Kravspecifikation för del 2

1. ✅ Börja använda basala **MLOps best-practices:**

   A. 📊 Ordna så att du har ett bra sätt att övervaka träningen, samt att spara (och hålla ordning på!) versioner av modellerna du tränar upp. (Tips: Foldrar, smart valda filnamn, etc).

   B. 💾 Kom ihåg att det troligen inte är den sista epokens checkpoint (sparad version) som är den bästa, p.g.a. overfitting om du kör många epochs. Du behöver själv se till att spara checkpoints lite då och då under träningen.

   C. 🗄️ Versionshantering av parametrar, kod, och checkpoints (modellens parametrar) blir snabbt väldigt viktigt när man börjar jobba seriöst med machine learning. Om man inte håller reda på vilken körning som gjorts med vilka parametrar, så blir det svårt att jämföra prestanda, vidareutveckla modellen senare, etc. Ingen vill ha en modell levererad där programmeraren inte minns hur den gjordes.

   D. 🚀 **Fördjupning/överkurs:** Vilka verktyg finns som kan hjälpa till att hålla ordning på versioner, körningar, resultat, hyper parameters, etc?

2. 🚀 **Fördjupning/överkurs:** Addera **performance metrics**, så att du kan se hur lång tid respektive tränings-körning tar. (En körning = Alla epochs av träning för en given modell med specificerad uppsättning hyper parameters).

3. ✅ Applicera lämpliga **data augmentation** methods för att artificiellt variera, och till och med "förstora" datamängden lite. Använd exempelvis skalning, rotation, färgvariation, brus, (spegelvändning?), etc. Här är lite repetition om data augmentation:

   A. Se avsnittet "Uppgift 1 Perceptron för OCR" längre ner i denna uppgift (del 3).

   B. Bra 10-minuters video om ämnet: Pytorch-Data-Augmentation-using-Torchvision

   C. Observera att det inte är säkert att data augmentation leder till bättre resultat i alla lägen, eftersom det beror på detaljerna i vad man bygger, och vilken data man har.

4. ✅ Det är nu dags att testa CNN istället för FFN. Byt ut de första lagren i modellen till **convolutional layers** för att ta upp translationsinvarians i bilden. Sök själv upp vad man brukar ha ***direkt efter varje convolutional layer*** för att hålla ordning på antalet dimensioner till nästa lager .

   > (Svårt ord "translations-invarians": translation=förflyttning, invarians="samma oavsett" ⇒ translationsinvarians = "spelar ingen roll var i bilden")

5. ✅ Experimentera därefter med att addera ett till (eller flera) convolutional layer(s). Tanken är att testa **olika nät-arkitekturer**. Använd Google/Chat för initial gissning för modell-arkitektur, men sedan är det trial-and-error som gäller. Att fundera på:

   A. ⚖️ Hur påverkas resultatet av olika modell-arkitektur? Var noga med att inte jämföra modeller med stor skillnad i antal parameters, om aspekter som tex lagertyp ska utvärderas.
   B. 🤔 Vilken typ av features detekteras typiskt av de senare convolutional-lagren jämfört med det första convolutional-lagret?
      a. Hint: Videon från lektion om convolutional layers: Convolutional-Neural-Networks-Explained. (Självfallet kan du också plotta weight-matrices från din egen modell också om du vill!)

6. ✅ Applicera lämpliga **regularization methods** för att minimera risken för overfitting och göra cost-function-landskapet fördelaktigt för back-prop.

   A. Exempel på sådana metoder är: drop-out, weight-decay, noise injection, batch normalization, etc.
   B. Mer info om regularization: understanding-regularization-with-pytorch

7. 🚀 **Fördjupning:** Skapa en lista av kombinationer av hyper parameters, och låt datorn träna din modell på nytt för samtliga hyperparameter settings i din lista. Detta kallas för **hyper-parameter tuning**, d.v.s. att via trial-and-error hitta inställningar som funkar.

   - Enklast möjliga hyper-param-tuning: Gör en lista med t.ex. 10 rader (en per körning), där varje rad specificerar alla hyper-parameters som kan vara svårt att gissa optimalt värde på:
     - Antal neuroner för respektive lager
     - Typ av activation function
     - Vilka lager som ska vara convolutional, storlek på convolutional kernel, antal convolutional kernels per lager, etc.
     - Learning rate / hyper-params till optimizer ADAM
     - Drop-out level
     - Settings för olika typer av data augmentation (rotation, skalning, noise, etc)
   - Det finns även en mängd automatiserade typer av hyper-param-tuning. Se exempelvis: https://en.wikipedia.org/wiki/Hyperparameter_optimization

In [2]:

```python
import os
import time
import logging
import random
import torch
import torch.nn as nn
import torch.optim as optim
from datetime import datetime
from torch.utils.data import DataLoader, random_split
from torchvision import datasets, transforms

# Define the perceptron neural-network model
class Perceptron(nn.Module):
    # Define the constructor
    def __init__(self, dropout_rate=0.25):
        super().__init__()

        self.dropout_rate = dropout_rate

        # Convolutional layers
        self.conv_layers = nn.Sequential(
            nn.Conv2d(1, 32, kernel_size=3, padding=1), # 32x32 with 1 paddi
            nn.BatchNorm2d(32), # Add batch normalization after convolution
            nn.ReLU(), # Apply ReLU activation function
            nn.MaxPool2d(kernel_size=2), # 14x14 output dimensions
            nn.Dropout2d(dropout_rate) # Add dropout to convolutional featur
        )

        # Fully connected layers
        self.fc_layers = nn.Sequential(
            nn.Linear(32 * 14 * 14, 128), # 128 neurons in the hidden layer
            nn.BatchNorm1d(128), # Add batch normalization after linear laye
            nn.ReLU(), # Apply ReLU activation function
            nn.Dropout(dropout_rate), # Add dropout to hidden layer
            nn.Linear(128, 10) # 10 neurons in the output layer
        )

    # Define the forward pass
    def forward(self, x):
        # Pass through the convolutional layers
        logits = self.conv_layers(x)

        # Flatten the output
        logits = logits.view(logits.size(0), -1)

        # Pass through the fully connected layers
        logits = self.fc_layers(logits)

        return logits

def train_model(model, training_dataset, testing_dataset, hyperparameters, m
    # Extract hyperparameters
    num_epochs = hyperparameters['num_epochs']
    learning_rate = hyperparameters['learning_rate']
    batch_size = hyperparameters['batch_size']
    l1_lambda = hyperparameters['l1_lambda']
    l2_lambda = hyperparameters['l2_lambda']
    dropout_rate = hyperparameters['dropout_rate']
    early_stopping_patience = hyperparameters['early_stopping_patience']
    criterion = nn.CrossEntropyLoss()
    optimizer = optim.Adam(model.parameters(), lr=learning_rate, weight_deca

    # Split training data into train and validation subsets
    training_subset_size = int(0.8 * len(training_dataset))
    validation_subset_size = len(training_dataset) - training_subset_size
    training_subset, validation_subset = random_split(training_dataset, [tra

    # Create DataLoaders
    train_loader = DataLoader(training_subset, batch_size=batch_size, shuffl
    validation_loader = DataLoader(validation_subset, batch_size=batch_size,
    testing_loader = DataLoader(testing_dataset, batch_size=batch_size, shuf
    # Create model directory
    model_dir = os.path.join('models', f'run_{run_id}', f'model_{model_id}')
    os.makedirs(model_dir, exist_ok=True)

    # Create checkpoints directory
    checkpoints_dir = os.path.join(model_dir, 'checkpoints')
    os.makedirs(checkpoints_dir, exist_ok=True)

    # Set up logging
    logging.basicConfig(level=logging.DEBUG, format='%(asctime)s - %(levelna
    logger = logging.getLogger()
    log_file = os.path.join(model_dir, f'model_{model_id}.log')
    fhandler = logging.FileHandler(filename=log_file, mode='a')
    formatter = logging.Formatter('%(asctime)s - %(name)s - %(levelname)s -
    fhandler.setFormatter(formatter)
    logger.addHandler(fhandler)

    # Log hyperparameters
    logger.info("=" * 100)
    logger.info(f"Model ID: {model_id}")
    logger.info(f"Run ID: {run_id}")
    logger.info(f"Training configuration:")
    logger.info(f"  Learning rate: {learning_rate}")
    logger.info(f"  Batch size: {batch_size}")
    logger.info(f"  L1 regularization: {l1_lambda}")
    logger.info(f"  L2 regularization: {l2_lambda}")
    logger.info(f"  Dropout rate: {dropout_rate}")
    logger.info(f"  Epochs: {num_epochs}")
    logger.info(f"  Early stopping patience: {early_stopping_patience}")
    logger.info(f"  Optimizer: Adam")
    logger.info(f"  Loss function: CrossEntropyLoss")

    # Training and validation loop
    training_start_time = time.time()
    best_val_loss = float('inf')
    best_model_path = None
    early_stopping_counter = 0
    epochs_without_improvement = 0

    for epoch in range(num_epochs):
        # Cache the start time of the epoch
        epoch_start_time = time.time()

        # Training phase
        model.train()
        running_train_loss = 0.0
        for x, y in train_loader:
            # Move data to device
            x, y = x.to(device), y.to(device)

            # Forward pass
            outputs = model(x)

            # Calculate base loss
            base_loss = criterion(outputs, y)
            # L1 regularization (L2 is handled by weight_decay)
            l1_reg = torch.tensor(0., device=device)
            for param in model.parameters():
                l1_reg += torch.norm(param, 1)

            # Total loss with L1 regularization
            loss = base_loss + l1_lambda * l1_reg

            # Backward pass and optimization
            optimizer.zero_grad()
            loss.backward()
            optimizer.step()

            # Update running loss
            running_train_loss += loss.item()

        # Calculate average loss
        avg_train_loss = running_train_loss / len(train_loader)

        # Print training loss
        logger.info("="*100)
        logger.info(f"Epoch [{epoch+1}/{num_epochs}]")
        logger.info(f"Training Loss: {avg_train_loss:.4f}")

        # Validation phase
        model.eval()
        running_val_loss = 0.0
        correct = 0
        total = 0
        with torch.no_grad():
            for x, y in validation_loader:
                # Move data to device
                x, y = x.to(device), y.to(device)

                # Forward pass
                outputs = model(x)
                loss = criterion(outputs, y)
                running_val_loss += loss.item()

                # Calculate accuracy
                _, predicted = torch.max(outputs, 1)
                total += y.size(0)
                correct += (predicted == y).sum().item()

        # Calculate average loss and accuracy
        avg_val_loss = running_val_loss / len(validation_loader)
        val_accuracy = 100 * correct / total

        # Print validation loss and accuracy
        logger.info(f"Validation Loss: {avg_val_loss:.4f}")
        logger.info(f"Validation Accuracy: {val_accuracy:.2f}%")

        # Save the checkpoint
        checkpoint_filename = f'epoch_{epoch+1}.pth'
        checkpoint_path = os.path.join(checkpoints_dir, checkpoint_filename)
        torch.save(model.state_dict(), checkpoint_path)

        # Update the best model if the current model has a lower validation
        if avg_val_loss < best_val_loss:
            best_model_path = checkpoint_path # Cache the path to the best m
            best_val_loss = avg_val_loss
            epochs_without_improvement = 0
            logger.info(f"New best validation loss: {best_val_loss:.4f}")
        else:
            epochs_without_improvement += 1
            logger.info(f"No improvement for {epochs_without_improvement} ep

        # Early stopping check
        if epochs_without_improvement >= early_stopping_patience:
            logger.info(f"Early stopping triggered after {epoch+1} epochs")
            break

        # Log the duration of the epoch
        epoch_end_time = time.time()
        epoch_duration = epoch_end_time - epoch_start_time
        logger.info(f"Epoch {epoch+1} duration: {epoch_duration:.2f} seconds

    training_end_time = time.time()
    training_duration = training_end_time - training_start_time
    logger.info(f"Training duration: {training_duration:.2f} seconds")
    if epochs_without_improvement >= early_stopping_patience:
        logger.info(f"Training stopped early at epoch {epoch+1}/{num_epochs}
    else:
        logger.info(f"Completed all {num_epochs} epochs")

    # Get the best model for testing
    logger.info(f"Loading best model from: {best_model_path}")
    model.load_state_dict(torch.load(best_model_path))

    # Testing loop
    testing_start_time = time.time()
    model.eval()
    running_test_loss = 0.0
    correct = 0
    total = 0
    with torch.no_grad():
        for x, y in testing_loader:
            # Move data to device
            x, y = x.to(device), y.to(device)

            # Forward pass
            outputs = model(x)
            loss = criterion(outputs, y)
            running_test_loss += loss.item()

            # Calculate accuracy
            _, predicted = torch.max(outputs, 1)
            total += y.size(0)
            correct += (predicted == y).sum().item()

    # Calculate average loss and accuracy
    avg_test_loss = running_test_loss / len(testing_loader)
    test_accuracy = 100 * correct / total

    # Log the results
    testing_end_time = time.time()
    testing_duration = testing_end_time - testing_start_time
    total_run_time = testing_end_time - training_start_time
    logger.info("="*100)
    logger.info(f"Testing duration: {testing_duration:.2f} seconds")
    logger.info(f"Total duration: {total_run_time:.2f} seconds")
    logger.info("="*100)
    logger.info(f"Best Model: {best_model_path}")
    logger.info(f"Best Validation Loss: {best_val_loss:.4f}")
    logger.info(f"Test Loss: {avg_test_loss:.4f}")
    logger.info(f"Test Accuracy: {test_accuracy:.2f}%")
    logger.info("="*100)

    return test_accuracy


if __name__ == "__main__":
    # Create a unique ID for the run
    run_id = datetime.now().strftime("%Y%m%d_%H%M%S")

    # Create directory structure
    os.makedirs('models', exist_ok=True)
    run_dir = os.path.join('models', f'run_{run_id}')
    os.makedirs(run_dir, exist_ok=True)

    # Set up logging
    logging.basicConfig(level=logging.DEBUG, format='%(asctime)s - %(levelna
    logger = logging.getLogger()
    log_file = os.path.join(run_dir, f"run_{run_id}.log")
    os.makedirs(os.path.dirname(log_file), exist_ok=True)
    fhandler = logging.FileHandler(filename=log_file, mode='a')
    formatter = logging.Formatter('%(asctime)s - %(name)s - %(levelname)s -
    fhandler.setFormatter(formatter)
    logger.addHandler(fhandler)

    # Log start of experiment
    logger.info("="*100)
    logger.info(f"Starting new experiment with run ID: {run_id}")
    logger.info("="*100)

    # Select device to run on
    device = torch.accelerator.current_accelerator().type if torch.accelerat
    logger.info(f"Using device: {device}")

    # Transformations for training data with data augmentation
    training_transform = transforms.Compose([
        transforms.RandomRotation(10),  # Rotate by up to 10 degrees
        transforms.RandomAffine(0, scale=(0.8, 1.2)),  # Random scaling
        transforms.ToTensor(),  # Convert to tensor
        transforms.Normalize((0.1307,), (0.3081,)),  # Normalize with MNIST
    ])
    # Load the MNIST dataset
    training_dataset = datasets.MNIST(root='./data', train=True, download=Tr
    testing_dataset = datasets.MNIST(root='./data', train=False, download=Tr
    logger.info(f"Loaded datasets - Training: {len(training_dataset)} sample

    # Hyperparameters options
    num_variants = 5
    num_epochs = 10
    batch_size_learning_rate_options = [(32, 0.0001), (64, 0.0001), (128, 0.
    l1_lambda_options = [0.0001, 0.00001, 0.000001]
    l2_lambda_options = [0.001, 0.0001, 0.00001]
    dropout_rate_options = [0.1, 0.25, 0.5]
    early_stopping_patience = 5

    hyperparameters_list = []
    for _ in range(num_variants):
        # Randomly select hyperparameters
        batch_size, learning_rate = random.choice(batch_size_learning_rate_o
        l1_lambda = random.choice(l1_lambda_options)
        l2_lambda = random.choice(l2_lambda_options)
        dropout_rate = random.choice(dropout_rate_options)

        # Construct the hyperparameters dictionary
        hyperparameters = {
            'num_epochs': num_epochs,
            'learning_rate': learning_rate,
            'batch_size': batch_size,
            'l1_lambda': l1_lambda,
            'l2_lambda': l2_lambda,
            'dropout_rate': dropout_rate,
            'early_stopping_patience': early_stopping_patience
        }

        # Ensure hyperparameters are unique
        if hyperparameters not in hyperparameters_list:
            hyperparameters_list.append(hyperparameters)

    # Track best model and accuracy
    best_model_accuracy = 0
    best_model_id = None

    # Track all model results
    model_results = []

    # Train a model for each set of hyperparameters
    for i, hyperparameters in enumerate(hyperparameters_list):
        # Create a unique ID for this model
        model_id = f"variant_{i+1}"

        # Log start of training for this model
        logger.info("="*100)
        logger.info(f"Starting training for model {model_id}")
        logger.info(f"Hyperparameters:")
        logger.info(f"Learning rate: {hyperparameters['learning_rate']}")
        logger.info(f"Batch size: {hyperparameters['batch_size']}")
        logger.info(f"L1 lambda: {hyperparameters['l1_lambda']}")
        logger.info(f"L2 lambda: {hyperparameters['l2_lambda']}")
        logger.info(f"Dropout rate: {hyperparameters['dropout_rate']}")
        logger.info(f"Early stopping patience: {hyperparameters['early_stopp

        # Initialize a fresh model for each set of hyperparameters with the
        model = Perceptron(dropout_rate=hyperparameters['dropout_rate']).to(

        # Train the model and get test accuracy
        accuracy = train_model(model, training_dataset, testing_dataset, hyp

        # Log the results for this model
        logger.info(f"Completed training for model {model_id}")
        logger.info(f"Test accuracy: {accuracy:.2f}%")

        # Store model results
        model_results.append({
            'model_id': model_id,
            'accuracy': accuracy,
            'hyperparameters': hyperparameters
        })

        # Track the best model
        if accuracy > best_model_accuracy:
            best_model_accuracy = accuracy
            best_model_id = model_id
            logger.info(f"New best model: {model_id} with accuracy {accuracy

    # Log summary of all models
    logger.info("="*100)
    logger.info("Training completed for all models")
    logger.info("Summary of model performances:")

    for result in sorted(model_results, key=lambda x: x['accuracy'], reverse
        logger.info(f"Model {result['model_id']}: Accuracy {result['accuracy
        logger.info(f"  Learning rate: {result['hyperparameters']['learning_
        logger.info(f"  Batch size: {result['hyperparameters']['batch_size']
        logger.info(f"  L1 lambda: {result['hyperparameters']['l1_lambda']}"
        logger.info(f"  L2 lambda: {result['hyperparameters']['l2_lambda']}"
        logger.info(f"  Dropout rate: {result['hyperparameters']['dropout_ra
        logger.info(f"  Early stopping patience: {result['hyperparameters'][
        logger.info("-" * 50)

    # Log best model details
    logger.info("="*100)
    logger.info(f"Best model: {best_model_id} with accuracy: {best_model_acc
    best_model_hp = next(result['hyperparameters'] for result in model_resul
    logger.info(f"Best model hyperparameters:")
    logger.info(f"  Learning rate: {best_model_hp['learning_rate']}")
    logger.info(f"  Batch size: {best_model_hp['batch_size']}")
    logger.info(f"  L1 lambda: {best_model_hp['l1_lambda']}")
    logger.info(f"  L2 lambda: {best_model_hp['l2_lambda']}")
    logger.info(f"  Dropout rate: {best_model_hp['dropout_rate']}")
    logger.info(f"  Early stopping patience: {best_model_hp['early_stopping_
    logger.info("="*100)
```

```
2025-04-27 21:09:47,943 - INFO - ===========================================
=========================================================
2025-04-27 21:09:47,944 - INFO - Starting new experiment with run ID: 202504
27_210947
2025-04-27 21:09:47,944 - INFO - ===========================================
=========================================================
2025-04-27 21:09:47,945 - INFO - Using device: cuda
2025-04-27 21:09:47,944 - INFO - Starting new experiment with run ID: 202504
27_210947
2025-04-27 21:09:47,944 - INFO - ===========================================
=========================================================
2025-04-27 21:09:47,945 - INFO - Using device: cuda
2025-04-27 21:09:48,009 - INFO - Loaded datasets - Training: 60000 samples,
Testing: 10000 samples
2025-04-27 21:09:48,010 - INFO - ===========================================
=========================================================
2025-04-27 21:09:48,010 - INFO - Starting training for model variant_1
2025-04-27 21:09:48,010 - INFO - Hyperparameters:
2025-04-27 21:09:48,011 - INFO - Learning rate: 0.0001
2025-04-27 21:09:48,011 - INFO - Batch size: 128
2025-04-27 21:09:48,011 - INFO - L1 lambda: 1e-06
2025-04-27 21:09:48,012 - INFO - L2 lambda: 1e-05
2025-04-27 21:09:48,012 - INFO - Dropout rate: 0.5
2025-04-27 21:09:48,012 - INFO - Early stopping patience: 5
2025-04-27 21:09:48,020 - INFO - ===========================================
=========================================================
2025-04-27 21:09:48,021 - INFO - Model ID: variant_1
2025-04-27 21:09:48,021 - INFO - Run ID: 20250427_210947
2025-04-27 21:09:48,022 - INFO - Training configuration:
2025-04-27 21:09:48,022 - INFO -   Learning rate: 0.0001
2025-04-27 21:09:48,023 - INFO -   Batch size: 128
2025-04-27 21:09:48,023 - INFO -   L1 regularization: 1e-06
2025-04-27 21:09:48,023 - INFO -   L2 regularization: 1e-05
2025-04-27 21:09:48,024 - INFO -   Dropout rate: 0.5
2025-04-27 21:09:48,024 - INFO -   Epochs: 10
2025-04-27 21:09:48,025 - INFO -   Early stopping patience: 5
2025-04-27 21:09:48,025 - INFO -   Optimizer: Adam
2025-04-27 21:09:48,025 - INFO -   Loss function: CrossEntropyLoss
2025-04-27 21:09:58,802 - INFO - ===========================================
=========================================================
2025-04-27 21:09:58,803 - INFO - Epoch [1/10]
2025-04-27 21:09:58,803 - INFO - Training Loss: 0.9644
2025-04-27 21:10:01,291 - INFO - Validation Loss: 0.5133
2025-04-27 21:10:01,292 - INFO - Validation Accuracy: 90.58%
2025-04-27 21:10:01,297 - INFO - New best validation loss: 0.5133
2025-04-27 21:10:01,297 - INFO - Epoch 1 duration: 13.27 seconds
2025-04-27 21:10:12,016 - INFO - ===========================================
=========================================================
2025-04-27 21:10:12,017 - INFO - Epoch [2/10]
2025-04-27 21:10:12,017 - INFO - Training Loss: 0.5255
2025-04-27 21:10:14,517 - INFO - Validation Loss: 0.3081
2025-04-27 21:10:14,518 - INFO - Validation Accuracy: 93.79%
2025-04-27 21:10:14,524 - INFO - New best validation loss: 0.3081
2025-04-27 21:10:14,524 - INFO - Epoch 2 duration: 13.23 seconds
2025-04-27 21:10:25,222 - INFO - ===========================================
=========================================================
2025-04-27 21:10:25,222 - INFO - Epoch [3/10]
2025-04-27 21:10:25,223 - INFO - Training Loss: 0.3868
2025-04-27 21:10:27,683 - INFO - Validation Loss: 0.2268
2025-04-27 21:10:27,683 - INFO - Validation Accuracy: 95.03%
2025-04-27 21:10:27,688 - INFO - New best validation loss: 0.2268
2025-04-27 21:10:27,689 - INFO - Epoch 3 duration: 13.16 seconds
2025-04-27 21:10:38,553 - INFO - ===========================================
=========================================================
2025-04-27 21:10:38,554 - INFO - Epoch [4/10]
2025-04-27 21:10:38,554 - INFO - Training Loss: 0.3232
2025-04-27 21:10:41,125 - INFO - Validation Loss: 0.1859
2025-04-27 21:10:41,126 - INFO - Validation Accuracy: 95.59%
2025-04-27 21:10:41,131 - INFO - New best validation loss: 0.1859
2025-04-27 21:10:41,131 - INFO - Epoch 4 duration: 13.44 seconds
2025-04-27 21:10:52,297 - INFO - ===========================================
=========================================================
2025-04-27 21:10:52,298 - INFO - Epoch [5/10]
2025-04-27 21:10:52,298 - INFO - Training Loss: 0.2828
2025-04-27 21:10:54,808 - INFO - Validation Loss: 0.1516
2025-04-27 21:10:54,809 - INFO - Validation Accuracy: 96.21%
2025-04-27 21:10:54,814 - INFO - New best validation loss: 0.1516
2025-04-27 21:10:54,814 - INFO - Epoch 5 duration: 13.68 seconds
2025-04-27 21:11:05,659 - INFO - ===========================================
=========================================================
2025-04-27 21:11:05,659 - INFO - Epoch [6/10]
2025-04-27 21:11:05,660 - INFO - Training Loss: 0.2563
2025-04-27 21:11:08,181 - INFO - Validation Loss: 0.1393
2025-04-27 21:11:08,181 - INFO - Validation Accuracy: 96.53%
2025-04-27 21:11:08,186 - INFO - New best validation loss: 0.1393
2025-04-27 21:11:08,187 - INFO - Epoch 6 duration: 13.37 seconds
2025-04-27 21:11:18,980 - INFO - ===========================================
=========================================================
2025-04-27 21:11:18,981 - INFO - Epoch [7/10]
2025-04-27 21:11:18,982 - INFO - Training Loss: 0.2350
2025-04-27 21:11:21,499 - INFO - Validation Loss: 0.1234
2025-04-27 21:11:21,500 - INFO - Validation Accuracy: 96.66%
2025-04-27 21:11:21,505 - INFO - New best validation loss: 0.1234
2025-04-27 21:11:21,505 - INFO - Epoch 7 duration: 13.32 seconds
2025-04-27 21:11:32,599 - INFO - ===========================================
=========================================================
2025-04-27 21:11:32,599 - INFO - Epoch [8/10]
2025-04-27 21:11:32,600 - INFO - Training Loss: 0.2281
2025-04-27 21:11:35,128 - INFO - Validation Loss: 0.1130
2025-04-27 21:11:35,129 - INFO - Validation Accuracy: 96.91%
2025-04-27 21:11:35,136 - INFO - New best validation loss: 0.1130
2025-04-27 21:11:35,137 - INFO - Epoch 8 duration: 13.63 seconds
2025-04-27 21:11:45,922 - INFO - ===========================================
=========================================================
2025-04-27 21:11:45,923 - INFO - Epoch [9/10]
2025-04-27 21:11:45,924 - INFO - Training Loss: 0.2183
2025-04-27 21:11:48,411 - INFO - Validation Loss: 0.1101
2025-04-27 21:11:48,412 - INFO - Validation Accuracy: 97.00%
2025-04-27 21:11:48,417 - INFO - New best validation loss: 0.1101
2025-04-27 21:11:48,418 - INFO - Epoch 9 duration: 13.28 seconds
2025-04-27 21:11:59,186 - INFO - ===========================================
=========================================================
2025-04-27 21:11:59,187 - INFO - Epoch [10/10]
2025-04-27 21:11:59,187 - INFO - Training Loss: 0.2123
2025-04-27 21:12:01,765 - INFO - Validation Loss: 0.1047
2025-04-27 21:12:01,765 - INFO - Validation Accuracy: 97.02%
2025-04-27 21:12:01,770 - INFO - New best validation loss: 0.1047
2025-04-27 21:12:01,771 - INFO - Epoch 10 duration: 13.35 seconds
2025-04-27 21:12:01,771 - INFO - Training duration: 133.75 seconds
2025-04-27 21:12:01,772 - INFO - Completed all 10 epochs
2025-04-27 21:12:01,772 - INFO - Loading best model from: models/run_2025042
7_210947/model_variant_1/checkpoints/epoch_10.pth
2025-04-27 21:12:02,488 - INFO - ===========================================
=========================================================
2025-04-27 21:12:02,489 - INFO - Testing duration: 0.71 seconds
2025-04-27 21:12:02,489 - INFO - Total duration: 134.46 seconds
2025-04-27 21:12:02,490 - INFO - ===========================================
=========================================================
2025-04-27 21:12:02,490 - INFO - Best Model: models/run_20250427_210947/mode
l_variant_1/checkpoints/epoch_10.pth
2025-04-27 21:12:02,491 - INFO - Best Validation Loss: 0.1047
2025-04-27 21:12:02,491 - INFO - Test Loss: 0.7590
2025-04-27 21:12:02,492 - INFO - Test Accuracy: 72.61%
2025-04-27 21:12:02,492 - INFO - ===========================================
=========================================================
2025-04-27 21:12:02,493 - INFO - Completed training for model variant_1
2025-04-27 21:12:02,493 - INFO - Test accuracy: 72.61%
2025-04-27 21:12:02,494 - INFO - New best model: variant_1 with accuracy 72.
61%
2025-04-27 21:12:02,494 - INFO - ===========================================
=========================================================
2025-04-27 21:12:02,495 - INFO - Starting training for model variant_2
2025-04-27 21:12:02,495 - INFO - Hyperparameters:
2025-04-27 21:12:02,496 - INFO - Learning rate: 0.0001
2025-04-27 21:12:02,496 - INFO - Batch size: 64
2025-04-27 21:12:02,496 - INFO - L1 lambda: 1e-06
2025-04-27 21:12:02,497 - INFO - L2 lambda: 1e-05
2025-04-27 21:12:02,497 - INFO - Dropout rate: 0.25
2025-04-27 21:12:02,497 - INFO - Early stopping patience: 5
2025-04-27 21:12:02,505 - INFO - ===========================================
=========================================================
2025-04-27 21:12:02,506 - INFO - Model ID: variant_2
2025-04-27 21:12:02,506 - INFO - Run ID: 20250427_210947
2025-04-27 21:12:02,507 - INFO - Training configuration:
2025-04-27 21:12:02,507 - INFO -   Learning rate: 0.0001
2025-04-27 21:12:02,508 - INFO -   Batch size: 64
2025-04-27 21:12:02,508 - INFO -   L1 regularization: 1e-06
2025-04-27 21:12:02,508 - INFO -   L2 regularization: 1e-05
2025-04-27 21:12:02,509 - INFO -   Dropout rate: 0.25
2025-04-27 21:12:02,509 - INFO -   Epochs: 10
2025-04-27 21:12:02,510 - INFO -   Early stopping patience: 5
2025-04-27 21:12:02,510 - INFO -   Optimizer: Adam
2025-04-27 21:12:02,510 - INFO -   Loss function: CrossEntropyLoss
2025-04-27 21:12:14,092 - INFO - ===========================================
=========================================================
2025-04-27 21:12:14,093 - INFO - Epoch [1/10]
2025-04-27 21:12:14,093 - INFO - Training Loss: 0.6118
2025-04-27 21:12:16,632 - INFO - Validation Loss: 0.2700
```

```
2025-04-27 21:12:16,633 - INFO - Validation Accuracy: 95.02%
2025-04-27 21:12:16,637 - INFO - New best validation loss: 0.2700
2025-04-27 21:12:16,638 - INFO - Epoch 1 duration: 14.13 seconds
2025-04-27 21:12:28,173 - INFO - ===========================================
=========================================================
2025-04-27 21:12:28,174 - INFO - Epoch [2/10]
2025-04-27 21:12:28,174 - INFO - Training Loss: 0.2741
2025-04-27 21:12:30,713 - INFO - Validation Loss: 0.1626
2025-04-27 21:12:30,714 - INFO - Validation Accuracy: 96.41%
2025-04-27 21:12:30,719 - INFO - New best validation loss: 0.1626
2025-04-27 21:12:30,720 - INFO - Epoch 2 duration: 14.08 seconds
2025-04-27 21:12:42,410 - INFO - ===========================================
=========================================================
2025-04-27 21:12:42,410 - INFO - Epoch [3/10]
2025-04-27 21:12:42,411 - INFO - Training Loss: 0.2018
2025-04-27 21:12:44,989 - INFO - Validation Loss: 0.1217
2025-04-27 21:12:44,990 - INFO - Validation Accuracy: 96.96%
2025-04-27 21:12:44,994 - INFO - New best validation loss: 0.1217
2025-04-27 21:12:44,995 - INFO - Epoch 3 duration: 14.27 seconds
2025-04-27 21:12:56,704 - INFO - ===========================================
=========================================================
2025-04-27 21:12:56,704 - INFO - Epoch [4/10]
2025-04-27 21:12:56,705 - INFO - Training Loss: 0.1718
2025-04-27 21:12:59,243 - INFO - Validation Loss: 0.1069
2025-04-27 21:12:59,244 - INFO - Validation Accuracy: 97.31%
2025-04-27 21:12:59,249 - INFO - New best validation loss: 0.1069
2025-04-27 21:12:59,249 - INFO - Epoch 4 duration: 14.25 seconds
2025-04-27 21:13:10,742 - INFO - ===========================================
=========================================================
2025-04-27 21:13:10,743 - INFO - Epoch [5/10]
2025-04-27 21:13:10,743 - INFO - Training Loss: 0.1530
2025-04-27 21:13:13,264 - INFO - Validation Loss: 0.0955
2025-04-27 21:13:13,264 - INFO - Validation Accuracy: 97.39%
2025-04-27 21:13:13,269 - INFO - New best validation loss: 0.0955
2025-04-27 21:13:13,270 - INFO - Epoch 5 duration: 14.02 seconds
2025-04-27 21:13:24,701 - INFO - ===========================================
=========================================================
2025-04-27 21:13:24,702 - INFO - Epoch [6/10]
2025-04-27 21:13:24,703 - INFO - Training Loss: 0.1426
2025-04-27 21:13:27,230 - INFO - Validation Loss: 0.0873
2025-04-27 21:13:27,231 - INFO - Validation Accuracy: 97.51%
2025-04-27 21:13:27,236 - INFO - New best validation loss: 0.0873
2025-04-27 21:13:27,236 - INFO - Epoch 6 duration: 13.97 seconds
2025-04-27 21:13:38,784 - INFO - ===========================================
=========================================================
2025-04-27 21:13:38,785 - INFO - Epoch [7/10]
2025-04-27 21:13:38,786 - INFO - Training Loss: 0.1314
2025-04-27 21:13:41,308 - INFO - Validation Loss: 0.0784
2025-04-27 21:13:41,309 - INFO - Validation Accuracy: 97.90%
2025-04-27 21:13:41,313 - INFO - New best validation loss: 0.0784
2025-04-27 21:13:41,314 - INFO - Epoch 7 duration: 14.08 seconds
2025-04-27 21:13:52,809 - INFO - ===========================================
=========================================================
2025-04-27 21:13:52,809 - INFO - Epoch [8/10]
2025-04-27 21:13:52,810 - INFO - Training Loss: 0.1263
2025-04-27 21:13:55,375 - INFO - Validation Loss: 0.0768
```

```
2025-04-27 21:13:55,375 - INFO - Validation Accuracy: 97.78%
2025-04-27 21:13:55,380 - INFO - New best validation loss: 0.0768
2025-04-27 21:13:55,380 - INFO - Epoch 8 duration: 14.07 seconds
2025-04-27 21:14:06,971 - INFO - ====================================================================================================
2025-04-27 21:14:06,971 - INFO - Epoch [9/10]
2025-04-27 21:14:06,972 - INFO - Training Loss: 0.1191
2025-04-27 21:14:09,554 - INFO - Validation Loss: 0.0690
2025-04-27 21:14:09,555 - INFO - Validation Accuracy: 98.03%
2025-04-27 21:14:09,560 - INFO - New best validation loss: 0.0690
2025-04-27 21:14:09,560 - INFO - Epoch 9 duration: 14.18 seconds
2025-04-27 21:14:21,036 - INFO - ====================================================================================================
2025-04-27 21:14:21,037 - INFO - Epoch [10/10]
2025-04-27 21:14:21,037 - INFO - Training Loss: 0.1159
2025-04-27 21:14:23,567 - INFO - Validation Loss: 0.0719
2025-04-27 21:14:23,568 - INFO - Validation Accuracy: 97.90%
2025-04-27 21:14:23,573 - INFO - No improvement for 1 epochs (patience: 5)
2025-04-27 21:14:23,573 - INFO - Epoch 10 duration: 14.01 seconds
2025-04-27 21:14:23,574 - INFO - Training duration: 141.06 seconds
2025-04-27 21:14:23,574 - INFO - Completed all 10 epochs
2025-04-27 21:14:23,574 - INFO - Loading best model from: models/run_20250427_210947/model_variant_2/checkpoints/epoch_9.pth
2025-04-27 21:14:24,328 - INFO - ====================================================================================================
2025-04-27 21:14:24,329 - INFO - Testing duration: 0.75 seconds
2025-04-27 21:14:24,329 - INFO - Total duration: 141.82 seconds
2025-04-27 21:14:24,330 - INFO - ====================================================================================================
2025-04-27 21:14:24,330 - INFO - Best Model: models/run_20250427_210947/model_variant_2/checkpoints/epoch_9.pth
2025-04-27 21:14:24,331 - INFO - Best Validation Loss: 0.0690
2025-04-27 21:14:24,331 - INFO - Test Loss: 0.5928
2025-04-27 21:14:24,332 - INFO - Test Accuracy: 81.12%
2025-04-27 21:14:24,332 - INFO - ====================================================================================================
2025-04-27 21:14:24,333 - INFO - Completed training for model variant_2
2025-04-27 21:14:24,334 - INFO - Test accuracy: 81.12%
2025-04-27 21:14:24,334 - INFO - New best model: variant_2 with accuracy 81.12%
2025-04-27 21:14:24,334 - INFO - ====================================================================================================
2025-04-27 21:14:24,335 - INFO - Starting training for model variant_3
2025-04-27 21:14:24,335 - INFO - Hyperparameters:
2025-04-27 21:14:24,336 - INFO - Learning rate: 0.0001
2025-04-27 21:14:24,336 - INFO - Batch size: 64
2025-04-27 21:14:24,337 - INFO - L1 lambda: 0.0001
2025-04-27 21:14:24,337 - INFO - L2 lambda: 1e-05
2025-04-27 21:14:24,337 - INFO - Dropout rate: 0.5
2025-04-27 21:14:24,338 - INFO - Early stopping patience: 5
2025-04-27 21:14:24,345 - INFO - ====================================================================================================
2025-04-27 21:14:24,345 - INFO - Model ID: variant_3
2025-04-27 21:14:24,346 - INFO - Run ID: 20250427_210947
2025-04-27 21:14:24,346 - INFO - Training configuration:
2025-04-27 21:14:24,347 - INFO -   Learning rate: 0.0001
```

```
2025-04-27 21:14:24,347 - INFO -   Batch size: 64
2025-04-27 21:14:24,347 - INFO -   L1 regularization: 0.0001
2025-04-27 21:14:24,348 - INFO -   L2 regularization: 1e-05
2025-04-27 21:14:24,348 - INFO -   Dropout rate: 0.5
2025-04-27 21:14:24,349 - INFO -   Epochs: 10
2025-04-27 21:14:24,349 - INFO -   Early stopping patience: 5
2025-04-27 21:14:24,350 - INFO -   Optimizer: Adam
2025-04-27 21:14:24,350 - INFO -   Loss function: CrossEntropyLoss
2025-04-27 21:14:35,903 - INFO - ===========================================
=========================================================
2025-04-27 21:14:35,904 - INFO - Epoch [1/10]
2025-04-27 21:14:35,905 - INFO - Training Loss: 1.2636
2025-04-27 21:14:38,448 - INFO - Validation Loss: 0.3768
2025-04-27 21:14:38,449 - INFO - Validation Accuracy: 92.55%
2025-04-27 21:14:38,453 - INFO - New best validation loss: 0.3768
2025-04-27 21:14:38,454 - INFO - Epoch 1 duration: 14.10 seconds
2025-04-27 21:14:50,191 - INFO - ===========================================
=========================================================
2025-04-27 21:14:50,192 - INFO - Epoch [2/10]
2025-04-27 21:14:50,192 - INFO - Training Loss: 0.7511
2025-04-27 21:14:52,723 - INFO - Validation Loss: 0.2383
2025-04-27 21:14:52,724 - INFO - Validation Accuracy: 94.62%
2025-04-27 21:14:52,729 - INFO - New best validation loss: 0.2383
2025-04-27 21:14:52,729 - INFO - Epoch 2 duration: 14.27 seconds
2025-04-27 21:15:04,190 - INFO - ===========================================
=========================================================
2025-04-27 21:15:04,190 - INFO - Epoch [3/10]
2025-04-27 21:15:04,191 - INFO - Training Loss: 0.6004
2025-04-27 21:15:06,729 - INFO - Validation Loss: 0.1803
2025-04-27 21:15:06,730 - INFO - Validation Accuracy: 95.60%
2025-04-27 21:15:06,734 - INFO - New best validation loss: 0.1803
2025-04-27 21:15:06,735 - INFO - Epoch 3 duration: 14.01 seconds
2025-04-27 21:15:18,262 - INFO - ===========================================
=========================================================
2025-04-27 21:15:18,262 - INFO - Epoch [4/10]
2025-04-27 21:15:18,263 - INFO - Training Loss: 0.5409
2025-04-27 21:15:20,832 - INFO - Validation Loss: 0.1600
2025-04-27 21:15:20,833 - INFO - Validation Accuracy: 95.82%
2025-04-27 21:15:20,838 - INFO - New best validation loss: 0.1600
2025-04-27 21:15:20,838 - INFO - Epoch 4 duration: 14.10 seconds
2025-04-27 21:15:32,333 - INFO - ===========================================
=========================================================
2025-04-27 21:15:32,334 - INFO - Epoch [5/10]
2025-04-27 21:15:32,334 - INFO - Training Loss: 0.5071
2025-04-27 21:15:34,885 - INFO - Validation Loss: 0.1445
2025-04-27 21:15:34,886 - INFO - Validation Accuracy: 96.00%
2025-04-27 21:15:34,890 - INFO - New best validation loss: 0.1445
2025-04-27 21:15:34,891 - INFO - Epoch 5 duration: 14.05 seconds
2025-04-27 21:15:46,507 - INFO - ===========================================
=========================================================
2025-04-27 21:15:46,508 - INFO - Epoch [6/10]
2025-04-27 21:15:46,508 - INFO - Training Loss: 0.4833
2025-04-27 21:15:49,095 - INFO - Validation Loss: 0.1327
2025-04-27 21:15:49,096 - INFO - Validation Accuracy: 96.26%
2025-04-27 21:15:49,101 - INFO - New best validation loss: 0.1327
2025-04-27 21:15:49,101 - INFO - Epoch 6 duration: 14.21 seconds
```

```
2025-04-27 21:16:00,554 - INFO - ===========================================
=========================================================
2025-04-27 21:16:00,555 - INFO - Epoch [7/10]
2025-04-27 21:16:00,555 - INFO - Training Loss: 0.4685
2025-04-27 21:16:03,103 - INFO - Validation Loss: 0.1229
2025-04-27 21:16:03,103 - INFO - Validation Accuracy: 96.67%
2025-04-27 21:16:03,108 - INFO - New best validation loss: 0.1229
2025-04-27 21:16:03,109 - INFO - Epoch 7 duration: 14.01 seconds
2025-04-27 21:16:14,654 - INFO - ===========================================
=========================================================
2025-04-27 21:16:14,654 - INFO - Epoch [8/10]
2025-04-27 21:16:14,655 - INFO - Training Loss: 0.4614
2025-04-27 21:16:17,200 - INFO - Validation Loss: 0.1232
2025-04-27 21:16:17,200 - INFO - Validation Accuracy: 96.53%
2025-04-27 21:16:17,205 - INFO - No improvement for 1 epochs (patience: 5)
2025-04-27 21:16:17,206 - INFO - Epoch 8 duration: 14.10 seconds
2025-04-27 21:16:28,722 - INFO - ===========================================
=========================================================
2025-04-27 21:16:28,723 - INFO - Epoch [9/10]
2025-04-27 21:16:28,724 - INFO - Training Loss: 0.4503
2025-04-27 21:16:31,257 - INFO - Validation Loss: 0.1147
2025-04-27 21:16:31,257 - INFO - Validation Accuracy: 96.73%
2025-04-27 21:16:31,262 - INFO - New best validation loss: 0.1147
2025-04-27 21:16:31,263 - INFO - Epoch 9 duration: 14.06 seconds
2025-04-27 21:16:42,802 - INFO - ===========================================
=========================================================
2025-04-27 21:16:42,802 - INFO - Epoch [10/10]
2025-04-27 21:16:42,803 - INFO - Training Loss: 0.4405
2025-04-27 21:16:45,350 - INFO - Validation Loss: 0.1101
2025-04-27 21:16:45,350 - INFO - Validation Accuracy: 96.83%
2025-04-27 21:16:45,355 - INFO - New best validation loss: 0.1101
2025-04-27 21:16:45,356 - INFO - Epoch 10 duration: 14.09 seconds
2025-04-27 21:16:45,356 - INFO - Training duration: 141.01 seconds
2025-04-27 21:16:45,356 - INFO - Completed all 10 epochs
2025-04-27 21:16:45,357 - INFO - Loading best model from: models/run_2025042
7_210947/model_variant_3/checkpoints/epoch_10.pth
2025-04-27 21:16:46,110 - INFO - ===========================================
=========================================================
2025-04-27 21:16:46,110 - INFO - Testing duration: 0.75 seconds
2025-04-27 21:16:46,111 - INFO - Total duration: 141.76 seconds
2025-04-27 21:16:46,111 - INFO - ===========================================
=========================================================
2025-04-27 21:16:46,112 - INFO - Best Model: models/run_20250427_210947/mode
l_variant_3/checkpoints/epoch_10.pth
2025-04-27 21:16:46,112 - INFO - Best Validation Loss: 0.1101
2025-04-27 21:16:46,112 - INFO - Test Loss: 0.4928
2025-04-27 21:16:46,113 - INFO - Test Accuracy: 96.58%
2025-04-27 21:16:46,113 - INFO - ===========================================
=========================================================
2025-04-27 21:16:46,114 - INFO - Completed training for model variant_3
2025-04-27 21:16:46,114 - INFO - Test accuracy: 96.58%
2025-04-27 21:16:46,114 - INFO - New best model: variant_3 with accuracy 96.
58%
2025-04-27 21:16:46,115 - INFO - ===========================================
=========================================================
2025-04-27 21:16:46,115 - INFO - Starting training for model variant_4
```

```
2025-04-27 21:16:46,115 - INFO - Hyperparameters:
2025-04-27 21:16:46,116 - INFO - Learning rate: 0.0001
2025-04-27 21:16:46,116 - INFO - Batch size: 32
2025-04-27 21:16:46,116 - INFO - L1 lambda: 1e-06
2025-04-27 21:16:46,117 - INFO - L2 lambda: 0.0001
2025-04-27 21:16:46,117 - INFO - Dropout rate: 0.5
2025-04-27 21:16:46,117 - INFO - Early stopping patience: 5
2025-04-27 21:16:46,124 - INFO - ===========================================
=========================================================
2025-04-27 21:16:46,125 - INFO - Model ID: variant_4
2025-04-27 21:16:46,125 - INFO - Run ID: 20250427_210947
2025-04-27 21:16:46,125 - INFO - Training configuration:
2025-04-27 21:16:46,126 - INFO -   Learning rate: 0.0001
2025-04-27 21:16:46,127 - INFO -   Batch size: 32
2025-04-27 21:16:46,127 - INFO -   L1 regularization: 1e-06
2025-04-27 21:16:46,127 - INFO -   L2 regularization: 0.0001
2025-04-27 21:16:46,128 - INFO -   Dropout rate: 0.5
2025-04-27 21:16:46,128 - INFO -   Epochs: 10
2025-04-27 21:16:46,129 - INFO -   Early stopping patience: 5
2025-04-27 21:16:46,129 - INFO -   Optimizer: Adam
2025-04-27 21:16:46,129 - INFO -   Loss function: CrossEntropyLoss
2025-04-27 21:16:59,404 - INFO - ===========================================
=========================================================
2025-04-27 21:16:59,405 - INFO - Epoch [1/10]
2025-04-27 21:16:59,406 - INFO - Training Loss: 0.7898
2025-04-27 21:17:02,092 - INFO - Validation Loss: 0.3410
2025-04-27 21:17:02,093 - INFO - Validation Accuracy: 92.63%
2025-04-27 21:17:02,098 - INFO - New best validation loss: 0.3410
2025-04-27 21:17:02,098 - INFO - Epoch 1 duration: 15.97 seconds
2025-04-27 21:17:15,312 - INFO - ===========================================
=========================================================
2025-04-27 21:17:15,313 - INFO - Epoch [2/10]
2025-04-27 21:17:15,313 - INFO - Training Loss: 0.4159
2025-04-27 21:17:17,950 - INFO - Validation Loss: 0.2074
2025-04-27 21:17:17,951 - INFO - Validation Accuracy: 94.55%
2025-04-27 21:17:17,956 - INFO - New best validation loss: 0.2074
2025-04-27 21:17:17,956 - INFO - Epoch 2 duration: 15.86 seconds
2025-04-27 21:17:31,162 - INFO - ===========================================
=========================================================
2025-04-27 21:17:31,163 - INFO - Epoch [3/10]
2025-04-27 21:17:31,163 - INFO - Training Loss: 0.3297
2025-04-27 21:17:33,817 - INFO - Validation Loss: 0.1599
2025-04-27 21:17:33,818 - INFO - Validation Accuracy: 95.72%
2025-04-27 21:17:33,823 - INFO - New best validation loss: 0.1599
2025-04-27 21:17:33,823 - INFO - Epoch 3 duration: 15.87 seconds
2025-04-27 21:17:46,925 - INFO - ===========================================
=========================================================
2025-04-27 21:17:46,925 - INFO - Epoch [4/10]
2025-04-27 21:17:46,926 - INFO - Training Loss: 0.2942
2025-04-27 21:17:49,624 - INFO - Validation Loss: 0.1410
2025-04-27 21:17:49,625 - INFO - Validation Accuracy: 96.16%
2025-04-27 21:17:49,631 - INFO - New best validation loss: 0.1410
2025-04-27 21:17:49,632 - INFO - Epoch 4 duration: 15.81 seconds
2025-04-27 21:18:03,071 - INFO - ===========================================
=========================================================
2025-04-27 21:18:03,072 - INFO - Epoch [5/10]
```

```
2025-04-27 21:18:03,072 - INFO - Training Loss: 0.2730
2025-04-27 21:18:05,717 - INFO - Validation Loss: 0.1268
2025-04-27 21:18:05,718 - INFO - Validation Accuracy: 96.47%
2025-04-27 21:18:05,724 - INFO - New best validation loss: 0.1268
2025-04-27 21:18:05,724 - INFO - Epoch 5 duration: 16.09 seconds
2025-04-27 21:18:19,046 - INFO - ===========================================
=========================================================
2025-04-27 21:18:19,047 - INFO - Epoch [6/10]
2025-04-27 21:18:19,047 - INFO - Training Loss: 0.2562
2025-04-27 21:18:21,699 - INFO - Validation Loss: 0.1150
2025-04-27 21:18:21,699 - INFO - Validation Accuracy: 96.78%
2025-04-27 21:18:21,704 - INFO - New best validation loss: 0.1150
2025-04-27 21:18:21,705 - INFO - Epoch 6 duration: 15.98 seconds
2025-04-27 21:18:34,869 - INFO - ===========================================
=========================================================
2025-04-27 21:18:34,870 - INFO - Epoch [7/10]
2025-04-27 21:18:34,871 - INFO - Training Loss: 0.2418
2025-04-27 21:18:37,510 - INFO - Validation Loss: 0.1126
2025-04-27 21:18:37,511 - INFO - Validation Accuracy: 96.71%
2025-04-27 21:18:37,516 - INFO - New best validation loss: 0.1126
2025-04-27 21:18:37,516 - INFO - Epoch 7 duration: 15.81 seconds
2025-04-27 21:18:50,728 - INFO - ===========================================
=========================================================
2025-04-27 21:18:50,729 - INFO - Epoch [8/10]
2025-04-27 21:18:50,730 - INFO - Training Loss: 0.2330
2025-04-27 21:18:53,378 - INFO - Validation Loss: 0.1097
2025-04-27 21:18:53,379 - INFO - Validation Accuracy: 96.84%
2025-04-27 21:18:53,384 - INFO - New best validation loss: 0.1097
2025-04-27 21:18:53,385 - INFO - Epoch 8 duration: 15.87 seconds
2025-04-27 21:19:06,585 - INFO - ===========================================
=========================================================
2025-04-27 21:19:06,586 - INFO - Epoch [9/10]
2025-04-27 21:19:06,586 - INFO - Training Loss: 0.2306
2025-04-27 21:19:09,242 - INFO - Validation Loss: 0.1018
2025-04-27 21:19:09,242 - INFO - Validation Accuracy: 96.91%
2025-04-27 21:19:09,247 - INFO - New best validation loss: 0.1018
2025-04-27 21:19:09,248 - INFO - Epoch 9 duration: 15.86 seconds
2025-04-27 21:19:22,462 - INFO - ===========================================
=========================================================
2025-04-27 21:19:22,463 - INFO - Epoch [10/10]
2025-04-27 21:19:22,464 - INFO - Training Loss: 0.2199
2025-04-27 21:19:25,131 - INFO - Validation Loss: 0.0949
2025-04-27 21:19:25,132 - INFO - Validation Accuracy: 97.24%
2025-04-27 21:19:25,137 - INFO - New best validation loss: 0.0949
2025-04-27 21:19:25,137 - INFO - Epoch 10 duration: 15.89 seconds
2025-04-27 21:19:25,138 - INFO - Training duration: 159.01 seconds
2025-04-27 21:19:25,138 - INFO - Completed all 10 epochs
2025-04-27 21:19:25,139 - INFO - Loading best model from: models/run_2025042
7_210947/model_variant_4/checkpoints/epoch_10.pth
2025-04-27 21:19:25,982 - INFO - ===========================================
=========================================================
2025-04-27 21:19:25,983 - INFO - Testing duration: 0.84 seconds
2025-04-27 21:19:25,984 - INFO - Total duration: 159.85 seconds
2025-04-27 21:19:25,984 - INFO - ===========================================
=========================================================
2025-04-27 21:19:25,985 - INFO - Best Model: models/run_20250427_210947/mode
```

```
l_variant_4/checkpoints/epoch_10.pth
2025-04-27 21:19:25,985 - INFO - Best Validation Loss: 0.0949
2025-04-27 21:19:25,986 - INFO - Test Loss: 0.6466
2025-04-27 21:19:25,986 - INFO - Test Accuracy: 81.94%
2025-04-27 21:19:25,986 - INFO - ===========================================
=========================================================
2025-04-27 21:19:25,987 - INFO - Completed training for model variant_4
2025-04-27 21:19:25,988 - INFO - Test accuracy: 81.94%
2025-04-27 21:19:25,988 - INFO - ===========================================
=========================================================
2025-04-27 21:19:25,989 - INFO - Starting training for model variant_5
2025-04-27 21:19:25,989 - INFO - Hyperparameters:
2025-04-27 21:19:25,990 - INFO - Learning rate: 0.0001
2025-04-27 21:19:25,990 - INFO - Batch size: 32
2025-04-27 21:19:25,990 - INFO - L1 lambda: 1e-05
2025-04-27 21:19:25,991 - INFO - L2 lambda: 0.0001
2025-04-27 21:19:25,991 - INFO - Dropout rate: 0.25
2025-04-27 21:19:25,992 - INFO - Early stopping patience: 5
2025-04-27 21:19:25,998 - INFO - ===========================================
=========================================================
2025-04-27 21:19:25,999 - INFO - Model ID: variant_5
2025-04-27 21:19:25,999 - INFO - Run ID: 20250427_210947
2025-04-27 21:19:26,000 - INFO - Training configuration:
2025-04-27 21:19:26,000 - INFO -   Learning rate: 0.0001
2025-04-27 21:19:26,001 - INFO -   Batch size: 32
2025-04-27 21:19:26,001 - INFO -   L1 regularization: 1e-05
2025-04-27 21:19:26,002 - INFO -   L2 regularization: 0.0001
2025-04-27 21:19:26,002 - INFO -   Dropout rate: 0.25
2025-04-27 21:19:26,002 - INFO -   Epochs: 10
2025-04-27 21:19:26,003 - INFO -   Early stopping patience: 5
2025-04-27 21:19:26,004 - INFO -   Optimizer: Adam
2025-04-27 21:19:26,004 - INFO -   Loss function: CrossEntropyLoss
2025-04-27 21:19:39,321 - INFO - ===========================================
=========================================================
2025-04-27 21:19:39,322 - INFO - Epoch [1/10]
2025-04-27 21:19:39,322 - INFO - Training Loss: 0.5892
2025-04-27 21:19:41,974 - INFO - Validation Loss: 0.2142
2025-04-27 21:19:41,975 - INFO - Validation Accuracy: 95.28%
2025-04-27 21:19:41,979 - INFO - New best validation loss: 0.2142
2025-04-27 21:19:41,980 - INFO - Epoch 1 duration: 15.98 seconds
2025-04-27 21:19:55,153 - INFO - ===========================================
=========================================================
2025-04-27 21:19:55,154 - INFO - Epoch [2/10]
2025-04-27 21:19:55,154 - INFO - Training Loss: 0.2850
2025-04-27 21:19:57,797 - INFO - Validation Loss: 0.1352
2025-04-27 21:19:57,797 - INFO - Validation Accuracy: 96.44%
2025-04-27 21:19:57,802 - INFO - New best validation loss: 0.1352
2025-04-27 21:19:57,803 - INFO - Epoch 2 duration: 15.82 seconds
2025-04-27 21:20:11,008 - INFO - ===========================================
=========================================================
2025-04-27 21:20:11,009 - INFO - Epoch [3/10]
2025-04-27 21:20:11,009 - INFO - Training Loss: 0.2319
2025-04-27 21:20:13,644 - INFO - Validation Loss: 0.1120
2025-04-27 21:20:13,644 - INFO - Validation Accuracy: 96.85%
2025-04-27 21:20:13,649 - INFO - New best validation loss: 0.1120
2025-04-27 21:20:13,650 - INFO - Epoch 3 duration: 15.85 seconds
```

```
2025-04-27 21:20:26,897 - INFO - ===========================================
=========================================================
2025-04-27 21:20:26,898 - INFO - Epoch [4/10]
2025-04-27 21:20:26,899 - INFO - Training Loss: 0.2070
2025-04-27 21:20:29,563 - INFO - Validation Loss: 0.0950
2025-04-27 21:20:29,564 - INFO - Validation Accuracy: 97.59%
2025-04-27 21:20:29,569 - INFO - New best validation loss: 0.0950
2025-04-27 21:20:29,570 - INFO - Epoch 4 duration: 15.92 seconds
2025-04-27 21:20:42,766 - INFO - ===========================================
=========================================================
2025-04-27 21:20:42,767 - INFO - Epoch [5/10]
2025-04-27 21:20:42,767 - INFO - Training Loss: 0.1925
2025-04-27 21:20:45,424 - INFO - Validation Loss: 0.0864
2025-04-27 21:20:45,425 - INFO - Validation Accuracy: 97.49%
2025-04-27 21:20:45,429 - INFO - New best validation loss: 0.0864
2025-04-27 21:20:45,430 - INFO - Epoch 5 duration: 15.86 seconds
2025-04-27 21:20:58,675 - INFO - ===========================================
=========================================================
2025-04-27 21:20:58,675 - INFO - Epoch [6/10]
2025-04-27 21:20:58,676 - INFO - Training Loss: 0.1898
2025-04-27 21:21:01,330 - INFO - Validation Loss: 0.0789
2025-04-27 21:21:01,331 - INFO - Validation Accuracy: 97.85%
2025-04-27 21:21:01,335 - INFO - New best validation loss: 0.0789
2025-04-27 21:21:01,336 - INFO - Epoch 6 duration: 15.91 seconds
2025-04-27 21:21:14,555 - INFO - ===========================================
=========================================================
2025-04-27 21:21:14,556 - INFO - Epoch [7/10]
2025-04-27 21:21:14,557 - INFO - Training Loss: 0.1771
2025-04-27 21:21:17,202 - INFO - Validation Loss: 0.0707
2025-04-27 21:21:17,202 - INFO - Validation Accuracy: 98.07%
2025-04-27 21:21:17,207 - INFO - New best validation loss: 0.0707
2025-04-27 21:21:17,208 - INFO - Epoch 7 duration: 15.87 seconds
2025-04-27 21:21:30,433 - INFO - ===========================================
=========================================================
2025-04-27 21:21:30,434 - INFO - Epoch [8/10]
2025-04-27 21:21:30,434 - INFO - Training Loss: 0.1711
2025-04-27 21:21:33,101 - INFO - Validation Loss: 0.0707
2025-04-27 21:21:33,101 - INFO - Validation Accuracy: 97.92%
2025-04-27 21:21:33,106 - INFO - New best validation loss: 0.0707
2025-04-27 21:21:33,107 - INFO - Epoch 8 duration: 15.90 seconds
2025-04-27 21:21:46,335 - INFO - ===========================================
=========================================================
2025-04-27 21:21:46,336 - INFO - Epoch [9/10]
2025-04-27 21:21:46,336 - INFO - Training Loss: 0.1665
2025-04-27 21:21:48,976 - INFO - Validation Loss: 0.0713
2025-04-27 21:21:48,977 - INFO - Validation Accuracy: 97.91%
2025-04-27 21:21:48,982 - INFO - No improvement for 1 epochs (patience: 5)
2025-04-27 21:21:48,982 - INFO - Epoch 9 duration: 15.88 seconds
2025-04-27 21:22:02,207 - INFO - ===========================================
=========================================================
2025-04-27 21:22:02,208 - INFO - Epoch [10/10]
2025-04-27 21:22:02,209 - INFO - Training Loss: 0.1618
2025-04-27 21:22:04,873 - INFO - Validation Loss: 0.0668
2025-04-27 21:22:04,874 - INFO - Validation Accuracy: 98.03%
2025-04-27 21:22:04,879 - INFO - New best validation loss: 0.0668
2025-04-27 21:22:04,879 - INFO - Epoch 10 duration: 15.90 seconds
```

```
2025-04-27 21:22:04,880 - INFO - Training duration: 158.88 seconds
2025-04-27 21:22:04,880 - INFO - Completed all 10 epochs
2025-04-27 21:22:04,881 - INFO - Loading best model from: models/run_2025042
7_210947/model_variant_5/checkpoints/epoch_10.pth
2025-04-27 21:22:05,722 - INFO - ===========================================
=========================================================
2025-04-27 21:22:05,723 - INFO - Testing duration: 0.84 seconds
2025-04-27 21:22:05,723 - INFO - Total duration: 159.72 seconds
2025-04-27 21:22:05,724 - INFO - ===========================================
=========================================================
2025-04-27 21:22:05,724 - INFO - Best Model: models/run_20250427_210947/mode
l_variant_5/checkpoints/epoch_10.pth
2025-04-27 21:22:05,724 - INFO - Best Validation Loss: 0.0668
2025-04-27 21:22:05,725 - INFO - Test Loss: 0.5400
2025-04-27 21:22:05,725 - INFO - Test Accuracy: 91.01%
2025-04-27 21:22:05,725 - INFO - ===========================================
=========================================================
2025-04-27 21:22:05,726 - INFO - Completed training for model variant_5
2025-04-27 21:22:05,727 - INFO - Test accuracy: 91.01%
2025-04-27 21:22:05,727 - INFO - ===========================================
=========================================================
2025-04-27 21:22:05,727 - INFO - Training completed for all models
2025-04-27 21:22:05,728 - INFO - Summary of model performances:
2025-04-27 21:22:05,728 - INFO - Model variant_3: Accuracy 96.58%
2025-04-27 21:22:05,729 - INFO -   Learning rate: 0.0001
2025-04-27 21:22:05,729 - INFO -   Batch size: 64
2025-04-27 21:22:05,730 - INFO -   L1 lambda: 0.0001
2025-04-27 21:22:05,730 - INFO -   L2 lambda: 1e-05
2025-04-27 21:22:05,730 - INFO -   Dropout rate: 0.5
2025-04-27 21:22:05,731 - INFO -   Early stopping patience: 5
2025-04-27 21:22:05,731 - INFO - -------------------------------------------
-------
2025-04-27 21:22:05,732 - INFO - Model variant_5: Accuracy 91.01%
2025-04-27 21:22:05,732 - INFO -   Learning rate: 0.0001
2025-04-27 21:22:05,733 - INFO -   Batch size: 32
2025-04-27 21:22:05,733 - INFO -   L1 lambda: 1e-05
2025-04-27 21:22:05,734 - INFO -   L2 lambda: 0.0001
2025-04-27 21:22:05,734 - INFO -   Dropout rate: 0.25
2025-04-27 21:22:05,734 - INFO -   Early stopping patience: 5
2025-04-27 21:22:05,735 - INFO - -------------------------------------------
-------
2025-04-27 21:22:05,735 - INFO - Model variant_4: Accuracy 81.94%
2025-04-27 21:22:05,736 - INFO -   Learning rate: 0.0001
2025-04-27 21:22:05,736 - INFO -   Batch size: 32
2025-04-27 21:22:05,737 - INFO -   L1 lambda: 1e-06
2025-04-27 21:22:05,737 - INFO -   L2 lambda: 0.0001
2025-04-27 21:22:05,738 - INFO -   Dropout rate: 0.5
2025-04-27 21:22:05,738 - INFO -   Early stopping patience: 5
2025-04-27 21:22:05,739 - INFO - -------------------------------------------
-------
2025-04-27 21:22:05,739 - INFO - Model variant_2: Accuracy 81.12%
2025-04-27 21:22:05,739 - INFO -   Learning rate: 0.0001
2025-04-27 21:22:05,740 - INFO -   Batch size: 64
2025-04-27 21:22:05,740 - INFO -   L1 lambda: 1e-06
2025-04-27 21:22:05,740 - INFO -   L2 lambda: 1e-05
2025-04-27 21:22:05,741 - INFO -   Dropout rate: 0.25
```

```
2025-04-27 21:22:05,741 - INFO -   Early stopping patience: 5
2025-04-27 21:22:05,741 - INFO - -------------------------------------------
-------
2025-04-27 21:22:05,742 - INFO - Model variant_1: Accuracy 72.61%
2025-04-27 21:22:05,742 - INFO -   Learning rate: 0.0001
2025-04-27 21:22:05,743 - INFO -   Batch size: 128
2025-04-27 21:22:05,744 - INFO -   L1 lambda: 1e-06
2025-04-27 21:22:05,744 - INFO -   L2 lambda: 1e-05
2025-04-27 21:22:05,744 - INFO -   Dropout rate: 0.5
2025-04-27 21:22:05,745 - INFO -   Early stopping patience: 5
2025-04-27 21:22:05,745 - INFO - -------------------------------------------
-------
2025-04-27 21:22:05,745 - INFO - ===========================================
=========================================================
2025-04-27 21:22:05,746 - INFO - Best model: variant_3 with accuracy: 96.58%
2025-04-27 21:22:05,746 - INFO - Best model hyperparameters:
2025-04-27 21:22:05,746 - INFO -   Learning rate: 0.0001
2025-04-27 21:22:05,747 - INFO -   Batch size: 64
2025-04-27 21:22:05,747 - INFO -   L1 lambda: 0.0001
2025-04-27 21:22:05,747 - INFO -   L2 lambda: 1e-05
2025-04-27 21:22:05,748 - INFO -   Dropout rate: 0.5
2025-04-27 21:22:05,748 - INFO -   Early stopping patience: 5
2025-04-27 21:22:05,749 - INFO - ===========================================
=========================================================
```