Identification of Fungi infected Leaf Diseases using Deep Learning Techniques

Sukanya S. Gaikwad[a], Shivanand S. Rumma[b], Mallikarjun Hangarge[c]

[a] Research Scholar, Dept. of Computer Science, Gulbarga University, Kalaburagi, Karnataka, India.
[b]Chairman, Dept. of Computer Science, Gulbarga University, Kalaburagi, Karnataka, India.
[c]Associate Professor, Dept. of Computer Science, Karnatak Arts, Science, and Commerce College, Bidar, Karnataka, India.
[a]gsukanya116@gmail.com, [b]shivanand_sr@yahoo.co.in, [c]mhangarge@yahoo.co.in

Article History: Received: 10 November 2020; Revised 12 January 2021 Accepted: 27 January 2021; Published online: 5 April 2021

Abstract: This paper presents CNN based model to identify fungi-affected leaf diseases of Psidium guajava. Identifying these leaf diseases at an early stage will help the farmers take significant precautions and prevent the disease from spread to other parts of the plant and the neighbouring plants. The dataset is collected from real environment of the farms. It has four categories of infected leaves (3971 in number) and they are Pseudocercospora leaf spot, Rust, Insect eaten leaf, and another category is of the healthy leaf of Guava (Psidium guajava). Then, applied CNN, AlexNet and SqueezeNet architectures to identify 3971 fungi infected leaves of Guava. SqueezeNet architecture shown 75.9% recognition accuracy as compared to the other two architectures.

Keywords: Psidium guajava, Leaf Diseases, Pseudocercospora leaf spot, Rust, Deep Learning, CNN, AlexNet, SqueezeNet.

1. Introduction

The population in India mainly depends on agriculture. India is a country where we have a varied type of temperature, weather and land. So growing of the fruits and vegetables is also varied at different parts of the country. Fruits are grown in home gardens, fruit farms, or on the roadways where the seed might have soon by wind or animals. Among many fruits grown, Guava is the fourth most popular fruit in India after Mango, Banana and Citrus. The scientific name of Guava is Psidium guajava, which is of Latin origin. The origin of Guava is assumed to be in Central America or Mexico. Slowly, it has been spread across the world by man and nature. In India, it was introduced by the Portuguese in the early 17th century. Guava is the most ancient fruits of India because of its hardy and prolific bearing nature, making it to be grown in most Indian states. Maharashtra is the first highest Guava producer, after which Bihar and Uttar Pradesh come second and third, respectively.

Guava is a good source of vitamin C & A, calcium, manganese and phosphorus. It is used in the production of jam, jellies, nectar etc. Leaves of Guava are used in curing diarrhea, tanning and dyeing. The fruit is consumed either raw or ripped with little of salt and pepper on it. This plant serves as a host to fungi, algae and bacteria found in the atmosphere, which affects the leaf's, stem, barks, fruit and twig of the plant. Most of these plants are lost due to fungi.

The traditional method of identifying these leaf diseases is tiresome, time-consuming and expensive as it requires the continuous monitoring of the plant by the experts and they have to screen the plant through the naked eye, which might sometimes be a wrong assumption of the disease. At current situation the recognition and classification of guava leaf diseases primarily relay on the plant pathologists and work experience of the farmers. In recent years, emergence of machine learning and deep learning techniques have aced in getting better results. Deep learning techniques have become a research attraction in the field of computer vision, pattern and image recognition such SVM and CNN. Hence, we aim to propose and develop a model using deep learning techniques to identify these fungi affected diseases at an early stage. Identifying these leaf diseases at an early stage will help the farmers take significant precautions and prevent the disease from spread to other parts of the plant and the neighboring plants.

We focus on three different categories of leaf diseases, i.e, Pseudocercospora leaf spot, Rust, insect eaten leaf and another category is of a healthy leaf of Guava. The dataset is collected from real time enviroment[12].

As there is too much of work done on publically available dataset. Hence we collect the dataset from a nearby farm of Kalaburagi district of Karnataka, India.

The rest of the paper is arranged as Section 2 describes the literature survey, Section 3 preparation of dataset, Section 4 narrates the proposed model, Section 5 describes the experimental results, comparative analysis in Section 6 and lastly conclusions and future work are summarized in Section 7**.**

2. Literature Survey

Md. Rasel Howlader *et al.,* [1] proposed a deep convolution neural network (D-CNN) based model to identify and classify guava leaf diseases on their dataset, namely BU Guava Leaf (BUGL2018). They classified three diseased leaves, i.e., algal leaf spot, whitefly and Rust, and another category are for the healthy leaf of Guava. Using their proposed model they achieved recognition accuracy of 98.74% on the test set.

Ibtesam Their *et al.,* [2] developed a Knowledge-Based System for identifying guava problems. Their main aim was to help farmers, specialists, and students diagnose the seven different types of guava diseases. The model was developed using CLIPS and Delphi XE10.2 languages. It does not require intensive training to be used; it is user friendly and is easy to use.

S. Arivazhagan and S.Vineth Ligi [3] introduced a deep learning-based method that identifies the leaf diseases of mango plant's. They have identified five different leaf diseases. The diseases studied are Anthracnose, Alternaria leaf spots, Leaf Gall, Leaf Webber, Leaf burn of Mango. The dataset consists of 1200 images of diseased and healthy mango leaves. The proposed CNN model achieves an accuracy of 96.67% for identifying leaf diseases.

Geetharamani G and Arun Pandian J [4] proposed a nine-layer deep CNN for identifying plant leaf diseases. They used an open dataset of 39 different classes of plant leaves. After extensive training of the model, they achieved an accuracy of 96.46%. They claim that the proposed model is better than the traditional machine learning approaches.

Xiaoxiao SUN *et al.,* [5] developed a CNN based model to recognize the seven different tea diseases of leaves. To obtain a high accuracy rate, they tuned the model with different hyperparameters like training with different epochs, dropout, and learning rate and obtained a good accuracy of 93.75%. They compared the developed model with SVM and BP and obtained 89.36% and 87.69% accuracy.

Rathan Kumar Veeraballi *et al.,* [6] proposed a CNN model for image classification of papaya leaf diseases, including papaya mosaic and papaya leaf curl. They used the pre-trained ResNet-50 model to train and classify the leaf images. An accuracy of 85% is obtained for the trained model, which proves that it is appropriate for a real-time environment.

Xiaofei Chao *et al.,* [7] developed an apple tree leaf diseases (ALTD) model based on DCNN. Six different classes of apple leaves were used for classification. They combined the Xception and DenseNet model features, which gives the highest recognition accuracy of 98.82%, which is higher than Inception-v3, Mobile Net, VGG-16, DenseNet-201, Xception, VGG-INCEP.

3. Preparation of Dataset

A proper dataset is required to analyze the performance of the deep learning techniques. We introduce our dataset collected from a nearby farm of Kalaburagi district of Karnataka, India. The dataset consists of three different categories of leaf diseases i.e, Pseudocercospora leaf spot, Rust, insect eaten leaf, and another category is of the healthy leaf of Guava. Two different smartphones and one digital camera is used to capture the images. A total of 3971 images are used for classifying the leaves. These images are classified into four different classes, three are for diseased class and the other is healthy leaf, which are described as below.

Pseudocercospora leaf spot: Pseudocercospora psidii fungi cause this disease. It appears as a small, roughly shaped, or vaguely circular dark brown border on the leaves upper surface. The fungus may grow, and in the center of lesions, gray tufts of mycelium may be visible. During wet conditions, leaves infection occurs, and it can also be spread by the splashing of water. We collected 727 images of pseudocercospora leaf spot images from 50 different species of guava plant, among which 510 are given for training and 217 are used for testing.

Rust: It is caused by the fungi Puccinia psidii. It affects both the leaf and fruit of the Guava plant. The leaves may appear distorted with orange to red pustules on the leaves upper and lower surface. Circular lesions on the leaves with yellow halos and dark borders are also visible. We collected 864 images of rust leaf images from 50 different species of guava, among which 606 are given for training and 258 are used for testing.

Insect eaten: It is caused by insect-eating of the leaves. It may appear anywhere around the guava leaf. Mostly it appears around the edges of leaves. We collected 324 images of rust leaf images from 50 different species of guava, among which 225 are given for training and 99 are used for testing.

Below are sample images of diseased and healthy leaves of the guava plant shown in Figureure 1 to 4.

Figure 1. Healthy Leaf

Figure 2. Pseudocercospora leaf spot

Figure 3. Rust

Figure 4. Insect eaten

Below Table 1 shows the overview of our dataset.

Table 1. Overview of Dataset

Name	Total Images	Training Images	Testing Images
Healthy	2056	1439	617
Insect eaten	324	225	99
Leaf spot	727	510	217
Rust	864	606	258
Total	3971	2780	1191

3. Proposed Model

In traditional machine learning algorithms we extract the features manually and then send it for classification, where as in deep learning algorithms the layers of CNN extracts the features from the data given by making use of the kernels. Which eliminates the need for manually extracting the features from the data. This has lead us to work on deep learning techniques. Hence, we propose a deep learning-based CNN model. It has sequential layers, and each layer uses the previous layer as input to the model. CNN requires minimal pre-processing and is very good at analyzing images. With its multilayered structure, CNN is good at separating the desired features.

The CNN model's basic building blocks are the Convolution layer, Pooling layer, Activation Function, and the Fully Connected layer. Below Figure 5 shows the architecture of the CNN model.

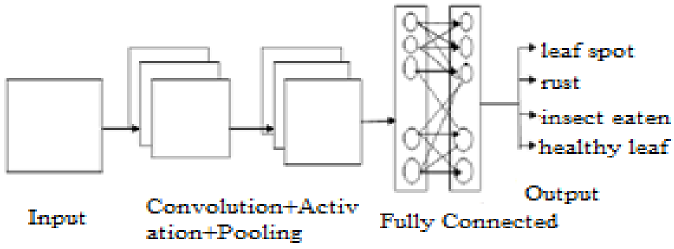

Figure 5. CNN architecture.

Depending upon the CNN architecture, there may be multiple convolutions, activation and pooling layers.

The image is used as an input layer, where pre processing of raw data takes place. Any size of image can be given for classification, the model will resize according to the models standard size which is 256x256 pixels. Prior to the beginning of convolution operation, the input layer contains images as pixel values. If the image is gray scale then input shape will be PxP and if it is color image as in our dataset then the input shape will be PxPxN, where N represents the dimensions i.e, R, G and B colors. These three different input matrices, i.e. R, G and B channels for every image in the dataset, are given input to the first convolution layer. The convolution operation is given as below in equation (1)

$$P_i = B_i + \sum_n Z_{in} * X_n \qquad (1)$$

Where P_i is the feature map, * is convolution operation, X_n is the input channel,Z_{in} is the kernels and B_i represents the bais value. Each input image matrix is convoluted, and batch normalization is applied. Batch normalization is done to standardize the raw inputs while feeding to the next layer. After every batch normalization layer, Relu activation is used. This activation function alleviates the problem of over fitting. The max-pooling operation is then applied to the output matrix whose equation (2) is given as below

$$h_j = \max_{i\ aj} f_i \qquad (2)$$

which is connected to the Fully Connected (FC) layer. The FC layer's residue is connected to the softmax function. The softmax operation is given as in equation (3)

$$O_x = S(\gamma)_x = \frac{e^{\gamma x}}{\sum_{n=1}^{N} e\gamma n} \qquad (3)$$

Where O_x is the output vector, S is softmax function taking N dimensional vector γ and outputs real values between 0 nad 1. The adam optimizer is used to optimize the algorithm. The learning rate is at 0.0001. The execution environment used is CPU. Image augmentation is used to reduce the over fitting of the model. It is a technique where the given image is resized, rotated, tilted, and zoomed to capture the essential features in all angles. This in turn will erase the problem of over fitting of the model. Lastly, we train the model with 25 epochs to get accuracy. Below Figureure 6 shows the proposed CNN model.

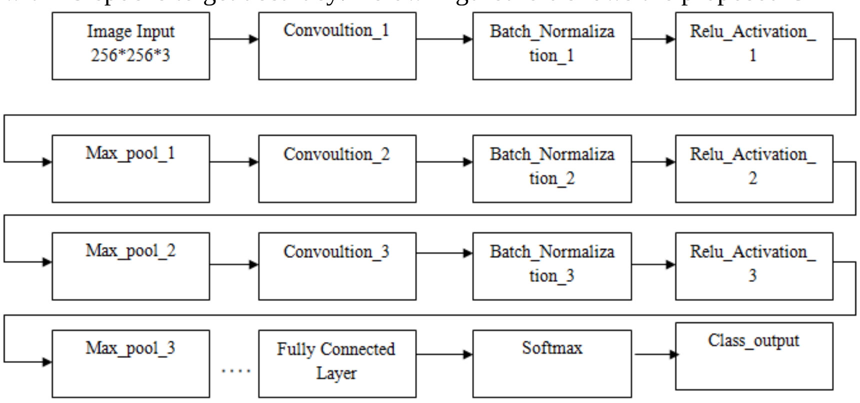

Figure 6. Proposed CNN model

5. Experimental Results

The dataset consists of 3971 images. The experiment conducted divides the dataset into a 70:30 split where 70% of the dataset we use for training and the remaining dataset, i.e. 30%, is used for testing purposes. The difficult task in identifying and classifying guava leaves is that the leaves with different diseases are very similar. Therefore, this similarity can lead the leaves to be mapped into the wrong classes. The CNN model trained through several iterations to classify 3971 images; we achieved an accuracy of 66.5%. The accuracy of the model is calculated as,

$$\text{Accuracy (\%)} = \frac{\text{Total number of images correctly classified}}{\text{Total number of images used for testing}} * 100$$

Class wise recognition accuracy of CNN model is given in the Table 2 below.

Table 2. Class-wise recognition accuracy of CNN

Sl. no.	Class Name	Accuracy in %
1	Leaf spot	44.6%
2	Rust	62.8%
3	Insect eaten	10.6%
4	Healthy	71.2%

The confusion matrix obtained for proposed CNN model is shown in the Table 3.

Table 3. Confusion matrix for proposed CNN model.

	Leaf spot	Rust	Insect eaten	Healthy
Leaf spot	**44**	66	2	108
Rust	16	**202**	8	41
Insect eaten	10	15	**10**	72
Healthy	28	43	0	**546**
			Average in %	**66.5%**

The below graph Figure 7 shows the accuracy obtained for training and validation data using the proposed CNN model. And the graph Figure 8 shows the loss for training and validation data using the proposed model.

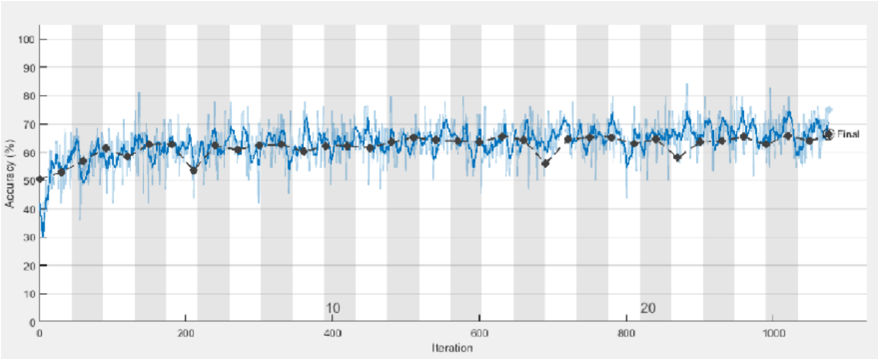

Figure 7. Accuracy for training(blue line) & validation data(black line)

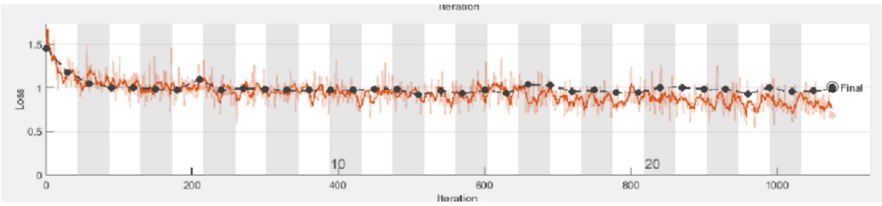

Figure 8. Loss for training(red line) and validation data(black line)

6. Comparative Analysis

There are many different pre-trained models. We have used AlexNet [8] and SqueezeNet [9] for compartive analysis with our proposed CNN architecture. Both networks are trained on GPU's and on 1000s of image categories. The experiment divided the dataset into 70:30 split where 70% of data is used for training and 30% data is used for testing. The Adam optimizer is used, and the learning rate is kept at 0.0001. And the model is trained for 25 epochs. These parameters are kept the same for all three models used. So using our dataset and after fine-tuning the hyperparameters, we have obtained the below results shown in Table 4.

Table 4. Comparative Analysis of models

Sl. No.	Architecture	Accuracy	Elapsed Time
1	Proposed model	66.5%	238 Mins 41 Sec
2	AlexNet	71%	152 Mins 40 Sec
3	SqueezeNet	75.9%	199 Mins 44 Sec

Class wise recognition accuracy of AlexNet and SqueezeNet model is given in the Table 5 and Table 6 respectively.

Table 5. Class-wise recognition accuracy of AlexNet

Sl. no.	Class Name	Accuracy in %
1	Leaf spot	44.5%
2	Rust	78.3%
3	Insect eaten	53.5%
4	Healthy	79.4%

Table 6. Class-wise recognition accuracy of SqueezeNet

Sl. no.	Class Name	Accuracy in %
1	Leaf spot	58%
2	Rust	75%
3	Insect eaten	47.2%

4	Healthy	86.2%

The confusion matrix obtained for AlexNet and SqueezeNet is shown in the Table 7 and Table 8 respectively.

Table 7. Confusion matrix for AlexNet.

	Leaf spot	Rust	Insect eaten	Healthy
Leaf spot	**110**	35	7	66
Rust	43	**184**	4	28
Insect eaten	24	7	**23**	43
Healthy	70	43	9	**529**
			Average in %	**71%**

Table 8. Confusion matrix for SqueezeNet.

	Leaf spot	Rust	Insect eaten	Healthy
Leaf spot	**105**	66	29	45
Rust	32	**201**	11	15
Insect eaten	12	7	**50**	28
Healthy	32	21	16	**548**
			Average in %	**75.9%**

It is clearly observed from the Table 4- Table 8 that SqueezeNet performs better than our proposed model and the AlexNet network. The time required to train the model is little more as compared to the AlexNet but recognition accuracy is good as compared to the other two models. Hence we claim that SqueezeNet performs better on the collected dataset of the Guava plant.

The proposed model differes from the pre-trained model in getting better results because of the pre-trained architectures are trained on 1000's of categories and after several iterations and hyper parameter tuning they have got good results. Hence the pre-trained model gave good results as comapred to the proposed model.

7. Conclusions and Future Work

The paper aims to identify and classify three different leaf diseases and one category of the guava plant's healthy leaf. A total of 3971 images are real time environment collected from 3 different sources. The proposed deep learning-based CNN model gives an accuracy of 66.5%. For comparative analysis, we have used two different pre-trained models, i.e., AlexNet and SqueezeNet. It is discovered from the comparative analysis that SqueezeNet performs better as compared to the other two models.

In the future, we try to hyper-tune the proposed model's parameters to achieve high classification accuracy on the available dataset.

8. Acknowledgement

This work is supported and funded by Karnataka Science and Technology Promotion Society (KSTePS), DST, GOVT. OF KARNATAKA.

References (APA)

[1]. M. R. Howlader, U. Habiba, R. H. Faisal and M. M. Rahman.:Automatic Recognition of Guava Leaf Diseases using Deep
[2]. Convolution Neural Network,2019 International Conference on Electrical, Computer and Communication Engineering
[3]. (ECCE), Cox'sBazar, Bangladesh, pp. 1-5, April – 2019.
[4]. Ibtesam Dheir, Samy S. Abu-Naser.:Knowledge Based System for Diagnosing Guava Problems, International Journal of
[5]. Academic Information Systems Research (IJAISR), Vol. 3 Issue 3, pp. 9-15, March – 2019.
[6]. S. Arivazhagan, S.Vineth Ligi.:Mango Leaf Diseases Identification Using Convolutional Neural Network, International
[7]. Journal of Pure and Applied Mathematics, Volume 120 No. 6 August – 2018.
[8]. Geetharamani G and Arun Pandian J.:Identification of plant leaf diseases using a nine-layer deep convolutional neural
[9]. network, Computers and Electrical Engineering, pp. 323-328, April – 2019.
[10]. Xiaoxiao SUN, Shaomin MU, Yongyu XU, Zhihao CAO, Tingting SU.:Image Recognition of Tea Leaf Diseases Based on
[11]. Convolutional Neural Network, International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), pp.
[12]. 304-309, Jinan, China, January – 2020.
[13]. Rathan Kumar Veeraballi, Muni Sankar Nagugari, Chandra Sekhara Rao Annavarapu, and Eswar Varma Gownipuram.:Deep
[14]. Learning Based Approach for Classification and Detection of Papaya Leaf Diseases, Intelligent Systems Design and
[15]. Applications, pp.291-302, vol 940, Springer, April – 2019.
[16]. Xiaofei Chao, Guoying Sun, Hongke Zhao, Min Li and Dongjian He.:Identification of Apple Tree Leaf Diseases
[17]. Based on Deep Learning Models, Symmetry, June – 2020.
[18]. Krizhevsky, A., Sutskever, I. & Hinton, G.: ImageNet classification with deep convolutional neural networks,
[19]. Advances in Neural Information Processing Systems 25, pp. 1090–1098, 2012.
[20]. Forrest N. Iandola, Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, Kurt Keutzer.:Squeezenet:
[21]. Alexnet-level Accuracy with 50x fewer parameters and <0.5MB Model Size, Computer Vision and Pattern Recognition,
[22]. February – 2016.
[23]. Smys, S., Joy Iong Zong Chen, and Subarna Shakya.:Survey on Neural Network Architectures with Deep Learning. Journal
[24]. of Soft Computing Paradigm (JSCP), pp.186-194, 2020.
[25]. Vijayakumar, T.: Posed Inverse Problem Rectification Using Novel Deep Convolutional Neural Network. Journal of
[26]. Innovative Image Processing (JIIP), pp.121-127, 2020.
[27]. Muhammad Hammad Saleem, Johan Potgieter and Khalid Mahmood Arif.: Plant Disease Detection and Classification by
[28]. Deep Learning. Plants. 8, pp.468, 2019.