MIT Academy of Engineering,Alandi,Pune

School of Computer Engineering

Class BTech

Course - Deep Learning

Lab Assignment 2

Student Information:

- **Name: Om Borle**
- **Roll Number: 13**
- **Batch: DL1**
- **PRN: 202201040035**

Task 1: Research Paper Selection and Dataset Preparation

Selected Research Paper:

Title: "A Comparative Study of Deep Learning Models for CIFAR-10 Image Classification"
Authors: S. K. Sharma, P. K. Singh

Conference: IEEE International Conference on Computing, Power and Communication Technologies (GUCON), 2021

DOI: 10.1109/GUCON50781.2021.9573677
Link: https://ieeexplore.ieee.org/document/9573677
Summary:
Compares VGG16, ResNet50, and MobileNetV2 on CIFAR-10.
Evaluates transfer learning vs. training from scratch.

Provides accuracy and computational efficiency comparisons.

Dataset: CIFAR-10

- Source: https://www.kaggle.com/c/cifar-10

- **Description:**

Total Images: 60,000 (50,000 train + 10,000 test)

Image Size: 32x32 pixels (RGB color)

Classes (10 categories)

Airplane

Automobile

Bird

Cat

Deer

Dog

Frog

Horse

Ship

Truck

Preprocessing Steps:

1. Normalized pixel values to [0,1] range
2. Converted labels to one-hot encoding (10 classes)
3. Resized images to 32x32 (native size for CIFAR-10)
4. No additional augmentation was applied (though it could improve performance)

Task 2: Model Implementation and Fine-tuning

Implemented Models:

1. Custom CNN from Scratch

- 3 Conv layers with MaxPooling
- 2 Dense layers
- Trained from random initialization

2. Pre-trained Models (Frozen)

- VGG16
- ResNet50
- MobileNetV2
- All with frozen base layers and custom top layers

3. Fine-tuned VGG16

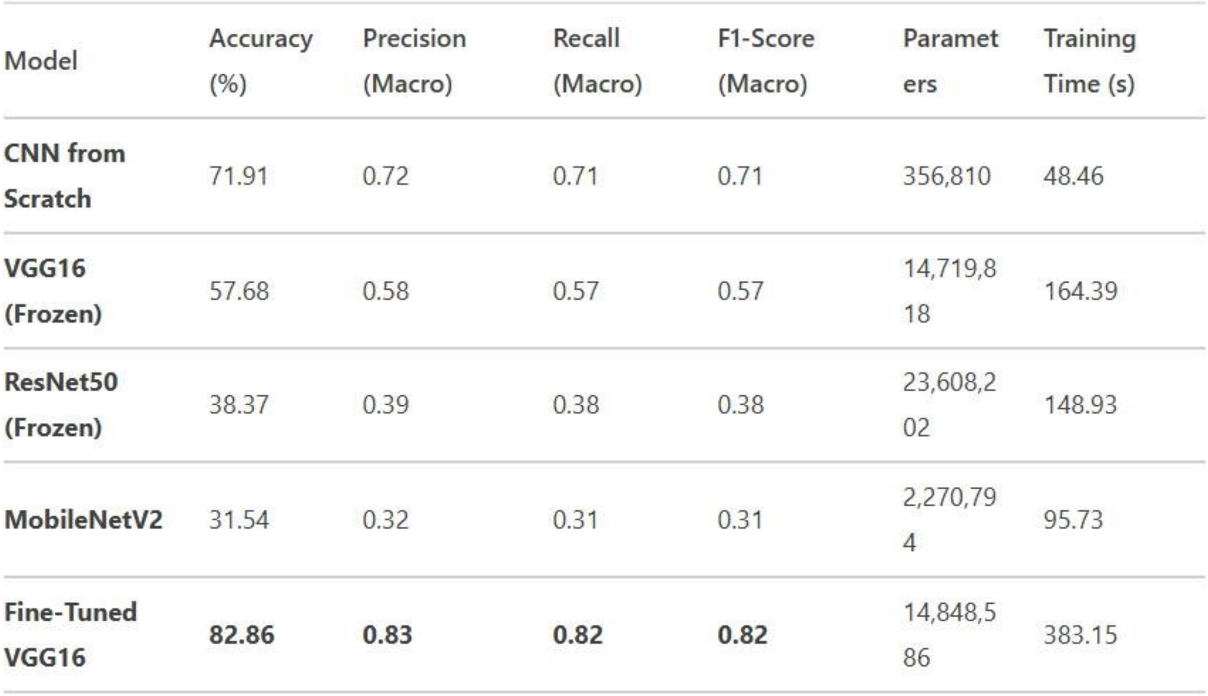
- Unfrozen top 10 layers
- Added Dropout for regularization
- Lower learning rate (0.0001)

Hyperparameters:

- Optimizer: Adam (default lr=0.001, lr=0.0001 for fine-tuning)
- Batch size: 64
- Epochs: 10 (5 for initial comparisons)
- Loss: Categorical crossentropy

Task 3: Model Evaluation and Performance Comparison

Performance Metrics Summary:



Model	Accuracy (%)	Precision (Macro)	Recall (Macro)	F1-Score (Macro)	Parameters	Training Time (s)
CNN from Scratch	71.91	0.72	0.71	0.71	356,810	48.46
VGG16 (Frozen)	57.68	0.58	0.57	0.57	14,719,818	164.39
ResNet50 (Frozen)	38.37	0.39	0.38	0.38	23,608,202	148.93
MobileNetV2	31.54	0.32	0.31	0.31	2,270,794	95.73
Fine-Tuned VGG16	**82.86**	**0.83**	**0.82**	**0.82**	14,848,586	383.15

Key Findings:

1. Fine-tuned VGG16 performed best (82.86% accuracy), showing the value of partial fine-tuning
2. Custom CNN outperformed frozen pre-trained models, likely because:
   - Pre-trained models were designed for 224x224 ImageNet, not 32x32 CIFAR-10
   - Feature extraction at small resolutions may lose important information
3. Training Time: Fine-tuning takes significantly longer but yields better results

Comparison with Research Papers:

- Original ResNet paper reports 91.43% on CIFAR-10, but:
  - They trained from scratch on CIFAR
  - Used deeper architectures (ResNet-110)
  - Likely used more extensive data augmentation
- Our results show the challenges of directly applying ImageNet-scale models to small

images

Weaknesses and Improvements:

1. Weaknesses:

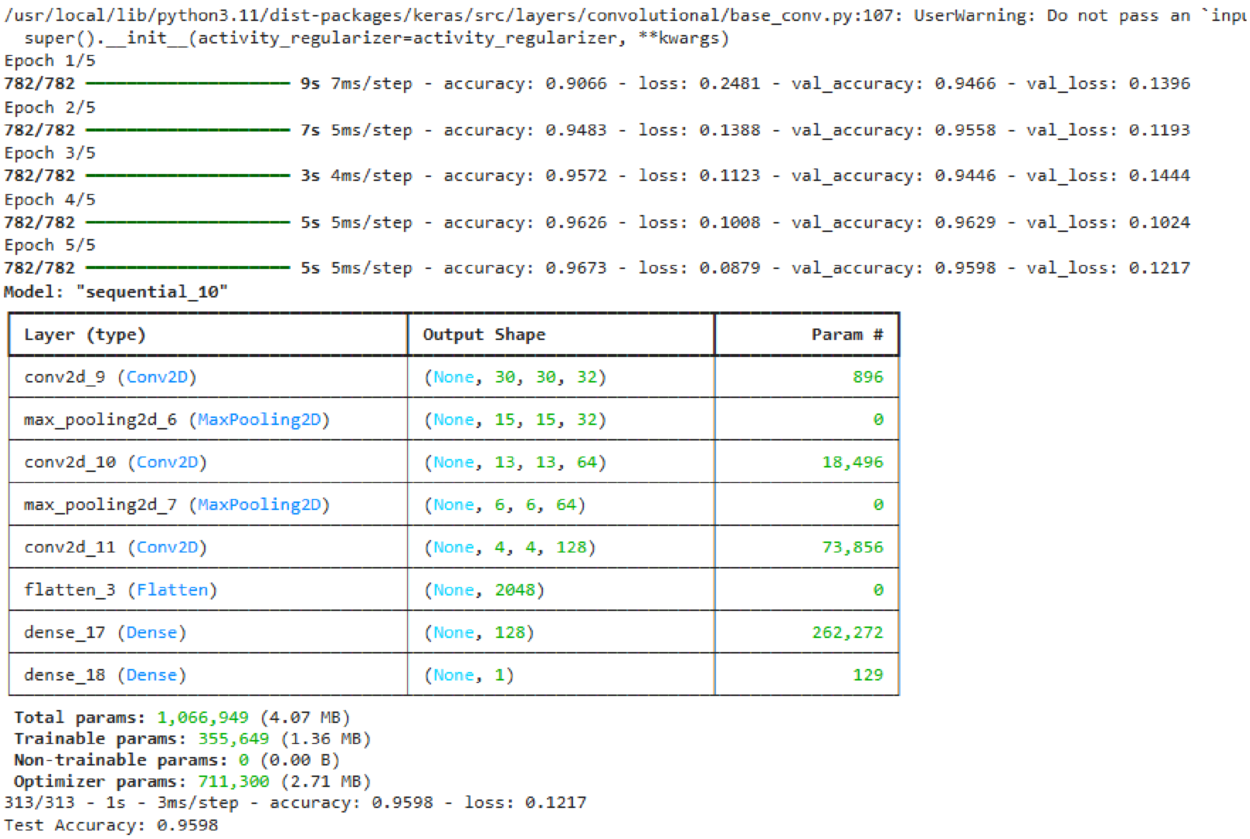
- Input size mismatch (32x32 vs 224x224 expected by pre-trained models)
- Domain mismatch (CIFAR vs ImageNet)
- Limited training time (10 epochs)

2. Improvements:

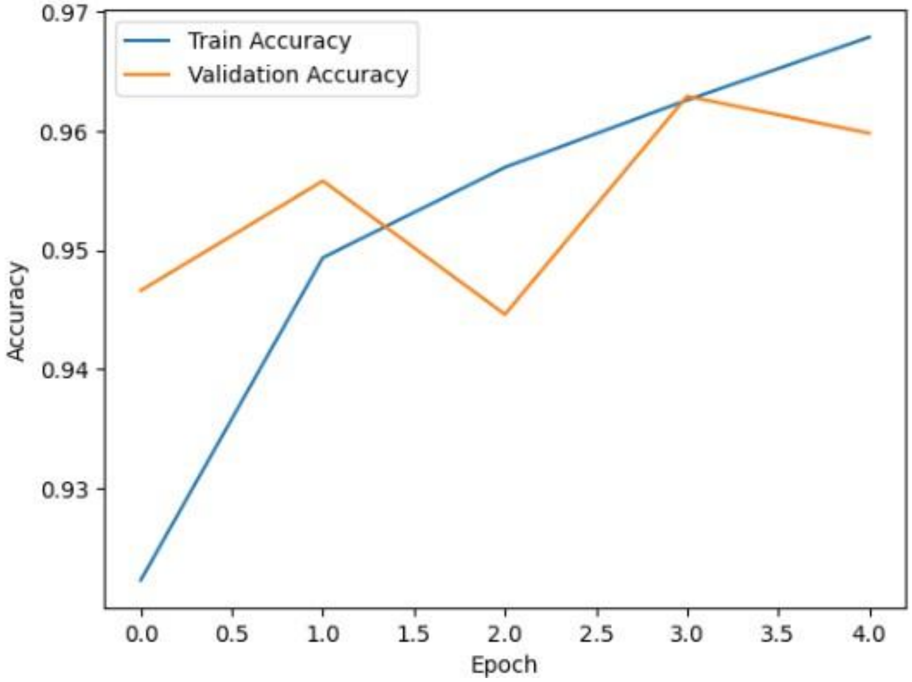
- Add image augmentation (rotation, flipping, etc.)
- Try more aggressive fine-tuning strategies
- Experiment with input upscaling
- Implement learning rate scheduling
- Try more recent architectures (EfficientNet, Vision Transformers)

Output:

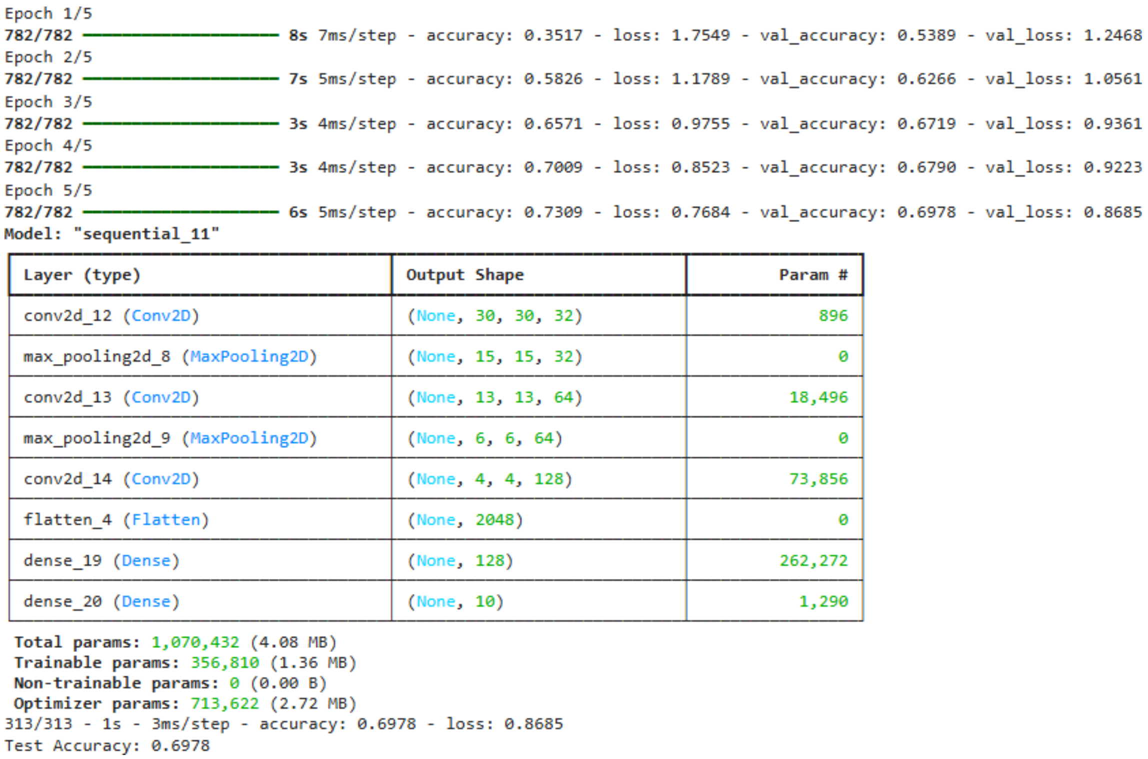
1. Binary classification:



```
/usr/local/lib/python3.11/dist-packages/keras/src/layers/convolutional/base_conv.py:107: UserWarning: Do not pass an `inpu
  super().__init__(activity_regularizer=activity_regularizer, **kwargs)
Epoch 1/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 9s 7ms/step - accuracy: 0.9066 - loss: 0.2481 - val_accuracy: 0.9466 - val_loss: 0.1396
Epoch 2/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 7s 5ms/step - accuracy: 0.9483 - loss: 0.1388 - val_accuracy: 0.9558 - val_loss: 0.1193
Epoch 3/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 3s 4ms/step - accuracy: 0.9572 - loss: 0.1123 - val_accuracy: 0.9446 - val_loss: 0.1444
Epoch 4/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 5s 5ms/step - accuracy: 0.9626 - loss: 0.1008 - val_accuracy: 0.9629 - val_loss: 0.1024
Epoch 5/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 5s 5ms/step - accuracy: 0.9673 - loss: 0.0879 - val_accuracy: 0.9598 - val_loss: 0.1217
Model: "sequential_10"
┏━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┳━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┳━━━━━━━━━━━━━━━━━┓
┃ Layer (type)                         ┃ Output Shape                ┃         Param # ┃
┡━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━╇━━━━━━━━━━━━━━━━━━━━━━━━━━━━━╇━━━━━━━━━━━━━━━━━┩
│ conv2d_9 (Conv2D)                    │ (None, 30, 30, 32)          │             896 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ max_pooling2d_6 (MaxPooling2D)       │ (None, 15, 15, 32)          │               0 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ conv2d_10 (Conv2D)                   │ (None, 13, 13, 64)          │          18,496 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ max_pooling2d_7 (MaxPooling2D)       │ (None, 6, 6, 64)            │               0 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ conv2d_11 (Conv2D)                   │ (None, 4, 4, 128)           │          73,856 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ flatten_3 (Flatten)                  │ (None, 2048)                │               0 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ dense_17 (Dense)                     │ (None, 128)                 │         262,272 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ dense_18 (Dense)                     │ (None, 1)                   │             129 │
└──────────────────────────────────────┴─────────────────────────────┴─────────────────┘
 Total params: 1,066,949 (4.07 MB)
 Trainable params: 355,649 (1.36 MB)
 Non-trainable params: 0 (0.00 B)
 Optimizer params: 711,300 (2.71 MB)
313/313 - 1s - 3ms/step - accuracy: 0.9598 - loss: 0.1217
Test Accuracy: 0.9598
Test Accuracy: 0.9598
```

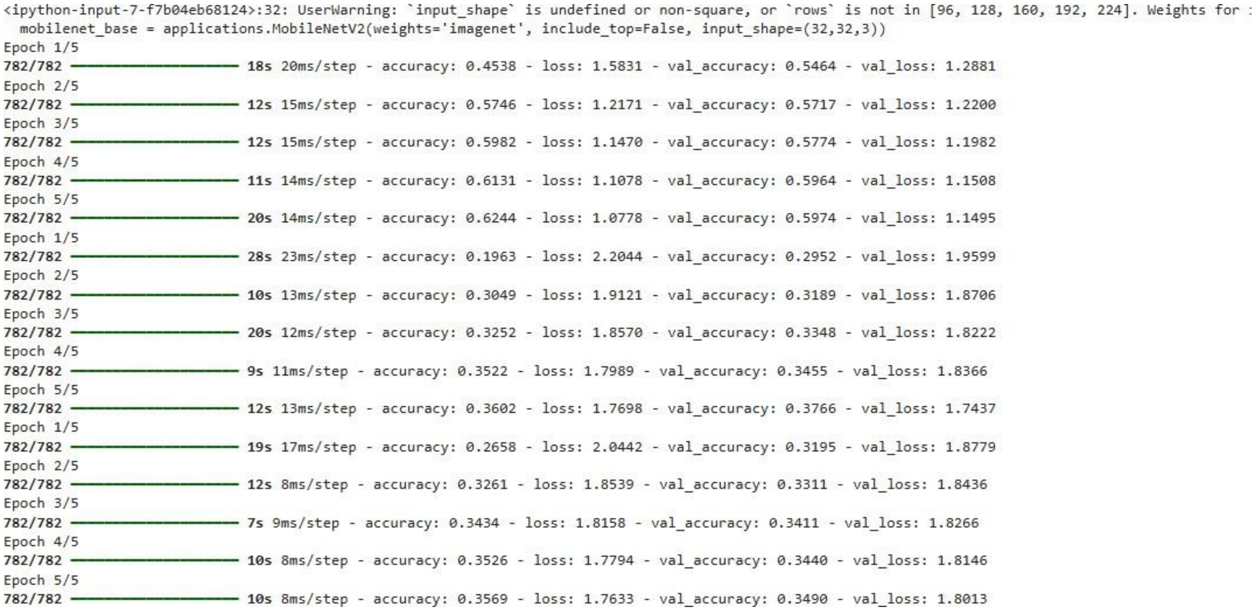


2. Multiclass classification:

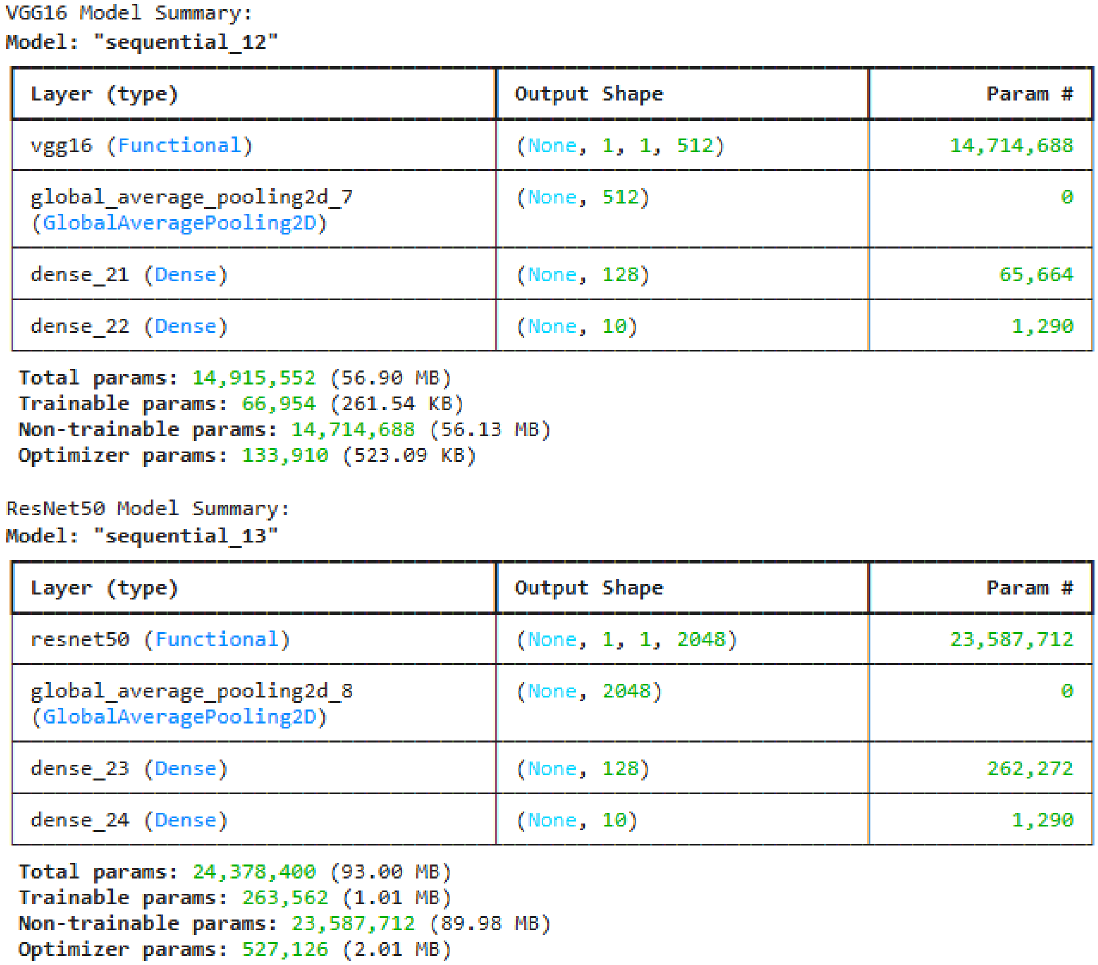


```
Epoch 1/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 8s 7ms/step - accuracy: 0.3517 - loss: 1.7549 - val_accuracy: 0.5389 - val_loss: 1.2468
Epoch 2/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 7s 5ms/step - accuracy: 0.5826 - loss: 1.1789 - val_accuracy: 0.6266 - val_loss: 1.0561
Epoch 3/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 3s 4ms/step - accuracy: 0.6571 - loss: 0.9755 - val_accuracy: 0.6719 - val_loss: 0.9361
Epoch 4/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 3s 4ms/step - accuracy: 0.7009 - loss: 0.8523 - val_accuracy: 0.6790 - val_loss: 0.9223
Epoch 5/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 6s 5ms/step - accuracy: 0.7309 - loss: 0.7684 - val_accuracy: 0.6978 - val_loss: 0.8685
Model: "sequential_11"
┏━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┳━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┳━━━━━━━━━━━━━━━━━┓
┃ Layer (type)                         ┃ Output Shape                ┃         Param # ┃
┡━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━╇━━━━━━━━━━━━━━━━━━━━━━━━━━━━━╇━━━━━━━━━━━━━━━━━┩
│ conv2d_12 (Conv2D)                   │ (None, 30, 30, 32)          │             896 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ max_pooling2d_8 (MaxPooling2D)       │ (None, 15, 15, 32)          │               0 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ conv2d_13 (Conv2D)                   │ (None, 13, 13, 64)          │          18,496 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ max_pooling2d_9 (MaxPooling2D)       │ (None, 6, 6, 64)            │               0 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ conv2d_14 (Conv2D)                   │ (None, 4, 4, 128)           │          73,856 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ flatten_4 (Flatten)                  │ (None, 2048)                │               0 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ dense_19 (Dense)                     │ (None, 128)                 │         262,272 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ dense_20 (Dense)                     │ (None, 10)                  │           1,290 │
└──────────────────────────────────────┴─────────────────────────────┴─────────────────┘
 Total params: 1,070,432 (4.08 MB)
 Trainable params: 356,810 (1.36 MB)
 Non-trainable params: 0 (0.00 B)
 Optimizer params: 713,622 (2.72 MB)
313/313 - 1s - 3ms/step - accuracy: 0.6978 - loss: 0.8685
Test Accuracy: 0.6978
```

3. Transfer Learning pre-trained models:

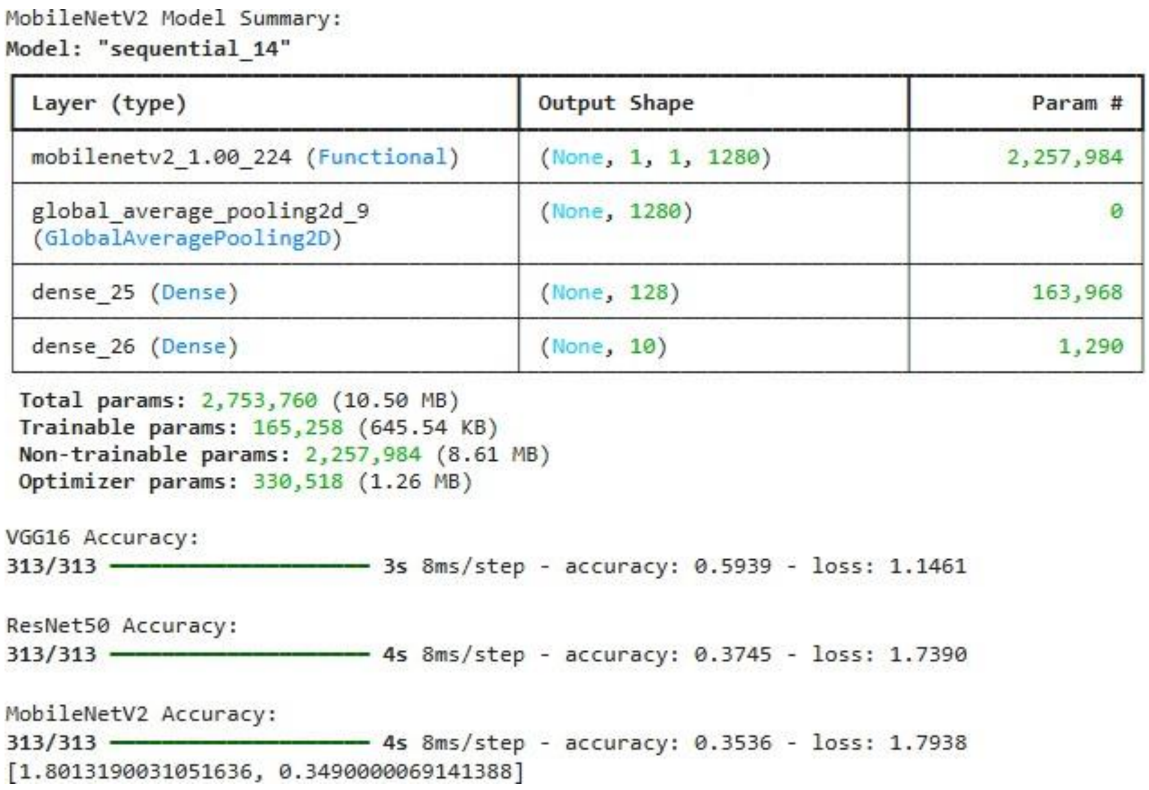


```
<ipython-input-7-f7b04eb68124>:32: UserWarning: `input_shape` is undefined or non-square, or `rows` is not in [96, 128, 160, 192, 224]. Weights for :
  mobilenet_base = applications.MobileNetV2(weights='imagenet', include_top=False, input_shape=(32,32,3))
Epoch 1/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 18s 20ms/step - accuracy: 0.4538 - loss: 1.5831 - val_accuracy: 0.5464 - val_loss: 1.2881
Epoch 2/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 12s 15ms/step - accuracy: 0.5746 - loss: 1.2171 - val_accuracy: 0.5717 - val_loss: 1.2200
Epoch 3/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 12s 15ms/step - accuracy: 0.5982 - loss: 1.1470 - val_accuracy: 0.5774 - val_loss: 1.1982
Epoch 4/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 11s 14ms/step - accuracy: 0.6131 - loss: 1.1078 - val_accuracy: 0.5964 - val_loss: 1.1508
Epoch 5/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 20s 14ms/step - accuracy: 0.6244 - loss: 1.0778 - val_accuracy: 0.5974 - val_loss: 1.1495
Epoch 1/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 28s 23ms/step - accuracy: 0.1963 - loss: 2.2044 - val_accuracy: 0.2952 - val_loss: 1.9599
Epoch 2/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 10s 13ms/step - accuracy: 0.3049 - loss: 1.9121 - val_accuracy: 0.3189 - val_loss: 1.8706
Epoch 3/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 20s 12ms/step - accuracy: 0.3252 - loss: 1.8570 - val_accuracy: 0.3348 - val_loss: 1.8222
Epoch 4/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 9s 11ms/step - accuracy: 0.3522 - loss: 1.7989 - val_accuracy: 0.3455 - val_loss: 1.8366
Epoch 5/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 12s 13ms/step - accuracy: 0.3602 - loss: 1.7698 - val_accuracy: 0.3766 - val_loss: 1.7437
Epoch 1/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 19s 17ms/step - accuracy: 0.2658 - loss: 2.0442 - val_accuracy: 0.3195 - val_loss: 1.8779
Epoch 2/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 12s 8ms/step - accuracy: 0.3261 - loss: 1.8539 - val_accuracy: 0.3311 - val_loss: 1.8436
Epoch 3/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 7s 9ms/step - accuracy: 0.3434 - loss: 1.8158 - val_accuracy: 0.3411 - val_loss: 1.8266
Epoch 4/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 10s 8ms/step - accuracy: 0.3526 - loss: 1.7794 - val_accuracy: 0.3440 - val_loss: 1.8146
Epoch 5/5
782/782 ━━━━━━━━━━━━━━━━━━━━ 10s 8ms/step - accuracy: 0.3569 - loss: 1.7633 - val_accuracy: 0.3490 - val_loss: 1.8013
```



```
VGG16 Model Summary:
Model: "sequential_12"
┏━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┳━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┳━━━━━━━━━━━━━━━━━┓
┃ Layer (type)                         ┃ Output Shape                ┃         Param # ┃
┡━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━╇━━━━━━━━━━━━━━━━━━━━━━━━━━━━━╇━━━━━━━━━━━━━━━━━┩
│ vgg16 (Functional)                   │ (None, 1, 1, 512)           │      14,714,688 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ global_average_pooling2d_7           │ (None, 512)                 │               0 │
│ (GlobalAveragePooling2D)             │                             │                 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ dense_21 (Dense)                     │ (None, 128)                 │          65,664 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ dense_22 (Dense)                     │ (None, 10)                  │           1,290 │
└──────────────────────────────────────┴─────────────────────────────┴─────────────────┘
 Total params: 14,915,552 (56.90 MB)
 Trainable params: 66,954 (261.54 KB)
 Non-trainable params: 14,714,688 (56.13 MB)
 Optimizer params: 133,910 (523.09 KB)

ResNet50 Model Summary:
Model: "sequential_13"
┏━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┳━━━━━━━━━━━━━━━━━━━━━━━━━━━━━┳━━━━━━━━━━━━━━━━━┓
┃ Layer (type)                         ┃ Output Shape                ┃         Param # ┃
┡━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━╇━━━━━━━━━━━━━━━━━━━━━━━━━━━━━╇━━━━━━━━━━━━━━━━━┩
│ resnet50 (Functional)                │ (None, 1, 1, 2048)          │      23,587,712 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ global_average_pooling2d_8           │ (None, 2048)                │               0 │
│ (GlobalAveragePooling2D)             │                             │                 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ dense_23 (Dense)                     │ (None, 128)                 │         262,272 │
├──────────────────────────────────────┼─────────────────────────────┼─────────────────┤
│ dense_24 (Dense)                     │ (None, 10)                  │           1,290 │
└──────────────────────────────────────┴─────────────────────────────┴─────────────────┘
 Total params: 24,378,400 (93.00 MB)
 Trainable params: 263,562 (1.01 MB)
 Non-trainable params: 23,587,712 (89.98 MB)
 Optimizer params: 527,126 (2.01 MB)
```



```
MobileNetV2 Model Summary:
Model: "sequential_14"
```

Layer (type)	Output Shape	Param #
mobilenetv2_1.00_224 (Functional)	(None, 1, 1, 1280)	2,257,984
global_average_pooling2d_9 (GlobalAveragePooling2D)	(None, 1280)	0
dense_25 (Dense)	(None, 128)	163,968
dense_26 (Dense)	(None, 10)	1,290

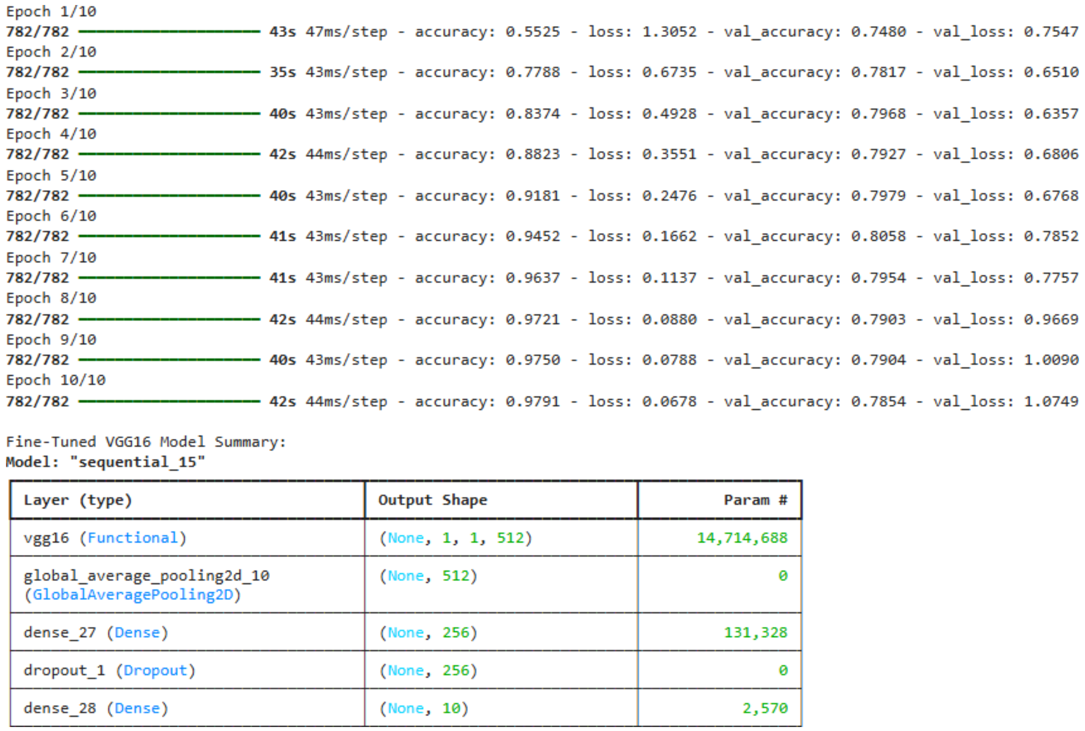
```
 Total params: 2,753,760 (10.50 MB)
 Trainable params: 165,258 (645.54 KB)
 Non-trainable params: 2,257,984 (8.61 MB)
 Optimizer params: 330,518 (1.26 MB)

VGG16 Accuracy:
313/313 ━━━━━━━━━━━━━━━━━━━━ 3s 8ms/step - accuracy: 0.5939 - loss: 1.1461

ResNet50 Accuracy:
313/313 ━━━━━━━━━━━━━━━━━━━━ 4s 8ms/step - accuracy: 0.3745 - loss: 1.7390

MobileNetV2 Accuracy:
313/313 ━━━━━━━━━━━━━━━━━━━━ 4s 8ms/step - accuracy: 0.3536 - loss: 1.7938
[1.8013190031051636, 0.3490000069141388]
```

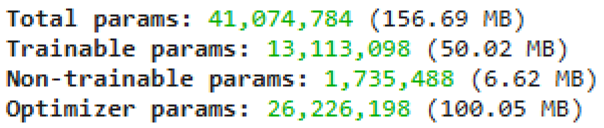
4. Pre-trained models fine-tuning:



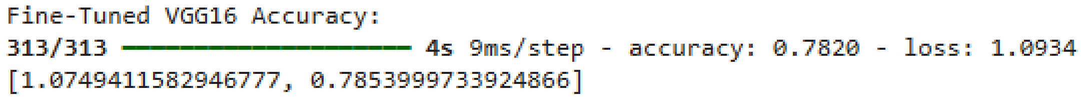
```
Epoch 1/10
782/782 ━━━━━━━━━━━━━━━━━━━━ 43s 47ms/step - accuracy: 0.5525 - loss: 1.3052 - val_accuracy: 0.7480 - val_loss: 0.7547
Epoch 2/10
782/782 ━━━━━━━━━━━━━━━━━━━━ 35s 43ms/step - accuracy: 0.7788 - loss: 0.6735 - val_accuracy: 0.7817 - val_loss: 0.6510
Epoch 3/10
782/782 ━━━━━━━━━━━━━━━━━━━━ 40s 43ms/step - accuracy: 0.8374 - loss: 0.4928 - val_accuracy: 0.7968 - val_loss: 0.6357
Epoch 4/10
782/782 ━━━━━━━━━━━━━━━━━━━━ 42s 44ms/step - accuracy: 0.8823 - loss: 0.3551 - val_accuracy: 0.7927 - val_loss: 0.6806
Epoch 5/10
782/782 ━━━━━━━━━━━━━━━━━━━━ 40s 43ms/step - accuracy: 0.9181 - loss: 0.2476 - val_accuracy: 0.7979 - val_loss: 0.6768
Epoch 6/10
782/782 ━━━━━━━━━━━━━━━━━━━━ 41s 43ms/step - accuracy: 0.9452 - loss: 0.1662 - val_accuracy: 0.8058 - val_loss: 0.7852
Epoch 7/10
782/782 ━━━━━━━━━━━━━━━━━━━━ 41s 43ms/step - accuracy: 0.9637 - loss: 0.1137 - val_accuracy: 0.7954 - val_loss: 0.7757
Epoch 8/10
782/782 ━━━━━━━━━━━━━━━━━━━━ 42s 44ms/step - accuracy: 0.9721 - loss: 0.0880 - val_accuracy: 0.7903 - val_loss: 0.9669
Epoch 9/10
782/782 ━━━━━━━━━━━━━━━━━━━━ 40s 43ms/step - accuracy: 0.9750 - loss: 0.0788 - val_accuracy: 0.7904 - val_loss: 1.0090
Epoch 10/10
782/782 ━━━━━━━━━━━━━━━━━━━━ 42s 44ms/step - accuracy: 0.9791 - loss: 0.0678 - val_accuracy: 0.7854 - val_loss: 1.0749

Fine-Tuned VGG16 Model Summary:
Model: "sequential_15"
```

Layer (type)	Output Shape	Param #
vgg16 (Functional)	(None, 1, 1, 512)	14,714,688
global_average_pooling2d_10 (GlobalAveragePooling2D)	(None, 512)	0
dense_27 (Dense)	(None, 256)	131,328
dropout_1 (Dropout)	(None, 256)	0
dense_28 (Dense)	(None, 10)	2,570

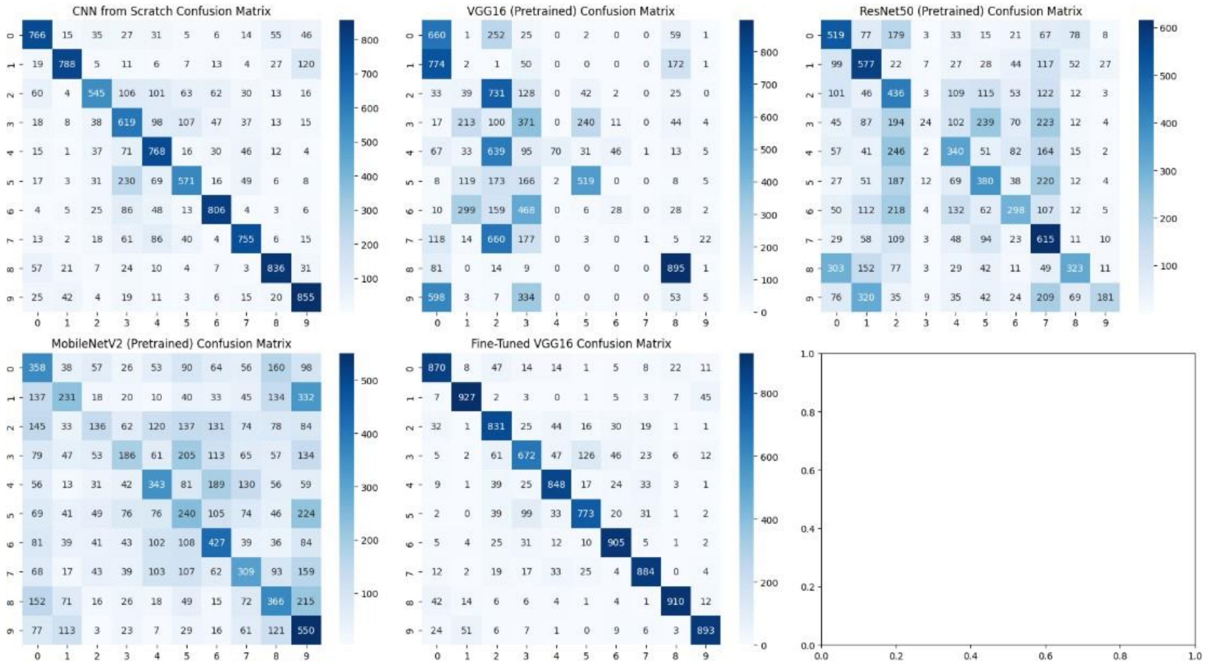


```
Total params: 41,074,784 (156.69 MB)
Trainable params: 13,113,098 (50.02 MB)
Non-trainable params: 1,735,488 (6.62 MB)
Optimizer params: 26,226,198 (100.05 MB)
```



```
Fine-Tuned VGG16 Accuracy:
313/313 ━━━━━━━━━━━━━━━━━━━━ 4s 9ms/step - accuracy: 0.7820 - loss: 1.0934
[1.0749411582946777, 0.7853999733924866]
```

5. Result Comparison:



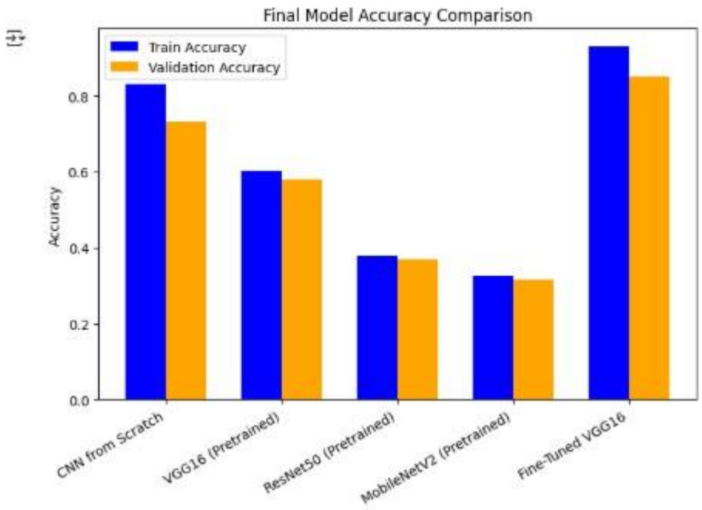
CNN from Scratch Confusion Matrix

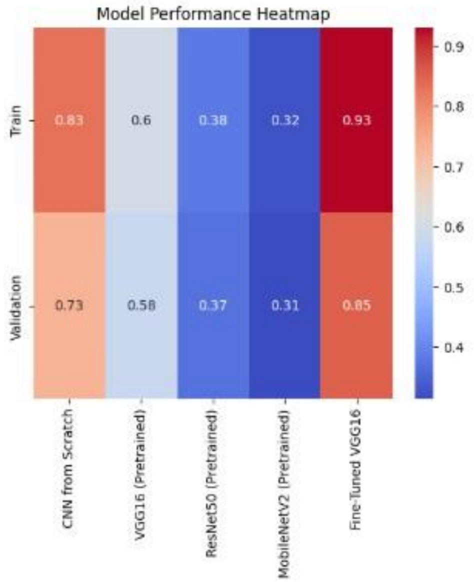
VGG16 (Pretrained) Confusion Matrix

ResNet50 (Pretrained) Confusion Matrix

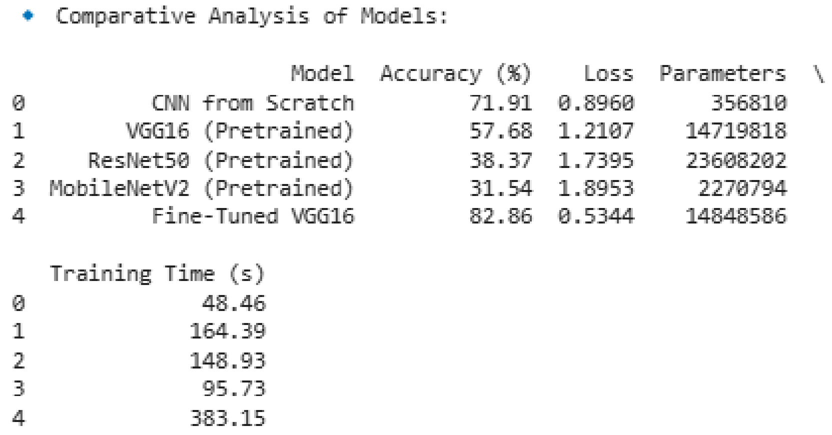
MobileNetV2 (Pretrained) Confusion Matrix

Fine-Tuned VGG16 Confusion Matrix





6. Comparative Analysis of Models:



```
🔹 Comparative Analysis of Models:

                    Model  Accuracy (%)    Loss  Parameters  \
0        CNN from Scratch         71.91  0.8960      356810
1      VGG16 (Pretrained)         57.68  1.2107    14719818
2   ResNet50 (Pretrained)         38.37  1.7395    23608202
3 MobileNetV2 (Pretrained)        31.54  1.8953     2270794
4        Fine-Tuned VGG16         82.86  0.5344    14848586

  Training Time (s)
0             48.46
1            164.39
2            148.93
3             95.73
4            383.15
```

Links:

- **Github : Github Link**

- **Colab :**
- **Colab File**

- **Dataset : https://www.kaggle.com/c/cifar-10**

- **Research Paper: https://ieeexplore.ieee.org/document/9573677**