MAT JOURNALS

Journal of Data Mining and Management
e-ISSN: 2456-9437
Volume 4Issue 2

Recognition and Classification of Leaf Disease in Potato Plants

Sanjay K Hoysala*[*1], Santhosh K S[1], Srihari D R[1], Suhith Shekar Chandra[1], Ajay Prakash B V[2]
[1]UG Students, [2]Associate Professor
[1,2]Department of Computer Science and Engineering, SJB Institute of Technology, Bangalore, Karnataka, India
***Email:** sanjaykumar.hoysala@gmail.com*
***DOI:** http://doi.org/10.5281/zenodo.3353806*

Abstract

Farming is one of the most important lifelines of the country. A nation's growth majorly depends on how advanced and effective their agricultural practices are in improving the crop yield. When a crop is grown many at times, farmers are unable to identify the health and wellbeing of the plant; they only recognize the problems when it becomes too late hence losing out on that year's expected yield. In this study, we have introduced a recognition and classification technique which is able to detect any ailments that the plant is suffering from at an early stage itself thus enabling the farmers to do the needful at a recoverable stage itself. To make the system as user-friendly as possible, we have provided a feature where the farmers are able to assess the health of the plant by providing a picture of the potato plants' leaf.

***Keywords:** Agriculture, convolutional neural network, image classification, plant disease*

INTRODUCTION

Digital image processing consists of using computer algorithms to process digital images. The processing of images leads to many useful outcomes that can be further used in many useful applications; some of these are as follows: image sharpening and restoration, medical field, remote sensing, transmission and encoding, machine/robot vision, colour processing, pattern recognition, video processing, microscopic imaging and many other applications.

Potato is one of the leading contributors of carbohydrates in one's nutrition. Hence, it plays a pivotal role in supplying energy to society. The growth and production of potato is the essence of the time. While growing this crop one must look after the common diseases this crop suffers from. There are majorly two diseases that cause significant yield loss they are Phytophthora infestans (late blight) and Alternaria solani (early blight). Detection of these diseases at an early stage can save the farmers from a lot of loss and sufferings. In the past decade, the main method of detection of these has been through naked eye observation. This method has been proved as ineffective and tedious. Hence, the introduction of image processing and detection tools has enabled continuous monitoring and prevention of these diseases possible. The main symptoms of these diseases are majorly visible on the plants leaves thus enabling imaging techniques combined with machine learning to offer a prediction of the disease the plant is suffering from. So, the end result of this study is to demonstrate an effective and guaranteed disease detection system for the potato plant.

LITERATURE SURVEY AND INFERENCES

Paper [1] presents the different ways in which images are processed, thus building a foundation before they are classified using different image classification algorithms. It includes gathering and

 Page 25-29 © MAT Journals 2019. All Rights Reserved

organizing data, building a predictive model which enables applications to recognize patterns in images. [2] LeNet describes the different advantages of using convolution neural networks. They have also explained the operation of local or sparse connectivity and shared weights used by the convolutional neural networks. Paper [3] explains the architecture of the convolution neural network. It explains how an image is converted into a convoluted image which is then passed through various layers such as convolution layer, pooling layer, fully connected layer and finally tells how the output is predicted based on the trained images. [4] Understanding the Convolutional Neural Network. It explains the conversion of an image to an array.[5] Thresholding and network backpropagation are used in this method. [6] PCA and ANN are used in this method. Dimensions of feature data are reduced using PCA that is nothing, but it increases the speed of NN by reducing the no. of neurons in the input layer. [7] Sometimes the object in the spot image cannot be fixed by thresholding. To avoid such situations, the authors of this paper have proposed the use of LTSRG-algorithm for segmentation of the image at hand. [8] Spectrum based algorithms have been used to diagnose cucumber leaf diseases. [9] Spectrometer, a device that measures light intensity in the electromagnetic spectrum, is used to classify the rubber tree disease.

PROPOSED METHODOLOGY

For an image to be understood by the computer, it must either be a raster or a vector image. Raster images are a sequence of pixels whose numerical values for colours are discrete, whereas, in vector images, the polygons are colour annotated. The first step in any image recognition process is to extract the important data and discard the rest. Before the model can start classifying, we must train the model by showing thousands of images, which must be classified in the later stages. To train the model with such huge datasets, we need to use machine learning algorithms. Next, we need neural networks to define a function that provides a required output based on the trained data. There are various image classification algorithms such as K-nearest neighbours (KNN), support vector machines (SVM), artificial neural networks (ANN), convolutional neural networks (CNN) and many others.

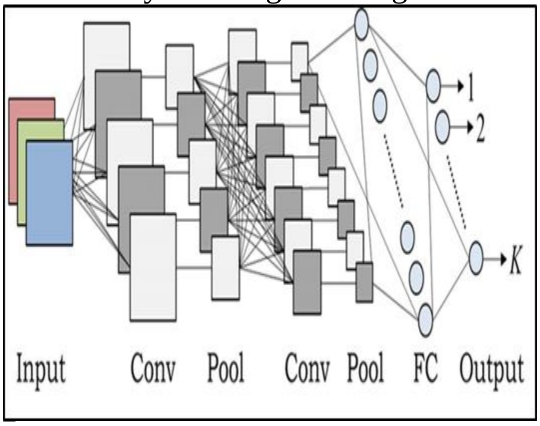
Neural networks have many similarities to convolutional neural networks(CNN). Convolutional neural networks consist of numerous neurons that possess learnable weights and biases. The neurons in the CNN receive some inputs then perform a dot product on them. The input is received and transformed by the neural network. Transformation in the neural network is done through a series of hidden layers which thus enables the model to make accurate predictions. Each hidden layer is made up of a set of neurons wherein these neurons are fully connected to all the neurons present in the previous layer. In a single layer function, all the neurons are completely independent and do not share any connections. The “output layer” is also known as the last fully connected layer, it also indicates the class scores in the classification settings. Convolutional neural networks can constrain the architecture because the input consists of an image. Unlike a regular neural network, the layers of neurons are arranged in 3 dimensions: width, height and depth.

A convolutional neural network is a deep learning algorithm and this type of artificial neural network is used in image recognition. It consists of some hidden layers along with input and output layers. The four hidden layers in CNN are convolution layer, pooling layer, ReLU layer and fully connected layer. There are many advantages of using CNN over other machine learning and deep learning

Page 25-29 © MAT Journals 2019. All Rights Reserved

MAT JOURNALS
Journal of Data Mining and Management
e-ISSN: 2456-9437
Volume 4Issue 2

algorithms. CNN incorporates local connectivity and weight sharing.



$$(f * g)(t) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(\tau)\, g(t-\tau)\, d\tau$$

Figure 1: *Architecture of CNN.*

The above function describes the convolutional neural function. A convolution is basically a combined integration of two functions; it shows how one function modifies the other. Fig. 1 shows the architecture of CNN. Once the image passes through the first convolution layer, the output of which is an activation map, the convolution layer then takes this activation map and extracts the relevant features from it. Following which, these features are passed on to the further stages, which are then stored in a convolution layer as a feature matrix.

Each of these filters gives a various feature which supports prediction making process. The number of parameters is further reduced by using pooling layers. There are different pooling techniques like max-pooling, sum pooling and average pooling. The technique to be used is decided based on the application at hand. Before the actual prediction is made, we use several convolutions and pooling layers. As the neural network gets denser, the prediction becomes better. The final layer of CNN is a fully connected layer, where the output from different layers is flattened and sent to transform the output into the desired number of classes.

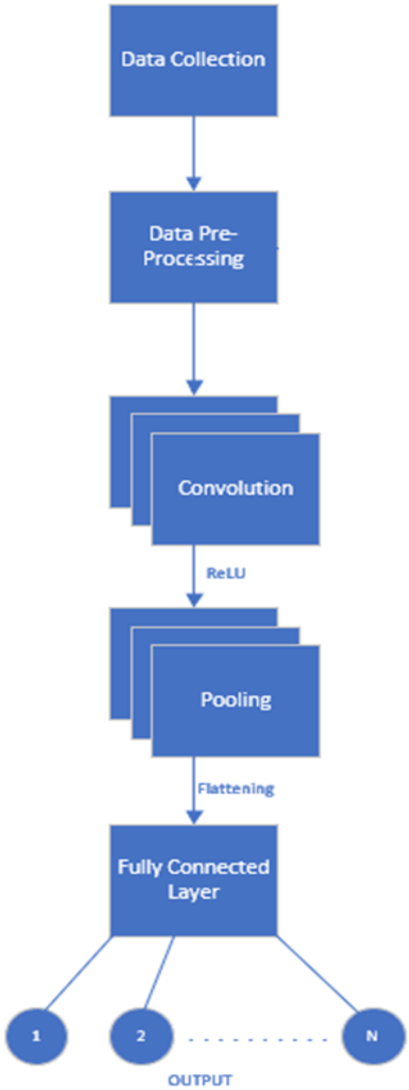


Figure 2: *Block diagram of the proposed method.*

IMPLEMENTATION AND RESULT

The software tool that we have used to build this model is Google Colab and the programming language used is Python 3.6. The dataset we have used consists of around 1000 images of healthy, early blight and late blight diseases of a potato leaf. We have trained the model by giving 1000 images of each as input, out of which 80% of it was used for training the model and the rest was used for testing the model. The model used converts an image to an array

Page 25-29 © MAT Journals 2019. All Rights Reserved

Journal of Data Mining and Management
e-ISSN: 2456-9437
Volume 4Issue 2

which in turn consists of height, weight and width of the image.

Early Blight disease *Late blight disease*

Figure 3: *Results.*

An image input matrix is of dimension h*w*d which is then multiplied with a filter (fh*fw*d), the output matrix dimension is (h-fh+1) *(w-fw+1) *1. The filter is also called as feature mapping which can be used for edge detection, blurring and sharpening of the image.
Sometimes, the filter does not fit the input image, at such stage we have two options:

- Pad the picture with zeros so that it will fit.
- Drop the part of the image where the filter did not fit. This is called as valid padding which keeps only valid part of the image.

0.0625	0.125	0.0625
0.125	0.25	0.125
0.0625	0.125	0.0625

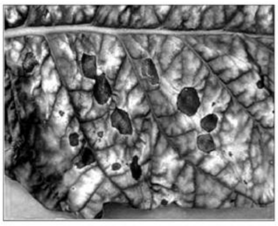


Figure 4: *Matrix value of the blurred image.*

0	-1	0
-1	5	-1
0	-1	0

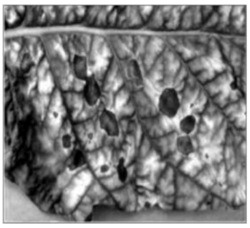


Figure 5: *Matrix value of the sharper image.*

The above matrix and the image (Fig.4 and 5) represent the different form of images, which is multiplied with the respective filter in order to get blur and sharpen images as shown. After applying the filter, ReLU is applied to the matrix, which is an activation function, it is very important because it does not saturate the image. The gradient will always be high if the neuron activates. The updates are effective, as long as it is not a dead neuron. Then many convolution layers are added to get the accurate prediction. The output from those layers is then flattened and fed into the fully connected layer (FC layer). Then the model applies all the activation functions to obtain the output which is the predicted image.

The model we built was 99.07% accurate. This accuracy was obtained by training it with around 700-800 images. The graph against training and valid accuracy is depicted in Figure 6.

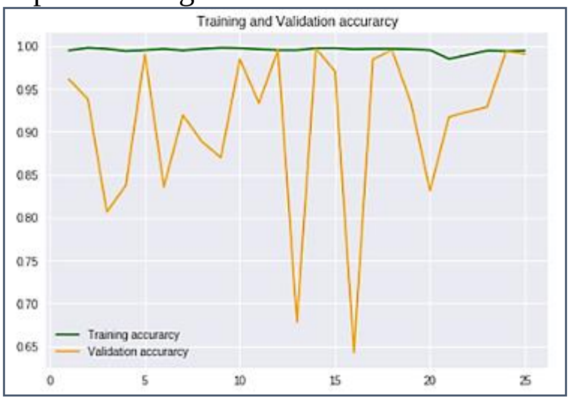


Figure 6: *Graph against training vs testing accuracy.*

```
[INFO] Calculating model accuracy
431/431 [==============================] - 2s 4ms/step
Test Accuracy: 99.0719266391409
```

Figure 7: *Accuracy of the model.*

CONCLUSION

The goal of this study is to identify the disease the potato plant is suffering from, by providing an image of the potato plant's leaf as input to the model. The model we have proposed is efficient, fast, accurate and has a high prediction rate. Farming is not an easy task by any means, it involves taking many risks. We hope to reduce

Page 25-29 © MAT Journals 2019. All Rights Reserved

these risks by enabling farmers to detect the health and longevity of the potato plant at an early stage itself. Thus enabling them to save the potato plant through precautions. A nation truly grows when all parts of its economy develop at an equal rate. It is our job as responsible citizens to make this happen.

ACKNOWLEDGEMENT
Research work has always been challenging and it comes with its own set of ups and downs but breaking these barriers and coming out as victorious in the end is what makes it worthwhile. Regarding this, we extend our sincere and heartfelt obligation towards all the personages who have helped us in this odyssey. We are forever indebted to Prof. Dr Ajay Prakash B V, Associate Professor, SJBIT, for his continuous guidance and encouragement to accomplish this work. Last, but not the least, we would like to thank the Department of Computer Science and Engineering present in our institute and all our friends who have stood by us and encouraged us in the completion of our work.

REFERENCES
1. What is the working of image recognition and how it is used? https://www.marutitech.com/working-image-recognition/.
2. Convolutional Neural Networks (LeNet) http://deeplearning.net/tutorial/lenet.html.
3. Architecture of Convolutional Network https://www.analyticsvidhya.com/blog/2017/06/architecture-of-convolutional-neural-networks-simplified-demystified/.
4. Understanding of CNN. https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148.
5. Sanjeev S Sannakki, Vijay S Rajpurohit, VB Nargund, Pallavi Kulkarni (2013), "Diagnosis and classification of grape leaf diseases using neural networks",*4th ICCCNT.*
6. Haiguang Wang, Guanlin Li, Zhanhong Ma, Xiaolong (2012), "Image recognition of plant diseases based on principal component analysis and neural networks",*8th International Conference on Natural Computation (ICNC 2012).*
7. Jun Pang, Zhong-ying Bai, Jun-chen Lai, Shao-kun Li (2011), "Automatic segmentation of crop leaf spot disease images by integrating local threshold and seeded region growing", *IEEE.*
8. Hong-ning Li, Jie Feng, Wei-ping Yang, Xiang-sheng Wu, Ze-dong Li, Wei Liu (2011), "Spectrum-based method for quantitatively detecting diseases on cucumber leaf", *IEEE.*
9. Hashim H, Haron MA, Osman FN, Al Junid, SAM (2010), "Classification of rubber tree leaf disease using spectrometer", *IEEE.*

Cite this article as: *Sanjay K Hoysala, Santhosh K S, Srihari D R, Suhith Shekar Chandra, & Ajay Prakash B V. (2019). Recognition and Classification of Leaf Disease in Potato Plants. Journal of Data Mining and Management, 4(2), 25–29. http://doi.org/10.5281/zenodo.3353806*

Page 25-29 © MAT Journals 2019. All Rights Reserved