*International Journal of Research in Advent Technology, Vol.7, No.4, April 2019*
*E-ISSN: 2321-9637*
*Available online at www.ijrat.org*

# A Machine Learning Approach for Estimating Crop Damage based on Leaf Disease Detection

Dr.T.M.N.Vamsi[1], Pavan Roy Jarjana[2], Suvarna Mogalapalli [3], Chakradhar Yadav Golagani[4]
*Department of Computer Science and Engineering, Engineering & technology Program*
*Gayatri Vidya Parishad College for Degree and PG Courses (A),Visakhapatnam, AP, India*
*[1]bioprativamsi@gmail.com,[2]rd062595@gmail.com@gmail.com, [3]suvarnamogalapalli18@gmail.com,*
*[4]chakradharyadav2345@gmail.com*

**Abstract-**Farmers rely on agriculture for their livelihood. Efficient production of crop depends on a lot of external factors that may result in loss to the farmer. Fortunately, Insurance companies have come to their aid by providing insurance based on certain criteria. However, this process is time consuming and requires manual verification to validate the conditions required for successful insurance claim. This paper aims to automate this process by enabling farmers to identify the disease or damage that occurred to the crop through the mobile application and obtain fast and secure insurance claims if the damage to the crop is more than the prescribed range. When the farmers have applied for the insurance the insurers will verify the details that are stored in the database through a desktop application and to estimate area affected by damage. This reduces the need for manual verification and increase secure and efficient insurance claims.

**Keywords-**Classification, Neural Networks, Convolutional Neural Networks, OCR.

## 1. INTRODUCTION

Classification is a supervised learning methodology in machine learning in which the computer program studies from the data input given to it and then uses this knowledge to classify new observations. Some of the applications are speech recognition, handwriting recognition, biometric identification, document classification. Different types of classification techniques include Logistic Regression, Naive Bayes Classifier, Support Vector Machine, Decision trees, Random Forest and Neural Networks. Among all these classifiers Neural Networks have the knack to learn and model non-linear and complex relationships by generalizing the model and predict unseen data.

An artificial Neural Networks or Neural network is a simulation of the human brain that constructs patterns from the data. A neural network consists of layers; each layer contains a predefined number of units (neurons). These units convert an input into some output vector. The input layer consists of units equivalent to the number of parts the input is split into (for example in case of an image the number of units in the input layer is equal to the count of pixels in the image and each pixel is given as input to each unit). These units then process the input by applying a nonlinear function and generate an output vector that is passed on to the next layer. In general, neural networks are Feed forward i.e, a unit feeds its output to all the units on the next layer thereby arriving at the point of least error as fast as possible. The weights generated during this process are assigned to the signals passing from units in one layer to the other, and these weights are tuned in training phase to design a model to a specific problem.

Optical character recognition (OCR), is an image processing technique where text from images like ASCII. OCR is mainly used to extract text from printed data records, passport documents, computer generated receipts, business cards, bank statements, mails, books or any suitable documentation. The text obtained in this way can be used in various fields like text mining, text-to-speech translation, machine translation and the data can be stored in a compact manner.

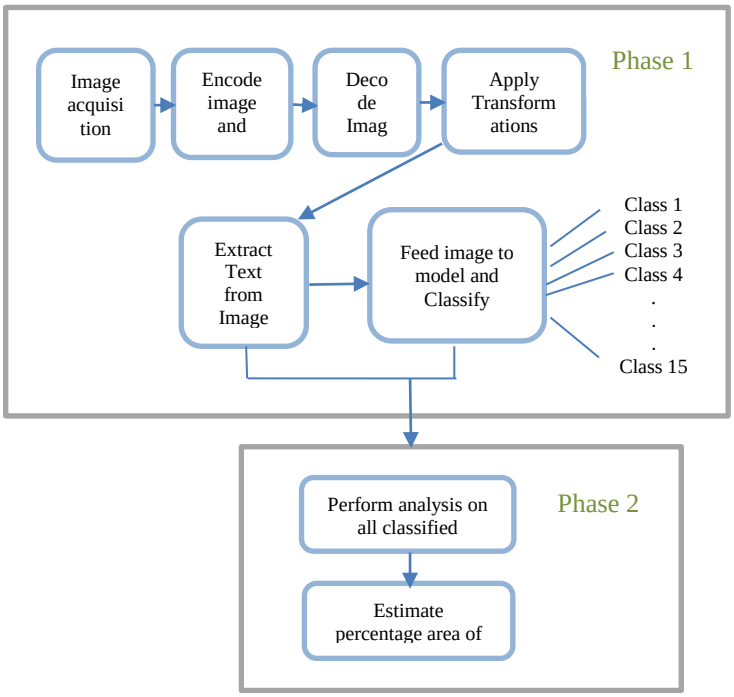


Fig 1: Data Flow Diagram

## 2. EXISTING METHODOLOGIES

Several research ideas have appeared for Agricultural Crop using different technological approaches over the past decades. These approaches address the merits and demerits of technological solutions and are trying to develop new approaches suitable for this domain.

The existing methodologies give farmers a real-time dashboard of soil conditions (moisture), wind speed, and temperature as well as integration with whether forecasts. Integration allows faster claims payments. To insurers it allows them to set policy conditions for risk control and select more complaint farmers and saves cost to mitigate insurance fraud.

Crop Insurance provided by the government which covers losses to farmers crops from natural events like drought, hail, flood and losses due to reduction of crop prices. A variety of Crop Insurance mobile applications can be used to calculate the Insurance Premium for crops taking into consideration various factors like crop area, loan and coverage amounts. But for the diseased crop, the farmer is not getting any insurance from the government since there is no appropriate crop monitoring system to claim the insurance for loss.

In [1], Convolutional Neural Network (CNN) model and Learning Vector Quantization (LVQ) algorithm-based method used for tomato leaf disease detection and classification. Color information is actively used for plant leaf disease researches. The filters are applied to three channels based on RGB components. The LVQ has been fed with the output feature vector of convolution part for training the network. In [2], CNN is proposed to improve the correctness of text recognition. In the proposed system, [1] is referred to detect the disease or damage of the crop and [2] is referred to uniquely identify the crop plants for insurance claim.

## 3. PROPOSED METHODOLOGY

The system has been proposed to provide farmers a means to claim insurance and also to find out what disease the plant has. In the first stage we take the input from the mobile application through camera module and input is used in two ways. The image is encoded and uploaded to the server where it is decoded and fed as input to the classification model where the disease is identified and simultaneously fed to another model where the text is recognized.

So, by performing the disease classification we can tell what disease the plant has and by recognizing the text pattern we can identify the plant id so that multiple images of the same plant can be considered as single image so that the integrity of the system increases and also helps the insurers to reduce their efforts in verifying the details about issuing the insurance.

The details that are entered by the farmers are verified by the insurers so that if the farmer is really eligible to claim the insurance or not.

**Algorithm:**

1. *Input image from camera*
2. *Encode image to base64*
3. *Upload to server*
4. *Transform image to numpy array, crop image*
5. *Classify plant and disease*
6. *Extract text from image*
7. *Map identified plant with text(unique id)*
8. *If image with same unique id exists, overwrite image*
9. *Calculate percentage area of diseased crops in total crop area for each individual crop type*

*Display result on mobile application*

## 4. RESULTS AND DISCUSSION

The disease classification is performed using Convolutional Neural Networks (CNN) and scored 98.53% accurate results. For disease classification, a subset of PlantVillage dataset is considered. Each image is of size 256*256. To plot the confusion matrix and calculate the accuracy 30 images from each class have been used. The model is able to correctly classify at least 28 out of the 30 images given as input.


Fig 2: Pepper bell bacterial spot


Fig 3: Pepper bell healthy


Fig 4: Potato early blight


Fig 5: Potato healthy


Fig 6: Potato late blight


Fig 7: Tomato bacterial spot


Fig 8: Tomato early blight


Fig 9: Tomato healthy


Fig 10: Tomato late blight


Fig 11: Tomato leaf mold


Fig 12: Tomato septoria leaf spot


Fig 13: Tomato spider mites two spotted spider mite


Fig 14: Tomato target spot


Fig 15: Tomato mosaic virus


Fig 16: Tomato yellow leaf curl virus

## 5. TABLES

Table 1. Classification results as confusion matrix

| | Pepper Bell Bacterial spot | Pepper Bell Healthy | Potato Early blight | Potato Healthy | Potato Late blight | Tomato Bacterial spot | Tomato Early blight | Tomato Healthy | Tomato Late blight | Tomato Leaf mold | Tomato Septoria leaf spot | Tomato Spider mites Two-spotted spider mite | Tomato target spot | Tomato mosaic virus | Tomato yellow leaf curl virus | **F-score Accuracy** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pepper Bell Bacterial spot | 29 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 96% |
| Pepper Bell Healthy | 0 | 30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100% |
| Potato Early blight | 0 | 0 | 30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100% |
| Potato Healthy | 0 | 0 | 0 | 28 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 94% |
| Potato Late blight | 0 | 0 | 0 | 0 | 30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100% |
| Tomato Bacterial spot | 0 | 0 | 0 | 0 | 0 | 29 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 96% |
| Tomato Early blight | 0 | 0 | 0 | 0 | 0 | 1 | 28 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 96% |
| Tomato Healthy | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100% |
| Tomato Late blight | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 30 | 0 | 0 | 0 | 0 | 0 | 0 | 100% |
| Tomato Leaf mold | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 30 | 0 | 0 | 0 | 0 | 0 | 100% |
| Tomato Septoria leaf spot | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 30 | 0 | 0 | 0 | 0 | 100% |
| Tomato Spider mites Two-spotted spider mite | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 28 | 0 | 0 | 0 | 96% |
| Tomato target spot | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 30 | 0 | 0 | 100% |
| Tomato mosaic virus | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 30 | 0 | 100% |
| Tomato yellow leaf curl virus | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 30 | 100% |
| **Average Accuracy** | | | | | | | | | | | | | | | | **98.53%** |

## 6. CONCLUSION AND FUTURE WORK

The proposed method efficiently identified the disease with 98.53% accuracy and scored 94% accuracy in gaining integrity of a plant. This method proved to be more effective than existing methods in detecting the disease of a plant and thereby estimating area of damaged crop.The model is able to correctly classify at least 28 out of the 30 images given as input.

The application designed takes the input from the mobile camera module. This includes some manual work which can be replaceable by a drone camera in future. The drone camera captures entire crop and it is helpful to detect the damage of the crop by external elements like fire, cyclone etc. Also, it is easier to capture every plant in the field with the drone camera. In future, the application will be extended to integrate drone camera module with the proposed system.

## REFERENCES

[1] Melike Sardogan, Adem Tuncer and Yunus Ozen "Plant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm", 3rd International Conference on Computer Science and Engineering (UBMK) sept,2018.

[2] Boya Wang, Jianqing Xu, Junbao Li, Cong Hu and Jeng-Shyang Pan, "Scene text recognition algorithm based on faster RCNN", IEEE, 22 February 2018.