BASIC CNN AND RESNET50 PERFORMANCE COMPARISON ON SMALL DATASET

Ege Kemal Germen

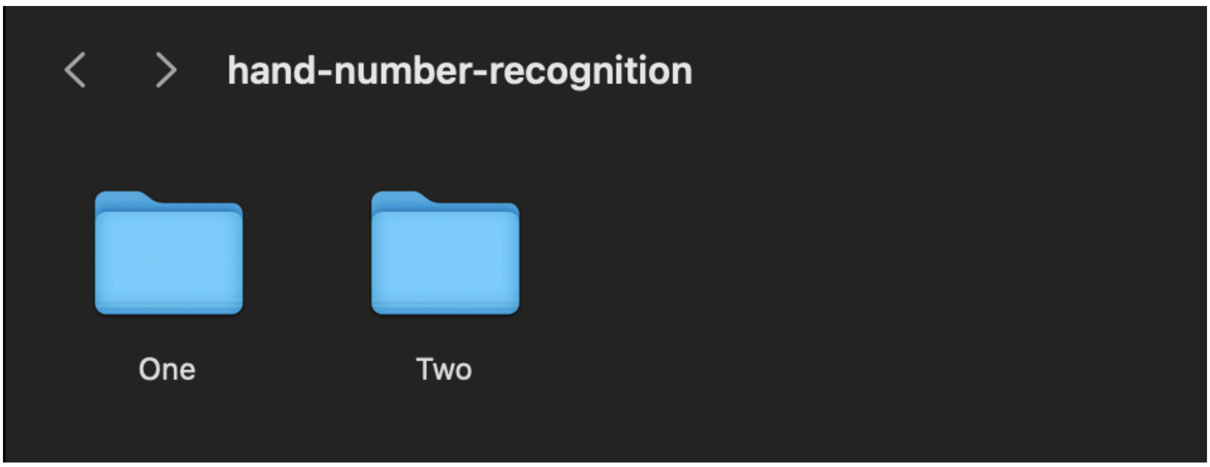
Dokuz Eylul University, Izmir, Turkey

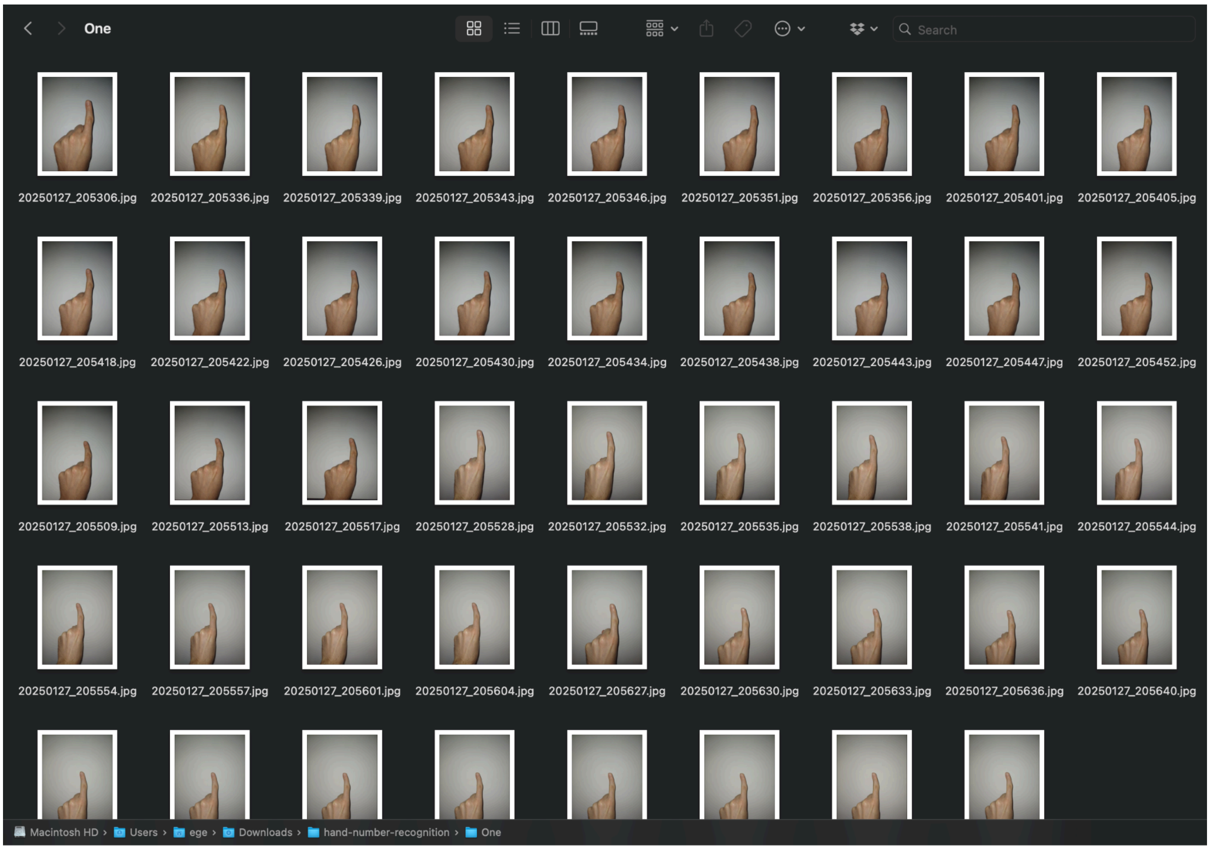
Introduction

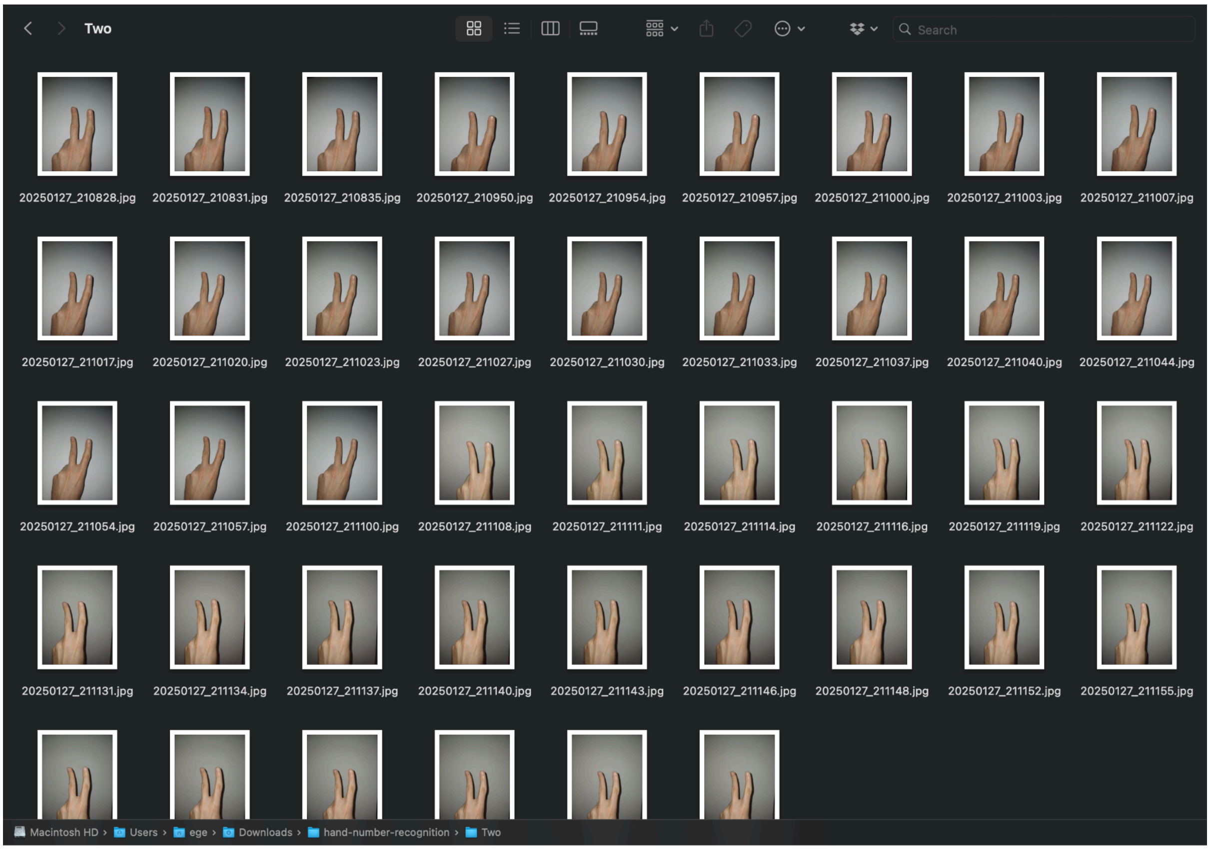
In this project we will compare two Convolutional Neural Networks' performance on small dataset: a basic 7-layer CNN model and the ResNet50 which were trained to recognize hand gestures representing the numbers 1 and 2.

Data Collection

The dataset for this project was created by manually gathering images corresponding to the two classes under study. Images were captured using a smartphone camera and organized into two folders, each representing a class. The images were preprocessed to have a consistent size and format suitable for input into both models. To improve the robustness of the dataset, data augmentation techniques such as rotation and flipping were applied, increasing the diversity of training samples.




One
Search
20250127_205306.jpg
20250127_205336.jpg
20250127_205339.jpg
20250127_205343.jpg
20250127_205346.jpg
20250127_205351.jpg
20250127_205356.jpg
20250127_205401.jpg
20250127_205405.jpg
20250127_205418.jpg
20250127_205422.jpg
20250127_205426.jpg
20250127_205430.jpg
20250127_205434.jpg
20250127_205438.jpg
20250127_205443.jpg
20250127_205447.jpg
20250127_205452.jpg
20250127_205509.jpg
20250127_205513.jpg
20250127_205517.jpg
20250127_205528.jpg
20250127_205532.jpg
20250127_205535.jpg
20250127_205538.jpg
20250127_205541.jpg
20250127_205544.jpg
20250127_205554.jpg
20250127_205557.jpg
20250127_205601.jpg
20250127_205604.jpg
20250127_205627.jpg
20250127_205630.jpg
20250127_205633.jpg
20250127_205636.jpg
20250127_205640.jpg
Macintosh HD › Users › ege › Downloads › hand-number-recognition › One


Two
Search
20250127_210828.jpg
20250127_210831.jpg
20250127_210835.jpg
20250127_210950.jpg
20250127_210954.jpg
20250127_210957.jpg
20250127_211000.jpg
20250127_211003.jpg
20250127_211007.jpg
20250127_211017.jpg
20250127_211020.jpg
20250127_211023.jpg
20250127_211027.jpg
20250127_211030.jpg
20250127_211033.jpg
20250127_211037.jpg
20250127_211040.jpg
20250127_211044.jpg
20250127_211054.jpg
20250127_211057.jpg
20250127_211100.jpg
20250127_211108.jpg
20250127_211111.jpg
20250127_211114.jpg
20250127_211116.jpg
20250127_211119.jpg
20250127_211122.jpg
20250127_211131.jpg
20250127_211134.jpg
20250127_211137.jpg
20250127_211140.jpg
20250127_211143.jpg
20250127_211146.jpg
20250127_211148.jpg
20250127_211152.jpg
20250127_211155.jpg
Macintosh HD › Users › ege › Downloads › hand-number-recognition › Two

Model Design

Basic CNN Model

The first model is a sequential CNN with the following architecture:

1. **Convolutional Layers**:
    - Layer 1: 32 filters, 3x3 kernel, ReLU activation, followed by 2x2 max pooling.
    - Layer 2: 64 filters, 3x3 kernel, ReLU activation, followed by 2x2 max pooling.
    - Layer 3: 128 filters, 3x3 kernel, ReLU activation, followed by 2x2 max pooling.
    - Layer 4: 128 filters, 3x3 kernel, ReLU activation, followed by 2x2 max pooling.
    - Layer 5: 128 filters, 3x3 kernel, ReLU activation, followed by 2x2 max pooling.

2. **Flattening Layer**: Converts the 3D feature maps into a 1D vector.

3. **Dense Layers**:
    - Dense Layer 1: 64 units with ReLU activation.
    - Dense Layer 2: Output layer with 1 unit and sigmoid activation for binary classification.

ResNet50 Model

The second model uses the pre-trained ResNet50 model as a feature extractor. It consists of:

1. **Pretrained Model**: The ResNet50 architecture pre-trained on the ImageNet dataset.

2. **Global Average Pooling Layer**: Reduces the dimensionality of feature maps from the ResNet50 output.

3. **Prediction Layer**: A dense layer with 1 unit and sigmoid activation for binary classification.

The ResNet50-based model leverages transfer learning to adapt the pre-trained weights to the dataset at hand.

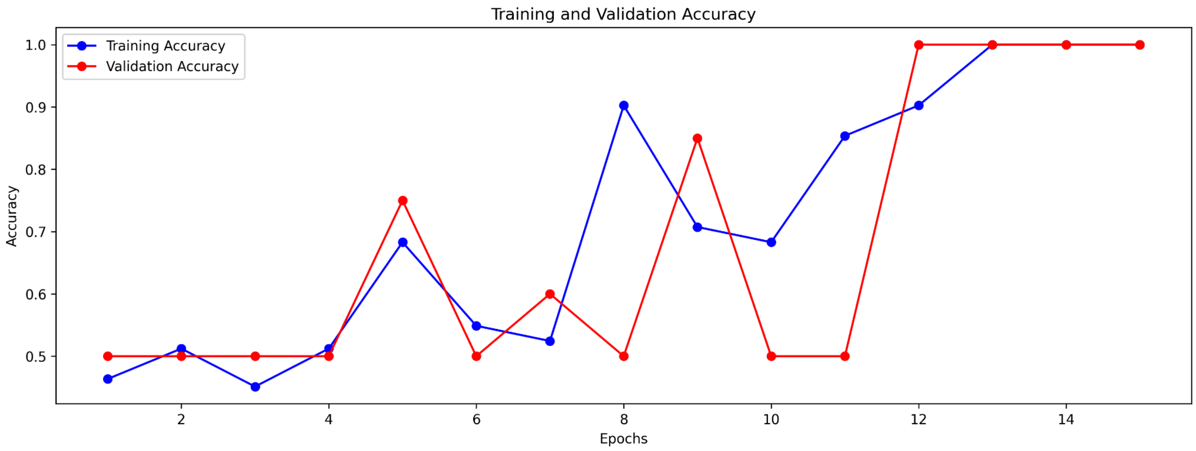
Training

Training and validation sets were split from the dataset. 80% of data is used for training, 20% of the data used for validation. Followings are used at training process:

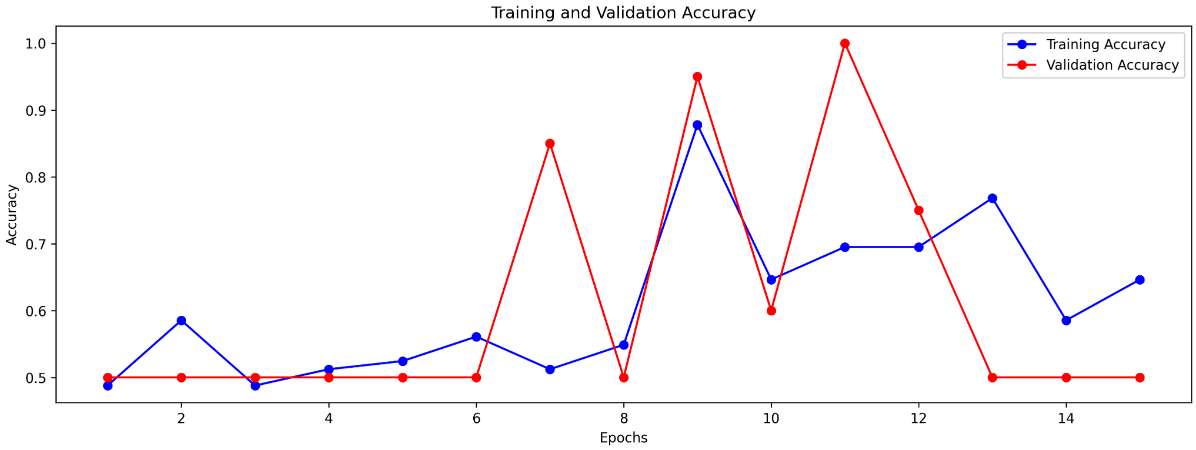
- Batch size: 16
- Epochs: 15
- Loss Function: Binary cross-entropy
- Optimizer: Adam

Results

Accuracy:

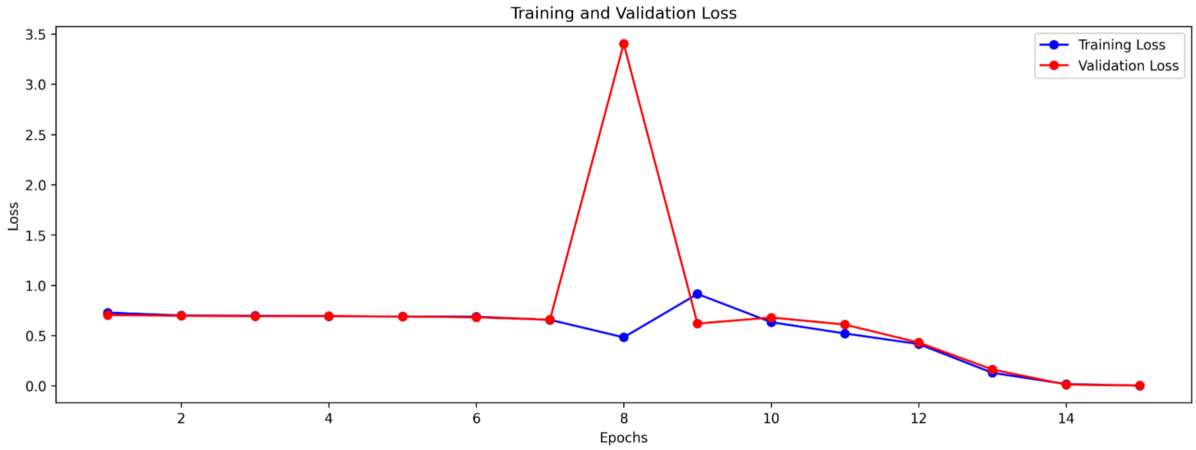


Accuracy of Basic CNN

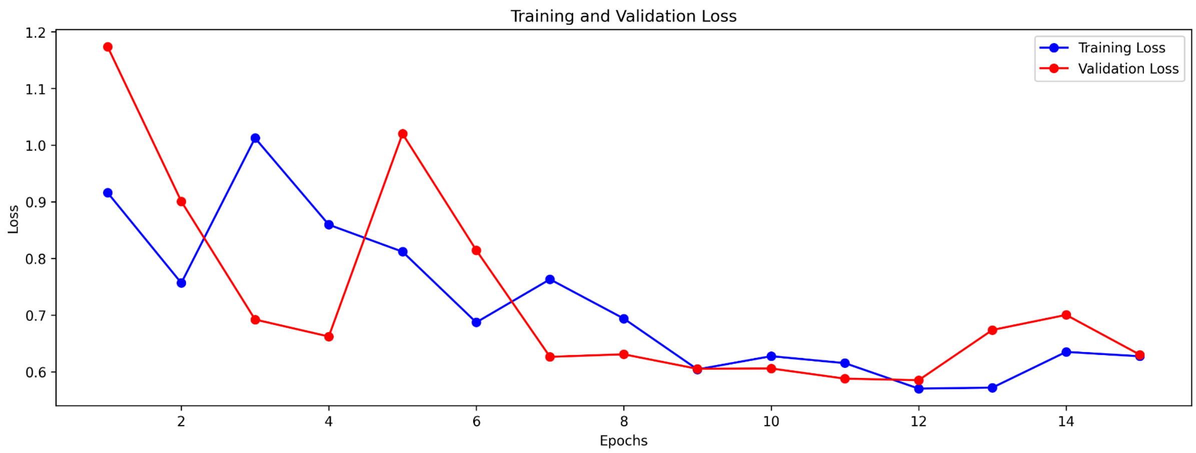


Accuracy of ResNet50

Loss:

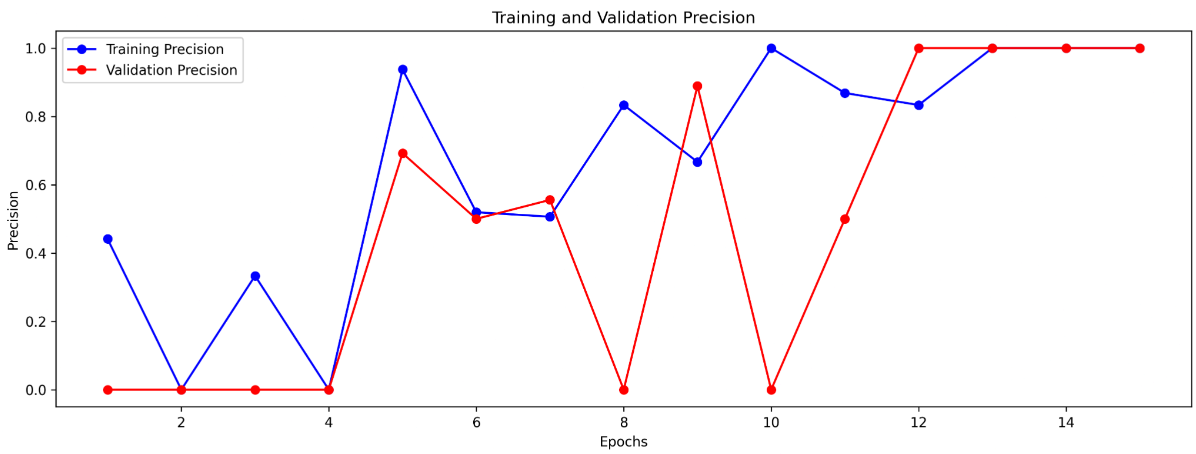


Loss of Basic CNN

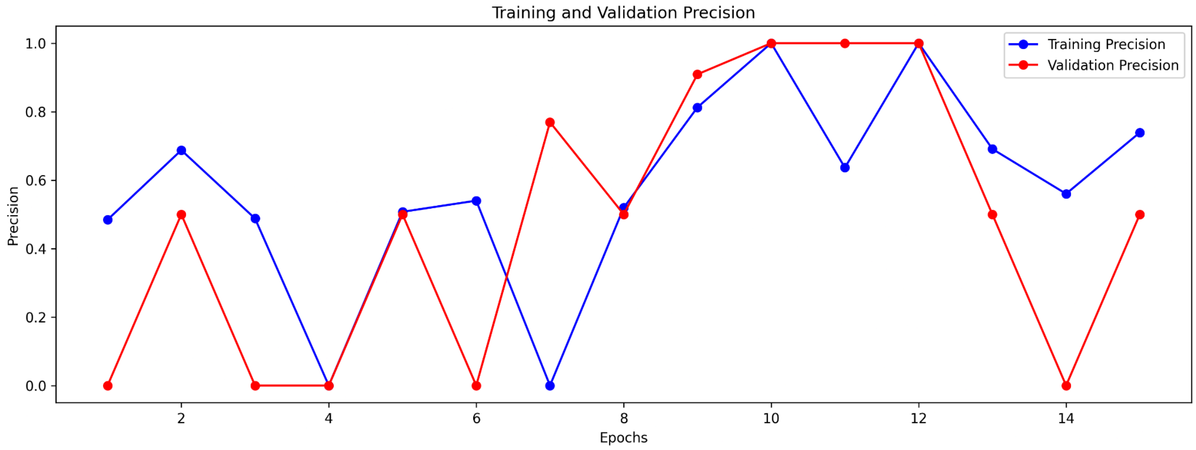


Loss of ResNet50

Precision:

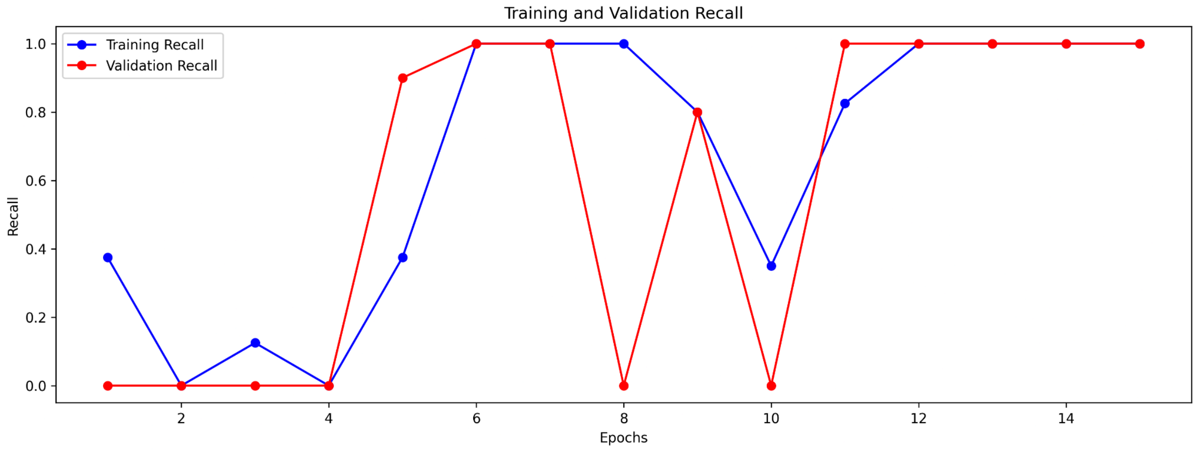


Precision of Basic CNN

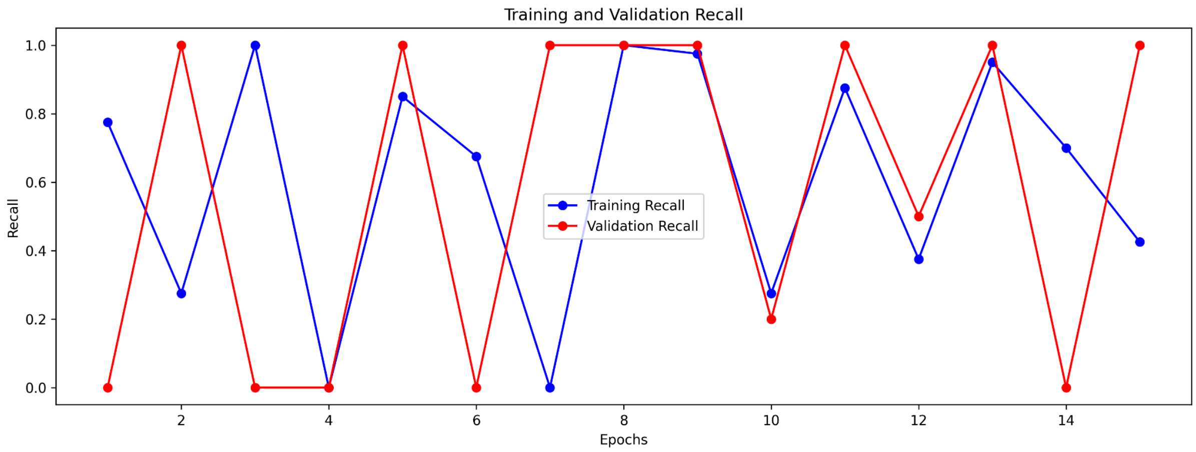


Precision of ResNet50

Recall:



Recall of Basic CNN



Recall of ResNet50

Conclusion

We can observe that the basic CNN demonstrates greater stability and achieves better accuracy for this specific task compared to ResNet50. This suggests that the simpler architecture of the basic CNN is more effective at capturing the dataset's features.