IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.133133

Potato Disease Detection using Deep Learning

R.Arthi[1], H.Mohammed Haarish[2]

Assistant Professor, PG& Research Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore 641006 Tamil Nadu India[1]

PG Student, PG& Research Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore 641006 Tamil Nadu India[2]

Abstract: In Bangladesh potato is one of the major crops. Potato cultivation has been very popular in Bangladesh for the last few decades. But potato production is being hampered due to some diseases which are increasing the cost of farmers in potato production. However, some potato diseases are hampering potato production that is increasing the cost of farmers. Our main goal is to diagnose potato disease using leaf pictures that we are going to do through advanced machine learning technology. This paper offers a picture that is processing and machine learning based automated systems potato leaf diseases will be identified and classified. Image processing is the best solution for detecting and analysing these diseases. In this analysis, picture division is done more than 2034 pictures of unhealthy potato and potato's leaf, which is taken from openly accessible plant town information base and a few pre prepared models are utilized for acknowledgment and characterization of sick and sound leaves. Among them, the program predicts with an accuracy of 99.23% in testing with25% test data and 75% train data. Our output has shown that machine learning exceeds all existing tasks in potato disease detection.

Keywords: Machine learning,VGG16, CNN, potato leaf

I. INTRODUCTION

Potato is one of the major harvests in our country. The production of potatoes in Bangladesh is very lower than compared to the other developed countries in the world. The production of potatoes is hampered by many kinds of Pests and diseases. So we can’t export potatoes to our expectations in the other countries. Among them early blight, leaf roll virus, scab, Hollow heart etc are the most terrible disease of potatoes at previous present and times in Bangladesh. In Bangladesh, the major area's farmers faced many hampers on this disease every year. The farmers and businessmen of our country are facing many problems with those diseases particularly in the case of export to other countries as Russia, Indonesia, Malaysia, Sri Lanka, Thailand, Hong Kong, Turkey, Vietnam, Maldives etc. Due to COVID-19, the price of potatoes increases day by day. Even a decade ago, production was below half a million tons. Now it is moving towards billions of tons. This success has brought Bangladesh to the top ten potato producing countries in Qatar.

The recognition was given by the Food and Agriculture Organization of the United Nations. According to a report of this organization, Bangladesh is in the eighth place with a production of 722 lakh10 thousand tons. Not only is this a wonderful achievement in production, potato is now one of the most profitable crops in the country. It has also become a means of earning foreign currency. Potatoes worth 33 million were exported last year. But before that Bangladesh had to import potatoes from 20countries of the world. We believe that, if we detect the disease of potato properly and provide the proper treatment, the production growth will increase to our expectation. We have taken the help of image processing to diagnose potato disease and potato leaf disease. Here we have used these parameters to diagnose the disease which can be identified by looking at the characteristics of the disease and the type of disease and we have tried to give more antidotes here. Accuracy is very good with training data in the project and accuracy with sample data is above 99%. So our project will be able to accurately diagnose potato disease and leaf disease. All through the proposed model, the CNN algorithm is utilized to recognize various kinds of potato infections, having7 classes of potato sicknesses and accomplished 99% accuracy rate.

II. RELATED WORKS

The proportion of light at wavelengths across the electromagnetic spectrum that is either absorbed, transmitted or reflected from a plant leaf is dependent on leaf structure, physiology and biochemistry. Since these elements are influenced by pests, pathogens and their associated induced diseases, the detection, differentiation and diagnosis of plant diseases is theoretically possible by non-destructive analysis of the light reflected from plant leaves.

DOI: 10.17148/IJARCCE.2024.133133

In this study the utility of analysis of light over the visible and near-infrared (400–1000 nm) portion of the spectrum to detect and distinguish between several economically important potato diseases, using either Partial Least Squares and Back Propagation Neural Network spectral calibration models was explored. Models could detect and distinguish between diseases with obvious foliar symptoms (blackleg and late blight), even pre-symptomatically, correctly classifying spectra from greenhouse experiments with an accuracy of 84.6%. When these diseases were analysed separately, models could distinguish between spectra from healthy and pre-symptomatic leaves, plus three classes of late blight lesion advancement with 92% accuracy. For blackleg, models distinguished between spectra from healthy, pre-symptomatic foliage and plants expressing blackleg symptoms with a 74.6% classification accuracy. However, models trained on spectra from whole-plant readings from field trials did not have this level of accuracy, with an r2 between target and model values of 0.66 for late blight, 0.31 for blackleg symptoms and 0.41 for healthy foliage. Regardless of greenhouse or field environment, models failed to detect or distinguish between diseases with subtle foliar impacts (black dot, powdery scab and Rhizoctonia diseases). While deployment of hand-held spectrometers for disease detection on a broad-acre scale is impractical, these findings could underpin methods to analyse hyperspectral imaging data with sub-plant resolution for incorporation into precision agriculture and Integrated Pest Management programmes for potato blackleg and late blight management.

Potato is one of the prominent food crops all over the world. In Bangladesh, potato cultivation has been getting remarkable popularity over the last decades. Many diseases affect the proper growth of potato plants. Noticeable diseases are seen in the leaf region of this plant. Two common and popular leaf diseases of the potato plants are Early Blight (EB) and Late Blight (LB). However, if these diseases were identified at an early stage it would be very helpful for better production of this crop. To solve this problem by detecting and analyzing these diseases image processing is the best option. This paper proposes an image processing and machine learning-based automatic system that will identify and classify potato leaf diseases. In this paper, image segmentation is done over 450 images of healthy and diseased potato leaf, which is taken from publicly available plant village database and seven classifier algorithms are used for recognition and classification of diseased and healthy leaves. Among them, The Random Forest classifier gives an accuracy of 97%. In this manner, our proposed approach leads to a path of automatic plant leaf disease detection.

With the enhancement in agricultural technology and the use of artificial intelligence in diagnosing plant diseases, it becomes important to make pertinent research to sustainable agricultural development. Various diseases like early blight and late blight immensely influence the quality and quantity of the potatoes and manual interpretation of these leaf diseases is quite time-taking and cumbersome. As it requires tremendously a good level of expertise, efficient and automated detection of these diseases in the budding phase can assist in ameliorating the potato crop production. Previously, various models have been proposed to detect several plant diseases. In this paper, a model is presented that uses pre-trained models like VGG19 for fine-tuning(transfer learning) to extract the relevant features from the dataset. Then, with the help of multiple classifiers results were perceived among which logistic regression outperformed others by a substantial margin of classification accuracy obtaining 97.8% over the test dataset.

III. PROPOSED APPROACH

VGG-16 | CNN model

The input to the network is image of dimensions (224, 224, 3). The first two layers have 64 channels of 3*3 filter size and same padding. Then after a max pool layer of stride (2, 2), two layers which have convolution layers of 256 filter size and filter size (3, 3). This followed by a max pooling layer of stride (2, 2) which is same as previous layer. Then there are 2 convolution layers of filter size (3, 3) and 256 filter. After that there are 2 sets of 3 convolution layer and a max pool layer. Each have 512 filters of (3, 3) size with same padding.This image is then passed to the stack of two convolution layers. In these convolution and max pooling layers, the filters we use is of the size 3*3 instead of 11*11 in AlexNet and 7*7 in ZF-Net. In some of the layers, it also uses 1*1 pixel which is used to manipulate the number of input channels. There is a padding of 1-pixel (same padding) done after each convolution layer to prevent the spatial feature of the image.

After the stack of convolution and max-pooling layer, we got a (7, 7, 512) feature map. We flatten this output to make it a (1, 25088) feature vector.After this there are 3 fully connected layer, the first layer takes input from the last feature vector and outputs a (1, 4096) vector, second layer also outputs a vector of size (1, 4096) but the third layer output a 1000 channels for 1000 classes of ILSVRC challenge, then after the output of 3rd fully connected layer is passed to softmax layer in order to normalize the classification vector. After the output of classification vector top-5 categories for evaluation. All the hidden layers use ReLU as its activation function. ReLU is more computationally efficient because it results in faster learning and it also decreases the likelihood of vanishing gradient problem.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.133133

The table below listed different VGG architecture. We can see that there are 2 versions of VGG-16 (C and D). There is not much difference between them except for one that except for some convolution layer there is (3, 3) filter size convolution is used instead of (1, 1). These two contains 134 million and 138 million parametersrespectively.

They are very successful in image recognition. The key part to understand, which distinguishes CNN from traditional neural networks, is the convolution operation. Having an image at the input, CNN scans it many times to look for certain features. This scanning (convolution) can be set with 2 main parameters: stride and padding type. As we see on below picture, process of the first convolution gives us a set of new frames, shown here in the second column (layer). Each frame contains an information about one feature and its presence in scanned image. Resulting frame will have larger values in places where a feature is strongly visible and lower values where there are no or little such features. Afterwards, the process is repeated for each of obtained frames for a chosen number of times. In this project I chose a classic VGG-16 model which contains only two convolution layers.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure-1 Configuration of CNN

The latter layer we are convolving, the more high-level features are being searched. It works similarly to human perception. To give an example, below is a very descriptive picture with features which are searched on different CNN layers. As you can see, the application of this model is face recognition. You may ask how the model knows which features to seek. If you construct the CNN from the beginning, searched features are random. Then, during training process, weights between neurons are being adjusted and slowly CNN starts to find such features which enable to meet predefined goal, i.e. to recognize successfully images from the training set. Between described layers there are also pooling (sub-sampling) operations which reduce dimensions of resulted frames. Furthermore, after each convolution we apply a non-linear function (called ReLU) to the resulted frame to introduce non-linearity to the model.

Eventually, there are also fully connected layers at the end of the network. The last set of frames obtained from convolution operations is flattened to get a one-dimensional vector of neurons. From this point we put a standard, fully-connected neural network. At the very end, for classification problems, there is a softmax layer. It transforms results of the model to probabilities of a correct guess of each class Apply the model and plot the graphs for accuracy and loss:

ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024
DOI: 10.17148/IJARCCE.2024.133133

We will compile the model and apply it using fit function. The batch size will be 2. Then we will plot the graphs for accuracy and loss. We got average validation accuracy of 1.00% and average training accuracy of 1.00%.

IV. RESULTS AND DISCUSSION

Dataset

We collected this information by taking pictures directly from the potato fields of the village. We were able to collect data on about seven types of diseases of potatoes and potato leaves. We collect total 2034 potato and potato leaves images. We basically split the data between potato leaves and potato disease. While collecting data, we noticed that diseases is more contagious in potato leaves. The class that divides our data is mentioned below:

- potato leaf roll virus
- Hollow heart of potato
- scab of potato
- Soft rot of potato
- Sutalipokarog
- Virus jonitoro

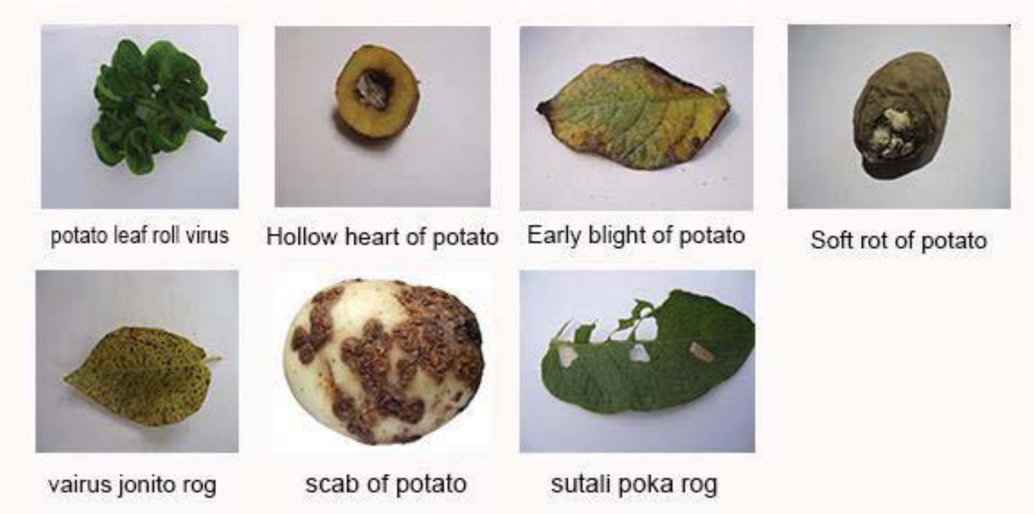

Figure-2 Sample Dataset

class	Disease name	Images (number)	Train data	Test data
01	early blight of potato	428	329	100
02	potato leaf roll virus	394	274	75
03	Hollow heart of potato	221	218	80
04	scab of potato	154	150	98
05	Soft rot of potato	238	200	66
06	Sutali poka rog	305	295	95
07	Virus jonito rog	294	275	120
total		2034		

Figure-3 Classification Analysis

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.133133

V. CONCLUSION AND FUTURE WORK

This paper mainly focuses on disease detection of potatoes from any surface by using machine learning (CNN). We found that VGG16 Architecture is the best way to perform this type of detection object. However, this model gains 100% of validation accuracy. We have a large amount of data set and to get the best accuracy, we have tried our best. We think this type of project will play a vital role in our agriculture sector. Most of the farmers of the village in Bangladesh are not literate and they can't know about the disease properly. They can't know the method of detecting disease. That's why the insect is destroying the potato and our farmers get to suffer from it. We think that, this work can change the situation of the potato grower in Bangladesh..

REFERENCES

[1]. Hillary Sanders, Joshua Saxe, Richard Harang, Cody Wild A Deep Learning Approach to Detecting Malicious Web Content in a Fast, Format-Independent Way. pp. 8-14 in Proceedings of the 2018 IEEE Symposium on Security and Privacy Workshops (SPW 2018), San Francisco, CA, USA, August 2.

[2]. Jie Wu, Longfei Wu, Xiaojiang Du Phishing Attacks on Mobile Computing Platforms: Effective Defense Schemes 6678-6691 in IEEE Transactions on Vehicular Technology, vol. 65, no. 8, 2016.

[3]. Praneesh, M., and R. Annamalai Saravanan. "Deep Stack Neural Networks Based Learning Model for Fault Detection and Classification in Sensor Data." Deep Learning and Edge Computing Solutions for High Performance Computing (2021): 101-110.

[4]. A. Kumar, R. S. Umurzoqovich, N. D. Duong, P. Kanani, A. Kuppusamy, M. Praneesh, and M. N. Hieu, ‘‘An intrusion identification and prevention for cloud computing: From the perspective of deep learning,’’ Optik, vol. 270, Nov. 2022, Art. no. 170044.

[5]. Muthukumar S, Dr.Krishnan .N, Pasupathi.P, Deepa. S, “Analysis of Image Inpainting Techniques with Exemplar, Poisson, Successive Elimination and 8 Pixel Neighborhood Methods”, International Journal of Computer Applications (0975 – 8887), Volume 9, No.11, 2010

[6]. Damian Bienkowskia, Matt J. Aitkenheadb, Alison K.Leesc, Christopher Gallaghera, Roy Neilsona “Detection anddifferentiation between potato diseases using calibration modelstrained with non-imaging spectrometry data” Computers andElectronics in Agriculture 167 (2019) 105056.

[7]. Monzurul Islam, AnhDinh, Khan Wahid ,PankajBhowmik“Detection of Potato Diseases Using Image Segmentation andMulticlass Support Vector Machine” , 2017 IEEE 30th CanadianConference on Electrical and Computer Engineering (CCECE).

[8]. Md. AsifIqbal and KamrulHasanTalukder,”Detection of PotatoDisease Using Image Segmentation and Machine Learning”

[9]. DivyanshTiwari, MritunjayAshish,NitishGangwar,AbhishekSharma, SuhanshuPatel ,Dr.SuyashBhardwaj.“Potato LeafDiseases Detection Using Deep Learning”. Proceedings of theInternational Conference on Intelligent Computing and ControlSystems (ICICCS 2020) IEEE Xplore Part Number: CFP20K74-ART; ISBN: 978-1-7281-4876-2.

[10]. FarabeeIslam,MdNazmulHoq,ProfessorDr.ChowdhuryMofizurRahman.” Application of Transfer Learning to DetectPotato Disease from Leaf Image”.

[11]. Athanikar, Girish and Ms.PritiBadar. “Potato Leaf DiseasesDetection and Classification System Mr.” (2016).

[12]. MalvikaRanjan, Manasi Rajiv Weginwar, NehaJoshi, A.B.Ingole,“Detection and Classification of Leaf Disease UsingArtificial Neural Network,” International Journal of TechnicalResearch and Applications, 2015, pp. 331–333.

[13]. F. T. Pinki, N. Khatun and S. M. M. Islam, “Content based paddyleaf disease recognition and remedy prediction using supportvector machine,” 2017 20th International Conference ofComputer and Information Technology (ICCIT), Dhaka, 2017,pp. 1–5.

[14]. Yao Q, Guan Z, Zhou Y, Tang J, Hu Y, Yang B, “Application ofsupport vector machine for detecting rice diseases using shapeand color texture features,” 2009 International Conference onEngineering Computation, IEEE, Hong Kong, 2009, pp. 79–83.

[15]. C. U. Kumari, S. Jeevan Prasad and G. Mounika, "Leaf DiseaseDetection: Feature Extraction with K-means clustering andClassification with ANN," 2019 3rd International Conference onComputing Methodologies and Communication (ICCMC),Erode, India, 2019, pp. 1095-1098.

[16]. Wiwart M, Fordonski G, Zuk-Golaszewska K, Suchowilska E“Early diagnostics of macronutrient deficiencies in three legumespecies by color image analysis,” Compute Electron Agric 65, 20