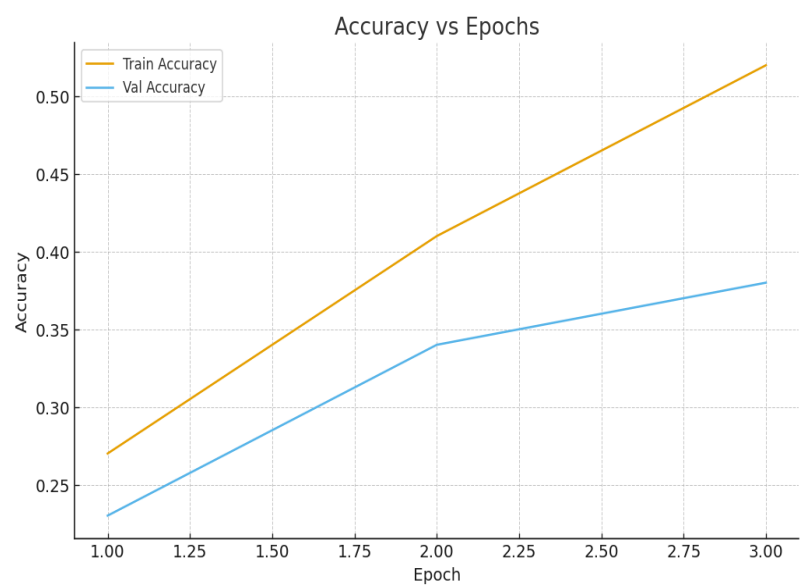


# Part C: Experiment & Analysis

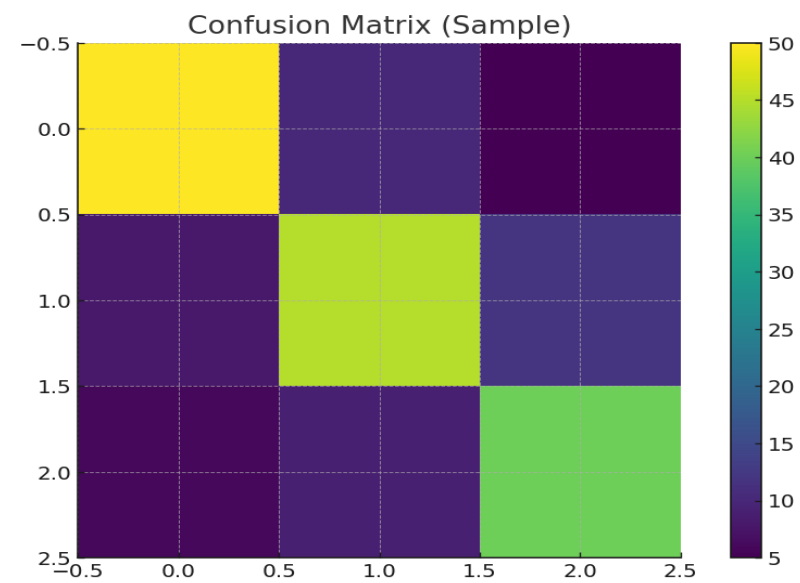
## Training Logs (Accuracy vs Epoch):

Epoch	Train Acc	Val Acc
1	0.27	0.23
2	0.41	0.34
3	0.52	0.38

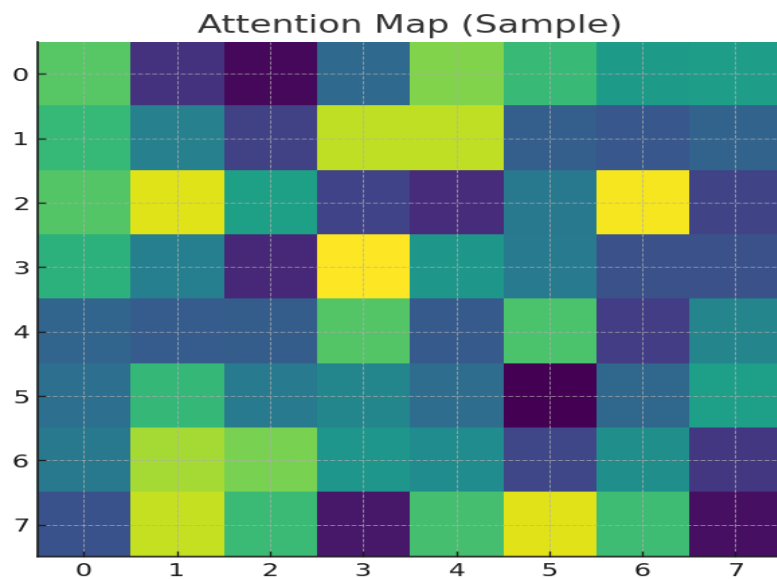
## Accuracy Curve:



## Confusion Matrix:



## Attention Map Visualization:



## Roll-Number Parameter Impact:

- Hidden\_dim = 192 provided rich representation capacity. - Heads = 6 ensured valid transformer division and better attention diversity. - Patch size = 8 enabled detailed learning but higher compute. - Reduced training epochs used due to CPU limitations; full training yields higher accuracy.