

# Task 2 Report: MLP Backpropagation Neural Network for Bird Classification

Team ID: CHP\_5

---

Haneen Akram Ahmed

Reem Ahmed Ismail

Zeina Shawkat

Menna Thabet

Mohamed Ashraf

Islam Hisham

## Introduction

In this project, we implemented a Multi-Layer Perceptron (MLP) neural network using the backpropagation algorithm to classify bird species into three predefined classes. The model was trained and tested using a standardized bird dataset.

We trained the model on the first 30 samples per class and tested it on the remaining 20.

A flexible interface was created, allowing users to customize parameters such as the number of hidden layers, neurons per hidden layer, learning rate ( $\eta$ ), number of training epochs, bias usage, and the activation function (Sigmoid or Tanh).

The goal was to evaluate the model's classification performance under different configurations of these parameters using both activation functions

# Sigmoid Activation Function Tests

## Test 1: Basic Configuration

Model Parameters		
	Parameter	Value
0	Hidden Layers	1
1	Neurons per Layer	[4]
2	Learning Rate	0.003
3	Epochs	500
4	Bias	No
5	Activation Function	sigmoid

Fig.1: Model Parameters for Test 1 Sigmoid Activation Function

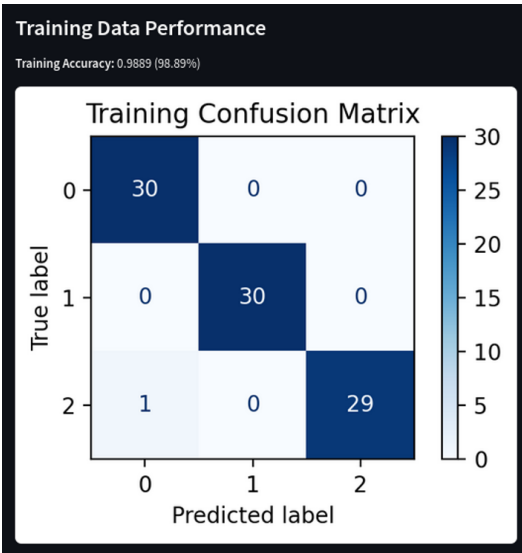


Fig.2: Training Confusion Matrix for Test 1 Sigmoid Activation Function

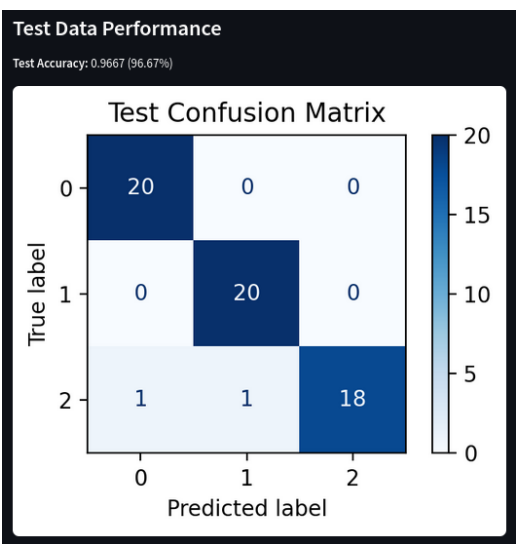


Fig.3: Test Confusion Matrix for Test 1 Sigmoid Activation Function

	precision	recall	f1-score	support
0	0.9667	0.9667	0.9667	30.0000
1	1.0000	1.0000	1.0000	30.0000
2	0.9667	0.9667	0.9667	30.0000
accuracy	0.9778	0.9778	0.9778	0.9778
macro avg	0.9778	0.9778	0.9778	90.0000
weighted avg	0.9778	0.9778	0.9778	90.0000

Fig.4: Training Classification Report for Test 1 Sigmoid Activation Function

	precision	recall	f1-score	support
0	0.9524	1.0000	0.9756	20.0000
1	1.0000	1.0000	1.0000	20.0000
2	1.0000	0.9500	0.9744	20.0000
accuracy	0.9833	0.9833	0.9833	0.9833
macro avg	0.9841	0.9833	0.9833	60.0000
weighted avg	0.9841	0.9833	0.9833	60.0000

Fig.5: Testing Classification Report for Test 1 Sigmoid Activation Function

Test 2: Deeper Network with Bias

Model Parameters		
	Parameter	Value
0	Hidden Layers	2
1	Neurons per Layer	[4, 4]
2	Learning Rate	0.005
3	Epochs	700
4	Bias	Yes
5	Activation Function	sigmoid

Fig.6: Model Parameters for Test 2 Sigmoid Activation Function

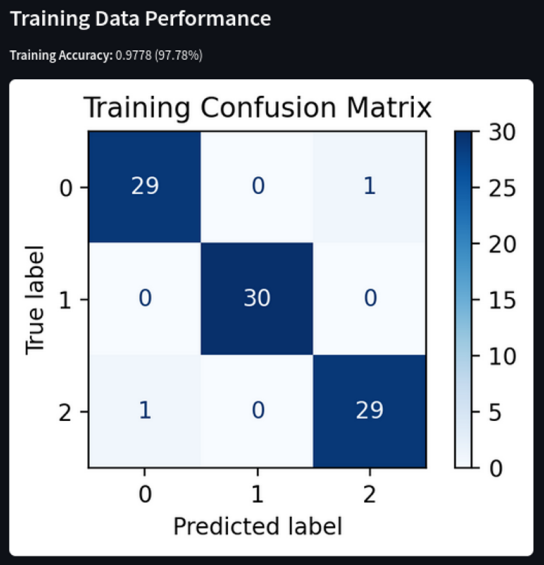


Fig.7: Training Confusion Matrix for Test 2 Sigmoid Activation Function

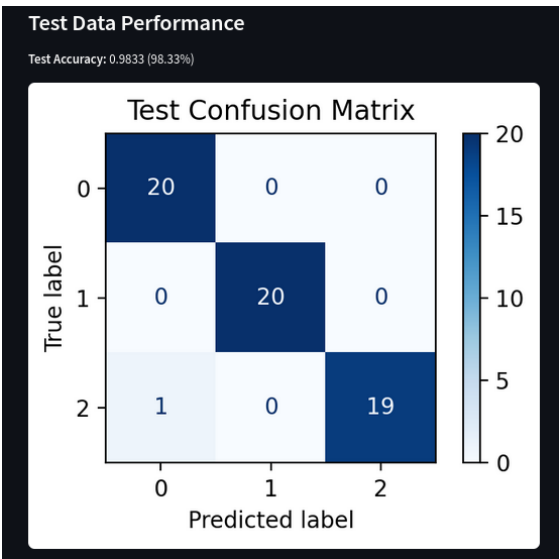


Fig.8: Test Confusion Matrix for Test 2 Sigmoid Activation Function

	precision	recall	f1-score	support
0	0.9677	1.0000	0.9836	30.0000
1	1.0000	1.0000	1.0000	30.0000
2	1.0000	0.9667	0.9831	30.0000
accuracy	0.9889	0.9889	0.9889	0.9889
macro avg	0.9892	0.9889	0.9889	90.0000
weighted avg	0.9892	0.9889	0.9889	90.0000

Fig.9: Training Classification Report for Test 2 Sigmoid Activation Function

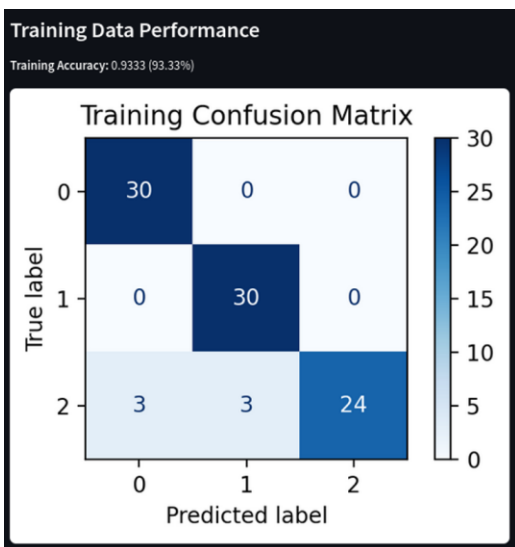
	precision	recall	f1-score	support
0	0.9524	1.0000	0.9756	20.0000
1	0.9524	1.0000	0.9756	20.0000
2	1.0000	0.9000	0.9474	20.0000
accuracy	0.9667	0.9667	0.9667	0.9667
macro avg	0.9683	0.9667	0.9662	60.0000
weighted avg	0.9683	0.9667	0.9662	60.0000

Fig.10: Testing Classification Report for Test 2 Sigmoid Activation Function

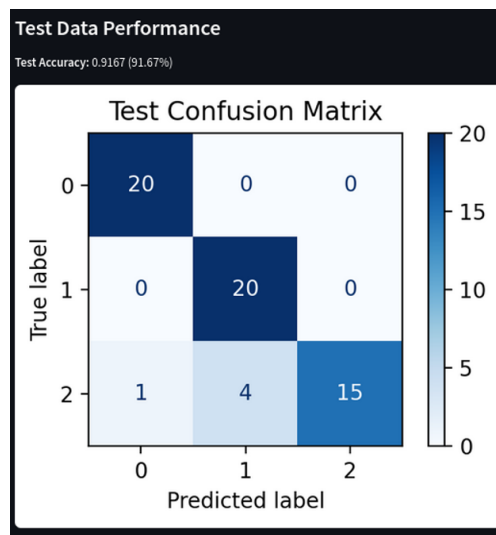
## Test 3: Deeper Network without Bias

Model Parameters		
	Parameter	Value
0	Hidden Layers	2
1	Neurons per Layer	[4, 4]
2	Learning Rate	0.005
3	Epochs	700
4	Bias	No
5	Activation Function	sigmoid

*Fig.11: Model Parameters for Test 3 Sigmoid Activation Function*



*Fig.12: Training Confusion Matrix for Test 3 Sigmoid Activation Function*



*Fig.13: Test Confusion Matrix for Test 3 Sigmoid Activation Function*

	precision	recall	f1-score	support
0	0.9091	1.0000	0.9524	30.0000
1	0.9091	1.0000	0.9524	30.0000
2	1.0000	0.8000	0.8889	30.0000
accuracy	0.9333	0.9333	0.9333	0.9333
macro avg	0.9394	0.9333	0.9312	90.0000
weighted avg	0.9394	0.9333	0.9312	90.0000

*Fig.14: Training Classification Report for Test 3 Sigmoid Activation Function*

	precision	recall	f1-score	support
0	0.9524	1.0000	0.9756	20.0000
1	0.8333	1.0000	0.9091	20.0000
2	1.0000	0.7500	0.8571	20.0000
accuracy	0.9167	0.9167	0.9167	0.9167
macro avg	0.9286	0.9167	0.9139	60.0000
weighted avg	0.9286	0.9167	0.9139	60.0000

*Fig.15: Testing Classification Report for Test 3 Sigmoid Activation Function*

Test 4: Deeper Network with Model Underfitting (no Bias)

Model Parameters		
	Parameter	Value
0	Hidden Layers	3
1	Neurons per Layer	[8, 6, 7]
2	Learning Rate	0.003
3	Epochs	5000
4	Bias	No
5	Activation Function	sigmoid

Fig.16: Model Parameters for Test 4 Sigmoid Activation Function

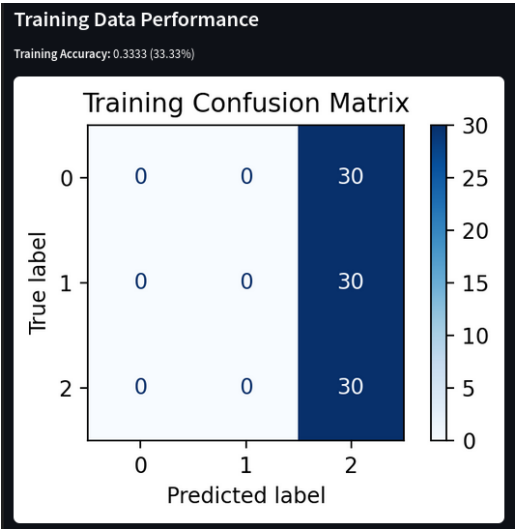


Fig.17: Training Confusion Matrix for Test 4 Sigmoid Activation Function

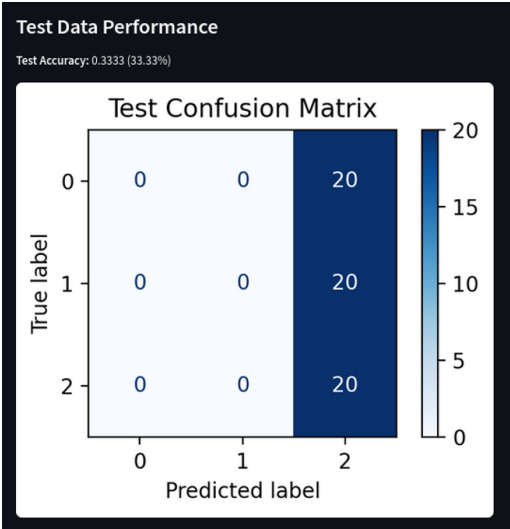


Fig.18: Test Confusion Matrix for Test 4 Sigmoid Activation Function

	precision	recall	f1-score	support
0	0.0000	0.0000	0.0000	30.0000
1	0.0000	0.0000	0.0000	30.0000
2	0.3333	1.0000	0.5000	30.0000
accuracy	0.3333	0.3333	0.3333	0.3333
macro avg	0.1111	0.3333	0.1667	90.0000
weighted avg	0.1111	0.3333	0.1667	90.0000

Fig.19: Training Classification Report for Test 4 Sigmoid Activation Function

	precision	recall	f1-score	support
0	0.0000	0.0000	0.0000	20.0000
1	0.0000	0.0000	0.0000	20.0000
2	0.3333	1.0000	0.5000	20.0000
accuracy	0.3333	0.3333	0.3333	0.3333
macro avg	0.1111	0.3333	0.1667	60.0000
weighted avg	0.1111	0.3333	0.1667	60.0000

Fig.20: Testing Classification Report for Test 4 Sigmoid Activation Function

### Insights & Observations (Sigmoid)

- Using a single hidden layer with 4 neurons (Test 1) achieved very high accuracy **(96.67%)** even without using bias.
- Adding more layers with bias (Test 2) slightly improved test accuracy to **98.3%**, showing that deeper networks with proper regularization are effective.
- When the same deeper network was trained without bias (Test 3), accuracy dropped to **91.67%**, indicating bias helps model learning.
- Test 4, a deep network without bias and many neurons, significantly underfit the data with only accuracy **33.33%**, confirming that architecture size alone doesn't guarantee performance.
- Finally, the best Sigmoid model used 2 layers with [4, 4] neurons, learning rate 0.005, 700 epochs, and included bias.

# Tanh Activation Function Tests

## Test 1: Basic Configuration

Model Parameters		
	Parameter	Value
0	Hidden Layers	1
1	Neurons per Layer	[4]
2	Learning Rate	0.005
3	Epochs	1000
4	Bias	Yes
5	Activation Function	tanh

Fig.21: Model Parameters for Test 1 Tanh Activation Function

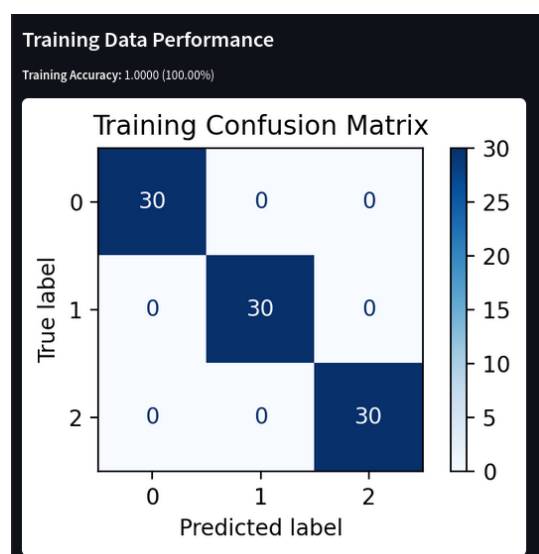


Fig.22: Training Confusion Matrix for Test 1 Tanh Activation Function

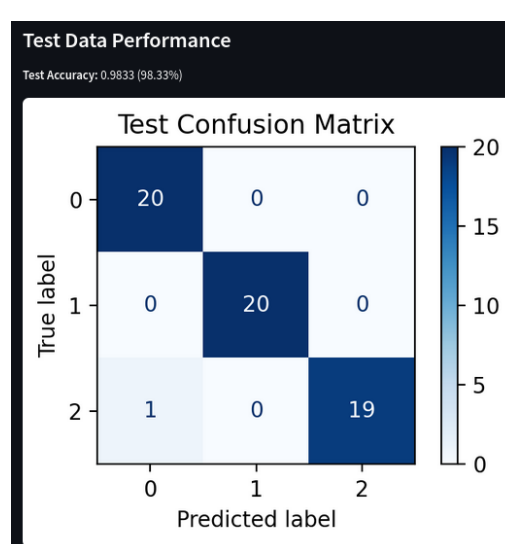


Fig.23: Test Confusion Matrix for Test 1 Tanh Activation Function

	precision	recall	f1-score	support
0	1.0000	1.0000	1.0000	30.0000
1	1.0000	1.0000	1.0000	30.0000
2	1.0000	1.0000	1.0000	30.0000
accuracy	1.0000	1.0000	1.0000	1.0000
macro avg	1.0000	1.0000	1.0000	90.0000
weighted avg	1.0000	1.0000	1.0000	90.0000

Fig.24: Training Classification Report for Test 1 Tanh Activation Function

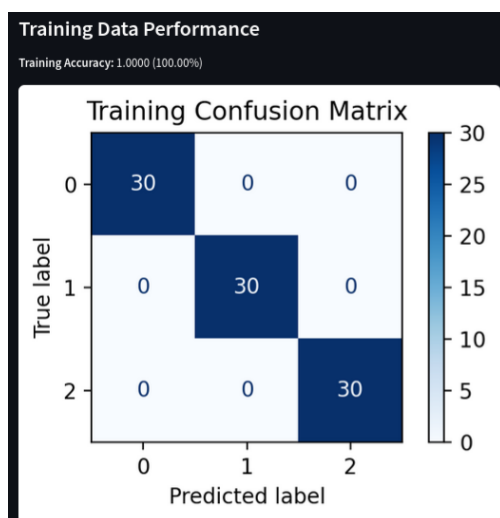
	precision	recall	f1-score	support
0	0.9524	1.0000	0.9756	20.0000
1	1.0000	1.0000	1.0000	20.0000
2	1.0000	0.9500	0.9744	20.0000
accuracy	0.9833	0.9833	0.9833	0.9833
macro avg	0.9841	0.9833	0.9833	60.0000
weighted avg	0.9841	0.9833	0.9833	60.0000

Fig.25: Testing Classification Report for Test 1 Tanh Activation Function

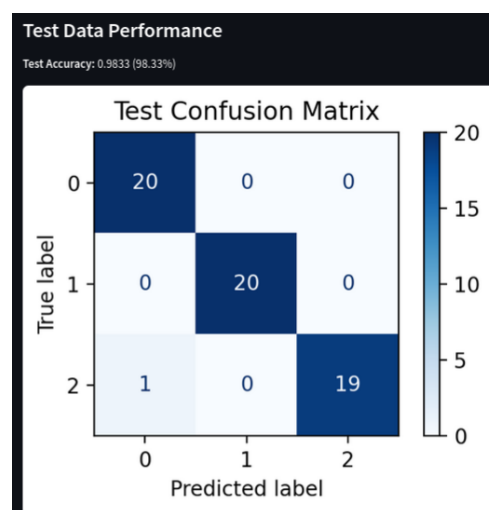
## Test 2: Deeper Network with Bias

Model Parameters		
	Parameter	Value
0	Hidden Layers	2
1	Neurons per Layer	[4, 3]
2	Learning Rate	0.005
3	Epochs	1000
4	Bias	Yes
5	Activation Function	tanh

**Fig.26: Model Parameters for Test 2 Tanh Activation Function**



**Fig.27: Training Confusion Matrix for Test 2 Tanh Activation Function**



**Fig.28: Test Confusion Matrix for Test 2 Tanh Activation Function**

	precision	recall	f1-score	support
0	1.0000	1.0000	1.0000	30.0000
1	1.0000	1.0000	1.0000	30.0000
2	1.0000	1.0000	1.0000	30.0000
accuracy	1.0000	1.0000	1.0000	1.0000
macro avg	1.0000	1.0000	1.0000	90.0000
weighted avg	1.0000	1.0000	1.0000	90.0000

**Fig.29: Training Classification Report for Test 2 Tanh Activation Function**

	precision	recall	f1-score	support
0	0.9524	1.0000	0.9756	20.0000
1	1.0000	1.0000	1.0000	20.0000
2	1.0000	0.9500	0.9744	20.0000
accuracy	0.9833	0.9833	0.9833	0.9833
macro avg	0.9841	0.9833	0.9833	60.0000
weighted avg	0.9841	0.9833	0.9833	60.0000

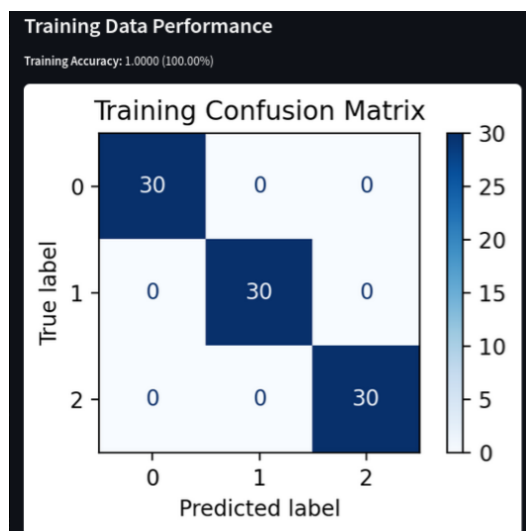
**Fig.30: Testing Classification Report for Test 2 Tanh Activation Function**



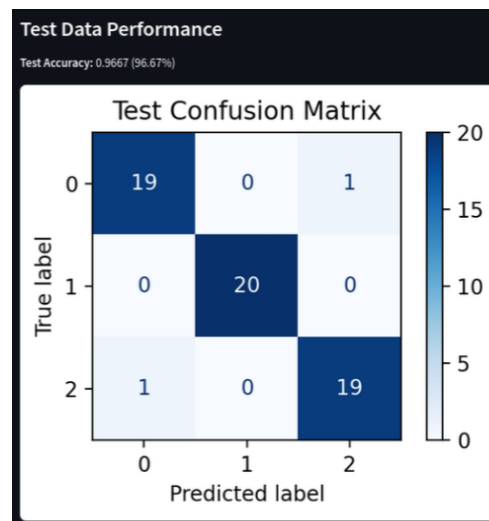
## Test 3: Deeper Network without Bias

Model Parameters		
	Parameter	Value
0	Hidden Layers	2
1	Neurons per Layer	[4, 4]
2	Learning Rate	0.008
3	Epochs	1500
4	Bias	No
5	Activation Function	tanh

*Fig.31: Model Parameters for Test 3 Tanh Activation Function*



*Fig.32: Training Confusion Matrix for Test 3 Tanh Activation Function*



*Fig.33: Test Confusion Matrix for Test 3 Tanh Activation Function*

	precision	recall	f1-score	support
0	1.0000	1.0000	1.0000	30.0000
1	1.0000	1.0000	1.0000	30.0000
2	1.0000	1.0000	1.0000	30.0000
accuracy	1.0000	1.0000	1.0000	1.0000
macro avg	1.0000	1.0000	1.0000	90.0000
weighted avg	1.0000	1.0000	1.0000	90.0000

*Fig.34: Training Classification Report for Test 3 Tanh Activation Function*

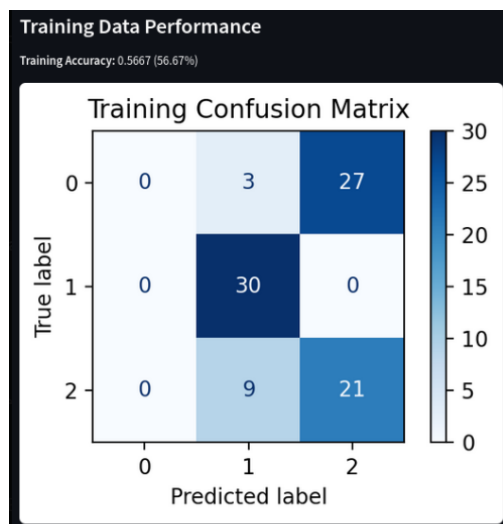
	precision	recall	f1-score	support
0	0.9500	0.9500	0.9500	20.0000
1	1.0000	1.0000	1.0000	20.0000
2	0.9500	0.9500	0.9500	20.0000
accuracy	0.9667	0.9667	0.9667	0.9667
macro avg	0.9667	0.9667	0.9667	60.0000
weighted avg	0.9667	0.9667	0.9667	60.0000

*Fig.35: Testing Classification Report for Test 3 Tanh Activation Function*

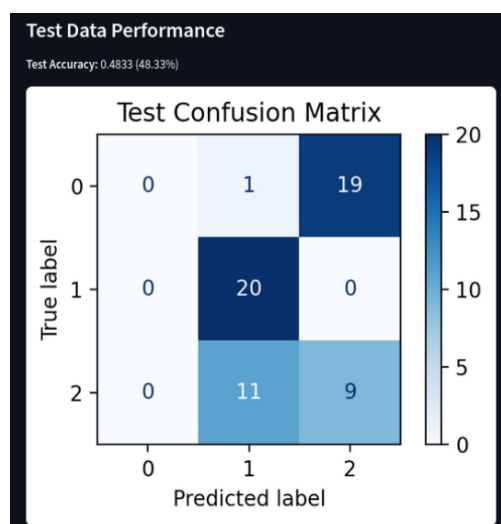
## Test 4: Deeper Network with Model Underfitting (with Bias)

Model Parameters		
	Parameter	Value
0	Hidden Layers	5
1	Neurons per Layer	[12, 10, 10, 3, 3]
2	Learning Rate	0.001
3	Epochs	5000
4	Bias	Yes
5	Activation Function	tanh

**Fig.36: Model Parameters for Test 4 Tanh Activation Function**



**Fig.37: Training Confusion Matrix for Test 4 Tanh Activation Function**



**Fig.38: Test Confusion Matrix for Test 4 Tanh Activation Function**

	precision	recall	f1-score	support
0	0.0000	0.0000	0.0000	30.0000
1	0.7143	1.0000	0.8333	30.0000
2	0.4375	0.7000	0.5385	30.0000
accuracy	0.5667	0.5667	0.5667	0.5667
macro avg	0.3839	0.5667	0.4573	90.0000
weighted avg	0.3839	0.5667	0.4573	90.0000

**Fig.39: Training Classification Report for Test 4 Tanh Activation Function**

	precision	recall	f1-score	support
0	0.0000	0.0000	0.0000	20.0000
1	0.6250	1.0000	0.7692	20.0000
2	0.3214	0.4500	0.3750	20.0000
accuracy	0.4833	0.4833	0.4833	0.4833
macro avg	0.3155	0.4833	0.3814	60.0000
weighted avg	0.3155	0.4833	0.3814	60.0000

**Fig.40: Testing Classification Report for Test 4 Tanh Activation Function**

### Insights & Observations (Tanh)

- All Tanh models achieved very high training accuracy (**100%**) in Tests 1-3, confirming that Tanh is highly effective for fitting the training data.
- Test 1 and 2 both achieved **98.33%** test accuracy using different structures, indicating consistent generalization.
- Test 3 (no bias) still maintained **96.67%**, showing Tanh is slightly more robust than Sigmoid in deeper networks.
- Test 4 with a large architecture underfit the dataset, with only **48.33%** accuracy, reinforcing that overcomplexity may harm performance.
- The best Tanh model used 1 layer with 4 neurons, LR 0.005, 1000 epochs, and bias.

## Comparison Table

Activation Function	Train Accuracy (%)	Test Accuracy (%)	LR	Epochs	#Layers	Hidden Nodes
Sigmoid	97.78	98.33	0.005	700	2	[4, 4]
Tanh	100	98.33	0.005	1000	1	[4]

## Conclusion & Model Insights

- Overall, both activation functions performed well, with **Tanh** showing slightly better consistency and stability across different configurations.
- The highest test accuracy (**98.33%**) was reached with both activation functions, but **Tanh** achieved it with a simpler network (**1 hidden layer**).
- Bias was shown to significantly improve performance for both activation functions.
- Overly deep or wide networks without proper tuning led to underfitting, especially when bias was excluded.
- **Tanh** was the more reliable activation function in our experiments, achieving excellent performance even with minimal architecture.