

1. Models

- 1.1 Previous Model(VGG16)

- The initial baseline model delivered 84.49% accuracy on a food dataset. It displayed a response time of 0.25 seconds, handling 3.93 transactions per second. It occupied 122.1 MB and used 964 MiB memory. The model struggled in classifying dessert, dairy, and bread categories, showing significant misclassifications among rice, soup, and egg items.

- 1.2 Submitted Models

- 1.2.1 Model A(**Xception**)

- This model contains over 22983219 parameters(2115595 trainable and 20867624 nontrainable). To improve its adaptability, it underwent diverse data augmentation techniques like interpolation, rotation, rescaling, shiftings, and horizontal and vertical flips. Initially compiled with Adam optimizer and trained for 24 epochs with a learning rate of 0.0001 and an additional 12 epochs and updated learning rate of 0.00001 for fine-tuning, it achieved a validation accuracy of 79%, improving to 84.2% after fine-tuning. Also, the last 5 layers were unfrozen as a part of the transfer learning process. The final accuracy on the evaluation set is 86.79%. Operational metrics recorded a 0.15-second response time and 4.66 transactions per second. This model underperforms in desert and dairy classification.

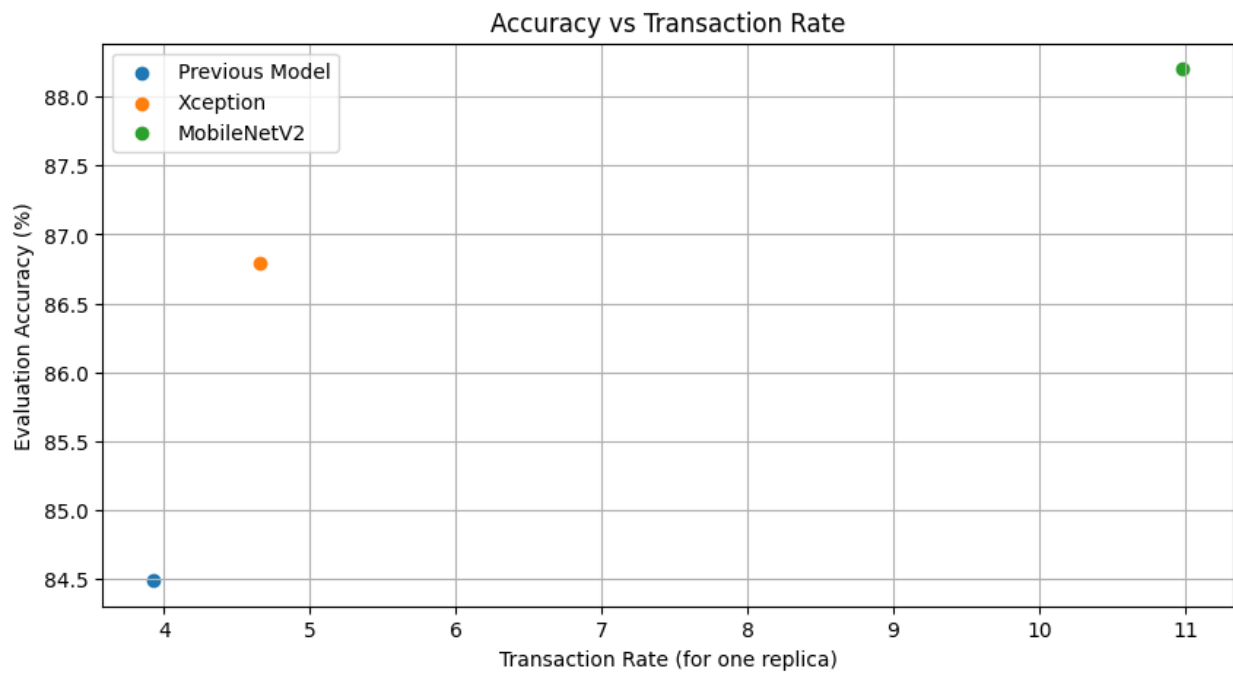
- 1.2.1 Model B(**MobileNetV2**)

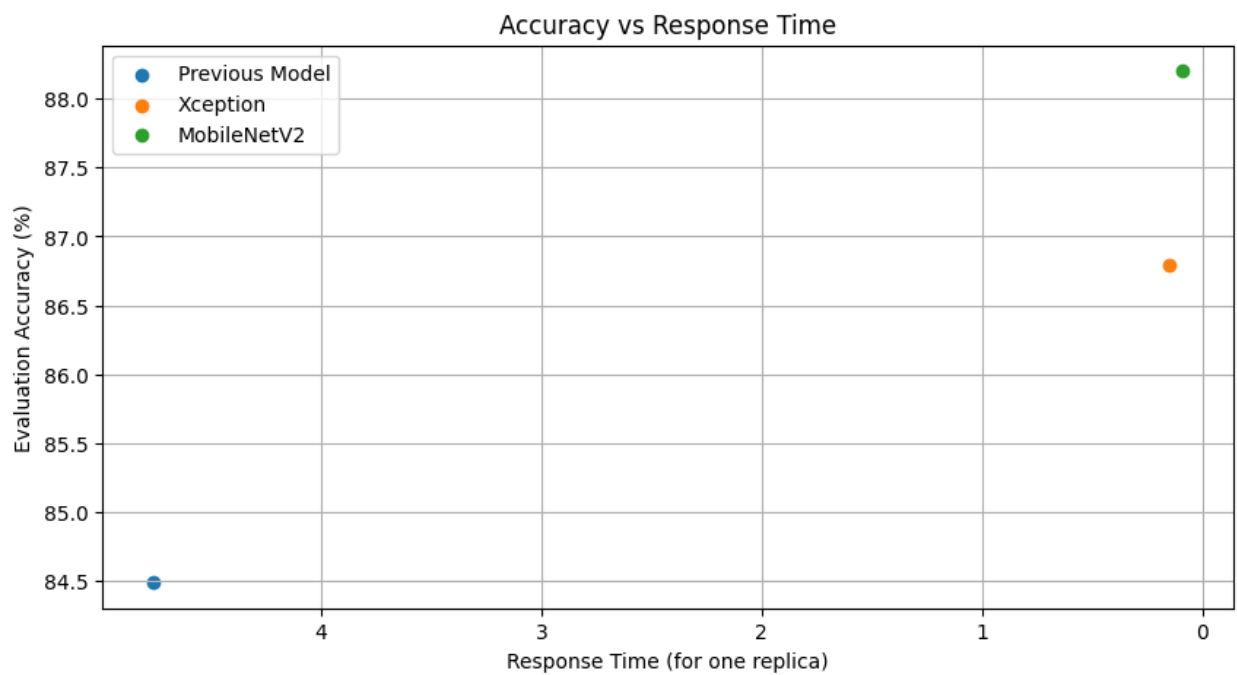
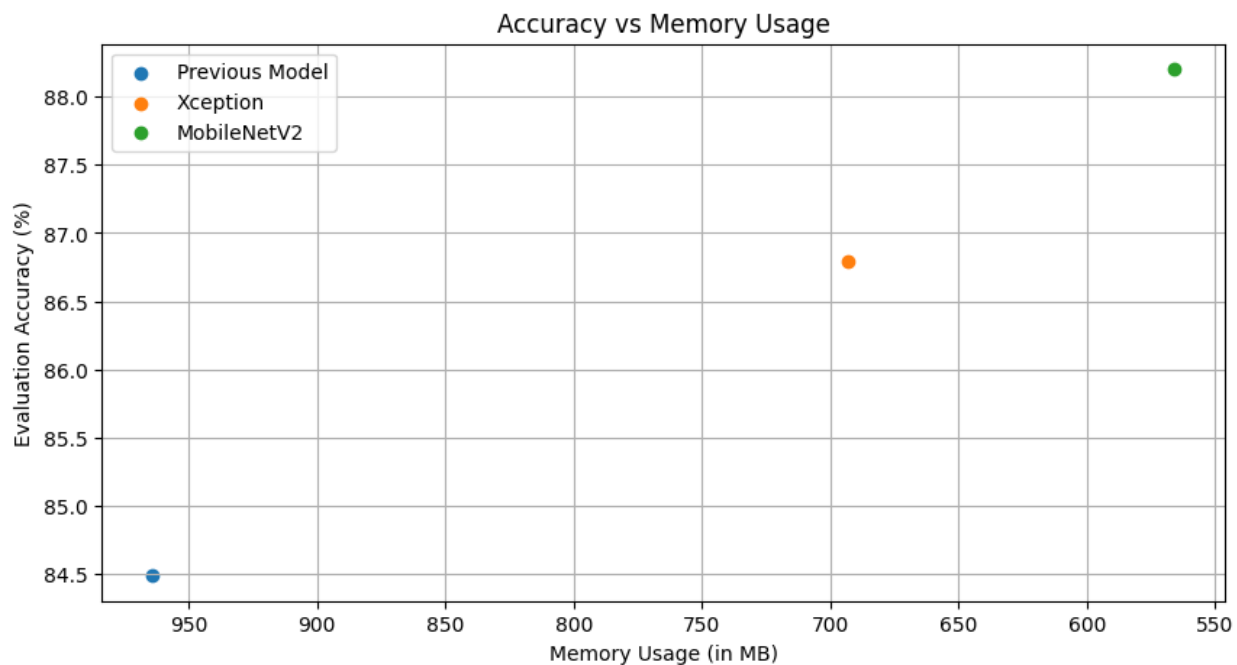
- This model contains over 3590219 parameters(1740427 trainable and 1849792 non-trainable). To improve its adaptability, it underwent diverse data augmentation techniques like interpolation, rotation, rescaling, shiftings, and horizontal and vertical flips. Initially compiled with Adam optimizer and trained for 24 epochs with a learning rate of 0.0003 and an additional 12 epochs and updated learning rate of 0.00003 for fine-tuning, it achieved a validation accuracy of 84.84%, improving to 86.56% after fine-tuning. Also, the last 5 layers were unfrozen as a part of the transfer learning process. The final accuracy on the evaluation set is 88.2%. Operational metrics recorded a superior 0.09-second response time and 10.98 transactions per second. However, it still underperformed in desert and dairy classification.

- 1.3 Model Summary

Base Model	Parameters	Training Epochs	Fine Tuning Epochs	Final Accuracy	Response Time	Memory Usage	Transactions Per Second	Model Size(MB)
Previous Model	15257419	24	12	84.49	4.76	964 MB	3.93	122.1
Xception	22983219	24	12	86.79	0.15	693 MB	4.66	134.7
MobileNet V2	3590219	24	12	88.2	0.09	566 MB	10.98	28.8

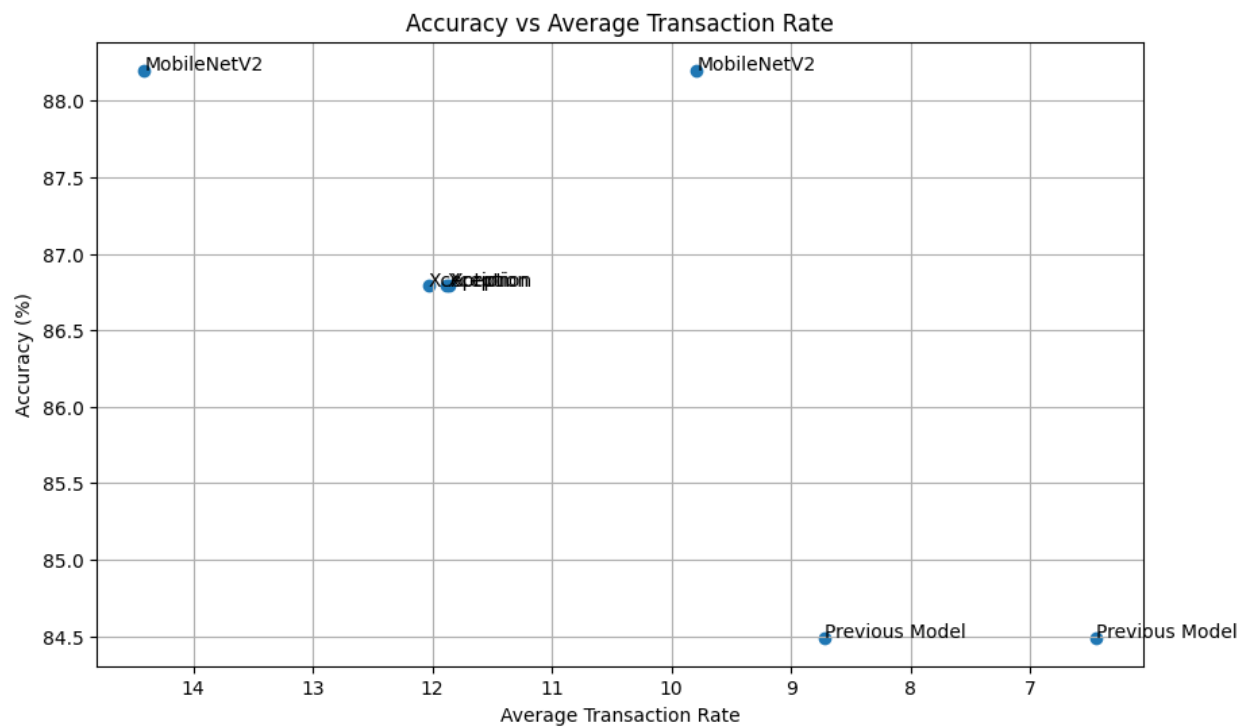
- Here **MobileNetV2** has the best overall metrics.

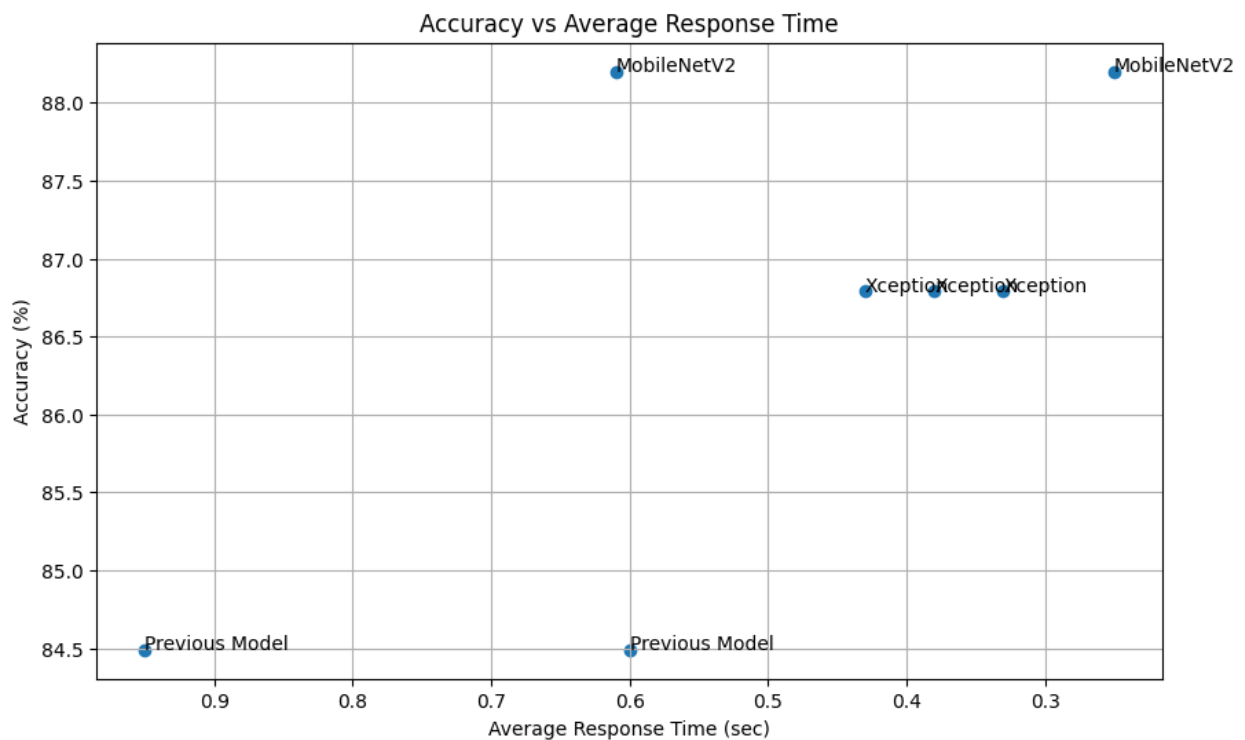
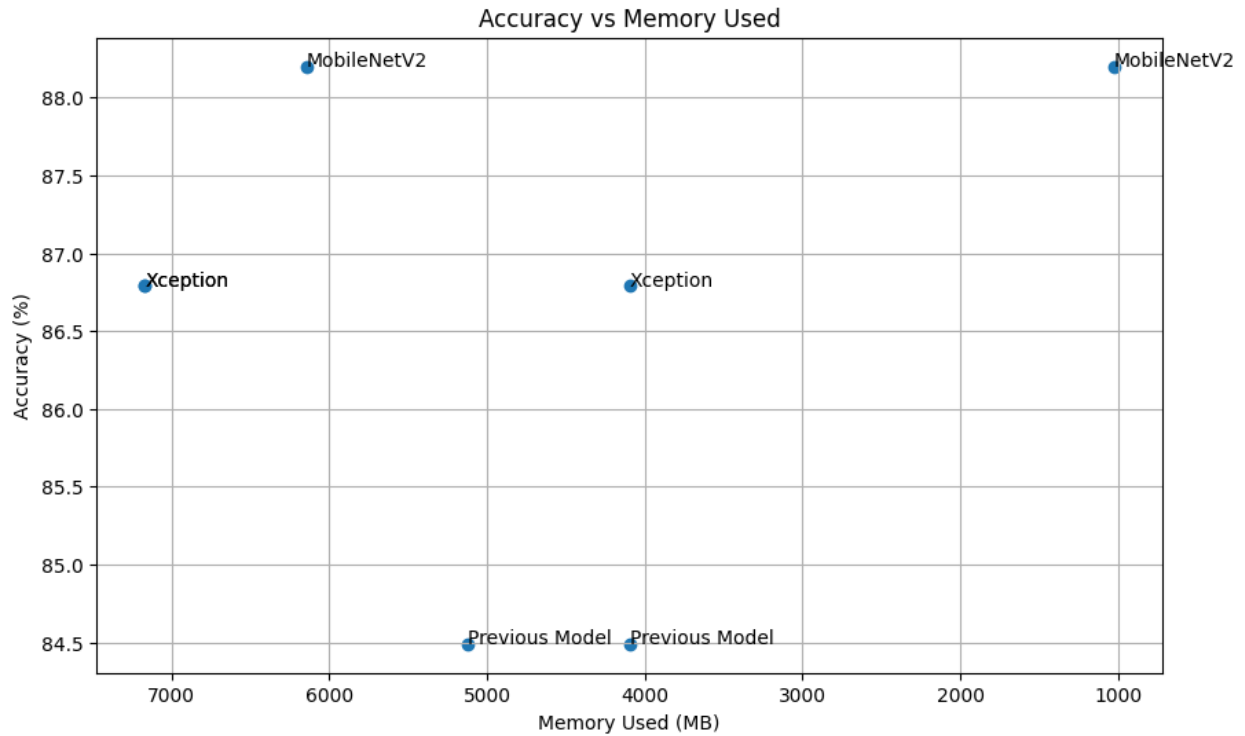




3. Deployments

Model	CPU Limit	RAM Limit(GB)	CPU Utilization Threshold(%)	Average Response Time(sec)	Average Transaction Rate
Previous Model	2	4	40	0.95	6.45
Previous Model	4	5	60	0.6	8.72
Xception	2	4	60	0.43	11.86
Xception	9	7	75	0.33	12.03
Xception	6	7	75	0.38	11.88
MobileNetV2	3	6	75	0.25	14.41
MobileNetV2	1	1	80	0.61	9.79





MobileNetV2 performs 1.5 times faster than the Baseline Model with half the required resources.

For Systems that are expected to receive a high footfall on their platforms to utilize this ML service, a CPU Limit of 3 and RAM of 6GB is good enough to scale the high number of incoming requests without failure with the **MobileNetV2 model**.

If there are budget constraints within the organization, 1 CPU and 1 GB RAM would be sufficient to meet this criterion, as it is still better than the original model with more resources, making it a robust cost-effective option to go for.

3. Conclusion

- MobileNetV2 performs well in all the metrics discussed above, thus becoming our model of choice. The setup with 1 CPU and 1GB RAM is a very attractive option to opt for.

Appendix A: Saved models

Base: <https://drive.google.com/file/d/1DjeFpMhxvrlKKe71JjCwOZ3vrhfDUQcf/view?usp=sharing>

Xception: <https://drive.google.com/file/d/10BLfNZGCinC3rbqMeTiFRcC-s-gVPMYB/view?usp=sharing>

MobileNetV2: <https://drive.google.com/file/d/1zOILNpXNmGAUcuHVBknrWUQPgZ9Q3VCD/view?usp=sharing>

Appendix B: Deployment files

Model	Yaml Link
Previous Model	https://drive.google.com/file/d/1ZkRU0YYPAX2O38vxwNLYVW9yy3nvUj4h/view?usp=sharing
Previous Model	https://drive.google.com/file/d/1_Y8Rz4F8OKg_7_tgX_1oBk1n23i1avcx/view?usp=sharing
Xception	https://drive.google.com/file/d/1Ebux8E9FMaaPmu89aHIUby8WDO-Tuhq7/view?usp=sharing
Xception	https://drive.google.com/file/d/1cGsbY-SHQY5kCEyBK7BbM9vikECWPWsD/view?usp=sharing
Xception	https://drive.google.com/file/d/1zJdNg4dJIB7sJZcJnO_nO1vRuEb1zTn-/view?usp=sharing
MobileNetV2	https://drive.google.com/file/d/1IT2X1-45ntFX7dUk2yr1dmyOrYzo9Jb7/view?usp=sharing
MobileNetV2	https://drive.google.com/file/d/1jlnLFrr-r-2fotGz5goJLGU-HQcNNggy/view?usp=sharing

