Average Accuracy, Loss, And F1 For Each Model During Training Process Across 8 Runs.