Model	iT	PT	DL	PT	DL	iT	PT	DL
From paper:	iTransformer			Patch TST		Ours		
lookback length	96	96	96	336	512	336		
Patch Len/Stride	-	12/*	-	16/8 to 24/2	-	16/8		
d_model	128-512	*	-	16-128	-	512		
d_ff	128-2048	*	-	128-256	-	512		
learning_rate	1e-3 to 5e-5	*	*	2.5e-4 to 1e-4	1e-2 to 1e-4	1E-04		
batchsize	16-32	*	*	24-128	8-32	4		
ETTh1	0.454	0.469	0.456	0.417	0.42	0.538	0.584	0.524
ETTh2	0.383	0.387	0.559	0.331	0.43	0.261	0.263	0.228
ETTm1	0.407	0.387	0.403	0.352	0.36	0.442	0.416	0.452
ETTm2	0.288	0.281	0.350	0.258	0.27	0.182	0.169	0.170
ECL	0.178	0.205	0.212	0.162	0.17	0.165	0.164	0.167
Exchange	0.360	0.367	0.354	-	-	0.462	0.544	0.390
Traffic	0.428	0.481	0.625	0.396	0.43	0.451	0.416	0.435
Weather	0.258	0.259	0.265	0.230	0.25	0.241	0.234	0.245
Average	0.345	0.353	0.403	0.307	0.33	0.343 (2.72%)	0.349 (4.35%)	0.326 (1.69%)

*Abbreviations: iT=iTransformer, PT=PatchTST, DL=DLinear, *=not available

Table 1: Comparison of different time series forecasting methods (DLinear, iTransformer, PatchTST) reported in prior papers versus our reimplementations under unified hyperparameter settings (rightmost group). Reported results often vary due to inconsistent lookback lengths and tuning strategies—e.g., the DLinear result in iTransformer uses a short lookback length of 96, which severely limits its capacity (e.g., only a 96×192 linear layer), leading to underfitting and degraded performance. Our unified setup (lookback=336) ensures a fairer and more challenging comparison across models, and outperforms similar reimplemented baselines by 1.69–4.35% in MSE.