# UE23CS352A: ML Lab Week 14: CNN Image Classification

**Student Name: Mohammed Shehzaad Khan**

**SRN: PES2UG23CS349**

**Course: UE23CS352A - Machine Learning**

**Date: November 20th , 2025**

---

## 1. Introduction

This laboratory exercise focused on implementing a Convolutional Neural Network (CNN) using PyTorch to classify hand gesture images into three categories: rock, paper, and scissors. The primary objective was to gain hands-on experience in building deep learning models for image classification tasks, including data preprocessing, model architecture design, training implementation, and performance evaluation. The lab utilized a dataset of 2,188 images, split into 80% training and 20% testing sets, to develop and validate a CNN capable of accurately recognizing hand gestures in the rock-paper-scissors game.

## 2. Model Architecture

The implemented CNN architecture consists of three convolutional blocks followed by a fully-connected classifier. Each convolutional block contains a Conv2d layer, ReLU activation function, and MaxPool2d layer with a kernel size of 2. The first convolutional layer accepts 3 input channels (RGB images) and produces 16 output channels using a 3×3 kernel with padding of 1. The second convolutional layer transforms 16 channels to 32 channels, while the third layer increases dimensionality from 32 to 64 channels, both using the same 3×3 kernel size and padding configuration. The MaxPooling layers reduce spatial dimensions by half after each convolutional block, progressively reducing the 128×128 input images to 16×16 feature maps. The fully-connected classifier begins with a Flatten layer that converts the 64×16×16 feature maps into a 16,384-dimensional vector. This is followed by a Linear layer reducing dimensions to 256, a ReLU activation, a Dropout layer with probability 0.3 for regularization, and a final Linear layer producing 3 output classes corresponding to rock, paper, and scissors predictions.

## 3. Training and Performance

The model was trained using the Adam optimizer with a learning rate of 0.001, which provides adaptive learning rates for efficient convergence. CrossEntropyLoss was selected as the loss function, appropriate for multi-class classification problems. Training was conducted for 10 epochs with a batch size of 32 images. The training process demonstrated excellent convergence, with the loss decreasing from 0.6375 in the first epoch to 0.0040 by the final epoch, indicating effective learning of discriminative features. Upon evaluation on the unseen test set of 438 images, the model achieved a test accuracy of 97.49%, correctly classifying 427 out of 438 images. This high accuracy demonstrates the model's strong generalization capability and its effectiveness in distinguishing between the three hand gesture classes.

## 4. Conclusion and Analysis

**The developed CNN model performed exceptionally well, achieving 97.49% test accuracy, which indicates robust feature learning and excellent generalization to unseen data. The low final training loss combined with high test accuracy suggests that the model successfully avoided overfitting, likely due to the inclusion of dropout regularization and appropriate model capacity. One challenge encountered during implementation was ensuring proper tensor dimensions throughout the network, particularly when transitioning from convolutional layers to fully-connected layers, which required careful calculation of the flattened feature size. To potentially improve model accuracy further, data augmentation techniques such as random rotations, flips, and brightness adjustments could be applied to increase training data diversity and model robustness. Additionally, implementing learning rate scheduling or early stopping mechanisms could optimize convergence and prevent overfitting in extended training scenarios. Overall, the lab successfully demonstrated practical skills in designing, implementing, and evaluating deep learning models for computer vision applications.**