False Smut Disease Detection in Paddy using Convolutional Neural Network

Nahid Hasan
City University
Dhaka, Bangladesh
contact.me.nahid@gmail.com

Tanzila Hasan
City University
Dhaka, Bangladesh
contact.me.onu@gmail.com

Shahadat Hossain
City University
Dhaka, Bangladesh
hossain_shahadat92@outlook.com

Md. Manzurul Hasan
American International University-Bangladesh (AIUB)
Dhaka, Bangladesh
manzurul@aiub.edu

ABSTRACT

Rice false smut (RFS) is the most severe grain disease affecting rice agriculture worldwide. Because of the various mycotoxins produced by the causal pathogen, Villosiclava virens, epidemics result in yield loss and poor grain quality (anamorph: Ustilaginoidea virens). As a result, the farmers' main concern is disease management measures that are effective, simple, and practical. This research proposes a model based on the Convolutional Neural Network (CNN), widely used for image classification and identification due to its high accuracy. We acquire data from actual rice farming fields and high-resolution RFS images from open-source data on the internet. To compare and validate our model's performance, we train and test it using online photos and actual photographs. As a result, our model provides 97 % accurate results for detecting the illness in actual photos. Finally, we evaluate and record all of the data for subsequent studies.

CCS CONCEPTS

• **Computing methodologies** → **Object identification**.

KEYWORDS

cnn, false smut detection image processing

ACM Reference Format:

Nahid Hasan, Tanzila Hasan, Shahadat Hossain, and Md. Manzurul Hasan. 2018. False Smut Disease Detection in Paddy using Convolutional Neural Network. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Rice (Oryza sativa) is the primary crop of many countries, including Bangladesh. In Bangladesh, 90% of all farmers are actively involved in rice farming, and paddy accounts for 95% of the country's food requirements. The value of paddy in socioeconomic growth is enormous. Paddy production, for example, employs a large number of people and generates significant foreign cash in Bangladesh by exporting paddy to other countries each year. According to USDA (United States Department of Agriculture) figures, Bangladesh produces 3.6 crore tons of rice. Furthermore, Bangladesh ranked fourth in rice production, after Indonesia, India, and China. There are several reasons to restrict rice production, with the disease in paddy grains being one of the most important. A survey was conducted in the paddy fields of NRRI, Cuttack, in 2019, and they discovered a significant loss of grain discoloration in the grown rice genotypes, with symptoms in the form of brown/black/ash-colored grains, mainly accompanied by chaffiness. The prevalence of grain discoloration ranged from 25% to 92% in different rice genotypes [1].

This illness has been reported in India, with the maximum infection rate, 85%, in Tamil Nadu [10]. False smut (Ustilaginoidea virens) is a severe crop-damaging pathogen. The disease's usual symptom is the formation of black fungal mycelium in rice grains, which is covered by yellow fungal growth in the field. Mature spores are orange in hue and eventually turn yellowish-green or greenish-black. In most cases, just a few grains in a panicle are contaminated, with the remainder being normal. This illness is more common when farmers grow paddy crops than when they first saw them. As a result, the grain has a low market price because of the damaged color and poor quality. As a result, the spread of false smut diseases is becoming more common. In recent years, an increase in the prevalence of RFS has been documented in most major rice-growing regions worldwide, including China, India, and the United States [12]. That is why we created a CNN system that can readily identify paddy phony smut sickness. The rest of this article's information have organized as follows. The second section discusses the connected works. Section 3 delves into the analysis approaches, while Section 4 describes our trial with the CNN model and shows the results of our model's performance on the dataset, as well as a discussion of it. Finally, before concluding, section 5 discusses our future research directions.

2 RELATED WORK

The study of diagnosis from a photograph is an intriguing one in agriculture. The technology for detecting rice disease has been established based on images of sick rice plants. They detect the extracted color, form, and texture for bacterial leaf blight, brown

Woodstock '18, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

spot, and leaf smoot. In terms of accuracy, 93.33% of the training dataset was corrected, while 73.33% of the test dataset was corrected [7]. One of the newest research subjects that sparked our interest in agriculture is diagnosing leaf pictures using image processing techniques. This research suggests utilizing an Optimized Deep Neural Network with the Jaya Algorithm to recognize and classify paddy leaf diseases. They took photos straight from the paddy field and identified four diseases: bacterial blight, blast, brown spot, and sheath rot. Furthermore, RGB photos have been transformed into HSV images enabling color-based background removal and masking in pre-processing. Furthermore, a clustering algorithm was employed to divide sick portions, standard parts, and background parts [8].

Identifying disease from tree leaf images is becoming an exciting field of study in agriculture. From the images of rice leaves, three diseases have been identified: Bacterial Leaf Blight, Brown Spot, and Leaf Smut. The Deep Convolutional Neural Network Alexnet model was employed for accurate feature extraction [9]. Rice is one of the most significant crops in the current context, as the economy is dependent on agricultural production and productivity. Moreover, even rice is a staple crop for most people in most nations. However, rice is grown in the field and is susceptible to bacterial and fungal infections. As a result, we must be more cautious in rice production and diagnosis. Three rice diseases were identified through image processing: Bacterial Leaf Blight, Brown Spot, and Leaf Smut. The Rice Leaf Disease dataset from the UCI Machine Learning Repository was employed, and the remaining neural networks obtained roughly 95.83% accuracy in classifying the pictures into the intended disease category [6]. These studies focus on three types of diseases: paddy blast, brown spot disease, and narrow brown spot disease. They were using the Matlab tool. The neural network is used to categorize paddy illnesses such as paddy blast, brown spot disease, narrow brown spot disease, and normal paddy leaf disease. As training pictures, ten samples of Blast Disease image, ten of Brown Spot Disease image, and ten of Narrow-Brown Spot Disease image are employed. Besides, the overall test result is 92.5% [13].

Rice is India's principal crop and has the most paddy agriculture area, including brown and white paddy cultivation. Paddy is grown in nearly every state in India. Agriculture employs more than three-quarters of India's population. The first cause of rice plant illness is fungal/bacterial assault, and the second is unexpected climatic change. Grain plant diseases or climate change can cause famine. It might also have a negative impact on the economy. Rice blasts, rice blight, brown spots, leaf smut, tungro, and sheath blight are the most common rice diseases. Rice disease is the most common concern for most farmers; thus, early diagnosis is critical, according to the [11].

3 METHODOLOGY

In this section we explain our workflow of analyses and required methods which we have used to observe our online and collected dataset and to detect Paddy False Smut disease (Figure 1).

3.1 False Smut Disease

Paddy is the major crop in many nations throughout the world. Thus its production and diagnosis are critical for us [9] There are

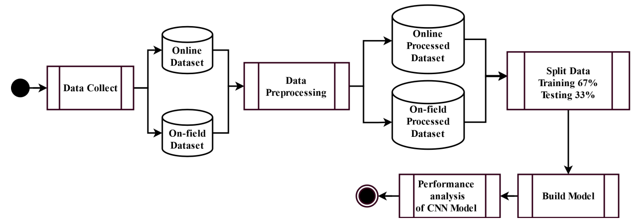


Figure 1: Work flow diagram

several diseases in paddy. That is why we lose paddy production every year. One of these is False Smut Disease. The fungal pathogen Ustilaginoidea virens, which produces both sexual ascospores and asexual chlamydospores in its life cycle, is the primary cause of Rice false smut disease, according to [2].

(a) False Smut Disease **(b) Healthy Rice**

Figure 2: Affected rice and healthy rice

3.2 Data Collection and Preprocessing

Dataset is the most crucial part of image processing. So, we should be careful when we collect data. The imbalanced dataset is one of the significant problems in dataset[5]. Our research works for two types of datasets: the online dataset (we collect images from the internet) and the On-field dataset (we collect images from the paddy field).In every dataset, we have two types of class Disease and Healthy. After collecting data, we have pre-processed our data. Then, we use image augmentation for re-scaling, crop, rotation, zooming, resizing images, formatting images.

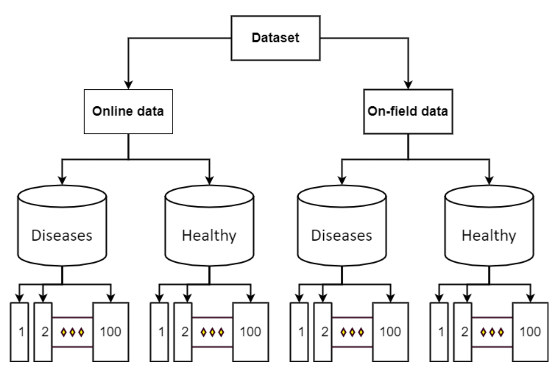


Figure 3: Dataset

3.3 Convolutional Neural Network

The Convolutional Neural Network (CNN) is a deep learning neural network class. In a nutshell, CNN is a robust machine learning technique for automatic image categorization processing. CNN is commonly utilized for image segmentation and is also used in image classification. Convolution neural network (CNN) plays an integral part in many classic image classification methods [4]. CNN's method includes some layers for detecting actual output. The input image is convolved through a filter collection in the convolution layers. We employ binary classification and merge each feature map with a fully linked network. That is why the sigmoid algorithm is used in our research.

4 RESULT AND PERFORMANCE ANALYSES

Metrices is a function that we use to evaluate the performance of our model. There are several metrics in CNN. We show accuracy metrics such as precision, specificity, sensitivity, and f1 score in Table 1. Accuracy metrics generate two local variables, total and count, used to determine the frequency with which Y pred and Y true fit [3]. In our research, we achieve an accuracy of 90% (on-field dataset) and 95% (off-field dataset) (online dataset). Precision is one metric used to assess the performance of a machine learning model. It is calculated by dividing the number of true positives by positive forecasts. Our research has obtained a precision of 93.75% (online dataset model) and 90% (on-field dataset model).

Table 1: Result Table

	Precision	Specificity	Sensitivity	F1 score	Accuracy
Online Dataset	93.75%	94.11%	93.75%	93.75%	95.45%
On-field Dataset	90.62%	91.17%	90.62%	90.62%	90.90%

Sensitivity metrics assess a model's ability to predict true positives in each accessible category. We achieve a sensitivity of 93.75% (online dataset model) and 90.62% (offline dataset model) (on-field dataset model). Specificity measures a model's ability to predict the actual negatives of each accessible category. We acquire a specificity of 94.11% (online dataset model) and 91.17% (offline dataset model) (on-field dataset model). The F1 score is a better metric than accuracy since it is the harmonic mean of precision and recall. The F1 score is 93.75% (online dataset model) and 90.62% (offline dataset model) (on-field dataset model). Confusion metrices are a summary of classification problem prediction outcomes. We provide confusion metrices for two datasets: one for online images and one for on-field images in Figure 6.

After fitting our model to both datasets (online and on-field), we obtain two curves: accuracy and loss curves (training and validation). The number of epochs in the accuracy curve is shown by the x-axis, while the y-axis represents the number of accuracies. Likewise, the x-axis represents the number of epochs in the loss curve, while the y-axis represents the number of validations.

5 FUTURE RESEARCH DIRECTION AND CONCLUSION

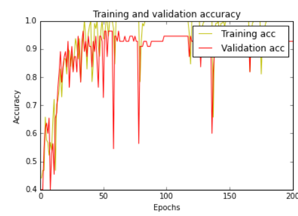
Any scientific work may be improved indefinitely. Nevertheless, every study has limitations, and our analyses are no exception. We identified the last stage of the fake smut illness based on healthy and diseased photos from both on-field and online databases. We shall endeavor to gather the second stage of this disease in our future expanded effort to assure the early diagnosis of false smut disease.

Paddy is Bangladesh's most important food crop, accounting for 80% of the country's total land area. Paddy agriculture employs 90% of the overall farmer population. The value of paddy in socioeconomic growth is enormous. Unfortunately, many paddies are destroyed by false smut disease. As a result, farmers must bear a large number of losses. Not just farmers, but the country's economy, have a significant influence. As a result, we have developed a model to detect false smut disease and discovered substantial results using two types of datasets. We hope that our study effort is the conception of a solution to one of Bangladesh's most pressing agricultural challenges and that it inspires other scholars to investigate agriculture's well-being.

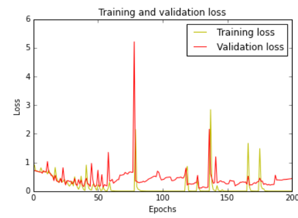
(a) Online dataset model confusion metrices

(b) On-field model dataset confusion metrices

Figure 4: Confusion metrices

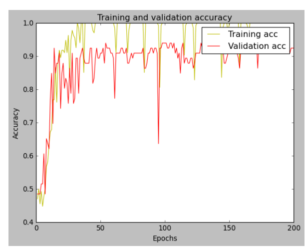


(a) Accuracy curved

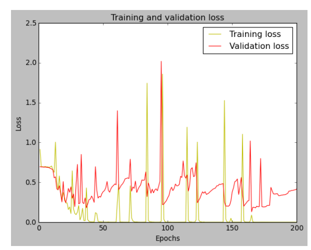


(b) Loss curved

Figure 5: Online Dataset model curved



(a) Accuracy curved



(b) Loss curved

Figure 6: On-field Dataset model curved

REFERENCES

[1] Mathew S Baite, S Raghu, SR Prabhukarthikeyan, U Keerthana, Nitiprasad N Jambhulkar, and Prakash C Rath. 2020. Disease incidence and yield loss in rice due to grain discolouration. *Journal of Plant Diseases and Protection* 127, 1 (2020), 9–13.

[2] A Biswas. 2001. False smut disease of rice: a review. *Environment and Ecology* 19, 1 (2001), 67–83.

[3] Peter A Flach. 2003. The geometry of ROC space: understanding machine learning metrics through ROC isometrics. In *Proceedings of the 20th international conference on machine learning (ICML-03)*. 194–201.

[4] Xinyu Lei, Hongguang Pan, and Xiangdong Huang. 2019. A Dilated CNN Model for Image Classification. *IEEE Access* 7 (2019), 124087–124095. https://doi.org/10.1109/ACCESS.2019.2927169

[5] Victoria López, Alberto Fernández, and Francisco Herrera. 2014. On the importance of the validation technique for classification with imbalanced datasets: Addressing covariate shift when data is skewed. *Information Sciences* 257 (2014), 1–13.

[6] Sanjay Patidar, Aditya Pandey, Bub Aditya Shirish, and A Sriram. 2020. Rice plant disease detection and classification using deep residual learning. In *International Conference on Machine Learning, Image Processing, Network Security and Data Sciences*. Springer, 278–293.

[7] Harshadkumar B Prajapati, Jitesh P Shah, and Vipul K Dabhi. 2017. Detection and classification of rice plant diseases. *Intelligent Decision Technologies* 11, 3 (2017), 357–373.

[8] S Ramesh and D Vydeki. 2020. Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm. *Information processing in agriculture* 7, 2 (2020), 249–260.

[9] Divvela Srinivasa Rao, N Kavya, S Naveen Kumar, L Yasaswi Venkat, and N Pranay Kumar. 2020. Detection and classification of rice leaf diseases using deep learning. *Int J Adv Sci Tech* 29, 03 (2020), 5868–5874.

[10] Prabira Kumar Sethy, Nalini Kanta Barpanda, Amiya Kumar Rath, and Santi Kumari Behera. 2020. Rice false smut detection based on faster R-CNN. *Indonesian Journal of Electrical Engineering and Computer Science* 19, 3 (2020), 1590–1595.

[11] Jitesh P Shah, Harshadkumar B Prajapati, and Vipul K Dabhi. 2016. A survey on detection and classification of rice plant diseases. In *2016 IEEE International Conference on Current Trends in Advanced Computing (ICCTAC)*. IEEE, 1–8.

[12] Wen-Ming Wang, Jing Fan, John Martin Jerome Jeyakumar, and Y Jia. 2019. Rice false smut: an increasing threat to grain yield and quality. *Protecting rice grains in the post-genomic era. London: IntechOpen* (2019), 89–108.

[13] Radhiah Zainon. 2012. *Paddy disease detection system using image processing*. Ph.D. Dissertation. UMP.