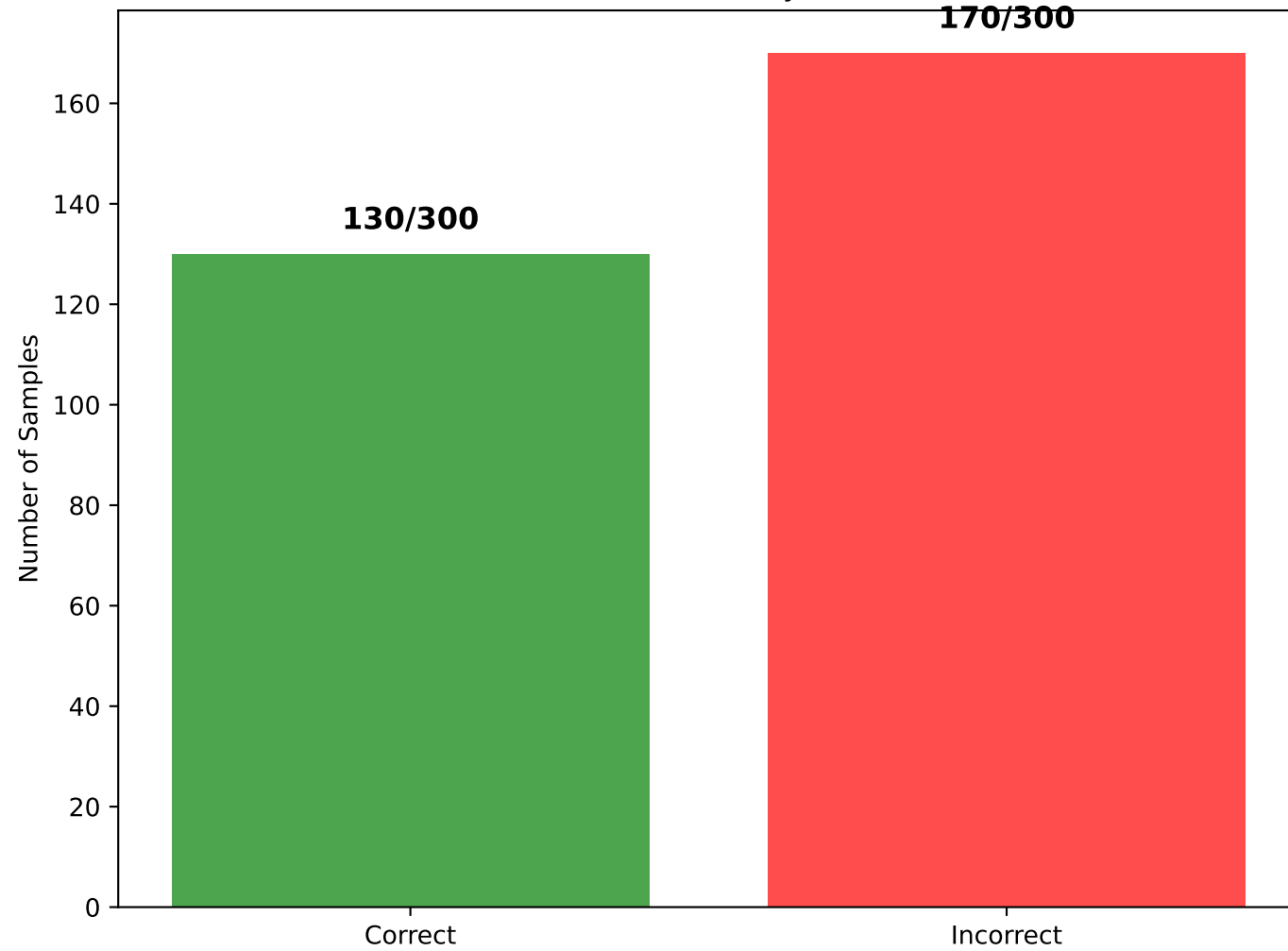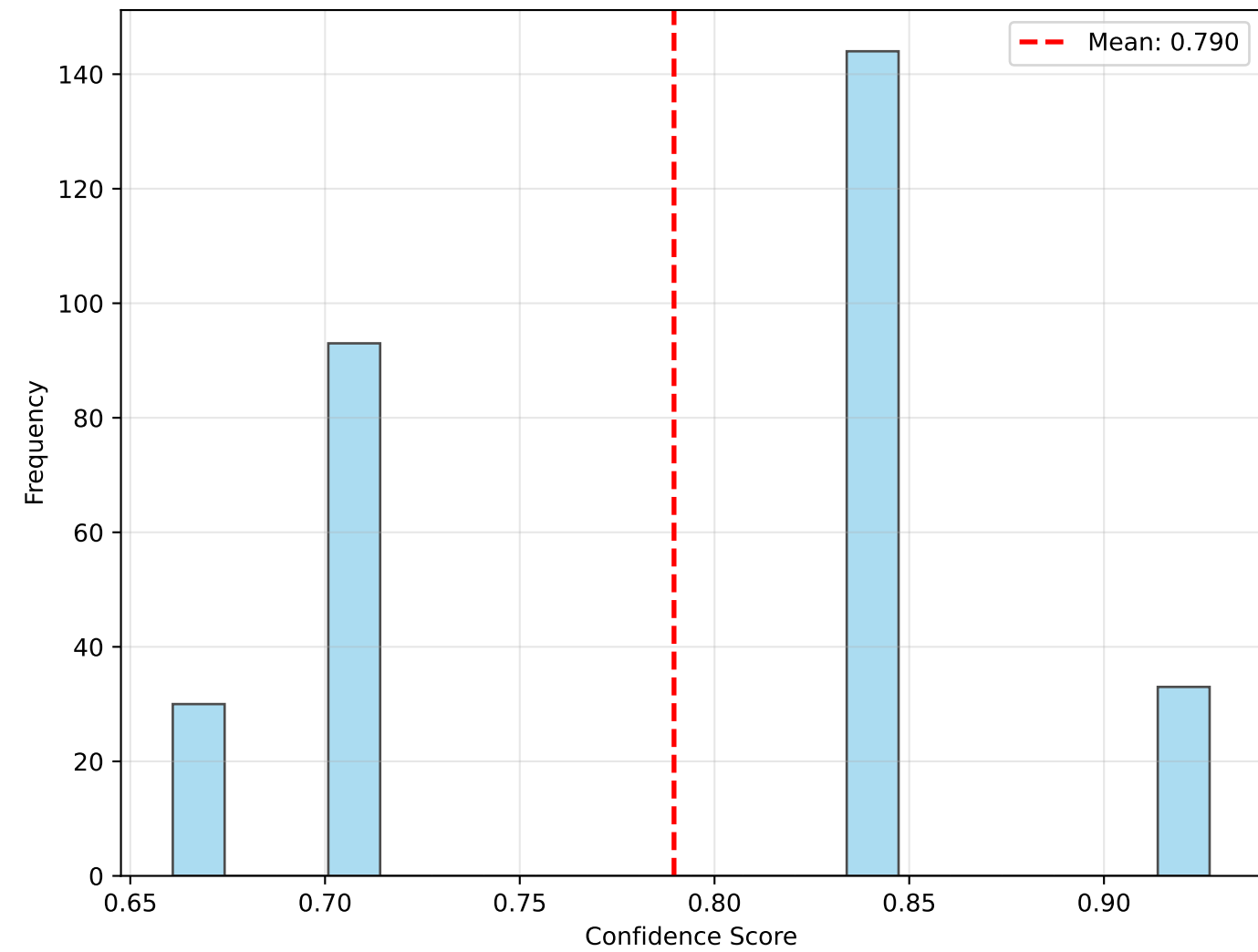# SO8T Dataset Evaluation Results
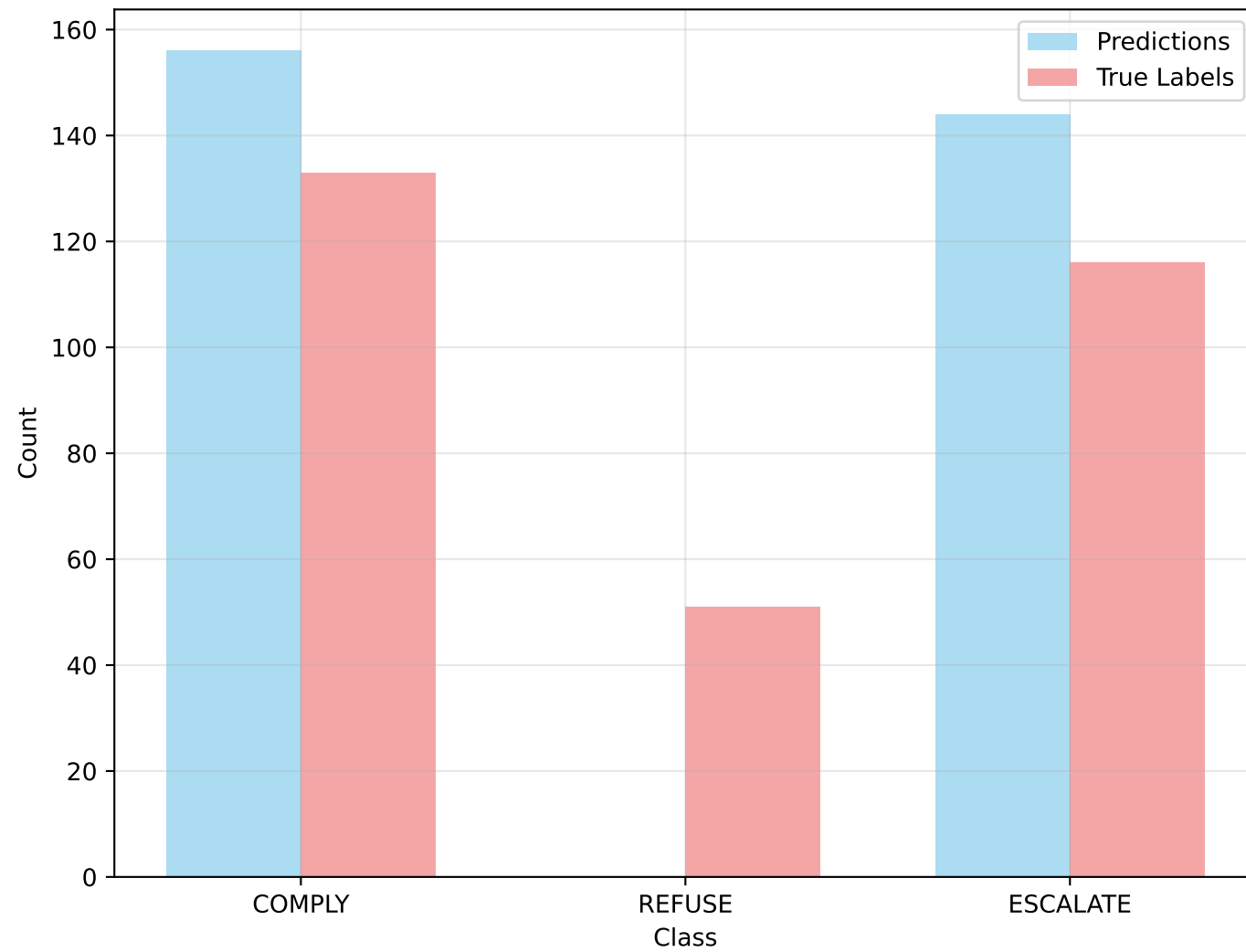
## Prediction Accuracy: 43.3%



## Confidence Distribution (Dataset)



Mean: 0.790

## Class Distribution: Predictions vs True Labels



Legend: Predictions, True Labels

## Confusion Matrix



| True Label \ Predicted Label | COMPLY | REFUSE | ESCALATE |
|---|---|---|---|
| COMPLY | 74 | 0 | 59 |
| REFUSE | 22 | 0 | 29 |
| ESCALATE | 60 | 0 | 56 |