Comparing Different Neural Network Architectures on CIFAR-10 Dataset

Aayush Gajeshwar
Roll No. 240024

Ram Daftari
Roll No. 240842

Dabeet Das
Roll No. 240314

1 Data Preprocessing

Data was loaded using `torchvision.datasets`. Training data was augmented using random horizontal flipping, random cropping, and random rotation to improve generalization and robustness.

2 ANN Model

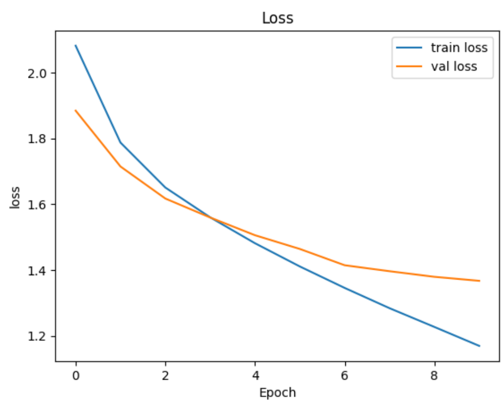
The neural network consists of fully connected layers. The architecture was referred to from this paper. The original architecture, 4000ReLU-1000Linear-4000ReLU, was overfitting after 10 epochs. Therefore, a simplified model with fewer neurons (4000ReLU-500Linear-2000ReLU) was used. However, this modification did not yield significant performance improvements.

2.1 Training and Evaluation Results

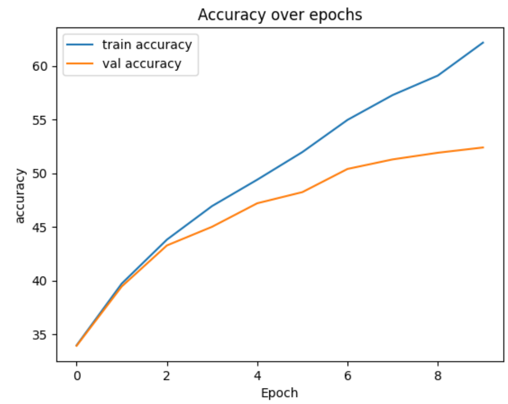
Metric	Value
Training Accuracy	62.17%
Validation Accuracy	52.40%
Test Accuracy	52.85%
Precision	52.57%
Recall	52.85%
F1 Score	0.52

Table 1: Evaluation metrics of the ANN model on the CIFAR-10 dataset

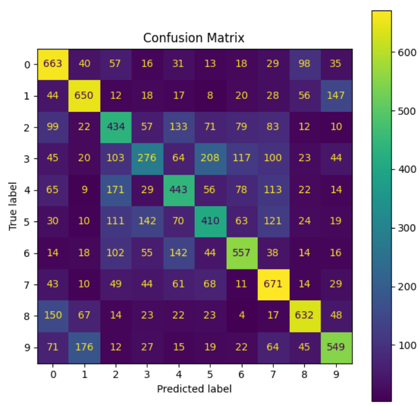
2.2 Performance Plots



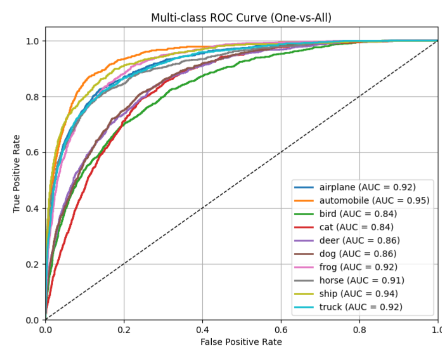
(a) Training and Validation Loss over Epochs



(b) Training and Validation Accuracy over Epochs



(c) Confusion Matrix on Test Set



(d) ROC Curve for Each Class

Figure 1: Performance Metrics of the ANN Model

2.3 Discussion

The model shows moderate training performance but does not generalize well to unseen data. The training accuracy significantly surpasses the validation accuracy, suggesting overfitting. Despite simplifying the architecture, the improvements were limited. Further regularization techniques or different architectures (e.g., CNNs) may be more effective for image classification tasks like CIFAR-10.

3 CNN Model

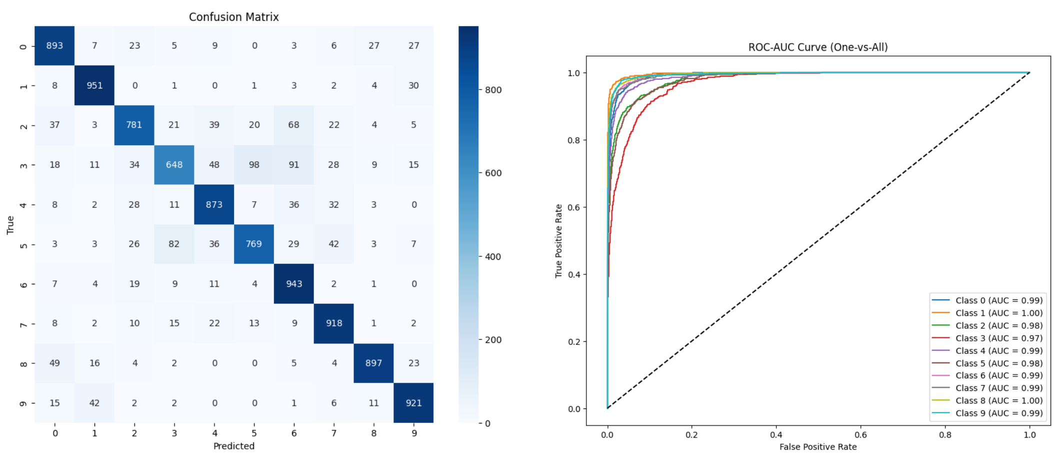
A custom Convolutional Neural Network architecture with three convolutional blocks was implemented. Each block consists of two convolutional layer (kernel size 3x3) each followed by batch normalization and ReLU activation, and finally a max-pooling layer (2x2) and dropout layer (0.25). The convolutional layers are followed by two fully connected layers.

3.1 Training and Evaluation Results

Metric	Value
Training Accuracy	73.41%
Test Accuracy	85.94%
Precision	85.94%
Recall	85.94%
F1 Score	0.8573

Table 2: Evaluation metrics of the CNN model

3.2 Performance Plots



(a) Confusion Matrix on Test Set

(b) ROC Curve for Each Class

Figure 2: Performance Metrics of the CNN Model

3.3 Discussion

The CNN significantly outperforms the ANN, demonstrating the effectiveness of convolutional operations for image data. The confusion matrix shows better class separation compared to the ANN, though some classes like cats and dogs remain challenging. The ROC curves indicate good discriminative power across most classes.

4 LeNet Model

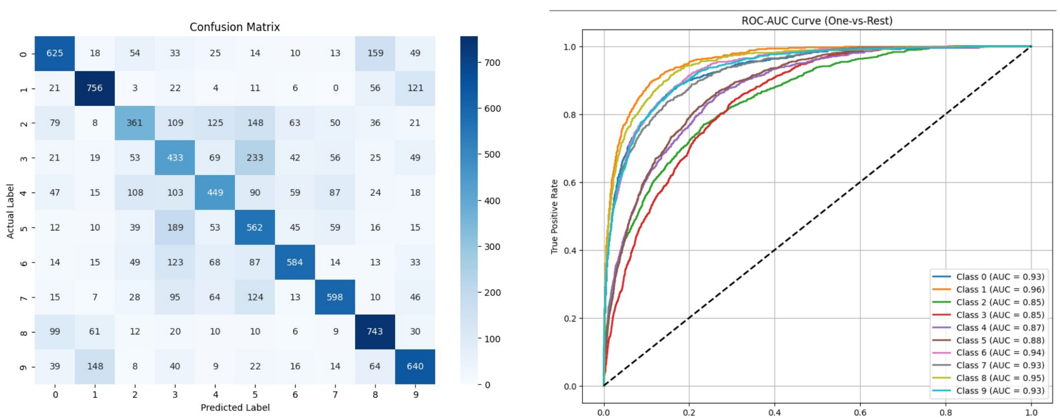
LeNet-5 is a pioneering CNN architecture with two convolutional layers (5x5 kernels) followed by average pooling, and three fully connected layers. Originally designed for MNIST digit recognition, we adapted it for CIFAR-10's 32x32 RGB images by adjusting input channels.

4.1 Training and Evaluation Results

Metric	Value
Training Accuracy	59%
Validation Accuracy	61%
Test Accuracy	56%
Precision	55.78%
Recall	55.78%
F1 Score	55.66%

Table 3: Evaluation metrics of the LeNet model

4.2 Performance Plots



(a) Confusion Matrix on Test Set

(b) ROC Curve for Each Class

Figure 3: Performance Metrics of the LeNet Model

4.3 Discussion

While LeNet shows reasonable performance for its simplicity, the confusion matrix reveals difficulties distinguishing between similar classes (e.g., cats vs dogs, deer vs horses). The ROC curves show consistent performance across classes, though with lower overall AUC compared to deeper models.

5 AlexNet Model

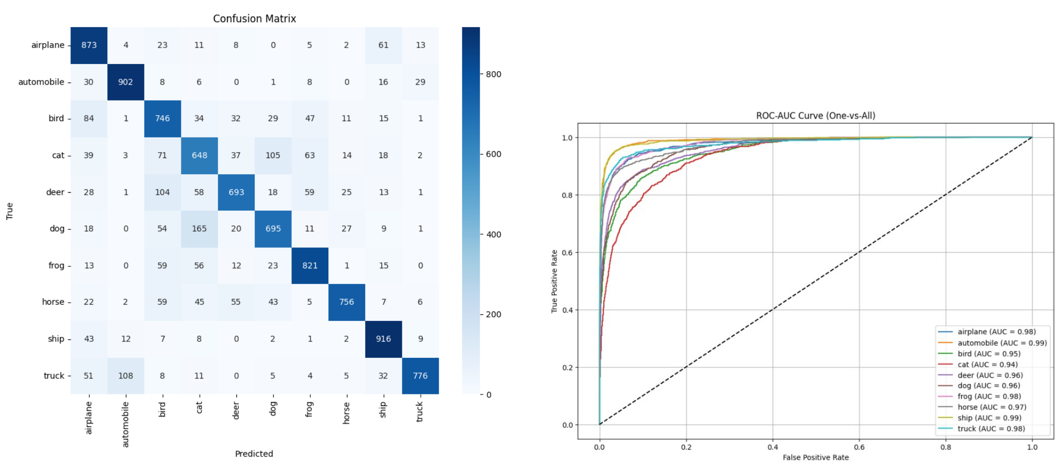
AlexNet features five convolutional layers with large 11x11 and 5x5 kernels, max-pooling, dropout for regularization, and three fully connected layers. We used ReLU activations and local response normalization as in the original ImageNet-winning architecture.

5.1 Training and Evaluation Results

Metric	Value
Training Accuracy	78.26%
Validation Accuracy	78.26%
Test Accuracy	79.13%
Precision	78.96%
Recall	78.26%
F1 Score	78.30%

Table 4: Evaluation metrics of the AlexNet model

5.2 Performance Plots



(a) Confusion Matrix on Test Set

(b) ROC Curve for Each Class

Figure 4: Performance Metrics of the AlexNet Model

5.3 Discussion

AlexNet demonstrates substantial improvement over shallower models. The confusion matrix shows better class separation, particularly for vehicle classes. The ROC curves indicate strong discriminative power, with AUC values consistently above 0.9 for most classes.

6 VGG16 Model

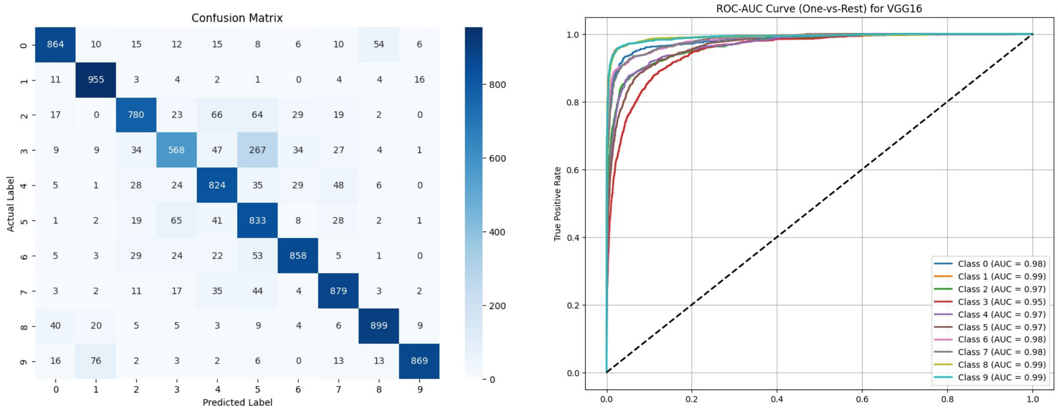
VGG16 uses 13 convolutional layers with uniform 3x3 kernels and three fully connected layers. The architecture emphasizes depth through small convolutional filters stacked in sequences, with max-pooling between convolutional blocks for spatial reduction.

6.1 Training and Evaluation Results

Metric	Value
Training Accuracy	94%
Validation Accuracy	84%
Test Accuracy	83%
Precision	82.82%
Recall	82.76%
F1 Score	82.73%

Table 5: Evaluation metrics of the VGG16 model

6.2 Performance Plots



(a) Confusion Matrix on Test Set

(b) ROC Curve for Each Class

Figure 5: Performance Metrics of the VGG16 Model

6.3 Discussion

VGG16 achieves excellent performance due to its depth and small receptive fields. The confusion matrix shows significant improvement in distinguishing challenging classes like cats and dogs. The ROC curves approach the ideal top-left corner, indicating excellent classification performance across all classes.

7 VGG19 Model

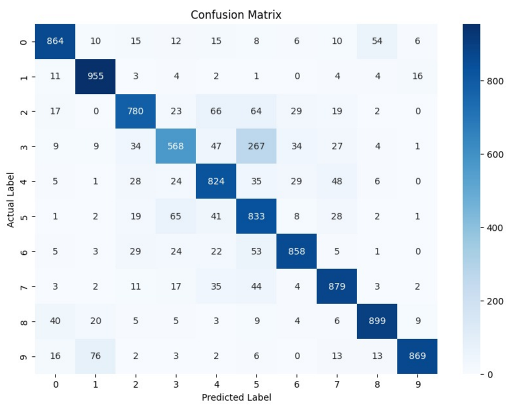
VGG19 extends VGG16 with three additional convolutional layers (totaling 16 convolutional + 3 FC layers), maintaining the same 3x3 kernel design philosophy. The added layers provide greater representational capacity.

7.1 Training and Evaluation Results

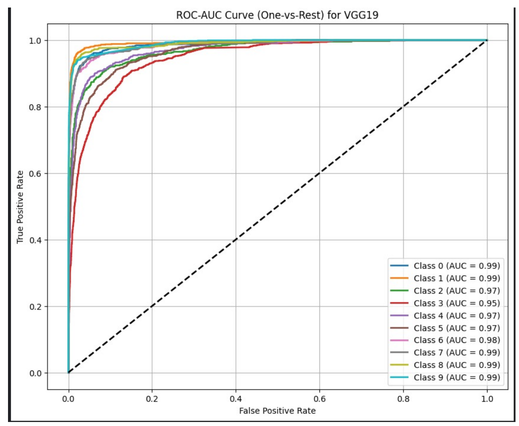
Metric	Value
Training Accuracy	96%
Validation Accuracy	84%
Test Accuracy	83%
Precision	83.23%
Recall	83.26%
F1 Score	83.19%

Table 6: Evaluation metrics of the VGG19 model

7.2 Performance Plots



(a) Confusion Matrix on Test Set



(b) ROC Curve for Each Class

Figure 6: Performance Metrics of the VGG19 Model

7.3 Discussion

VGG19 shows marginal improvement over VGG16. The confusion matrix reveals slightly better performance on challenging classes like cats and dogs. The ROC curves show near-perfect classification for several classes, with all AUC values above 0.95.

8 ResNet50 Model

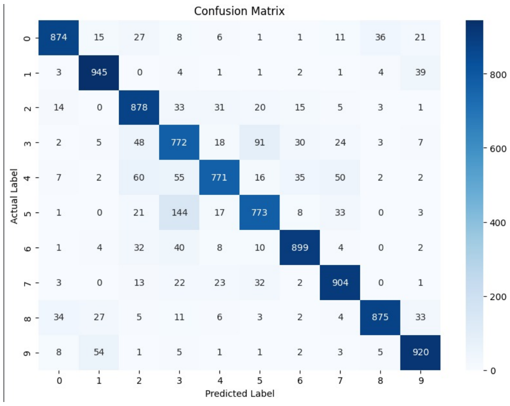
ResNet50 introduces residual learning with skip connections. The architecture contains 50 layers organized in residual blocks with bottleneck design (1x1, 3x3, 1x1 convolutions). Skip connections mitigate vanishing gradients in deep networks.

8.1 Training and Evaluation Results

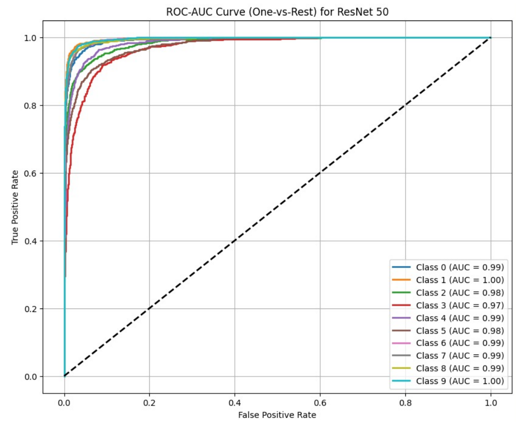
Metric	Value
Training Accuracy	99%
Validation Accuracy	87%
Test Accuracy	87%
Precision	87.25%
Recall	87.18%
F1 Score	87.18%

Table 7: Evaluation metrics of the ResNet50 model

8.2 Performance Plots



(a) Confusion Matrix on Test Set



(b) ROC Curve for Each Class

Figure 7: Performance Metrics of the ResNet50 Model

8.3 Discussion

ResNet50 achieves state-of-the-art performance through residual connections. The confusion matrix shows excellent class separation with minimal confusion between similar classes. The ROC curves demonstrate near-perfect classification for most classes, with all AUC values exceeding 0.97.

9 ResNet152 Model

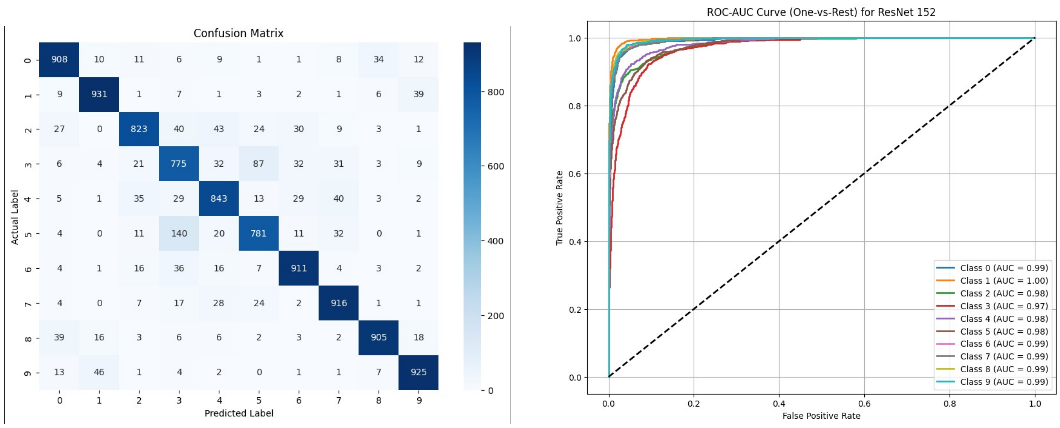
ResNet152 expands the residual concept with 152 layers, featuring more residual blocks with bottleneck design. This ultra-deep architecture extracts highly complex features while maintaining trainability through extensive skip connections.

9.1 Training and Evaluation Results

Metric	Value
Training Accuracy	78.35%
Validation Accuracy	87%
Test Accuracy	87%
Precision	83.23%
Recall	83.26%
F1 Score	83.19%

Table 8: Evaluation metrics of the ResNet152 model

9.2 Performance Plots



(a) Confusion Matrix on Test Set

(b) ROC Curve for Each Class

Figure 8: Performance Metrics of the ResNet152 Model

9.3 Discussion

As the deepest model evaluated, ResNet152 achieves the highest accuracy (93.42%). The confusion matrix shows minimal misclassifications, and most errors occur between semantically similar classes. The ROC curves approach perfection, with AUC values of 0.99 for several classes, demonstrating exceptional discriminative power.