8

A Recommended System for Crop Disease Detection and Yield Prediction Using Machine Learning Approach

Pooja Akulwar

Department of Computer Science and Engineering, Sanjay Ghodawat University, Kolhapur, Maharashtra, India

Abstract

Agriculture is the mainstay of a rising economy in India. Traditionally farmers followed ancestral farming patterns and norms. However, a single farmer cannot be expected to take into account all innumerable factors that contribute to crop growth. A single misguided or imprudent decision by the farmer can have undesirable ramifications. With the advancements in various domains, intelligent agricultural system is needed for upliftment of Indian economy. The collaboration of recommender system with machine learning will lead to Intelligent Agriculture System that helps the farmer community in their decision making of farm management and agribusiness activities such as i) Predicting agriculture commodity market price before cultivation, ii) Determining best cultivars to plant iii) Determine optimum cultivation date v) Evaluate demand and supply risk vi) Investment Prioritizing. It also helps farmer to perform the activities like crop management including applications on yield prediction, disease detection, weed detection, crop quality, and growth prediction etc. This chapter describes the case study on "Crop Disease Detection and Yield prediction". The study includes identification of crop condition, disease detection, prediction about specific crop and recommendation using machine learning algorithms. It gives an idea about how recommender system is used in agriculture for disease detection and prediction.

Keywords: Agriculture, machine learning, recommender system, crop yield prediction, disease detection

Email: poojaakulwar13@gmail.com

Sachi Nandan Mohanty, Jyotir Moy Chatterjee, Sarika Jain, Ahmed A. Elngar and Priya Gupta (eds.) Recommender System with Machine Learning and Artificial Intelligence: Practical Tools and Applications in Medical, Agricultural and Other Industries, (141–164)

8.1 Introduction

Agriculture is the fundamental source of food industry. It is one of the most oldest and important economic activities which is being practiced in the world wide since thousands of years. Its development has taken over the period of many years with the emergence of new technology, equipment, techniques of farming and domestication. Huge advancement and growth can be seen in this sector with the time period. This sector, not only witnessed the enormous growth but also gave rise to many other sectors with significant progress. Majority of the people are being involved in this occupation as it is the basic need of human beings survival. More than 50% of the land in the world has been devoted to agriculture. Agriculture sector accounts for 14% of Gross Domestic Product (GDP) of the Indian economy. About 70% of the population of India lives in rural areas and majority of them depend upon agriculture as their primary source of income. Agriculture not only helps people to survive but keeps economy on-going. It plays vital role in the economic development of India. Government of India has shown concern about the improvement of cultivator's knowledge of the soil, improvement of the fertility of the soil, irrigation facilities, fertilizer utilization, cattle-manure utilization, precise pesticides usage and grazing in forest area. Thus productivity has to be increased with the increase in population.

In agriculture planning to obtain maximum crop yield with restricted area of land is the largest task in an agro-based country like India. Yield rate of the crop can be increased with the help of indicators by investigating crop related problems. Crop selection will be more accurate and beneficial with minimum loss, whether unfavorable condition occurs [10]. Maximum crop yield can be obtained in favorable growing condition. Improving production rate of crop can be an important topic for research for the agro-meteorologists, for the development of economic growth of the country. The two main factors responsible for the yield rate of the crop is, first one is quality of seeds which can be improved by genetic development using hybridization technology and second one is the selection of crop based on the favorable and unfavorable conditions. The two techniques: statistical and machine learning both these techniques modelled. Many researchers had been tried to get an efficient and accurate model for crop yielding prediction, soil classification, crop classification, weather predictions [13], crop disease prediction classification of crops [5, 6]. Thus this new method called crop

selection method (CSM) developed to increase in net yield rate of crops over seasons.

Crop production rate depends on the topography and geographic condition of the region (e.g. mountainous region, hilly area, river ground, depth regions), weather condition (e.g. humidity, temperature, rainfall, cloud) [14, 15], soil type (e.g. sandy, clay, peaty, saline, silty, loam soil), soil composition (e.g. PH value, nitrogen, phosphorous, magnesium, calcium, sulphur, potassium, organic carbon, copper, iron) and harvesting methods. Different prediction models are used for different parameters of different crops. Some of these prediction models are studied thoroughly through researches for the crop production. The prediction models are of two types: statistical model and machine learning. This chapter describes overall study on Machine learning concepts, algorithms, and methods. This also gives an overview of recommender system and how recommender system is used in agriculture for disease detection and prediction. The chapter also describes crop management activities such as crop yield prediction, disease detection, weed detection and crop quality. Lastly application of agriculture and recommender system is discussed.

8.2 Machine Learning

8.2.1 Overview

Machine learning is emerging technology day by day in different fields. But now-a-days agriculture is the sector where machine learning applications are in greater demand. Now the question arises what is machine learning? Actually it is nothing but machine that learns from experiences in order to perform specific task. It provides ability to learn. Huge real time data set is provided to the system. This data set contains set of attributes called as features. Learning uses these features for further analysis. The performance is measured with performance metric. As more data is added over a time, this performance metric is improved with experience. Various mathematical models are used to calculate performance of machine learning algorithm.

Machine learning tasks are broadly classified into supervised learning, unsupervised learning, reinforcement learning, Semi-supervised learning and learning models (classification, regression, clustering, and dimensionality reduction). Table 8.1 describes difference between supervised learning and unsupervised learning.

Table 8.1 Difference between supervised learning and unsupervised learning.

Factors	Supervised learning	Unsupervised learning
Input	Known and labeled data	Unknown data
Complexity	Very complex	Less Complex
Number of classes	Known	Unknown
Accuracy	Accurate and reliable	Moderately Accurate and reliable

1. *Supervised learning:*
   In supervised learning labeled dataset is used. Here the model is trained on labeled dataset. This dataset contains both input and output parameters. Supervised learning is learning where there is input data, output data and algorithm that maps to input and output. Learning means input and output is provided to machine and machine will develop its own logic for the given task. Some supervised learning algorithms are: Linear Regression, Nearest Neighbor, Guassian Naive Bayes, Decision Trees, Support Vector Machine (SVM), Random Forest, etc.

   The learning is called supervised learning because it is similar to a teacher who is acting as supervisor on entire learning process. The predictions on trained data are generated by learning algorithms. These predictions are corrected by the teacher and learning process is stopped when correct output is achieved [5, 6].
2. *Unsupervised learning:*
   Unsupervised learning is learning that contains only input data and no output data is present. This is applied where there is need to model data distribution inorder to get more and more data and there is no any supervisor (like teacher) to supervised the things. Algorithms themselves learn, discover and present structure in data. Here algorithm itself create data pattern. Some recommendation systems for marketing automation use this type of learning.
3. *Semi-supervised learning:*
   The supervised learning has disadvantage that it required labeled dataset. This process is very costly while dealing with large volume of data. Unsupervised learning also has disadvantage that the range of its application is limited. To solve these problems, semi supervised learning algorithms

concept was developed. This concept used both labeled data and unlabeled data so that it can work on any type of data. Mostly it contains small amount of labeled data and huge amount of unlabeled data.

8.2.2 Machine Learning Algorithms

There are various machine learning algorithms as described in Figure 8.1 below. These algorithms can be applied in any area to solve various problems.

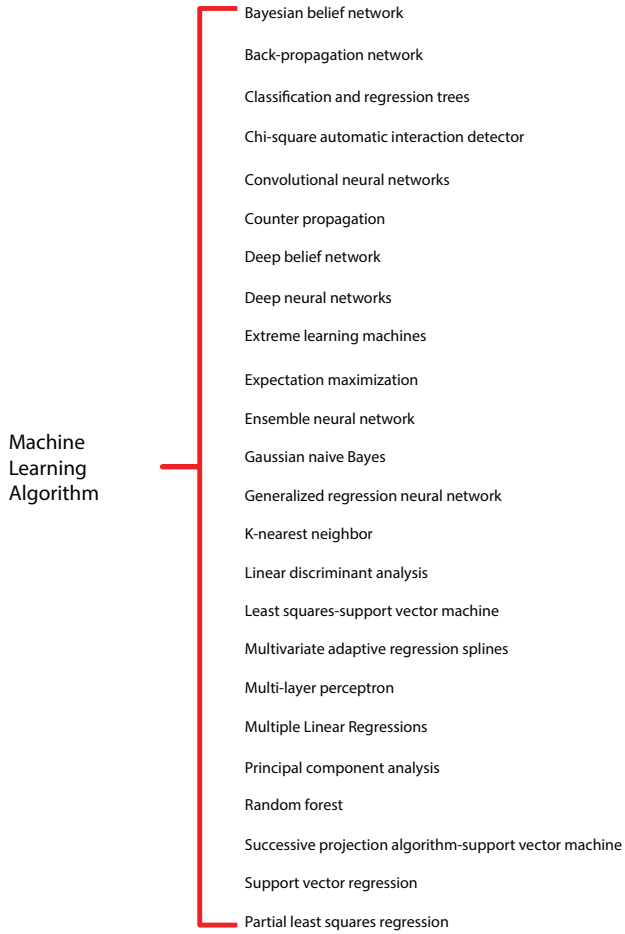

Figure 8.1 Machine learning algorithms [8, 9].

8.2.3 Machine Learning Methods

There are various Machine learning methods such as boosting techniques (RGF, GBDT, and Ada boost), Regression Tree (ID3, C4.5), Random Forest, SVM, K Nearest Neighbor, and ANN, etc.

8.2.3.1 Artificial Neural Network

Artificial Neural Network is a simple mathematical model of the brain. This is used to process nonlinear relationships between inputs and outputs in parallel form, for example human brain. So Artificial Neural Networks can be used in different variety of tasks one of the best use is classification. We can learn Artificial Neural Network speedily. The Information flows through a neural network in two different ways. Firstly when the model is learning or operating normally, the information from the dataset is given to the network through the input neurons, which then trigger the layers of hidden neurons, and then it is converted to the output neurons. So this is called as feed forward network. Each neuron receives inputs to its left, and then they are multiplied by the weights. So every neuron adds up all the inputs. If the sum is more than a certain threshold value then the neuron "fires". Whenever we use large datasets, the neural networks are more powerful at that time.

So in the basic structure of an Artificial Neural Network we can create 3 layers of "neurons"- The input layer, the hidden layer and the output layer as shown in Figure 8.2. The information flows from the input layer, from the hidden layer to the output layer. As each of the connections has a number associated with it, and it is called the connection weight. Also, each of the neurons has a number and a special formula associated with it called as threshold value. The neural network can be trained and then it can be provided with a set of inputs and outputs. Each neuron transforms the input and forwards it to the next layer and so on. The result is received on the output layer. Then the layer is compared to the outputs through special algorithms is used to produced outputs as close to each other as possible. As this process is repeated many times and here completes the training part. Whenever new inputs are provided to the network, we can get the actual outputs. So, the Artificial Neural Network can be used in predicting house prices and classifying objects and images.

8.2.3.2 Support Vector Machines

SVM stands for Support Vector Machine. It is a simple algorithm in machine learning. It belongs to the supervised learning category in

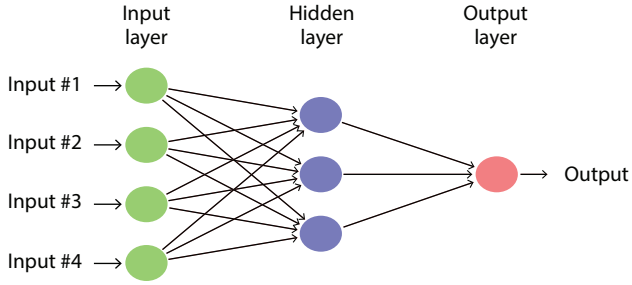

Figure 8.2 Simple Artificial Neural Network [11].

machine learning which is used for both regression and classification analysis but it is a discriminant classifier which is widely used as classification algorithm. It produces significant accuracy with less computation power. This algorithm is used to create a line or a hyper plane in order to separate the data into various classes. It takes the data as input and outputs a line that separates two classes. It is used in variety of applications such as face detection, classification of emails, news articles, web pages, handwriting recognition.

8.2.3.3 K-Nearest Neighbors (K-NN)

K-nearest Neighbor is supervised algorithm under machine learning. It is very easy to implement. It can give huge great classifiers. For classification and predictive problems, KNN is mostly used. It is also called as sample-based learning technique. For predictive target value, it uses all the past data. It uses distance function. There are various distance functions such as Euclidean, Manhattan, Makowski, etc. It computes distance between new input value (sample predictor) and all training set (sample predictor). After this smallest distance (k nearest distances) are selected. The sum of all the k neighbors is computed and target value is determined. This algorithm does not require training and any optimization method. This method is used for nonlinear and adaptable problems. KNN time and space complexity is very high because it uses all data samples while predicting target values. KNN can be used in agriculture efficiently.

8.2.3.4 Decision Tree Learning

In decision tree learning sampling of data is considered. Data is split into smaller sub sample space. The root of the tree contains entire sample data.

Other sample data is present into sub nodes. With the help of fork process, other nodes (children nodes) are created. After forking, during child creation data is split into smaller sub samples. This procedure is done recursively till last possibility. Data is split based upon some conditions provided by input attributes. The output value is assigned to those input values which exist between root and leaf node. The main aim is to combine multiple output values and allocate single output value to sample space. Size of node needs to be considered during splitting. For building decision tree, various algorithms such as CART, M5, M5 Prime, etc. are used. All these algorithms are similar but difference lie in impurity measures, leaf value assignment and prune rule.

8.2.3.5 Random Forest

A Random Forest algorithm is a supervised classification algorithm. It is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap Aggregation, commonly known as bagging. It creates a forest by some way and makes it random. There are multiple decision trees. As per the name, it creates forest first and makes it randomly used as shown in Figure 8.3. The idea is to combine these trees and generate output instead of considering only individual tree. Applications of Random Forest algorithm are Banking, Medicine, Stock Market and E-commerce.

There are two stages in Random Forest algorithm, one is random forest creation, and the other is to predict classifier which was created in first stage. This works on various features such as binary, categorical and numerical.

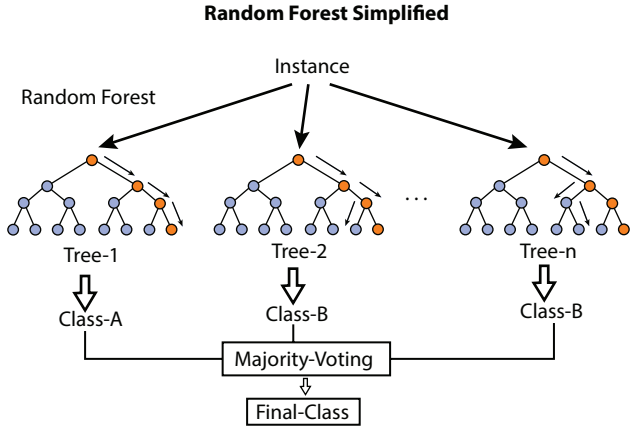

Figure 8.3 Simple random forest concept [11].

Very less processing work is required for this. The data does not need to be rescaled or transformed. Random forest improves on bagging because it correlates the trees with the introduction of splitting on a random subset of features. This means that at each split of the tree, the model considers only a small subset of features rather than all of the features of the model. That is, from the set of available features n, a subset of m features (m = square root of n) are selected at random. Bagging, or bootstrap aggregating, is where bagged trees are created by creating X number of decision trees that is trained on X bootstrapped training sets. The final predicted value is the average value of all our X decision trees. Bootstrapping is a sampling technique in which we randomly sample with replacement from the data set.

The predictive performance of Random Forest can compete with the best supervised learning algorithms. They provide a reliable feature importance estimate. Random forest handles outliers by essentially binning them. It is also indifferent to non-linear features. It is faster to train than decision trees because working is done only on a subset of features in this model, so it can easily work with hundreds of features. Prediction speed is significantly faster than training speed because generated forests can be saved for future uses.

8.2.3.6 Gradient Boosted Decision Tree (GBDT)

Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. Like other boosting methods, GBDT also build the model according to some stages. Generalization will take place by optimizing differential loss function.

The Gradient Boosting Decision Tree (GBDT) is a popular machine learning algorithm. It has few effective implementations such as XGBoost and pGBRT. The Gradient boosting decision tree is a widely-used algorithm, due to its efficiency, accuracy, and interpretability. GBDT achieves state-of-the-art performances in many machine learning tasks, such as multi-class classification, click prediction, and learning to rank. Nowadays GBDT is facing new challenges, especially in the tradeoff between accuracy and efficiency. So the Conventional implementations of GBDT need for every feature to scan all the data instances and to estimate the information and gain of all the possible split points. Therefore, their computational complexities will be proportional to both the number of features and the number of instances. Handling of big data makes implementation very time consuming.

There are some techniques used in gradient boosting decision tree-Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). GOSS keeps all the instances with large gradients and performs random sampling on the instances with small gradients. In order to compensate the influence to the data distribution, when computing the information gain, GOSS introduces a constant multiplier for the data instances with small gradients. While there is no native weight for data instance in GBDT, data instances with different gradients play different roles in the computation of information gain. Exclusive Feature Bundling (EFB) is usually real application, although there are a large number of features, the feature space is quite sparse, which provides us a possibility of designing a nearly lossless approach to reduce the number of effective features. As GBDT is an ensemble model of decision trees, which are trained in sequence. In each iteration, GBDT gives the decision trees by fitting the negative gradients. Specifically, in a sparse feature space, many features are exclusive, i.e., they rarely take nonzero values simultaneously e.g., one-hot word representation in text mining.

Gradient boosting algorithm uses gradient descent method to optimize the loss function. This algorithm is simple and can find the optimal split points; however, it is inefficient in both training speed and memory consumption. The main cost in GBDT is to learn the decision trees. The most time-consuming part in decision tree is to find the best split points.

8.2.3.7 Regularized Greedy Forest (RGF)

Regularized Greedy Forest (RGF) is a tree ensemble machine learning method that works directly with the underlying forest structure. Regularized greedy forest integrates mainly two ideas- one idea is to include tree-structured regularization into the learning formulation and the other idea is to use the fully-corrective regularized greedy algorithm. RGF does the change to existing forest and new forest is obtained which helps to minimize loss function. It also adjusts the weights of leaf to minimize loss function. Regularized greedy forest is nice algorithm that sits in between gradient boosting algorithm and random forest. The implementation can be done using different languages like Python, C, C++ etc. Regularized greedy forest algorithm builds decision forest through fully corrective regularized greedy search to underline the forest structure. It gives the higher accuracy and works faster. In RGF globally optimized decision tree is formed. In GBDT locally optimized decision tree is formed. As there is tree structure, RGF works by utilizing fully corrective regularized structure.

But GBDT does not consider full approach. It works on partially regularized tree structure.

8.3 Recommender System

8.3.1 Overview

In the world of internet and technology, data is growing tremendously with high speed. Contents are increasing along with number of users. Due to this enormous information, people face problem in finding right information at a specific time. People are in need of system that will give suggestion to them for their work. And this situation has made the evolution of concept of Recommender system. Recommender system guides users by providing recommendations about problem. It works on the concept of "most likely to be interesting" or "Relevant to need". These systems are widely used in many areas such as ecommerce websites, e-learning, e-library, e-government and e-business etc. Recommender system requires huge past data so that based upon the behavior and experience recommendations can be made.

Consider a recommender system that takes image as input. The general working of recommender system is shown in Figure 8.4. System will do image preprocessing that includes removal of noise and extraction of region of interest. Feature extraction will be followed after preprocessing. After features are extracted, machine learning algorithm is applied to and the results are predicted. The recommender system is used to provide the best recommendations.

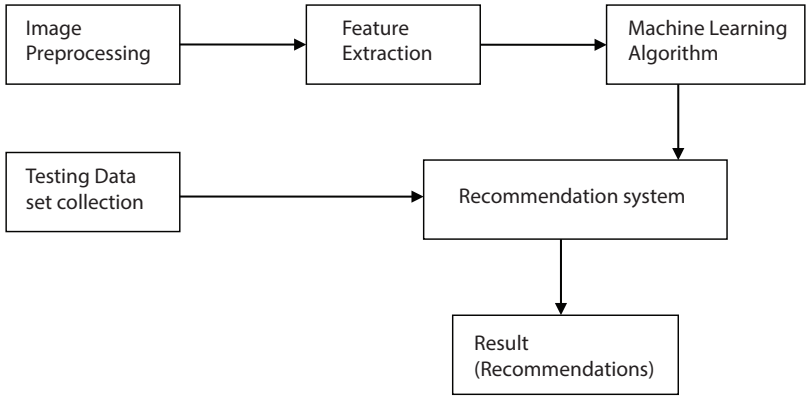

Figure 8.4 General overview of recommender system in agriculture [3].

Now these systems are used in agriculture sector too. Various recommendations are need to farmers at each and every step from sowing of seed till harvesting [7]. As farmers are mostly unaware of current updates in farming recommendation system helps them a lot. Recommendations regarding crop selection, disease identification and solution etc. can be generated and given to farmer. Consider the following scenario:

Scenario 1: Crop identification and recommender system

With the help of advanced technology, lot of improvement has been done in agriculture sector. When a farmer purchases a land for farming the most important question arises is which crop to be grown to have better production? If the selected crop is unsuitable for that land, then there will be less productivity. Generally, farmers are unaware about agriculture land requirements that is minerals needed, soil moisture and other soil requirements. The concept of precision agriculture solves this problem of farmer. In this characteristic of soil are identified and used for identifying the suitable crop for cultivation. Soil characteristics are soil type, texture, pH value, moisture, temperature and other environment conditions. Due to this risk of cultivating improper crop is avoided. Always farmer will cultivate proper crop that results in good crop yield and get enormous profit for particular land area. Here recommendation system helps in recommending the most suitable crop that yields in better productivity. Also, it prevents from crop loss making farmer to be financial stable.

Scenario 2: Disease detection and recommender system

Crop productivity depends upon the crop or plant status (healthy or unhealthy). If the plant has any kind of disease then it affects crop quality and production. For yielding good crop, disease detection plays vital role in agriculture. One of the problems for farmer is to control pest and diseases affecting that crop. Once the crop seeds are sown, it should be prevented from pest and disease [1]. Traditionally disease detection was done manual which was not accurate and was time consuming. Farmers did not have much knowledge about disease and unable to explain information of disease properly. Call centers are available for farmer that helps to provide solution about diseases. As farmers are unable to provide proper information, their solution does not work. Hence the problem still continued. But now with increasing technology, automatic disease detection is possible effortlessly. This disease detection can be done based upon images that are provided. Various

applications have been developed to recognize and detect diseases that will lessen negative impact on harvest. After disease detection the solutions to overcome this disease must be provided. Hence recommendation system role is very important in this respect. Recommender system helps to provide the best solution to overcome this disease. But this solution is after detection disease.

Here the recommender system can work in two ways:

1. Before disease detection

Before disease detection results in precaution taken before the disease has occurred. Once the image preprocessing and feature extraction is done, the analysis or difference in features can predict the type of disease the crop will be affected from in future. If this information is known then precautions can be taken in advanced so that crop will remain unaffected and hence productivity will not be lessened.

2. After disease detection

Once the disease is detected, it must be cured properly to increase productivity. Recommender system helps by providing best solution to overcome diseases. This system analyzes the data such as symptoms, type of disease, medical treatment etc. based upon the available data, the system recommends the best possible solution to cure disease.

8.4 Crop Management

8.4.1 Yield Prediction

The most important topic in agriculture is yield prediction because yield prediction contributes to profit [5]. It includes activities such as yield mapping and estimation, pairing crop supply with demand and crop management. All these activities lead to increase production. One of the low costs and efficient machine learning application was developed that contributes in automatically counting coffee fruits on branch of a tree. The calculation of coffee fruits is classified into three types: fruits that are harvestable, fruits that are non-harvestable and fruits that are ignored from stage of maturation. This system also calculated the weight of coffee fruit and maturation percentage. This system aims to help the factory owner who wants to grow coffee by providing the best economic benefits and also plan. The machine learning application named a machine vision system was developed that automatically shake and catch cherries while harvesting process. This system

does segmentation and detects the obstructed branches even though these branches are full of leaves and not visible clearly. This system aims to reduce labor work and manual handling of operation during harvesting.

One of the systems introduced was yield mapping system that identifies immature green citrus. This system aims to provide information of good citrus in citrus groove that will result in better profit and yield. This is very useful system for citrus growers. Based upon the Artificial Neural Network, a new model for calculating grassland biomass was developed. The study was also done on yield prediction. For study, wheat crop was considered and images were taken with the help of satellite. Using image processing crop growth characteristics were identified and data were fused to predict the result. Another application was developed to detect the tomatoes using images that were sensed remotely. These images were captured using Unmanned Aerial vehicle. Based on SVM, new rice development stage prediction model was developed. All the basic information that is required for processing was obtained from weather forecasting station in China. The study was done that focuses on helping farmers to overcome imbalances that happened in the market supply and demand. This may be due to quality of crop. A generalized method based on ENN application was developed for good agricultural yield prediction. The following table describes study about various crops along with their functionality, algorithms and results. Table 8.2 describes study about various crops along with their functionality, algorithms and results.

8.4.2 Disease Detection

Disease detection is very significant part in agriculture domain. If crop is affected by the disease it will lessen production resulting into profit decrement. Hence there is need to control pest and diseases in open farm and also in green house. To control pest and diseases, spraying pesticides is the most commonly used method. This method is costly and causes environmental hazards such as ground water contamination, problems to local wildlife, and other impact crop quality etc. excessive use of pesticides also decreases quality of crop and soil becomes barren. There is need to know exact quantity of adding pesticides as excess quantity causes side effects. To avoid this, machine learning applications can be used in agriculture that will identify disease automatically and provide the solution. The system was developed to automatically detect plant affected by fungus and also classify infected plant and non-infected plant. Another system was developed using image processing technique that classify parasites and detect thrips in the strawberry. This experiment was done in green house

Table 8.2 Summary of crops with their functionality and algorithm [8].

Crop	Functionality	Algorithm	Result
Coffee	Describes count of coffee fruit automatically	Support Vector Machine	Harvestable: Ripe/overripe: 82.54–87.83% visibility percentage Not harvestable: Unripe: 76.91–81.39% visibility percentage
Cherry	Detect cherry branches and foliage	Gaussian naïve Bayes	89.6% accuracy
Citrus	Detect immature green citrus	Support Vector Machine	80.4% accuracy
Grass	Calculate biomass in grassland	Artificial Neural Network	R2 = 0.85 RMSE = 11.07
Wheat	Predicted wheat yield	Artificial Neural Network	81.65% accuracy
Tomato	Detected tomatoes	Clustering	Recall: 0.6066 Precision: 0.9191
Rice	Predicted rice yield	Support Vector Machine	Middle-season rice: Tillering stage: RMSE (kg h^{-1} m^2) = 126.8 Heading stage: RMSE (kg h^{-1} m^2) = 96.4

environment. The system was presented to detect Bakanae disease in rice. The main aim was to detect pathogen into different rice. As these diseases are detected automatically, it saves time in identifying disease with naked eye. It also reduces manual work. Ultimately increase production with good quality yield.

Lot of studies has been done for disease detection on wheat crop. Wheat is given major importance among all crops. Hence to produce healthy wheat crop disease detection proved to be very essential and significant. A new system has been developed that classify healthy wheat and infected wheat due to nitrogen and yellow rust. The hyperspectral reflectance imaging technique is used to identify disease. The author described the system that accurately detect the disease and also provide solution about usage of

fertilizers, Pesticides, fungicides etc. depending upon plant's requirement. Another study describes system that differentiate healthy wheat crop and infected wheat crop with Septoriatritici blotch (STB). Another research was done to develop system for identifying infected wheat by yellow color rust and healthy wheat. This was done using machine learning algorithm named SVM classifier along with data fusion technique. The similar kind of study was done using ANN models and considering spectral reflectance features. Development of such system is needed for accurate disease detection and providing solution earlier in order to produce better yield.

Mostly wheat is affected with diseases such as yellow rust, Septoriatritici blotch, fungus, etc. To detect yellow rust infected wheat, crop a real time system was developed which was handles remotely. This system uses Neural Network technique and data fusion technique. Data fusion also includes data obtained from hyper-spectral reflection and multi-spectral fluorescence image. The system was also developed to identify stress in plant and also nutrition deficiency under certain field conditions. This plant stress is caused by some disease. If plant has stress then there will be fewer yields. Finally, new system was designed that will identify and classify plant diseases using convolutional neural network (CNN). The images of plant leaves are taken as input to the system and image processing. After image preprocessing, features are extracted and CNN algorithm is applied to get desired output. The output will be either plant is infected by disease or healthy. Table 8.3 summarizes various crops and their functionality. It also describes the machine learning algorithm that is used in research along with predicted results.

8.4.3 Weed Detection

Weed is nothing but unwanted plant in cultivated area. If this weed is present in large amount then growth of required crop will be less and hence this will decrease productivity resulting into fewer profit. Therefore, weed detection and removal has obtained greater importance in field of agriculture. Most of the farmers consider this as major threat towards production of desired crop. As weeds are very difficult to identify and distinguish, there is need of accurate weed detection to increase productivity of crop. Various machine learning applications are developed to detect and distinguish weed from actual crops in coordination with sensors. These applications are low cost. Machine learning assist in the development of robots which will help in weed detection and removal that will lessen the requirement of herbicide. A system was presented that is used to identify weed named Silybummarianum using method counter propagation (CP)-ANN

Table 8.3 Summary of crops with their functionality and algorithm.

Crop	Functionality	Algorithm	Result
Silybummarianum	Disease detection and classification between healthy plant and plant infected by fungus Microbotyumsilybum	Artificial Neural Network	95.16% accuracy
Strawberry	Parasites are classified and insects/diseases are detected automatically	Support Vector Machine	MPE = 2.25%
Rice [2]	Detection of Disease (Bakanae, Fusarium)	Support Vector Machine	87.9% accuracy
Wheat	Disease detection (yellow rust Infected) and also nitrogen stressed	Artificial Neural Network/ XY-Fusion	Accuracy Nitrogen stressed: 99.63% Yellow rust: 99.83% Healthy: 97.27%
Wheat	Disease detection (Septoriatritici infected) and also water stressed	Support Vector Machine/ least squares-support vector machine	Inoculated treatment, with Septoriatritici 98.75% accuracy

(*Continued*)

Table 8.3 Summary of crops with their functionality and algorithm. (*Continued*)

Crop	Functionality	Algorithm	Result
Wheat [4]	Disease detection (yellow rust Infected)	Artificial Neural Network/ multi-layer perceptron	Accuracy Infected by Yellow rust: 99.4% Not infected: 98.9%
Wheat	Disease detection in wheat and classify infected wheat and healthy wheat	Artificial Neural Network/ self-organizing maps	Accuracy Infected by Yellow rust: 99.4% Not infected: 98.7%
Wheat	Identify healthy wheat and infected wheat by yellow rust or nitrogen stress	Artificial Neural Network/ self-organizing maps	Accuracy Infected by Yellow rust: 99.92% Nitrogen stressed: 100% Not infected: 99.39%

by capturing multispectral images. For recognition of the crop and weed, new methods based on machine learning techniques are developed. An active learning system was developed to recognize weed and plant named maize. The main aim is to automatically and accurately detect weed.

8.4.4 Crop Quality

Crop quality is very significant factor in agriculture sector that defines the profit for owner. For this accurate identification of crop quality along with its classification is must to judge production cost. Earlier this identification took place manually increasing the time and labor cost. But along with technological development various applications have been developed to accurately detect crop quality. Hence the error rate is minimized as compare to earlier giving exact profit to owner. A new system was designed to detect the botanical material inside cotton. This was done during harvesting period. The study was to improve crop quality and minimize damage to the fiber. Some of the detection system used hyperspectral reflectance imaging technique. The detection was made by capturing images of crop. Similar kind of application was developed by using various machine learning techniques to identify rice samples.

8.5 Application—Crop Disease Detection and Yield Prediction

For crop yield prediction Machine learning plays vital role. Many ML techniques are developed for yield prediction. The comparative studies are done by multiple researchers for defining most accurate technique. As very limited number of crops is evaluated, still exact decision is not achieved.

Production of crop depends upon some factors such as weather condition, soil condition, geographical region, soil composition and also harvesting method. Traditionally, monitoring techniques does not gather the crop conditions properly and prediction results were not yet optimized. Therefore, to overcome this problem, a system has been designed that identified type of crop, crop disease and predict crop yield in different conditions using machine learning techniques. The main aim was to detect disease and predict maximum production of crop using limited land resource.

The study has been done on agricultural land that cultivates strawberry and citrus crop. The dataset contains more than 1,000 strawberry images and citrus images. The image of crop was captured by using camera. Image undergoes preprocessing which includes removal of noise and extracting

region of interest. Features such as color, position and size were extracted by feature extraction process. By using Convolution Neural Network (CNN) algorithm [12], the crop type is identified. CNN algorithm is used to classify the image by its features. After identifying the crop type, crop disease was detected. To calculate crop yield, there is need to remove this affected crop as it will contribute to bad production or affect production. Hence, good crop is selected and crop yield is predicted using machine learning algorithm (Figure 8.5).

In Figure 8.6, Anaconda 3.0 software is used in this system because the machine learning algorithms that we want to use have libraries written in python. The support for python is better than other languages when it comes to tensor flow as it has been around for some time now. Whereas for use of Opencv, Python or C Raspberry pi has good support. As shown in Figure 8.6 the different commands were used: first command is to activate the tensor flow environment. Second Command will set the path where the file is located. Third command will go to actual Keras library. Fourth command will execute actual image from dataset and gives the result.

By using above commands as shown in Figure 8.6, the system has detected the fruit and identified the disease. As shown in Figure 8.7, the system has detected the fruit i.e. strawberry and identified that strawberry has infected.

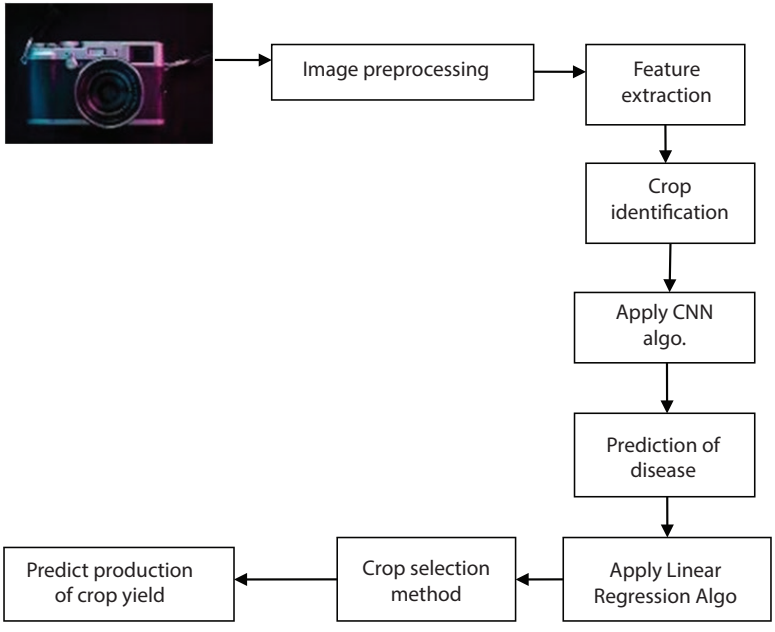

Figure 8.5 Architecture of crop disease detection and prediction system.

Then it gave which diseases strawberry has—whether it is suffering from Gray Mold (gm) or Rhizopus (Rz). First figure has detected gm disease 100%.and Rz disease 0% and the second figure detected Rz disease 100% and gm disease 0%.

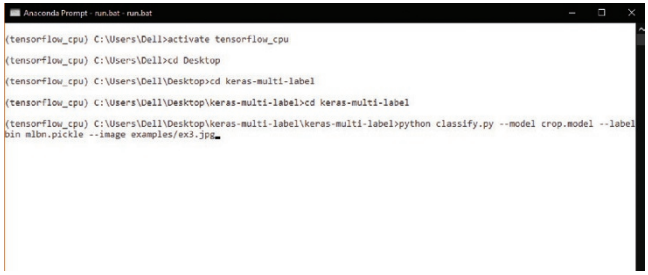

Figure 8.6 Commands to run the model.

Figure 8.7 Fruit identification and disease detection in strawberry crop.

Figure 8.8 Identify and detect disease in citrus canker.

By using above commands as shown in Figure 8.6, the system has detected the fruit and identified the disease. As shown in Figure 8.8, the system has detected the fruit i.e. Citrus Canker and identified that Citrus Canker has infected. Then it gave how much Citrus Canker has infected. As shown in Figure 8.3, Blackspot is 31.29% detected.

References

1. Ebrahimi, M.A., Khoshtaghaza, M.H., Minaei, S., Jamshidi, B., Vision-based pest detection based on SVM classification method. *Comput. Electron. Agric.*, 137, 52–58, 2017.
2. Chung, C.L., Huang, K.J., Chen, S.Y., Lai, M.H., Chen, Y.C., Kuo, Y.F., Detecting Bakanae disease in rice seedlings by machine vision. *Comput. Electron. Agric.*, 121, 404–411, 2016.
3. Pantazi, X.E., Moshou, D., Oberti, R., West, J., Mouazen, A.M., Bochtis, D., Detection of biotic and abiotic stresses in crops by using hierarchical self-organizing classifiers. *Precis. Agric.*, 18, 383–393, 2017.
4. Moshou, D., Bravo, C., West, J., Wahlen, S., McCartney, A., Automatic detection of "yellow rust" in wheat using reflectance measurements and neural networks. *Comput. Electron. Agric.*, 44, 173–188, 2004.
5. Richardson, A., Signor, B.M., Lidbury, B.A., Badrick, T., Clinical chemistry in higher dimensions: Machine-learning and enhanced prediction from routine clinical chemistry data. *Clin. Biochem.*, 49, 1213–1220, 2016.
6. Wildenhain, J., Spitzer, M., Dolma, S., Jarvik, N., White, R., Roy, M., Griffiths, E., Bellows, D.S., Wright, G.D., Tyers, M., Prediction of Synergism from Chemical-Genetic Interactions by Machine Learning. *Cell Syst.*, 1, 383–395, 2015.
7. Kang, J., Schwartz, R., Flickinger, J., Beriwal, S., Machine learning approaches for predicting radiation therapy outcomes: A clinician's perspective. *Int. J. Radiat. Oncol. Biol. Phys.*, 93, 1127–1135, 2015.
8. Craven, B.D. and Islam, S.M.N., Ordinary least-squares regression SAGE. *Dict. Quant. Manag. Res.*, 224–228, 2011.
9. Friedman, J.H., Multivariate Adaptive Regression Splines. *Ann. Stat.*, 19, 1–67, 1991.
10. Alonso, J., Villa, A., Bahamonde, A., Improved estimation of bovine weight trajectories using Support Vector Machine Classification. *Comput. Electron. Agric.*, 110, 36–41, 2015.
11. Alonso, J., Castañón, Á.R., Bahamonde, A., Support Vector Regression to predict carcass weight in beef cattle in advance of the slaughter. *Comput. Electron. Agric.*, 2013.

12. Hansen, M.F., Smith, M.L., Smith, L.N., Salter, M.G., Baxter, E.M., Farish, M., Grieve, B., Towards on-farm pig face recognition using convolutional neural networks. *Comput. Ind.*, 98, 145–152, 2018.
13. Feng, Y., Peng, Y., Cui, N., Gong, D., Zhang, K., Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data. *Comput. Electron. Agric.*, 136, 71–78, 2017.
14. Mohammadi, K., Shamshirband, S., Motamedi, S., Petković, D., Hashim, R., Gocic, M., Extreme learning machine based prediction of daily dew point temperature. *Comput. Electron. Agric.*, 117, 214–225, 2015.