X, Bryant C, Senthilnath J. Integrated phenology and climate in rice yields prediction using machine learning methods. Ecol Indic. 2021;120: 106935. https://doi.org/10.1016/j.ecolind.2020.106935.

51. Gyamerah S, Ngare P, Ikpe D. Probabilistic forecasting of crop yields via quantile random forest and Epanechnikov Kernel function. Agric For Meteorol. 2020. https://doi.org/10.1016/j.agrformet.2019.107808.

52. Habaragamuwa H, Ogawa Y, Suzuki T, Masanori T, Kondo O. Detecting greenhouse strawberries (mature and immature), using deep convolutional neural network. Eng Agric Environ Food. 2018;11:127–38. https://doi.org/10.1016/j.eaef.2018.03.001.

53. Haghverdi A, Leib B, Washington-Allen R, Ayers P, Buschermohle M. Perspectives on delineating management zones for variable rate irrigation. Comput Electron Agric. 2015;117:154–67. https://doi.org/10.1016/j.compag.2015.06.019.

54. Han J, Zhang Z, Cao J, Luo Y, Zhang L, Li Z, Zhang J. Prediction of winter wheat yield based on multi-source data and machine learning in china. Remote Sensing. 2020. https://doi.org/10.3390/rs12020236.

55. Huang K. Application of artificial neural network for detecting phalaenopsis seedling diseases using color and texture features. Comput Electron Agric. 2007;57:3–11. https://doi.org/10.1016/j.compag.2007.01.015.

56. Huang Y, Chen Z, Yu T, Huang X, Gu X. Agricultural remote sensing big data: Management and applications. J Integr Agric. 2018;17:1915–31. https://doi.org/10.1016/S2095-3119(17)61859-8.

57. Ingeli M, Galambosova J, Prokeinova R, Rataj V. Application of clustering method to determine production zones of field. Acta Technol Agric. 2015;18:42–5. https://doi.org/10.1515/ata-2015-0009.

58. Jain M, Mondal P, DeFries R, Small C, Galford G. Mapping cropping intensity of smallholder farms: a comparison of methods using multiple sensors. Remote Sensing Environ. 2013;134:210–23. https://doi.org/10.1016/j.rse.2013.02.029.

59. Jeong J, Resop J, Mueller N, Fleisher D, Yun K, Butler E, Timlin D, Shim K, Gerber J, Reddy V, Kim S. Random forests for global and regional crop yield predictions. PLoS ONE. 2016. https://doi.org/10.1371/journal.pone.0156571.

60. Ji Z, Pan Y, Zhu X, Wang J, Li Q. Prediction of crop yield using phenological information extracted from remote sensing vegetation index. Sensors. 2021;4:1406. https://doi.org/10.3390/s21041406.

61. Jiang Q, Wang QFZ. Study on delineation of irrigation management zones based on management zone analyst software. In: Jiang Q, editor. Computer and computing technologies in agriculture IV. CCTA 2010 IFIP advances in information and communication technology, vol. 346. Berlin: Springer; 2011. p. 4559–66. https://doi.org/10.1007/978-3-642-18354-6_50

62. Johnson D. An assessment of pre-and within-season remotely sensed variables for forecasting corn and soybean yields in the united states. Remote Sensing Environ. 2014;141:116–28. https://doi.org/10.1016/j.rse.2013.10.027.

63. Kamal K, Yin Z, Wu M, Wu Z. Depthwise separable convolution architectures for plant disease classification. Comput Electron Agric. 2019. https://doi.org/10.1016/j.compag.2019.104948.

64. Kamilaris A, Kartakoullis A, Prenafeta-Boldú F. A review on the practice of big data analysis in agriculture. Comput Electron Agric. 2017;143:23–37. https://doi.org/10.1016/j.compag.2017.09.037.

65. Kamir E, Waldner F, Hochman Z. Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. ISPRS J Photogramm Remote Sens. 2020;160:124–35. https://doi.org/10.1016/j.isprsjprs.2019.11.008.

66. Khalili E, Kouchaki S, Ramazi S, Ghanati F. Machine learning techniques for soybean charcoal rot disease prediction. Front Plant Sci. 2021. https://doi.org/10.3389/fpls.2020.590529.

67. Kim N, Lee Y. Machine learning approaches to corn yield estimation using satellite images and climate data: a case of Lowa state. J Korean Soc Surv Geod Photogramm Cartogr. 2016;34:383–90. https://doi.org/10.7848/ksgpc.2016.34.4.383.

68. Kim N, Ha K, Park N, Cho J, Hong S, Lee Y. A comparison between major artificial intelligence models for crop yield prediction: case study of the midwestern united states, 2006–2015. ISPRS Int J Geoinform. 2019. https://doi.org/10.3390/ijgi8050240.

69. Kitchen N, Sudduth K, Myers D, Drummond S, Hong S. Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. Comput Electron Agric. 2005;46:285–308. https://doi.org/10.1016/j.compag.2004.11.012.

70. Klerk L, Jakku E, Labarthe P. A review of social science on digital agriculture, smart farming and agriculture 4.0: new contributions and a future research agenda. NJAS Wageningen J Life Sci. 2019. https://doi.org/10.1016/j.njas.2019.100315.

71. Klompenburg T, Kassahun A, Catal C. Crop yield prediction using machine learning: a systematic literature review. Comput Electron Agric. 2020. https://doi.org/10.1016/j.compag.2020.105709.

72. Koch B, Khosla R, Frasier W, Westfall D, Inman D. Economic feasibility of variable-rate nitrogen application utilizing site-specific management zones. Agron J. 2004;96:1572–80. https://doi.org/10.2134/agronj2004.1572.

73. Kouadio L, Deo R, Byrareddy V, Adamowski J, Mushtaq S, Nguyen VP. Artificial intelligence approach for the prediction of robusta coffee yield using soil fertility properties. Comput Electron Agric. 2018;155:324–38. https://doi.org/10.1016/j.compag.2018.10.014.

74. Kritikos M. Precision agriculture in europe: legal, social and ethical considerations. science and technology options assessment. Scientific foresight unit (STOA) of the European parliament, brussels pe 603.207. 2017.

75. Kurtulmus F, Lee W, Vardar A. Immature peach detection in colour images acquired in natural illumination conditions using statistical classifiers and neural network. Precis Agric. 2014;15:57–79. https://doi.org/10.1007/s11119-013-9323-8.

76. Kussul N, Lavreniuk M, Skakun S, Shelestov A. Deep learning classification of land cover and crop types using remote sensing data. Geosci Remote Sens Lett. 2017;14:778–82. https://doi.org/10.1109/LGRS.2017.2681128.

77. Lioutas E, Charatsari C. Big data in agriculture: does the new oil lead to sustainability? Geoforum. 2020;109:1–3. https://doi.org/10.1016/j.geoforum.2019.12.019.

78. Lioutas ED, Charatsari C, Rocca GL, Rosa MD. Key questions on the use of big data in farming: an activity theory approach. NJAS Wageningen J Life Sci. 2019. https://doi.org/10.1016/j.njas.2019.04.003.

79. Liu B, Zhang Y, He D, Li Y. Identification of apple leaf diseases based on deep convolutional neural networks. Symmetry. 2017. https://doi.org/10.3390/sym10010011.

80. Liu L, Dong Y, Huang W, Du X, Ma H. Monitoring wheat fusarium head blight using unmanned aerial vehicle hyperspectral imagery. Remote Sens. 2020. https://doi.org/10.3390/rs12223811.

81. Ma H, Jing Y, Huang W, Shi Y, Dong Y, Zhang J, Liu L. Integrating early growth information to monitor winter wheat powdery mildew using multi-temporal Landsat-8 imagery. Sensors. 2018. https://doi.org/10.3390/s18103290.

82. Mahlein A, Alisaac E, Masri AA, Behmann J, Dehne H, Oerke E. Comparison and combination of thermal, fluorescence, and hyperspectral imaging for monitoring fusarium head blight of wheat on spikelet scale. Sensors. 2019. https://doi.org/10.3390/s19102281.

83. Maimaitijiang M, Sagan V, Sidike P, Hartling S, Esposito F, Fritschi F. Soybean yield prediction from UAV using multimodal data fusion and deep learning. Remote Sens Environ. 2020. https://doi.org/10.1016/j.rse.2019.111599.

84. Martinez-Casasnovas J, Escola A, Arno J. Use of farmer knowledge in the delineation of potential management zones in precision agriculture: a case study in maize (Zea mays L.). Agriculture. 2018. https://doi.org/10.3390/agriculture8060084.

85. Mathur SBR, Shukla A, Suresh K, Prakash C. Spatial variability of soil properties and delineation of soil management zones of oil palm plantations grown in a hot and humid tropical region of southern India. Catena. 2018;165:251–9. https://doi.org/10.1016/j.catena.2018.02.008.

86. Mauro AD, Greco M, Grimaldi M. A formal definition of big data based on its essential features. Libr Rev. 2016;65:122–35. https://doi.org/10.1108/LR-06-2015-0061.

87. Metwally M, Shaddad S, Liu M, Yao R, Abdo A, Li P, Jiao J, Chen X. Soil properties spatial variability and delineation of site-specific management zones based on soil fertility using fuzzy clustering in a hilly field in Jianyang, Sichuan, China. Sustainability. 2019;2019. https://doi.org/10.3390/su11247084.

88. Mohanty S, Hughes D, Salathe M. Using deep learning for image-based plant disease detection. Front Plant Sci. 2016;7:1–10. https://doi.org/10.3389/fpls.2016.01419.

89. Mucherino A, Papajorgji P, Pardalos PM. A survey of data mining techniques applied to agriculture. J Operational Res. 2009;9:121–40. https://doi.org/10.1007/s12351-009-0054-6.

90. Nawar S, Corstanje R, Halcro G, Mulla D, Mouazen A. Delineation of soil management zones for variable-rate fertilization: a review. Adv Agron. 2017;143:175–245. https://doi.org/10.1016/bs.agron.2017.01.003.

91. 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. https://doi.org/10.3390/rs12234000.

92. Newton J, Nettle R, Pryce J. Farming smarter with big data: Insights from the case of Australia's national dairy herd milk recording scheme. Agric Syst. 2020. https://doi.org/10.1016/j.agsy.2020.102811.

93. Ngo M, Kechadi T. Electronic farming records-a framework for normalising agronomic knowledge discovery. Comput Electron Agric. 2021. https://doi.org/10.1016/j.compag.2021.106074.

94. Ngo QH, Le-Khac NA, Kechadi T. Predicting soil pH by using nearest fields. In: Bramer M, Petridis M, editors. Artificial Intelligence XXXVI. SGAI 2019. Lecture notes in computer science, vol. 11927. Cham: Springer; 2019. https://doi.org/10.1007/978-3-030-34885-4_40.

95. Ngo VM, Kechadi MT Crop knowledge discovery based on agricultural big data integration. In: Proceedings of the 4th International conference on machine learning and soft computing, association for computing machinery. New York; ICMLSC. 2020. https://doi.org/10.1145/3380688.3380705

96. Ngo VM, Le-Khac N, Kechadi T. Data warehouse and decision support on integrated crop big data. Int J Bus Process Integr Manag. 2020. https://doi.org/10.1504/IJBPIM.2020.113115.

97. Oliveira I, Cunha R, Silva B, Netto M. A scalable machine learning system for pre-season agriculture yield forecast. In: the 14th IEEE eScience Conference. 2018. https://doi.org/10.1109/eScience.2018.00131

98. Oliver D, Bartie P, Heathwaite A, Pschetz L, Quilliam R. Design of a decision support tool for visualising E. coli risk on agricultural land using a stakeholder-driven approach. Land Use Policy. 2017;66:227–34. https://doi.org/10.1016/j.landusepol.2017.05.005.

99. Ortega R, Santibanez O. Determination of management zones in corn (Zea mays L.) based on soil fertility. Comput Electron Agric. 2007;58:49–59. https://doi.org/10.1016/j.compag.2006.12.011.

100. Ouzemou J, Harti A, Lhissou R. AEl-Moujahid, Bouch N, El-Ouazzani R, Bachaoui E, El-Ghmari A, Crop type mapping from pansharpened Landsat 8 NDVI data: a case of a highly fragmented and intensive agricultural system. Remote Sens Appl Soc Environ. 2018. https://doi.org/10.1016/j.rsase.2018.05.002.

101. Pantazi X, Moshou D, Mouazen A, Alexandridis T, Kuang B. Data fusion of proximal soil sensing and remote crop sensing for the delineation of management zones in arable crop precision farming. In: CEUR Workshop Proceedings. CEUR-WS. 2015. p. 765–776.

102. Pantazi X, Moshou D, Alexandridis T, Whetton R, Mouazen A. Wheat yield prediction using machine learning and advanced sensing techniques. J Comput Electron Agric. 2016;121:57–65. https://doi.org/10.1016/j.compag.2015.11.018.

103. Patricio D, Rieder R. Computer vision and artificial intelligence in precision agriculture for grain crops: a systematic review. Comput Electron Agric. 2018;153:69–81. https://doi.org/10.1016/j.compag.2018.08.001.

104. Pivoto D, Waquil P, Talamini E, Finocchio C, Corte V, Mores G. Scientific development of smart farming technologies and their application in Brazil. Inform Process Agric. 2018;5:21–32. https://doi.org/10.1016/j.inpa.2017.12.002.

105. Poppe K, Wolfert S, Verdouw C, Verwaart T. Information and communication technology as a driver for change in agri-food chains. Eurochoices. 2013;12:60–5.

106. Qin F, Liu D, Sun B, Ruan L, Ma Z, Wang H. Identification of alfalfa leaf diseases using image recognition technology. PLoS ONE. 2016. https://doi.org/10.1371/journal.pone.0168274.
107. Rafii F, TKechadi. Collection of historical weather data: Issues with missing values. In: Proceedings of the 4th International conference on smart city applications, association for computing machinery. New York; 2019. https://doi.org/10.1145/3368756.3368974
108. Ramos P, Prieto F, Montoya E, Oliveros C. Automatic fruit count on coffee branches using computer vision. Comput Electron Agric. 2017;137:9–22. https://doi.org/10.1016/j.compag.2017.03.010.
109. Raza M, Harding C, Liebman M, Leandro L. Exploring the potential of high-resolution satellite imagery for the detection of soybean sudden death syndrome. Remote Sens. 2020. https://doi.org/10.3390/rs12071213.
110. Reyes J, Wendroth O, Matocha C, Zhu J. Delineating site-specific management zones and evaluating soil water temporal dynamics in a farmer's field in Kentucky. Vadose Zone J. 2019;18:1–19. https://doi.org/10.2136/vzj2018.07.0143.
111. Rezapour S, Jooyandeh E, Ramezanzade M, Mostafaeipour S, Jahangiri M, Issakhov A, Chowdhury S, Techato K. Forecasting rainfed agricultural production in arid and semi-arid lands using learning machine methods: a case study. Sustainability. 2021;13:4607. https://doi.org/10.3390/su13094607.
112. Reznik T, Lukas V, Krivanek Z, Kepka M, Herman L, Reznikova H. Disaster risk reduction in agriculture through geospatial (big) data processing. ISPRS Int J Geoinform. 2017. https://doi.org/10.3390/ijgi6080238.
113. Rijswijk K, Klerk L, Turner J. Digitalisation in the New Zealand agricultural knowledge and innovation system: Initial understandings and emerging organisational responses to digital agriculture. NJAS Wageningen J Life Sci. 2019. https://doi.org/10.1016/j.njas.2019.100313.
114. Ji R, Min J, Wang Y, Cheng H, Zhang H, Shi W. In-season yield prediction of cabbage with a hand-held active canopy sensor. Sensors. 2017. https://doi.org/10.3390/s17102287.
115. Rosa LCL, Feitosa R, Happ P, Sanches ID, da Costa GOP. Combining deep learning and prior knowledge for crop mapping in tropical regions from multi-temporal SAR image sequences. Remote Sens. 2019. https://doi.org/10.3390/rs11172029.
116. RuB G, Krus R. Exploratory hierarchical clustering for management zone delineation in precision agriculture. In: Industrial conference on data mining ICDM 2011: advances in data mining. Applications and theoretical aspects. Lecture notes in computer science book series (LNCS, volume 6870). 2011. p. 161–173. https://doi.org/10.1007/978-3-642-23184-1_13
117. Sa I, Ge Z, Upcroft FDB, Perez T, Mccool C. Deepfruits: a fruit detection system using deep neural networks. Sensors. 2016. https://doi.org/10.3390/s16081222.
118. Sa I, Popovic M, Khanna R, Chen Z, Lottes P, Liebisch F, Nieto J, Stachniss C, Walter A, Siegwart R. Weedmap: a large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. Remote Sens. 2018. https://doi.org/10.3390/rs10091423.
119. Sabzi S, Abbaspour-Gilandeh Y. Using video processing to classify potato plant and three types of weed using hybrid of artificial neural network and particle swarm algorithm. Measurement. 2018;126:22–36. https://doi.org/10.1016/j.measurement.2018.05.037.
120. Sakamoto T. Incorporating environmental variables into a modis-based crop yield estimation method for United states corn and soybeans through the use of a random forest regression algorithm. ISPRS J Photogramm Remote Sens. 2020;160:208–28. https://doi.org/10.1016/j.isprsjprs.2019.12.012.
121. Schwalbert R, Amado T, Corassa G, Pott L, Prasad P, Ciampitti I. Satellite-based soybean yield forecast: integrating machine learning and weather data for improving crop yield prediction in southern brazil. Agric For Meteorol. 2020. https://doi.org/10.1016/j.agrformet.2019.107886.
122. Sengupta S, Lee W. Identification and determination of the number of immature green citrus fruit in a canopy under different ambient light conditions. Biosyst Eng. 2014;117:51–61. https://doi.org/10.1016/j.biosystemseng.2013.07.007.
123. Senthilnath J, Dokania A, Kandukuri M, Ramesh K, Anand G, Omkar S. Detection of tomatoes using spectral-spatial methods in remotely sensed RGB images captured by UAV. Biosyst Eng. 2016;146:16–32. https://doi.org/10.1016/j.biosystemseng.2015.12.003.
124. Shafi U, Mumtaz R, Garcia-Nieto J, Hassan S, Zaidi S, Iqbal N. Precision agriculture techniques and practices: from considerations to applications. Sensors. 2019. https://doi.org/10.3390/s19173796.
125. Sibiya M, Sumbwanyambe M. A computational procedure for the recognition and classification of maize leaf diseases out of healthy leaves using convolutional neural networks. AgriEngineering. 2019;1:119–31. https://doi.org/10.3390/agriengineering1010009.
126. Singh A, Jones S, Ganapathysubramanian B, Sarkar S, Mueller D, Sandhu K, Nagasubramanian K. Challenges and opportunities in machine-augmented plant stress phenotyping. Trends Plant Sci. 2021;25:53–69. https://doi.org/10.1016/j.tplants.2020.07.010.
127. Singh S, Ganapathysubramanian B, Sarkar S, Singh A. Deep learning for plant stress phenotyping: trends and future perspectives. Trends Plant Sci. 2018;23:883–98. https://doi.org/10.1016/j.tplants.2018.07.004.
128. Sivakumar ANV, Li J, Scott S, Psota E, Jhala A, Luck J, Shi Y. Comparison of object detection and patch-based classification deep learning models on mid- to late-season weed detection in UAV imagery. Remote Sens. 2020. https://doi.org/10.3390/rs12132136.
129. Sladojevic S, Arsenovic M, Culibrk AAD, Stefanovic D. Deep neural networks based recognition of plant diseases by leaf image classification. Computl Intell Neurosci. 2016. https://doi.org/10.1155/2016/3289801.
130. Soma K, Bogaardt M, Poppe K, Wolfert S, Beers G, Urdu D, Kirova MP, Thurston C, Belles CM. Research for agri committee. impacts of the digital economy on the food chain and the cap. Policy department for structural and cohesion policies. European parliament. Brussels; 2019.
131. Song Q, Hu Q, Zhou Q, Hovis C, Xiang M, Tang H, Wu W. In-season crop mapping with GF-1/WFV data by combining object-based image analysis and random forest. Remote Sens. 2017. https://doi.org/10.3390/rs9111184.
132. Song X, Wang J, Huang W, Liu L, Yan G, Pu R. The delineation of agricultural management zones with high resolution remotely sensed data. Precis Agric. 2009;10:471–87. https://doi.org/10.1007/s11119-009-9108-2.

133. Speranza E, Ciferri R, Grego C, Vicente L. A cluster-based approach to support the delination of management zones in precision agriculture. In: IEEE 10 th International Conference on eScience. 2014. https://doi.org/10.1109/eScience.2014.42,
134. Speranza E, Ciferri R, Ciferri C. Clustering approaches and ensembles applied in the delineation of management classes in precision agriculture. In: Proceedings of the XVII GEOINFO, November 2016. Campos do Jordao; 2016. p. 27-30.
135. Stombaugh T, Shearer S. Equipment technologies for precision agriculture. J Soil Water Conserv. 2000;55:6–11.
136. Su J, Liu C, Coombes M, Hu X, Wang C, Xu X, Li Q, Chen LGW. Wheat yellow rust monitoring by learning from multispectral UAV aerial imagery. Comput Electron Agric. 2018;155:157–66. https://doi.org/10.1016/j.compag.2018.10.017.
137. Tagarakis A, Liakos V, Fountas S, Koundouras S, Gemtos T. Management zones delineation using fuzzy clustering techniques in grapevines. Prec Agric. 2013;14:18–39.
138. Taylor S, Veal M, Grift T, Mcdonald T, Corley F. Precision forestry-operational tactics for today and tomorrow. In: In: 25th annual Meeting of the council of Forest Engineers. Auburn: Auburn University; 2002.
139. Too E, Yujian L, Njuki S, Yingchun L. A comparative study of fine-tuning deep learning models for plant disease identification. Comput Electron Agric. 2019;161:272–9. https://doi.org/10.1016/j.compag.2018.03.032.
140. Tripathi R, Shahid ANM, Lal B, Gautam P, Raja R, Mohanty S, Kumar A, Panda B, Sahoo R. Delineation of soil management zones for a rice cultivated area in Eastern India using fuzzy clustering. Catena. 2015;133:128–36. https://doi.org/10.1016/j.rse.2016.03.010.
141. Vallentin C, Dobers E, Itzerott S, Kleinschmit B, Spengler D. Delineation of management zones with spatial data fusion and belief theory. Prec Agric. 2010;21:802–30. https://doi.org/10.1007/s11119-019-09696-0.
142. Vendrusculo L, Kaleita A. Modeling zone management in precision agriculture through fuzzy c-means technique at spatial database. In: Proceedings of the 2011 ASABE Annual International Meeting Sponsored by ASABE. Gault House, Louisville, Kentucky. August 7-10. 2016. p. 350–359. https://doi.org/10.13031/2013.38168
143. Veys C, Chatziavgerinos F, AlSuwaidi A, Hibbert J, Hansen M, Bernotas G, Smith M, Yin H, Rolfe S, Grieve B. Multispectral imaging for presymptomatic analysis of light leaf spot in oilseed rape. Plant Methods. 2019. https://doi.org/10.1186/s13007-019-0389-9.
144. Villa P, Bresciani M, Pinardi RBM, Giardino C. A rule-based approach for mapping macrophyte communities using multi-temporal aquatic vegetation indices. Remote Sens Environ. 2015;171:218–33. https://doi.org/10.1016/j.rse.2015.10.020.
145. Vrindts E, Mouazen A, Reyniers M, Maertens K, Maleki M, Ramon H, Baerdemaeker JD. Management zones based on correlation between soil compaction, yield and crop data. Biosyst Eng. 2005;92:419–28. https://doi.org/10.1016/j.biosystemseng.2005.08.010.
146. Wiseman L, Sanderson J, Zhang A, Jakku E. Farmers and their data: an examination of farmers' reluctance to share their data through the lens of the laws impacting smart farming. NJAS Wageningen J Life Sci. 2019. https://doi.org/10.1016/j.njas.2019.04.007.
147. Wolfert S, Sorensen C, Goense D. Precision forestry-operational tactics for today and tomorrow. In: Global Conference (SRII). San Jose: Annual SRII. IEEE; 2014. p. 266–73.
148. Wolfert S, Verdouw C, Bogaardt M. Big data in smart farming: a review. Agric Syst. 2017;153:69–80. https://doi.org/10.1016/j.agsy.2017.01.023.
149. Xue J, Su B. Significant remote sensing vegetation indices: a review of developments and applications. J Sensors. 2017. https://doi.org/10.1155/2017/1353691.
150. Yamamoto K, Togami T, Yamaguch N. Super-resolution of plant disease images for the acceleration of image-based phenotyping and vigor diagnosis in agriculture. Sensors. 2017. https://doi.org/10.3390/s17112557.
151. Yan L, Zhou S, Cifang W, Hongyi L, Feng L. Classification of management zones for precision farming in saline soil based on multi-data sources to characterize spatial variability of soil properties. Trans Chin Soc Agric Eng. 2007;23:84–9.
152. You J, Li X, Low M, Lobell D, Ermon S. Deep gaussian process for crop yield prediction based on remote sensing data. In: the Thirty-First AAAI Conference on Artificial Intelligence. AAAI Publications. 2017. p. 4559–4566.
153. Zan X, Zhang X, Xing Z, Liu W, Zhang X, Su W, Liu Z, Zhao Y, Li S. Automatic detection of maize tassels from UAV images by combining random forest classifier and VGG16. Remote Sens. 2020. https://doi.org/10.3390/rs12183049.
154. Zhang X, Shi L, Jia X, Seielstad G, Helgason C. Zone mapping application for precision farming: a decision support tool for variable rate application. Prec Agric. 2010;11:103–14. https://doi.org/10.1007/s11119-009-9130-4.
155. Zhang X, Han L, Dong Y, Shi Y, Huang W, Han L, Gonzalez-Moreno P, Ma H, Ye H, Sobeih T. A deep learning-based approach for automated yellow rust disease detection from high-resolution hyperspectral UAV images. Remote Sens. 2019. https://doi.org/10.3390/rs11131554.
156. Zheng Q, Huang W, Cui X, Shi Y, Liu L. New spectral index for detecting wheat yellow rust using sentinel-2 multispectral imagery. Sensors. 2018. https://doi.org/10.3390/s18030868.
157. Zhou Y, Luo J, Feng L, Zhou X. DCN-based spatial features for improving parcel-based crop classification using high-resolution optical images and multi-temporal SAR data. Remote Sens. 2019. https://doi.org/10.3390/rs11131619.

Publisher's Note