ADVANCED MACHINE LEARNING -2
PROJECT REPORT

INTRODUCTION

In this study, we figure out the effects of network design and training sample size on image classification performance using a dataset of cats and dogs.
The goal of this project is to construct and contrast two models:

- A Convolutional Neural Network (CNN) trained from scratch
- The ImageNet dataset was used to pre-train a VGG16 transfer learning model.

The objective is to determine which network is better suited for various data situations and to comprehend how dataset size affects model performance.

DATASET DESCRIPTION

The dataset is originally taken from Kaggle Dogs vs Cats , But for this assignment I used the dataset provided on canvas which is smaller than compared to the Kaggle Dataset.

The Canvas version includes a smaller curated sample to make training possible with limited time and computational resources, whereas the full Kaggle dataset has 25,000 images (12,500 cats and 12,500 dogs).

I created three different training sample sizes:

- 1000 images per class
- 1500 images per class
- 2000 images per class

Validation and test sets were kept constant at 500 images per class each.
All images were resized to **150 X 150 pixels** before being fed into the network.

EXPERIMENTAL SETUP

Environment

- The models were implemented using **Google Colab** with **TensorFlow** and **Keras** libraries.
- GPU acceleration was enabled for faster training.

Preprocessing

- Images were rescaled (rescale=1./255) for normalization, converting pixel values from [0, 255] to [0, 1].
- **Data augmentation** was applied to training images to increase variability and reduce overfitting
- Rotation: ±40°
- Width and height shifts: up to 20%
- Shear: 0.2
- Zoom: 0.2
- Horizontal flip: True

Data Generators

- Keras ImageDataGenerator was used to automatically load and batch images from directory structures.

MODEL ARCHITECTURES

CNN Model (Trained from Scratch)

This model was built manually using the Keras Sequential API with:

- Four convolutional layers (with ReLU activations)
- MaxPooling layers for spatial reduction
- Dropout (0.5) for regularization
- Dense layers leading to a sigmoid output for binary classification

Configuration:

Optimizer: RMSprop (learning rate = 1e-4)

Loss Function: Binary Crossentropy

Metrics: Accuracy

VGG16 Model (Transfer Learning)

- VGG16 is a convolutional Neural Network architecture developed by the Visual Geometry Group (VGG) at the University of Oxford.
- It is called *VGG16* because it has 16 trainable layers — 13 convolutional layers and 3 fully connected layers.
- Unlike traditional CNNs, VGG16 uses many small 3×3 convolutional filters stacked deeper, which allows it to learn complex visual patterns efficiently.

Working Principle

- In this project, the pre-trained VGG16 model (trained on ImageNet) was used as a feature extractor.
- The top (fully connected) layers were removed, and new layers were added:
  Flatten → Dense(256, ReLU) → Dropout(0.5) → Dense(1, Sigmoid)
- This approach is known as Transfer Learning, where a model trained on a large dataset is fine-tuned for a smaller, related dataset**.**

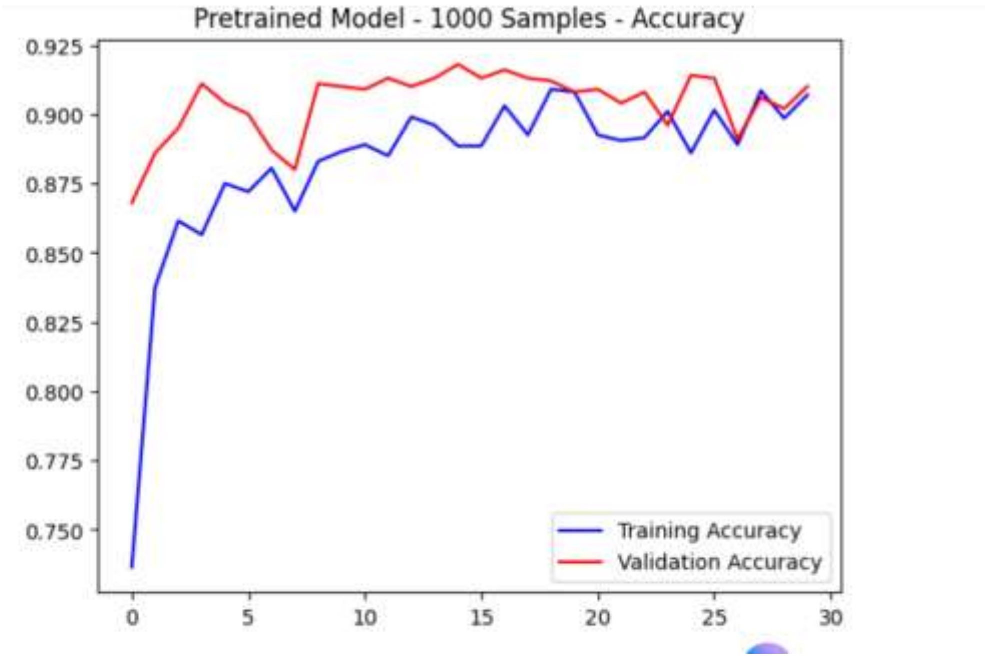
TRAINING DETAILS

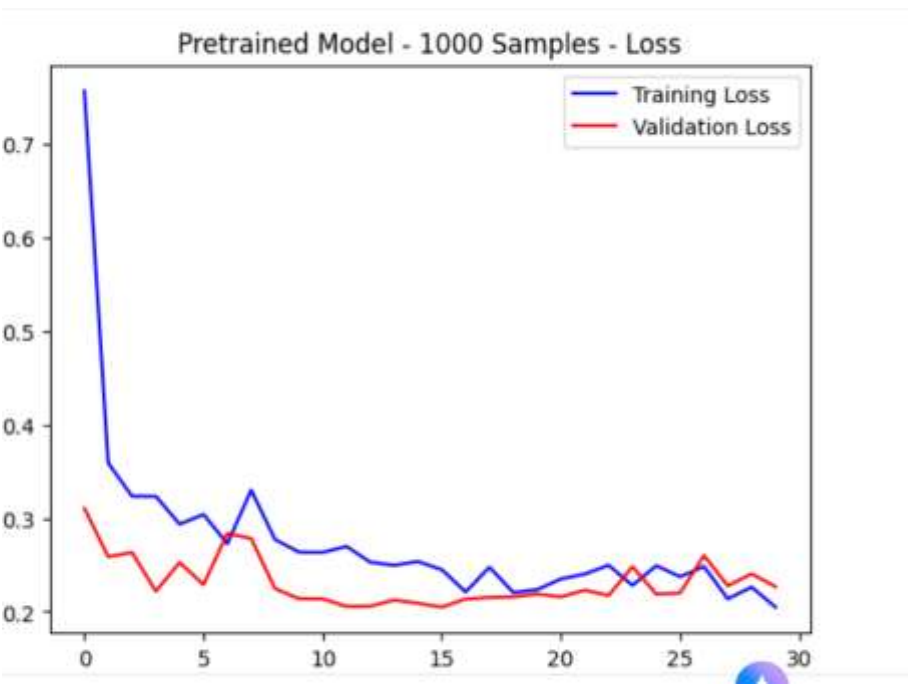
Each model — CNN (trained from scratch) and VGG16 (transfer learning) — was trained on three progressively larger subsets of the Canvas Cats vs Dogs dataset to study the effect of training sample size on performance.
The training subsets contained:

- 1000 images per class (2000 total) - small dataset
- 1500 images per class (3000 total) - medium dataset
- 2000 images per class (4000 total) - large dataset

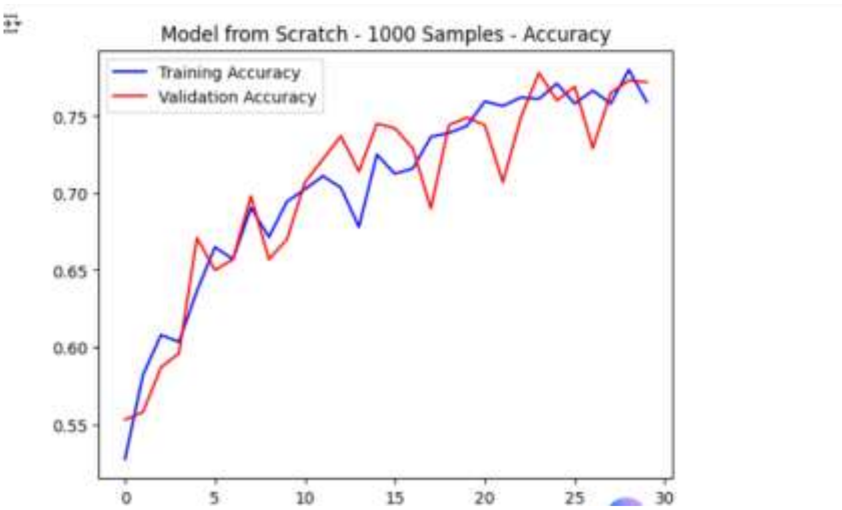
For each sample size, the validation and test sets were kept constant with 500 images per class, ensuring that performance comparisons were fair and consistent across all experiments.

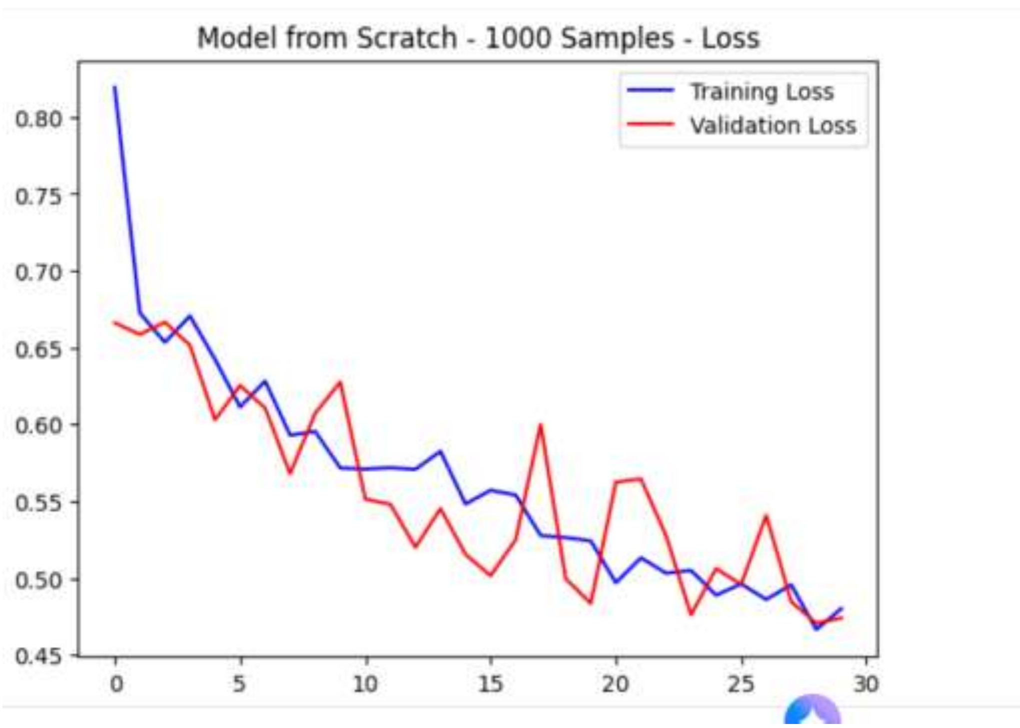
RESULTS SUMMARY
Experiment 1 – 1000 Images per Class (VGG16)



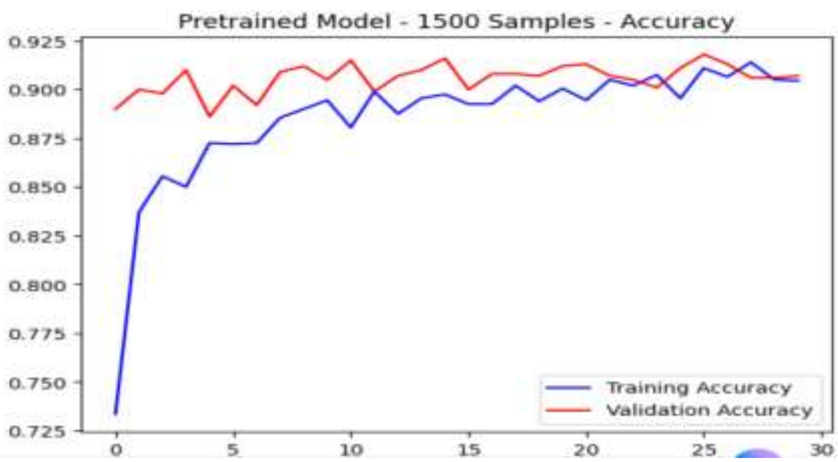


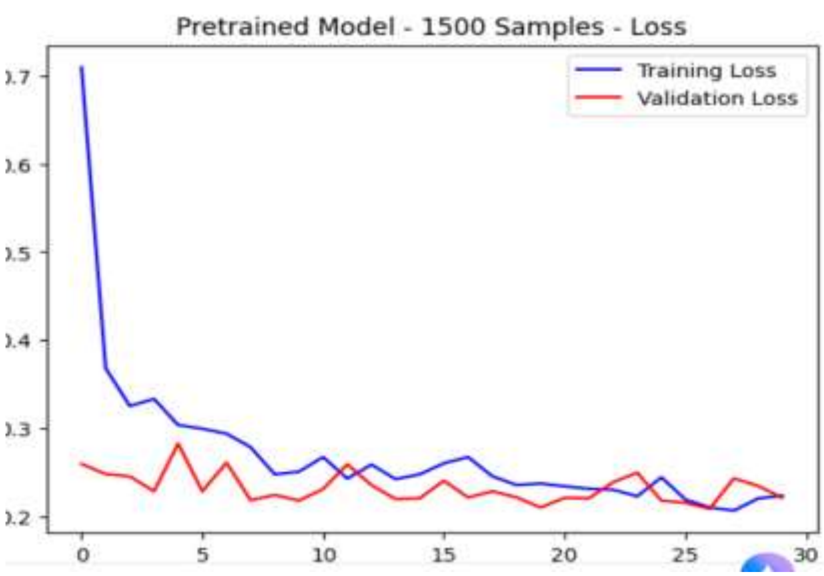
Experiment 2 – 1000 Images per Class (CNN from Scratch)



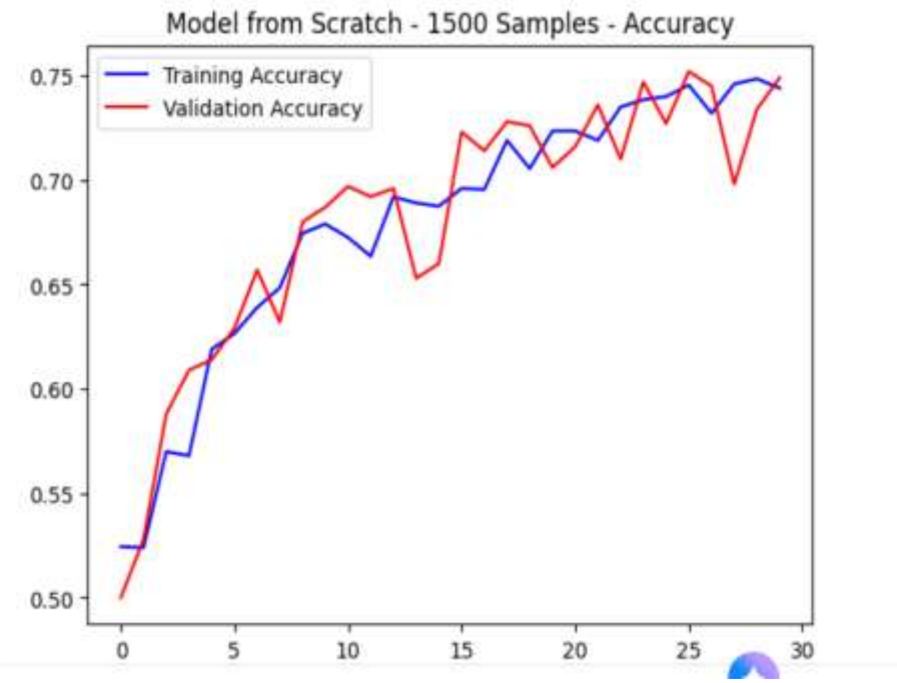


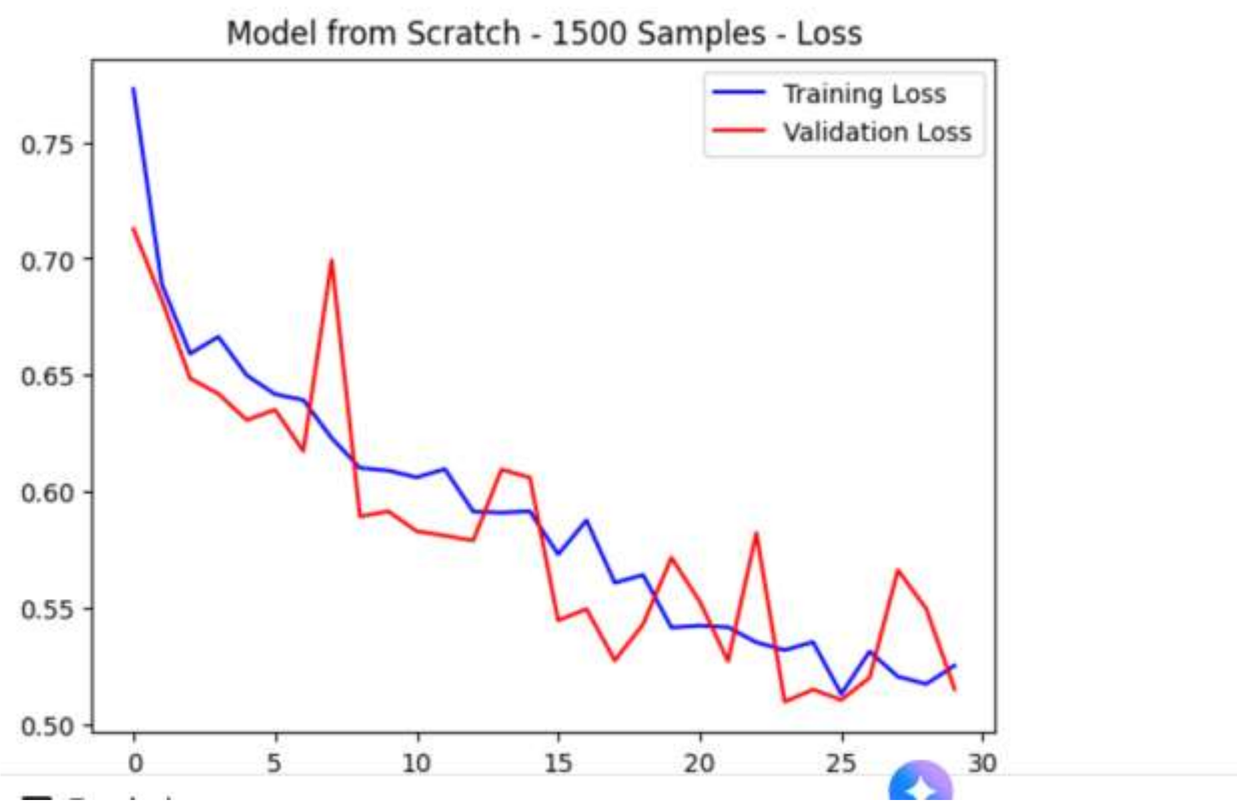
Experiment 3 – 1500 Images per Class (VGG16)



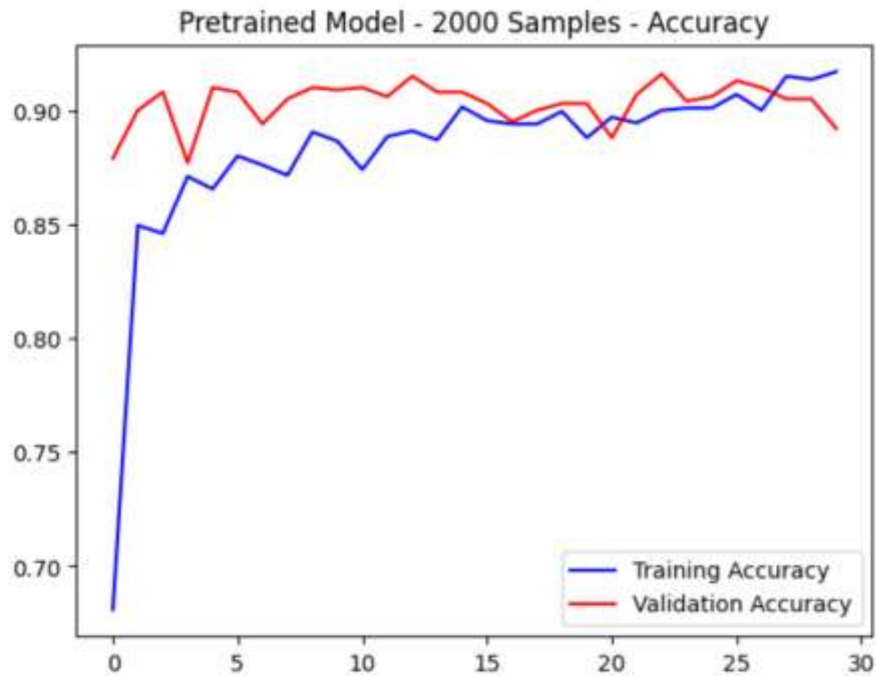


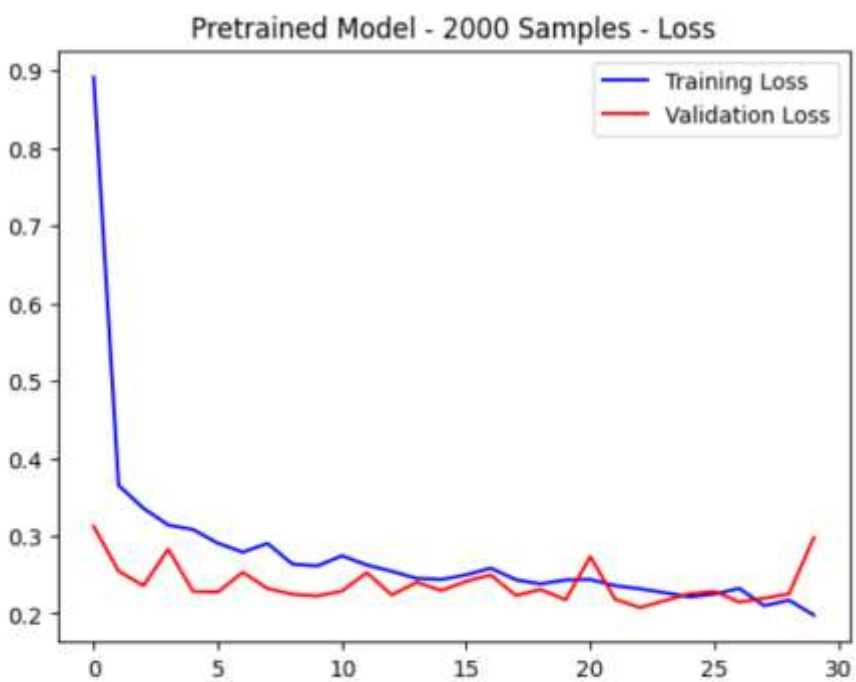
Experiment 4 – 1500 Images per Class (CNN from Scratch)



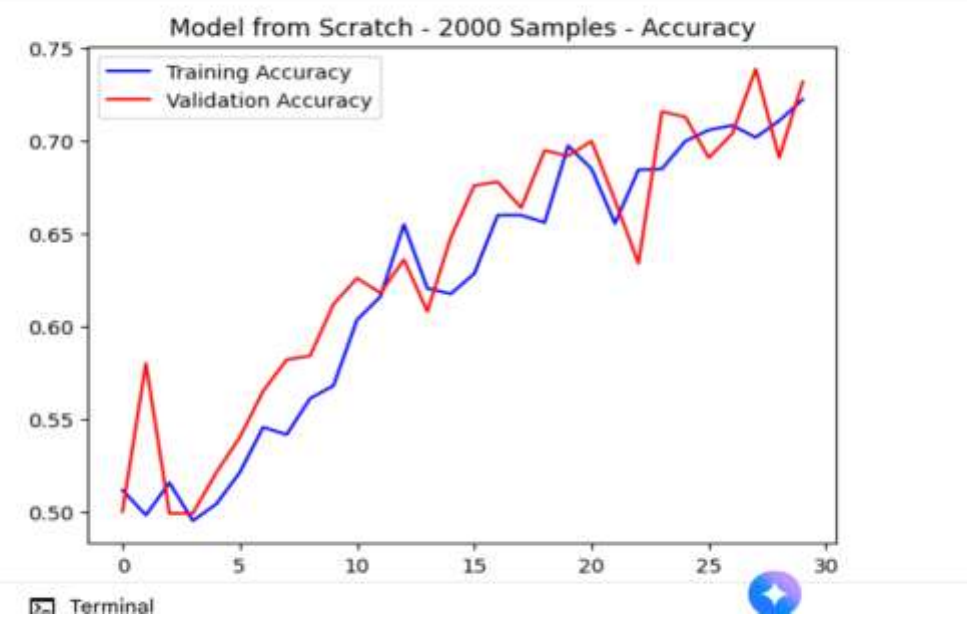


Experiment 5 – 2000 Images per Class (VGG16)





Experiment 6 – 2000 Images per Class (CNN from Scratch)



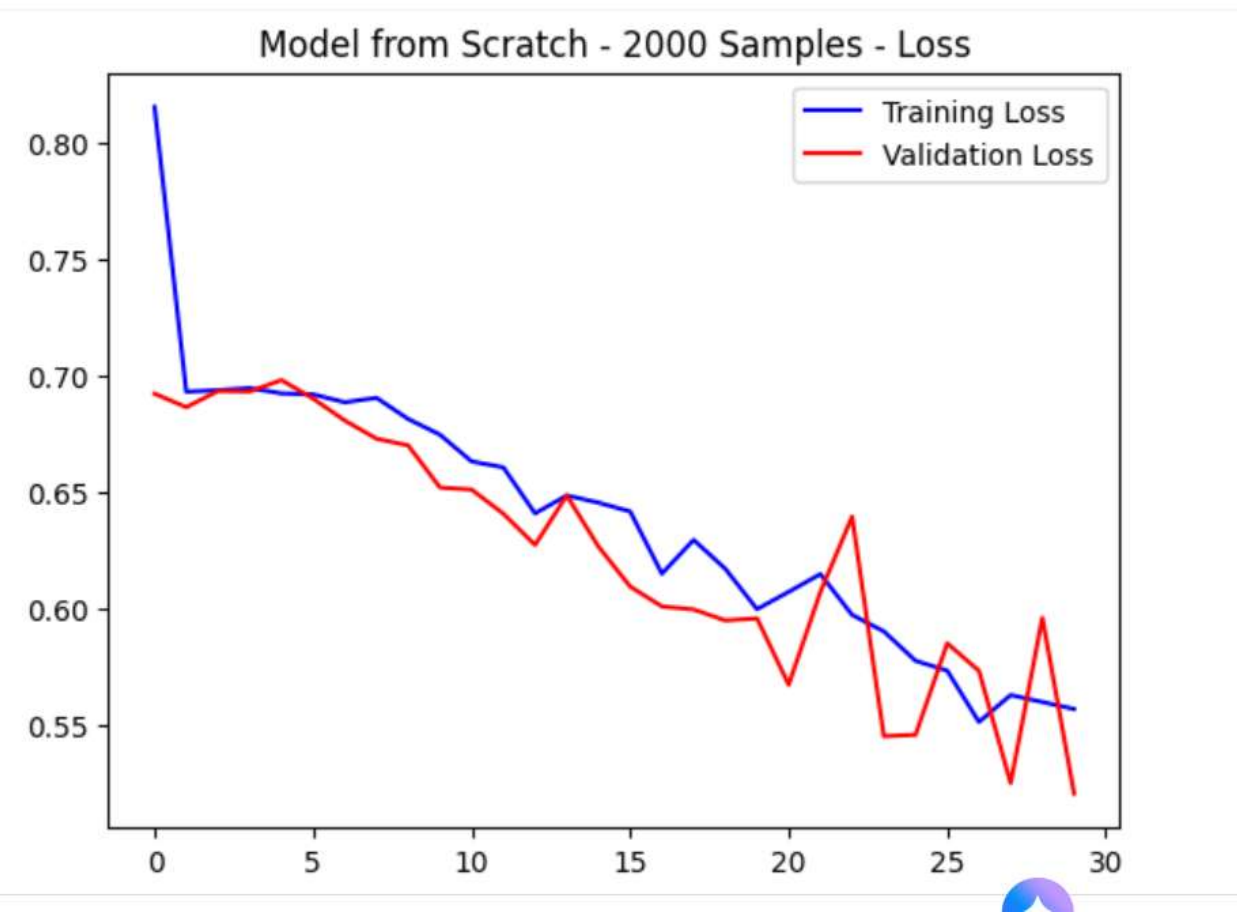
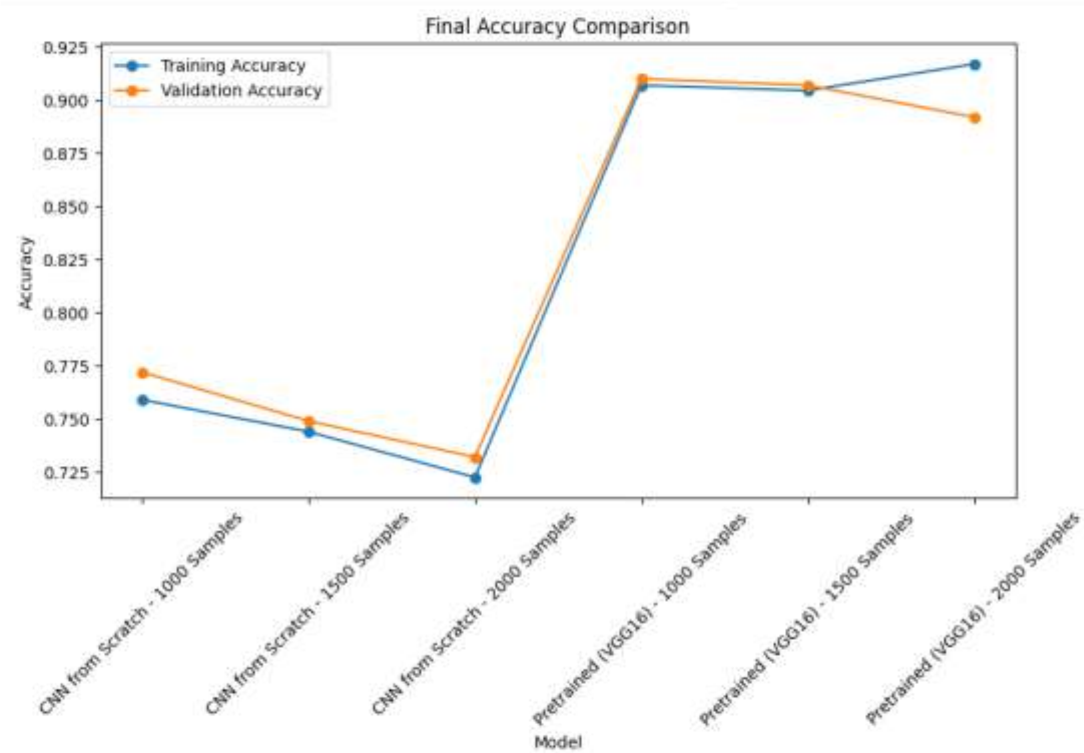


TABLE :

Training Sample Size	Model from Scratch - Training Accuracy	Model from Scratch - Validation Accuracy	Model from Scratch - Training Loss	Model from Scratch - Validation Loss	Pretrained Model - Training Accuracy	Pretrained Model - Validation Accuracy	Pretrained Model - Training Loss	Pretrained Model - Validation Loss
1000	0.7625	0.7480	0.4959	0.4952	0.9115	0.9130	0.2103	0.2318
1500	0.7065	0.7450	0.5788	0.5232	0.9075	0.9100	0.2180	0.2265
2000	0.7475	0.7800	0.4990	0.4756	0.8960	0.9130	0.2334	0.2114



FINAL SUMMARY:

The graphs show a significant difference in picture categorization performance between pre trained and scratch-trained models. Models that are trained from scratch show little progress even when the training sample size is increased, with poor accuracy (about 74–76%) and high loss (around 0.48–0.54). On the other hand, pre-trained models show the effectiveness of transfer learning by achieving high accuracy (over 90%) and low loss (around 0.20–0.25) even with fewer sample numbers. When compared to using pre-trained models, which converge more successfully and generalize significantly better on the "Cats vs. Dogs" dataset, this sharp difference emphasizes how ineffective training from scratch is. The pre-trained method uses less data to attain high performance and produces better outcomes.