Deep Learning for Plant Disease Identification

Padala Akhil Reddy

Company: Feynn Labs

Batch-8-23-S-B

Task 0

27-02-2023

Abstract:

Plant diseases are a severe threat to crop production, and consequently can compromise global food security. Therefore, it is necessary for farmers to monitor their crops, identify crop diseases at their early stages and effectively deal with them to minimise the crop damage. However, depending on the agricultural plot size, when there are a large number of plants and many types of diseases which can affect plants, identifying which disease is affecting which plant can be a laborious task. Therefore, it is proposed to use deep learning models to aid the timely identification of plant disease.

1. Problem Statement:

Crop loss worldwide is estimated to be $220 billion USD or 14.1% of crop production due to plant diseases[1]. Plant diseases can be caused by fungi, bacteria, viruses, nematodes and also abiotic factors like the environment. While climate change and population growth are creating more food demand, plant diseases are affecting both the quantity and quality of the crop yield, which is also economically undesirable to the farmers.

Rapid detection and diagnostic procedures at the early stages of the diseases can effectively prevent the further spread of plant diseases. So here, a reliable technique should be employed to instantly detect a diseased plant before the disease advances and spreads to other crops.

2. Customer Needs Assessment:

To understand the needs of the customer, first we need to understand the objectives of plant disease management[2]:

- First-level objective: Reduce the crop losses caused by spread of the disease.
- Second-level objective: Reduce the level of infection caused by the pathogen or by abiotic factors.
- Third-level objectives: a. Reduce the level of initial exposure to infection.
  b. Reduce the rate of disease development.
  c. Reduce the duration of the epidemic.

- Fourth-level objectives are basically the measures directly taken by the farmer to control the epidemic (like using required chemicals to control the infestation).

Farmers can take two types of measures to control the epidemic:

- Preventive measures: For example, using pathogen-free seeds, treating seeds with fungicides, pesticides and hot water.
- Curative measures: For example, removal of parasitic plants or weeds, isolation of affected plants, using fungicides, bactericides, etc.

The farmer will take his own preventive measures based on his knowledge and recommendations from various sources. The main issue that the farmer has to deal with, is on how to quickly identify the affected crops and take timely curative measures.

3. Target Specifications and Needs Statement:

The following are the design criteria which address customer needs that should be kept in mind while designing the product:

- Should report as few false negatives as possible (Ideally, the recall of the classification model should be very high).
- Able to detect the disease even in its very early stages.
- Should be able to detect a large variety of diseases.
- Should be usable in a smartphone.
- Ideally it should take only a single photo of the diseased plant to show that the plant is affected.
- If a certain disease is detected, then the application should give some information about the disease as well as recommend the correct fertiliser/pesticide/fungicide, along with the correct amount to be used.

Also, it should be kept in mind that the majority of the target customers will be small farmers who are not very tech-savvy, so the UI of the application should be as simple to use as possible. It should not have too many features which can end up confusing the customer.

4. External Search:

The following sources were used for gaining a better understanding of the problem statement:

- Plant Disease - Crop Loss.
- Crop Diseases: Types, Control and Prevention.
- Plant Disease Management Strategies.

There is a dataset called PlantVillage which is quite representative of what kind of data should be collected for training our model:
https://www.tensorflow.org/datasets/catalog/plant_village

This dataset contains 54,303 images of different classes containing healthy and diseased leaves of various plants. Some of the classes are:

- Apple_healthy.
- Apple_scab.
- Apple_black_rot.
- Apple_cedar_apple_rust.
- Grape_healthy.
- Grape_black_measles.
- Grape_black_rot.
- Grape_leaf_blight.
- Potato_healthy.
- Potato_early_blight.
- Potato_late_blight.
- Tomato_healthy.
- Tomato_early_blight.
- Tomato_late_blight.
- Tomato_leaf_mold.
- Tomato_mosaic_virus.
- Pepper_healthy.
- Pepper_bacterial_spots.
- Corn_healthy.
- Corn_common_rust.
- Corn_northern_leaf_blight.
- Corn_gray_leaf_spot.

Sample images taken from the dataset:

Now let us discuss the shortcomings of this dataset. This is very important to arrive at a better understanding of what kind of data should be collected, which will determine how well our product will work:

- That the photos in this dataset are taken in controlled lab environments where the background is uniform. Whereas the photos taken in the farms will have far more complex backgrounds. Also, merely by looking at the image, the model might learn that there is a disease but not exactly where the disease is affecting the plant. Diseases can affect different parts of the plant like fruits, stems, leaves, etc. All that must be accounted for while preparing the dataset.
- Another major drawback of this dataset is that each class has an equal number of images which is very difficult to achieve practically. Some diseases might be quite rare and endemic and it can be quite challenging to do proper data collection for each class.
- When each class has an unequal number of images then the classification problem becomes quite hard to solve.

There is another dataset called PlantDoc which addresses some of the issues faced while working with PlantVillage dataset[3]. However, it has very few images (2598) compared to PlantVillage, has fewer classes and is also an imbalanced dataset (different number of images per class).

5. Benchmarking:

There are two apps in the market which address the customer needs which we are also trying to address:

- Plantix: It can detect more than 500 plant diseases for 30 major crops. This app can quickly diagnose plant-diseases by processing images. It also contains information about fertilisers, pesticides, seeds and nearby stores to purchase them. Our prototype takes the most inspiration from the plantix app.

One feature which is not there in this app but can be included in our product is that the app should be able to show that there is no disease detected in the images where there are no plants or leaves, instead of just showing “Unknown disease”.

Also, the app does not cover enough varieties of the crops (only 30), so we can collect images of more varieties of crops and of the diseases which affect them.

- Agrio: Although this app is not as popular as Plantix, it has some additional features compared to it like accurate weather prediction, digital scouting and life-cycle tracking of pests, compared to Plantix. Notably it leverages satellite imaging for plant health monitoring.

6. Applicable Patents:

On searching for patents in our product area, we found patents which address only specific problem statements which are different from what we are trying to solve with our product. However, we also found one patent which can be incorporated into our product:

- **Patent CN114140403A:** Plant leaf disease detection method based on convolutional neural network[4].

This technique explained in this patent is similar to the one which we want to implement in our product. So we can incorporate this patent if required.

7. Applicable Regulations/Laws:

- The Digital Personal Data Protection Bill, 2022[5].
- The Pesticides Management Bill, 2008[6].
- The Fertiliser(Control) Order, 1985[7].

Considering that our product will be mainly used in India, we are considering primarily those laws which are enacted by the Government of India. If our product is launched in other countries, then we should also consider the laws enacted by their governments too. Some of them are listed below:

- Data Protection Laws and Regulations USA 2022-23[8].
- US Federal Insecticide, Fungicide and Rodenticide Act[9].
- Data Protection Laws in the EU[10].

8. Applicable Constraints:

- Collecting images of various types of crops and diseases which affect them is a laborious task. First we need to identify the common types of crops and all the diseases which account for almost all the infections of the crops.
- Another requirement is that we need people with domain knowledge to correctly identify the disease in the images and label them correctly.
- We also need to collect data from experts on how to manage each disease for each plant, and which pesticides and fertilisers to use.
- As the diseases may mutate, new diseases may emerge and old diseases may become too less spread to be irrelevant over time. The dataset must be updated from time to time.
- People who are experienced in working with computer-vision models are needed to work on the product, and also keep making changes in the model and its dataset to improve performance.
- It must be kept in mind that the application is real-time. So, even though we have used a deep learning model for classification, the speed of prediction must be high to give a better user experience.

- The average user is not so tech-savvy, so the UI must be as simple as possible and easy to use.

9. Business Opportunity:

- Agriculture is the most important sector of India's economy which uses a large portion of arable land in the country.
- Millions of farmers have to manually check their crops for any symptoms, but if the product is adopted, the diagnosis becomes more convenient and efficient.
- Given the proportion of India's workforce engaged in agriculture, the number of smartphone users and the low cost of data access, the product sees a huge potential for wide adoption.
- As the product gets widely adopted, large scale data collection can be done. The data collected is of high value for research and policy making given the current scenario of climate change, soil degradation and population growth creating more pressure on the agriculture sector.

10. Business Idea:

- We will follow the B2C business model.
- The app in which the model is to be deployed will be free for the users.
- The one possible way of monetisation in this business model is to provide access to the plant disease data that is being collected to various stakeholders at a certain price.

Data that is being collected about the crops when the users take photos through the app will be of different types.

- First, the photos uploaded by the user will be collected in a database.
- Data will also be collected about the user's location and this will be used to create mapping of the user's history of dealing with crop diseases, with his location. This will lead us to the correlation between disease occurrence and geography.
- Also, the data on which crops which are grown by the user at what times of the year will be collected implicitly along with the raw data from the user.

- This data is of very high value and can be shared to various institutions, research groups and governments to aid them in policymaking with respect to agriculture. As climate change and environmental pollution will drive changes in cropping patterns of the farmers. The data collected will be very important to track such changes and make necessary interventions.

It should be noted that in the initial stages of the product, accuracy is not going to be high, so it would be counter-productive to charge a fee from the users. Once the product is widely adopted, a lot of data collected to enrich the dataset used to train the model and also improvements made in the model itself, it would then be feasible to charge a fee from the users.

Another monetisation model while keeping the application free for the end users, is to display advertisements, but in a non-intrusive way (on the side of the display, without occupying too much space, and also without interrupting the user). The amount of revenue generated from this model depends upon how widely the application is used.

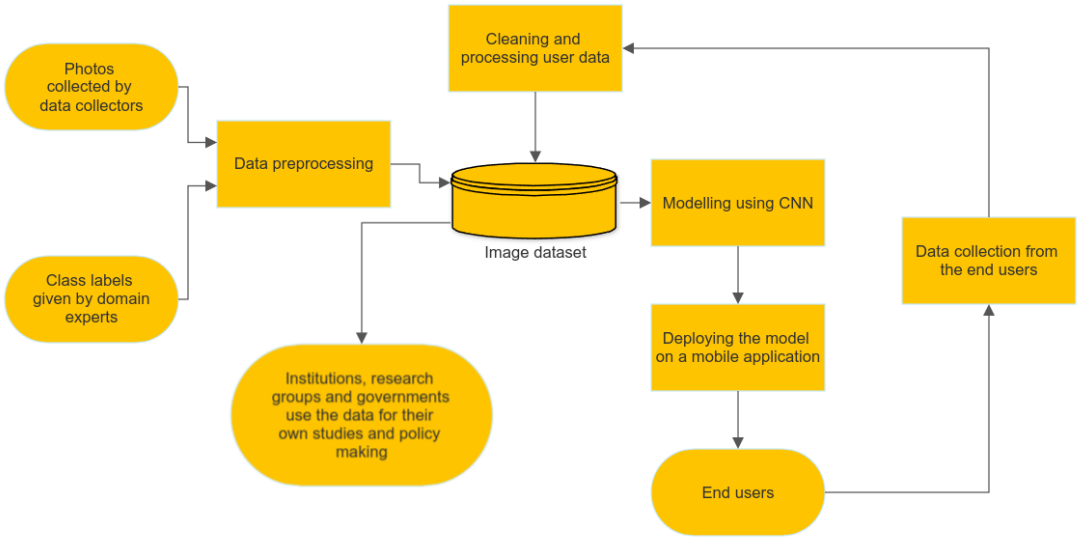
11. Concept Generation:

The following is the process by which we came up with the product idea:

- The problem which we are trying to solve is to identify whether a plant is affected by a disease or not based on its input features. So it is a classification problem.
- This classification problem can be solved in various ways. For example, we can train an SVM which takes in numerical input entered by the user and outputs the class whether the plant is healthy or affected by a disease.We can also use satellite imaging for the same.
- However, we must keep in mind that the app must be as easy to use as possible. Handling of numerical values by a not so tech savvy user can be highly error-prone and can therefore give wrong outputs. The user might not also be able to handle satellite imagery easily.
- After exploring a few ideas on how to solve the classification problem like the ones discussed above, we came to the conclusion that identifying the disease using convolutional neural networks(CNNs) on crop images is currently the best option to solve the classification problem in the context of our product design.

12. Product Prototype:

The following is a schematic of the product we would like to build:



Apart from the main feature discussed above (classifying the image as healthy or diseased), we can also have the following features in our product:

- A discussion forum where farmers can discuss the problems affecting their crops.
- Information about nearby stores which sell fertilisers/pesticides recommended to the farmer.

13. Conclusion:

The development of AI will have a great impact on the agriculture sector, contributing to food security and financial well-being of the farmers. Implementing our product which uses deep learning for plant disease prediction has the potential to revolutionise food production, because it gives the following benefits:

- Reduction of unnecessary pesticide usage, which is good for public health in the long run.
- As there is less requirement to physically handle the infected plants (like touching), and there is less usage of pesticides and fungicides, it is economically desirable and involves less occupational hazard for the farmer. Also, usage of less pesticides and fungicides also means less chances of pathogens developing resistance.
- Improvement in the quality and the quantity of the crop yield, which is economically desirable for the farmer.
- The farmers can track the seasonal patterns of how various diseases spread, and take appropriate precautions.
- Crop disease epidemics can be tracked and can be dealt with much faster.
- Reduction in delay in identifying emerging strains of diseases.

Given the fact that a large section of India's workforce is engaged in agricultural activities, wide adoption of smartphones as well as low cost of internet access, the proposed product idea can be converted to a feasible business model.

14. References:

1. https://www.cphdforum.org/index.php/2022/05/26/plant-disease-crop-loss/
2. Plant Disease Management Strategies.
3. PlantDoc Dataset.
4. Patent CN114140403A: Plant leaf disease detection method based on convolutional neural network
5. The Digital Personal Data Protection Bill, 2022.
6. Pesticides: Law, Registration and FDI in India.
7. The Fertiliser(Control) Order, 1985.
8. Data Protection Laws and Regulations USA 2022-23.
9. US Federal Insecticide, Fungicide and Rodenticide Act.
10. Data Protection in the EU.
11. How do free apps make money?