INNOVATING AGRICULTURE THROUGH EARLY DISEASE IDENTIFICATION IN LEAVES USING CNN

1st Y. Yekavarnika
dept of COMPUTER SCIENCE and ENGINEERING
SRI VASAVI ENGINEERING COLLEGE
Tadepalligudem, INDIA
20A81A0564@sves.org.in

2nd G. Nataraj, MTech
dept of COMPUTER SCIENCE and ENGINEERING
SRI VASAVI ENGINEERING COLLEGE
Tadepalligudem, INDIA
nataraj.gudapaty@srivasaviengg.ac.in

3rd K. Uma Sai Satya Sri
dept of COMPUTER SCIENCE and ENGINEERING
SRI VASAVI ENGINEERING COLLEGE
Tadepalligudem, INDIA
20A81A026@sves.org.in

4th Durga Abhinay
dept of COMPUTER SCIENCE and ENGINEERING
SRI VASAVI ENGINEERING COLLEGE
Tadepalligudem, INDIA
20A81A0535@sves.org.in

5th P. Suvarna Raju
dept of COMPUTER SCIENCE and ENGINEERING
SRI VASAVI ENGINEERING COLLEGE
Tadepalligudem, INDIA
20A81A048@sves.org.in

6th B. Jyothika Sai Anupama
dept of COMPUTER SCIENCE and ENGINEERING
SRI VASAVI ENGINEERING COLLEGE
Tadepalligudem, INDIA
20A81A0506@sves.org.in

7th I. Satya Anil Kumar
dept of COMPUTER SCIENCE and ENGINEERING
SRI VASAVI ENGINEERING COLLEGE
Tadepalligudem, INDIA
21A85A0506@sves.org.in

***Abstract*—Agriculture is the prime occupation sectors of the nation. Adapting technology plus digitalization is very crucial for the realm of agriculture to benefit the farmer and consumers parallel. Utilizing modern technology for adapting to and consistently monitoring healthcare aids in the early detection of diseases, offering an alternative interpretation. To tackle the timely identification of diseases affecting plants, our approach focuses on examining leaf health. Because the leaf's primary function is to produce nutrients through photosynthesis and support the plant's growth. Leaf diseases caused by bacteria or other pathogens can negatively impact agricultural yields. Timely detection and precise identification of these illnesses are essential for mitigating extensive crop devastation and financial setbacks. Our online platform offers a versatile solution, adept at identifying leaf diseases across various crops such as potato, tomato, wheat, and corn, catering to a wide range of agricultural settings. The emergence of leaf diseases has prompted the innovation of computational approaches utilizing ConvNets to effectively diagnose and manage plant health issues which attains an accuracy of above 95 %. CNN provides a more convenient approach to image classification and object recognition. It uses leveraging principles of linear algebra and matrix multiplication for the identification of patterns within an image. In this, the ConvNets model is trained to recognize patterns and features associated with specific diseases, enabling accurate and early identification.**

***Keywords*—Leaf Diseases, Image Processing, Deep Learning Techniques, Features, Patterns, Convolutional Neural Network(CNN) algorithm, Accuracy, Detection.**

I. INTRODUCTION

A crucial segment of the Indian economy is agriculture, employing nearly 50 % of the country's workforce. India is the world's largest producer of pulses, rice, wheat, spices, and spice products. The economic well-being of farmers relies heavily on the caliber of their produce, reliant on optimal plant growth and yield. Detecting plant diseases is crucial for maintaining crop health and maximizing agricultural productivity. Plants are prone to illnesses that impede their growth and disrupt the delicate balance of the farming environment. Utilizing automatic disease detection techniques proves advantageous in identifying plant diseases at their nascent stages. Manifestations of diseases affecting plants are evident in various parts such as leaves. Manual detection of these diseases through leaf images is laborious, necessitating the development of computational methods to automate the procedure for identifying and categorizing diseases involving a systematic approach to detection and classification.

In agriculture, timely identification and precise assessment of diseases affecting plants stand as crucial factors, pivotal for safeguarding food security and promoting sustainable cultivation practices. Conventional methods relying on human expertise are time-consuming, labor-intensive, and prone to errors. Leveraging advancements in Machine Learning and Deep Learning, automated disease detection system utilizing

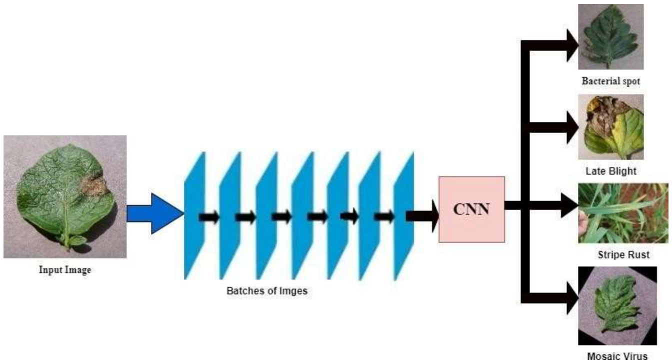

Fig. 1. Diagrammatic representation of all steps

ConvNets (CNN). I t has showcased exceptional prowess across a wide array of image recognition assignments, including medical imaging and object recognition, making them well-suited for the intricate task of identifying diseases of leaves derived from images of leaves..

The study presented in reference [7] introduces an innovative approach that employs 2D CNN to detect diseases in tomato and cotton plants by analyzing images of their leaves. The superiority of 2D CNN over current methods with reference to localized feature extraction, efficiency, robustness, adaptability, and end-to-end learning capabilities positions it as the preferred choice for disease identification. These distinctive features elevate CNNs to the forefront of image analysis and understanding, making them the superior option for a spectrum of techniques in machine vision applications.

II. LITERATURE REVIEW

Sammy V. Militante, Bobby D. Gerado, and Nanette V. Dionsio [1] conducted an extensive investigation into the The identification of plant leaves and their recognition of diseases using DL techniques in their research. In their literature review, they explored a range of methodologies for detection, analyzing a diverse dataset of over 35,000 images that embrace both well and tainted plant leaves. The collected datasets, as highlighted in their article, are invaluable for our detection endeavors.

Prajwala TM, Alla Pranathi, Kandiraju Sai Ashirtha, Nagarathna B. Chittaragi, and Shashidhar G. Koolagudi [2] directed their attention to the tomato leaf disease exposure in particularly, utilizing ConvNets (CNN).In their literature review, they underscored the consequence of CNN in image classification, specifically highlighting its effectiveness in extracting attributes for accurately classifying input plant leaf images.

Amrita S. Tulshan and Natasha Raul [3] conducted a study focusing on plant leaf detection utilizing machine learning. Their emphasis was on the crucial role of image preprocessing, segmentation, and feature drawing out in the detection process.

Shyamtanu Bhowmik, Anjan Kumar Talukar, and K. Andarapumar Sarma [4] explored the identification of diseases in tea leaves employing ConvNets (CNN). Their exploration highlighted the utilization of ConvNets for automatic attribute extraction our implementation of detecting plant diseases as an advanced approach.

Kshyanaprava Panda Panigrahi, Abhaya Kumar Sahoo, and Himansu Das [5] introduced a ConvNets (CNN) approach for identifying corn leaves malady, specifically tailored to meet the requirements of digital agricultural systems. Given our incorporation of corn as a key target plant, their work facilitated the seamless integration of corn leaf detection through ConvNets (CNN) in our project.

Kawcher Ahmed, Tasmia Rahman Shaidhi, Syed Md. Irfanul Alam, and Sifat Momen [6] explored rice leaf maladies detection employing ML techniques. Their study illuminated the varieties of malady in rice crops and the corresponding techniques used. Drawing from their insights, we integrated pertinent methodologies into our project for the resourceful recognition of diseases of the leaf of rice through ConvNets (CNN).

Eftekhar Hossain, Md. Farhad Hossain, and Mohammad Anisur Rahaman [8] presented a color and texture-based intended for the discovery and categorization of leaf-related diseases using a KNN classifier. Their work guided our consideration of crucial techniques and methodologies, especially focusing on color and texture-based approaches.

Assad Souleyman Doutoum and Bluent Tugrul [9] provided a ample review of leaf syndrome identification and categorization through DL techniques. Their overview offered a foundational understanding of diseases affecting plant leaves, aiding our project's implementation by providing insights intodisease types, effects, and detection methods.

Sherly Pushpa Annabel, T. Annapoorani, and P. Deepalaklakshmi [10] presented an analysis on ML leaf syndrome identification and categorization. This review high-lighted the diverse methodologies in ML required for identifying diseases in different plant leaves, offering precious supervision for our implementation.

Amandeep Singh Manider lal Singh [11] contributed in-sights on programmed Blast disease recognition from paddy leaf using a color-slicing approach. Their spotlight on diseasesin paddy fields, including the inception of the Leaf Color Chart (LCC), influenced our understanding of approaches based on color for disease recognition in paddy plants.

III. PROPOSED METHODOLOGY

This portion provides further details on the exact methods employed in the project to predict diseases in leaves. It also presents a visual representation of the project's workflow in Figure 2.

A. Datasets

In our project, we have in use datasets in the Kaggle website. The four datasets that we serene Potato, Tomato, Corn, and Wheat. Potato has 3 classes (Early_Blight, Late_Blight, Healthy) with 2152 images. Tomato has 10 classes (Bacterial_Spot, Early_Blight, Late_Blight, Septoria_Leaf Spot, Leaf mould, Spider_Mites, Target_Spot, Tomato Mosaic_Virus, Yellow Leaf_Curl_Virus, Healthy) with 22930 images. Wheathas 3 classes (Septoria, Stripe rust, Healthy) with 407 images. Corn has 4 classes (Blight, Common rust, Grey leaf spot,

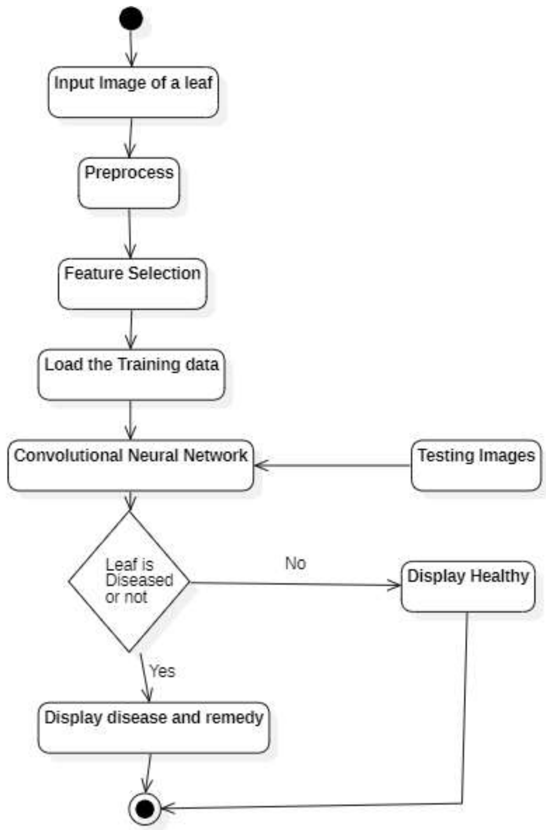

Fig. 2. Workflow of Project

Healthy) with 4188 images. The following Table 1contains a clear explanation of the datasets.

Name of the leaf dataset	Number of classes	Name of the diseases	Number of Images
Potato	3 Classes	• Early blight • Late blight	2152
Corn	4 Classes	• Blight • Common rust • Grey leaf spot	4188
Tomato	10 Classes	• Bacterial spot • Early blight • Late blight • Septoria leaf spot • Leaf mould • Spider mites • Target spot • Tomato mosaic virus • Yellow leaf curl virus	22930
Wheat	3 Classes	• Septoria • Stripe rust	407

Table. 1. Dataset Features

B. Data Preprocessing

Data preprocessing is a crucial step involving multiple procedures aimed at optimizing images for neural network models, particularly for tasks like classification, detection, or segmentation. Leveraging tools such as the Image Data Generator, found in libraries like TensorFlow's Keras API, facilitates seamless and efficient preprocessing and intensification of picture data, suspending the requirement for labor-intensive. This proves particularly advantageous in deep learning applications, where data augmentation acting a central role in enhancing model performance by introducing variability into the training dataset. This, in turn, mitigates overfitting concerns and bolsters generalization.

When employing an Image Data Generator for dataset preprocessing, the initial step involves specifying parameters tailored to the task. These parameters encompass actions such as rescaling pixel values (e.g., dividing by 255 for a pixel value range of 0 to 1), image rotation, horizontal or vertical shifting, zooming in or out, and horizontal or vertical flipping, among others. The principal intention of these transformations is to generate a more diverse set of training examples from the original dataset. By simulating various perspectives, lighting conditions, and sizes, these transformations enable the Model to adapt to the complexities it may encounter in real-world scenarios.

Fig. 3. Visualization of one of the batches of a corn leaf dataset

C. Prediction

A ConvNets (CNN) emerges as a potent deep learning algorithm, meticulously crafted for tasks centered around image acknowledgment and dispensing. Comprising compound mantles, as well as convolutional mantles, pooling mantles, and fully connected mantles, CNNs excel in discerning intricate patterns within images.

1. *Convolutional Layers::* The model incorporates several convolutional layers, each followed by a max-pooling layer. These layers adeptly learn to haul out attributes from inputted dataset, operating at various levels of abstraction.

- Within the convolutional layer, we enhance model performance by employing diverse filter sizes, either within the same layer or across different layers, to capture features at different scales. Accumulating more convolutional mantles can enable the association connection to grasp more complex patterns, though this must be balanced against the risks of overfitting and increased computational cost.

2. *Pooling Layer ::* Pooling layers play a key role in reducing the dimensions of feature maps, effectively decreasing the amount of specifications to learn and the computational load on the network.

- This layer focuses on summarizing features within a region of the feature map, enabling subsequent operations to work on summarized features rather than precisely positioned ones. This enhances the model's resilience to variations in feature positions within the input image.

3. *Flatten Layer ::* The Flatten layer transforms the output of the preceding convolutional mantle into a one-dimensional vector, preparing it for input into the dense layers.

- In the perspective of a convolutional mantle, thislayer focus on converting the multi-dimensional output into a one- dimensional array, facilitating its amalgamation into the fully associated mantles, well- known as dense mantles.

During training, the CNN undergoes learning to recognize indicative patterns and features in images, signaling the presence of diseases. Once trained, the CNN acts as a predictive tool for identifying whether a new leaf image is healthy or diseased. Bypassing the image through the network, the CNN analyzes it and produces a prediction, indicating the likelihood of disease presence. This predictive capability proves invaluable for farmers and agricultural professionals, enabling early disease detection and management. Ultimately, this technology aids in preventing crop losses and enhancing overall plant health.

4. Evaluation

In a conducted study [12], a predictive model was developed by relying solely on the proportion of each color channel (Red, Green, and Blue) hauled out from the RGB represented values from the affected area of rice leaves through image processing. This model aimed to classify diseases by feeding these percentages values into a Naïve_Bayes classifier, resulting in the categorization of diseases into classes: Bacterial_Leaf_Blight, Rice_Blast, and Brown_Spot. Notably, the model exhibited an accuracy surpassing 89 %.

Our proposed methodology represents a significant advancement in accuracy when compared to baseline models. Through thorough experimentation and meticulous fine-tuning, we have achieved a commendable accuracy rate of 97 % in identifying leaf diseases across various plant species. Ad- ditionally, we conducted comparative analyses with existing methodologies to corroborate the effectiveness and superiority of our approach.

IV. RESULT AND ANALYSIS

• This section elucidates the outcomes of our model. Initially, we partitioned the dataset into training data (80 %),test data (10 %), and validation data (10 %).

Figure 4 portrays the training and validation precision performance over epochs for our machine learning model. The x-axis signifies the amount of epochs, representing training sessions or iterations, while the y-axis depicts accuracy as a percentage – the fraction of accurate predictions made bythe model.

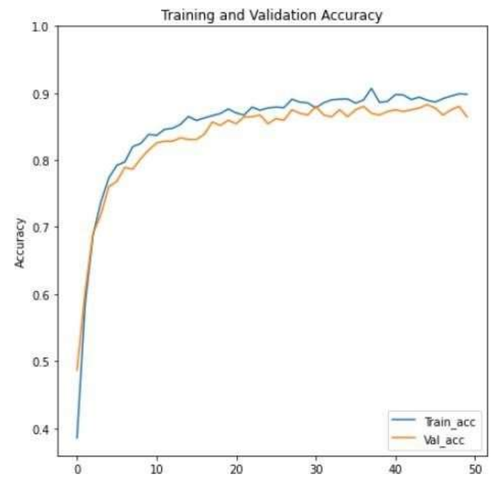

Fig. 4. Accuracy VS Epochs

- • The blue line illustrates the training accuracy, reflecting the model's proficiency on the data it is skilled on.
- The orange line represents the validation accuracy, signifying the predictive model's achievement on a distinct dataset it has not encountered during training.

This evaluation is crucial for assessing the model's generalization to unseen data and preventing overfitting on training dataset. Key observations from Figure 4 include:

- The training accuracy generally ascends with an increasing number of epochs, indicating the model's learning from the training data.
- • The validation accuracy initially rises but may plateau or slightly decrease, hinting at potential overfitting to the training data after a certain epoch count.
- . Determining the optimal number of epochs relies on factors like model's complexity, the size and quality of the training statistics and the specific problem addressed.
- It is generally cautioned against fixing the amount of epochs arbitrarily, without assessing the model's performance on a validation set. Continuous evaluation of unseen data is imperative for robust model training.
  The confusion matrix shows the amount of actual corn leaf diseases (blight, Common_Rust, Gray_Leaf_Spot, and healthy) in comparison to the model's predictions. In the confusion matrix, the rows represent the actual standards and the columns represent the predicted values. For example, the top row (Blight) shows that out of 120 actual blight cases, the model correctly predicted 108 blight cases.

  In our project, the ConvNets (CNN) model we developed with accuracy for potato is 98%, tomato is 94%, wheat is 99 %, and corn is 95%. Our model outperformed the accuracy levels reported in existing research utilizing traditional algorithms. Previous research has been relied on. methods such as SVM, decision trees, KNN, and logistic regression, our

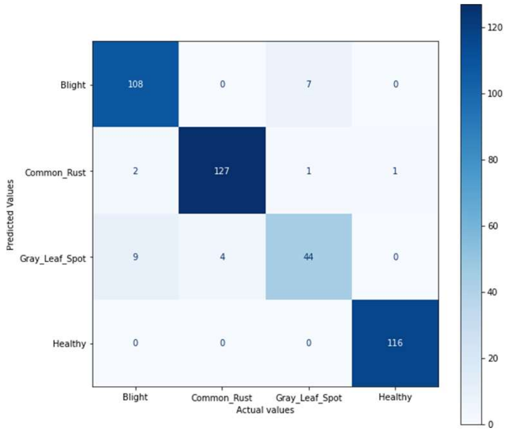

Fig. 5. confusion matrix visualization of a corn leaf disease classificationmodel's performance.

CNN model demonstrates superior capability in identifying and classifying leaf diseases. This advance- ment is attributed to CNN's ability to autonomously learn intricate features from leaf images through deep learning structural design, especially adept at handling visual data. Unlike conventional algorithms that might struggle with the complexity and variability of leaf patterns and disease symptoms, our CNN model excels by extracting and learning from these nuanced details, leading to more accurate diagnosis. This significant improvement underscores the potential of CNNs in revolutionizing the pasture of agricultural disease detection, offering a more reliable and efficient approach to safe- guarding crop health.

Name of the Leaf	Accuracy
Potato	98%
Tomato	94%
Wheat	99%
Corn	95%

Table. 2. Accuracy of leaves

V. CONCLUSION

Plant diseases represent a substantial threat to agricultural productivity, often resulting in significant losses if not promptly addressed. The adoption of leaf detection technology offers a proactive solution, enabling of disease identification of diseases at their early stages and mitigating potential damages. Our approach is grounded in image dispensing techniques, informed by comprehensive research and security of diverse plant diseases and leaf variations.

This paper introduces a narrow approach, specifically employing a ConvNets (CNN), for the recognition of distinct leaf diseases: potato, tomato, wheat, and corn. The identified diseases encompass Early_Blight, Late_Blight, septoria, Mosaic_Virus, yellow_Leaf_Curl_Virus, Grey_Leaf Spot, and Spider mites, among others. Our objective is to democratize disease detection in agriculture by developing a user-friendly web application accessible to a wide audience. The simplicity and efficacy of our design empower users, regardless of expertise, to accurately diagnose plant diseases. By facilitating easy access to disease detection capabilities, our application contributes to the progress of agriculture and supports the improvement of a healthier and more sustainable society.

VI. REFERENCES

[1] Militante, Sammy V., Bobby D. Gerardo, and Nanette V. Dionisio. ”Plant leaf identification and disease recognition using deep learning.” In 2019 IEEE Eurasia conference on IOT, communication, and engineering (ECICE), pp. 579-582. IEEE, 2019.

[2] Kaushik, M., P. Prakash, R. Ajay, and S. Veni. ”Tomato leaf disease recognition using a ConvNets (CNN) with data augmentation.” In 2020 5th International Conference on Communication and Electronics Systems(ICCES), pp. 1125-1132. IEEE, 2020.

[3] Applalanaidu, Majji V., and G. Kumaravelan. ”A review of ML approaches in plant leaf diseasedetection and classification.” In 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), pp. 716-724.IEEE, 2021.

[4] Bhowmik, Shyamtanu, Anjan Kumar Talukdar, and Kan- darpa Kumar Sarma. ”Detection of disease in tea leaves using ConvNets (CNN).” In 2020 Advanced Communication Technologies and Signal Processing (ACTS), pp. 1-6. IEEE, 2020.

[5] Panigrahi, Kshyanaprava Panda, Abhaya Kumar Sahoo, and Himansu Das. ”A CNN approach for corn leaves disease recognition to sustain digital agricultural system.” In 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), pp. 678- 683. IEEE, 2020.

[6] Ahmed, Kawcher, Tasmia Rahman Shahidi, Syed Md Irfanul Alam, and Sifat Momen. ”Rice leaf disease recognition using ML techniques. ” In 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), pp. 1-5. IEEE, 2019.

[7] Peyal, Hasibul Islam, Md Nahiduzzaman, Md Abu Hanif Pramanik, Md Khalid Syfullah, Saleh Mohammed Shahriar, Abida Sultana, Mominul Ahsan, Julfikar Haider, Amith Khandakar, and Muhammad EH Chowd- hury. ”Plant Disease Classifier: Detection of Dual-Crop Diseases using Lightweight 2D CNN Architecture.” IEEE Access(2023).

[8] Padilla, Dionis A., Ramon Alfredo I. Pajes, and Jerome t. De Guzman. “Detection of corn leaf diseases using a convolutional neural network with openMP implementation.” In 2020 IEEE 12th International Conference on Humanoid Nanotechnology Information Technology Communication and Control Environment, and Management(HNICEM), PP. 1-6. IEEE,2020.

[9] Doutoum, Assad S., and Bulent Tugrul. “A Review of Leaf Diseases Detection and Classification by Deep Learning.” IEEE Access (2023).

[10] Annabel, L. Sherly Puspha, T. Annapoorani, and P. Deepalakshmi. ”Machine learning for plant leaf dis- ease detection and classification–a review.” In 2019 International Conference on Communication and Signal Processing (ICCSP), pp. 0538-0542. IEEE, 2019.

[11] Singh, Amandeep, and Maninder Lal Singh. “Auomated blast disease identification from paddy plant leaf—A color slicing approach.” In 2018 7th International Conference on Industrial Technology and Management (ICITM), pp. 339-344. IEEE, 2018.

[12] T. Islam, M. Sah, S. Baral, and R. RoyChoudhury, “A faster technique on rice disease detection using image processing of pretentious area in agro-field,” in 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 62–66, IEEE, 2018.