Using machine learning to identify and diagnose crop disease

Megan Long, John Innes Centre, UK

1 Introduction

As the population of the world continues to grow, one of the biggest challenges facing food security is crop disease (Strange and Scott, 2005). Disease can have devastating effects on the yield of a crop, in some cases causing major losses where food quality is concerned. Not only is this a problem on a global scale, but it also has effects on a local scale for individual farmers. In poorer areas of the world, farming is the main or only source of income for many families

For any farmer, being able to detect and identify diseases in their plants is hugely important for the mitigation of potential losses. The problem with this, however, is that identifying crop diseases often requires specialist knowledge which is not always readily available to all farmers and can be expensive. Even with specialist knowledge, there are multiple factors that make it even more challenging. Many diseases appear with similar symptoms, meaning they are easily confused with one another. Furthermore, it is not uncommon for multiple diseases to be present at any one time, making the task of distinguishing them even more difficult.

http://dx.doi.org/10.19103/AS.2022.0107.22

Chapter taken from: Lobsey, C. and Biswas, A. (ed.), *Advances in sensor technology for sustainable crop production*, pp. 285–306, Burleigh Dodds Science Publishing, Cambridge, UK, 2023, (ISBN: 978 1 78676 977 0; www.bdspublishing.com)

Over recent years, machine learning techniques have been developed to assist with the identification and classification of diseases on a number of different crop types. Machine learning uses algorithms to learn a specific task from data without the solution or the process being explicitly defined by a human. It learns to recognise patterns in the data and make predictions based on this knowledge.

Multiple different methods have been used over the years, but in this chapter, we will focus on deep learning. This has become one of the most widely used machine learning methods over recent years, tackling problems in healthcare, self-driving cars and natural language processing as well as crop disease detection and a vast array of other use cases. In this chapter, we introduce deep learning for image analysis and discuss the key successes and pitfalls of using these methods for the identification of plant diseases.

2 A quick introduction to deep learning

A common machine learning technique is artificial neural networks (ANN), which are inspired by the biological neural network of the brain. They follow a logic structure that is meant to mimic the way the human brain thinks and draws conclusions. Over recent years, deep learning has taken centre stage in the machine learning world. Deep learning takes the ANNs further, making them deeper by adding layers. A typical machine learning ANN contains an input layer, an output layer and perhaps one hidden layer in between. In comparison, deep learning ANNs or deep learning networks contain many hidden layers.

The deep learning networks that are used for the identification of crop diseases are often a type of convolutional neural network (CNN) (LeCun et al., 1999) trained to perform their task by image analysis and classification. As before, each network has an input layer where the data (in this case images) is fed into the network and an output layer, which is where any predictions are given. Between these are a number of hidden layers which perform feature extraction. CNNs are used because of their strong ability to extract useful features from the images.

Feature extraction is where the network learns features about the images throughout the training process, allowing it to make predictions about the data. Earlier hidden layers learn low-level features, for example, lines and edges. As the network progresses through the hidden layers, the features extracted become more complex, which helps it to make a prediction about the image, for example, a classification of which disease is present in an input image. Figure 1 shows a simplified representation of the deep learning workflow for image analysis.

Training a network for image analysis requires a large dataset of images to work with, ideally tens of thousands of images depending on the problem. The image dataset is split into smaller datasets, usually train, validation and test sets; however, some studies use only a train set and test set. The split of the dataset between these sets varies over different works, but the bulk of the images is always contained within the train set, with a smaller amount in the validation and test sets.

When training a model, the train set of images is fed into the network in small batches (e.g. 64 images at a time). Once all the images have been through once, this is an epoch. The network will be trained for a certain number of epochs, as defined by the programmer. This number is often picked by taking an educated guess based on previous research and experiments in similar fields. Many studies will try multiple experiments with different numbers of epochs before finding the number that yields the best results for their data.

Between every epoch, the current network parameters are evaluated against the validation set to ensure that the training process is not overfitting to the data in the train set. If it were overfitting, then it would be learning features that are specific to the images in the train set and would not be able to generalise to new data of the same type, as in the test set. For example, if there were photos of a certain disease taken with light-coloured soil in the background, and these all ended up in the train set of images, the validation set with photos of that disease taken with dark soil in the background would highlight this. The network parameters would then readjust to ensure that it is not using that soil information from the train set for classifying that disease, but rather the disease information on the plant instead. Ideally, the train, validation and test tests would all contain both colours of soil.

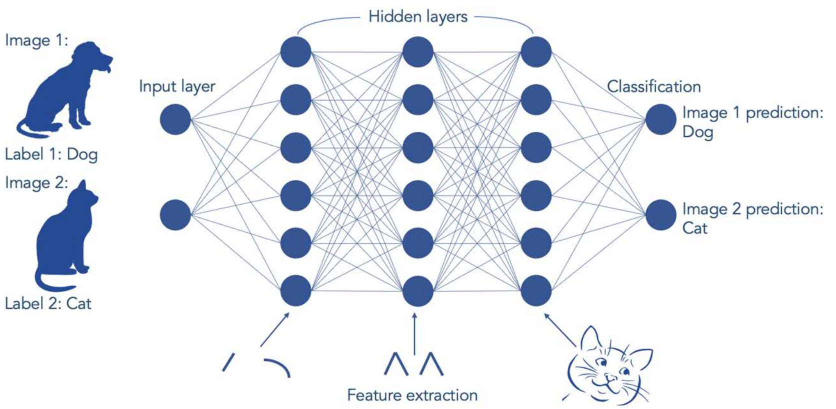

Figure 1 An artistic representation of a deep learning workflow for image classification. There is an input layer that takes in labelled images and an output layer that gives classification predictions. In between, there are multiple hidden layers that perform feature extraction. Earlier in the network basic features are learned, such as lines or edges, and further through the network they get more and more complicated until they learn features which allow the network to make its predictions.

Throughout training, the network is constantly adjusting its internal parameters after each batch and epoch to allow it to better make predictions about the images. As it learns, the accuracy of the predictions increases until it reaches a peak at the end of training. At this point, the network can be evaluated on the test set of images. This is a set of images of the same kind as contained in the train and validation sets, but that the network has never seen before. This shows how the trained network performs on brand new images to give a final accuracy rating. The whole process from start to finish can take a long time, from hours to days, to even months! This depends on multiple factors, such as the computing power available, the size of the network and the size of the dataset.

A lot of studies begin their experiments by using transfer learning with their datasets. This is a method that takes the knowledge learned by a previously trained network and applies it to the new problem. The main advantage of this is that it is relatively quick compared to training a deep learning network from scratch. Some examples of networks often used for transfer learning are AlexNet (Krizhevsky et al., 2012), GoogLeNet (Szegedy et al., 2014), VGGNet (Simonyan and Zisserman, 2015), ResNet (He et al., 2016), Inception V4 (Szegedy et al., 2016a) and MobileNet (Howard et al., 2017).

Transfer learning uses a network which has been fully trained on a large dataset (often the ImageNet dataset described in section 3). The pre-trained network is divided into two parts; the convolutional base, which is the part that performs feature extraction on the images, and the fully connected classifier, which forms predictions about the images. Depending on the method used, all or parts of the network are repurposed for the new dataset. In some cases, only the network structure is used and is retrained for the new problem without using the pre-trained knowledge.

3 Preparation of data for deep learning experiments

One of the biggest challenges for the successful application of machine learning techniques for the identification of plant diseases (or any image classification task for that matter) is the availability of data. The majority of these methods require large datasets of labelled or annotated images, which can be time-consuming to collect and process. For example, with plant-disease detection, it is necessary to have a large number of images for each disease for each plant species that is being modelled.

One of the most famous, and largest, datasets used for image analysis with deep learning is the ImageNet dataset (Deng et al., 2009). This dataset was created for use with object recognition software. The full dataset contains more than 14 million images with over 20 000 categories; however, a smaller subset of this has been used in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2015). This challenge ran annually between 2010 and 2017, encouraging participants to develop and improve computer vision techniques for image classification and object recognition. Multiple winning networks created for this competition over the years are now used as the starting point for hundreds of deep learning problems, including the problem of crop disease detection.

Collecting a dataset of images for use with any deep learning problem is not quite as easy as simply gathering as many images as possible by any means. It is important to ensure that the dataset contains appropriate information for the required use case. The rest of this section will discuss the factors to consider when preparing a dataset for plant-disease recognition and classification. These factors include a range of conditions, controlled versus uncontrolled capture conditions, image quality issues, the number of images required and labelling and annotation requirements.

The most widely known, and one of the only openly available datasets used for the recognition of plant diseases, is the Plant Village dataset (Hughes and Salathe, 2016). This is a collection of almost 88 000 images taken in controlled conditions with 38 categories, each corresponding to a plant-disease pair. Each image contains a single diseased or healthy leaf taken from the plant and placed on a neutral background and photographed under different lighting conditions. While this dataset was ground-breaking in the field of plant-disease detection when it was first created, the use of controlled conditions in the photos means that it is not comprehensive enough to be useful for an automated system in field conditions.

The Plant Village dataset was useful for demonstrating the potential of deep learning methods for the classification of plant diseases; however, in order to create a model that will be useful in realistic growth conditions, it is now important to collect datasets which accurately represent those conditions. PlantDoc (Singh et al., 2020) is a dataset created to cover many of the diseases present in the Plant Village dataset, but the images instead cover real field conditions. Here, images were downloaded from the internet and checked by members of the team before being added to the new dataset. This resulted in almost 3000 images, spanning 27 of the categories from Plant Village (any classes with fewer than 50 samples were removed for this dataset). This is a step in the right direction, but there is a distinct possibility of misclassified samples within the dataset due to them being taken from internet searches. Also, with it still being a relatively small dataset spanning a lot of classes, there is still a high chance that not all conditions are being covered.

For studies that are looking at building a model for a certain crop, it is unlikely that there are already datasets openly ready and available for use. This means that, for each case, there will be a large collection operation required prior to any numerical experiments. The result is usually a relatively small dataset with few categories (in some cases only two: diseased or healthy). While these can be useful for the problem at hand, there is still some way to go with generating a larger dataset to be used in a wider variety of cases.

The collection of a dataset which sufficiently covers each category can be a time-consuming task, often requiring the specialist knowledge of an expert pathologist and multiple volunteers to take the pictures. Furthermore, it is not simply a case of capturing a large number of images for each category, but also including a representative range of conditions. If the model is to be used for identifying diseases in the field, then the range of typical conditions that could be encountered in the field need to be represented. This includes:

- the variation in crop varieties/species – for example, different leaf colours or sizes;
- state of the crop – seedling, adult, flowering, mature with seed;
- stage and severity of the disease – early to late infection, mild to severe symptoms;
- weather and lighting conditions – full sun, sun and cloud, overcast, rain, etc.;
- background information – this needs to be consistent throughout the dataset. Having one class with different background information to the rest (e.g. glasshouse instead of field) will cause issues in training;
- image qualities – focus, depth of field, range of angles.

The main point to remember when creating a dataset for deep learning is that the conditions present need to be consistent between classes. Any class containing conditions which are not present in the others, for example, one class having sky in the background whereas no other does, will cause the network to learn the wrong information about that class and classify it based on the presence of sky, rather than the disease information.

Another factor to consider if working with real condition images is the diversity of background information, which might contribute negatively to the training process by distracting from the features that are of interest. If the images are collected in a field, for example, this may not be too much of an issue as the field conditions are likely to be relatively uniform. However, if the images are of a plant species which grows in various wild locations, then a vast array of background information can be expected. Where possible, the full range of diverse background conditions should be represented in images across all classes.

The number of images is also important. The number of images to aim for per category will depend on the complexity of the problem at hand. A simpler problem, for example, a binary classification problem of healthy or diseased, will require fewer images than a classification problem with multiple diseases with similar symptoms. However, the general rule of thumb with deep learning datasets is the more data, the better (ideally hundreds, if not thousands, of images per category in our opinion). The more images the network has to learn from, the better its performance is likely to be. It is also best if the data is relatively well balanced between each category, so the network does not learn a bias towards one class due to it having significantly more training samples than the others.

One technique to increase the number of images in a dataset where it is not possible to collect more is data augmentation. Augmenting the data involves performing multiple transformations on each image to add new samples to the dataset. For example, an image may be mirrored, flipped horizontally or vertically, rotated, or shifted to create tens of new images from a single sample. The main drawback of this is that there is no actual new data created, just variations of existing data. This means that the original dataset still needs to contain enough variation so that the network can learn enough to form predictions. There are other methods for working with smaller datasets; however, where possible it is always better to collect more data.

After collecting all available data, it will then need to be labelled and collated into a full dataset. For best results, a pathologist will need to label each image with the correct category, either as the images are taken, or by going through all data and assigning categories later. This of course can be incredibly time consuming and can result in misclassifications within the dataset if a pathologist is not available. In cases where different visualisation techniques are being used with the dataset (see Section 5), it may also be necessary to annotate the data with further information (e.g. a bounding box around a disease lesion). This often has to be done manually on each image and is a huge undertaking. Once all labelling and annotation is complete, the data can be sorted into the train, validation and test sets and start being used with a deep learning model.

4 Crop disease classification

A common use of deep learning methods for crop disease detection is to classify images of diseased plants into pre-defined classes. This can be a hugely difficult task for multiple reasons. For most, if not all, diseases there is a wide variation in visible symptoms throughout the life cycle of the disease, with some symptoms being more common than others. Furthermore, there can be a lot of similarities in visible symptoms of multiple diseases. Consequently, less common symptoms can easily be misclassified as another disease.

CNNs (LeCun et al., 1999) are a type of deep learning network which have become popular for image classification of plant diseases (Boulent et al., 2019). Many studies utilise pre-defined CNN structures for their work. A few examples that occur over and over again throughout the literature are AlexNet (Krizhevsky et al., 2012), GoogLeNet (Szegedy et al., 2014) and Inception (Szegedy et al., 2016b); however, there are plenty of others which provide a starting point for almost all of the studies we will discuss. These pre-defined networks are all CNNs with different numbers of layers and additional features to aid with feature extraction.

A good place to start here is with the Plant Village dataset. Multiple studies use the whole or part of this dataset with their work. Mohanty et al. (2016) aimed to show the viability of deep learning networks for the classification of a range of different diseases. They performed the first deep learning experiments using the Plant Village dataset with two different pre-trained networks: AlexNet and GoogLeNet. Through training these two networks both from scratch and using transfer learning, with a range of image processing techniques and train-test splits (the split of data between training and testing), they returned near-perfect accuracy with their best-performing method. Using a pre-trained GoogLeNet, full-colour images and an 80-20 train-test split the accuracy reached 99.34%.

Much like Mohanty et al., Brahimi et al. (2017) also used transfer learning with the two pre-trained networks AlexNet and GoogLeNet. In this study, however, rather than using the entire Plant Village dataset, a subset of images containing only diseased tomato leaves was used. Both studies utilised the networks by transfer learning and by training from scratch in an attempt to compare the results from both methods. In the same way, as in the work of Mohanty et al., the best results gained in Brahimi et al.'s work were of extremely high accuracy, reaching 99.18% accuracy in classifying tomato diseases. Again, this result came from the use of GoogLeNet with pre-training, although they do not specify the train-test split.

Another study that made use of a subset of the Plant Village dataset is that of Amara et al. (2017). They used only the banana leaf images in their work with the LeNet (Lecun et al., 1998) architecture. Although using a previously defined network, they did not use a pre-trained version, rather the architecture was trained from scratch with the banana leaf images. They used a range of train-test splits with both coloured and grayscale images. It was shown that the networks that used coloured images always outperform those without, thus showing the importance of colour information for the problem. Using a train-test split of 80-20, the network achieved an accuracy of 98.61%, another extremely promising result.

Too et al. (2019) took the whole of the Plant Village dataset and evaluated the performance of multiple pre-trained networks in classifying the diseases. They used transfer learning with some fine-tuning of VGG16, Inception V4, Resnet with 50, 101 and 152 layers and DenseNets (Huang et al., 2018). DenseNets was the best performer having gained an almost perfect accuracy of 99.75%.

This almost perfect accuracy is a common occurrence in studies which use only images from the Plant Village dataset. Although comprehensive in that it covers a wide range of diseases and plant species, the images within are not representative of those which would be found in real growth situations. They contain images of leaves taken from the plant and placed on a plain background, thus eliminating any background information, which obviously would not be the case in the field. The high accuracies gained in these studies are impressive; however, it is unknown how any of the models would perform when confronted with real field data.

Ferentinos (2018) demonstrated the issues of the Plant Village dataset for field use in their work. They made use of multiple pre-trained networks within his study; AlexNet, AlexNetOWTBn (Krizhevsky, 2014), GoogLeNet, Overfeat (Sermanet et al., 2013) and VGGNet. The dataset used contains images taken from Plant Village as 'lab condition' images, and was supplemented with more images taken in the field. This resulted in a dataset of 87 848 images sorted into 58 classes, some that contained just lab conditions, others that contained just field conditions and some with both. The most successful architecture in this study was the VGG network, which gained an accuracy of 99.53% on unseen images. Due to the presence of both lab condition and field condition images within the dataset used, Ferentinos (2018) experimented with training on laboratory condition images and testing on field condition images and vice versa. The accuracy of classification in these experiments was significantly lower than with the mixture of images for training. Training on field images and testing on laboratory images resulted in an accuracy of 65.69%, whereas the other way around resulted in an accuracy of only 33.27%. These figures emphasise the importance of including all relevant conditions within a training set for use in practice.

Although the Plant Village dataset is used regularly throughout the literature, there are plenty of studies which make use of data acquired elsewhere. Sladojevic et al. (2016) created a large dataset of images (over 30 000 in 15 classes) by taking pictures from internet searches. The dataset included a class for just healthy leaves and also a class with just background images. The reason for this was to train their network to differentiate leaves from their surroundings. The network used for this study was the pre-trained CaffeNet (Jia et al., 2014) model. Using this method, they gained a classification accuracy on their dataset of 96.3%. They concluded that the accuracy for individual categories was slightly lower in the classes which contained fewer images. Another thing to note about this study is how the images were collected. As they were taken straight from the internet, it is possible that some of the images have been wrongly classified which would have affected the accuracy of the network.

A study by Lu et al. (2017) used a relatively small dataset of rice disease images (500 images) to train CNNs inspired by LeNet and AlexNet architectures. Although they did not use the actual networks for either training from scratch or transfer learning, they did create a very similar network to those already defined. The accuracy gained for this network was 95%; while still a very encouraging result, this is slightly lower than many of the results discussed before. A reason for this could be to do with the size of the dataset used; with only 500 images spanning 10 categories, it could be hard for the network to learn all the characteristics present in each of the categories.

Alongside their own network modelled on a combination of AlexNet and GoogLeNet, Liu et al. (2018) utilised four pre-trained networks on their apple leaf disease identification problem; AlexNet, GoogLeNet, VGGNet and ResNet. They compared their network results to those obtained through transfer learning with the pre-trained networks and found that their model outperforms the known networks. The final accuracy recorded for their network was 97.62%, a percentage point higher than the next best performer VGGNet. Many studies make use of pre-defined networks; however, Liu et al. show that in some cases, defining a new network will gain a better performance. Often new networks will be inspired by one or several of the widely known networks (like in Lu et al., 2017), but this might be the best way to get all the best components for tackling the problem.

The train-test split is important for ensuring a network has enough data to learn from, while also having enough to for evaluating its performance. It is also important to include validation where possible. Often the validation is incorporated into the train part of the split when described in the literature.

Oppenheim et al. (2019) experimented with different train-test splits to find the best combination for their work detecting potato tuber disease. Their dataset contained 2465 images with 4 diseased and 1 uninfected category. They found that, unsurprisingly, more training data increased accuracy. The model that performed best on the test data used a 90-10 train-test split and gained an accuracy of 95.8%. Many studies elect to stick to an 80-20 split in the training and test data, in this case the higher amount of training images may improve training, but the lower amount of test images may not have contained enough images to fully show the performance of the network considering the size of the original dataset. A 90-10 split may be more suited to a larger dataset where the test set would contain more images.

The studies discussed in this section have shown the great potential for deep learning to be used for crop disease detection. The Plant Village dataset was a breakthrough in the field, which has seen multiple networks classify its images with incredibly high accuracy. Furthermore, other works have used more complex images while still gaining promising results. There is a lot of room for expanding these techniques for use with more diseases and more crop types.

5 Different visualisation techniques

Not all studies use only the image and the class label in their work. There are plenty of studies which use different visualisation techniques to try and improve the abilities of deep learning for plant-disease recognition tasks. As there are many different techniques being used throughout the field, we will discuss only a few in this section.

The first technique we will mention is the use of segmentation of images. This is where images are taken as usual and are cropped into smaller pieces. Ramcharan et al. (2017) collected a dataset of cassava leaf images with disease or pest damage in real field conditions. They ended up with two usable datasets: the first containing 2756 original cassava leaf images and a second with 15 000 segmented images containing smaller leaf sections (leaflets). Transfer learning with Inception V3 was used with three different classifiers on both datasets, with the best performer giving an accuracy of 93% on the leaflet dataset. It is not surprising that the larger dataset yielded better results as the networks were able to train on more data, plus the images were a lot less complex than those of the full cassava leaves.

In a slightly different approach, Ma et al. (2018) used image segmentation techniques to cut out the visible symptoms in images of cucumber diseases. These techniques were able to distinguish between lesion and background information and so crop away everything but the lesion. This resulted in a collection of images of only the visible symptoms placed on plain black backgrounds. They collected their dataset using images from the Plant Village dataset and from forestry images (https://www.forestryimages.org/) as well as supplementing them with their own images collected in real field conditions. After segmenting, they augmented the data to give 14 208 images for training a deep CNN. They gained an accuracy of 93.4% using their model, which outperformed AlexNet when they were compared. This is one of the lowest accuracy scores in all the studies using Plant Village images. It seems that the segmentation methods used here do not add anything to the process other than a higher preparation time for the dataset.

From the two approaches taken by Ramcharan et al. (2017) and Ma et al. (2018), it is clear that in some cases segmentation of images can be a useful tool. For complex images with lots of background information, segmenting the images to create smaller, less complicated samples will likely yield better classification results. However, for images taken in controlled conditions, segmenting may be adding an unnecessary, time-consuming step to the process. It is best to look at each problem on a case-by-case basis and decide based on the data whether the additional time injection for segmentation will be worthwhile.

Zhang et al. (2019a) also used a different method for classifying diseased plant images. They took the three channels of RBG images and fed them each into a different CNN for each channel. The outputs of the three CNNs were then sent into a fully connected classifier together to get a single classification. The idea was to utilise all the colour information in each image for better classification power. Each of the CNNs had the same architecture; however, different learned weights were obtained due to the different colour information in each of the channels; red, green and blue. Different CNN architectures (their own model, DNN, LeNet-5 (Lecun et al., 1998) and GoogLeNet) were tested on the Plant Village tomato disease images, while a cucumber disease image dataset containing 500 images were used to test different train-test splits. The best results gained in this study used a 70-30 train-test split with their own network architecture and gave a classification accuracy of 94.27%. Although this is still a high disease recognition rate, it is not as high as some of the other studies show suggesting that better results can be gained when all three channels are fed into the same network.

Some studies like to include lesion location information in their work. For example, DeChant et al. (2017) created a dataset containing 1834 images of healthy and northern leaf blight-infected leaves of maize plants. In the infected images, each lesion was annotated with a line which was used as extra information for training the networks. They used a three-step method for classifying their images, first taking small portions of the images and training to detect the presence of lesions in these images. The second step used the networks trained in step one to create heat maps of the probability of parts of the image containing a lesion. The final step used these heatmaps to train a network to classify an image into either containing lesions (infected) or not (healthy). This method managed to get an accuracy in the full image classification of 96.7%. This result is extremely encouraging considering the use of images containing much background information. It does however only evaluate the method on two classes, healthy or diseased. It would be interesting to see how the accuracy would be affected by using this technique with more classes.

The work by DeChant et al. (2017) also made use of heatmaps. These can be an especially useful tool for ensuring that a network is functioning correctly. A heatmap shows the parts of the image that the network predicts to be most likely to contain disease information. These parts appear ‘hotter’, so on a blue–red scale, red sections are more likely to contain a lesion, whereas blue is the background material. From these heatmaps it is easy to pick out issues in the data; for example, if there is a red patch on a piece of background material in an image, it indicates that the network is probably not using the correct information to drive its classifications but is using background information instead.

Published by Burleigh Dodds Science Publishing Limited, 2023.

There are various other visualisation techniques used for classifying crop diseases, each with their own merits and pitfalls. It would take a long time to go over them all! We have added a paper to the 'Where to look for more information' section which contains plenty of references to studies using different visualisation methods in their work for anyone who wishes to find out more.

6 Hyperspectral imaging for early disease detection

A relatively new addition to the field of deep learning is the use of hyperspectral imagery. Hyperspectral images capture information across the electromagnetic spectrum, not just the visual spectrum. In many cases there can be 'fingerprints' left by certain objects or substances (such as a disease) in different spectrums, which are not visible in regular conditions. It is for this reason that hyperspectral imagery has started to be deployed for the early detection of crop diseases. The idea is that, before there are visible symptoms on the plant, there may be unseen effects across other spectrums that would allow for early diagnosis and treatment before the disease got more severe.

Studies by Wang et al. (2019) and Jin et al. (2018) took hyperspectral images of sweet peppers with tomato spotted wilt virus and wheat with fusarium head blight, respectively. They both performed their experiments by taking the pixels of the images to train their networks. The pixels were labelled with background, diseased or healthy. This method required a small number of full hyperspectral images due to the number of pixels contained within each image. Both studies gain promising results with their small datasets, with Wang et al. gaining an accuracy of 96.25% on images of plants taken prior to disease symptoms being visible.

Nagasubramanian et al. (2018) worked with hyperspectral images of soybean crops with and without charcoal rot. They collected a dataset of 111 images, which were split into smaller data patches to create a larger dataset of 1090 training images and 539 test images. There appears to be a large bias towards the healthy class in their dataset; however, this did not appear to cause any problems with training. They used a CNN as their model, which gained an accuracy of 95.73% when presented with the test images.

In a similar experiment to Nagasubramanian, but with *Aphis gossypii* Glover infection of cotton leaves, Yan et al. (2021) used CNNs with hyperspectral images. They performed multiple experiments with RBG images and full hyperspectral images, comparing the performance of the CNNs against other machine learning methods. In all cases, they found that the CNNs gained the highest accuracies. They also found that the hyperspectral images gave better results than just RGB images.

An interesting study by Zhang et al. (2019b) utilised an unmanned aerial vehicle for collecting their hyperspectral images. Five images were taken of full wheat plots, with and without yellow rust, from a height of 30 m. These images were segmented into 15 000 smaller hyperspectral image blocks which were used for training and evaluating their network, which combined features of Inception and ResNet. Each block was labelled either as containing rust, healthy plants or other (e.g. soil or road). After the model had been trained on 10 000 of the data blocks, it achieved a classification accuracy of 85% on the remaining 5000 test data blocks. The next step in their process involved mapping the data blocks back onto the original images. Here the sections that were predicted as rust areas were highlighted in red to show an infection map over the entire plot. This is extremely useful for showing the infection levels in a full plot and methods like this could possibly be utilised for scoring the amount of disease present as well as classification.

Although there is not a great number of studies already utilising this technology for plant-disease detection, it shows great promise and there is plenty of room for growth in the future. As always, the availability of data is a bottleneck for advancement. Hyperspectral imagining requires a lot more preparation than simply collecting photos with a regular camera. For example, environmental factors such as light and heat levels can have an effect on the different wavelengths, thus affecting the collected images. Furthermore, hyperspectral imaging cameras can be a costly investment, meaning they are not readily available for everyone to be able use in their work. A positive, however, is that for some methods, these studies tend not to require as large a dataset as those using regular visible spectrum images. A smaller number of images can be segmented or taken at pixel level to give a large dataset for training and evaluation.

7 Case study: identification and classification of diseases on wheat

In this section we will discuss some recent advances. In this work we used deep learning methods for the identification and classification of wheat diseases in field conditions. Wheat is a staple crop, vital for feeding many people across the world. Therefore, it is important to be able to control any diseases and mitigate any yield losses.

Alongside a healthy category, we included four of the most commercially important wheat diseases for the UK and other countries in our dataset: Yellow rust, otherwise known as stripe rust, caused by the basidiomycete fungus *Piccunia striiformis* f.sp. *tritici (Pgt)* (Liu and Hambleton, 2010). Septoria, otherwise known as Septoria leaf blotch or just Septoria, caused by the ascomycete fungus *Zymoseptoria tritici* (formerly *Mycosphaerella graminicola)*

(Hardwick et al., 2001). Brown rust, caused by *Puccinia triticina* (Goyeau et al., 2006; Bolton et al., 2008) and powdery mildew, caused by *Blumeria graminis* (Dubin and Duveiller, 2011).

The most familiar symptoms of yellow rust are the easily recognisable yellow/orange pustules which form in stripe patterns on the leaves of wheat. Later in its life cycle, however, when the yellow/orange pustules fall off, necrotic lesions with black telia remain on the leaf. It is at this stage that yellow rust can be easily mistaken for mature Septoria which appears as necrotic lesions which follow the veins of the leaf containing many small, black pycnidia. To make matters even more complicated, brown rust appears as orange/brown pustules on wheat leaves, which can be similar in appearance to the early-stage yellow rust pustules. Figure 2 shows examples of these diseases on wheat leaves. Although visibly different from the other three diseases, the white powdery pustules of powdery mildew still cause problems in many parts of the world, making it an important foliar wheat disease and worthy of inclusion in a wheat disease detection model.

Our first important step to creating our model was the collection of a viable dataset of images to use for training. For our model to be useful in the field, it needed to be trained using images which cover the range of conditions which would be encountered in the field, for example, different weather and light conditions, different varieties and colours of wheat, growth stages of the plants and life cycle stage of the diseases. In this experiment, we were looking at only a

Figure 2 Examples of wheat leaf diseases. (a) Brown rust appears with orange/brown pustules along the leaf, which can be confused with the yellow/orange pustules of (b) yellow rust. (c) Yellow rust where the orange pustules have fallen off leaves necrotic lesions that follow the veins of the leaf. This is easily mistaken with (d) mature Septoria, which also appears as necrotic lesions that follow the veins of the leaf.

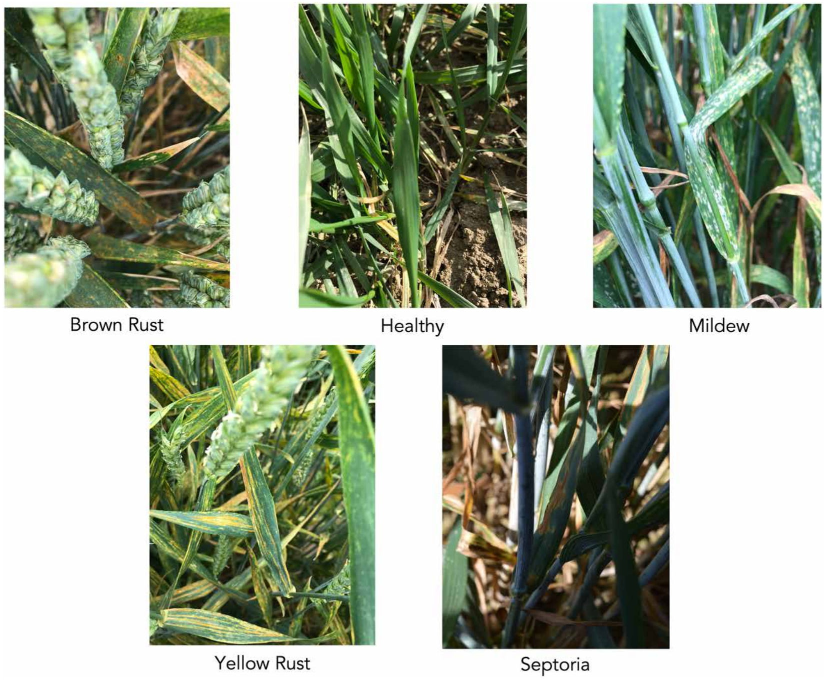

Figure 3 Example wheat disease images collected for the dataset. The images contain complex background information including soil, other plants and shadow. From left to right: top – brown rust, healthy, mildew; bottom – yellow rust, Septoria.

single disease at a time, therefore the sites for photographing diseases needed to be carefully selected by a pathologist to contain only one disease at a time.

Having collected a significant number of images for each category, the photos taken went through manual quality control. Any images which were blurry, contained no important information or where the important information was obstructed were removed. Images which were thought to contain multiple diseases or where it was not clear which disease was present were also removed. The resulting dataset contained between 2000 and 5000 images per category, providing a vast array of conditions and complex background information, as would be expected in the field. Figure 3 shows example images from the dataset.

We first used transfer learning with four pre-trained networks: VGG16 (Simonyan and Zisserman, 2015), Inception V3 (Szegedy et al., 2016b), Mobilenet (Howard et al., 2017), Xception (Chollet, 2017). Each had been pre-trained using the ImageNet dataset. These experiments allowed us to determine whether a deep learning model would be able to learn to classify these diseases with such complex input data. The results for each pre-trained model were between 85% and 92% classification accuracy.

These results gave us the confidence that deep learning methods would work for this problem. We then developed a bespoke model for this problem using a CNN architecture. The model took several days to train on all of the available training data, before being evaluated on the test dataset. When challenged with the new images from the test dataset, our trained model performed with a classification accuracy of over 97%.

We decided to compare the performance of our model against that of a selection of trained pathologists. Five participants, with differing backgrounds and specialisations, were included in the experiment. A smaller subset of the dataset was used taken from the test set, including a number of images that were incorrectly classified by the network. Each participant was shown the images one by one on a computer screen, they were asked to assign a tag corresponding to their classification of the image. Their classifications were collected and compared with those from the network and the results showed that the network outperformed each member of the group, gaining the highest classification accuracy.

Deep learning networks with the appropriate architecture have the power to deal with real field images containing complex background information. Our resultant wheat disease classification network will be useful for identifying a disease in the field, and so making it easier to take appropriate action. For the purpose of breeding for disease resistance, it would be beneficial for a model to be able to quantify the amount of disease as well as identify and classify different symptoms.

8 Conclusion and future trends

The control of crop diseases is becoming more important than ever as the population of the world continues to increase. For farmers and agronomists, the first step to controlling a disease is identification, which, without access to a trained pathologist, can be difficult in itself. Deep learning methods open the door to automated crop disease detection, which will be game-changing for the treatment and prevention of diseases across the globe.

Recent research had made great strides in classifying images of diseased plants taken in controlled conditions, gaining some extremely high accuracy results. There has also been plenty of progress in using real field images to train networks. It is clear that deep learning techniques are more than capable of handling the task of disease detection and classification. The main obstacle holding back this area of research is the availability of data. Collecting a useful dataset with sufficient samples, covering enough conditions can be a challenging and time-consuming task.

There are a few directions that this research could take in the future. The first is looking at crops which are infected with multiple diseases. At present, very few, if any, studies use images that contain multiple diseases, and it is easy to see why as the presence of multiple diseases complicates the problem substantially. However, it is a very common occurrence in the field to have more than one disease present. Therefore, it is important that deep learning networks are able to handle this if they are to be deployed in the field to aid farmers and agronomists.

Another direction that could be taken in the future is the use of deep learning for quantifying the amount of disease present as well as classifying it. This will be an extremely valuable tool for breeders looking to breed varieties of staple crops with resistance to important diseases. Currently, pathologists are required to spend a lot of time scoring the diseases manually on thousands of plots, so automating this process would be beneficial for them in freeing up time for other important tasks. There are currently a few studies that have started working on this problem, but there is plenty of room for growth and improvement.

This work was supported by the BBSRC Norwich Research Park Biosciences Doctoral Training Partnership as a CASE Award, [grant number BB/S507428/1], in collaboration with Limagrain UK, KWS UK Ltd and RAGT Seeds Ltd.

9 Where to look for more information

- A clear introduction to deep learning and how to create your own deep learning networks – 'Deep Learning with Python' by Francois Chollet.
- About dataset size and variation – Barbedo (2018) 'Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification' in *Computers and Electronics in Agriculture*.
- More references to different visualisation techniques – Saleem et al. (2019) 'Plant Disease Detection and Classification by Deep Learning' in *Plants*.

10 References

Amara, J., Bouaziz, B. and Algergawy, A. (2017). A deep learning-based approach for banana leaf diseases classification. Datenbanksysteme für Business, Technologie und Web – Workshopband, pp. 79–88.

Bolton, M. D., Kolmer, J. A. and Garvin, D. F. (2008). Wheat leaf rust caused by Puccinia triticina, *Molecular Plant Pathology* 9(5), 563–575. doi:10.1111/j.1364-3703.2008.00487.x.

Boulent, J., Foucher, S., Théau, J. and St-Charles, P. L. (2019). Convolutional neural networks for the automatic identification of plant diseases, *Frontiers in Plant Science* 10, 941. doi:10.3389/fpls.2019.00941.

Brahimi, M., Boukhalfa, K. and Moussaoui, A. (2017). Deep learning for tomato diseases: classification and symptoms visualization, *Applied Artificial Intelligence* 31(4), 299–315. doi:10.1080/08839514.2017.1315516.

Chollet, F. (2017). Xception: deep learning with depthwise separable convolutions, *arXiv:1610.02357 [cs]* [Preprint]. Available at: http://arxiv.org/abs/1610.02357 (Accessed: 3 March 2021).

DeChant, C., Wiesner-Hanks, T., Chen, S., Stewart, E. L., Yosinski, J., Gore, M. A., Nelson, R. J. and Lipson, H. (2017). Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning, *Phytopathology* 107(11), 1426–1432. doi:10.1094/PHYTO-11-16-0417-R.

Deng, J., Dong, W., Socher, R., Li, L., Li, K. and Fei-Fei, L. (2009). ImageNet: a large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255. doi:10.1109/CVPR.2009.5206848.

Dubin, H. J. and Duveiller, E. (2011). Fungal, bacterial and nematode diseases of wheat: breeding for resistance and other control measures. In: *The World Wheat Book, A History of Wheat Breeding* Angus, W., Bonjean, A. and van Ginkel, M. (Eds), pp. 1131–1191.

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis, *Computers and Electronics in Agriculture* 145, 311–318. doi:10.1016/j.compag.2018.01.009.

Goyeau, H., Park, R., Schaeffer, B. and Lannou, C. (2006). Distribution of pathotypes with regard to host cultivars in French wheat leaf rust populations, *Phytopathology* 96(3), 264–273. doi:10.1094/PHYTO-96-0264.

Hardwick, N. V., Jones, D. R. and Slough, J. E. (2001). Factors affecting diseases of winter wheat in England and Wales, 1989–98, *Plant Pathology* 50(4), 453–462. doi:10.1046/j.1365-3059.2001.00596.x.

He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778. doi:10.1109/CVPR.2016.90.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H. (2017). MobileNets: efficient convolutional neural networks for mobile vision applications, *arXiv:1704.04861 [cs]* [Preprint]. Available at: http://arxiv.org/abs/1704.04861 (Accessed: 3 March 2021).

Huang, G., Liu, Z., van der Maaten, L. and Weinnberger, K. Q. (2018). Densely connected convolutional networks, *arXiv:1608.06993 [cs]* [Preprint]. Available at: http://arxiv.org/abs/1608.06993 (Accessed: 28 September 2021).

Hughes, D. P. and Salathe, M. (2016). An open access repository of images on plant health to enable the development of mobile disease diagnostics, *arXiv:1511.08060 [cs]* [Preprint]. Available at: http://arxiv.org/abs/1511.08060 (Accessed: 5 March 2021).

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S. and Darrell, T. (2014). Caffe: convolutional architecture for fast feature embedding, *arXiv:1408.5093 [cs]* [Preprint]. Available at: http://arxiv.org/abs/1408.5093 (Accessed: 27 September 2021).

Jin, X., Jie, L., Wang, S., Qi, H. and Li, S. (2018). Classifying wheat hyperspectral pixels of healthy heads and fusarium head blight disease using a deep neural network in the wild field, *Remote Sensing* 10(3), 395. doi:10.3390/rs10030395.

Krizhevsky, A. (2014). One weird trick for parallelizing convolutional neural networks, *arXiv:1404.5997 [cs]* [Preprint]. Available at: http://arxiv.org/abs/1404.5997 (Accessed: 27 September 2021).

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks, *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. Red Hook, NY, USA: Curran Associates Inc. (NIPS'12), pp. 1097–1105.

Lecun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998). Gradient-based learning applied to document recognition, *Proceedings of the IEEE* 86(11), 2278–2324. doi:10.1109/5.726791.

LeCun, Y., Haffner, P., Bottou, L. and Bengio, Y. (1999). *Shape, Contour and Grouping in Computer Vision. International Workshop on Shape, Contour and Grouping in Computer Vision*, Springer Verlag, pp. 319–345. doi:10.1007/3-540-46805-6_19.

Liu, B., Zhang, Y., He, D. and Li, Y. (2018). Identification of apple leaf diseases based on deep convolutional neural networks, *Symmetry* 10(1), 11. doi:10.3390/sym10010011.

Liu, M. and Hambleton, S. (2010). Taxonomic study of stripe rust, Puccinia striiformis sensu lato, based on molecular and morphological evidence, *Fungal Biology* 114(10), 881–899. doi:10.1016/j.funbio.2010.08.005.

Lu, Y., Yi, S., Zeng, N., Liu, Y. and Zhang, Y. (2017). Identification of rice diseases using deep convolutional neural networks, *Neurocomputing* 267, 378–384. doi:10.1016/j.neucom.2017.06.023.

Ma, J., Du, K., Zheng, F., Zhang, L., Gong, Z. and Sun, Z. (2018). A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network, *Computers and Electronics in Agriculture* 154, 18–24. doi:10.1016/j.compag.2018.08.048.

Mohanty, S. P., Hughes, D. P. and Salathé, M. (2016). Using deep learning for image-based plant disease detection, *Frontiers in Plant Science* 7, 1419. doi:10.3389/fpls.2016.01419.

Nagasubramanian, K., Jones, S., Singh, A. K., Singh, A., Ganapathysubramanian, B. and Sarkar, S. (2018). *Explaining Hyperspectral Imaging Based Plant Disease Identification: 3D CNN and Saliency Maps*, arXiv:1804.08831 [cs.CV]. Available at: https://arxiv.org/abs/1804.08831 (Accessed: 7 September 2022).

Oppenheim, D., Shani, G., Erlich, O. and Tsror, L. (2019). Using deep learning for image-based potato tuber disease detection, *Phytopathology* 109(6), 1083–1087. doi:10.1094/PHYTO-08-18-0288-R.

Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J. and Hughes, D. P. (2017). Deep learning for image-based cassava disease detection, *Frontiers in Plant Science* 8, 1852. doi:10.3389/fpls.2017.01852.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C. and Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge, arXiv:1409.0575 [cs] [Preprint]. Available at: http://arxiv.org/abs/1409.0575 (Accessed: 21 September 2021).

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R. and LeCun, Y. (2013). OverFeat: integrated recognition, localization and detection using convolutional networks, *International Conference on Learning Representations (ICLR) (Banff)* [Preprint].

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition, *arXiv:1409.1556 [cs]* [Preprint]. Available at: http://arxiv.org/abs/1409.1556 (Accessed: 3 March 2021).

Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S. and Batra, N. (2020). PlantDoc: a dataset for visual plant disease detection. 249–253. doi:10.1145/3371158.3371196.

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D. and Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification, *Computational Intelligence and Neuroscience* 2016, 3289801. doi:10.1155/2016/3289801.

Strange, R. N. and Scott, P. R. (2005). Plant disease: a threat to global food security, *Annual Review of Phytopathology* 43(1), 83–116. doi:10.1146/annurev.phyto.43.113004.133839.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. (2014). Going deeper with convolutions, *arXiv:1409.4842 [cs]* [Preprint]. Available at: http://arxiv.org/abs/1409.4842 (Accessed: 28 May 2021).

Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A. (2016a). Inception-v4, inception-resnet and the impact of residual connections on learning, *arXiv:1602.07261 [cs]* [Preprint]. Available at: http://arxiv.org/abs/1602.07261 (Accessed: 27 September 2021).

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2016b). Rethinking the inception architecture for computer vision, doi:10.1109/CVPR.2016.308.

Too, E. C., Yujian, L., Njuki, S. and Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification, *Computers and Electronics in Agriculture*, 161, pp. 272–279. doi:10.1016/j.compag.2018.03.032.

Wang, D., Vinson, R., Holmes, M., Seibel, G., Bechar, A., Nof, S. and Tao, Y. (2019). Early detection of tomato spotted wilt virus by hyperspectral imaging and outlier removal auxiliary classifier generative adversarial nets (OR-AC-GAN), *Scientific Reports* 9(1), 4377. doi:10.1038/s41598-019-40066-y.

Yan, T., Xu, W., Lin, J., Duan, L., Gao, P., Zhang, C. and Lv, X. (2021). Combining multi-dimensional convolutional neural network (CNN) With visualization method for detection of aphis gossypii Glover infection in cotton leaves using hyperspectral imaging, *Frontiers in Plant Science* 12, 604510. doi:10.3389/fpls.2021.604510.

Zhang, S., Huang, W. and Zhang, C. (2019a). Three-channel convolutional neural networks for vegetable leaf disease recognition, *Cognitive Systems Research* 53, 31–41. doi:10.1016/j.cogsys.2018.04.006.

Zhang, X., Han, L., Dong, Y., Shi, Y., Huang, W., Han, L., González-Moreno, P., Ma, H., Ye, H. and Sobeih, T. (2019b). A deep learning-based approach for automated yellow rust disease detection from high-resolution hyperspectral UAV images, *Remote Sensing* 11(13), 1554. doi:10.3390/rs11131554.