Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Plant Disease Prediction using Machine Learning

***Vivek Kumar Soni and **Dr Prerna Soni**
*Assistant Professor,
Department of Computer Science EngineeringSSIPMT, Raipur (C.G)
**Assistant Professor Department of Biotechnology
SPCAS PG College Nawapara, Rajim, District- Raipur (Chhattisgarh)
*Corresponding author Email-eklavya.vivek@gmail.com

Article Info
Page Number: 9739 – 9747
Publication Issue:
Vol 71 No. 4 (2022)

Article History
Article Received*: 15 September 2022*
Revised: *25 October 2022*
Accepted: *14 November 2022*
Publication: *21 December 2022*

Abstract: — The major population of Indians is relies on agriculture. The production of crop and income and yield are directly affected by plant diseases. Prediction of plant diseases is very important step to arrest the loss so researchers are included for such a mechanism which can overcome this problem with the modern approaches of science like machine learning techniques. The present paper includes collection of leaf randomly with healthy and diseased leaf and identifying those leaves using prepared data sets. It involves CNN Convolution Neural Network which will be able to predict whether a plant is suffering from a disease or not. The proposed paper uses different layers and other hyperparameters for building, training and testing this classification model.

Keywords: Convolution Neural Network, plant disease, healthy and diseased leaf, testing,training

I. Introduction

Plant diseases are important components to determine the production of plants for farmers. Observation and identification of diseases in plant is very important as to arrest the deprivation in the production, manually it is very difficult, and hence machine learning technique is now a day's major breakthrough for this observation. Current years, machine learning has emerges as major developmentin the area of artificial intelligence. Conventional method needs excessive quantityof labor, expertise within the plant diseases. Inspite of, some experimental techniques in machine learning have been implemented for increasing the rate of identification with accurate of results. Deep learning is a experimental technology which implies machine vision techniques for obtaining digital images to find diseased or disease free plants [1].

In this present work, we have described the technique for the detection of plant diseases with the help of their leaves pictures. Machine learning is a sub part of artificial intelligence which works automatically or gives instructions to do a particular task. The main aim of machine learning is to understand the training data and fit that training data into models that should be useful to the people. So it can assist in good decisions making and predicting the correct output using the large amount of training data. The colour of leaves, amount of damage to leaves, area of the leaf, texture parameters are used for classification.

Fig 1: Diseased plant with infected leaf

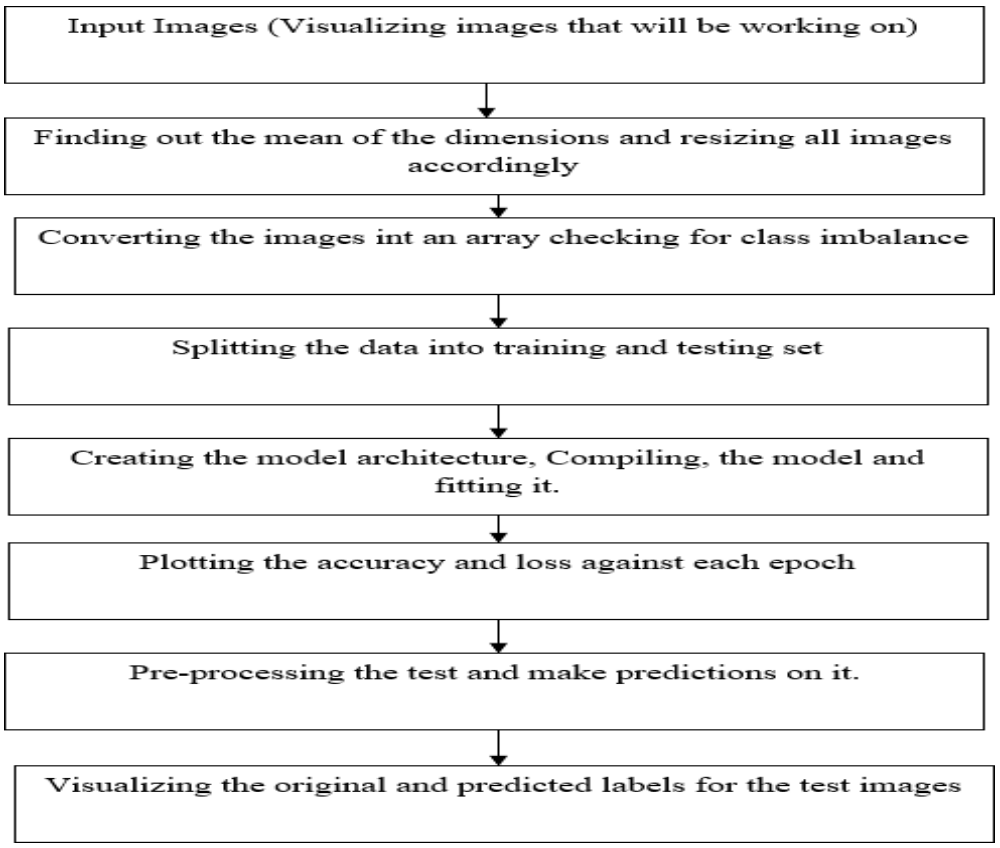
II. METHODOLOGY

In the different parts of plants like root, stem, leaves, the susceptible part for several diseases are leaves. This is because of some environmental conditions e.g. temperature, moisture, stress, humidity etc. If this condition is unfavourable then different infections like viral, fungal and bacterial diseases manifested by plant in the form of diseases. The recognition of these symptoms due to same pattern is very difficult hence to control the loss from plant diseases [7] early prediction is required. In image processing some steps are followed like Pre-processing, Feature Extraction and Classification.

Machine learning means feeding machine with necessary data to make it learn. There are two types of machine learning supervised learning or predictive learning and second one is unsupervised learning or descriptive learning. In current years successful application of machine learning in various field like traffic detection [2], medical Image Recognition [3], Scenario text detection [4], expression recognition [5], face Recognition [6], etc make it more versatile.

In this paper we will create a convolutional neural network that is a CNN model [8] and it will be able to predict a disease of leaves with the image that we supply. Wewill use different layers and other hyper parameters for building, training and testing this multiclass classification model.

A machine learning method for classifying plant disease is projected in this paper.
Used methods are following:



1. Input image: - These are the images that will be working on i.e. leaf of the plant.

2. Finding out the mean of the dimension and resizing all images accordingly: - Reducing the size of an image means changing its dimensions by removing its pixels. Scaling up an image increases the number of its pixels but lowers quality. Either way, the image's aspect ratio changes, which results in distortion.As we are making a CNNmodel. we will import all the required layers, activations, optimizers, etc.Now we willobserve some of the images that are there in our dataset. We will plot 16 images.

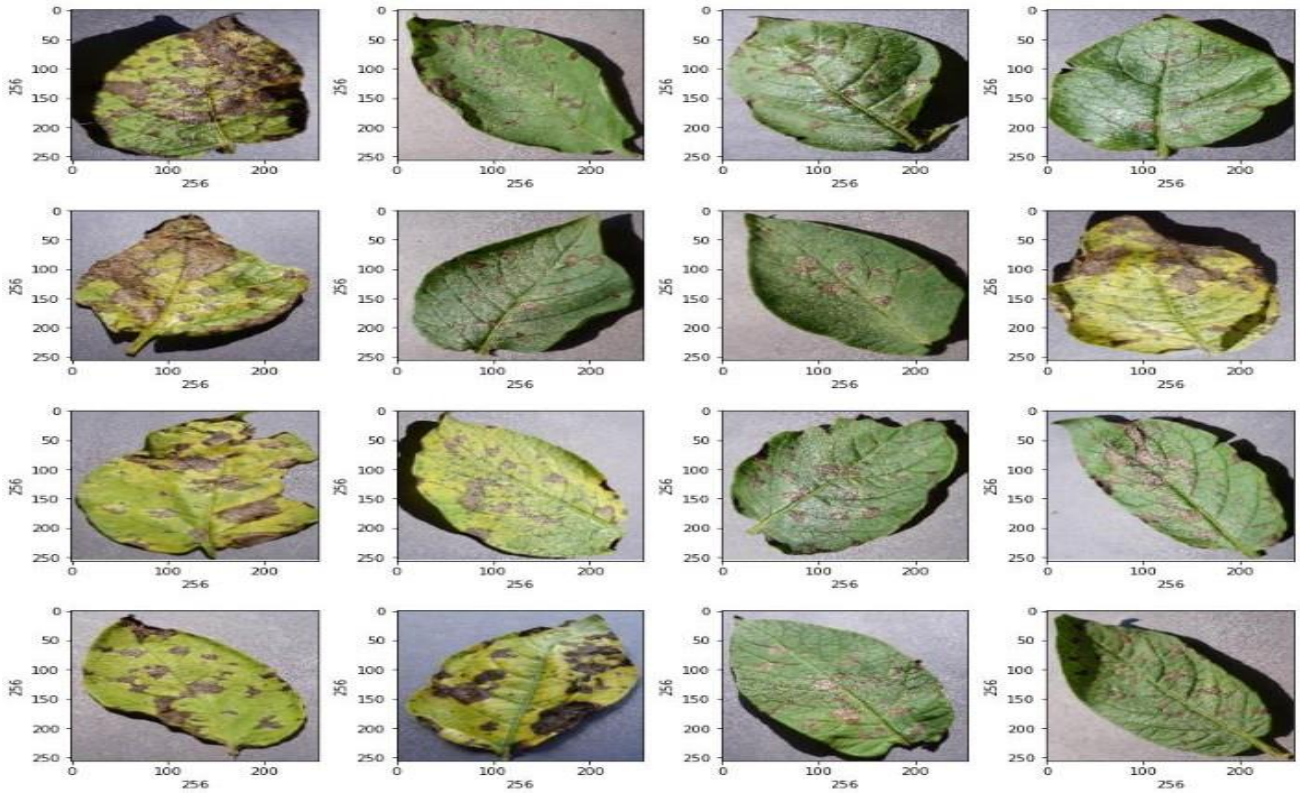


Figure 2: Different types of Leaf images in Dataset

3. Converting the images into Array: Converting an image to an array is an important task to train a machine learning model based on the features of an image [9].So,after visualizing the images moving forward and create a function which will convert the images into an array in order to normalize our dataset.
4. Class imbalance: -Class imbalance[10] is a problem that occurs in machine learning classification problems. There is a bias or skewness towards the majority class presentin the target. We will also observe the number of images under different classes to seeif the dataset is balanced or not. We will observe the shape of the image. Checking thetotal number of the images which is the length of the labels list.
5. Splitting the Data: - Data splitting [12] is an important aspect of data science, particularly for creating models based on data. This technique helps ensure the creation of data models and processes that use data models such as machine learning are accurate. we will split the dataset into testing and training data. We have taken test size as 0.2 so as data will be divided into 80% training and 20% testing data. Now we will normalize the dataset of our images. As pixel values ranges from 0 to 255 so we will divide each image pixel with 255 to normalize the dataset.
6. Creating a Network architecture for the model compiling it and fitting it.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 256, 256, 32)      896
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 85, 85, 32)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 85, 85, 16)        4624
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 42, 42, 16)        0
_________________________________________________________________
flatten (Flatten)            (None, 28224)             0
_________________________________________________________________
dense (Dense)                (None, 8)                 225800
_________________________________________________________________
dense_1 (Dense)              (None, 3)                 27
=================================================================
Total params: 231,347
Trainable params: 231,347
Non-trainable params: 0
```

While compiling the model we need to set the type of loss which will be Binary Crossentropy [11] for our model along with this we also need to set the optimizer and the metrics respectively.

7. Plotting the accuracy and Loss against each Epoch: - We will split the dataset into validation and training data Fitting the model with the data and finding out the accuracy at each epoch to see how our model is learning. Now we will train our model on 10 epochs and a batch size of 128.During each epoch we can see in figure how the model is performing by viewing the training and validation accuracy.

Experimental Results

First,getsomeimagesfromthisdatasetandextractfeaturesfromtheseimages.Different classes for different diseases where Class0representscornrust, Class1representsPotato-Early_blight,class2representsBacterialBlight,andclass3representsTomato-Bacterial_spot.Theresultafterapplyingthepretreatmentprocedureisshowninfigure5.

1. Plotting the model accuracy for training and validation.

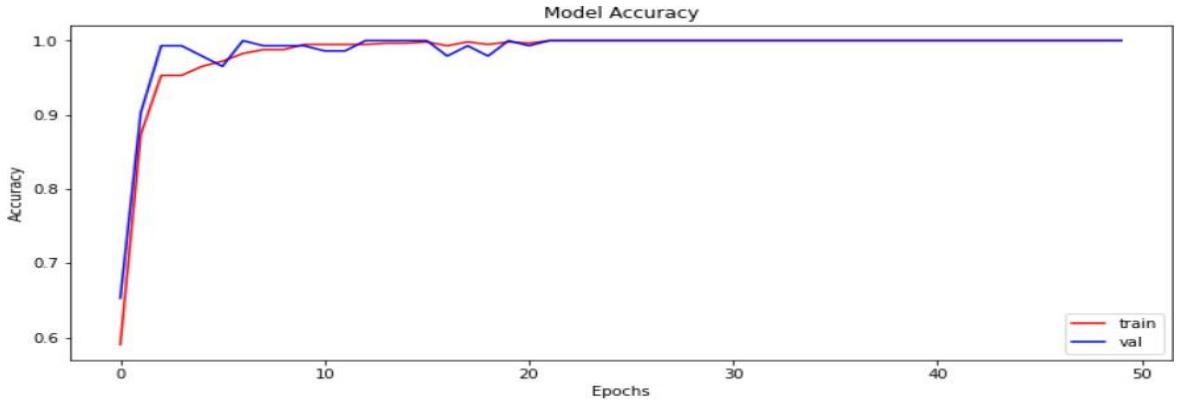


Figure 3: Model Accuracy

2. Evaluating the model to know the accuracy of the model.

```
[INFO] Calculating model accuracy
6/6 [==============================] - 0s 30ms/step - loss: 0.0072 - accuracy: 0.9944
Test Accuracy: 99.44444298744202
```

3. We will use our model to predict predicting the testing dataset label. Printing out theoriginal and the predicted label.

```
Originally :  Potato-Early_blight
Predicted :  Potato-Early_blight
```

Figure 5: Potato Leaf

Figure 6: Tomato Leaf with Bacterial _spot

Discussion

We started with loading the dataset into google colab using google drive and visualizing the images. Normalizing is an important step when working with any type of dataset. After that we created a CNN Model which is further used for predicting the plant diseases using the image supplied to model. This model is highly beneficial as it can be used by different agricultural firms and farmers to increase their yield and stop wastage of crops due to disease.

References

1. Lee SH, Chan CS, Mayo SJ, Remagnino P. How deep learning extracts and learns leaf features for plant classification. Pattern Recogn. 2017;71:1–13.
2. Yang D, Li S, Peng Z, Wang P, Wang J, Yang H. MF-CNN: traffic flow prediction using convolutional neural network and multi-features fusion. IEICE Trans Inf Syst. 2019;102(8):1526–36.

3. Sundararajan SK, Sankaragomathi B, Priya DS. Deep belief cnn feature representation based content based image retrieval for medical images. J Med Syst. 2019;43(6):1–9.
4. Melnyk P, You Z, Li K. A high-performance CNN method for offline handwritten chinese character recognition and visualization. Soft Comput. 2019;24:7977–87.
5. Li J, Mi Y, Li G, Ju Z. CNN-based facial expression recognition from annotated rgb-dimages for human–robot interaction. Int J Humanoid Robot. 2019;16(04):504–5.
6. Venu, S., Kotti, J., Pankajam, A., Dhabliya, D., Rao, G. N., Bansal, R., . . . Sammy, F. (2022). Secure big data processing in multihoming networks with AI-enabled IoT. Wireless Communications and Mobile Computing, 2022 doi:10.1155/2022/3893875
7. Kumar S, Singh SK. Occluded thermal face recognition using bag of CNN(BoCNN).IEEE Signal Process Lett. 2020;27:975–9.
8. S. Yun, W. Xianfeng, Z. Shanwen, and Z. Chuanlei, "Plant based crop disease recognition with leaf image features and meteorological data," International Journal of Agricultural and Biological Engineering, vol. 8, no. 4, p. 60, 2015
9. Xinyu Lei, Hongguang Pan∗ ,Xiangdong Huang "A dilated CNN model for image classification" DOI 10.1109/ACCESS.2019.2927169, IEEE Access College of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an 710049, China
10. Mark S. Nixon, Alberto S. Aguado" Feature Extraction & Image Processing for Computer Vision" Publisher: Academic Press, Year: 2012 ISBN: 0123965497,978012396549
11. Shaza M. Abd Elrahman1 and Ajith Abraham "A Review of Class Imbalance Problem" Journal of Network and Innovative Computing ISSN 2160-2174, Volume 1 (2013) pp. 332-340
12. Dr.A.Usha Ruby , PrasannavenkatesanTheerthagiri , Dr.I.Jeena Jacob , Dr.Y.Vamsidhar "Binary cross entropy with deep learning technique for Image classification"A.Usha Ruby et al., International Journal of Advanced Trends in Computer Science and Engineering, 9(4), July – August 2020, 5393 – 5397
13. Delwendeelianebirba, "A Comparative study of data splitting algorithms for machine learning model selection" degree project in computer science and engineering, second cycle, 30 credits stockholm, sweden 2020
14. Timande, S., & Dhabliya, D. (2019). Designing multi-cloud server for scalable and secure sharing over web. International Journal of Psychosocial Rehabilitation, 23(5), 835-841. doi:10.37200/IJPR/V23I5/PR190698

Biography

Vivek Kumar Soni is an Assistant Professor in Department of Computer Science and Engineeringat SSIPMT, Raipur, Chhattisgarh, India. He has 11 years teaching experience including other academic institution such as Century Cement College and B.M College of Durg, Chhattisgarh, India. He received his M.Tech(Computer Science Engineering) from CSVTU, Bhilai, Chhattisgarh, India in 2014. Also received his MCA (Computer Application)

from BIT Durg (C.G), India. He has published 7 research articles in leading journals. He has presented 11 research papers in different national and international conferences and seminars. He has also organized one national conference in area of Recent Innovations in Emerging Computing Technologies. His research interests include Machine Learning, Computer Vision, Image Processing and Signal Processing

Dr.Prerna Soni is an Assistant Professor in Department of Biotechnology at Seth Phoolchand Agrawal Smriti Post-graduate College, Nawapara-Rajim, Raipur, Chhattisgarh, India. She have 17 years teaching experience including other academic institution such as Govt. Science College and Vipra College of Bilaspur, Chhattisgarh, India. She received her Ph.D. (Life Science) from Guru Ghasidas Central University, Bilaspur, Chhattisgarh, India in 2013. She received her M. Phil. (Biotechnology) from Vinayka Mission University Tamilnadu, India. She has published three chapters in the studies on medicinal plants. She has published 14 research articles in leading journals. She holds two awards such as “Kaumi Ekta Award” and “ShikshakRatnaSamta” Award. She is the member of the Board of the Studies and Indian Botanical Society Association. She has presented 21 research papers in different national and international conferences and seminars, She has also organized one national conference in area of impact of emerging science and technology in society. Her research interests include plant tissue culture, secondary metabolites, antimicrobial activity of medicinal plants, immunological techniques and instrumentation.