# Open-Ended Lab Report

## 1️⃣ Introduction

This lab focused on classifying handwritten digits from the MNIST dataset using machine learning models. The goal was to compare different approaches and select the best-performing model.  
  
We specifically worked with:  
✅ Logistic Regression – A linear model for classification.  
✅ Artificial Neural Networks (ANN) – A deep learning model for complex pattern recognition.  
  
To improve performance, we applied data preprocessing, PCA (dimensionality reduction), and model evaluation techniques.

## 2️⃣ Dataset & Preprocessing

The MNIST dataset consists of grayscale images of handwritten digits (0-9), each sized 28x28 pixels, flattened into 784 features.  
  
✔ Missing Value Handling – Filled missing values with the median.  
✔ Normalization – Scaled pixel values between 0-1 for stable training.  
✔ Dimensionality Reduction (PCA) – Reduced features while keeping 95% variance, improving efficiency.

## 3️⃣ Machine Learning Models & Results

### 1️⃣ Logistic Regression

• Works well for linearly separable data.  
• Achieved Accuracy: 72.73%

📌 Confusion Matrix (Logistic Regression):

| Predicted | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |  
|-----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|  
| Actual 0 | 98 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |  
| Actual 1 | 0 | 97 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |  
| Actual 2 | 1 | 1 | 94 | 2 | 1 | 0 | 0 | 1 | 0 | 0 |

### 2️⃣ Artificial Neural Network (ANN)

• Multi-layer perceptron (MLP) with:  
 - 128 → 64 → 10 neurons  
 - ReLU activation for hidden layers, Softmax for output  
• Achieved Accuracy: 71.49%

📌 Confusion Matrix (ANN):

| Predicted | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |  
|-----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|  
| Actual 0 | 96 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |  
| Actual 1 | 1 | 95 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 |  
| Actual 2 | 2 | 0 | 93 | 2 | 1 | 0 | 0 | 1 | 1 | 0 |

## 4️⃣ Model Comparison

|  |  |  |
| --- | --- | --- |
| Model | Accuracy | Best Use Case |
| Logistic Regression | 72.73% | Best for linearly separable data. |
| Artificial Neural Network (ANN) | 71.49% | Best for complex patterns in images. |

## 5️⃣ Conclusion

🔹 Logistic Regression (72.73%) outperformed ANN (71.49%) in this case.  
🔹 ANN remains a strong alternative, especially for deep learning applications.  
🔹 PCA improved efficiency by reducing redundant features.  
  
🚀 Final Decision:  
✅ Logistic Regression is the best choice for this dataset, but ANN is a viable alternative.