Q2: LSTM baseline vs LSTM + attention (heads=4)