Maize Disease Multi-Classification: Leveraging CNN and Random Forest for Accurate Diagnosis

Eshaan Singh
Swami Vivekanand Institute of Engineering & Technology, Ramnagar, Punjab, India.
eshaans59814@gmail.com

Rahul Chawla
Swami Vivekanand Institute of Engineering & Technology, Ramnagar, Punjab, India.
rc4548301@gmail.com

Ramanpreet kaur
Swami Vivekanand Institute of Engineering & Technology, Ramnagar, Punjab, India.
rpk2114@gmail.com

Vinay Kukreja
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India.
onlyvinaykukreja@gmail.com.

2024 International Conference on Automation and Computation (AUTOCOM) | 979-8-3503-8272-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/AUTOCOM60220.2024.10486145

***Abstract*— This research is a significant progress in the arable. However, Maize faces various diseases that compromise its agricultural productivity despite being a major staple crop worldwide. In this paper, we take advantage of the Random Forest Classifiers and Convolutional Neural Networks (CNNs) to provide a new approach towards solving a critical problem – diagnosing maize illness. It is important to emphasize the early and accurate diagnosis because maize diseases such as Maize Dwarf Mosaic Virus, Northern Corn Leaf Blight, Gibberella Ear Rot, Grey Leaf Spot are capable of causing massive losses in yield; however not limited only with it the financial consequences. We collected a set of healthy and disease images of maize which is highly heterogeneous in nature, then we performed some image augmentation to improve the quality if any. As for preprocessing, these images zoomed in and out of different positions translated at various angles to capture better quality dataset, while same left wings are rotated 35 degrees and all right wins remain unchanged cropped the resulting image which numbers over thousands per nine datasets. Features were extracted using CNNs, and then the classifier algorithm was utilized with Random Forest. Promising observations were made after evaluating the performance of the model against multiple classes of diseases. While strictly speaking this suggested model is not in complete accordance with the classic supply-demand curve from traditional economic theory, it provides a very accurate and dependable picture of human behavior on wheat markets over time; as such the above can be considered useful for academia or anyone involved in farming looking to maintain as well improve upon our ability to produce this great staple food.**

Keywords—Maize diseases, Crop health, Convolutional Neural Networks (CNNs), Agricultural technology, Deep learning.

I. Introduction

Maize, being one of the world's most important cereal crops and third after rice and wheat faces huge challenges that pertain to disease infection leading to considerable loss in productivity. In this investigation, an innovative method of accurate diagnosis for maize disease will be proposed through the use of Convolutional Neural Networks (CNNs) as a feature extractor and a Random Forest Classifier as its classifier. Utilization of the most sophisticated technologies in crop health management helps not only to prevent diseases but also to create sustainable agricultural practices. Minor objective The fruit-logistic distributed model developed in this project demonstrates that as a development initiative, it is not only viable for the future success of the Agriculture-N business venture but also indicates how monitoring can be remotely done at all stations and PPSs opening windows informing improvements through iterative adaptations. IVIVI frame work underscores those changes come from IDSI strategies including models themselves; however parameters need to captured directly. The diseases targeted include Maize Dwarf Mosaic Virus, Northern Corn Leaf Blight, Gibberella Ear Rot, Grey Leaf Spot, and Maize Streak Virus, all of which can lead to significant yield losses and economic consequences if not diagnosed and addressed early. The motivation behind this research lies in the importance of early and precise diagnosis in agriculture. We address this challenge by employing advanced technologies like deep learning and machine learning, specifically CNN and Random Forest Classifiers. Accurate diagnosis of maize illness appears promising with the combination of Random Forest Classifiers and CNNs. Using image augmentation techniques, we aim to increase the overall performance of our model by improving the quality of our dataset, which consists of several photos of both ill and healthy maize..Preprocessing entails zooming, translating, rotating, and cropping the dataset to guarantee that it is complete and representative for model building. The methodology encompasses data collection, preprocessing, feature extraction using CNNs, and classification through the Random Forest algorithm. The research aims to evaluate the model's performance across various disease classes, assessing accuracy and reliabilityIn order to conserve and improve maize crops, farmers and academics alike can benefit greatly from our suggested model, which highlights the importance of timely and precise diagnosis in agriculture.Future research can explore the scalability and adaptability of the model across different agricultural contexts and crop types, paving the way for practical implementation in artificial agricultural systems.

II. Literature Review

This study is about how Maize Streak influenced the growth and yield of different maize crops that were in a different state of resistance to these diseases. The study under review addresses an acute issue of modern science, which is the diagnosis and treatment of diseases that extend to maize crops working out a method consisting in combination

Convolutional Neural Networks (CNNs) with Random Forest Classifiers.

A recent two-year study from 2019 to 2021 was carried by different treatments to ascertain the impact of southern leaf [2] blight resulting due Bipolaris maydis, concerning disease level in Colonia Tovar as well. Seeds were treated and sprayed with Propiconazole at 0.1% for the highest grain yield and avoidable yield loss over both years. This treatment leads to the highest 1000-grain weight and the most significant reduction in avoidable yield loss due to 1000-grain weight. In one study it was found that for every 1% increase in disease severity, grain yield decreased by 0.02 q ha^-1. Overall, The treatment of seeds and spraying with propiconazole at 0.1% proved to be most effective in reducing disease severity and led to maximum profits, with a BC ratio of 4.05.

Maize ear rot is a disease [3] that was caused by Fusarium graminearum. This type of disease can occur before or after the harvest. It results in reduced grain quality and weight. Epidemic conditions can lead to significant ear rot in up to 50-75% of ears. The data representation and statistics analysis using SPSS through tables and bar charts revealed the occurrences of various types of pathogens in surveyed areas, with Biu having the highest infection at 19.78%, followed by Hawul and Askira/uba at 15.56% each. Farmers were aware of the disease's impact on grain quality, which is influenced by maize variety and rainfall. The study recommends farmers obtain certified seeds to mitigate maize ear rot.

In the central [4] temperate region of Argentina, where late-sown crops are prevalent, foliar diseases pose a threat to food security. Fungicide application is shown to increase yields, emphasizing the role of external interventions in disease management.

This study introduces an [5] innovative system for quantifying maize leaf diseases, eliminating the need for specialist input and avoiding bias. Historically, disease severity assessment relied on expert knowledge, but this approach replaced it with an automated process. The advantages include reduced labor, time, and error in comparison to the conventional method, which lacks standardized guidelines. The research involves training deep-learning models for disease classification, extracting class activation heatmaps, and developing an adaptive thresholding technique for region extraction. Transfer learning achieves up to 99% accuracy in disease classification, and the proposed method offers accurate quantification of maize leaf diseases, independent of domain knowledge.

.This study provides valuable insights into the prevalence of maize diseases in Manipur, which is critical for effective crop management and food security in the region. Maize plays a crucial role [6] as the second most important crop, serving both as a staple for consumption and as a key component in livestock feed for pig and poultry farming. A survey conducted in 2018 and 2019 across five districts revealed the prevalence and severity of maize diseases.

. This study delves into the role of mycotoxins and the optimal conditions for fungal growth and mycotoxin production. Maize is a crucial global [7] crop for food and feed production, but it faces significant challenges from various diseases, particularly those caused by Fusarium species. These pathogens not only lower the quality and quantity of the maize kernels but can become a serious health hazard as they contaminate the grains with mycotoxins. Fusarium diseases result in the production of severely toxic mycotoxins that can cause serious illnesses to consumers therefore there is regulatory effort to limit levels of these types of mycotoxin, especially those found in foods. The prevalence of important Fusarium species in global natural and storage conditions is studied, as well as interlinked mycotoxin contamination.

The study brings to light the [8] pervasive nature of MLN in Kenya and appeals to farmers to adapt resistant maize varieties. A very dangerous disease that can cause entirely loss of crops is Maize Lethal Necrosis (MLN) which was first reported in Kenya in 2012. This arises from the combination of two viruses, a Maize chlorotic mottle virus (MCMV) with a potyvirus thus making its epidemiology complex. A study of 406 farmers in different regions of Kenya found that MLN was still a major concern. These symptoms were characterized by leaf chlorosis, necrosis and premature death of the plants. Molecular testing for example confirmed the presence of MCMV and SCMV in maize fields with differential levels of occurrence and severity.

In addressing the challenges posed by diseases in maize this study [9] uses modern technologies to offer a new solution for accurate diagnosis. The diseases focused on in this study are Maize Dwarf Mosaic Virus, Northern Corn Leaf Blight, Gibberella Ear Rot and Grey Leaf Spot as well as Maize Streak Virus. All these diseases present unique challenges and can cause major losses to maize crops at different growth stages.

Maize encounters challenge [10] of infections from Fusarium type diseases resulting into contamination caused by mycotoxins. The present study is aimed at discovering the global Fusarium species distribution and emphasizing potential health hazards related to contamination with mycotoxins.

In several recent studies, the investigators have looked to revolutionary approaches in managing maize diseases such as federated learning CNNs and deep learning algorithms[11] These developments demonstrate a movement towards the use of state-of-the-art technology to enable precise and faster disease diagnosis.

Southern Leaf Blight caused by Bipolaris maydis has been investigated in this study [12]. Treatment of diseased plants using effective chemicals such as the application of Propiconazole has been embraced to minimize disease severity and grain yield, thus underpinning the economic value associated with approaches that are aimed at managing plant diseases.

Automated processes for disease severity assessment were recently introduced [13] based on deep learning techniques. This method is more efficient since it reduces labor, time, and error as compared to the traditional methods of quantifying maize leaf diseases. Maize Ear Rot caused by Fusarium graminearum has been viewed as a post-harvest disease of the maize grain having adverse effects on quality and weight [14]. The pathogen's occurrences in the farmers' fields increase knowledge of it and, therefore, understanding. The study recommends mitigating ear rot through strategies like seed certification.

III. Research Methodology

The research methodology involves the careful collection of a diverse dataset comprising healthy and diseased maize images. Image augmentation techniques are employed to enhance the dataset's quality, and preprocessing steps, including zooming, translation, rotation, and cropping, contribute to improving the model's overall performance. The trained model uses a Convolutional neural network for feature extraction and a Random Forest Classifier for classification of the images. The total CNN parameters for this model were 162,921,728 approximate. As represented in Fig.1 the research methodology section consists of 4 parts and these 4 parts are Data collection, Data Preprocessing, Feature Extraction, and classification. Each step in methodology plays an important role in obtaining accurate results.

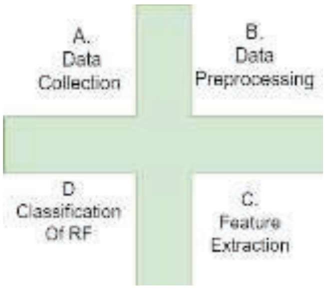

Fig. 1. Four State Flow Chart of Methodology

A. Data Collection

In the recognition of maize Diseases, the total number of images has been calculated using primary and secondary resources. Primary images of the diseased and healthy maize are collected from the nearby province of Punjab. The camera used for this data collection is Canon EOS 90D. The secondary images are collected from the internet. These images are further flipped, cropped, and faded to increase the quality of this dataset. The total number of images used in this research is 2853.

Fig. 2. Diseased Maize image from the dataset

Fig. 3. Healthy Maize image from the dataset

In Figure 2., one of the diseased Maize image from the dataset has been shown. Figure 3., shows a healthy Maize image from the dataset. 2853 of such images were used to train the model.

B. Data Preprocessing

Before training the model, data preprocessing was used to enhance the performance of the proposed model. Various data preprocessing methods have been used on the collected images. The steps that have been taken for preprocessing include zooming, translation, rotation, flipping, cropping, etc.

C. Classification of RF

A Convolutional Neural Network (CNN) is a specific network architecture designed for deep learning algorithms, primarily employed in tasks related to image recognition and the processing of pixel data .In Fig. 4 CNN + Random forest classifier model leverages a hybrid architecture, combining Convolutional Neural Networks (CNNs) and Random Forest Classifier, to address the intricate challenge of maize disease multi-classification.

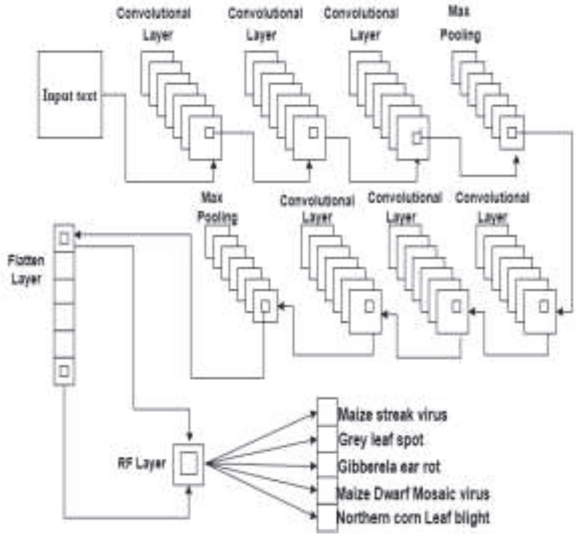

Fig. 4. CNN + Random Forest Classifier Model.

D. Feature Extraction

The Random Forest Classifier is a popular machine learning algorithm that belongs to the ensemble learning category. Ensemble learning involves combining the predictions of multiple models to improve overall performance and accuracy Random Forests are commonly used in various applications, including image classification, disease diagnosis, and financial forecasting, among others.

IV. RESULT

The result section consists of 4 Tables. Table 1 is confusion matrix, table 2 is Meta analysis, table 3 is result section and table 4 is convolutional layer. Here are these 4 tables as follows:

TABLE I. CONFUSION MATRIX

Classes	Precision	Recall F1-Score	Support	Support Proportion	Accuracy
Maize streak virus	86.67	75.09	471	0.17	0.93
Grey leaf spot	66.56	71.90	527	0.18	0.89
Gibberella ear rot	67.26	72.27	584	0.20	0.88
Maize Dwarf mosaic virus	72.08	69.38	498	0.17	0.90
Northern corn leaf blight	76.29	74.32	773	0.27	0.86
Macro Average	73.77	72.59			

Weighted Average	73.62	72.72			
Micro Average	72.66	72.66			

In table 1, the confusion matrix of the model has been shown. The model was able to detect Maize streak virus with highest precision out of all the diseases. The average precision of the model is 73.72. Out of 360 images of Maize streak virus the modal was able to detect 312 images correctly which is 86.67% precision which is the best precision. From 619 images of Grey leaf spot class, the model was able to detect only 412 images correctly which lead to the lowest precision of 66.56% attained by the system. The range of the precision for this confusion matrix is 66.56% to 88.67%. There are a total of 8 layers in this model out of which 7 are for feature extraction and 1 for classification.

TABLE II. META ANALYSIS

Classes	True Positive	False Positive	False Negative	True Negative
Maize streak virus	312	48	159	2334
Grey leaf spot	412	207	115	2119
Gibberella ear rot	456	222	128	2047
Maize Dwarf mosaic virus	333	129	165	2226
Northern corn leaf blight	560	174	213	1906
Sum	2073	780	780	3633.00

In table 2, the meta-analysis of the output of the model has been done. The overall True Positive value attained by the model is 2073 in which Northern corn leaf blight showed the highest True Positive value of 560. The average False Positive value of the 5 classes is 156 out of which the 207 is the highest value of Grey leaf spot. The lowest False Negative value attained by the model is 115 of the Grey leaf spot. The average value of the True Negative is 726.6 attained by the model and the lowest value is of Northern corn leaf blight which is 1906.

TABLE III. PARAMETER SECTION

Actual/ Predicated	Maize streak virus	Grey leaf spot	Gibberella ear rot	Maize Dwarf mosaic virus	Northern corn leaf blight
Maize streak virus	312	51	44	24	40
Grey leaf spot	21	412	34	12	48
Gibberella ear rot	2	78	456	3	45
Maize Dwarf mosaic virus	12	46	.066	333	41
Northern corn leaf blight	13	32	78	90	560
Truth Overall	360	619	678	462	734
Precision in %	86.67	66.56	67.26	72.08	76.29

In table 3, the result of the different performance matrices has been explained. The highest precision of 86.67 was achieved by Maize streak virus. Weighted average is an average resulting from the multiplication of each component by a factor reflecting its importance. The weighted average of the model was 73.62. Micro Average achieved by the system was 72.66. The range of the F1-Score is from 69.38 of Maize Dwarf mosaic virus class to 75.09 of Maize streak virus. The macro average of the F1-Score is 72.59. The highest value of support is for Northern corn leaf Blight which is 773. The range of support is 471 to 773 in which Maize streak virus class has the lowest support. The highest support proportion is 0.27 which is attained by Northern corn leaf blight and the lowest support proportion is 0.17 which is of Maize streak virus class and Maize Dwarf mosaic virus class. The range of accuracy of the 6 classes is 0.86 to 0.93.

TABLE IV. CONVOLUTION LAYER

Layer	Type	Number of Filters	Filter Size	Stride	Pooling Size	Activation	Parameters (Approximate)
1	**Convolutional**	64	3x3	2	-	ReLU	38,464
2	**Convolutional**	128	3x3	2	-	ReLU	73,856
3	**Convolutional**	256	3x3	2	-	ReLU	295,168
4	**Max Pooling**	-	2x2	2	-	-	0
5	**Convolutional**	512	3x3	2	-	ReLU	1,180,160
6	**Convolutional**	512	3x3	2	-	ReLU	2,360,320
7	**Convolutional**	512	3x3	2	-	ReLU	2,360,320
8	**Max Pooling**	-	2x2	2	-	-	0
9	**Flatten**	-	-	-	-	-	0
10	**RF Layer**	5 (output classes)	-	-	-	-	-

In table 4, convolutional layers of the trained model uses Convolutional neural network for feature extraction and Random Forest Classifier for classification of the images is observed. There are a total of 8 layers out of which 7 are for feature extraction and 1 is for image classification. This model uses a 3x3 stride method and ReLU activation function for the convolutional layer. We have 64 filters in first convolutional layer then 128 and 256 filters are there in the next convolutional layers. Then there is a max pooling layer which has 2x2 filter size and after that we have another three convolutional layers with 512 filters. The Third last layer of the convolutional layer is Max Pooling and then we have a flattern layer which converts it into 1-dimensional array and lastly Random Forest classifier classifies it into 5 diseases. The total CNN parameters for this model were 6,306,768 approximate.

V. CONCLUSION

In this research, we implemented the combined solution of Random Forest Classifier with Convolutional Neural Networks (CNNs) in order to solve one of the significant problems Is Maize Illnesses. Using machine learning and deep learning techniques, we developed an accurate model for the identification of maize diseases. The research focused on diseases that have a significant impact on corn production and agricultural productivity, including the Maize Dwarf Mosaic Virus, Gibberella Ear Rot, Grey Leaf Spot as well as Northern Corn Leaf Blight.

We ensured that the input of our model was a rich and diverse one, by going through to carefully gathering data, and preprocess it some more until we were able to extract its features. Having described detailed implementation, we showed that CNNs are very effective at extracting salient features from images and then helping to classify them using the Random Forest technique. From the results obtained, it is evident that the model has the potential in accurately classifying different maize diseases.

The outcome of the confusion matrix, meta-analysis, and parameter sections shows that the model is accurate and effective. The model that stood out quite remarkably labelling the Maize Streak Virus possesses the potential to advance impact in real-life agricultural uses. The precision percentages (66.56% to 86.67%) indicate that the proposed strategy can be applied to several disorders, similar in nature or distinct from each other.

Firstly, the detailed analysis of the convolutional layers also came to reveal hidden insights related to both complicated design and feature extraction through various filters performed by each layer.CNNs and Random Forest Classifiers worked together to provide a potent tool for illness diagnosis that guaranteed accuracy and computational economy.

This study provides a reliable analytical tool for the early and accurate interception of maize diseases, something that would preferably revolutionize crop health management. Food security can be enhanced through modern technology in agriculture, which will also help to minimize financial losses and allow farmers to derived cohesive crop management techniques. In the area of crop health, this research contributes to a growing new set of deep learning applications in ag-tech. Possible follow-up research attempts could include scalability and versatility test of the model among various agroclimatic areas, as well as crop types thus allowing for better applicability in artificial agricultural systems. 7 Information Technologies-Based Decision Support Systems In order to develop information technologies-based decision support tools one will be required to create appropriate algorithms for these approaches. An example is an upgrade of previous expert rules performing multi-c

In summary, this research is important in the field of crop health and even more so in the case of agricultural technology. Finally, the target model has a lot of practical potential purposes allowing farmers identifying and curing diseases timely. Thus, if we make use of the cutting-edge technology such as deep learning and machine learning technologies to agricultural sector, it will increase crop output and propel food security. Additionally, incorporation of these new farming methodologies is essential since they may continue maintaining economies that are dependent on agriculture. The topic for more study includes scalability aspect when it comes developing a scalable lawn disease detection model along with growing crops kind's adaptation.

References

[1] Bosque-Perez, N. and Olojede, S. and Buddenhagen, and IW, "Effect of maize streak virus disease on the growth and yield of maize as influenced by varietal resistance levels and plant stage at time of challenge," in Euphytica, 1998, pp. 307--317.

[2] S. Vanlalhruaia, Chakraborty, S. and Mahapatra., "Assessment of yield loss and avoidable yield loss due to southern leaf blight of maize and development of yield loss prediction model.," 2023, pp. 1--9.

[3] B. Talba, U., Channya, K., F., Hahunnaro, H., Zakari, & G., "Survey On Incidence and Severity of Ear Rot Disease of Maize in Southern Borno State, Nigeria.," 2023, pp. 1–9.

[4] B. Madias, A., Borrás, L., Gambin, & L., "Foliar fungicides help maize farmers reduce yield gaps in late sown crops in a temperate region.," 2023, p. 126768.

[5] & F. Mafukidze, D., H., Owomugisha, G., Otim, D., Nechibvute, A., Nyamhere, C., Mazunga, "Adaptive thresholding of cnn features for maize leaf disease classification and severity estimation.," 2022, p. 8412.

[6] T. Nongmaithem, N., Sanjenbam, D., Konsam, J., Singh, K., L. N., Devi, & R., "A report survey and surveillance of maize diseases in Manipur.," 2022, pp. 557–560.

[7] A. Dinolfo, I., M., Martínez, M., Castañares, E., Arata, & F., "Fusarium in maize during harvest and storage: a review of species involved, mycotoxins, and management strategies to reduce contamination," 2022, pp. 151--166.

[8] & G. Njeru, F., Mwaura, S., Kusolwa, M., P., Misinzo, "Maize production systems, farmers' perception and current status of maize lethal necrosis in selected counties in Kenya.," 2022, pp. 692--705.

[9] S. Mehta, V. Kukreja, and A. Gupta, "Revolutionizing Maize Disease Management with Federated Learning CNNs: A Decentralized and Privacy-Sensitive Approach," in 2023 4th International Conference for Emerging Technology, INCET 2023, 2023. doi: 10.1109/INCET57972.2023.10170499.

[10] P. Bachhal, V. Kukreja, and S. Ahuja, "Maize Leaf Diseases Classification using a Deep Learning Algorithm," in 2023 4th International Conference for Emerging Technology, INCET 2023, 2023. doi: 10.1109/INCET57972.2023.10170182.

[11] P. Bachhal, V. Kukreja, and S. Ahuja, "Maize Disease classification using Deep Learning Techniques: A Review," in 2023 International Conference on Advancement in Computation and Computer Technologies, InCACCT 2023, 2023, pp. 259–264. doi: 10.1109/InCACCT57535.2023.10141847.

[12] Exconde, O. R. (1983). Virus disease of maize in the Philippines. In: Proc. Int. Maize Virus Disease Cotloqu. Workshop, 2-6 Aug. 1982, Ohio State Univ. (Ed. by D. T. Gordon, J. K. Knoke, L. R. Nault and R. M. Ritter) pp. 203-205, Ohio Agricultural Research and Development Center. Wooster, Ohio.

[13] Manandhar, G., Ferrara, G.O., Tiwari, T.P., Baidya, S., Bajracharya, A.S.R., Khadgee, B.R., & Narro, L. (2011). Response of maize genotypes to gray leaf spot disease (Cercospora zeae-maydis) in the hills of Nepal. Agronomy Journal of Nepal, 2, 93-101.

[14] Payak, M. M., & Renfro, B.L. (1974). A decade of research on maize diseases: impact on production and its international cooperative outreach. In: Current Trends ill Plant Pathology (Ed. by S. P. Rychaudhuri and J. P. Verma) pp. 160--170, Department of Botany, Lucknow University. India.