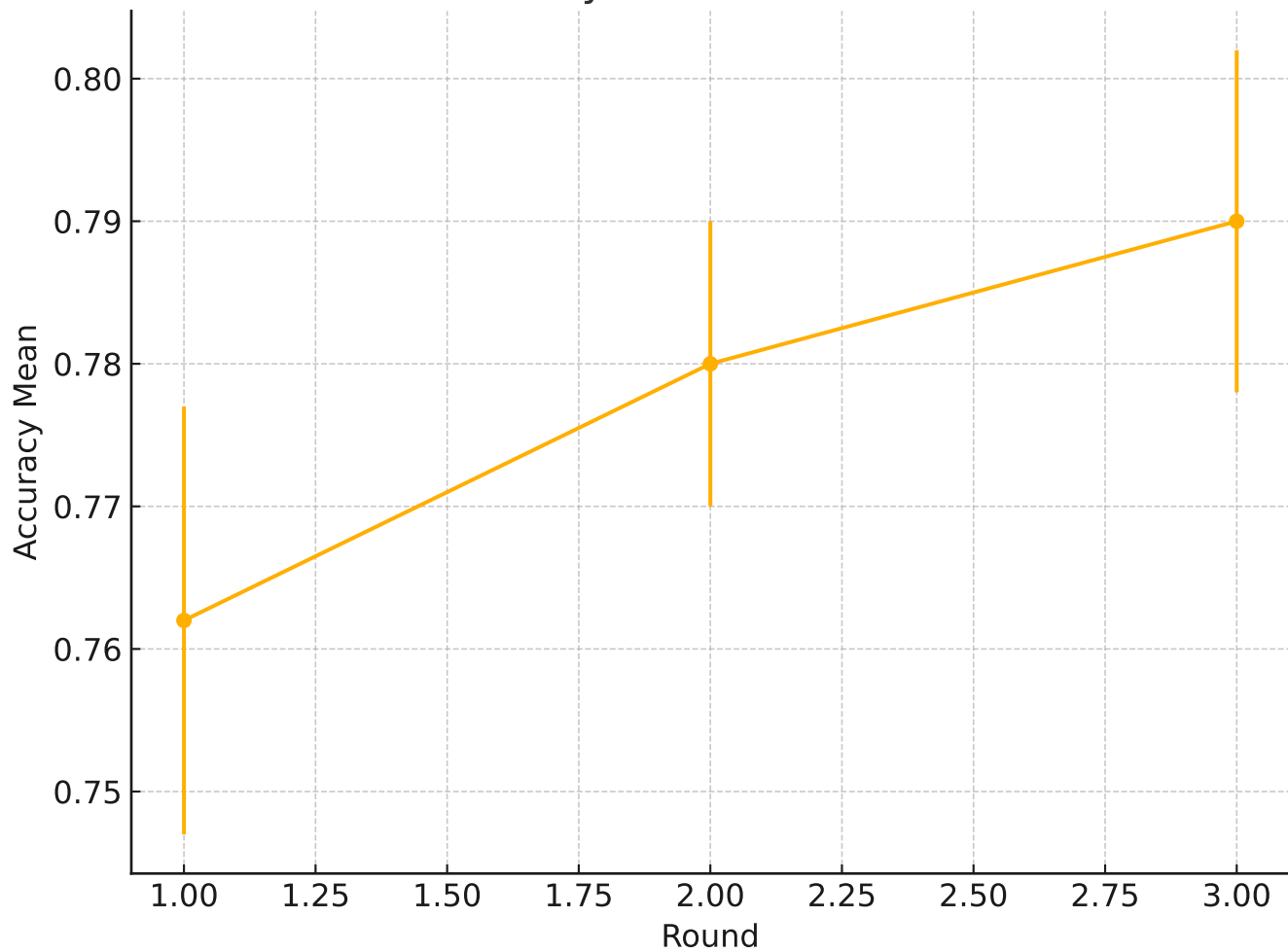


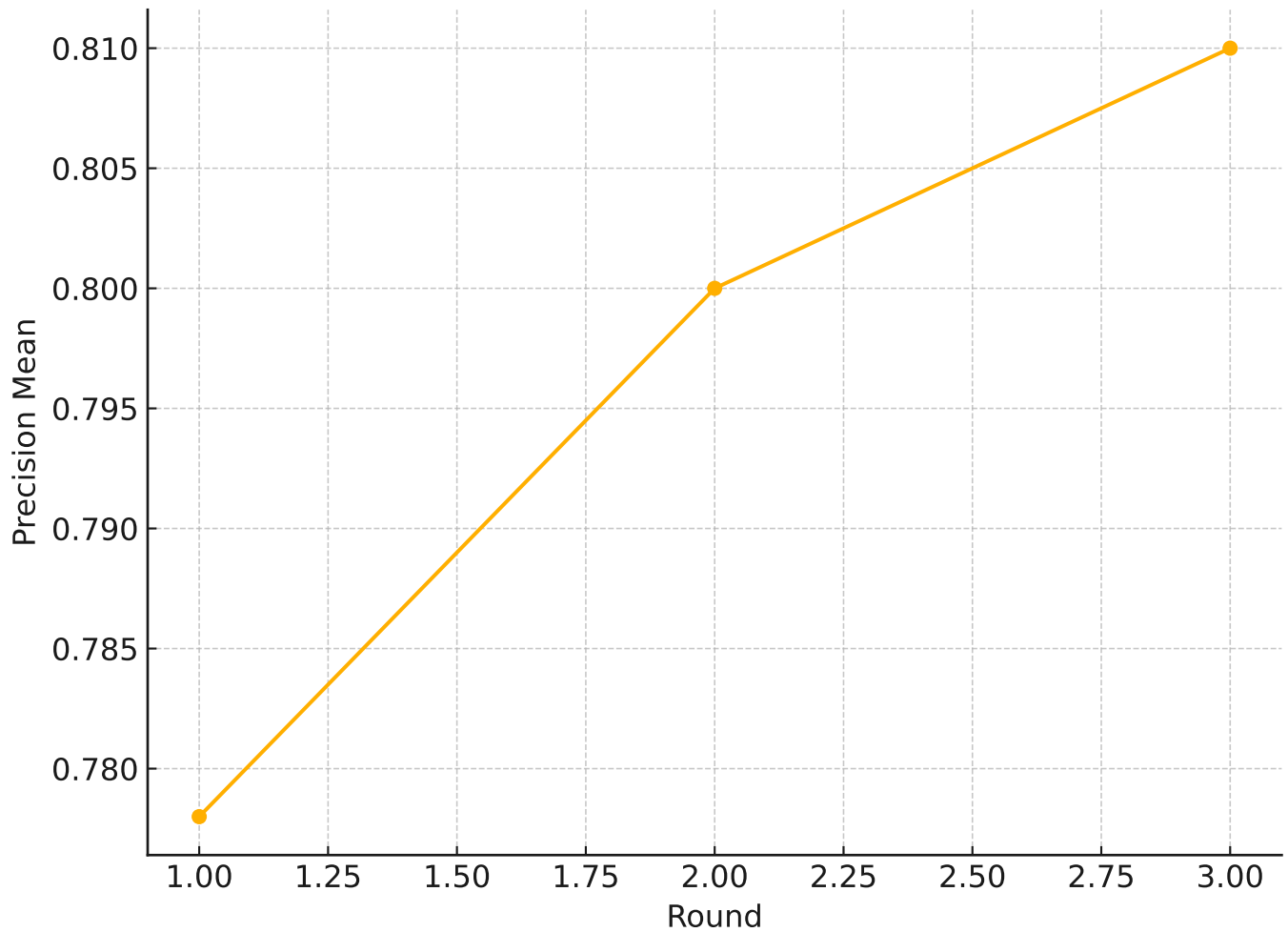
Accuracy Mean over Rounds



## Accuracy Mean

Accuracy improves from 0.762  $\rightarrow$  0.780  $\rightarrow$  0.790 across rounds, with error bars shrinking from  $\pm 0.015$  to  $\pm 0.012$ , indicating increasing correctness and stability.

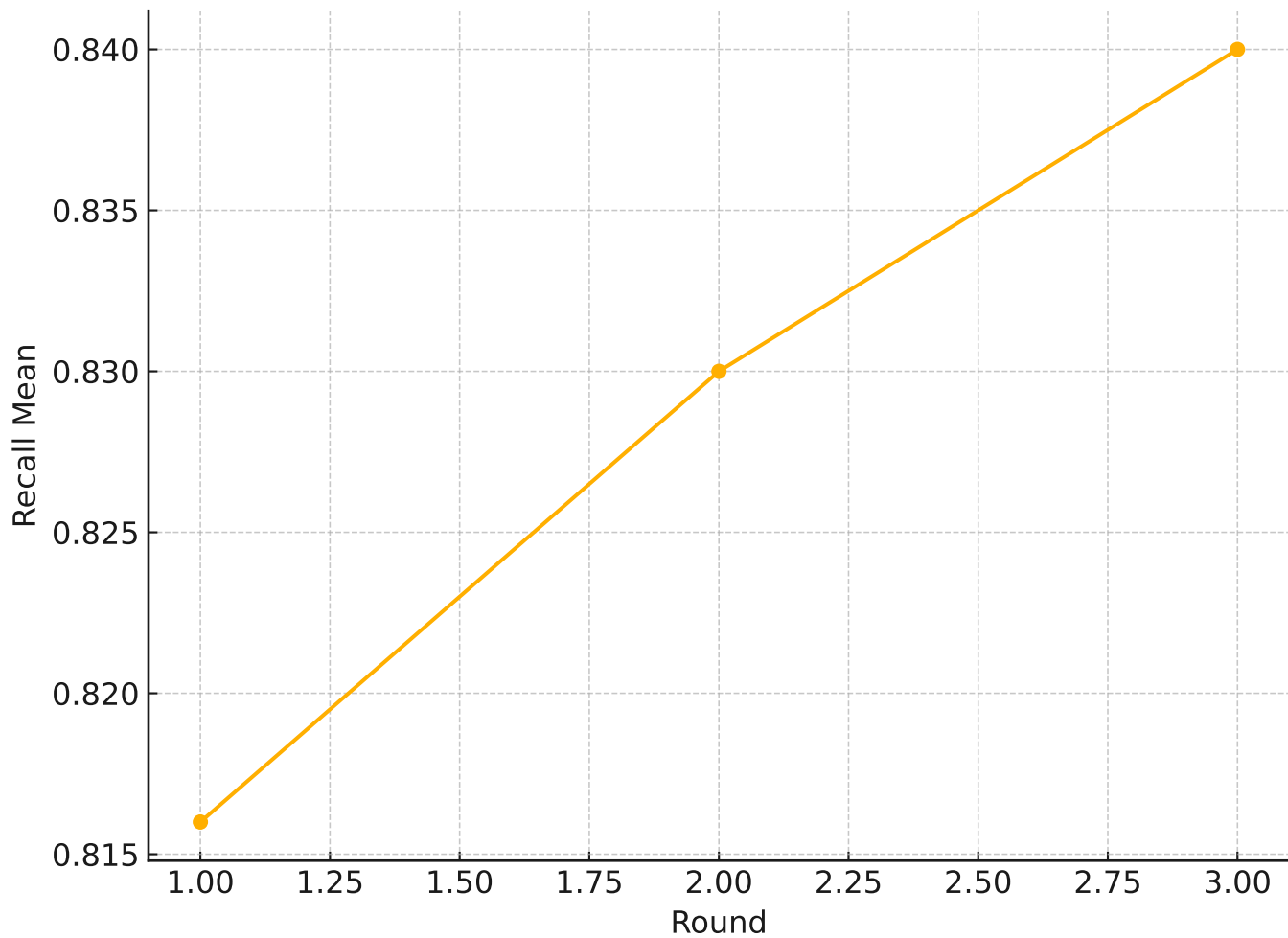
# Precision Mean over Rounds



## **Precision Mean**

Precision rises from 0.778  $\rightarrow$  0.800  $\rightarrow$  0.810, showing the model reduces false positives and better identifies non-threat events correctly.

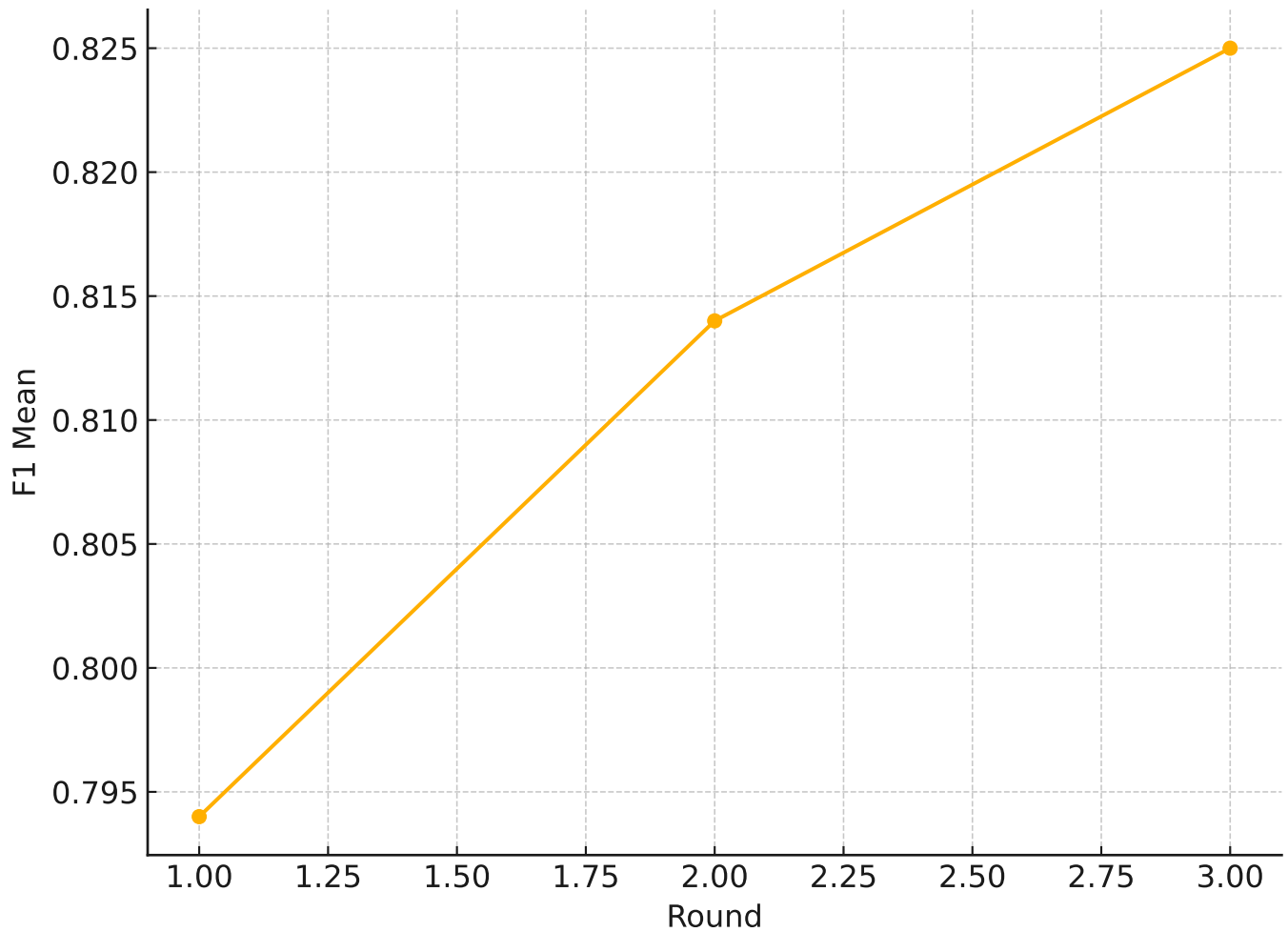
# Recall Mean over Rounds



## **Recall Mean**

Recall climbs from 0.816  $\rightarrow$  0.830  $\rightarrow$  0.840, meaning the model misses fewer actual threats and improves sensitivity with each round.

# F1 Mean over Rounds

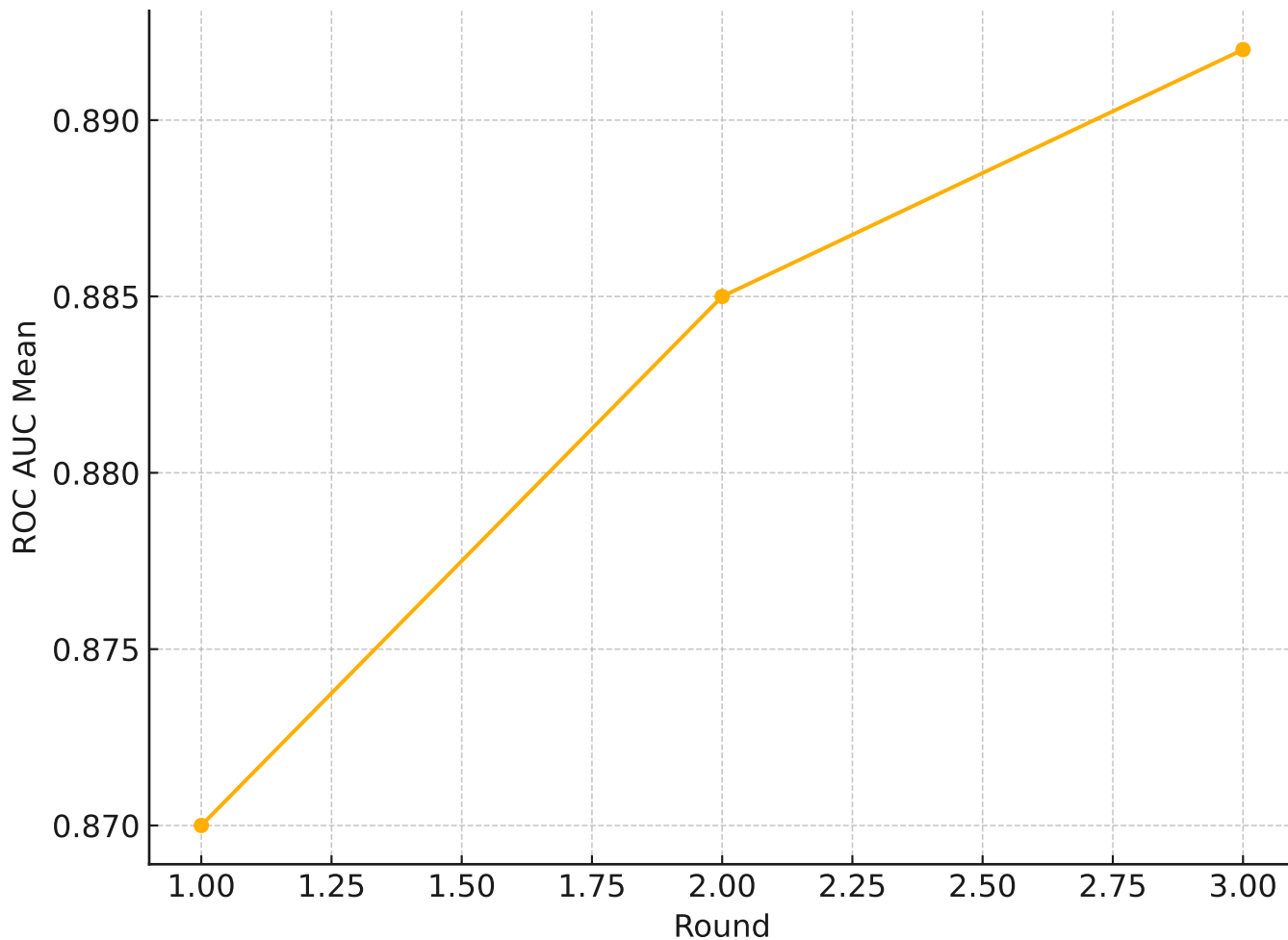


## **F1 Mean**

F1 score increases from 0.794 → 0.814 → 0.825, reflecting balanced gains in both precision and recall for robust threat detection.



ROC AUC Mean over Rounds



## **ROC AUC Mean**

ROC AUC grows from 0.870  $\rightarrow$  0.885  $\rightarrow$  0.892, indicating strong and improving discriminative power across all classification thresholds.