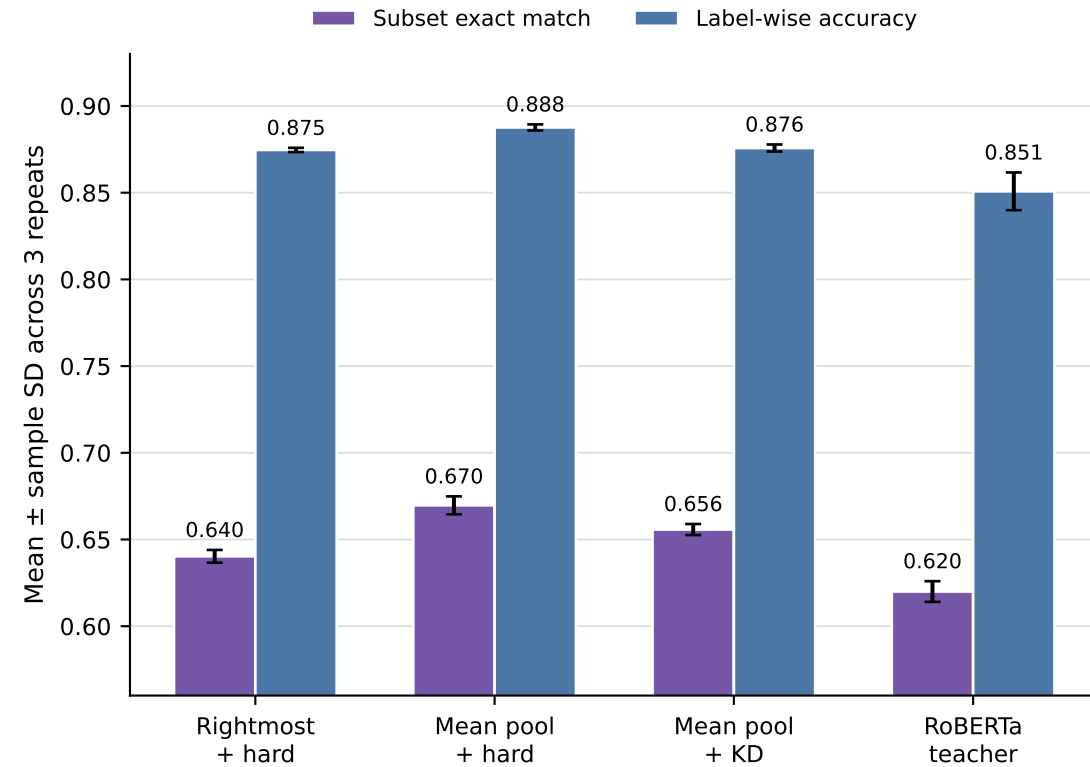
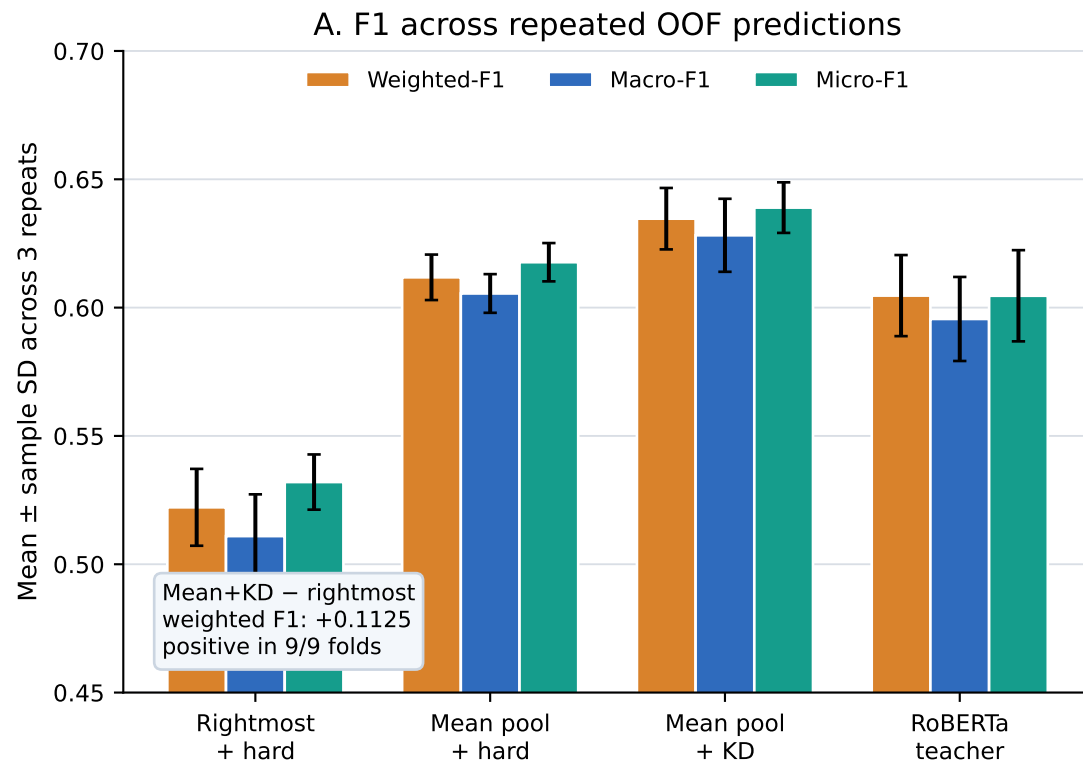


MALINT internal matched study: all Qwen student arms keep depth fixed at 4L

B. Accuracy-type metrics



3×3 matched OOF, 1,100 development rows per repeat, threshold 0.30. The RoBERTa teacher is an external control. This is a 4L interface result—not 4L-vs-8L/24L depth superiority; label-wise accuracy is imbalance-sensitive.