Automatic Diseases Detection and Classification in Maize Crop using Convolution Neural Network

Kifayat Ullah*[1], Muhammad Asim Jan[1], Ali Sayyed[1]
[1]Computer Science Department, CECOS University of IT and Emerging Sciences, Peshawar, Pakistan
*kifayat@cecos.edu.pk

ABSTRACT

Maize is a major crop in Pakistan and it plays an important role in the economy of the country. However different types of plant diseases could affect both the quality and quantity of maize production. To cope with such issues, the majority of the farmers still depend on traditional methods, which are expensive, time-consuming, laborious, and not very effective. To address the issues, we proposed a Convolution Neural Network (CNN) based solution for the detection and classification of different types of maize diseases. We used a publicly available free dataset of 4000 images. The images were classified into four categories. The first three categories represent the Common rust, Cercospora leaf spot grey, and Northern leaf blightand diseases, while the last category represents normal leaves. To test, implement, and evaluate the performance of our proposed method, we used a MATLAB simulation environment. We also compared our results with two other solutions, available in the literature. Our solution achieved 96.53 % accuracy. From the results, we concluded that the proposed method could be used for the automatic detection and classification of different types of maize diseases.

Keywords: Convolution Neural Network (CNN), Image processing, Maize diseases.

1. INTRODUCTION

Plants play an important role in our daily life as they produce grains, vegetables, fruits, and seeds. More than 240,500 species of plants have been found. Besides providing food, plants are also important for the environment. It also provides oxygen. Maize is one of the top-most produced crops in the world, along with rice and wheat [1]. Its grains are used in many nutrition products. However, certain diseases on maize crops lead to a reduction in the quality and quantity of its production [2]. In maize, the three most common diseases are Common rust, Cercospora leaf spot grey, and Northern leaf blight. The effect of these diseases on the leave is shown in the figure. 1. Monitoring of leaf diseases play an important role in the successful cultivation of the crop in the farms.
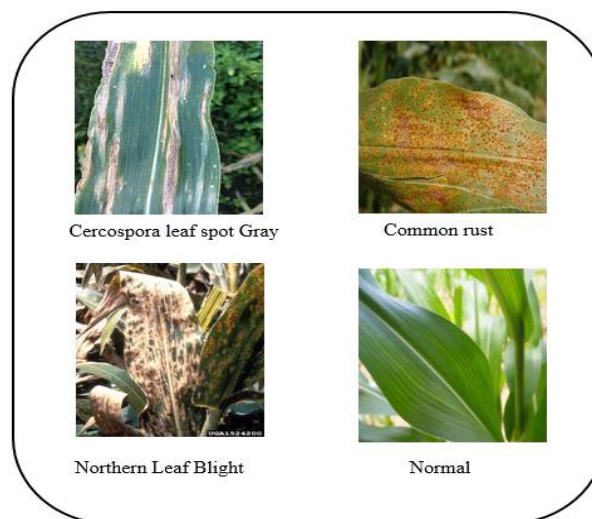

Figure 1: Different types of maize diseases

Traditionally, the farmers find different types of maize diseases by simply observing the maize plants. However, most of the times, the farmers have no idea about the particular type of disease and its prevention methods. Hence, they approach experts to sort out the issues. In some cases, they even need to travel long distances to find an expert, which could be very expensive and time-consuming.

The adoption of advanced technological solutions, especially deep learning techniques, for the detection and categorization of plant diseases is an emerging research area. These solutions help in monitoring large fields of crops by detecting the symptoms of certain kinds of diseases automatically [3]. Deep learning techniques achieve very useful results in different fields, e.g., robotics, smart transportation, smart cities, bio-informatics, and computer vision. For this research work, we used a famous deep learning technique called Convolution Neural Network (CNN).

In Pakistan, maize is a major crop with great significance. It contributed to the economy of the country. However, the production of this crop has reduced to a great extent, due to certain types of diseases. Early detection of maize diseases is

required for the production of high-quality crops. To address the challenges of maize plants diseases detection and classification, researchers are trying to adopt different techniques, for instance, support vector machine, image processing, and Artificial Neural Network (ANN). To contribute to this area, we proposed a CNN-based solution for automatic maize diseases detection and classification. Our proposed solution will use the deep learning approach while focusing on maize leaves images. The proposed solution is efficient in terms of achieving high accuracy and low loss. To this end, the main objective of our proposed work is to design an efficient, and accurate solution for maize crop diseases detection and classification by using CNN. To achieve our objective, we implement, test, evaluate, and compare the result of our proposed solution with two other existing solutions.

2. LITERATURE REVIEW

The work presented in [4] provides a detailed survey of different deep learning techniques to address various agricultural and food production challenges. The authors examined the particular agricultural and food problems. Additionally, they discussed the specific models and frameworks, used in deep learning. They also consider the preprocessing of user data and the overall performance achieved according to the metrics used in different studies. Moreover, the study compared deep learning with other existing popular techniques, concerning differences in classification or regression performance. They concluded that deep learning techniques provide high accuracy and outperform over commonly used existing image processing techniques.

The work in [5] focused on the deep learning models for plant disease detection and diagnosis. The authors developed a CNN based model to perform plant disease detection and diagnosis using simple leaf images of healthy and diseased plants. After training the model, it starts to perform experiment and give a success rate of 99 percent. The significantly high success rate makes the model very useful. Based on that high level of performance, it becomes evident that CNN are highly suitable for the automated detection and diagnosis of plant diseases through the analysis of simple leaves images.

The research presented in [6] described a methodology for early and accurate plant disease detection, using ANN and advanced image processing techniques. The proposed approach is based on ANN, classifier for classification, and Gabor filter for feature extraction. They also achieves better results with a recognition rate of up to 91 percent.

The work in [7] focuses on disease detection in apples plants through K-mean clustering and image analysis techniques. To classify and recognize different types of diseases, they used the texture and color features, which generally appear in normal and affected areas of the plants.

Another research, proposed in [8], performed a comparison-based study of conventional multiple regression, ANN, and Support Vector Machine (SVM). From their finding, they concluded that SVM based regression approach led to a better description of the relationship between the environmental conditions and disease level which could be useful for disease management.

In the research presented in [9], a model for disease detection in crops was developed. They used a public dataset of 54,306 images of diseased and healthy plant leaves. They trained the model by using deep CNN to identify 14 different crop species and types of diseases. However, the performance of the proposed solution a major drawback, as the disease identification process works slowly.

The work presented in [10] used image processing techniques to identify different types of plant disease from a simple image. Identifying disease can lead to quicker interventions that can be implemented to reduce the effects of crop diseases. This work used a public dataset of 86,147 images of diseased and healthy plants, a deep convolution network, and semi-supervised methods to classify crop species and disease.

Another model for the detection of plant disease through some automatic technique was proposed in [11]. The proposed model has the advantage, as it reduces laborious work of monitoring in big farms. The proposed model automatically detects the symptoms of diseases when they appear on plant leaves. In addition, this work also proposed an algorithm for the automatic detection and classification of plant leaf diseases.

3. METHODOLOGY & EXPERIMENTAL SETUP

In this section, we present our proposed methodology and experimental setup.

3.1 Proposed Methodology

For this research work, we followed a systematic research methodology. The different steps of the methodology are explained in the following subsections and are also depicted in the figure. 2.

Input Image
The input to our proposed model is an image of the leaf of a maize plant. This image will then goes through several steps, which are explained in the following subsections.
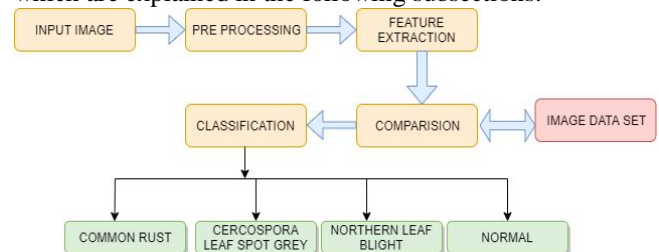

Figure 2: Step-wise research methodology

Preprocessing

Normally, the images contain undesirable features, i.e., clamor, obscure, conflicting, and missing information. To improve the quality of pictures, pre-processing is required. A few strategies for pre-processing of an image include information standardization, histogram leveling, and picture improvement such a clamor evacuation, contrast alteration, and picture honing. In our proposed solution the honed pictures were changed from Red Green Blue (RGB) to Hue-Saturation-Intensity (HSI). The main reason for this conversion is to limit the number of hues.

Features Extraction

We extract several features from the images. The capacities, alongside their marks, were utilized by the classifier to analyze the various pictures and characterize them into appropriate classes.

Image Dataset

We used a publicly available dataset of 4000 pictures. The dataset was downloaded from the kaggle website. The pictures were arranged into four distinct categories. The three categories represent a disease type while the froth category consists of normal leaves.

Comparison

The processed image is then compared with the images in the dataset. Based on the comparison, the image is categorized and is placed in one of the four categories. For this to happen, the model needs to be trained. We used CNN model for this purpose.

Classification

This step is the core of our research. As the primary objective of this work is to detect and categorize the type of disease. Our proposed solution precisely achieved this objective and we will show it in the results and discussion section.

3.2 Experimental Setup

To implement our proposed solution, carry out experiments, and evaluate the performance we used a state-of-the-art professional simulation environment called MATLAB. In addition to MATLAB, we also used computer vision libraries for implementation. We conducted the experiments on leaf images of maize crop, which contained both normal and diseased leaves. As mentioned in the methodology subsection, we used a publicly available dataset of 4000 images. The images are classified into four different categories. The information about each category is also given in the table. 1.

Table 1: Categories of images

S#	Category	Number of images
1	Normal	1000
2	Cercospora	1000
3	Northern leaf blight	1000
4	Common rust	1000
	Total images	4000

4. RESULTS AND DISCUSSION

In this section, we present the results of our experimental work. The focus of our experiments was on calculating the accuracy of the proposed solution. In addition, we also compared the results of our experiments with two other solutions, available in the literature.

There are four types of possibilities to classify an image in a particular class.

False Negative

When an image is classified in a disease class but actually belongs to the normal class.

False Positive

When an image is classified in a normal class but actually belongs to a disease class.

True Negative

When an image is classified in a disease class and it actually belongs to that class.

True Positive

When an image is classified in a normal class and it actually belongs to that class.

We used the following basic equations to do a simple analysis of our results.

$$FPR = \left(\frac{FP}{TN+FP}\right) \ldots\ldots\ldots\ldots(1)$$

$$FNR = \left(\frac{FN}{TP+FN}\right)\ldots\ldots\ldots\ldots(2)$$

$$F - Measure = 2 * (FPR * FNR)/(FPR * FNR)\ldots(3)$$

$$Accurcy = \left(\frac{TP}{TotaltestImages}\right) * 100 \quad \ldots(4)$$

In these equations, FPR stands for False Positive Rate, FP stands for False Positive, TN stands for True Negative, FN stands for False Negative and TP stands for True Positive.

4.1 Categorization of Maize Leaves

Our proposed solution can efficiently categorize the leaves into the respective category. The four categories that we considered are Rust, Cercospora leaf spot grey, Northern leaf blight, and normal (healthy). A visual result of the different categories is shown in the figure. 3.
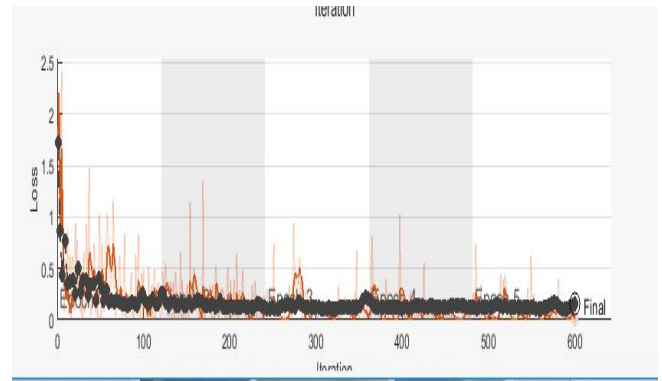
Figure. 5: Loss in proposed solution

Figure 3: Visual results of leaves classification

4.2 Accuracy

To find the efficiency of our proposed solution, we considered the accuracy parameter. The accuracy of a machine learning classification algorithm is one of the most important ways to measure how efficiently the algorithm classifies an input image. In other words, we can say that accuracy is the number of correctly predicted data points out of all the data points. In our proposed solution, we achieved an accuracy of above 96 percent, which means that our proposed solution was able to accurately classify the images in the respective category 96.53 % of the time. The graph depicted in the figure. 4 further elaborate our results. It is important to note that an epoch is one complete cycle through the dataset. We consider 5 epochs for our experiments.
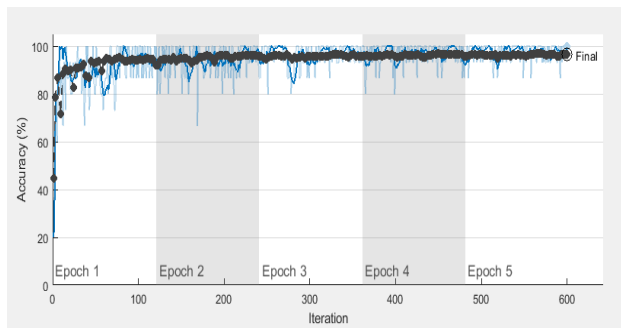
4.4 Comparison with other solutions

Finally, we compare the performance of our proposed solution with two existing solutions, in terms of accuracy. The results in figure. 6 reveal that our proposed solution outer performed, as compared to KNN and Ensemble. In the case of KNN, the accuracy is 80.2 percent, while in the case of Ensemble the accuracy is 84.6 percent. However, it can be seen that in our case, the achieved accuracy is 96.53 percent.
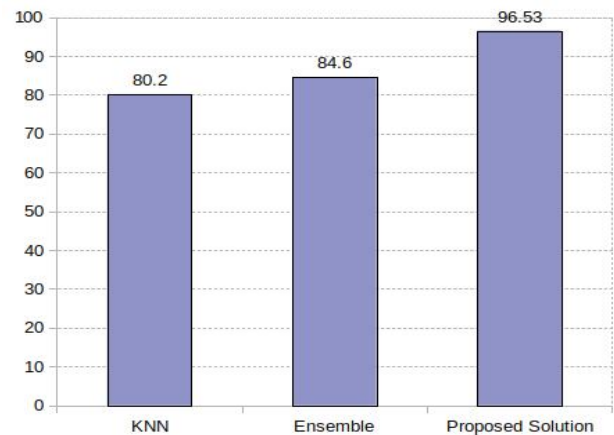

Figure 6 Comparison with different solutions

Figure 4: Accuracy of proposed solution

4.3 Loss

Another important parameter to measure the effectiveness of a solution is called Loss. Normally we measure these parameters in percentage. But loss can also be measured in terms of the number of errors made by the model. Another difference is that for accuracy we are interested in achieving a maximum number, while in case of loss, we expect that our model will generate a low value. Hence, we can say that the objective of any model should be to minimize the loss. To evaluate the performance of our proposed solution, in term of loss, we provide test samples to the model and then we measure the loss. The obtained results are shown in figure 5. From the graph it can be seen that a very small amount of data (less than 2.5) was lost, at the start of the experiments.

5. CONCLUSION AND FUTURE WORK

Maize is a major crop of Pakistan. It contributes to the economy of the country in several ways. However, several diseases can affect the productivity of the maize crop. Traditional approaches of plants diseases are expensive, time-consuming, and laborious. To automate the process of maize leave disease detection and classification, we proposed a CNN-based solution. We used a publicly available dataset of 4000 images to train our model. We divided the images into four different categories of which the three categories represent diseased leaves while the fourth category shows the normal \ healthy leaves. Our proposed solution detect and classify several types of maize diseases. We implement the solution in a MATLAB simulation environment and performed several experiments to test and evaluate its performance. We further compare the obtained results, in terms of accuracy with two other solutions. Our proposed

solution achieved an accuracy of 96.53 %. The obtained results reveal the efficiency of our proposed solution for the detection and classification of maize diseases.

In the near future, we pretend to include other types of maize plant diseases. We also have a plan to implement our solution in real-world scenarios. We will work on the extension of the proposed solution to other types of crops, e.g., wheat, and rice, etc.

References

[1]   J Huerta-Espino, J., Singh, R., Germán, S., McCallum, B., Park, R., Chen, W., Bhardwaj, S. and Goyeau, H., **"Global status of wheat leaf rust caused by Puccinia triticina"**. Euphytica, 179(1), pp.143-160, 2011.

[2]   S. Weizheng, W. Yachun, C. Zhanliang, and W. Hongda, **"Grading method of leaf spot disease based on image processing,"** in 2008 International Conference on Computer Science and Software Engineering, 2008.

[3]   T. Rumpf, A.-K. Mahlein, U. Steiner, E.-C. Oerke, H.-W. Dehne, and L. Plümer, **"Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance,"** Computers and electronics in agriculture, vol. 74, no. 1, pp. 91-99, 2010.

[4]   A. Kamilaris and F. X. Prenafeta-Boldú, **"Deep learning in agriculture: A survey,"** Comput. Electron. Agric., vol. 147, pp. 70–90, 2018.

[5]   K. P. Ferentinos, "**Deep learning models for plant disease detection and diagnosis,"** Comput. Electron. Agric., vol. 145, pp. 311–318, 2018.

[6]   Kulkarni, A.H. and A. Patil, **"Applying image processing technique to detect plant diseases,"** International Journal of Modern Engineering Research, vol. 2(5): pp. 3661-3664, 2012.

[7]   S. Bashir, **"Remote area plant disease detection using image processing,"** IOSR j. electron. commun. eng., vol. 2, no. 6, pp. 31–34, 2012.

[8]   R. Kaundal, A. S. Kapoor, and G. P. S. Raghava, **"Machine learning techniques in disease forecasting: a case study on rice blast prediction,"** BMC Bioinformatics, vol. 7, p. 485, 2006.

[9]   S. P. Mohanty, D. P. Hughes, and M. Salathé, **"Using deep learning for image-based plant disease detection,"** Front. Plant Sci., vol. 7, p. 1419, 2016.

[10]   Cortes, E., **"Plant Disease Classification Using Convolutional Networks and Generative Adverserial Networks,"** Stanford University Reports, Stanford, 2017,

[11]   T. Rumpf, A.-K. Mahlein, U. Steiner, E.-C. Oerke, H.-W. Dehne, and L. Plümer, "**Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance,"** Comput. Electron. Agric., vol. 74, no. 1, pp. 91–99, 2010.