Model Performance Comparison (Balanced Accuracy vs. Model Size)