# Introduction

Crops provide us with food that is used as a staple food all over the world. However, managing the plant diseases can be a very hard topic especially if the farmer Is new to whole industry and do not know what must be done to prevent and control the diseases. Due to this problem with the proper disease identification a solution can be the use of the deep learning where a model is trained to find out the diseases to prevent the diseases. This report presents an analysis of the plant disease detection methods using deep learning that can be used and evaluates their effectiveness solving the problems faced by farmers.

## Problem description, context and motivation

Plant diseases can cause damages which can be irreversible if they are not identified at the right moment. Traditional methods such as relying on the human vision could be Too late for working out a solution as there are time consuming and the human eyes may not notice the subtle notes of the early stages of the diseases to prevent these.

If the diseases are not identified and treated there can be many stakeholders that can get effected. The main group includes the farmers as they will lose what they have invested to the crop because of this loss consumers will have to pay higher prices which can put strain on the governments as they will have to arrange solutions for these problems. Moreover, other parties such as food processors may also have to face the ripple effects of these problems.

The problem of plant diseases is common happening when it comes to crops and can happen anywhere in the world and can depend on the weather, plant type, and how much the area is affected by plant diseases.

Solving this problem can ensure that the plants get the necessary treatments at the necessary times to ensure that the farmers get the full yield of their crops and to ensure the food security. Moreover, building a system that can identify the diseases can help new farmers who are entering the industry to correctly identify diseases and use the appropriate amounts of pesticides which is good for both the consumers and the farmers too.

## Aims

The primary aim of this project is to develop an accurate and efficient system for classifying the plant diseases especially in potato and wheat plants as these are the most common types of crops that are in the United Kingdom. Furthermore, this system also provides the recommendations for the users on what must be done to overcome the situation. Moreover, the project aims to make a web application for the users to upload the photos and get an classification of the diseases and get an recommendation for the diseases that is model is trained on.

## Objectives

The main objective of this project is to make a web application for identifying plant diseases in wheat and potato and provide recommendations to the users. To achieve the main objective the project needs to collect the necessary datasets that is needed to train the models which is intended to be done through the free to access websites such as Kaggle and free to use to use datasets. Furthermore, after collecting the datasets the training of Convolutional Neural Network (CNN) and testing the models for their accuracy. Additionally, after all the objectives above has been achieved the recommendation system will be built to provide the necessary recommendations for the user.

## Legal

From the legal standpoint data protection laws must be taken into consideration such as collecting the data from the users and other free to use websites. Intellectual property rights must also be taken into consideration such as when using open-source tools and algorithms.

## Social

This project aims to close the gap between farmers and the latest technology that is available by making detection systems more efficient and affordable in the ways of correctly identifying the diseases and providing the necessary solutions for the farmers before it is too late for something can be done or vast amounts of pesticides have to be used in order to eradicate the diseases which can be harmful for everyone that is involved. This system also helps new farmers as they may lack the experience that is needed for the detection of the diseases.

## Ethical

When dealing with the ethical concerns of this project the datasets the researcher collect must ensure that it contains equal amount of information in all the classes that they want to find out otherwise, the system may have bias introduced into the dataset which can lead to improper classification of the diseases in the trained models. Additionally, this project uses images of plants which must be collected from a human controlled environment which means that we must take the legal consent form the farmers before moving forward.

## Professional

From the professional perspective, this project aims to keep record of what has been done in the forms of flowcharts, diagrams and other suitable formats to achieve the same results if this was to replicate by another party. This project also aims to connect with the other professionals in the field to get the necessary recommendations for the purpose of providing a proper outcome for the users.

## Background

This problem of plant diseases is a challenge that has been there since the start of agriculture which even have led countries to shortages of food supplies in the past. In fact, [1, pp. 837–842] shows that 40% of the damage is caused by the plant diseases which can cost up to US$220 billion worldwide. However, with the evolving technology concepts such as using deep learning is getting widely popular which can help to classify the diseases and provide the necessary steps for solving the problem where it is impossible for humans to detect the diseases until the diseases are spread to a wider area of the crop.

## Project overview

This report is organized as follows, the introduction will give the basic outline of the problem, the aims it tries to achieve and objectives of this project. The document also highlights the legal social, ethical and professional aspects of the project and finally the background. The literature review of this project will highlight the key research that has been done in the area and the existing methods for plant disease detection. The technology review of this project will also highlight what technologies are being used to achieve this project. The methodology section will help the readers understand how the technology that is talked about in the technology review is applied.

# Literature review

Plant diseases pose a critical risk for the plant health which can reduce or destroy the crop output which may also have effects on other sectors which depend on the agriculture. Early detection of the plant diseases is the key for reducing the impact of plant diseases this will ensure that the farmers and other stakeholders can necessarily actions to mitigate the problems that are faced by these.

Before the introduction of machine learning and deep learning farmers and other stakeholders used to visual inspection, which is simple, yet time consuming process given the size of the area they may decide to investigate before giving a final wording of the disease/s. Furthermore, the decisions that is taken by stakeholders can be different from one to another which can lead to multiple types of unwanted treatments being applied on the crop which can increase the cost for farmers and unwanted use of pesticides can be harmful.[2, pp. 569–596] shows that use of recommended amount pesticides can alter the natural soil properties and harm good soil microorganisms which can lead to reduced soil fertility and productivity. Furthermore, [3] shows that the use of pesticides has linked to causing chronic kidney diseases in humans.

Introduction of machine learning, deep learning and computer vision the detection of plant diseases completely changed. Before the introduction of deep learning, which is a subset of machine learning was the one that was used to do disease identification, but deep learning dramatically increased the accuracy and the effectiveness of plant disease detection [4, p. 468].

When it comes to data collection which is considered as a main part of the system. Some researchers tend to collect the data from the real-world scenarios while some researchers tend to grow the plants by themselves and inoculate them with the virus [5, pp. 56683–56698]. However, the images that has been collected from the real world may have noise in them. According to [6, p. 1177225] noise can decrease the model accuracy by 5% to 40% in tasks such as classification. Additionally, the model even achieves a significant accuracy just by looking at the background of the image which can indicate the reliance on the background pixels rather than the actual disease features[7]. However, to overcome the problem of fitting to the noise different pre-processing techniques can be considered such as image clipping where the cropping to get the interested image region after that image smoothing can be done to smooth out the picture and image enhancement which is done by increasing the contrast [8, pp. 768–771].

The accuracy of the plant disease classification mainly depends on how much of data that can be captured from the plant leaves. [9, pp. 229–242] researchers have used

The other problem that is highlighted by the other researchers is that the unavailability of enough datasets to train the models. Due to that the researchers have used data augmentation, where it increases the training data which can help to reduce the problem of overfitting and improving generalization to unseen data[10, pp. 913–920]. For the plant disease detection [5, pp. 56683–56698] these researchers’ purpose two methods. First method is the use of traditional augmentation where they purpose methods such as the use of tensile rotation, adjustment resolution image translation distribution balance the repercussion of using this method includes poor quality, inadequate diversity and unevenness. The second method they purpose Is the use of Generate Adversarial Networks (GANS), a method that has been introduced by [11, pp. 139–144] it is an artificial intelligence algorithm that solve the generative modelling problem by studying the training examples. The main reason this method is used is because to generate synthetic samples with the same characteristics as the given training distribution. This method Is being used extensively to generate samples and many other variations based on the original GANS has been published such as DCGAN (Deep Convolutional Generative Adversarial Network), CGAN (Conditional Generative Adversarial Network).