CHESS PIECE CLASSIFICATION

Starting Model Hyper-Parameters:

- Number of Epochs: 15
- Batch Size: 32
- Topology: [7225, 32, 16, 12, 6]
- Criterion: Cross Entropy Loss
- Optimiser: AdamW
- Learning Rate: 0.001
- Activation Function: ReLU
- Accuracy: 78.91%

Topology (5 Tests per Model)		
Name	*Layer Structure [Input, Hidden, Output]*	*Accuracy*
Model1	[7225, 32, 16, 12, 6]	Average: 78.91% | Highest: 82.12%
Model2	[7225, 64, 32, 16, 6]	Average: 81.26% | Highest: 84.25%
Model3	[7225, 512, 128, 6]	Average: 81.51% | Highest: 85.81%
Model4	[7225, 512, 256, 128, 64, 32, 6]	Average: 80.13% | Highest: 86.50%
Model5	[7225, 32, 32, 32, 6]	Average: 78.11% | Highest: 84.81%
Model6	[7225, 64, 64, 64, 6]	Average: 78.63% | Highest: 81.38%
Model7	[7225, 128, 128, 128, 6]	Average: 77.16% | Highest: 80.50%

Model1, Model3 and Model4 show a high dispersion around 8% to 10% whereas Model2 had the dispersion only around ±1% except the one with highest result. Models 5 to 7 performed poorly. All Models can be found in the folder model.py so the primary code stays clear.

Optimalization (5 Tests per Model)		Learning Rate (5 Tests per Model)	
Type	*Accuracy*	*Value*	*Accuracy*
SGD	Average: 56.11% | Highest: 60.38%	0.01	Average: 59.85% | Highest: 63.38%
ASGD	Average: 54.70% | Highest: 56.44%	0.005	Average: 71.57% | Highest: 74.12%
Adam	Average: 79.71% | Highest: 82.56%	0.003	Average: 77.92% | Highest: 79.31%
AdamW	Average: 81.48% | Highest: 83.65%	0.001	Average: 81.20% | Highest: 83.25%
Adadelta	Average: 47.69% | Highest: 48.31%	0.0005	Average: 80.84% | Highest: 83.81%

Optimizer Adam performed a bit worse than AdamW, where AdamW is also more stable.
When we set Learning Rate (LR) to 0.05 the model learns only pawns because of their dominance in the dataset. We can see that LR around 0.001 seems to be the perfect middle.

Activation Functions (5 Tests per Model)	
Type	*Accuracy*
ReLU	Average: 84.95% | Highest: 88.44%
Tanh	Average: 63.87% | Highest: 66.12%
Sigmoid	Average: 56.99% | Highest: 60.44%

ReLU did performed the best for this topology, since Tanh & Sigmoid are usually used with fewer hidden layers and significantly less neurons per layer.

Final Model Hyper-Parameters:

- Number of Epochs: 15
- Batch Size: 32
- Topology: [7225, 512, 128, 6]
- Criterion: Cross Entropy Loss
- Optimiser: AdamW
- Learning Rate: 0.001
- Activation Function: ReLU
- Accuracy: 85.44%