Class imbalance

Training vs Test Accuracy
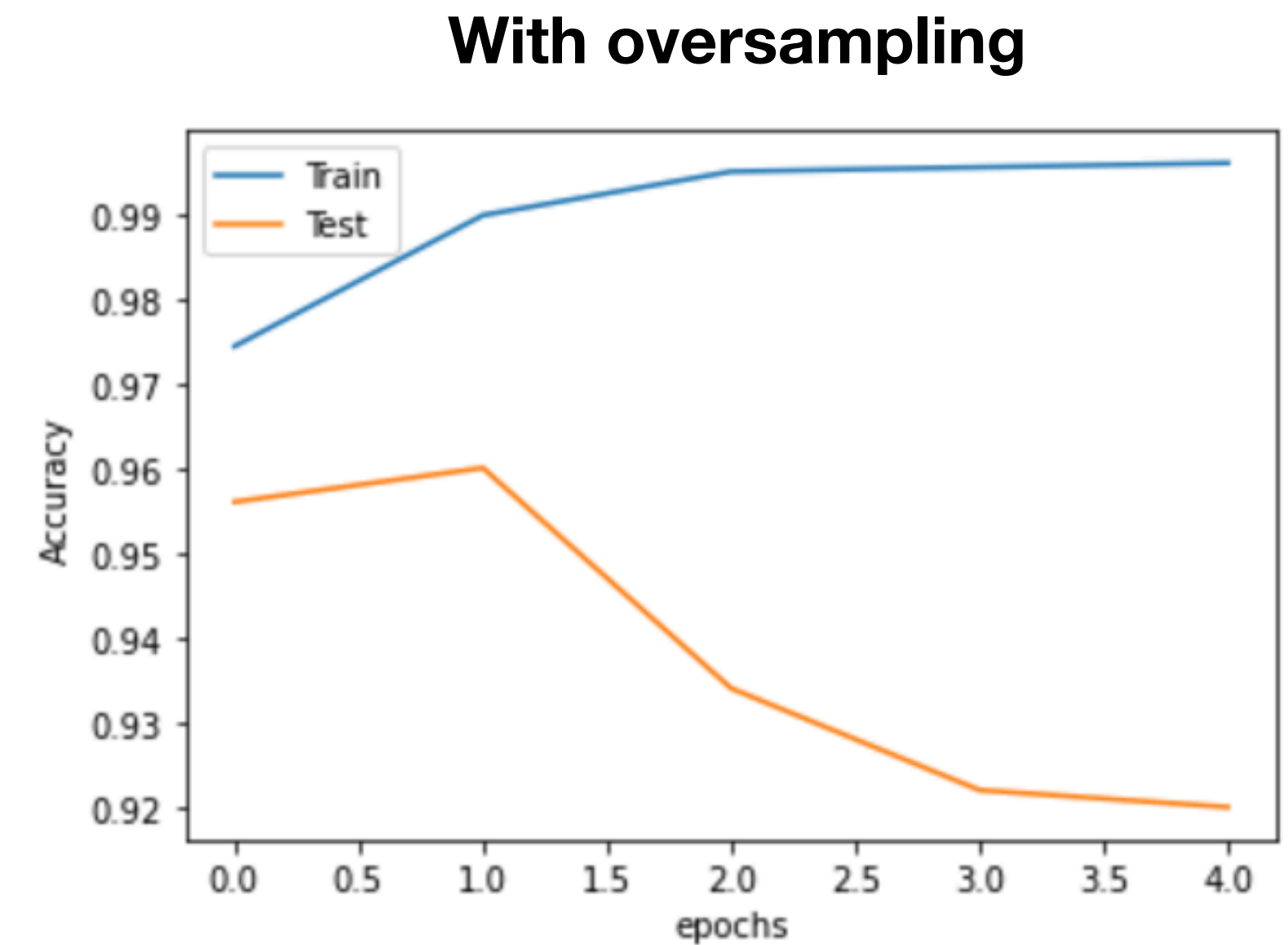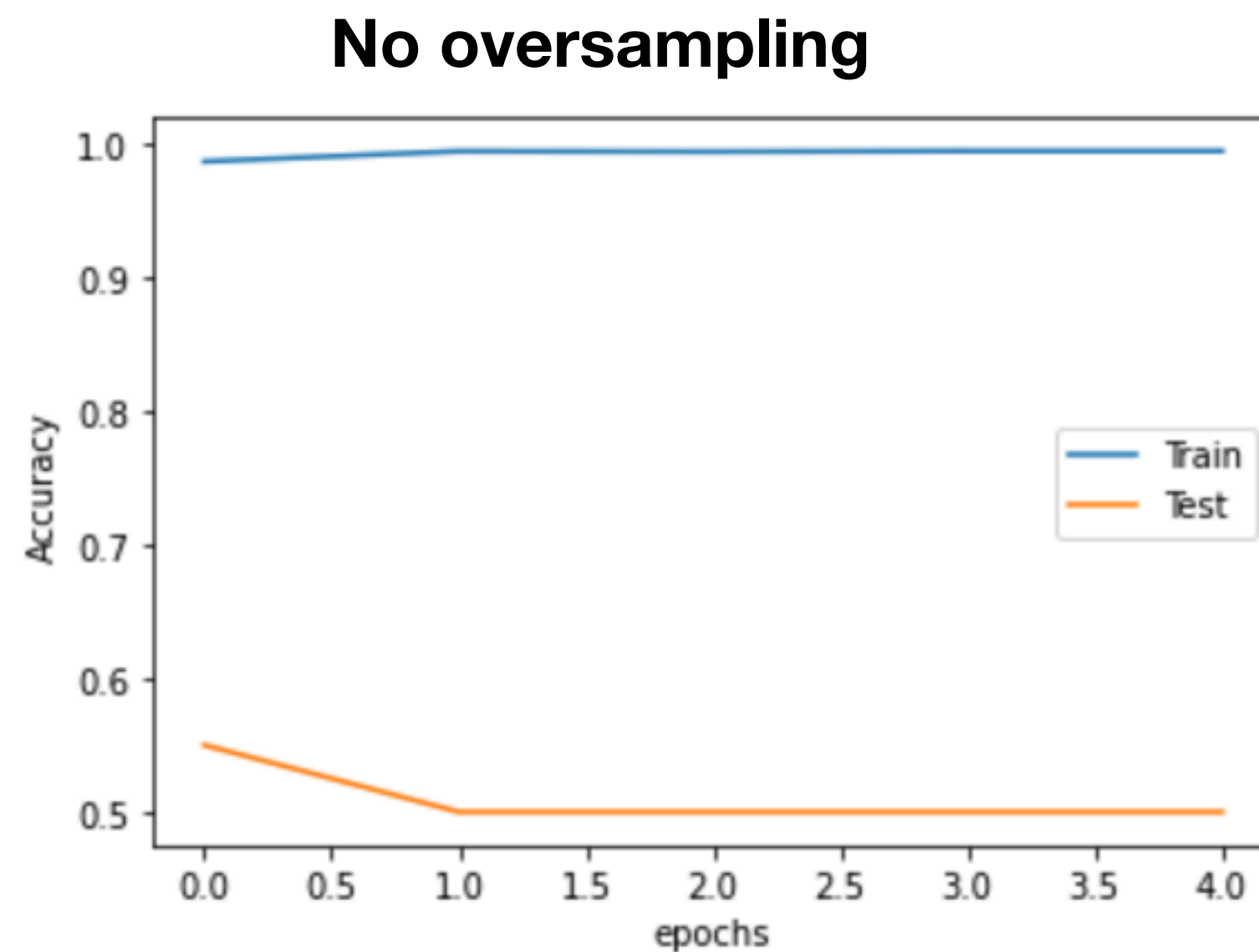

No oversampling

With oversampling
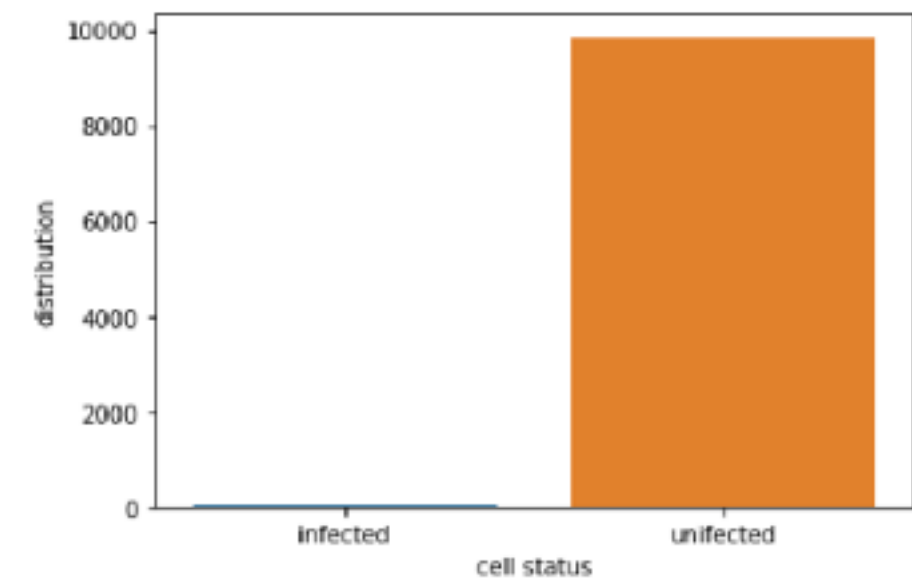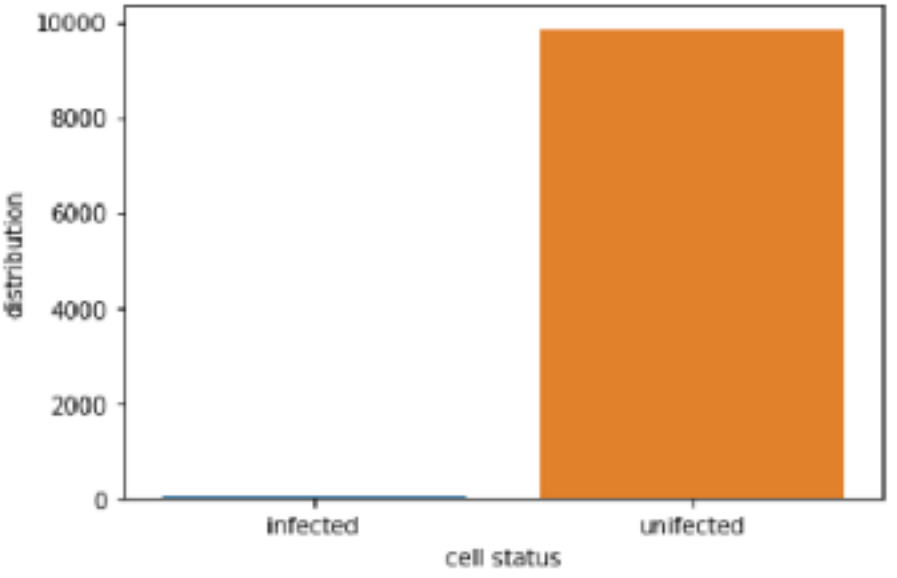
Training samples:
- **9850 Uninfected**
- **50 Infected**

Test samples:
-250 infected
-250 Uninfected
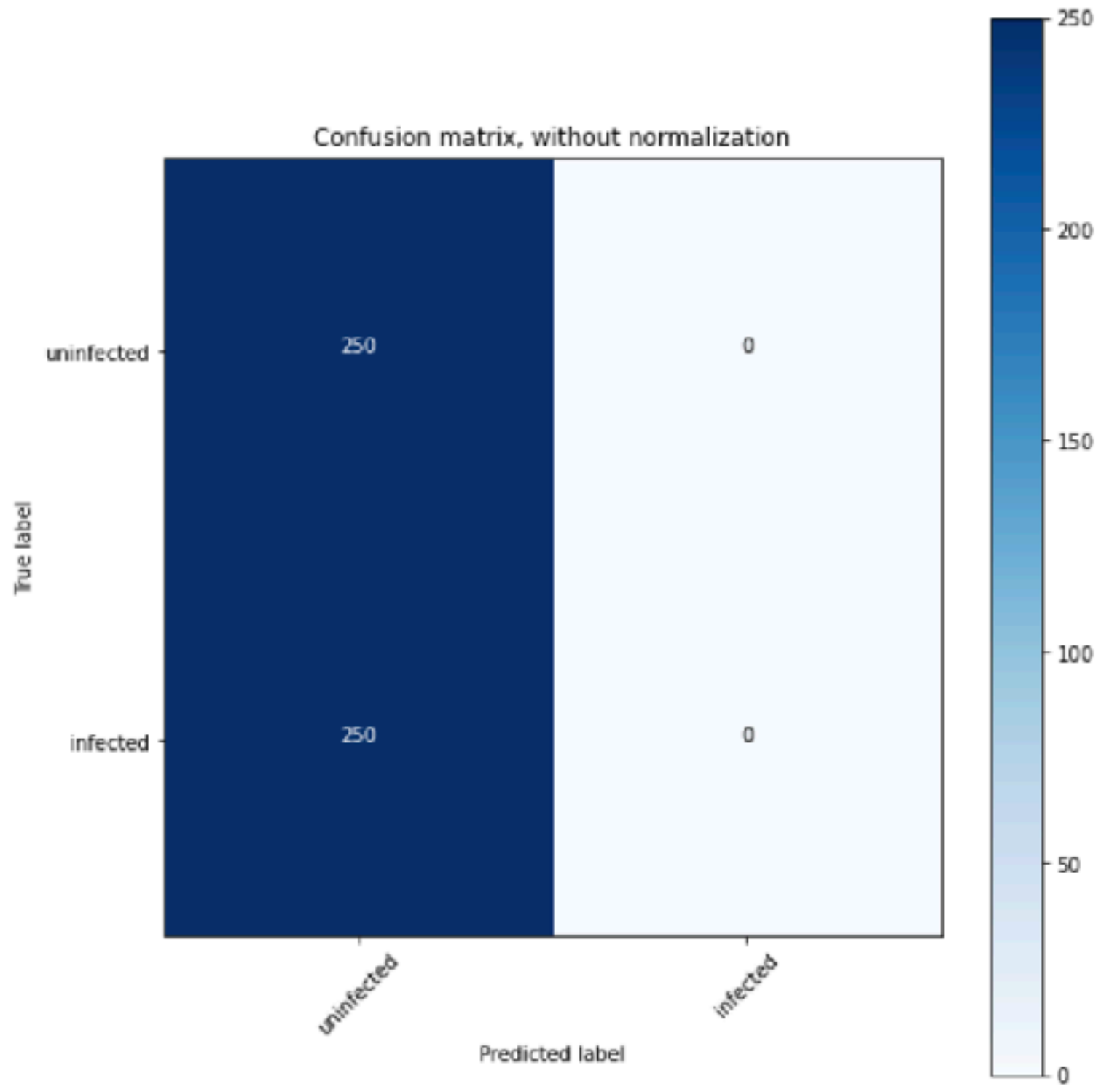
Early stopping matters a lot for imbalance dataset

Training samples:
- 9850 Uninfected
- 50 Infected

Test samples:
- 250 infected
- 250 Uninfected

No oversampling

Confusion matrix, without normalization

With oversampling

Normalized confusion matrix