Full Report on "Deep Learning for Plant Identification and Disease Classification from Leaf Images: Multi-prediction Approaches"

November 2, 2024

0.1 Introduction

The use of deep learning has transformed numerous fields, and it holds very high potential in agriculture, specifically for plant pathology. Traditional methods of identifying plant species and diagnosing diseases are often manual, requiring significant effort and time. With the development of Convolutional Neural Networks, automated image-based tasks have become feasible, particularly for the tasks like leaf disease detection and plant identification. Leaf images are ideal input data points since leaf characteristics, including colour, shape, and texture, are distinctive enough to identify both species and disease indicators. Despite the rapid development of CNNs, regular approaches typically focus on single-task models, addressing either species identification or disease classification, rather than combining both tasks into a single model.
The study hypothesizes that a multi-task CNN could perform better than single-task models by using shared information across tasks. The proposed Generalised Stacking Multi-output CNN (GSMo-CNN) is a multi-output model that can handle both species and disease classification in only one framework.

0.2 Problem Statement

There is a critical need for efficient and accurate tools in plant disease diagnostics. Most of the existing methods based on CNNs are limited to single-task classification, which ignores the semantic relationships between plant species and disease types. Plant species and their specific diseases often have complex interdependencies. For example, some diseases only appear in particular species. A model that can simultaneously identify species and classify diseases could theoretically achieve higher accuracy.

Objective: This paper develops a single CNN model that can carry out both tasks simultaneously: plant identification and disease classification. The hypothesis is that the multi-prediction model will be more effective and efficient than traditional single-task approaches. In pursuit of this goal, this study considers four different models of multi-task deep learning: multi-model, multi-label, multi-output, and multi-task, making a detailed comparison of each to identify the best-performing model.

0.3 Methodology: Step-by-Step Pipeline for the Multi-Prediction CNN Model

Thissubsectioncoversthe methodologythroughwhichthe studywasconductedwithstep-by-steptechnicaldescrip one canknowhow thedesignofpipelineidentifiesplant species andclassifiesplantdiseasefromtheimages.

0.3.1 Background on the Problem and Approach

Objective: The objective is to develop a deep learning model that can recognize the species of a plant and classify any diseases found using only an image of a plant leaf. The challenge is to combine these tasks into a single, efficient model that performs both tasks at once, making it a multi-prediction model.

Traditional Approach Limitations: Historically, the separated model was either used in identifying the plant species or classifying the disease. Combining both tasks may result in efficiency, and better accuracy as well as saving some computational complexity.
Solution: The researchers propose a Generalised Stacking Multi-output CNN (GSMo-CNN), a single model with separate outputs for plant species and disease, but that allows learning to be shared between these tasks.

0.3.2 Pipeline Overview

The GSMo-CNN model uses a pipeline that involves several key stages: data preparation, CNN architecture selection, multi-task learning, and final output generation. Here's a breakdown of each step:
Step 1: Data Preparation

0.3.3 Collection of Datasets

A set of images of leaves are collected from the dataset including Plant Village, Plant Leaves, and PlantDoc. In each of these images, information related to the species of plant is mentioned and, in case a disease affects the plant, information regarding the disease also accompanies.

0.3.4 Pre-processing the Images

The sizes of all the images are standardized into one dimension so that while being fed to the CNN, they become the same dimension, such as 256x256 pixels.
Images are standardized in color and intensity to avoid the model from getting biased by variations in lighting or inconsistencies in the background.

0.3.5 Data Partitioning

The dataset is divided into three segments
Training Set: 70% images, which are used for training the model.
Validation Set: 10% images, which are used for tuning model parameters.
Test Set: 20% images, which are used to evaluate final performance.

Step 2: CNN Backbones for Feature Extraction

In Deep Learning, the CNN backbones are the primary network structures that take essential features from images, such as patterns, textures, and shapes:

InceptionV3: useful in multi-scale feature extraction.
ResNet101: Known for handling deep layers without losing information.
MobileNetV2 and EfficientNet: Lightweight models suitable for mobile applications.

Each CNN backbone captures the process of an image into features, such as spots or veins on the leaf for distinction of species and disease.

0.4 Step 3: Multi-Prediction Methods

The research offers classification possible multi-task architectures which provide a better architecture of identifying species and diseases.

0.4.1 Multi-Model CNN:

Two CNNs, one used for species identification and another used for disease classification.
Cons: More computationally expensive since it runs through each image twice

0.4.2 Multi-Label CNN:

Combines species and disease into a single label, like “apple_scab" or “tomato_blight."
Downside: As species and diseases increase, the number of labels becomes too large, making the model difficult to manage.

0.4.3 Multi-Output CNN:

A single CNN with two separate outputs-one output for species, another for disease. The backbone layers will share the two tasks so that the model learns feature information that can benefit not only species identification but also disease classification.

0.4.4 Multi-Task CNN:

Proposed GSMo-CNN This is a preprint and has not been peer-reviewed.A GSMo-CNN is a type of Multi-Output CNN, yet the way that it stacks the prediction layers is completely different from any of the models mentioned above:
Local Predictions: A CNN predicts local species and disease on a temporary basis.
Stack Output Layers: These predictions are used as temporary values to be input back into the model and used to bias the model's final predictions.
GSMo-CNN Overcomes this Limitation: GSMo-CNN stacks the outputs so that each component of the output spaces shares species and disease information, providing higher accuracy and confidence in species and disease classification capabilities.

Step 4: Architecture of the GSMo-CNN Model
1. Feature Extraction Layers:
Convolution layers comprisethefrontendof GSMo-CNNwhichextractssalientinformationfromimagesintheformofs and texturesandareusedonalltasks.Theelements allow for derivingthe model'sinformationontheleaf shape,thevei andcolourdistribution**.**

2. Temporary Prediction Layers:
The modelhasdistinct layers,whichareused in generating∗temporary predictions∗ for each task. For instance, the temporary prediction layer may first indicate that the leaf belongs to an"apple"species affected by"rust"disease.

3.Stacked Output Layers:
The temporary predictionswillbeaddedtothefeature that has beenextracted;thus,theyare againfedinto the model.Forexample,itimprovesupon itspredictionsbecauseofbeing inthecontextofbeing in arelationship with othertasksorjobs.The modelforinstancemight changeitspredicteddiseasewhenspecific diseasesin that speciesare notin thefirst predictedspecies itwasable to find in the data.

4. Final Prediction Layers:
Speciesanddiseasearetheoutcomesofthestackedstructurefromwhichthefinalpredictionlayergetstherefinedresults.
Only theselastpredictions arecountedtoestimatetheperformanceof the model.
Step 5: Training the GSMo-CNN Model

0.4.5 Loss Functions:

At each level of prediction, a separate ∗cross-entropy loss function∗ measures the accuracy of prediction in each layer.
The model reduces these losses to increase accuracy at each step of the prediction pipeline.

0.4.6 Weight Tuning:

The model has four balancing weights that are used to adjust how much contributions of different layers are being made. This brings the model to decide how often should it rely on temporary predictions versus final predictions.

0.4.7 Optimization:

It runs several cycles with the update of various weights and minimizing loss function for optimal prediction ability employing accuracy and F1 scores as measures.

0.5 Step 6: Inference and Prediction

When a new image is fed into the GSMo-CNN model during inference (testing):

0.5.1 Initial Feature Extraction:

The CNN extracts features from the image, just as it did during training.

0.5.2 Temporary Prediction:

Temporary predictions for both tasks are made, guiding the model on possible species and disease categories.

0.5.3 Final Prediction Using Stacked Layers:

The final prediction layers generate the ultimate species and disease classification by refining the temporary predictions.
Only these final predictions are used in the results, as they represent the model's most accurate assessment.

0.6 Step 7: Evaluation Metrics

To assess model performance, the researchers used the following metrics:
1.Accuracy: Measures the percentage of correct predictions out of the total predictions.
2.F1-score: Balances precision (how many of the predicted positive cases were true) and recall (how many true cases were identified by the model).
3.False Positive Rate (FPR): Assesses the number of incorrect positive predictions, helping identify any bias in the model.

0.7 Summary of the Methodology Pipeline

The seven main phases of the GSMo-CNN pipeline include:
Dataset Preparation: Standardization of Images and Splitting for the training, validation, and testing of the data
Backbone CNN: Determination of either the inceptionV3 or the resnet as it performs the tasks of feature extraction or any other
multi-output architecture design: Single multi-task strategy will do, stacking more output layers increases the accuracy.
GSMo-CNN design. Utilize the following aspects; convolutional, temporary prediction, and stacked layers
Training the Model by optimization using minimum loss value and adjustment of weights values.
6. Infer Predictions: Obtain more accurate predictions of both species and disease.
7. Evaluation of Performance: Test with accuracy, F1-score, FPR.
The stacked architecture enables GSMo-CNN to treat the two tasks as one in a compact structure. The architecture of the stacked model leverages relationships between tasks, hence improving accuracy. It forms a great base for further development of such applications for practical use in agricultural fields, especially in field diagnostics.

0.8 Detailed Comparison of Datasets Used in the Study

Three datasets were appropriately chosen which are: Plant Village, Plant Leaves, and PlantDoc. These datasets have been chosen to test the robustness of various CNN models, especially the proposed GSMo-CNN. Each of these datasets has a different nature, testing the model under controlled, diverse, and real-world conditions, respectively. More details follow about each dataset, its composition, purpose, and the challenges it presents to the models.
Plant Village
Description: Plant Village is one of the most popular datasets used for plant pathology research. There are 54,305 images of leaf samples in this dataset, which comprises images from 14 species of plants across 22 disease categories. The capturing conditions were controlled, like laboratory settings, with a simple plain homogeneous background and superior illumination, thus ensuring that the texture, color, and patterns of the leaves could be easily observable.

0.9 Data Composition:

Species Diversity: Includes popular agricultural plants such as apple, blueberry, grape, and tomato.
Disease Variety: Covers common diseases like apple scab, grape black rot, and tomato early blight. Each species may have multiple disease categories, adding to the classification complexity.
Background Uniformity: All images have consistent lighting and background, making it easier for models to detect and classify features without the noise or interference present in natural settings.
Purpose: The primary usage of this dataset is for benchmarking; that is, to determine the ideal performance of a model in perfect conditions. It's uniform, thus providing CNNs with good accuracy as it has a very little amount of interference or noise due to its background.
Challenges: Although Plant Village is an excellent benchmark, the experimentally controlled setup of this cannot replicate the complexity of an agricultural field with varied lighting, background or similar environmental conditions.

Performance Benchmark: Under controlled conditions, the CNN models show the highest accuracy on the Plant Village dataset as against any other dataset. In that, the results also point out that InceptionV3 leads with *accuracy of 98.7%* with *0.97 F1-score*. Such results have reflected the capability of every model in feature extraction and classification when no noise exists.

0.10 PlantDoc

Description: PlantDoc is a collection of 2,598 images showcasing 13 plant species with 17 diseases. It was recorded in different settings, such as natural settings, variable illumination, and various orientations, considering the multiple real-world conditions that agricultural practitioners are likely to face.
Data Composition:
Species and diseases representation: The dataset contains species such apple, cherry, corn and tomato, along with the diseases corresponding to these species.
Complex Background: The images are having high background noise i.e soil, other plants and in some images farming tools which makes it difficult for CNNs to distinguish leaf features with accuracy.
Variable angles and variable lighting: The images in PlantDoc are of different angles and with different lighting than Plant Village, resulting in shadow effects, overexposed and underexposed in some images.
Another dataset suitable for evaluating the CNNs is PlantDoc, established as a testbed for the reliability and adaptiveness of the model under real-world condition. It gives a measure for the ability of CNNs to generalize out of otherwise perfect conditions seen in lab.
Issue: Environmental noise of PlantDoc such as hazy backgrounds and variable light conditions are the potential challenges for CNNs. Hence models should distinguish the leaf and its surroundings correctly and be able to adapt to the noise and be able to achieve a high classification accuracy.

0.11 Performance Benchmark:

PlantDoc saw the lowest accuracy scores across all models, underscoring the dataset's complexity. InceptionV3 maintained the highest performance with 94.2% accuracy and a 0.93 F1-score, but all models experienced noticeable performance drops compared to their results on Plant Village and Plant Leaves.

0.12 Key Observations

1.Controlled vs. Real-world Conditions:
Models perform best on Plant Village due to its clean, controlled setup, where there are minimal distractions, and disease symptoms are clearly visible.
Performance drops on PlantDoc highlight the additional challenge posed by real-world conditions, where models must contend with noise, shadows, and various other confounding factors.

0.12.1 Impact of Background Complexity:

PlantDoc's complex backgrounds present a significant challenge, especially for models not optimized for high noise levels. CNNs must differentiate disease-related features from unrelated background elements. Plant Leaves, with moderate variability, serves as a good dataset, providing insights into a model's ability when moving from controlled set up to real-world set up.

0.13 Species and Disease Diversity:

Though preparing Plant Village is much more complicated and allows the preparation to achieve the greatest accuracy of species and diseases, a comparable preparation is easier. However, the models are

tested somewhat under more realistic scenarios than the Plant Village does, enabling them to bridge laboratory and field data.

0.13.1 Performance Trend Across Datasets:

Since datasets are relatively uncontrolled along with the diversity in field conditions, accuracy down goes. InceptionV3 exhibits the maximum performance on all the datasets. For this type of model which trains without diversity in laboratory settings alone, it is observed at times that these might drastically fail in the real-world scenario due to some noisy backgrounds; so, the requirement of variance at times of training datasets gets pertinent

0.14 Conclusion

The dataset comparison clearly shows that while controlled datasets like Plant Village are excellent for baseline performance evaluation, real-world datasets such as PlantDoc are essential for evaluating a model's practical ability. Plant Leaves is a transitioning dataset providing a moderate complexity in the background. Differences in accuracy across these datasets highlight the need for the diversity of datasets for well-generalizing CNN models-the need that is particularly imperative for a model used to predict in real-life scenarios.

Dataset	Images	Plant Species	Disease Types	Background Complexity	Typical Performance	InceptionV3 Accuracy	InceptionV3 F1-Score
Plant Village	54,305	14	22	Controlled, consistent	High	**98.7%**	**0.97**
Plant Leaves	4,502	12	22	Moderately varied	Moderate	**96.5%**	**0.95**
PlantDoc	2,598	13	17	Field-based, highly variable	Challenging	**94.2%**	**0.93**

Detailed Comparison of CNN Performance

1. InceptionV3

It should have special features like different information is being extracted at the same time with varied filters of different sizes with adaptability towards complex data.

Approximately 23 million parameters. This places a balance between model complexity and computational efficiency.

Performance:

Plant Village: Achieved 98.7% accuracy and 0.97 F1-score

Plant Leaves: Maintained high accuracy of 96.5% with a 0.95 F1-score,

PlantDoc: Scored 94.2% accuracy and 0.93 F1-score,

Advantages of InceptionV3-It shows high adaptability and handling of complex patterns, good for large-scale applications related to plant pathology as it can suit practical purposes in the way of such datasets as provided in this problem of the PlantDoc dataset.

2. ResNet101

Special Features: The residual block makes it easier for the network to learn identity functions and prevents degradation in performance with an increase in the depth of the network by connecting layers skipping one or more.

It has about 44.5 million parameters, thus making it more computationally expensive than InceptionV3 but allows for deeper learning.

Performance:

Plant Village: Achieved 97.8% accuracy with a 0.96 F1-score.

Plant Leaves: Scored 94.9% accuracy and 0.92 F1-score

PlantDoc: Scored 90.5% accuracy and 0.88 F1-score

Advantages: ResNet101's depth and residual connections make it powerful for feature-rich tasks, though it requires high computational resources and may not be ideal for real-time field applications due to its complexity.

3. MobileNetV2

Features: The model uses the inverted residual structure with bottleneck layers and depth wise separable convolutions, that significantly reduces the number of parameters, while improving the speed without compromising much accuracy.

Parameters: Approximately 3.4 million, making it the lightest model in this comparison

Performance

Plant Village: 91.4% accuracy and 0.89 F1-score.

Plant Leaves: 89.2% accuracy with 0.87 F1-score, performs quite well on basic variations, but is not as effective for high-resolution images.

PlantDoc: 85.0% accuracy and 0.82 F1-score but fails to generalize well for real-world variability.

Strengths: It is light in weight and deployable on mobile devices; yet it compromises some accuracy and is less robust in high variable conditions like those experienced in PlantDoc.

4. EfficientNet (B0)

Special Features: This architecture uses squeeze-and-excitation blocks, enhancing its ability to focus more on important features while adding less computational overhead.

Parameters: Approximately 5.3 million parameters, light in weight, and thus efficient.

Plant Village: 93.6% accuracy and 0.91 F1-score, reliable but less effective than deeper networks.

Plant Leaves: 90.4% accuracy and 0.89 F1-score, showing a balanced performance but less effective for high-detail tasks.

PlantDoc: 88.3% accuracy and 0.85 F1-score, showing some difficulty in noisy environments.

Advantages: EfficientNet is a good tradeoff between efficiency and performance; it is suitable for the applications where computational resources are limited but a higher accuracy is still desired than offered by MobileNetV2.

5. VGG16

Special Features: Although it does not contain the complex connections of the latest models, VGG16

is consistent due to deep architecture and uniform layer structure.
Parameters: It is about 138 million, which makes it computationally intensive and requires a lot of memory and processing power.
Performance:
Plant Village: The accuracy is 94.5% and the F1-score is 0.93.
Plant Leaves: Accuracy is 92.0% with an F1-score of 0.90.
PlantDoc: Achieved 87.6% accuracy with a score of 0.84 F1. In the real-world, this system fails to capture the variability.
Advantages: Although VGG16 can be used for the purpose of high-accuracy tasks in controlled environments, the computational requirements are high and make it less useful for real-time applications.
Major Insights
InceptionV3 was the overall top winner in all datasets and indeed performed very well, although not at the level needed, in the real dataset PlantDoc.
ResNet101 performed very well compared to other models but at a slight degradation of quality in real-world environments.
MobileNetV2 for mobile applications where the required resources are limited but comes at the cost of complex situations with compromised accuracy
For situations involving both compactness and precision as requirements, EfficientNet tends to balance efficiency with respect to performance.
Although reliable, VGG16 is not efficient when dealing with large datasets, as it is highly computational intensive.

Conclusion
The best backbone model for multi-task CNNs in application such as precision agriculture in terms of performance and efficiency across controlled and real-world datasets is ∗InceptionV3∗.

Model	Parameters (millions)	Plant Village Accuracy	Plant Village F1-Score	Plant Leaves Accuracy	Plant Leaves F1-Score	PlantDoc Accuracy	PlantDoc F1-Score
InceptionV3	23	**98.7%**	**0.97**	**96.5%**	**0.95**	**94.2%**	**0.93**
ResNet101	44.5	97.8%	0.96	94.9%	0.92	90.5%	0.88
MobileNetV2	3.4	91.4%	0.89	89.2%	0.87	85.0%	0.82
EfficientNet	5.3	93.6%	0.91	90.4%	0.89	88.3%	0.85
VGG16	138	94.5%	0.93	92.0%	0.90	87.6%	0.84