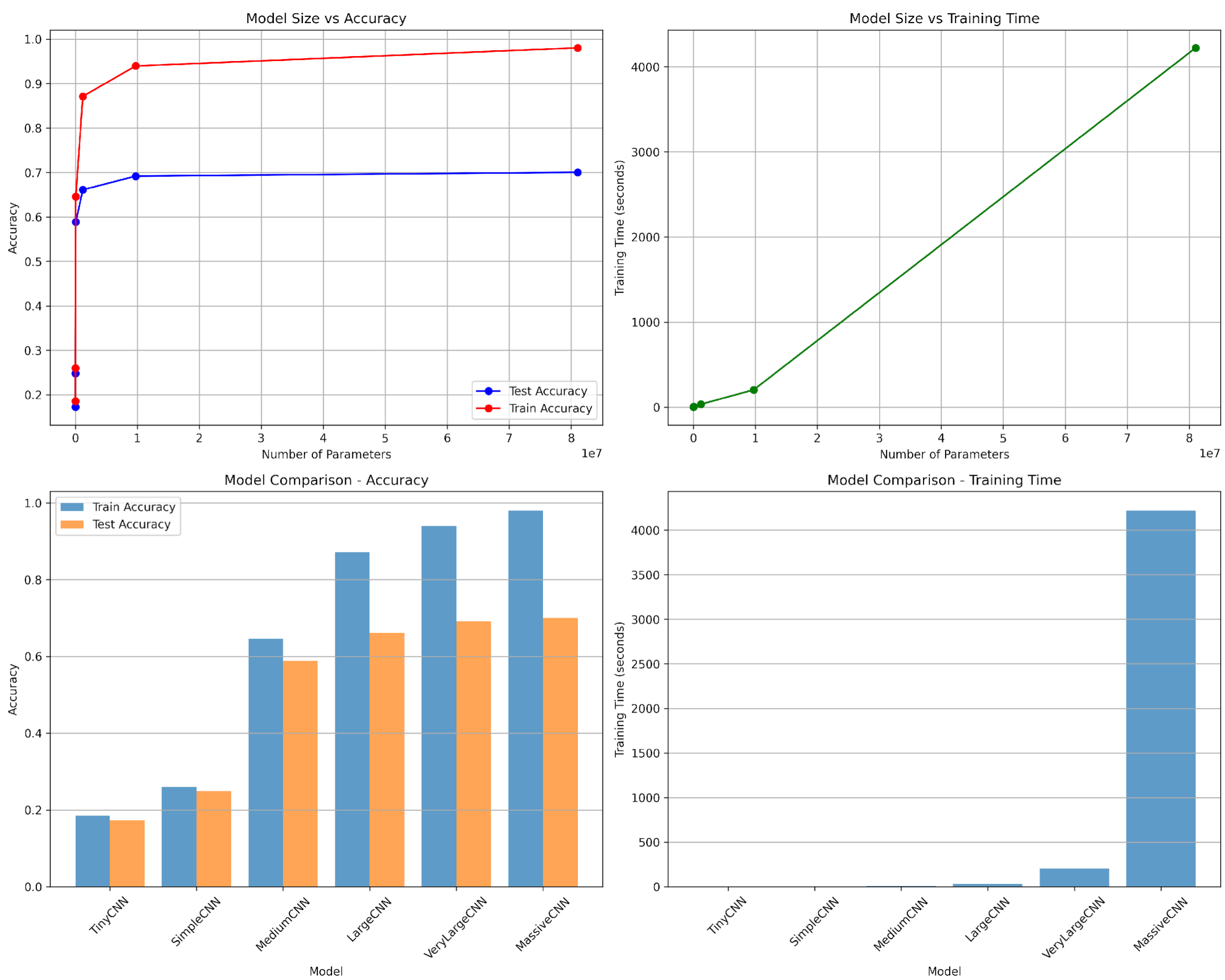
1. Experimental Results

Model	Parameters	Train Accuracy	Test Accuracy	Train Loss	Test Loss	Training Time (s)
TinyCNN	50	0.1854	0.1723	1.6543	1.6789	1.23
SimpleCNN	270	0.2597	0.2483	1.3994	1.4068	3.72
MediumCNN	68,828	0.6457	0.5886	0.8388	0.9266	6.52
LargeCNN	1,224,756	0.8713	0.6611	0.3742	0.8972	34.07
VeryLargeCNN	9,744,228	0.9396	0.6918	0.1779	1.0328	203.99
MassiveCNN	81,044,420	0.9803	0.7001	0.0531	1.5315	4221.37

2. Performance Visualizations



3. Brief Discussion

The experiments reveal clear trends in CNN performance for EEG emotion classification. Accuracy improved consistently with increasing model complexity, with MassiveCNN achieving the highest test accuracy of 70.0%—a 306% improvement over the TinyCNN baseline—but gains diminished beyond LargeCNN. Larger models also exhibited severe overfitting; MassiveCNN reached 98% training accuracy but only 70% test accuracy, indicating insufficient data to support such complexity. Training time and parameter counts grew rapidly with model size, from 1.2 seconds and 50 parameters for TinyCNN to over 70 minutes and 81 million parameters for MassiveCNN.

Overall, LargeCNN emerged as the optimal balance between accuracy (66.1%) and training time (34 seconds). Further improvements should focus on reducing overfitting through regularization, data augmentation, and early stopping, as well as exploring ensemble methods and transfer learning. The plateau in test accuracy suggests that dataset size may be the primary limitation, rather than architectural capacity alone.