A Comparative Study

On different existing CNN models through Transfer Learning

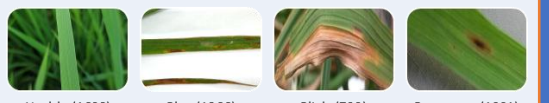
Prepared By:
Kshitij Ayush
Aditya Das
Bibek Kundu

Problem Statement

- The objective of this study is to conduct a comparative analysis of various standard **Convolutional Neural Network (CNN)** architectures to determine which model performs best for **Rice Leaf Disease** image classification task. The study aims to evaluate models such as **VGGNet**, **ResNet**, **Inception**, and **MobileNet**, comparing their accuracy, training time, and resource efficiency on the same dataset and under similar experimental conditions. The findings will help identify the most suitable CNN architecture for optimal performance in terms of accuracy and computational cost.

Work Flow Diagram

Original

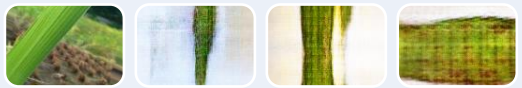


Healthy(1623) Blast(1366) Blight(722) Brownspot(1091)

Geometric Augmentation

Healthy(8115) Blast(6830) Blight(3609) Brownspot(5453)

WGAN-GP



Healthy (8115) Blast (8115) Blight (8115) Brownspot (8115)

P1

P2

P3

Base Feature Extractor

Classifier Head

Loss Function **(CCE)**

Learning Rate **(1e-5)**

ADAM Optimiser

Hold Out (80-20)

*Cost Sensitive Learning Techniques used

Accuracy

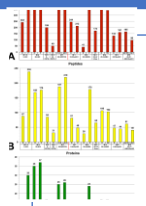
VALIDATION SET

Accuracy

TEST SET

Dataset and Preprocessing

Model Training



Evaluation Metrics

Dataset and Preprocessing

- Original Dataset :
  - Original images belonging to four classes which are **Healthy**, **Bacterial blight**, **Blast** and **Brownspot**. Applied with Standard Normalisation and Rescaled to (224 X 224) pixels.

- Geometric Augmentation:
  - Basic geometric augmentation like flips, rotation, shifts etc to create five images per image of the original image.

- Artificial Data:
  - **WGAN-GP**(**W**asserstein **G**enerative **A**dversarial **N**etwork with **G**radient **P**enalty): a generative model where two neural networks, a generator and a discriminator, compete against each other to create new, realistic data.

Dataset Description

Original dataset (OGD)


Healthy
•1623 images


Brownspot
•1091 images


Blast
•1366 images

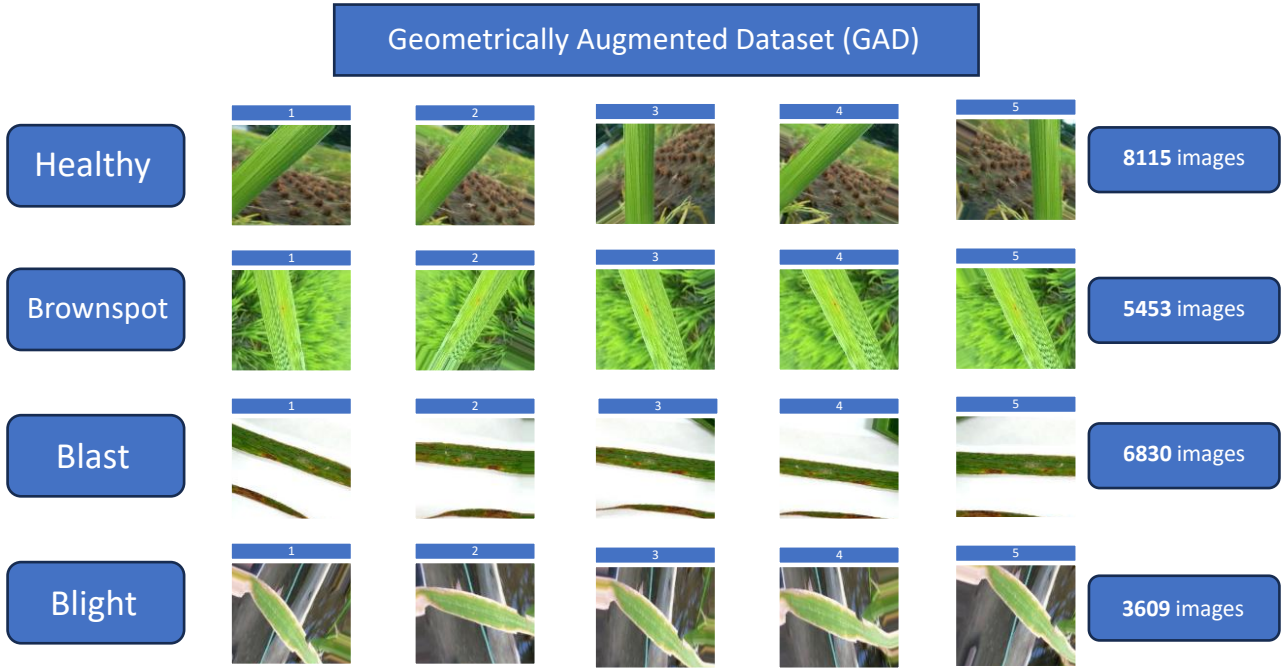

Blight
•722 images

*Applied with Standard Normalisation and Rescaled to (224 X 224) pixels

Geometric Augmentation

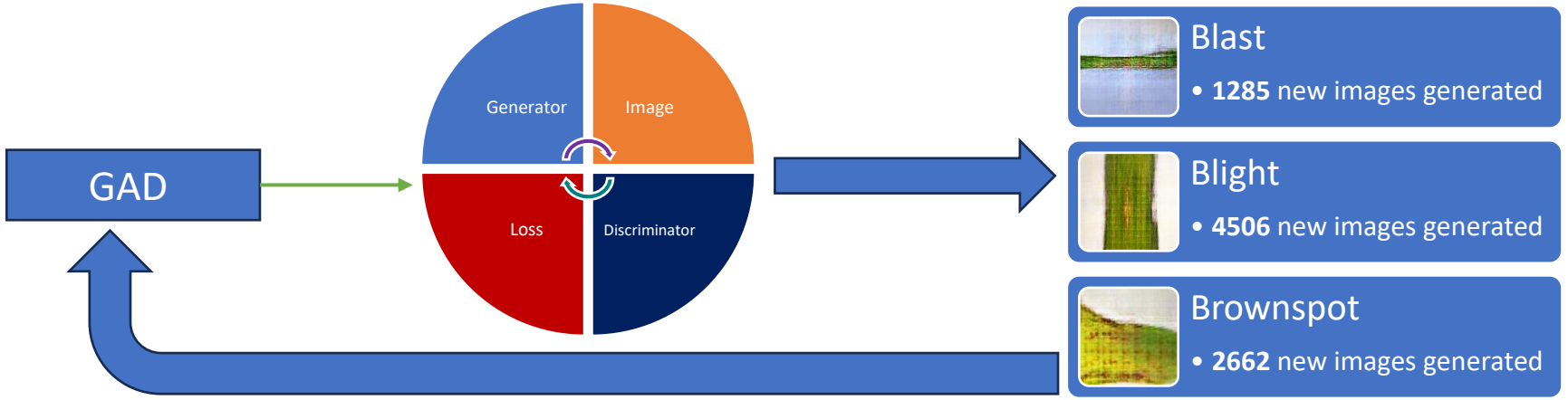
*done to increase the total size of the dataset by generating **5 images** per image of the original class

Geometrically Augmented Dataset (GAD)



- Geometric Augmentation was done to increase the overall size of the dataset.
- While geometric augmentation increased dataset size, it also amplified the **class imbalance**.
- The original dataset exhibits a notable class imbalance, with a ratio of approximately **2.25 (Healthy) : 1.51 (Brownspot) : 1.89 (Blast) : 1 (Blight)**.
- Advanced techniques were added to build a more robust model.

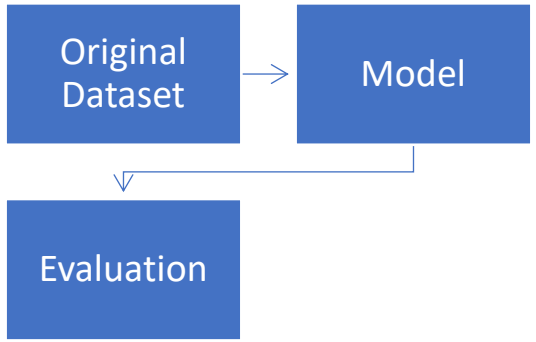
WGAN-GP : Generates new, highly realistic synthetic images by training two competing neural networks in a stable way to solve **class imbalance**.



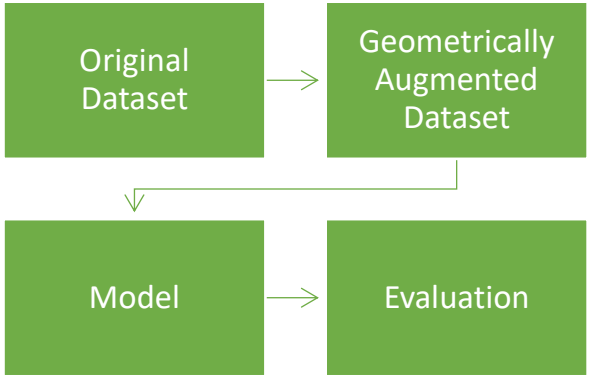
Thus, now each class has exactly **8115 images** and class imbalance have been solved.

Data Pipelines

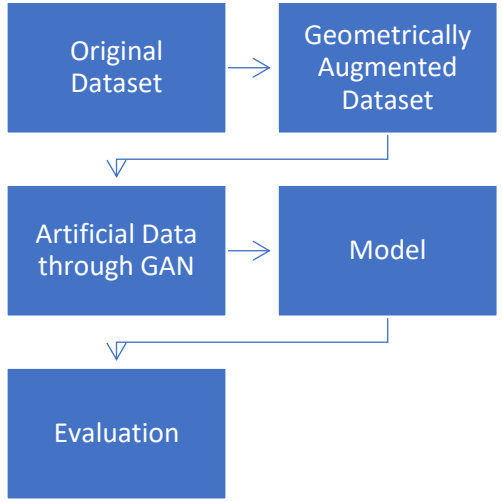
PIPELINE 1



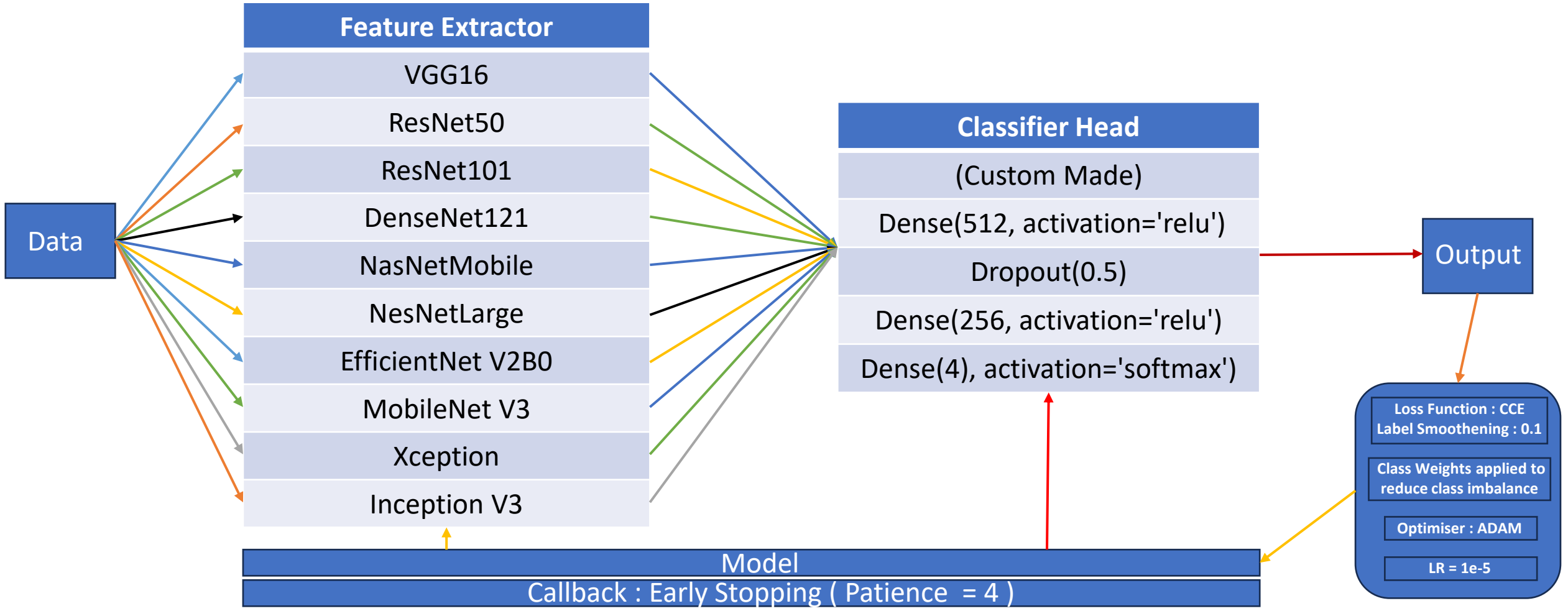
PIPELINE 2



PIPELINE 3



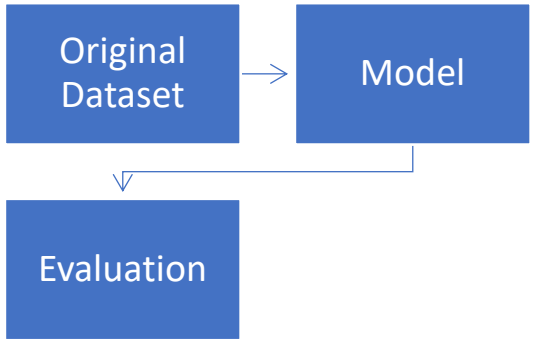
Model Training

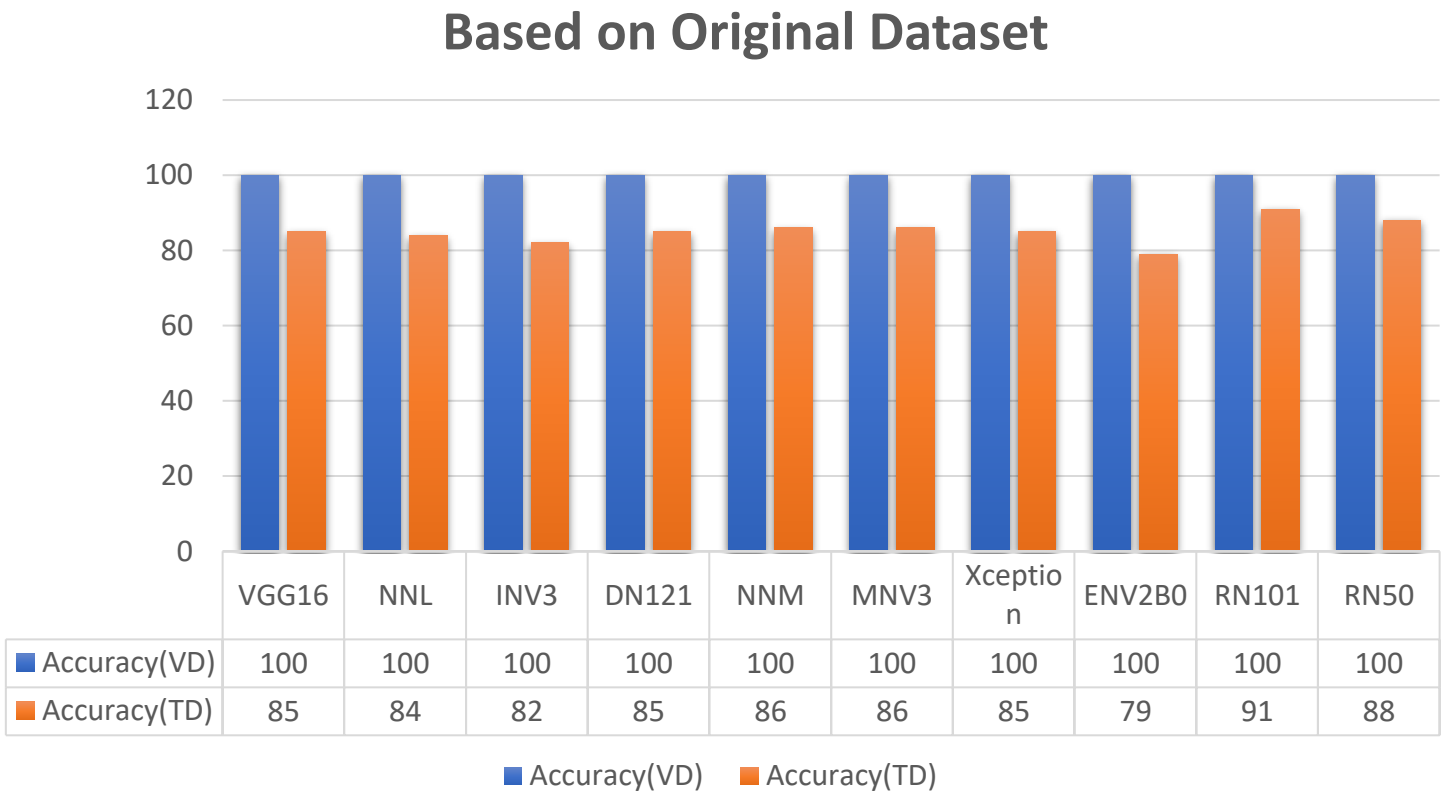


	Feature Extractor	Classifier Head
Phase 1 (30 epoch)	Frozen	Unfrozen
Phase 2 (60 epoch)	Unfrozen	Unfrozen

Evaluation Metrics

Work Flow



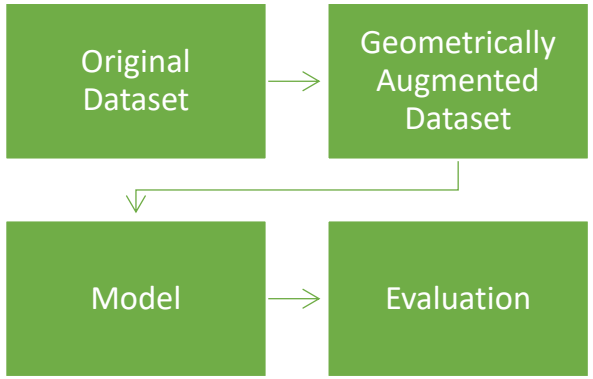


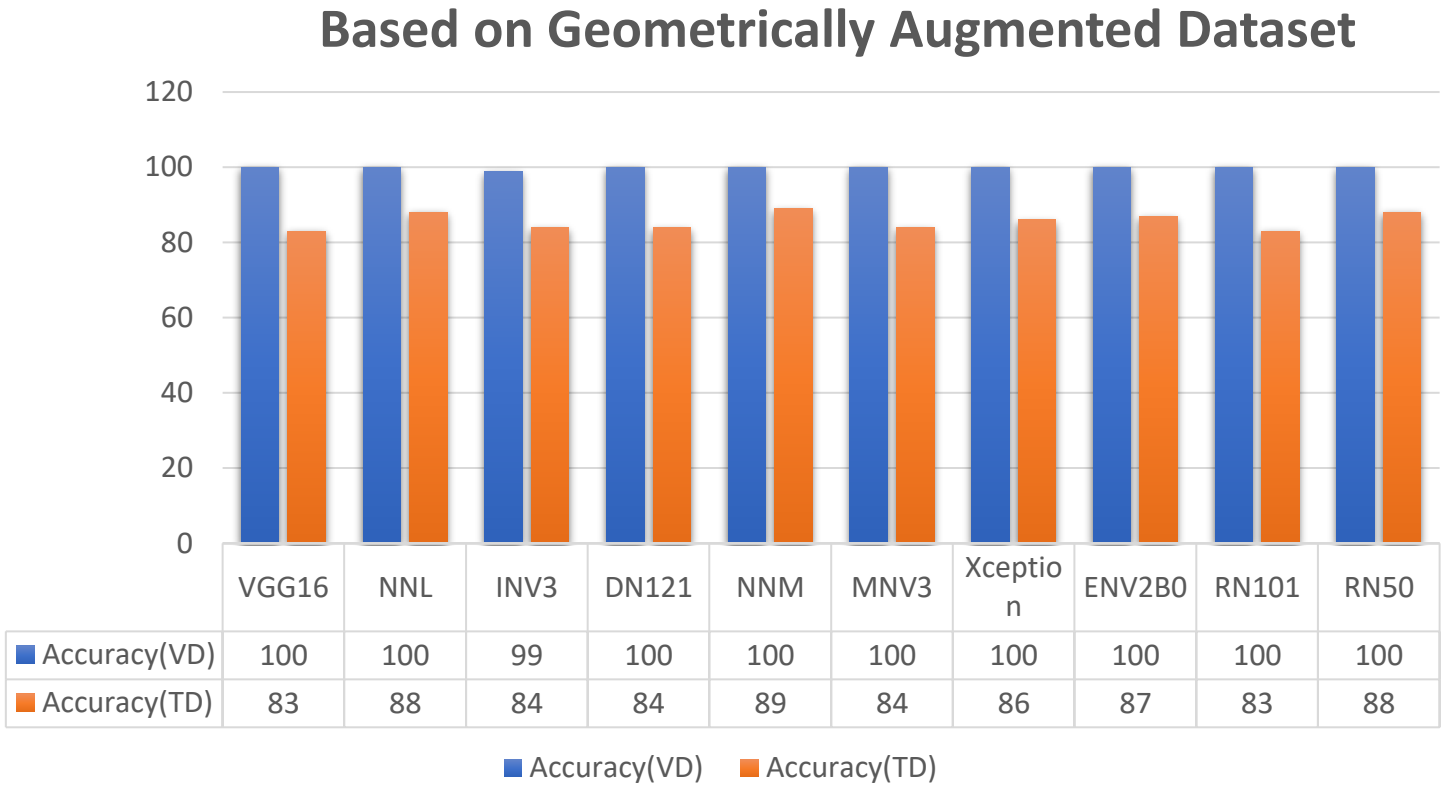
VD: Validation Dataset, **TD**: Test Dataset

- All CNN models achieved **100%** accuracy on the **Validation Dataset**.
- **RN101** achieved **highest accuracy** (**91%)** on **Test Dataset**.
- **RN50** stood at the **second place** with **88%** accuracy on **Test Dataset**.
- **NNM** and **MNV3** both achieved **86%** accuracy on **Test Dataset** holding the **third place**.

Evaluation Metrics

Work Flow



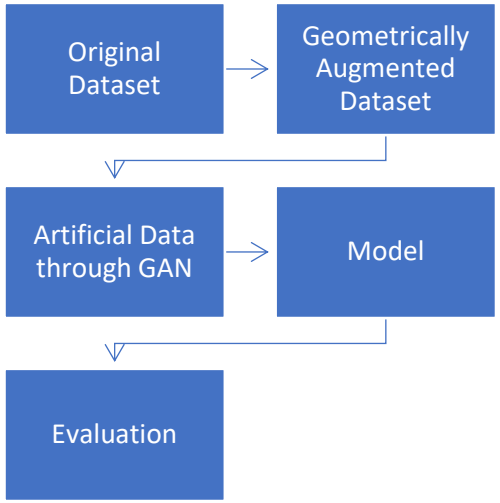


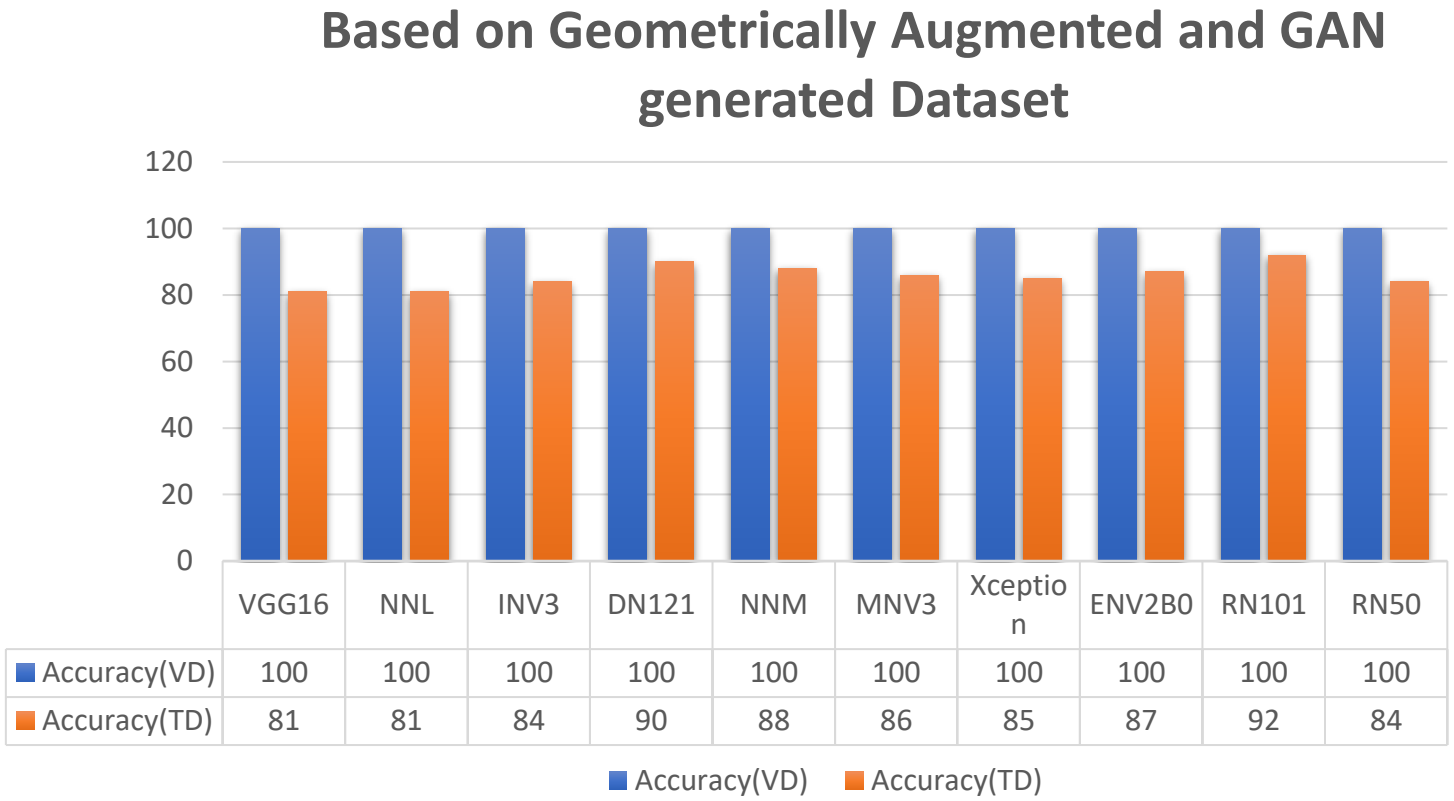
VD: Validation Dataset, **TD**: Test Dataset

- All CNN models achieved **99% to 100%** accuracy on the **Validation Dataset**.
- **NNM** achieved **highest accuracy** (**89%**) on **Test Dataset**.
- **RN50 and NNL** stood at the **second place** with **88%** accuracy on **Test Dataset**.
- **ENV2B0** achieved **87%** accuracy on **Test Dataset** holding the **third place**.

Evaluation Metrics

Work Flow





	VGG16	NNL	INV3	DN121	NNM	MNV3	Xception	ENV2B0	RN101	RN50
Accuracy(VD)	100	100	100	100	100	100	100	100	100	100
Accuracy(TD)	81	81	84	90	88	86	85	87	92	84

VD: Validation Dataset, **TD**: Test Dataset

- All CNN models achieved **100%** accuracy on the **Validation Dataset**.
- **RN101** achieved **highest accuracy** (**92%)** on **Test Dataset**.
- **DN121** stood at the **second place** with **90%** accuracy on **Test Dataset**.
- **NNM** achieved **88%** accuracy on **Test Dataset** holding the **third place**.

Conclusion

- All CNN models achieved **99–100% accuracy on Validation Dataset**.
- **Test accuracy dropped** to the range of **79–92%**, showing a gap in generalization.
- **ResNet101** consistently performed best, reaching **92% test accuracy** with augmented + GAN data.
- **DenseNet121** (90%) and **NasNetMobile** (89%) also showed strong results.
- Adding **geometric augmentation + GAN-generated data** improved balance across classes and boosted test performance.
- The gap between validation and test results emphasizes the need for **robust data preparation**, not just model depth.
- **Best Model:** *ResNet101 with GAN-augmented dataset* → most reliable and accurate.
- **Positive Impact of GAN:** solved class imbalance, improved diversity, and enhanced generalization ability of CNN models.