Cassava Plant Disease Classification Using a CNN-Swin Hybrid Model

**Abstract**

This study investigates the effectiveness of Convolutional Neural Networks (CNNs) and Swin Transformers, specifically VGG19, ResNet, EfficientNetV2L, and Swin Base for detecting and classifying cassava plants diseases. We utilized PyTorch for training and testing multiple Swin and CNN architectures, focusing on maximizing each model's configuration for best performance. The results show that the EfficientNetV2L and Swin based hybrid model surpassed other models, achieving a F1-score of 79% and demonstrating superior capability in dealing with the image data efficiently. However, the F1 score for Cassava Bacterial Blight (CBB) was significantly low on all architectures, suggesting challenges in detecting certain disease manifestations effectively. The findings show the potential of utilizing advanced CNNs and Swin Transformers for improving cassava disease diagnosis but also highlight the necessity for more varied training datasets to enhance model robustness.

**1 Introduction**

Cassava, a.k.a. Manihot Esculenta is a crucial staple crop across tropical and subtropical regions of the world. Its significance is particularly pronounced in Africa, Asia, and Latin America, where it serves as a primary source of carbohydrates for over half a billion people. However, the productivity and quality of cassava are severely hampered by various plant diseases, including cassava mosaic disease (CMD), cassava brown streak disease (CBSD), cassava bacterial blight (CBB), and cassava green mottle disease (CGMD). These diseases not only cause substantial losses to yield but also threaten food security in regions which rely on cassava.

The traditional methods of disease detection, which primarily involve manual inspection by farmers and experts, are time-consuming, labor-intensive, and often inaccurate due to the difficulty of identifying early disease symptoms. Therefore, there is a critical need for more efficient, and automatic methods that can improve disease management and reduce losses. In this regard, Convolutional Neural Networks (CNNs), Swin Transformers, and utilizing Transfer Learning in image processing has opened new avenues for rapid and accurate disease detection and classification.

**2 Dataset and Libraries**

We utilized the Crop Diseases Classification dataset available on Kaggle, comprising nearly 18,000 images of various crops affected by different diseases. The dataset offers a diverse range of crops and labels for diseased and healthy plants, making it suitable for training and evaluating our model's performance across different scenarios. Below is the dataset as well as the number of samples per class.

<https://www.kaggle.com/datasets/mexwell/crop-diseases-classification>

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | CBB | CBSD | CGM | CMD | Healthy | Total |
| # of Images | 921 | 1,831 | 1,993 | 11,027 | 2,166 | 17,938 |

**2.1 Libraries Used**

* Pytorch
* Torchvision
* Timm
* PIL
* Transformers

**3 Experiments**

In the experimental setup, multiple deep learning models were trained using the PyTorch library to classify images of cassava diseases. The experiment involved several models with varying configurations. The models were trained with a .001 learning rate using SGD and Adam optimizers and Cross Entropy loss. The dataset was divided into 3 sections randomly. 70% for training, 20% validation and 10% for testing. Then the data was preprocessed by first normalizing the data and reshaping it for the appropriate model.

The models were already pre-trained thus, it lowered the computational difficulty substantially. To fit our dataset, the models were fine tuned until the validation score converged for multiple epochs. Finally, the models were tested. The experiments were also performed with a GTX3060. Due to RAM constraint and large size of the models, different batches sizes were used for different models. To ensure that the models can be compared properly, the randomized portion for the preprocessing such as RandomFlip was removed to ensure consistency.

The output was then compared to the correct labels to produce the 5x5 confusion matrix which allowed us to calculate all of the accuracy metrics below.

**3.1 Results**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Class Name** | **Metric** | **VGG** | **ResNet** | **EfficientNetV2L** | **Swin Base 224** | **CombinedVGGEfNet** | **SwinEfficientNetComb** |
| Cassava Bacterial Blight (CBB) | Accuracy | 96.32% | 95.87% | 96.60% | 96.49% | 96.71% | 96.99% |
|  | F1 Score | 50.75% | 44.78% | 65.54% | 62.72% | 64.67% | 69.32% |
| Cassava Brown Streak Disease (CBSD) | Accuracy | 94.48% | 94.70% | 95.54% | 95.37% | 95.54% | 95.93% |
|  | F1 Score | 74.29% | 77.00% | 80.77% | 79.71% | 79.90% | 82.06% |
| Cassava Green Mottle (CGM) | Accuracy | 94.65% | 94.42% | 94.76% | 94.65% | 95.32% | 95.43% |
|  | F1 Score | 73.33% | 72.83% | 77.83% | 72.57% | 79.71% | 78.97% |
| Cassava Mosaic Disease (CMD) | Accuracy | 93.03% | 92.75% | 94.65% | 93.53% | 94.53% | 94.81% |
|  | F1 Score | 94.27% | 93.88% | 95.44% | 94.75% | 95.34% | 95.66% |
| Healthy | Accuracy | 90.52% | 89.35% | 92.25% | 92.08% | 91.36% | 92.19% |
|  | F1 Score | 67.56% | 65.08% | 70.00% | 68.44% | 69.18% | 70.34% |

**Global Results**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Macro Precision** | **Macro Recall** | **Macro F1 Score** | **Top 2 Accuracy** |
| VGG | 84.49% | 68.90% | 79.16% | 72.04% | 94.01% |
| ResNet | 83.55% | 69.21% | 75.68% | 70.71% | 94.70% |
| Swin | 86.06% | 72.86% | 79.51% | 75.64% | 95.26% |
| EfficientNetV2L | 86.89% | 78.07% | 77.80% | 77.91% | 96.37% |
| CombinedVGGEfNet | 86.73% | 77.33% | 78.66% | 77.76% | 94.59% |
| SwinEfficientNetComb | 87.67% | 78.28% | 80.42% | 79.27% | 95.87% |

**4 Conclusion**

As seen by the results, the EfficientNetV2L performance is the best among all of the models and combinations. With over a 96.37% top 2 accuracy rate, it provides a great first step for better solving the classification problem for cassava diseases. However, there are great limitations to the results gathered. First, the dataset was generated from a specific camera and environment. Next, the accuracy for F1 score for CBB class remains very low. Larger and more diverse dataset for each of the classes will be needed to better train the model.

**References**

[1] Liu, Z., Lin, Y., Cao, Y., et al. (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, 10-17 October 2021, 10012-10022. https://doi.org/10.1109/ICCV48922.2021.00986

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. https://doi.org/10.1109/CVPR.2016.90

[3] Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv.org. https://doi.org/10.48550/arxiv.1409.1556 [4] Tan, M., & Le, Q. V. (2021). EfficientNetV2: Smaller Models and Faster Training.arXiv.org. <https://doi.org/10.48550/arxiv.2104.00298>

[4] Tan, M., & Le, Q. V. (2021). EfficientNetV2: Smaller Models and Faster Training.arXiv.org. <https://doi.org/10.48550/arxiv.2104.00298>