Research Paper 1 (as per Excel Sheet)

Overall Summary:

Title - DiaMOSPlant: A Dataset for Diagnosis and Monitoring Plant Disease
Source - ResearchGate
Type - Journal Paper
Publisher - MDPI
Year of Publication - 2021
Dataset Distribution Ratio - 3006 images of leaves and 499 images of fruits
Dataset Splitting Ratio - 7 : 2 : 1 (training : validation : testing)
Method Applied - EfficientNetB0
Performance Metrics - Training Accuracy - 81.13%, Validation Accuracy - 82.82%, Testing Accuracy - 83.38%, Precision - 81.14%, Recall - 83.38% and F1 Score - 82.23%

This paper introduces a new dataset called DiaMOS Plant which consists of 3505 images of pear fruit and leaves affected by four diseases. The dataset was collected under realistic field conditions to maximize variability. The authors provide a benchmark for classifying diseases using convolutional neural networks and discuss limitations of existing plant disease datasets.

The authors analyze several existing plant disease datasets and find that they have limitations in size, representativeness, and completeness. They recommend that future datasets should have the following properties:

• Larger sample sizes with more disease classes to improve generalization
• Samples collected under field conditions to better represent real-world variability
• Images capturing disease symptoms at multiple stages of progression
• Images taken under different lighting and angle conditions
• Complete annotations with bounding boxes and multiple disease labels
• Performance baselines to enable development of new methods

The DiaMOS Plant dataset aims to address some of these recommendations by capturing images throughout the growing season under different lighting conditions to provide a more representative dataset for plant disease classification.

The dataset has 3006 images of leaves alone which are further classified into 43 Healthy Leaves, 884 Spotted Leaves, 54 Curled Leaves, 2025 Leaves with Curls.

The pictures were gathered using different devices including a smartphone (Honor 6x) and DSLR camera (Canon EOS60D), thus the images present two types of resolutions, 2976 X 3968 and 3456 X5184 respectively.

The methodology employed in this research paper to classify the diseases leaves is machine learning and deep learning. The concepts of machine and deep learning are used to develop models that can classify and recognize foliar diseases based on the images of leaves. In particular from the performance stats it can be seen that the three networks EfficientNetB0, InceptionV3, and MobileNetV2 have a better generalization capacity than the VGG19 and ResNet50 networks. In fact CNN algorithms like EfficientNetB0, InceptionV3 and MobileNetV2 obtained an accuracy for the test set of 83.38 %, 82.72 %, 83.06 % respectively, while ResNet50 of 56.67 %, and VGG19 of 71.76 %. It was also seen that MoblileNetV2 tends to converge faster. The f1-score ratio does not show notable differences in performance, reporting a high value for EfficientNetB0, InceptionV3, and MobileNetV2.

The dataset was trained and tested with two different optimisers namely RMSProp and the Adam optimiser.

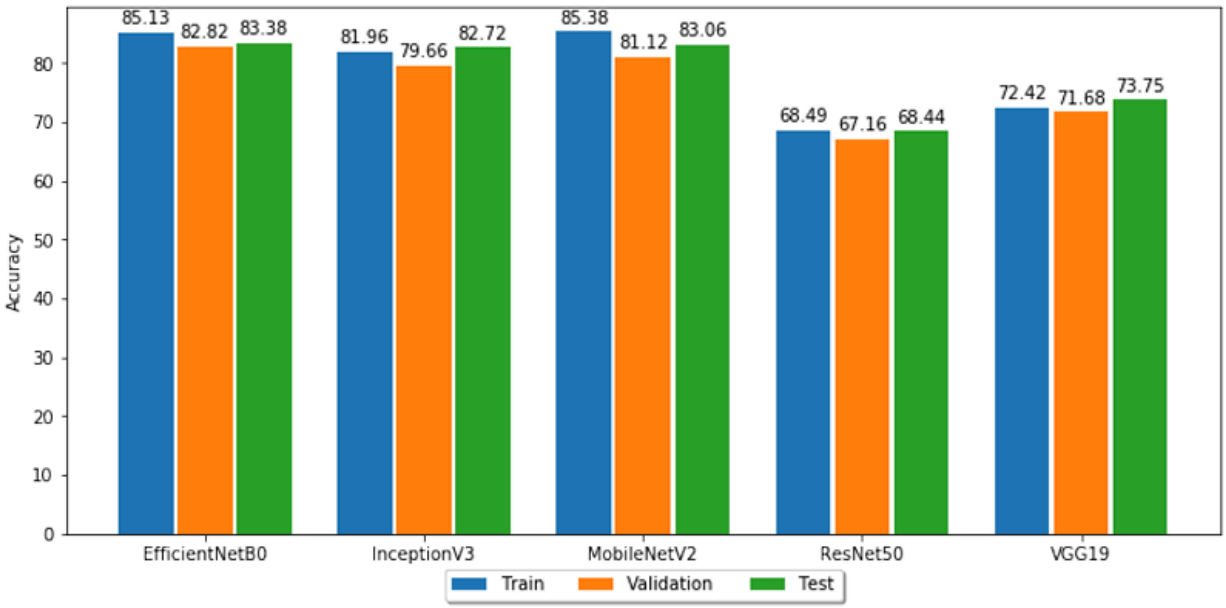
The RMSProp or the Root Mean Square Propagation optimiser is an adaptive learning rate optimization algorithm designed for training neural networks like in this one. It is particularly well-suited for non-stationary objectives, which are common in deep learning. It is widely used in deep learning due to its ability to adapt learning rates based on the history of gradients and its effectiveness in dealing with the scenarios involving large and complex datasets.

The Adam also known as Adaptive Moment Estimation optimiser is an algorithm for optimizing the training of machine learning models, particularly neural

networks. It combines the advantages of two other popular optimization algorithms: AdaGrad and RMSProp. Overtime Adam has become one of the most popular optimization algorithms in the field of machine learning due to its adaptive nature and robust performance across a variety of tasks.

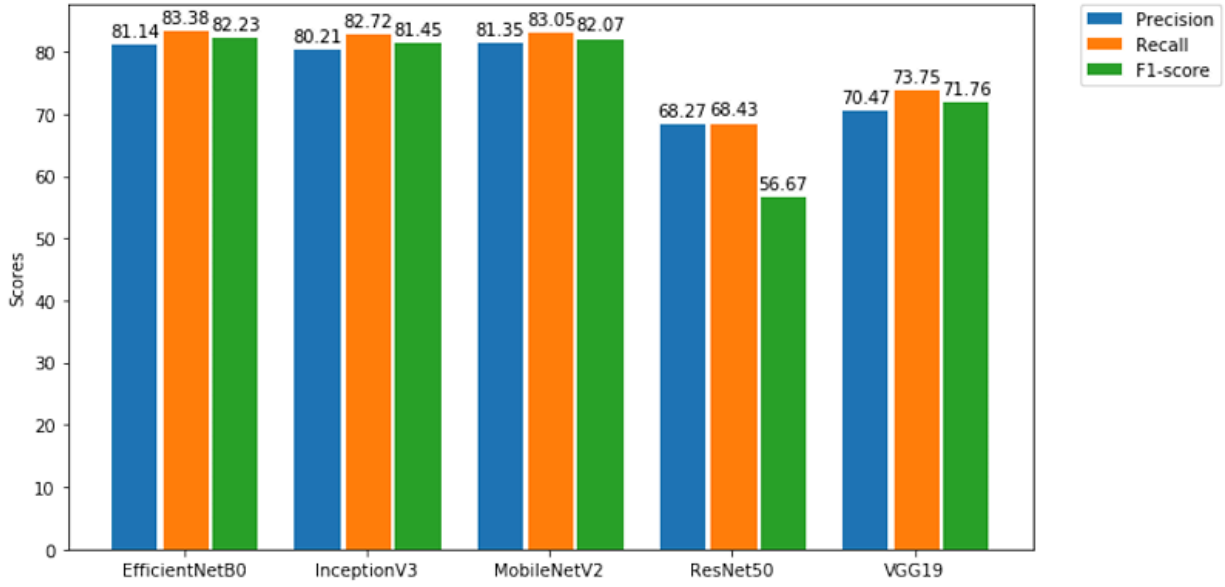
When the dataset was trained with these two optimisers in mind, the following metric stats we obtained:

With RMSProp:



CNN	*RMSprop*		
	Train Acc(%)	Validation Acc (%)	Test Acc (%)
EfficientNetB0	81.13	82.82	83.38
InceptionV3	81.96	79.66	82.72
MobileNetV2	**85.38**	81.12	83.06
ResNet50	68.49	67.16	68.44
VGG19	72.42	71.68	73.75

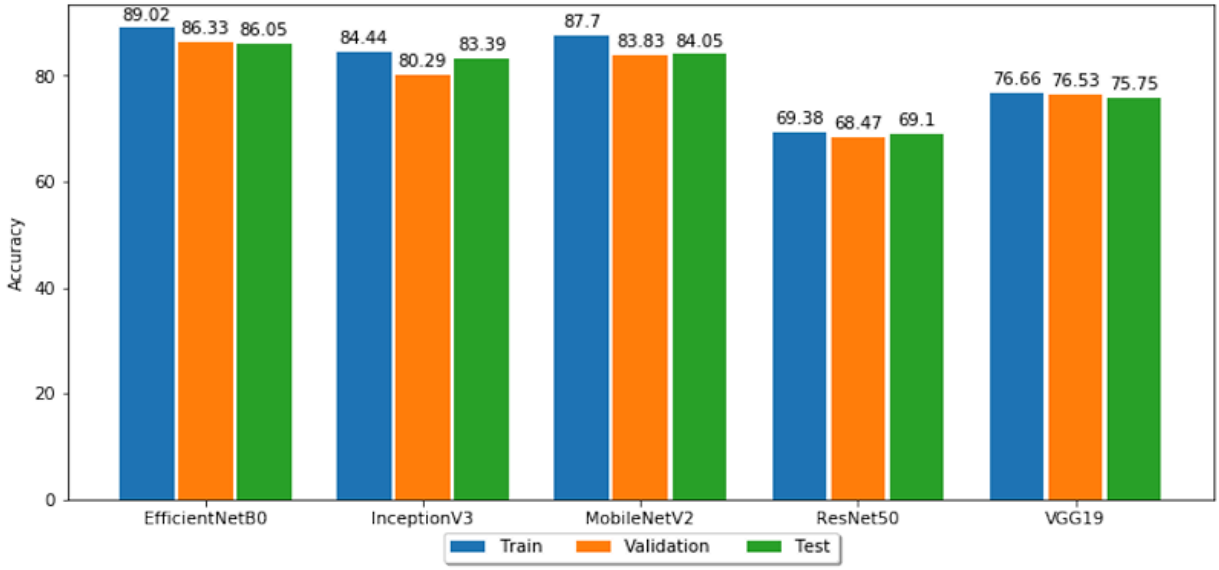
Accuracy obtained with RMSprop optimizer respectively in the training set, validation set and test set in the task of classifying the "healthy", "slug", "curl", "spot" classes.



	RMSprop		
CNN	**Precision (%)**	**Recall (%)**	**F1-score (%)**
EfficientNetB0	81.14	83.38	82.23
InceptionV3	80.21	82.72	81.45
MobileNetV2	81.35	83.05	82.07
ResNet50	68.27	68.43	56.67
VGG19	70.47	73.75	71.76

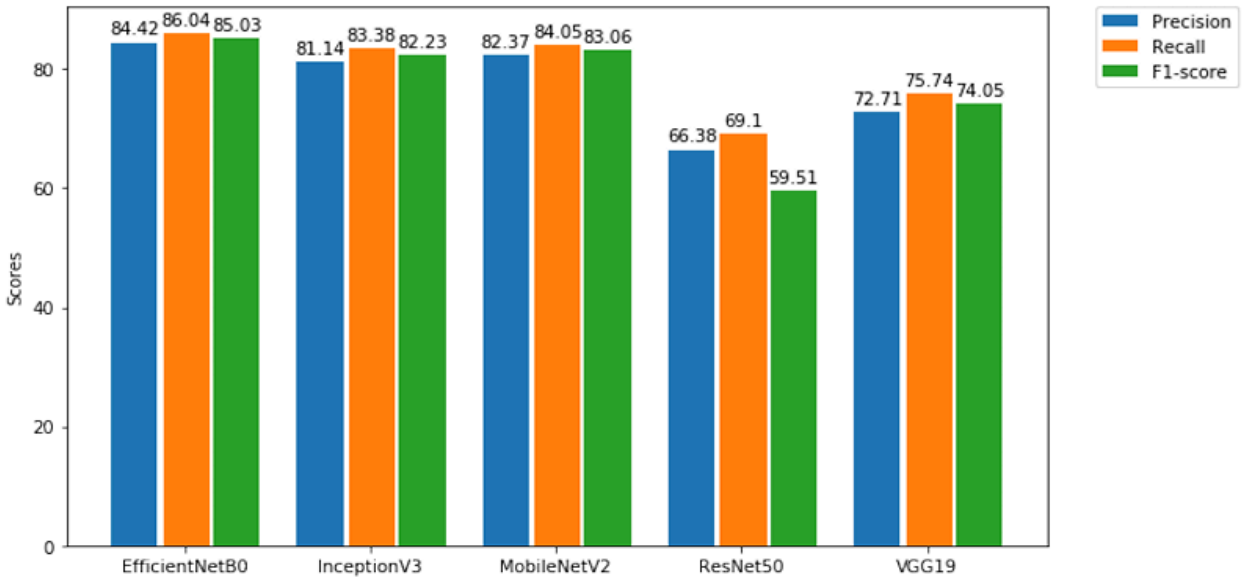
Precision, Recall, and F1-score reported with RMSprop optimizer on test set in the task of classifying the "healthy", "slug", "curl", "spot" classes.

With Adam:



CNN	*Adam*		
	Train Acc(%)	**Validation Acc (%)**	**Test Acc (%)**
EfficientNetB0	**89.02**	86.33	86.05
InceptionV3	84.44	80.29	83.39
MobileNetV2	87.70	83.83	84.05
ResNet50	68.38	68.47	69.10
VGG19	76.66	76.53	75.75

Accuracy obtained with Adam optimizer respectively in the training set, validation set and test set in the task of classifying the "healthy", "slug", "curl", "spot" classes.



	Adam		
CNN	**Precision (%)**	**Recall (%)**	**F1-score (%)**
EfficientNetB0	84.42	86.04	85.03
InceptionV3	81.14	83.38	82.23
MobileNetV2	82.37	84.05	83.06
ResNet50	66.38	69.10	59.51
VGG19	72.71	75.74	74.05

Precision, Recall, and F1-score reported with Adam optimizer on test set in the task of classifying the "healthy", "slug", "curl", "spot" classes.

In conclusion, the DiaMOS Plant Dataset offers a valuable resource for the research community, aiding in the development of robust AI models for plant disease diagnosis and monitoring.
This particular research paper also underscores the need for more comprehensive datasets to advance in the field of digital agriculture.