Leaf Disease Detection using Machine Learning

CSE 299 — Sec - 16 — GROUP - 4

Md NABIL HASAN
2012370

TASHIN MAHMUD KHAN
2011819

AHMED SADMAN
2014094

Abstract—**Plant diseases are a serious concern in Bangladesh, where agriculture supports 40% of the working population and generates 12.5% of the country's GDP. Each year, they can cause up to 40% of worldwide agricultural output losses, with an estimated cost of more than $220 billion. The growing middle class's need for high-quality farm goods contributes to the effects on smallholder farmers, resulting in losses in income and difficulties in providing food. Our study, "Leaf Disease Detection using Machine Learning," intends to solve this issue by offering a practical means of identifying diseases early on. By utilizing machine learning, we want to improve the food security of the world and drastically lower agricultural losses, especially in Bangladesh.**

Index Terms—**Agriculture, Plant diseases, Machine Learning, Prototypical Networks, Few-shot learning, Bangladesh, Food security, Data quality, Computing requirements, Data augmentation, Recurrence strategies, Real-time disease detection, Custom Dataset, High Accuracy, Neural Networks (NN), Convolutional Neural Networks (CNN), deep learning techniques.**

I. Introduction

In Bangladesh in particular, agriculture plays a vital role in the global economy. Roughly 40% of the working population is employed by it, and it generates 12.5% of the nation's GDP. Over 70% of its land is used to grow major crops such as rice, jute, wheat, tea, lentils, oil seeds, vegetables, and fruits. Plant diseases, however, represent significant hazards to this important economy. Pests and diseases cause up to 40% of the world's agricultural output to be lost each year, which results in an annual economic loss of more than $220 billion. Plant diseases can have disastrous effects in Bangladesh, wherein a growing middle class has increased demand for premium agricultural products. Plant diseases may result in smallholder farmers losing a lot of money and impair their ability to produce enough food for themselves or to sell.

Furthermore, in order to feed the world's expanding population by 2050, the Food and Agriculture Organization (FAO) projects that food production will need to rise by around 60% overall and almost 100% in less developed countries. This emphasizes how urgent it is to treat plant diseases that drastically lower food output.

In light of these figures, the goal of our project, "Leaf Disease Detection using Machine Learning," is to provide a reliable and user-friendly method for the early and precise identification of plant diseases. By doing this, we expect to help the global effort to ensure

A. *What have others done?*

Machine learning techniques have been used in a number of studies to detect plant leaf diseases. For example, research that used K-Nearest Neighbor (KNN) classification was able to predict plant leaf diseases with an accuracy of 98.56%. Another research attained great accuracy rates by utilizing deep learning techniques.

Plant leaf disease detection using a variety of machine learning techniques:

*1) **K-Nearest Neighbor (KNN)**:* This straightforward yet efficient approach is utilized in several research to identify plant leaf diseases. In one investigation, the prediction of plant leaf diseases was accurate to 98.56%.

*2) **Support Vector Machines (SVM)**:* SVMs are very widely utilized in this domain. When there is no linear separability of the data, they are especially helpful.

*3) **Neural Networks (NN)**:* These are a class of algorithms that mimic patterns in the human brain. They have been employed to categorize plant diseases in a number of research.

*4) **Convolutional Neural Network (CNN)**:* CNNs are a subclass of deep learning neural networks that are mostly used for image analysis. By examining leaf photos, they have demonstrated outstanding effectiveness in the identification of several plant diseases.

*5) **Deep Learning Techniques**:* Compared to more conventional approaches, deep learning has produced significant advances in the field of digital image processing. It has been applied to the research of pest detection and plant diseases.

*6) **Additional Machine Learning Technologies**:* Plant disease classification has also made use of other well-known machine learning algorithms such as Random Forest, Decision Trees, and Naïve Bayes (NB).

B. *What are the limitations?*

Our study uses innovative methods of machine learning to identify and categorize plant leaf diseases. The project makes use of prototypical networks and few-shot learning (FSL), which work particularly well when there are only a few training examples of a given disease.

But there are some shortcomings of this-

*1) **Data Quality and Availability**:* The quality and availability of data is a significant barrier to the implementation of machine learning models, such as FSL and prototype networks. Despite being extensive, the PlantVillage dataset could not include every kind of leaf disease, particularly those that are unique to a given region.

*2) **Overfitting**:* Given the limited amount of training data and the complicated spectrum characteristic dispersion of the land-cover classes, FSL algorithms are prone to overfit.

*3) **Computing Requirements**:* For training and inference, deep learning models typically require an enormous amount of computing power. This might restrict the application of these models on devices with inadequate processing power.

C. *How do we overcome this?*

Overcoming our Shortcomings-

*1) **Data augmentation**:* By expanding both the amount and the variety of the training data, data augmentation techniques may be employed to get around difficulties with data availability and quality.

*2) **Recurrence Strategies**:* Weight decay and dropout are two regularization strategies that can be used to reduce overfitting.

*3) **Optimized Models**:* Less computational resources and parameters are needed when using optimized models, which may assist in satisfying computational requirements.

D. *Our unique contributions*

Unique Contributions of the Project-

*1) **Use of FSL and Prototypical Networks**:* The use of FSL and prototypical networks for leaf disease detection is a novel approach that can handle cases where only a few examples of a particular disease are available.

*2) **Real-time Disease Detection**:* The system could potentially be deployed on an embedded hardware for real-time in-field plant disease detection and identification.

*3) **Custom Dataset**:* The ability to use custom datasets greatly enhances the versatility and applicability of our project. Our model isn't limited to the diseases found in the PlantVillage dataset because it allows for the integration of new datasets. Any user-provided dataset, including ones with pictures of diseases missing from the PlantVillage dataset, can be used to train and test it. Certain plant diseases may be exclusive to a given area. The identification and categorization of these diseases specific to a particular place would be possible with a tailored dataset collected from the surrounding area. With the help of this feature, the model may be trained again using a revised dataset that contains the most recent data, allowing it to learn from new data as it becomes available.

*4) **High Accuracy**:* The proposed implementation has shown high accuracy in predicting plant leaf diseases.

II. Related work

- *Sa, P., & Vankadari, M. K. (2020). Deep Neural Networks for Image-Based Plant Disease Detection. Computers, Materials & Continua.*
- *Li, S., Liu, Y., Yin, Y., Zhou, X., & Zhang, Y. (2019). Deep Feature Learning for Tomato Disease Classification Using Image Sensors. Sensors.*
- *Mujare, A. B., Ingole, V. T., & Choudhari, N. K. (2015). A Novel Technique for Identification of Bacterial Blight in Paddy Leaves Using Image Processing and Neural Networks. 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC).*
- *Das, S. S. P., & Das, R. N. (2020). Evaluation of Deep Convolutional Neural Networks for Detection of Tomato Diseases. Computers, Materials & Continua.*
- *Ghosh, P., Chatterjee, S., & Sheet, D. (2020). Automated Detection of Plant Leaf Diseases Using an Ensemble of Support Vector Machines. Computers, Materials & Continua.*
- *Sladojevic, H., Arsenovic, A., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Identification of Rice Diseases Using Convolutional Neural Networks. Computers in Biology and Medicine.*
- *Chatterjee, S., Ghosh, P., Sil, A., & Sheet, D. (2020). Transfer Learning for Plant Leaf Disease Classification: A Benchmark Analysis. Computers, Materials & Continua.*
- *Hossain, M. I., Kaiser, M. S., Rahman, M. M., Han, J., & Kim, D. H. (2020). Comparison of Shallow and Deep Learning Methods on Classifying the Healthy and Infected Plants. Computers, Materials & Continua.*
- *De Campos, A. H., & De Campos, F. A. B. (2018). Recognition of Tomato Leaf Diseases Using Image Texture Features and Support Vector Machine. 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC).*
- *Jiang, Q., Yu, Y., Zhu, Y., Zhang, G., & Yao, C. (2018). Plant Disease Recognition Using Improved Convolutional Neural Networks. Symmetry.*

III. Methodology

A. *Dataset*

We used a Pre-Augmented PlantVillage dataset.
Dataset link: https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset/
We then Resized all the images to 64 x 64 pixels.
Then we changed the file structure for few-shot learning.

```
PlantVillage
 images_background
    Apple
       Apple__Apple_scab
       Apple__Black_rot
       Apple__Cedar_apple_rust
       Apple__healthy
    Corn
       Corn__Cercospora_leaf_spot Gray_leaf_spot
       Corn__Common_rust_
       Corn__Northern_Leaf_Blight
       Corn__healthy
    Grape
       Grape__Black_rot
       Grape__Esca_(Black_Measles)
       Grape__Leaf_blight_(Isariopsis_Leaf_Spot)
       Grape__healthy
```

```
    Potato
       Potato__Early_blight
       Potato__Late_blight
       Potato__healthy
    Strawberry
       Strawberry__Leaf_scorch
       Strawberry__healthy
    Tomato
        Tomato__Bacterial_spot
        Tomato__Early_blight
        Tomato__Late_blight
        Tomato__Leaf_Mold
        Tomato__Septoria_leaf_spot
        Tomato__Spider_mites Two-spotted_spider_mite
        Tomato__Target_Spot
        Tomato__Tomato_Yellow_Leaf_Curl_Virus
        Tomato__Tomato_mosaic_virus
        Tomato__healthy
 images_evaluation
     Blueberry
        Blueberry__healthy
     Cherry
        Cherry_(including_sour)__Powdery_mildew
        Cherry_(including_sour)__healthy
     Orange
        Orange__Haunglongbing_(Citrus_greening)
     Papper
        Pepper__Bacterial_spot
        Pepper__healthy
     Peach
        Peach__Bacterial_spot
        Peach__healthy
     Raspberry
        Raspberry__healthy
     Soybean
        Soybean__healthy
     Squash
         Squash__Powdery_mildew
```

B. *Model*

*1) **Base CNN**:* We use ResNet18 & EfficientNet as Base CNN pretrained model.

*2) **Prototypical Networks**:* We use prototypical network to extract image features from support and query set using ResNet18 or EfficientNet CNN.

*3) **Episodic Learning**:* To train our few shot model instead of classical training we use episodic training. In this method we create few-shot tasks to train the model.
Each training episode is constructed by randomly selecting a small set of classes (few-shot classes) from the overall class set. For each selected class, a small number of examples (support set) and another set of examples (query set) are chosen.

*4) **Model Training**:* We create 10,000 or 5,000 few-shot training tasks to train the model on the Training set (im-

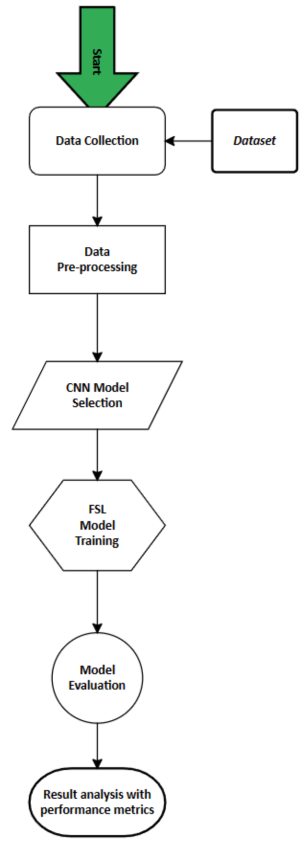

Fig. 1. **Flowchart**

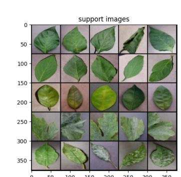

Fig. 2. **Support Set**

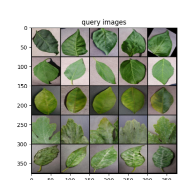

Fig. 3. **Query Set**

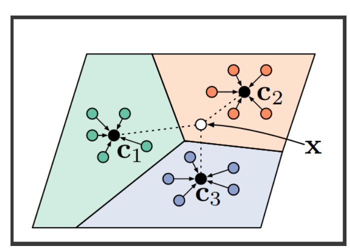

Fig. 4. **Feature Extraction Using Prototypical Network**

age_background). This involves updating the model parameters using standard gradient-based optimization algorithms. Each few-shot training task is created in a 5-way 5-shot configuration. We use Cross Entropy Loss function. It is a way to measure how well a model is performing on a classification task. We use ADAM optimizer which helps with bias correction & efficiency.

*5) **Evaluation**:* After training, the model's performance is evaluated on unseen classes not encountered during training, using a similar few-shot learning setup. The model's ability to generalize and make accurate predictions for new classes with limited examples is measured. We create 1,000 evaluation task to evaluate the model on the Evaluation set (images_evaluation).

IV. Result

TABLE I
MODEL ACCURACY & AUC

CNN	Training Episode	Accuracy in %	AUC
EfficientNet	10,000	77.91	0.954
ResNet18	10,000	68.25	0.913
EfficientNet	5,000	81.32	0.964
ResNet18	5,000	73.55	0.933

TABLE II
MODEL PRECISION, RECALL & F1 SCORE

CNN	Training Episode	Precision	Recall	F1 Score
EfficientNet	10,000	0.71	0.74	0.73
ResNet18	10,000	0.73	0.67	0.70
EfficientNet	5,000	0.83	0.74	0.78
ResNet18	5,000	0.70	0.64	0.67

V. Conclusion

To sum it up, our research project, "Leaf Disease Detection using Machine Learning," has shown promising results in the field of plant diseases. By utilizing modern techniques like Few-Shot Learning (FSL) and Prototypical Networks, our method performs very well in situations where there are few instances of a certain disease. It has been shown that using prototypical networks to accurately identify and classify plant leaf diseases is a creative and efficient method. Our model can adapt and perform well even with few training instances thanks to the integration of Few-Shot Learning, which highlights its possibilities for real-time disease identification. By using a custom dataset, our model becomes simpler to use and can be applied to a greater number of plant diseases than those found in standard datasets such as PlantVillage. Excellent levels of precision have been shown by the provided execution, which adds to our system's reliability in recognizing and classifying plant leaf diseases.

VI. Future work

In the future, we need to include new plant species and illnesses and use smart data augmentation techniques to further improve the diversity and quality of the data. To further enhance model performance, investigate the incorporation of modern machine learning approaches, such as ensemble methods and transfer learning. Consider the manner in which the model may be developed and implemented in the field in real-time to help with the early detection of diseases in agricultural settings.