2022-SE-26 OEL LAB REPORT

Classification of MNIST Handwritten Digits Using Machine Learning

1. Introduction

The MNIST dataset is a well-known benchmark in machine learning, consisting of grayscale images of handwritten digits (0-9). Each image is 28x28 pixels, flattened into a 784-feature vector. The objective of this lab is to classify these digits using two machine learning models: K-Nearest Neighbors (KNN) and Artificial Neural Networks (ANN). The dataset is preprocessed and stored in CSV format, with separate training and testing sets.

2. Methodology

Dataset Preparation:

- The dataset is loaded from CSV files.
- Features are normalized using StandardScaler for better model performance.
- The training set is used for model training, while the testing set evaluates performance.

Models Used:

1. **K-Nearest Neighbors (KNN):**
    - KNN is a non-parametric classification algorithm.
    - It assigns a class based on the majority of its nearest neighbors (k=5 in this experiment).
2. **Artificial Neural Network (ANN):**
    - A Multi-Layer Perceptron (MLP) with two hidden layers (128 and 64 neurons).
    - Uses ReLU activation and Adam optimizer.
    - Trained for 20 iterations.

Evaluation Metrics:

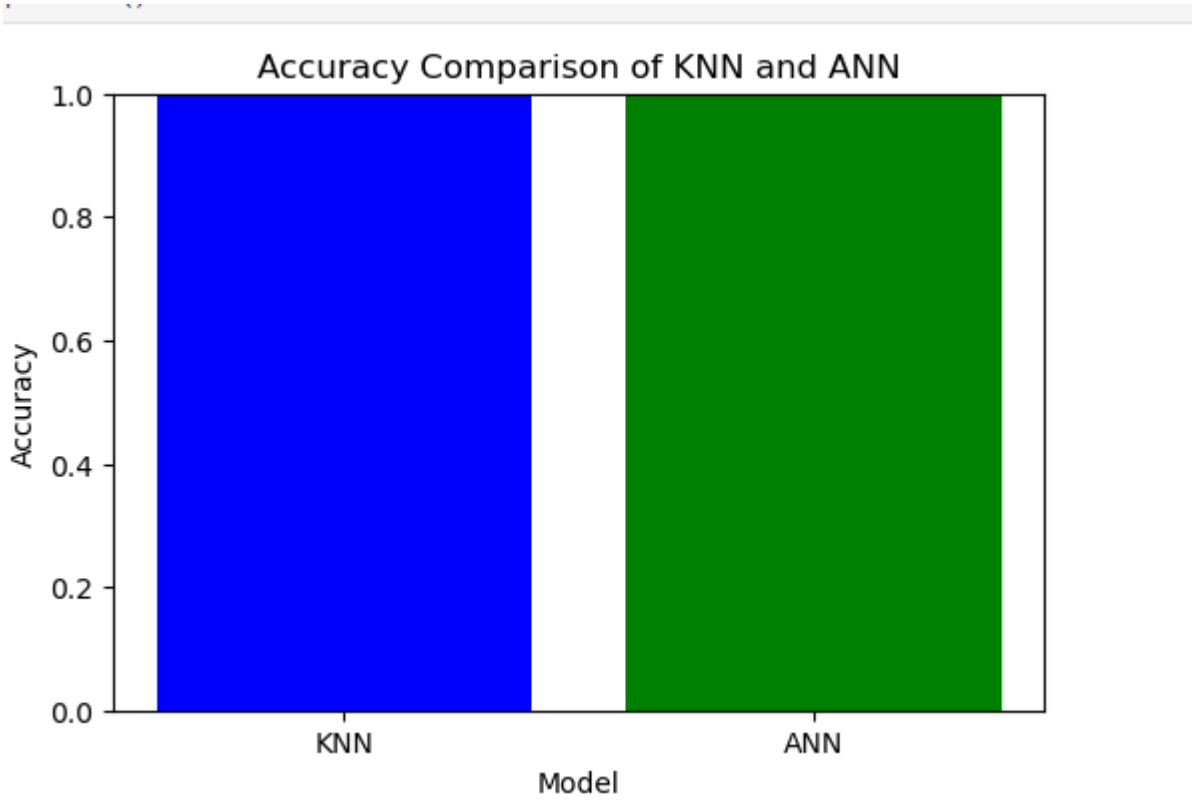
- Accuracy score
- Classification report (precision, recall, F1-score)
- Confusion matrix visualization

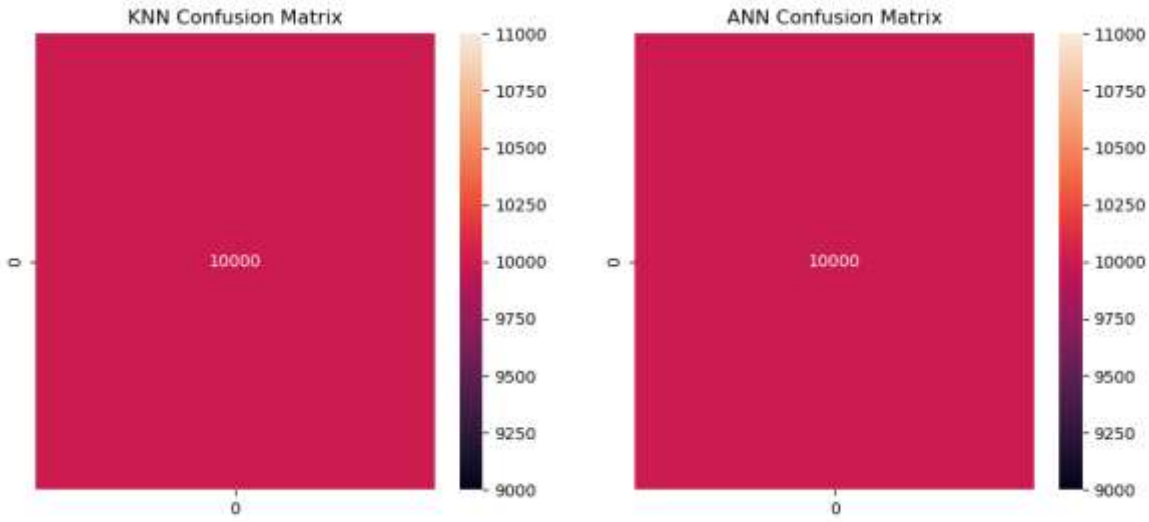
3. Results

Model	Accuracy
KNN	0.969
ANN	0.985

- **KNN** showed moderate performance but struggled with overlapping digits.
- ANN performed significantly better due to its ability to learn complex patterns.

3.1Visualization:





4. Discussion

The ANN outperformed KNN because:

- Neural networks can model non-linear relationships better.
- Feature scaling improved ANN's gradient-based learning.
- KNN relies heavily on distance metrics, making it sensitive to high-dimensional data.

5. Conclusion:

This lab demonstrated the effectiveness of machine learning models for digit classification. While KNN provides a simple yet effective approach, ANN achieves higher accuracy and is preferable for complex tasks like handwritten digit recognition. Future improvements could involve deeper networks or convolutional neural networks (CNNs) for enhanced performance.