Paddy Leaf diseases identification on Infrared Images based on Convolutional Neural Networks

Petchiammal A, Briskline Kiruba S, D. Murugan

Manonmaniam Sundaranar University, Tirunelveli, India
{petchiammalmtech17,kiruba.briskline}@gmail.com, dmurugan@msuniv.ac.in

Abstract. Agriculture is the mainstay of human society because it is an essential need for every organism. Paddy cultivation is very significant so far as humans are concerned, largely in the Asian continent, and it is one of the staple foods. However, plant diseases in agriculture lead to depletion in productivity. Plant diseases are generally caused by pests, insects, and pathogens that decrease productivity to a large scale if not controlled within a particular time. Eventually, one cannot see an increase in paddy yield. Accurate and timely identification of plant diseases can help farmers mitigate losses due to pests and diseases. Recently, deep learning techniques have been used to identify paddy diseases and overcome these problems. This paper implements a convolutional neural network (CNN) based on a model and tests a public dataset consisting of 636 infrared image samples with five paddy disease classes and one healthy class. The proposed model proficiently identified and classified paddy diseases of five different types and achieved an accuracy of 88.28%

Keywords: Convolutional Neural Network, Classification, Pre Processing, Infrared thermography, Rice diseases.

1   Introduction

Agriculture, paddy production plays a vital role in the country's economy. It is one of the most powerful tools to end extreme destitution, raise shared prosperity, and feed a predicted 9.7 billion people in 2050 [1]. Agriculture is a crucial sector for the incomes of various fields in the world. But, crop cultivation affected frequent climate changes, other weather factors, and crop diseases reduced agricultural growth and food security challenges. Another crucial role in crop diseases and pesticides destroy the crop and harm biodiversity. Paddy is one of the leading food crops in India because it is the preeminent food that most people in India need. The paddy farmer faced biotic stress because paddy diseases severely affect paddy production by 70% yield reduction.

The traditional methods are also in practice to detect and prevent paddy diseases. Most of the fungal diseases in paddy can easily be identified by monitoring leaves regularly. So, lots of agricultural experts and experienced experts are in need. However, crop protection experts are found to be very limited. Therefore, the prevention and treatment of paddy diseases, especially the early diagnosis and monitoring of paddy diseases, is excellent for improving paddy cultivation based on the automated system. Various computer vision and deep learning techniques have been used to identify many plant diseases [18]. In [19], this method identified the powdery mildew disease in the early stage based on machine learning. Similarly, SVM with a radial basis function [7], artificial neural networks [20], and Gaussian mixture models [8] have been developed to identify crop diseases. In many recent studies, Deep CNN and SVM were used to detect rice disease in its early stages to provide the appropriate treatment in time based on RGB images only [9][14][11].

This paper developed a Convolutional Neural Network (CNN) based classification model for identifying rice disease. The raw infrared images used to test our model were collected from a public data repository. However, pre-processing techniques were used before experimenting with our model, including data cleaning, class validation, and temperature data extraction. And also apply suitable data augmentation techniques to create many image samples, which are required for any deep learning-based methods. The experimental results showed an accuracy of 88.28%. The proposed model is coded in Python Programming language as it is considered the best choice for the convolutional neural network using open source libraries such as Keras. All the experiments ran on the Google Colab cloud platform with GPUs.

The rest of the paper is organized as follows. In Section 2, we summarize the related works. Section 3 presents the proposed rice disease identification methodology followed by the dataset details in Section 4. Section 5 presents the experimental setup and results, followed by a conclusion in Section 6.

2 Related Work

Several approaches have already been used to precisely identify or classify diseases of the various plants through image classification. Almost all models have used different image processing methods, machine, deep learning techniques, etc. In [17] proposed technique is used to identify paddy diseases from the infected area by using feature extraction based on data mining and images processing. In [15][16], proposed technique detects the paddy disease by feature extraction based on an SVM classifier for image processing.

In [6], the proposed model for region identification of rice diseases identified the diseases and identified the infected area using an image processing technique. Classification based on machine learning models like Naive Bayes and Support Vector machines was used and classified into different categories. The Canny edge detection algorithm was used to track the edge and get the histogram value from images to identify the diseases. The model identified the crop diseases based on periodically monitoring the crop cultivated area [5]. In [12], this model has used Scale Invariant Feature Transform, which extracted the features from paddy images to identify the paddy disease and classify it. It has provided high accuracy with the help of SVM and K Nearest Neighbors classifiers.

Analyzed different image classification algorithms and presented detailed studies based on automated identification system techniques [13]. The proposed model, which follows a predictive approach, is used to improve paddy production and helps farmers get a high yield of paddy crops based on machine learning techniques like clustering and decision trees [3]. To automatically identify the disease in paddy leaves by using the extraction method of hybridized gray scale co-occurrence matrix, DWT and SIFT developed an approach based on image processing techniques. This paper has classified the disease and normal leaf based on the various classifiers like SVM, Naive Bayesian, Back propagation, and KNN [10]. This model used DCNN architectures to identify rice plant diseases and achieved high levels of classification accuracy [4][2]. Although, they did not use an Infrared thermal image of the paddy for classification. The study considers all these issues; we have presented the five most affected paddy diseases and healthy leaves image classification by CNN classifier, which gave higher accuracy.

3 Methodology

3.1 Convolutional Neural Network architecture

In this paper, we have developed a CNN-based model to perform automated paddy disease classification based on infrared thermal images. This work will be a promising one that fulfills the requirement of farmers in India with better machine language algorithms. The block diagram of the proposed CNN is shown in Fig.1. After extracting the temperature, it is given as input to the classification approaches and then it gives the accurate type of the diseases.

We have used the Keras sequential API, where we added one layer at a time, starting from the input. The size and stride of the convolution determined the size of the output image. For instance, the 3×3 kernel of convolution with stride 1, converts a 3×3 image to a 2×2 pixel. The convolution above is applied to the thermal image. Normal images have height, width, and depth. But, in thermal images it has four dimensions. In other words, they are four-dimensional, where the first two dimensions are width and height, the third is depth, and the fourth is the channel of temperature. In-depth refers to the different color components of the images: red, green, and blue. Hence, most images will begin with a depth of 3, but the fourth value will not be binary in thermal images. It will usually be between 0 and 255. Typically, it is assumed that convolution is applied to the depth and temperature of the image. So, 3×3 convolutions are applied to an image of depth 3 and temperature 1 we have taken an input consisting of $3 \times 3 \times 3 \times 1$ pixels.
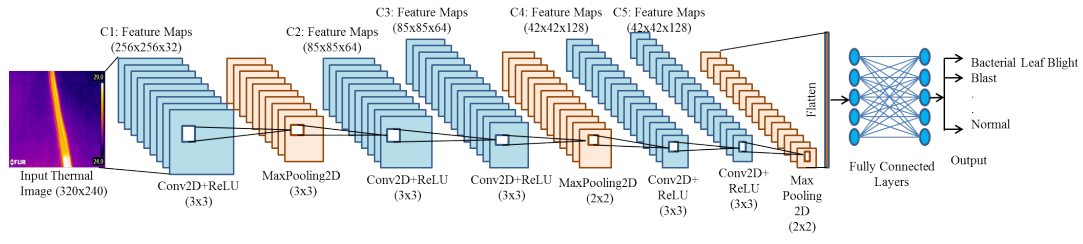
Fig. 1. CNN architecture for Rice disease classification.

The convolution output applied to a single input data will still be a single pixel. The result of the convolution applied to an entire three-dimensional image will be a two-dimensional image; this means that the image removed its third and fourth channels. Increased the number of conv2D layers to build a deeper model.

– Increase the number of filters to learn more features
– Added dropout for regularization
– Added more dense layers

Composition of the CNN model: Our CNN model consists of five layers: the first part is the "input thermal image." Its first convolutional layer is filtered with 32 kernels of size 3×3. Then, a 3×3 max-pooling layer is added after the first convolutional layer. The next convolutional layer contains 64 convolution kernels of size 3×3, and we have used a batch normalization layer to automatically standardize the inputs in a model and improve the accuracy and stability of neural networks. The next convolutional layer contains 64 convolution kernels of size 3×3, followed by a 2×2 max-pooling layer. This layer is used between Conv2D layers in CNN architecture for reducing the number of parameters and computation time in the network, controlling over fitting by progressively reducing the spatial size of the network. After, added two convolutional layers containing 128 convolution kernels of size 3×3 and a batch normalization layer. The last is composed of a max-pooling layer. Two dense connectivity strategies improve the usage efficiency of feature maps, enhancing the diagnostic performance for paddy leaf diseases and a 6-way Softmax layer. The Fig. 1 gives a pictorial representation of a simple convolutional neural network.

4 DATA SET

We collected images of the disease of paddy plants from Kaggle. This dataset contains five different diseases of rice and one healthy leaf, which are the most commonly affected diseases in India and other countries. The total number of samples is 636. These images are captured by a thermal camera (FLIR E8). These are- Bacterial Leaf Blight, Blast, Hispa, Leaf folder, Leaf spot, and healthy leaves are shown in Fig. 2.

4.1 Thermal image pre-processing

Deep learning-based image classification needs a lot of images to train the model. But it is not always possible to get a suitable image data set on that problem domain like a data set of paddy disease images. So, in image pre-processing steps, randomly transform the original IR images via a series of random translation rotations, etc., without reducing its quality from images and enhancing the quality of images.

4.2 Image Data Augmentation

Training CNNs requires substantial data. The more data CNNs have to learn, the more needed features they can obtain. Since the original paddy leaf image dataset collected in this study is insufficient, it is necessary to expand the dataset by different methods to distinguish the different
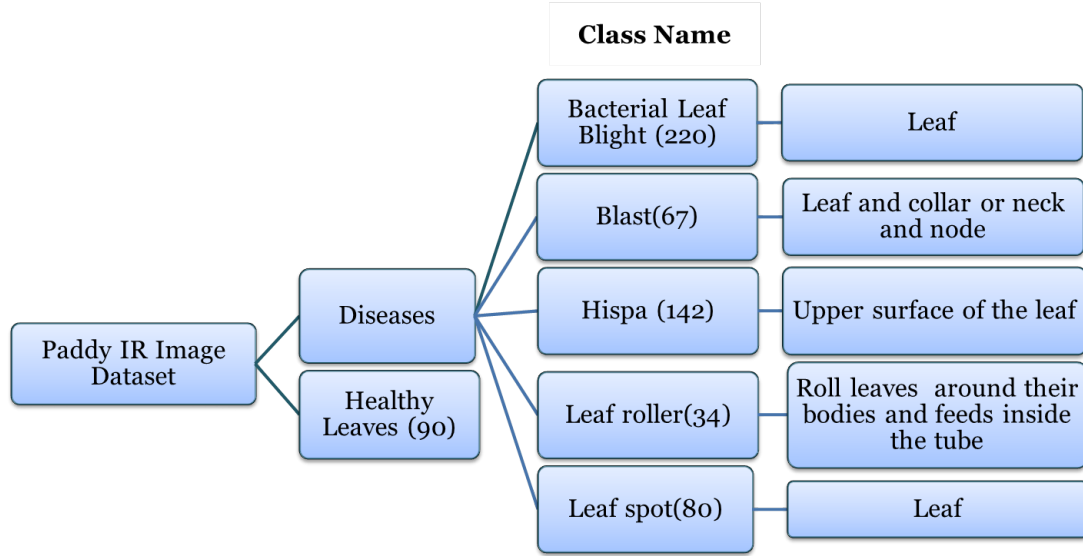
Fig. 2. List of rice diseases and their description.

disease categories. The deep learning neural network provides the capability to fit needed models using image data augmentation via the ImageDataGenerator class. A range of methods are used, as well as pixel scaling methods. Some ways followed are Image shifts, Image flips, mage rotations, Image brightness, and Image zoom.

5 EXPERIMENTAL SETUP AND RESULTS

After data augmentation and preprocessing, six hundred thirty-six included the paddy thermal images in the training set. Five common paddy leaf diseases were trained and classified. In the experiment, acquired training pictures were classified in advance and placed in the folder corresponding to the name of the disease category so that the respective names were taken as the category "labels." Before training, the weight decay was 0.0001 and used the Adam optimizer for iterations. After that was recorded with high accuracy at the epochs of 175, and the training model was saved. The code ran on Google Colab with Keras, TensorFlow, OpenCV, etc. Performance metrics are available to evaluate several performance metrics' paddy disease detection techniques. This chapter uses high detection accuracy to differentiate the number of classes correctly. The classification accuracy of the model was measured using the formula.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Where TP is true positives, TN is true negatives, FP is false positives and FN is false negatives.

5.1 Optimal model selection

We have executed the proposed CNN model based on the epoch to select the best model. An epoch is a hyperparameter that defines a single pass through the complete training set when training a deep learning model. The highest accuracy of 88.28% was achieved by the proposed CNN model with 175 epochs using the augmented data set.
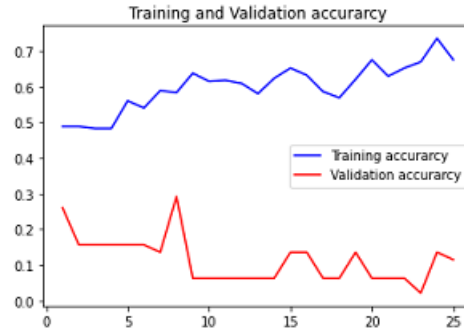
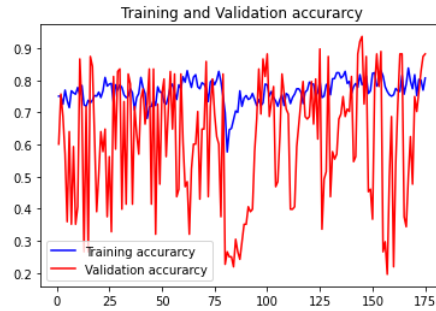**Fig. 3.** Comparison of training and validation accuracy (when epochs 25)

Fig. 4. Comparison of training and validation accuracy (when epochs 175).

5.2 Results and discussion

The data is divided into two sets: training and testing. They evaluated the final stage ran on a Google Colab. We have collected a data set of images of paddy diseases, including the five most commonly affected diseases. We use 80% images (510) of our data set for the training stage to train the model. 126 images are separated, 20% of the testing data set. We observed both the training and validation accuracy. Since then, we have used epochs 175. Finally, the proposed model has achieved a prediction accuracy of 88.28% when epoch 175 that CNN works well on a small data set for perfectly classifying an image due to its kernel trick. Therefore, CNN has been used in this proposed model. From Fig. 3 and 4, it can be observed that both the training and validation accuracy. Since then, we have set epochs 25 and 175.

6 CONCLUSION

Identification and classification of paddy disease manually are impossible for farmers. It takes a lot of time and effort to identify diseases manually. Hence, we developed an effective approach to identify and classify the disease from the paddy plant image. In this paper, the proposed method is aimed to create an automated recognized system to classify images of paddy diseases through the execution of an AI. We have used five diseases of paddy plants which are the most affected diseases. Then we built an image data set of paddy diseases. 80% of the data set was used for training purposes, and we achieved 88.28% prediction accuracy in the training phase. This proposed identity system fruitfully classified 126 images of a test data set of paddy plant diseases. The main goal for future work will be to develop a complete design and will be compared with other epochs based on training and testing set. Second, the model will be compared with other deep learning models. Third, a large amount of the original paddy Infrared thermal image dataset will be used.

References

1. Agriculture, and food, `https://www.worldbank.org/en/topic/agriculture/overview`

2. Ahmed, K., Shahidi, T.R., Alam, S.M.I., Momen, S.: Rice leaf disease detection using machine learning techniques. In: 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI). pp. 1–5. IEEE (2019)

3. Arumugam, A.: A predictive modeling approach for improving paddy crop productivity using data mining techniques. Turkish Journal of Electrical Engineering and Computer Sciences **25**(6), 4777–4787 (2017)

4. Atole, R.R., Park, D.: A multiclass deep convolutional neural network classifier for detection of common rice plant anomalies. International Journal of Advanced Computer Science and Applications **9**(1) (2018)

5. Badage, A.: Crop disease detection using machine learning: Indian agriculture. Int. Res. J. Eng. Technol **5**(9), 866–869 (2018)

6. Barik, L.: A survey on region identification of rice disease using image processing. Int J Res Sci Innov **5**(1) (2018)

7. Camargo, A., Smith, J.: An image-processing based algorithm to automatically identify plant disease visual symptoms. Biosystems engineering **102**(1), 9–21 (2009)

8. Casanova, J.J., O'Shaughnessy, S.A., Evett, S.R., Rush, C.M.: Development of a wireless computer vision instrument to detect biotic stress in wheat. Sensors **14**(9), 17753–17769 (2014)

9. Ferentinos, K.P.: Deep learning models for plant disease detection and diagnosis. Computers and electronics in agriculture **145**, 311–318 (2018)

10. Gayathri Devi, T., Neelamegam, P.: Image processing based rice plant leaves diseases in thanjavur, tamilnadu. Cluster Computing **22**(6), 13415–13428 (2019)

11. Hasan, M.J., Mahbub, S., Alom, M.S., Nasim, M.A.: Rice disease identification and classification by integrating support vector machine with deep convolutional neural network. In: 2019 1st international conference on advances in science, engineering and robotics technology (ICASERT). pp. 1–6. IEEE (2019)

12. Jagan Mohan, K., Balasubramanian, M., Palanivel, S.: Detection and recognition of diseases from paddy plat leaf. Int J Comput Appl **144**, 12 (2016)

13. Jayanthi, G., Archana, K., Saritha, A.: Analysis of automatic rice disease classification using image processing techniques. International Journal of Engineering and Advanced Technology **8**(3S) (2019)

14. Jiang, F., Lu, Y., Chen, Y., Cai, D., Li, G.: Image recognition of four rice leaf diseases based on deep learning and support vector machine. Computers and Electronics in Agriculture **179**, 105824 (2020)

15. Liang, W.j., Zhang, H., Zhang, G.f., Cao, H.x.: Rice blast disease recognition using a deep convolutional neural network. Scientific reports **9**(1), 1–10 (2019)

16. Pavithra, S., Priyadharshini, A., Praveena, V., Monika, T.: Paddy leaf disease detection using svm classifier. International Journal of communication and computer Technologies **3**(1), 16–20 (2015)

17. Rautaray, S.S., Pandey, M., Gourisaria, M.K., Sharma, R., Das, S.: Paddy crop disease prediction—a transfer learning technique. International Journal of Recent Technology and Engineering **8**(6), 1490–1495 (2020)

18. Singh, A., Ganapathysubramanian, B., Singh, A.K., Sarkar, S.: Machine learning for high-throughput stress phenotyping in plants. Trends in plant science **21**(2), 110–124 (2016)

19. Wspanialy, P., Moussa, M.: Early powdery mildew detection system for application in greenhouse automation. Computers and Electronics in Agriculture **127**, 487–494 (2016)

20. Zhu, S., Xia, X., Zhang, Q., Belloulata, K.: An image segmentation algorithm in image processing based on threshold segmentation. In: 2007 third international IEEE conference on signal-image technologies and internet-based system. pp. 673–678. IEEE (2007)