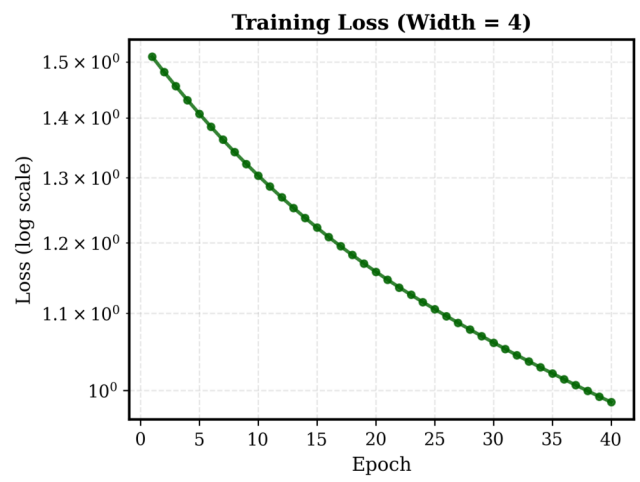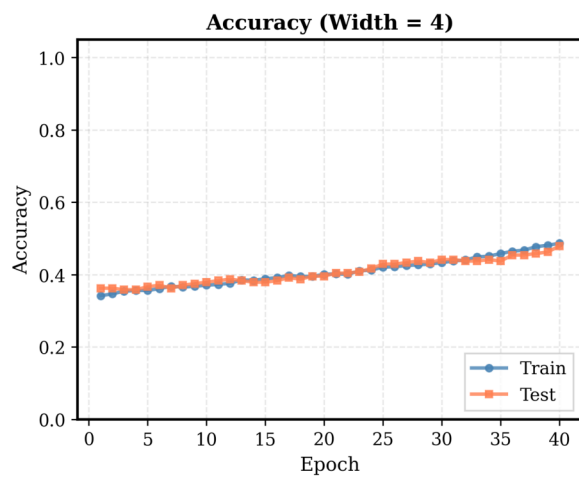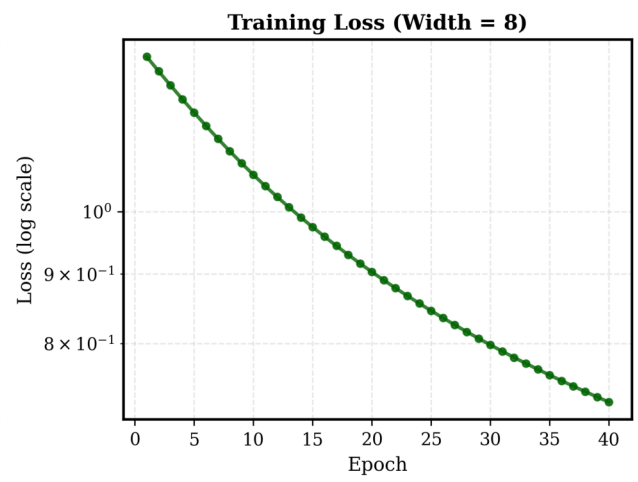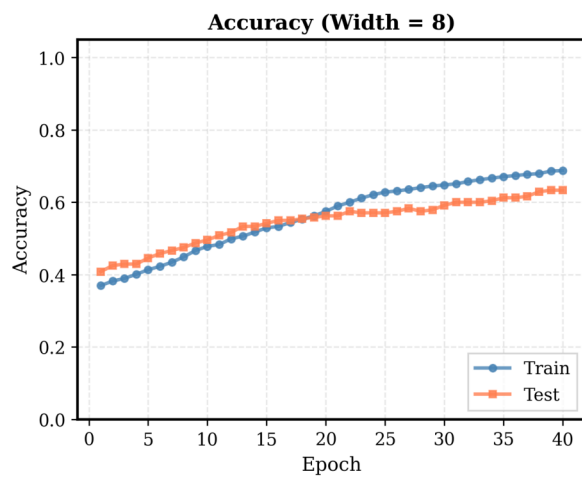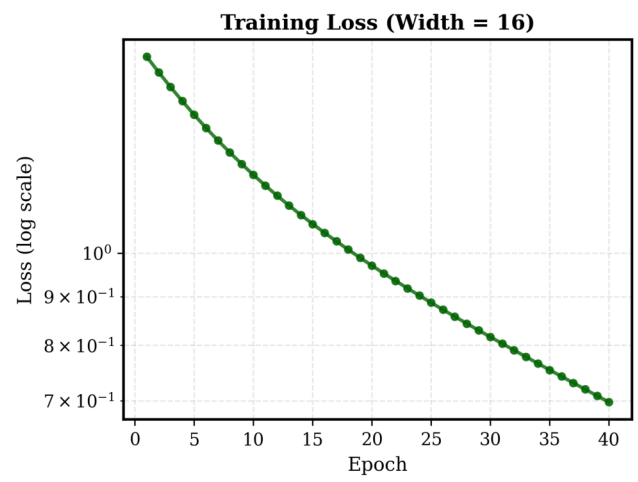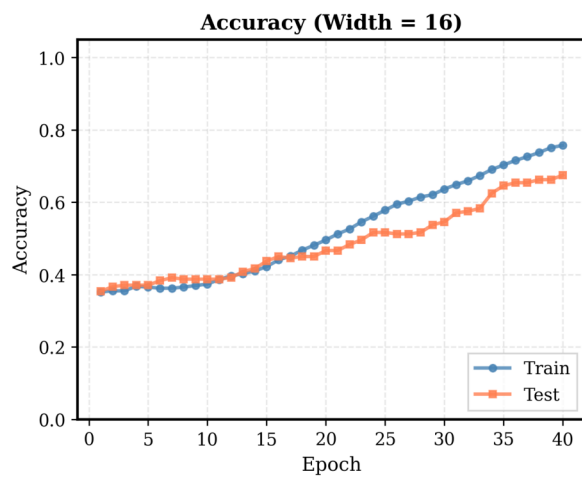# How Hidden Layer Width Influences the Performance of a

This tutorial studies how the number of neurons in a single hidden layer affects learning, generalisation and training behaviour. We test widths: 4, 8, 16, 32, 64, 128.
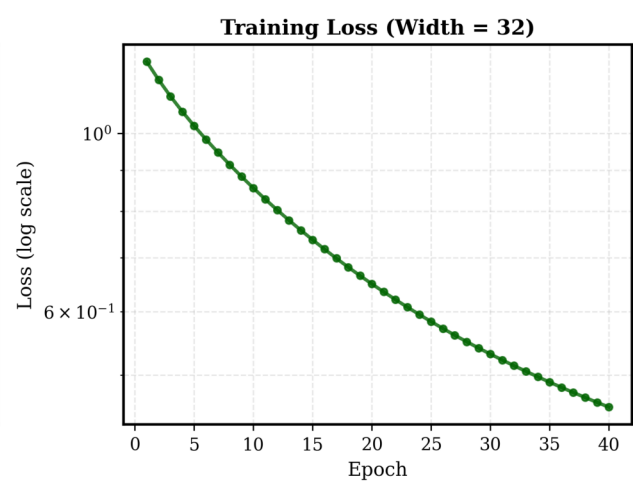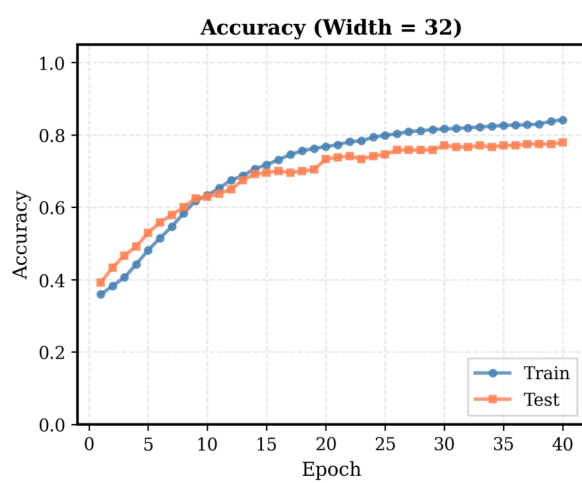
Key ideas: width controls model capacity. Small widths lead to underfitting, very large widths often lead to overfitting. The universal approximation theorem guarantees representational ability with sufficient width but says nothing about generalisation.
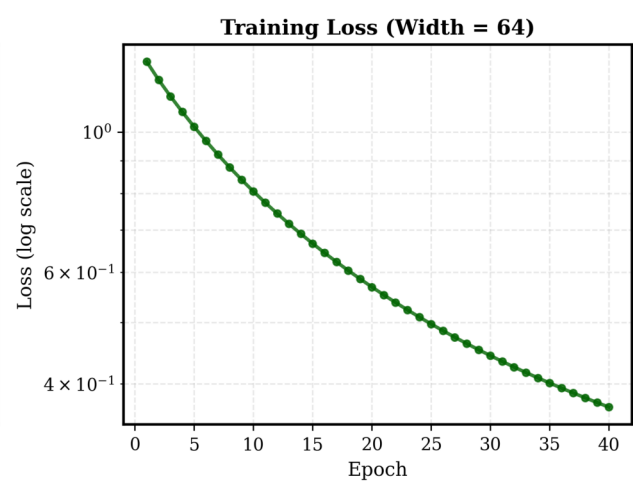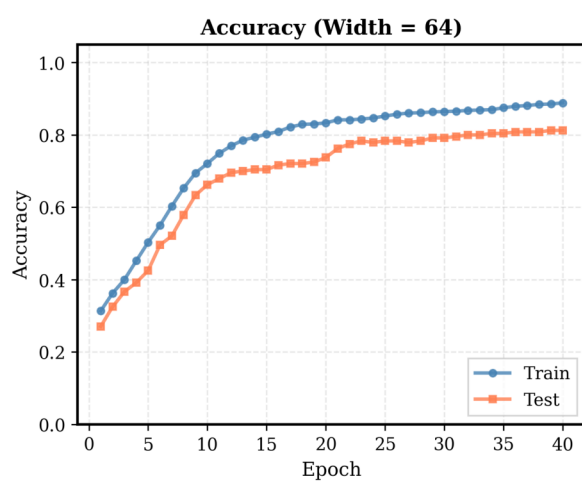
Experimental setup: synthetic classification dataset (sklearn.make_classification), 1200 examples, 20 features, 3 classes. StandardScaler applied. Each MLP trained for 40 epochs with Adam optimizer (lr=0.001). Metrics: train/test accuracy and training loss.
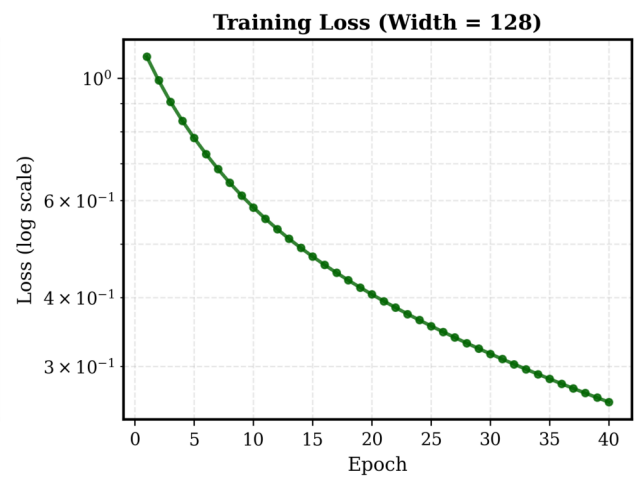
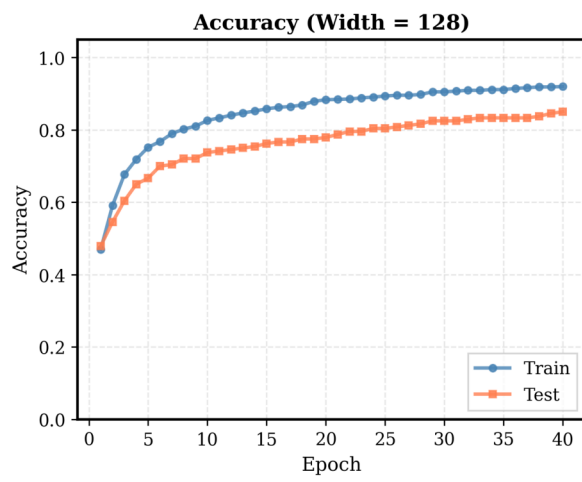**Accuracy (Width = 4)**   **Training Loss (Width = 4)**

Accuracy (Width = 8) — **Training Loss (Width = 8)**

Accuracy (Width = 16) — **Training Loss (Width = 16)**

Accuracy (Width = 32) — **Training Loss (Width = 32)**

Accuracy (Width = 64)

Training Loss (Width = 64)

Accuracy (Width = 128)

Training Loss (Width = 128)

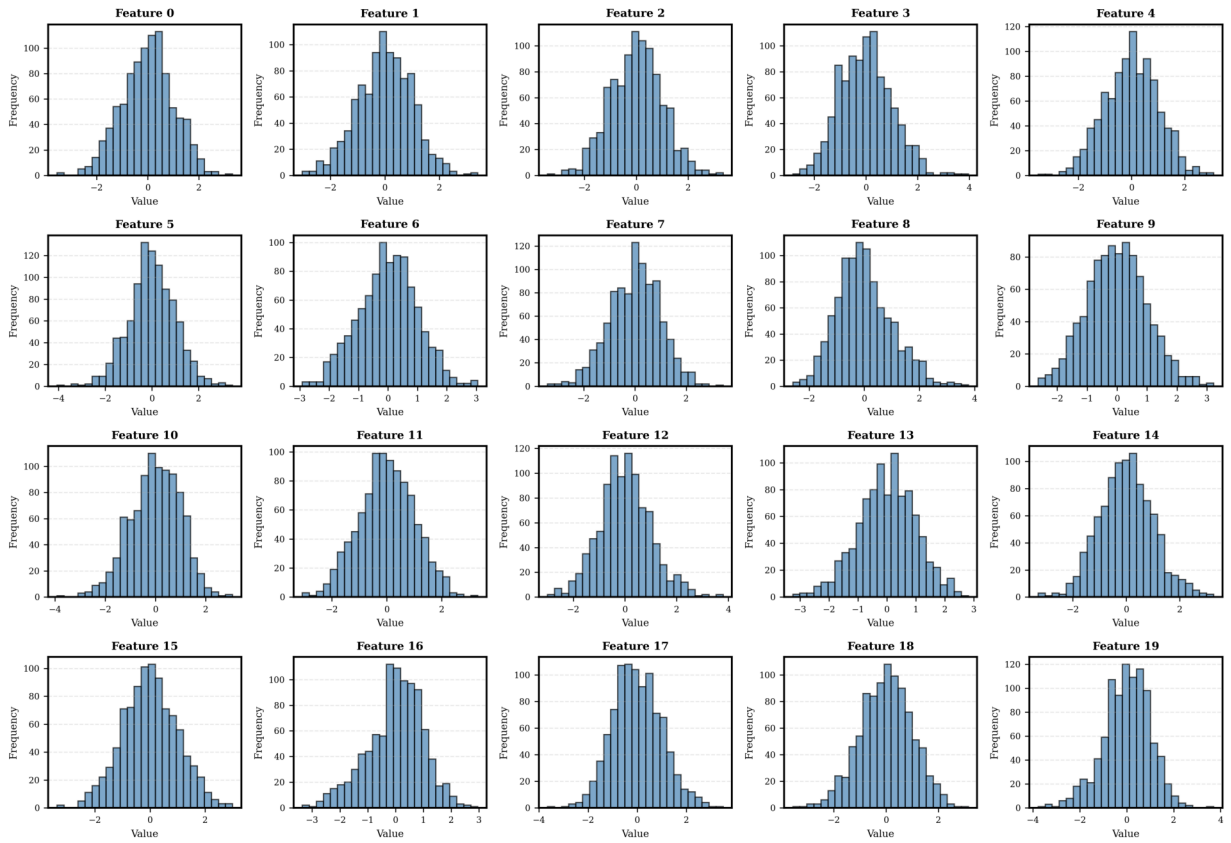**Training Set Class Distribution**
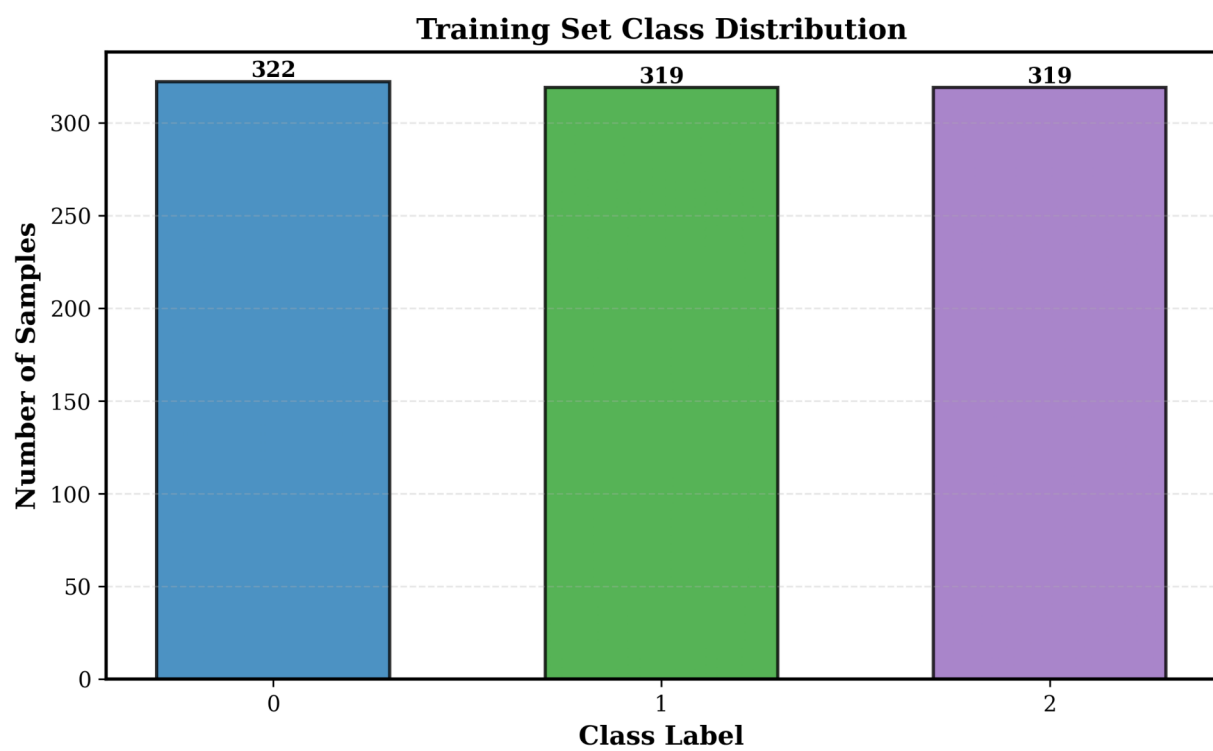
Feature Distributions (Training Set)

Feature Box Plots (Training Set)
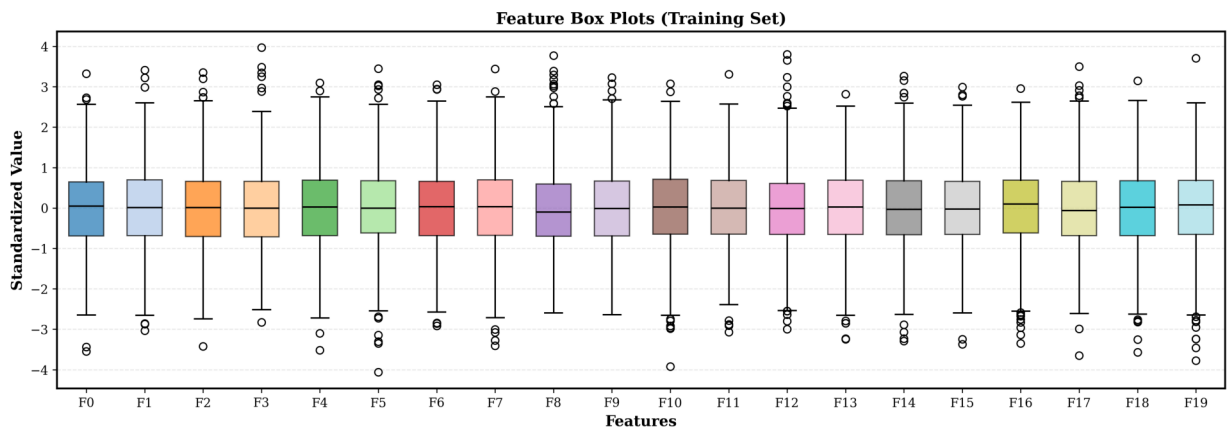
Feature Correlation Heatmap

Train-Test Data Distribution (PCA Projection)

Final Test Accuracy by Hidden Layer Width
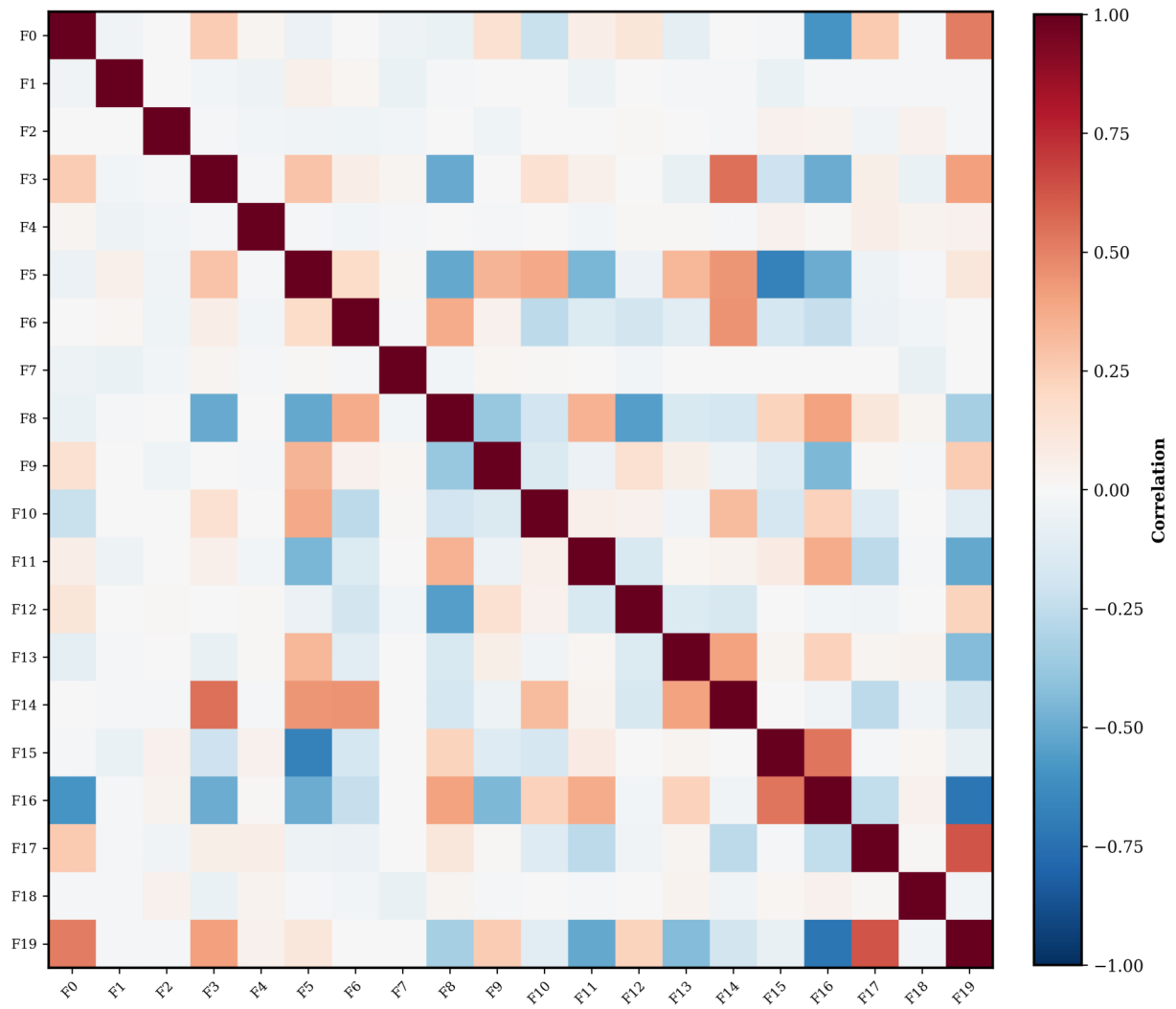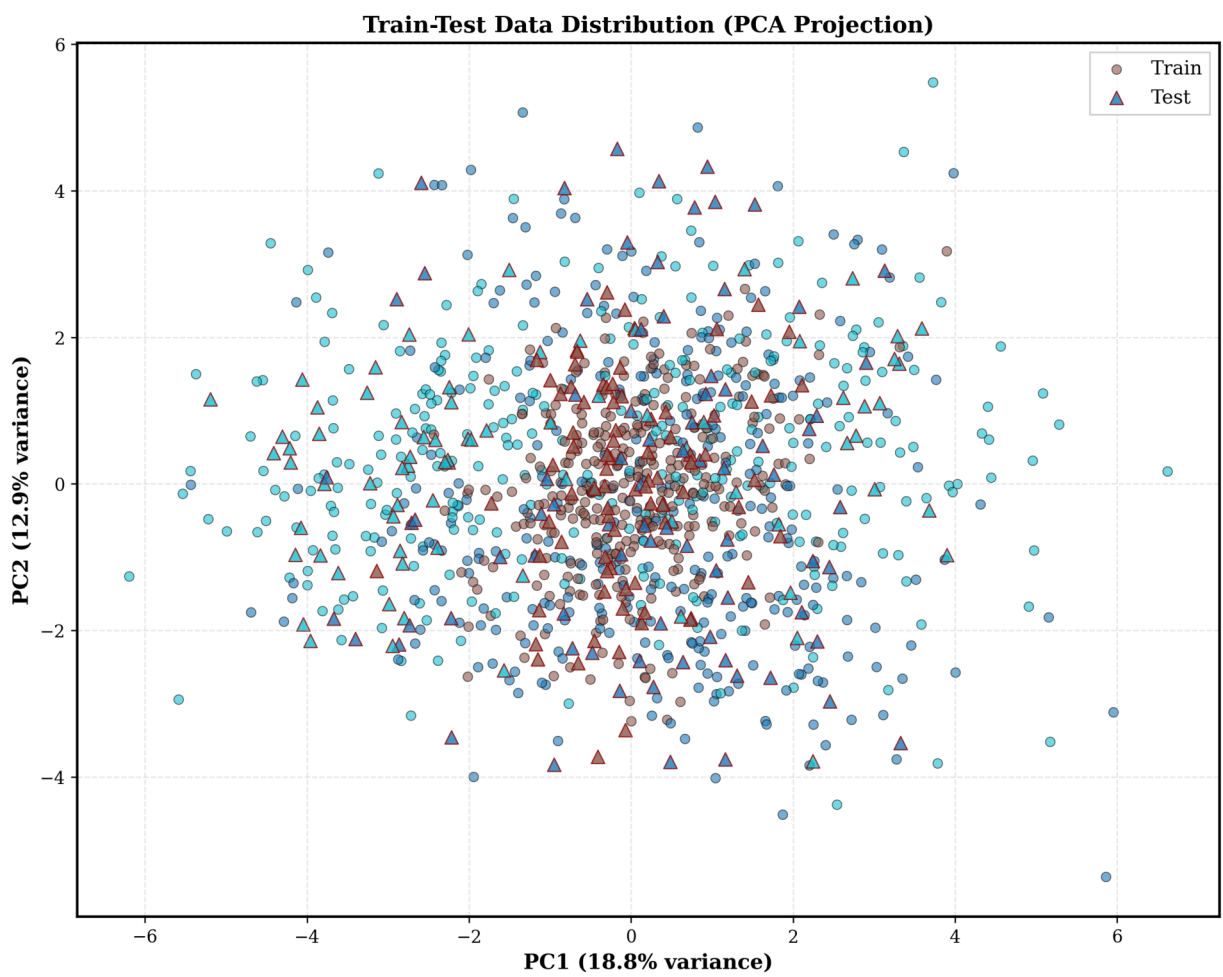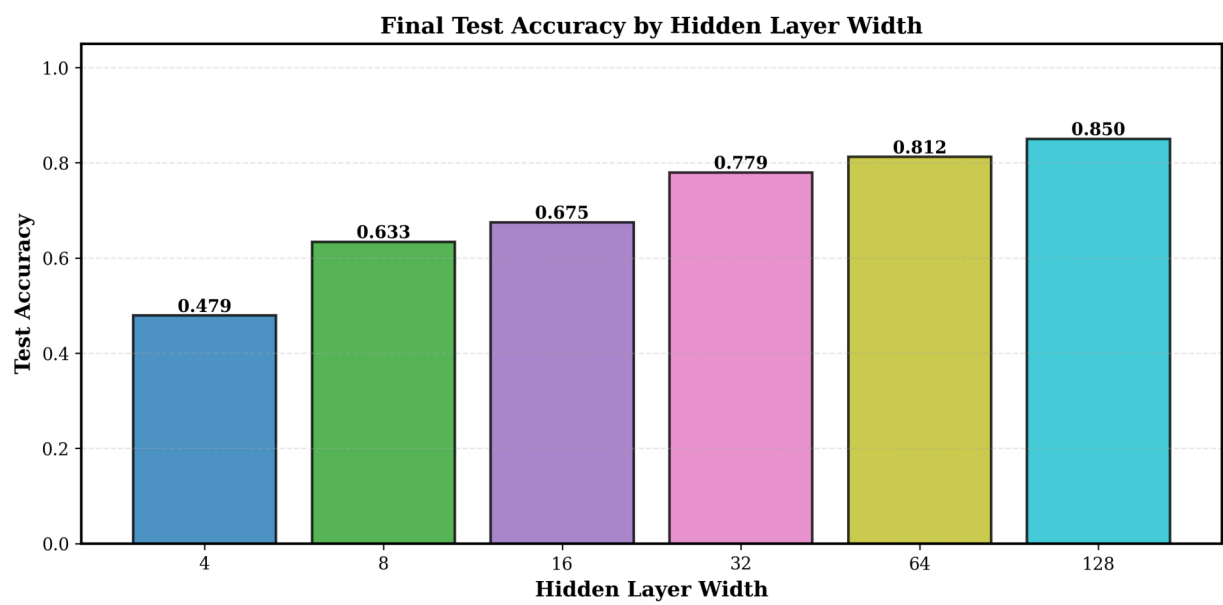
Dataset Summary

Metric	Value
Total Samples	1200
Training Samples	960
Test Samples	240
Number of Features	20
Number of Classes	3
Test Size Ratio	0.2 (20%)
Preprocessing	StandardScaler
Class Distribution	Stratified
Random Seed	42