# A literature survey on the Use of Convolutional Neural Networks in Agriculture

Andreas Kamilaris and Francesc X. Prenafeta-Boldú

IRTA Torre Marimon, Institute for Food and Agricultural Research and Technology (IRTA)

Torre Marimon, Caldes de Montbui, Barcelona

**Abstract:** Deep learning constitutes a modern technique for image processing with large potential. As it has been successfully applied in various areas, it has recently entered also the domain of agriculture. In this paper, we have conducted a survey of research efforts that employ convolutional neural networks (CNN), which constitute a specific class of deep learning, applied to various agricultural and food production challenges. We examine the agricultural problems under study, models employed, sources of data used and the overall precision achieved according to the performance metrics used by the authors. We compare CNN with other existing techniques, and we list advantages and disadvantages of using CNN in agriculture. Moreover, we discuss the future potential of this technique, and we describe our personal experiences after employing CNN to approximate a problem of identifying missing vegetation from a sugar cane plantation in Costa Rica. Our overall findings indicate that CNN constitute a promising technique with high performance in terms of precision and classification accuracy, outperforming existing commonly-used image processing techniques. However, the success of each CNN model is highly dependent on the quality of the dataset used for training.

**Keywords:** Convolutional Neural Networks, Deep learning, Agriculture, Survey, Smart Farming, Food Systems.

## 1. Introduction

Smart farming (Tyagi, 2016) is important for tackling various challenges of agricultural production such as productivity, environmental impact, food security and sustainability (Gebbers & Adamchuk, 2010). As the global population is continuously growing (Kitzes, et al., 2008), a large increase of food production must be achieved as well (FAO, 2009). This increase of food production must be accompanied with protection of the natural ecosystems by means of using sustainable farming procedures. Food needs to maintain a high nutritional value while its security must be ensured around the world (Carvalho, 2006).

To address these challenges, the complex, multivariate and unpredictable agricultural ecosystems need to be better understood. This would be achieved by continuously monitoring, measuring and analyzing various physical aspects and phenomena. The deployment of new information and communication technologies (ICT) for small-scale crop/farm management and larger-scale ecosystems' observation will facilitate this task, enhancing management and decision/policy making by context, situation and location awareness (Kamilaris, A., Blasi, Torrellas, & Prenafeta-Boldú, 2017).

Emerging ICT technologies relevant for understanding agricultural ecosystems include remote sensing (Bastiaanssen, Molden, & Makin, 2000), Internet of Things (IoT) (Weber & Weber, 2010), cloud computing (Hashem, et al., 2015) and big data analysis (Chi, et al., 2016), (Kamilaris, Kartakoullis, & Prenafeta-Boldú, 2017). Remote sensing, by means of satellites, airplanes and unmanned aerial vehicles (UAV) (i.e. drones) provides large-scale snapshots of the agricultural environments. It has several advantages when applied to agriculture, being a well-known, non-destructive method to collect information about earth features. Remote sensing data may be obtained systematically over very large geographical areas, including zones inaccessible to human exploration. IoT uses advanced sensor technology to measure various parameters at the field while cloud computing is used for collection, storage, pre-processing and modeling of huge amounts of data coming from various, heterogeneous sources. Finally, big data analysis is used in combination with cloud computing for real-time, large-scale analysis of data stored on the cloud (Waga & Rabah, 2014), (Kamilaris, Gao, Prenafeta-Boldú, & Ali, 2016).

The four aforementioned technologies (remote sensing, IoT, cloud computing and big data analysis) could create novel applications and services that could improve agricultural productivity, increasing food security, e.g. by better understanding climatic conditions and changes.

A large subset of the volume of data collected through remote sensing and IoT involve images. Images can provide a complete picture of the agricultural fields, and image analysis could address a variety of challenges (Liaghat & Balasundram, 2010), (Ozdogan, Yang, Allez, & Cervantes, 2010). Hence, image analysis is an important research area in the agricultural domain and intelligent analysis techniques are used for image identification/classification, anomaly detection etc., in various agricultural applications (Teke, Deveci, Haliloğlu, Gürbüz, & Sakarya, 2013), (Saxena & Armstrong, 2014).

From those, the most common sensing method is satellite-based, using multi-spectral and hyperspectral imaging. Synthetic aperture radar (SAR), thermal and near infrared (NIR) cameras have been used in a lesser extent (Ishimwe, Abutaleb, & Ahmed, 2014), while optical and X-ray imaging have been applied in fruit and packaged food grading (Saxena & Armstrong, 2014). The most common techniques used for image analysis include machine learning (K-means, support vector machines (SVM) and artificial neural networks (ANN), amongst others), wavelet-based filtering, vegetation indices (NDVI) and regression analysis (Saxena & Armstrong, 2014).

Besides the aforementioned techniques, deep learning (DL) (LeCun, Bengio, & Hinton, 2015) is a modern approach with much potential and success in various domains where it has been employed (Najafabadi, et al., 2015), (Wan, et al., 2014). DL belongs to the research area of machine learning, and it is similar to ANN (Schmidhuber, 2015). However, DL constitutes a "*deeper*" neural network that provides a hierarchical representation of the data by means of various convolutions. This allows better learning capabilities in terms of capturing the full complexity of the real-life task under study, and thus the trained model can achieve higher classification accuracy.

In this survey we examine problems that employ a particular class of DL named convolutional neural networks (CNN), defined as deep, feed-forward artificial neural networks. CNN extend classical ANN by adding more "*depth*" into the network, as well as various convolutions that allow data representation in a hierarchical way (Schmidhuber, 2015), (LeCun & Bengio, 1995). CNN have been applied successfully in various visual imagery-related problems (Szegedy, et al., 2015).

The motivation for preparing this survey stems from the fact that CNN have been recently employed in agriculture with growing popularity and success. The fact that today exist more than 20 research efforts employing CNN for addressing various agricultural problems, encouraged the authors to prepare this survey. As CNN constitute probably the most popular and widely-used technique in agricultural research today, in problems related to image analysis, this survey focuses on this specific subset of DL models and techniques. A more general overview of the use of DL in agriculture is performed in (Kamilaris & Prenafeta-Boldú, 2018). To the authors' knowledge, this is the first survey in the agricultural domain that focuses on this particular practice, while a small number of more general surveys do exist (Deng & Yu, 2014), (Wan, et al., 2014), (Najafabadi, et al., 2015), presenting and analyzing related work in other research domains and application areas.

The aim of this research was to introduce the technique of CNN, as a promising and high-potential approach for addressing various challenges in agriculture related to computer vision. Besides analyzing the state of the art work at the field, a practical example of CNN applied in identifying missing vegetation based on aerial images has been presented in order to further illustrate the benefits and shortcomings of this technique.

## 2. Methodology

The bibliographic analysis involved two steps: a) collection of related work; and b) detailed review and analysis of this work.

In the first step, a keyword-based search for conference papers and articles was performed between August-September 2017 by the authors. Sources were the scientific databases IEEE Xplore and ScienceDirect, as well as the web scientific indexing services Web of Science and Google Scholar. As search keywords, the following query was used:

["deep learning" | "convolutional neural networks"] AND ["agriculture" OR "farming"]

In this way, we filtered out papers referring to CNN but not applied to the agricultural domain. From this effort, 27 papers were initially identified. Restricting the search for papers with appropriate application of the CNN technique and meaningful findings, the initial number of papers was reduced to 23. The following three criteria had been used to define appropriate application of CNN:

1. Use of CNN or CNN-based approach as the technique for addressing the problem under study.
2. Target some problem or challenge related to agriculture.
3. Show practical results by means of some well-defined performance metrics that indicate the success of the technique used.

As the authors of the related work under study used different performance metrics to evaluate their CNN models, we list and describe these metrics below:

- RMSE: Standard deviation of the differences between predicted values and observed values.
- F1 Score: The harmonic mean of precision and recall. For multi-class classification problems, F1 gets averaged among all the classes.

- QM: Obtained by multiplying sensitivity (proportion of pixels that were detected correctly) and specificity (which proportion of detected pixels are truly correct) (Douarre, Schielein, Frindel, Gerth, & Rousseau, 2016).
- RFC: Ratio of a predicted count of fruits by a CNN model, vs. the actual count performed offline by the authors or by experts (Rahnemoonfar & Sheppard, 2017), (Chen, et al., 2017).
- LC: A score related to the rank of the correct species in the list of retrieved species during the LifeCLEF 2015 Challenge (Reyes, Caicedo, & Camargo, 2015).

In the second step, the 23 papers selected from the first step were analyzed one-by-one, considering the following research questions: (a) which was the problem they addressed, (b) approach employed, (c) sources of data used and (d) which has been the overall precision. We also recorded: (e) whether the authors had compared their CNN-based approach with other techniques, and (f) which the difference in performance was. Examining how CNN perform in relation to other existing techniques was a critical aspect of this survey, as it would be the main indication of CNN effectiveness and performance. We note that it was difficult if not impossible to compare between different research works, due to the many assumptions and design selections made by the authors (i.e. datasets, performance metrics, CNN models, parameter setting etc.). Thus, we recorded the comparisons between CNN and other classical techniques (i.e. SVM, ANN, random forests etc.), *as performed by the authors at each research paper separately, using the same training/testing dataset and performance metric to compare with*.

## 3. Convolutional Neural Networks

In machine learning, CNN constitute a class of deep, feed-forward artificial neural networks that has successfully been applied to computer vision applications (Schmidhuber, 2015), (LeCun & Bengio, 1995).

In contrast to ANN whose training requirements in terms of time are impractical in some large-scale problems, CNN can learn complex problems particularly fast because of weight sharing and more complex models used, which allow massive parallelization (Pan & Yang, 2010). CNN can increase their probability of correct classifications, provided there are adequately large datasets (i.e. hundreds up to thousands of measurements, depending on the complexity of the problem under study) available for describing the problem. CNN consist of various convolutional, pooling and/or fully connected layers (Canziani, Paszke, & Culurciello, 2016). The convolutional layers act as feature extractors from the input images whose dimensionality is then reduced by the pooling layers, while the fully connected layers act as classifiers. Usually, at the last layer, the fully connected layers exploit the high-level features learned, in order to classify input images into predefined classes (Schmidhuber, 2015).

The highly hierarchical structure and large learning capacity of CNN models allow them to perform classification and predictions particularly well, being flexible and adaptable in a wide variety of complex challenges (Oquab, Bottou, Laptev, & Sivic, 2014).

CNN can receive any form of data as input, such as audio, video, images, speech, and natural language (Karpathy, et al., 2014), (Kim, 2014), (Abdel-Hamid, et al., 2014), (Kamilaris & Prenafeta-Boldú, 2017). CNN have been applied successfully by numerous organizations in various domains, such as the web (i.e. personalization systems, online chat robots), health (i.e. identification of diseases from MRI scans), disaster management (i.e. identifications of disasters by remote sensing images), post services (i.e. automatic reading of addresses), car industry (i.e. autonomous self-driving cars) etc.

An example CNN architecture (Simonyan & Zisserman, 2014) is displayed in Figure 1. As the figure shows, various convolutions are applied at some layers of the network, creating different representations of the learning dataset, starting from more general ones at the first larger layers and becoming more specific at the deeper layers. A combination of convolutional layers and dense layers tends to present good precision results.

There exist various "successful" popular architectures which researchers may use to start building their models instead of starting from scratch. These include AlexNet (Krizhevsky, Sutskever, & Hinton, 2012), VGG (Simonyan & Zisserman, 2014) (displayed in Figure 1), GoogleNet (Szegedy, et al., 2015) and Inception-ResNet (Szegedy, Ioffe, Vanhoucke, & Alemi, 2017). Each architecture has different advantages and scenarios where it is more appropriately used (Canziani, Paszke, & Culurciello, 2016). It is also worth noting that almost all of the aforementioned architectures come along with their weights pre-trained, which means that their network has already been trained by some dataset and has thus learned to provide accurate recognition for some particular problem domain (Pan & Yang, 2010). Common datasets used for pre-training DL architectures include ImageNet (Deng, et al., 2009) and PASCAL VOC (PASCAL VOC Project, 2012)

Moreover, there are various tools and platforms that allow researchers to experiment with DL (Bahrampour, Ramakrishnan, Schott, & Shah, 2015). The most popular ones are Theano, TensorFlow, Keras (which is an API on top of Theano and TensorFlow), Caffe, PyTorch, TFLearn, Pylearn2 and the Deep Learning Matlab Toolbox. Some of these tools (i.e. Theano, Caffe) incorporate popular architectures such as the ones mentioned above (i.e. AlexNet, VGG, GoogleNet), either as libraries or classes.

## 4. CNN Applications in Agriculture

In Appendix I, we list the identified relevant works, indicating the particular problem they address, the agricultural area involved, sources of data used, overall precision achieved, details over the CNN-based implementation as well as comparisons with other techniques, wherever available.

### 4.1 Areas of Use

Twelve areas have been identified in total, with the popular ones being plant and leaf-based disease detection (3 papers), land cover classification (3 papers), plant recognition (3 papers), fruits counting (4 papers) and weed identification (3 papers).

It is remarkable that all papers have been published after 2014, indicating how recent and modern this technique is, in the domain of agriculture. More precisely, six papers have been published in 2017, ten in 2016, six in 2015 and one in 2014.

The large majority of these papers deal with image classification and identification of areas of interest, including detection of obstacles (Steen, Christiansen, Karstoft, & Jørgensen, 2016), (Christiansen, Nielsen, Steen, Jørgensen, & Karstoft, 2016) and fruit counting (Rahnemoonfar & Sheppard, 2017), (Sa, et al., 2016), while some other papers focus on predicting future values such as corn yield (Kuwata & Shibasaki, 2015) and soil moisture content at the field (Song, et al., 2016).

From another perspective, most papers (19) target crops, while few works consider the issues of land cover (3 papers) and livestock agriculture (1 paper).

**4.2 Data Sources**

Observing the sources of data used to train the CNN model at each paper, large datasets of images are mainly used, containing in some cases thousands of images (e.g. (Mohanty, Hughes, & Salathé, 2016), (Reyes, Caicedo, & Camargo, 2015)). From these image datasets, some originate from well-known and publicly-available datasets such as PlantVillage, LifeCLEF, MalayaKew and UC Merced, while others are produced by the authors for their research needs (Sladojevic, Arsenovic, Anderla, Culibrk, & Stefanovic, 2016), (Rahnemoonfar & Sheppard, 2017), (Bargoti & Underwood, 2016), (Xinshao & Cheng, 2015), (Sørensen, Rasmussen, Nielsen, & Jørgensen, 2017). Papers dealing with land cover and crop type classification employ a smaller number of images (e.g. 10-100 images), produced by UAV (Lu, et al., 2017) or satellite-based remote sensing (Kussul, Lavreniuk, Skakun, & Shelestov, 2017). A particular paper investigating segmentation of root and soil uses images from X-ray tomography (Douarre, Schielein, Frindel, Gerth, &

Rousseau, 2016). Moreover, some projects use historical text data, collected either from repositories (Kuwata & Shibasaki, 2015) or field sensors (Song, et al., 2016). In general, the more complicated the problem to be solved (e.g. large number of classes to identify), the more data is required.

### 4.3 Performance Metrics and Overall Precision

Various performance metrics have been employed by the authors, with percentage of correct predictions (CA) on the validation or testing dataset being the most popular one (16 papers, 69%). Others include RMSE (3 papers), F1 Score (3 papers), Quality Measure (QM) (Douarre, Schielein, Frindel, Gerth, & Rousseau, 2016), ratio of total fruits counted (RFC) (Chen, et al., 2017) and the LifeCLEF metric (LC) (Reyes, Caicedo, & Camargo, 2015). The aforementioned metrics have been defined in Section 2.

The majority of related work employs CA, which is generally high (i.e. above 90%), indicating the successful application of CNN in various agricultural problems. The highest CA statistics have been observed in the works of (Chen, Lin, Zhao, Wang, & Gu, 2014), (Lee, Chan, Wilkin, & Remagnino, 2015) and (Steen, Christiansen, Karstoft, & Jørgensen, 2016), with accuracies of 98% or more.

### 4.4 Technical Details

From a technical side, almost half of the research works (12 papers) employed popular CNN architectures such as AlexNet, VGG and Inception-ResNet. The other half (11 papers) experimented with their own architectures, some combining CNN with other techniques, such as principal component analysis (PCA) and logistic regression (Chen, Lin, Zhao, Wang, & Gu, 2014), SVM (Douarre, Schielein, Frindel, Gerth, & Rousseau, 2016), linear regression (Chen, et al., 2017), LMC classifiers (Xinshao & Cheng, 2015) and macroscopic cellular automata (Song, et al., 2016).

Regarding the frameworks used, all the studies that employed some well-known architecture had also used a DL framework, with Caffe being the most popular (10 papers, 43%), followed by deeplearning4j (1 paper), and Tensor Flow (1 paper). Five research works developed their own software, while some authors built their own models on top of Theano (3 papers), Pylearn2 (1 paper), MatConvNet (1 paper) and Deep Learning Matlab Toolbox (1 paper). A possible reason for the wide use of Caffe is that it incorporates various CNN frameworks and datasets which can be then used easily by the users.

It is worth-mentioning that some of the related works that possessed only small datasets to train their CNN models (Bargoti & Underwood, 2016), (Sladojevic, Arsenovic, Anderla, Culibrk, & Stefanovic, 2016), (Sørensen, Rasmussen, Nielsen, & Jørgensen, 2017), exploited data augmentation techniques (Krizhevsky, Sutskever, & Hinton, 2012) to enlarge artificially the number of training images using label-preserving transformations, such as translations, transposing, reflections and altering the intensities of the RGB channels.

Furthermore, the majority of related work included some image pre-processing steps, where each image in the dataset was resized to a smaller size, before being used as input to the model, such as 256x256, 128x128, 96x96, 60x60 pixels, or converted to grayscale (Santoni, Sensuse, Arymurthy, & Fanany, 2015). Most of the studies divided their data randomly between training and testing/verification sets, using a ratio of 80-20 or 90-10 respectively. Also, various learning rates have been recorded, from 0.001 (Amara, Bouaziz, & Algergawy, 2017) and 0.005 (Mohanty, Hughes, & Salathé, 2016) up to 0.01 (Grinblat, Uzal, Larese, & Granitto, 2016). Learning rate is about how quickly a network learns. Higher values help avoiding being stuck in local minima. A general approach used by many of the evaluated papers is to start out with a high learning rate and lower it as the training goes on. The learning rate is very dependent on the network architecture.

Finally, most of the research works that incorporated popular CNN architectures took advantage of *transfer learning* (Pan & Yang, 2010), which leverages the already existing knowledge of some related task in order to increase the learning efficiency of the problem under study, by fine-tuning pre-trained models. When it is not possible to train a network from scratch due to having a small training dataset or addressing a quite complex problem, it is useful that the network is initialized with weights from another pre-trained model. Pre-trained CNN are models that have already been trained on some relevant dataset with possibly different number of classes. These models are then adapted to the particular challenge and dataset. This method was followed in (Lu, et al., 2017), (Douarre, Schielein, Frindel, Gerth, & Rousseau, 2016), (Reyes, Caicedo, & Camargo, 2015), (Bargoti & Underwood, 2016), (Steen, Christiansen, Karstoft, & Jørgensen, 2016), (Lee, Chan, Wilkin, & Remagnino, 2015), (Sa, et al., 2016), (Mohanty, Hughes, & Salathé, 2016), (Christiansen, Nielsen, Steen, Jørgensen, & Karstoft, 2016), (Sørensen, Rasmussen, Nielsen, & Jørgensen, 2017), for the VGG16, DenseNet, AlexNet and GoogleNet architectures.

**4.5 Performance Comparison with Other Approaches**

The eighth column of Appendix I presents whether the authors of related work compared their CNN-based approach with other techniques used for solving their problem under study. CNN scored 1-4% better CA in comparison to SVM (Chen, Lin, Zhao, Wang, & Gu, 2014), (Lee, Chan, Wilkin, & Remagnino, 2015), (Grinblat, Uzal, Larese, & Granitto, 2016), 3-11% improved CA in comparison to unsupervised feature learning (Luus, Salmon, van den Bergh, & Maharaj, 2015), 2-44% better CA in relation to local shape and color features (Dyrmann, Karstoft, & Midtiby, 2016 ), (Sørensen, Rasmussen, Nielsen, & Jørgensen, 2017), 2% better CA (Kussul, Lavreniuk, Skakun, & Shelestov, 2017) or 18% less RMSE (Song, et al., 2016) compared to multilayer perceptrons.

Moreover, CNN achieved 6% higher CA than random forests (Kussul, Lavreniuk, Skakun, & Shelestov, 2017), 2% better CA than Penalized Discriminant Analysis (Grinblat, Uzal, Larese, & Granitto, 2016), 41% improved CA when compared to ANN (Lee, Chan, Wilkin, & Remagnino, 2015) and 24% less RMSE compared to Support Vector Regression

(Kuwata & Shibasaki, 2015).

Furthermore, CNN reached 25% better RFC than an area-based technique (Rahnemoonfar & Sheppard, 2017), 30% higher RFC than the best texture-based regression model (Chen, et al., 2017), 84.3% better F1 score in relation to an algorithm based on local decorrelated channel features and 3% higher CA compared to a Gaussian Mixture Model (GMM) (Santoni, Sensuse, Arymurthy, & Fanany, 2015).

In only one case (Reyes, Caicedo, & Camargo, 2015), CNN showed worse performance than other techniques. This was against a technique involving local descriptors to represent images together with KNN as classification strategy (20% lower LC).

## 5. Discussion

Our analysis has shown that CNN offer superior performance in terms of precision in the vast majority of related work, based on the performance metrics employed by the authors, with GMM being a technique with comparable performance in some cases (Santoni, Sensuse, Arymurthy, & Fanany, 2015), (Reyes, Caicedo, & Camargo, 2015). Although our work under study is relatively small, we can speculate that in most of the agricultural challenges a satisfactory precision has been observed, especially in comparison to other techniques employed to solve the same problem. This indicates a successful application of CNN in various agricultural domains. In particular, the areas of plant and leaf disease detection, plant recognition, land cover classification, fruit counting and identification of weeds belong to the categories where the highest precision has been observed.

Although CNN have been associated with computer vision and image analysis (which is also the general case in this survey), we have observed two related works where CNN-based models have been trained based on field sensory data (Kuwata & Shibasaki, 2015) and a combination of static and dynamic environmental variables (Song, et al., 2016). In both cases, the performance (i.e. RMSE) was better than other techniques under study.

When comparing performance in terms of precision/accuracy, it is of paramount importance to adhere to same experimental conditions, i.e. datasets and performance metrics (when comparing CNN with other techniques), as well as architectures and model parameters (when comparing studies both employing CNN). From the related work under study, 15 out of the 23 papers (65%) performed direct, valid and correct comparisons among CNN and other commonly-used techniques. Hence, we suggest that some of the findings in Section 4.5 need to be considered with caution.

### 5.1 Advantages/Disadvantages of Convolutional Neural Networks

Except from the improvements in precision observed in the classification/prediction problems at the surveyed works, there are some other important advantages of using CNN in image processing. Previously, traditional approaches for image classification tasks had been based on hand-engineered features, whose performance and accuracy affected heavily the overall results. Feature engineering (FE) is a complex, time-consuming process which needs to be altered whenever the problem or the dataset changes. Thus, FE constitutes an expensive effort that depends on experts' knowledge and does not generalize well (Amara, Bouaziz, & Algergawy, 2017). On the other hand, CNN do not require FE, as they locate the important features automatically through the training process. Quite impressively, in the case of fruit counting, the model learned explicitly to count (Rahnemoonfar & Sheppard, 2017). CNN seem to generalize well (Pan & Yang, 2010) and they are quite robust even under challenging conditions such as illumination, complex background, size and orientation of the images, and different resolution (Amara, Bouaziz, & Algergawy, 2017).

Their main disadvantage is that CNN take sometimes much longer time to train. However, after training, their testing time efficiency is much faster than other methods like SVM or KNN (Chen, Lin, Zhao, Wang, & Gu, 2014), (Christiansen, Nielsen, Steen, Jørgensen, &

Karstoft, 2016). Another important disadvantage (see also Section 5.2) is the need for large datasets (i.e. hundreds or thousands of images), and their proper annotation, which is sometimes a delicate procedure that must be performed by domain experts. Our personal experimentation with CNN (see Section 5.4) reveals this problem of poor data labeling, which could create significant reduction in performance and precision achieved. Other disadvantages include problems that might occur when using pre-trained models on similar and smaller datasets (i.e. a few hundreds of images or less), optimization issues because of the models' complexity, as well as hardware restrictions.

### 5.2 Dataset Requirements

A considerable barrier in the use of CNN is the need for large datasets, which would serve as the input during the training procedure. In spite of data augmentation techniques, which could augment some dataset with label-preserving transformations, in reality at least some hundreds of images are required, depending on the complexity of the problem under study. In the domain of agriculture, there do not exist many publicly-available datasets for researchers to work with, and in many cases researchers need to manually develop their own sets of images. This could require many hours or days of work.

Researchers working with remote sensing data have more options, because of the availability of images provided by satellites such as MERIS, MODIS, AVHRR, RapidEye, Sentinel, Landsat etc. These datasets contain multi-temporal, multi-spectral and multi-source images that could be used in problems related to land and crop cover classification.

### 5.3 Future of Deep Learning in Agriculture

Our survey has shown that only 12 agriculture-related problems (see Section 4.1) have been approximated by CNN. It is interesting to see how CNN would behave also in other agricultural-related problems, such as crop phenology, seeds identification, soil and leaf nitrogen content, irrigation, plants' water stress detection, water erosion assessment, pest detection and herbicide use, identification of contaminants, diseases or defects on food, crop hail damage and greenhouse monitoring. Intuitively, since many of the aforementioned research areas employ data analysis techniques with similar concepts and comparable performance to CNN (i.e. linear and logistic regression, SVM, KNN, K-means clustering, Wavelet-based filtering, Fourier transform), then it would be worth to examine the applicability of CNN to these problems too.

Other possible application areas could be the use of aerial imagery (i.e. by means of drones) to monitor the effectiveness of the seeding process, increase the quality of wine production by harvesting grapes at the right moment for best maturity levels, monitor animals and their movements to consider their overall welfare and identify possible diseases, and many other scenarios where computer vision is involved.

As noted before, all the papers considered in this survey made use of basic CNN architectures, which constitute only a specific, simpler category of DL models. Our study did not consider/include more advanced and complex models such as Recurrent Neural Networks (RNN) (Mandic & Chambers, 2001) or Long Short-Term Memory (LSTM) architectures (Gers, Schmidhuber, & Cummins, 2000). These architectures tend to exhibit dynamic temporal behavior, being able to remember (i.e. RNN) and also to forget after some time or when needed (i.e. LSTM). An example application could be to estimate the growth of plants, trees or even animals based on previous consecutive observations, to predict their yield, assess their water needs or prevent diseases from occurring. These models could find applicability in environmental informatics too, for understanding climatic change, predicting weather conditions and phenomena, estimating the environmental impact of various physical or artificial processes etc.

### 5.4 Personal Experiences from a Small Study

To better understand the capabilities and effectiveness of CNN, we performed a small experiment, addressing a problem not touched upon by related work, that of identifying missing vegetation from a crop field, as shown in Figure 2. As depicted in the figure, areas labeled as (1) are examples of sugar cane plants, while areas (2) are examples of soil. Areas labeled as (3) constitute examples of missing vegetation, i.e. it is soil while it should have been sugar cane. Finally, areas (4) are examples of irrelevant image segments.

We used a dataset of aerial photos from a sugar cane plantation in Costa Rica, prepared by the company INDIGO Inteligencia Agrícola[1]. Data has been captured by a drone in May, 2017 from a geographical area of three hectares based on a single crop field. We split this data into 1,500 80x80 images, and then experts working at INDIGO annotated (labelled) every image as "*Sugar cane*", "*Soil*" or "*Other*", where the latter could be anything else in the image such as fences, irrigation infrastructures, pipes etc. The VGG architecture (Simonyan & Zisserman, 2014) was used on the Keras/Theano platform. Data was split randomly into training and testing datasets, using a ratio of 85-15. To accelerate training, a pre-trained VGG model was used, based on the ImageNet dataset (Deng, et al., 2009). Our results are presented at the confusion matrix in Figure *3*. The VGG model has achieved a CA of 79.20%, after being trained for 15 epochs at a time duration of 11 hours using a desktop PC with an Intel Core i7 quad-processor@2GHz and 6GB of RAM. Trying to investigate the sources of the relatively large error (in comparison to related work under study as listed in Appendix I), we observed, together with experts from INDIGO, the following (main) labelling issues, illustrated in Figure 4:

- Some images were mislabeled as "Other" while they actually represented "Sugar cane" snapshots (4% of the 20.80% total error).
- Some images mislabeled as "Other" while they represented "Soil" (8% of the error).
- Some images mislabeled as "Soil" while being "Sugar cane" (2% of the error).

Based on the above, the error would have been reduced from 20.80% to 6-8% by more careful labeling. This experiment emphasizes the importance of proper labeling of the training dataset, otherwise CA can deteriorate significantly. Sometimes, as this experiment

[1] INDIGO Inteligencia Agrícola. https://www.indigoia.com/

revealed, the annotation procedure is not trivial because there are images which could belong to multiple labels (see Figure *4*). In particular, labeling of images as “Soil” or “Sugar cane” is sometimes really difficult, because it depends on the percentage of plants or more generally vegetation in the image. Increasing image size makes the problem even worse, as labelling ambiguity increases. Thus, the variation among the classes of the training dataset is also an important parameter that affects the learning efficiency of the model. Perhaps a solution for future work would be to automate labeling by means of the use of a vegetation index, together with orientation or organization of the plants in the image, which could indicate a pattern of plantation or the soil in between. Still with these constraints, this experiment indicates that CNN can constitute a reliable technique for addressing this particular problem. The development of the model and its learning process do not require much time, as long as the dataset is properly prepared and correctly labeled.

Our aim is that this survey would motivate more researchers to experiment with CNN and deep learning in general, applying them for solving various agricultural problems involving classification or prediction, related not only to computer vision and image analysis, but more generally to data analysis. The overall benefits of CNN are encouraging for their further use towards smarter, more sustainable farming and more secure food production.

## 6. Conclusion

Our findings indicate that CNN reached high precision in the large majority of the problems they have been used, scoring higher precision than other popular image processing techniques. Their main advantages are the ability to approximate highly complex problems quite effectively, and that they do not need feature engineering before-hand. Our personal experiences after employing CNN to approximate a problem of identifying missing vegetation from a sugar cane plantation in Costa Rica revealed that the successful application of CNN is highly dependent on the size and quality of the dataset used for training the model, in terms of variance among the classes and labeling accuracy.

For future work, we plan to apply the general concepts and best practices of CNN, as described through this survey, to other areas of agriculture where this modern technique has not yet been adequately used. Some of these areas have been identified in the discussion section.

## Acknowledgments

This research has been supported by the P-SPHERE project, which has received funding from the European Union's Hori- zon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No 665919.

## References

Abdel-Hamid, O., Mohamed, A. R., Jiang, H., Deng, L., Penn, G., & Yu, D. (2014). Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on audio, speech, and language processing, 22*(10), 1533-1545.

Amara, J., Bouaziz, B., & Algergawy, A. (2017). A Deep Learning-based Approach for Banana Leaf Diseases Classification. (págs. 79-88). Stuttgart: BTW workshop.

Bahrampour, S., Ramakrishnan, N., Schott, L., & Shah, M. (2015). Comparative study of deep learning software frameworks. *arXiv preprint arXiv, 1511*(06435).

Bargoti, S., & Underwood, J. (2016). Deep Fruit Detection in Orchards. *arXiv preprint arXiv, 1610*(03677).

Bastiaanssen, W., Molden, D., & Makin, I. (2000). Remote sensing for irrigated agriculture: examples from research and possible applications. *Agricultural water management, 46*(2), 137-155.

Canziani, A., Paszke, A., & Culurciello, E. (2016). An Analysis of Deep Neural Network Models for Practical Applications. *arXiv preprint arXiv, 1605*(07678).

Carvalho, F. P. (2006). Agriculture, pesticides, food security and food safety. *Environmental science & policy, 9*(7-8), 685-692.

Chen, S. W., Shivakumar, S. S., Dcunha, S., Das, J., Okon, E., Qu, C., & Kumar, V. (2017). Counting Apples and Oranges With Deep Learning: A Data-Driven Approach. *IEEE Robotics and Automation Letters, 2*(2), 781-788.

Chen, Y., Lin, Z., Zhao, X., Wang, G., & Gu, Y. (2014). Deep Learning-Based Classification of Hyperspectral Data. *IEEE Journal of Selected topics in applied earth observations and remote sensing, 7*(6), 2094-2107.

Chi, M., Plaza, A., Benediktsson, J. A., Sun, Z., Shen, J., & Zhu, Y. (2016). Big data for remote sensing: challenges and opportunities. *Proceedings of the IEEE, 104*(11), 2207-2219.

Christiansen, P., Nielsen, L. N., Steen, K. A., Jørgensen, R. N., & Karstoft, H. (2016). DeepAnomaly: Combining Background Subtraction and Deep Learning for Detecting Obstacles and Anomalies in an Agricultural Field. *Sensors , 16*(11), 1904.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. (págs. 248-255). Miami, FL, USA: IEEE Conference on

Computer Vision and Pattern Recognition (CVPR).
Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and Trends in Signal Processing, 7*(3-4), 197-387.
Douarre, C., Schielein, R., Frindel, C., Gerth, S., & Rousseau, D. (2016). Deep learning based root-soil segmentation from X-ray tomography. *bioRxiv, 071662*.
Dyrmann, M., Karstoft, H., & Midtiby, H. S. (2016 ). Plant species classification using deep convolutional neural network. *Biosystems Engineering, 151*, 72-80.
FAO. (2009). *How to Feed the World in 2050.* Rome: Food and Agriculture Organization of the United Nations.
Gebbers, R., & Adamchuk, V. I. (2010). Precision agriculture and food security. *Science, 327*(5967), 828-831.
Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation, 12*(10), 2451-2471.
Grinblat, G. L., Uzal, L. C., Larese, M. G., & Granitto, P. M. (2016). Deep learning for plant identification using vein morphological patterns. *Computers and Electronics in Agriculture, 127*, 418-424.
Hashem, I., Yaqoob, I., Anuar, N., Mokhtar, S., Gani, A., & Khan, S. (2015). The rise of “big data” on cloud computing: Review and open research issues. *Information Systems, 47*, 98-115.
Ishimwe, R., Abutaleb, K., & Ahmed, F. (2014). Applications of thermal imaging in agriculture—A review. *Advances in Remote Sensing, 3*(3), 128.
Kamilaris, A., & Prenafeta-Boldú, F. (April de 2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture, 147*(1), 70-90.
Kamilaris, A., & Prenafeta-Boldú, F. X. (2017). Disaster Monitoring using Unmanned Aerial Vehicles and Deep Learning. *Disaster Management for Resilience and Public Safety Workshop, in Proc. of EnviroInfo 2017.* Luxembourg.
Kamilaris, A., A., A., Blasi, A., Torrellas, M., & Prenafeta-Boldú, F. (2017). Estimating the Environmental Impact of Agriculture by means of Geospatial and Big Data Analysis: The Case of Catalonia. *Proc. of EnviroInfo, From Science to Society* (págs. 39-48). Luxembourg: Springer, Cham.
Kamilaris, A., Gao, F., Prenafeta-Boldú, F. X., & Ali, M. I. (2016). Agri-IoT: A semantic framework for Internet of Things-enabled smart farming applications. *3rd World Forum on Internet of Things (WF-IoT)* (págs. 442-447). Reston, VA, USA: IEEE.
Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. *Computers and Electronics in Agriculture, 143*(1), 23-37.
Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. *Proc. of the Conference on Computer Vision and Pattern Recognition (CVPR)* (págs. 1725-1732). Columbus, OH, USA: IEEE .
Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint.* arXiv:1408.5882.
Kitzes, J., Wackernagel, M., Loh, J., Peller, A., Goldfinger, S., Cheng, D., & Tea, K. (2008). Shrink and share: humanity's present and future Ecological Footprint. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 363*(1491), 467-475.
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 1097-1105.
Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geoscience and Remote Sensing Letters, 14*(5), 778-782.
Kuwata, K., & Shibasaki, R. (2015). Estimating crop yields with deep learning and remotely sensed data. (págs. 858-861). Milan, Italy: IEEE International Geoscience and Remote Sensing Symposium (IGARSS).

LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks, 3361*(10).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444.

Lee, S. H., Chan, C. S., Wilkin, P., & Remagnino, P. (2015). Deep-plant: Plant identification with convolutional neural networks. (págs. 452-456). Quebec city, Canada: IEEE International Conference on Image Processing (ICIP).

Liaghat, S., & Balasundram, S. K. (2010). A review: The role of remote sensing in precision agriculture. *American journal of agricultural and biological sciences, 5*(1), 50-55.

Lu, H., Fu, X., Liu, C., Li, L. G., He, Y. X., & Li, N. W. (2017). Cultivated land information extraction in UAV imagery based on deep convolutional neural network and transfer learning. *Journal of Mountain Science, 14*(4), 731-741.

Luus, F. P., Salmon, B. P., van den Bergh, F., & Maharaj, B. T. (2015). Multiview deep learning for land-use classification. *IEEE Geoscience and Remote Sensing Letters, 12*(12), 2448-2452.

Mandic, D. P., & Chambers, J. A. (2001). *Recurrent neural networks for prediction: learning algorithms, architectures and stability.* New York: John Wiley.

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in plant science, 7*.

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data, 2*(1), 1.

Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and transferring mid-level image representations using convolutional neural networks. *Conference on Computer Vision and Pattern Recognition (CVPR)* (págs. 1717-1724.). Columbus, OH, USA: IEEE .

Ozdogan, M., Yang, Y., Allez, G., & Cervantes, C. (2010). Remote sensing of irrigated agriculture: Opportunities and challenges. *Remote sensing, 2*(9), 2274-2304.

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering , 22*(10), 1345-1359.

PASCAL VOC Project. (2012). *The PASCAL Visual Object Classes*. Obtenido de http://host.robots.ox.ac.uk/pascal/VOC/

Rahnemoonfar, M., & Sheppard, C. (2017). Deep Count: Fruit Counting Based on Deep Simulated Learning. *Sensors, 17*(4), 905.

Reyes, A. K., Caicedo, J. C., & Camargo, J. E. (2015). Fine-tuning Deep Convolutional Networks for Plant Recognition. Toulouse: CLEF (Working Notes).

Sørensen, R. A., Rasmussen, J., Nielsen, J., & Jørgensen, R. (2017). Thistle detection using convolutional neural networks. Montpellier, France: EFITA Congress.

Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., & McCool, C. (2016). Deepfruits: A fruit detection system using deep neural networks. *Sensors, 16*(8), 1222.

Santoni, M. M., Sensuse, D. I., Arymurthy, A. M., & Fanany, M. I. (2015). Cattle Race Classification Using Gray Level Co-occurrence Matrix Convolutional Neural Networks. *Procedia Computer Science, 59*, 493-502.

Saxena, L., & Armstrong, L. (2014). A survey of image processing techniques for agriculture. Perth, Australia: Proceedings of Asian Federation for Information Technology in Agriculture, Australian Society of Information and Communication Technologies in Agriculture.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks, 61*, 85-117.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv, 1409*(1556).

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Computational intelligence and neuroscience, 2016*.

Song, X., Zhang, G., Liu, F., Li, D., Zhao, Y., & Yang, J. (2016). Modeling spatio-temporal distribution of soil moisture by deep learning-based cellular automata model. *Journal of Arid Land, 8*(5), 734-748.

Steen, K. A., Christiansen, P., Karstoft, H., & Jørgensen, R. N. (2016). Using Deep Learning to Challenge Safety Standard for Highly Autonomous Machines in Agriculture. *Journal of Imaging, 2*(1), 6.

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. (págs. 4278-4284). AAAI.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., & Rabinovich, A. (2015). Going deeper with convolutions. (págs. 1-9). Boston, MA, USA: IEEE conference on computer vision and pattern recognition.

Teke, M., Deveci, H. S., Haliloğlu, O., Gürbüz, S. Z., & Sakarya, U. (2013). A short survey of hyperspectral remote sensing applications in agriculture. Istanbul, Turkey: 6th International Conference on Recent Advances in Space Technologies (RAST), IEEE.

Tyagi, A. C. (2016). Towards a Second Green Revolution. *Irrigation and Drainage, 65*(4), 388-389.

Waga, D., & Rabah, K. (2014). Environmental conditions' big data management and cloud computing analytics for sustainable agriculture. *World Journal of Computer Application and Technology, 2*(3), 73-81.

Wan, J., Wang, D., Hoi, S. C., Wu, P., Zhu, J., Zhang, Y., & Li, J. (2014). Deep learning for content-based image retrieval: A comprehensive study. (págs. 157-166). Orlando, FL: Proceedings of the 22nd ACM international conference on Multimedia, ACM.

Weber, R. H., & Weber, R. (2010). *Internet of Things* (Vol. 12). New York, NY, USA: Springer.

Xinshao, W., & Cheng, C. (2015). Weed seeds classification based on PCANet deep learning baseline. (págs. 408-415). IEEE Signal and Information Processing Association Annual Summit and Conference (APSIPA).

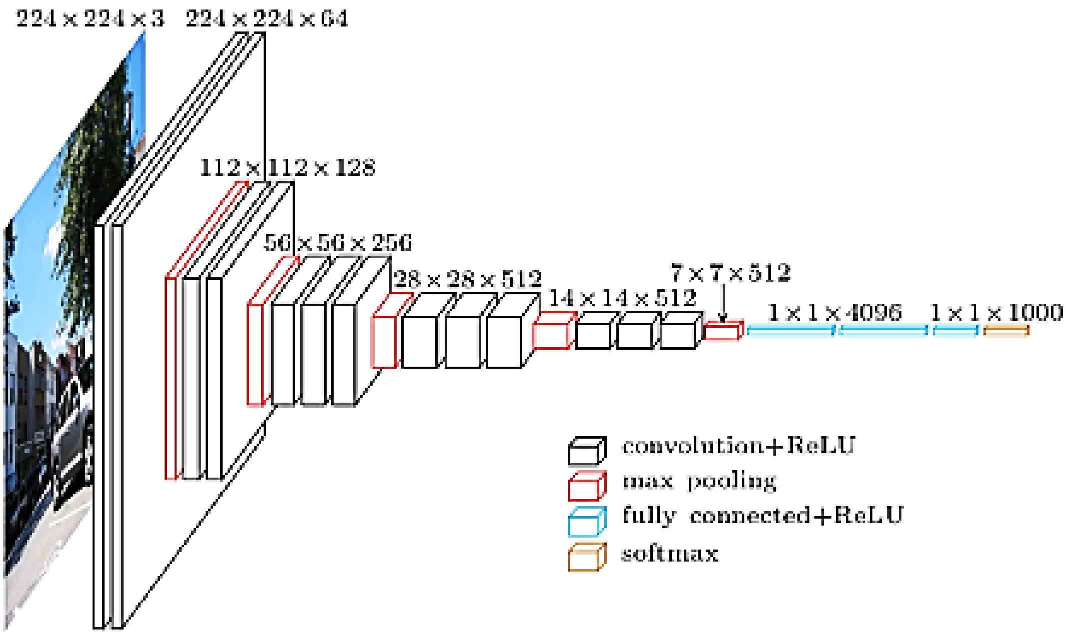


Figure 1: An example CNN architecture (VGG *(Simonyan & Zisserman, 2014)*).


Figure 2: Identification of missing vegetation from a crop field. Areas labeled as (1) represent examples of sugar cane plants while areas labeled as (2) constitute examples of soil. Areas (3) depict missing vegetation examples, i.e. it should have been sugar cane but it is soil. Finally, areas (4) are examples of “Others”, being irrelevant image segments.

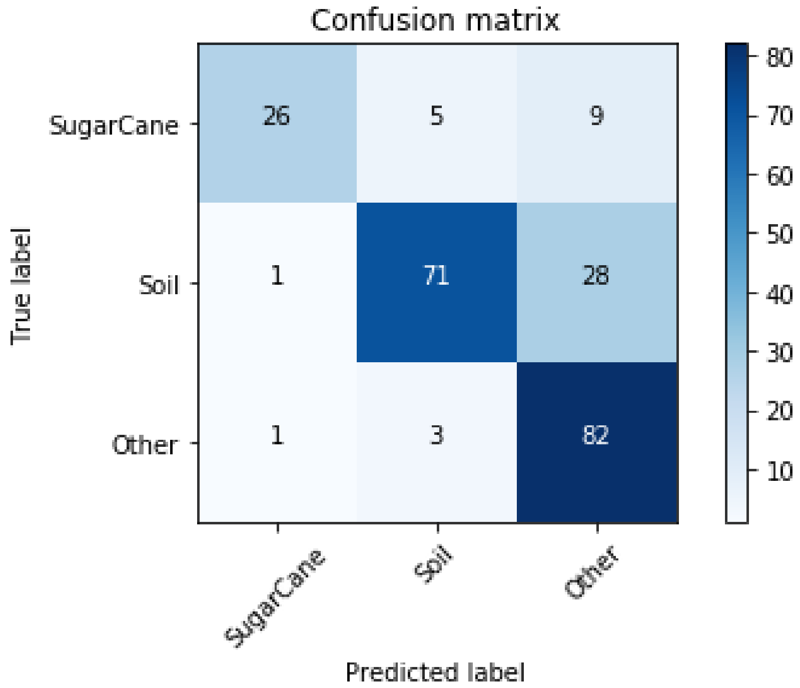


Figure 3: Confusion matrix of the results of identifying missing vegetation from a crop field.

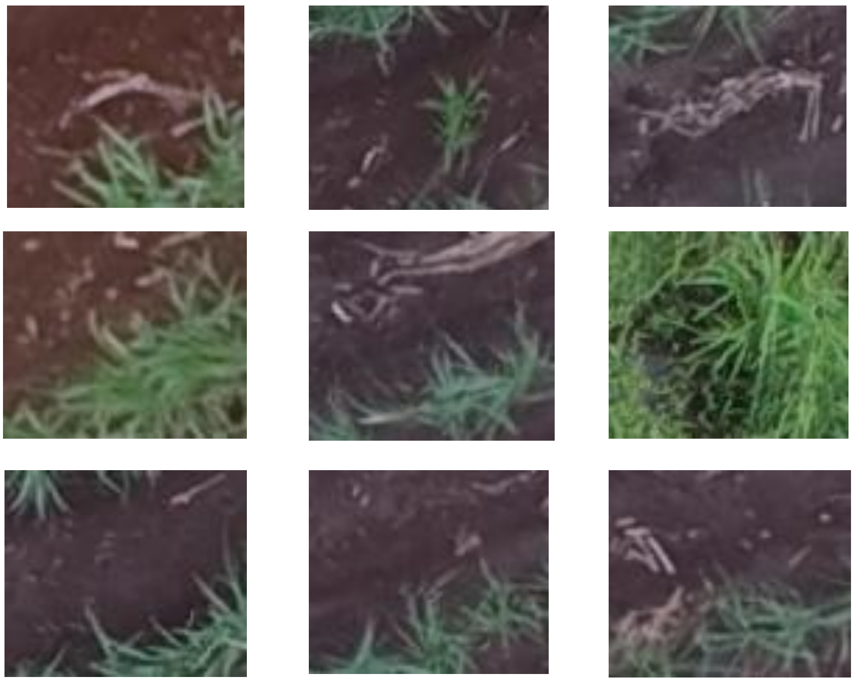

Figure 4: Incorrect labels of the vegetation dataset. Examples of images mislabeled as "Other" while should have been labeled "Soil" (top). Examples of images mislabeled as "Other" while should have been labeled "Sugar Cane" (middle). Examples of images mislabeled as "Soil" while they represented "Sugar Cane" (bottom).

Appendix I: Applications of deep learning in agriculture.

| No. | Agri Area | Problem Description | Data Used | Precision | DL Model Used | Frame work Used | Comparison with other technique | Ref. |
|---|---|---|---|---|---|---|---|---|
| 1. | Leaf disease detection | 13 different types of plant diseases out of healthy leaves | Authors-created database containing 4,483 images | 96.30% (CA) | CaffeNet | Caffe | Better results than SVM (no more details) | (Sladojevic, Arsenovic, Anderla, Culibrk, & Stefanovic, 2016) |
| 2. | Plant disease detection | Identify 14 crop species and 26 diseases | PlantVillage public dataset of 54,306 images of diseased and healthy plant leaves | 0.9935 (F1) | AlexNet, GoogleNet | Caffe | Substantial margin in standard benchmarks with approaches using hand-engineered features | (Mohanty, Hughes, & Salathé, 2016) |
| 3. | | Classify banana leaves' diseases | Dataset of 3,700 images of banana diseases obtained from the PlantVillage dataset | 96%+ (CA), 0.968 (F1) | LeNet | deeplearning4j | Methods using hand-crafted features not generalize well | (Amara, Bouaziz, & Algergawy, 2017) |
| 4. | Land cover classification | Identify 13 different land-cover classes in KSC and 9 different classes in Pavia | A mixed vegetation site over Kennedy Space Center (KSC), FL, USA, and an urban site over the city of Pavia, Italy | 98.00% (CA) | Hybrid of PCA, Autoencoder and logistic regression | Developed by the authors | 1% more precise than RBF-SVM | (Chen, Lin, Zhao, Wang, & Gu, 2014) |
| 5. | | Identify 21 land-use classes containing a variety of spatial patterns | UC Merced land-use data set | 93.48% (CA) | Author-defined | Theano | Unsupervised feature learning (UFL) (82-90%) and SIFT (85%) | (Luus, Salmon, van den Bergh, & Maharaj, 2015) |
| 6. | | Extract information about cultivated land | Images from UAV at the areas Pengzhou County and Guanghan County, Sichuan | 88-91% (CA) | Author-defined | N/A | N/A | (Lu, et al., 2017) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Province, China. | | | | | |
| 7. | Crop type classification | Classification of crops wheat, maize, soybeans sunflower and sugar beet | 19 multi-temporal scenes acquired by Landsat-8 and Sentinel-1A RS satellites from a test site in Ukraine | 94.60% (CA) | Author-defined | Developed by the authors | Multilayer perceptron (92.7%), Random Forests (88%) | (Kussul, Lavreniuk, Skakun, & Shelestov, 2017) |
| 8. | Plant Recognition | Recognize 7 views of different plants: entire plant, branch, flower, fruit, leaf, stem and scans | LifeCLEF 2015 plant dataset, which has 91,759 images distributed in 13,887 plant observations | 48.60% (LC) | AlexNet | Caffe | 20% worse than local descriptors to represent images and KNN, dense SIFT and a Gaussian Mixture Model | (Reyes, Caicedo, & Camargo, 2015) |
| 9. | | Recognize 44 different plant species | MalayaKew (MK) Leaf Dataset which consists of 44 classes, collected at the Royal Botanic Gardens, Kew, England | 99.60% (CA) | AlexNet | Caffe | SVM (95.1%), ANN (58%) | (Lee, Chan, Wilkin, & Remagnino, 2015) |
| 10. | | Identify plants from leaf vein patterns of white, soya and red beans | 866 leaf images provided by INTA Argentina. Dataset divided into three classes: 422 images correspond to soybean leaves, 272 to red bean leaves and 172 to white bean leaves | 96.90% (CA) | Author-defined | Pylearn 2 | Penalized Discriminant Analysis (PDA) (95.1%), SVM and RF slightly worse | (Grinblat, Uzal, Larese, & Granitto, 2016) |
| 11. | Segmentation of root and soil | Identify roots from soils | Soil images coming from X-ray tomography | Quality measure QM=0.23 (simulation) QM=0.57 (real roots) | Author-defined CNN with SVM for classification | MatConvNet | N/A | (Douarre, Schielein, Frindel, Gerth, & Rousseau, 2016) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 12. | Crop yield estimation | Estimate corn yield of county level in U.S. | Corn yields from 2001 to 2010 in Illinois U.S., downloaded from Climate Research Unit (CRU), plus MODIS Enhanced Vegetation Index | RMSE=6.298 | Author-defined | Caffe | Support Vector Regression (SVR) (RMSE=8.204) | (Kuwata & Shibasaki, 2015) |
| 13. | Fruit counting | Predict number of tomatoes in the images | 24,000 synthetic images produced by the authors | 91% (RFC) 1.16 (RMSE) on real images, 93% (RFC) 2.52 (RMSE) on synthetic images | Inception-ResNet | Tensor Flow | Area-based technique (ABT) (66.16%), (RMSE=13.56) | (Rahnemoonfar & Sheppard, 2017) |
| 14. | | Map from input images of apples and oranges to total fruit counts | 71 1280×960 orange images (day time) and 21 1920×1200 apple images (night time) | 0.968 (RFC), 13.8 (L2) for oranges 0.913 (RFC), 10.5 (L2) for apple | CNN (blob detection and counting) + linear regression | Caffe | Best texture-based regression model (ratio of 0.682) | (Chen, et al., 2017) |
| 15. | | Fruit detection in orchards, including mangoes, almonds and apples | Images of three fruit varieties: apples (726), almonds (385) and mangoes (1,154), captured at orchards in Victoria and Queensland, Australia. | F1-scores of 0.904 (apples) 0.908 (mango) 0.775 (almonds) | Faster Region-based CNN with VGG16 model | Caffe | ZF network (F1-scores of 0.892 0.876 and 0.726 for the apples, mangoes and almonds respectively) | (Bargoti & Underwood, 2016) |
| 16. | | Detection of sweet pepper and rock melon fruits | 122 images obtained from two modalities: color (RGB) and Near-Infrared (NIR) | 0.838 (F1) | Faster Region-based CNN with VGG16 model | Caffe | Conditional Random Field to model color and visual texture features (F1 =0.807) | (Sa, et al., 2016) |
| 17. | Obstacle detection | Identify ISO barrel-shaped obstacles in row crops and grass mowing | 437 images from authors' experiments and recordings | 99.9% in row crops and 90.8% in grass mowing (CA) | AlexNet | Caffe | N/A | (Steen, Christiansen, Karstoft, & Jørgensen, 2016) |
| 18. | | Detect obstacles that are distant, | Background data of 48 | 0.72 (F1) | AlexNet and VGG | Caffe | Local de-correlated | (Christiansen, |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | heavily occluded and unknown | images and test data of 48 images from annotations of humans, houses, barrels, wells and mannequins | | | | channel features (F1= 0.113) | Nielsen, Steen, Jørgensen, & Karstoft, 2016) |
| 19. | Identification of weeds | Classify 91 weed seed types | Dataset of 3,980 images containing 91 types of weed seeds | 90.96% (CA) | PCANet + LMC classifiers | Developed by the authors | Better results than feature extraction techniques (no details) | (Xinshao & Cheng, 2015) |
| 20. | | Classify weed from crop species based on 22 different species in total. | Dataset of 10,413 images, taken mainly from BBCH 12-16 containing 22 weed and crop species at early growth stages | 86.20% (CA) | Variation of VGG16 | Theano-based Lasagne library for Python | Local shape and color features (42.5% and 12.2% respectively) | (Dyrmann, Karstoft, & Midtiby, 2016 ) |
| 21. | | Identify thistle in winter wheat and spring barley images | 4,500 images from 10, 20, 30, and 50m of altitude captured by a Canon PowerShot G15 camera | 97.00% (CA) | DenseNet | Caffe | (Color feature-based) Thistle-Tool (95%) | (Sørensen, Rasmussen, Nielsen, & Jørgensen, 2017) |
| 22. | Prediction of soil moisture content | Predict the soil moisture content over an irrigated corn field | Soil data collected from an irrigated corn field (an area of 22 sq. km) in the Zhangye oasis, Northwest China. | RMSE=6.77 | Deep belief network-based macroscopic cellular automata (DBN-MCA) | Developed by the authors | Multi-layer perceptron MCA (MLP-MCA) (18% RMSE reduction) | (Song, et al., 2016) |
| 23. | Cattle race classification | Practical and accurate cattle identification from 5 different races | 1,300 images created by the authors | 93.76% (CA) | Gray Level Co-occurrence Matrix CNN (GLCM-CNN) | Deep Learning Matlab Toolbox | CNN without extra inputs (89.68%), Gaussian Mixture Model GMM (90%) | (Santoni, Sensuse, Arymurthy, & Fanany, 2015) |