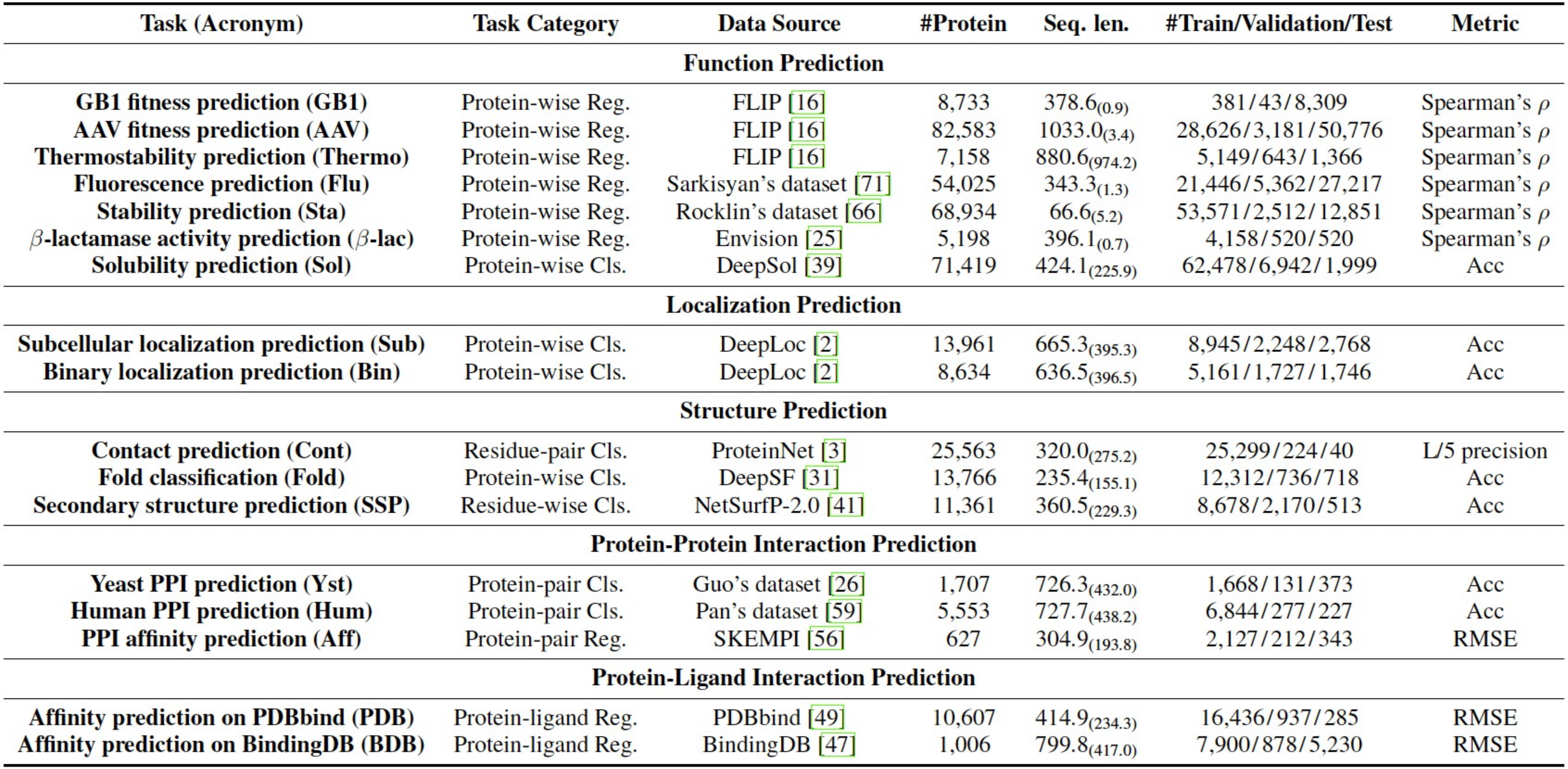
PEER: A Comprehensive and Multi-Task Benchmark for Protein Sequence Understanding

Minghao Xu

Mila – Quebec AI Institute

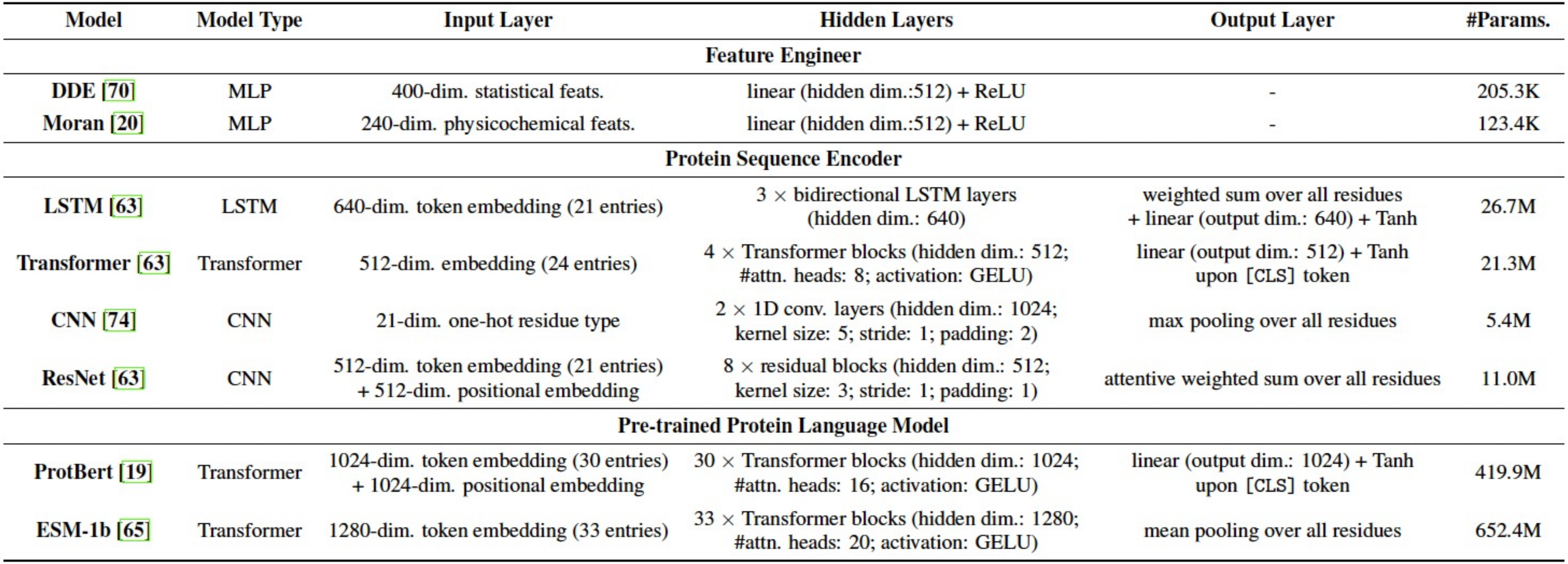
Acknowledgements: Zuobai Zhang, Jiarui Lu, Zhaocheng Zhu, Yangtian Zhang, Chang Ma, Runcheng Liu, Jian Tang

Benchmark Protein Sequence Understanding with Multiple Task Types



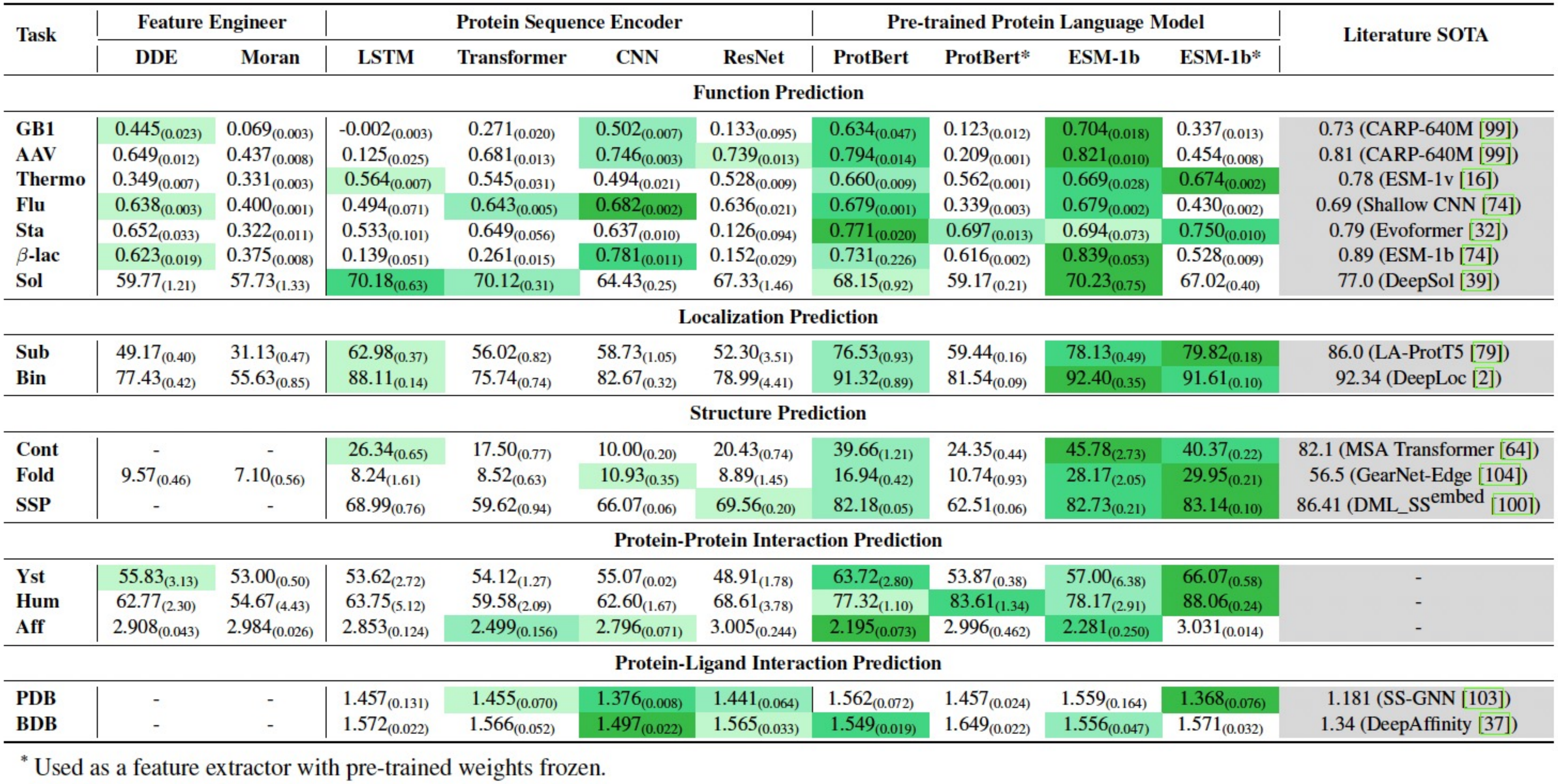
Task (Acronym)	Task Category	Data Source	#Protein	Seq. len.	#Train/Validation/Test	Metric
Function Prediction						
GB1 fitness prediction (GB1)	Protein-wise Reg.	FLIP [16]	8,733	$378.6_{(0.9)}$	381/43/8,309	Spearman's ρ
AAV fitness prediction (AAV)	Protein-wise Reg.	FLIP [16]	82,583	$1033.0_{(3.4)}$	28,626/3,181/50,776	Spearman's ρ
Thermostability prediction (Thermo)	Protein-wise Reg.	FLIP [16]	7,158	$880.6_{(974.2)}$	5,149/643/1,366	Spearman's ρ
Fluorescence prediction (Flu)	Protein-wise Reg.	Sarkisyan's dataset [71]	54,025	$343.3_{(1.3)}$	21,446/5,362/27,217	Spearman's ρ
Stability prediction (Sta)	Protein-wise Reg.	Rocklin's dataset [66]	68,934	$66.6_{(5.2)}$	53,571/2,512/12,851	Spearman's ρ
β-lactamase activity prediction (β-lac)	Protein-wise Reg.	Envision [25]	5,198	$396.1_{(0.7)}$	4,158/520/520	Spearman's ρ
Solubility prediction (Sol)	Protein-wise Cls.	DeepSol [39]	71,419	$424.1_{(225.9)}$	62,478/6,942/1,999	Acc
Localization Prediction						
Subcellular localization prediction (Sub)	Protein-wise Cls.	DeepLoc [2]	13,961	$665.3_{(395.3)}$	8,945/2,248/2,768	Acc
Binary localization prediction (Bin)	Protein-wise Cls.	DeepLoc [2]	8,634	$636.5_{(396.5)}$	5,161/1,727/1,746	Acc
Structure Prediction						
Contact prediction (Cont)	Residue-pair Cls.	ProteinNet [3]	25,563	$320.0_{(275.2)}$	25,299/224/40	L/5 precision
Fold classification (Fold)	Protein-wise Cls.	DeepSF [31]	13,766	$235.4_{(155.1)}$	12,312/736/718	Acc
Secondary structure prediction (SSP)	Residue-wise Cls.	NetSurfP-2.0 [41]	11,361	$360.5_{(229.3)}$	8,678/2,170/513	Acc
Protein-Protein Interaction Prediction						
Yeast PPI prediction (Yst)	Protein-pair Cls.	Guo's dataset [26]	1,707	$726.3_{(432.0)}$	1,668/131/373	Acc
Human PPI prediction (Hum)	Protein-pair Cls.	Pan's dataset [59]	5,553	$727.7_{(438.2)}$	6,844/277/227	Acc
PPI affinity prediction (Aff)	Protein-pair Reg.	SKEMPI [56]	627	$304.9_{(193.8)}$	2,127/212/343	RMSE
Protein-Ligand Interaction Prediction						
Affinity prediction on PDBbind (PDB)	Protein-ligand Reg.	PDBbind [49]	10,607	$414.9_{(234.3)}$	16,436/937/285	RMSE
Affinity prediction on BindingDB (BDB)	Protein-ligand Reg.	BindingDB [47]	1,006	$799.8_{(417.0)}$	7,900/878/5,230	RMSE

Baseline Models



Model	Model Type	Input Layer	Hidden Layers	Output Layer	#Params.
			Feature Engineer		
DDE [70]	MLP	400-dim. statistical feats.	linear (hidden dim.:512) + ReLU	-	205.3K
Moran [20]	MLP	240-dim. physicochemical feats.	linear (hidden dim.:512) + ReLU	-	123.4K
			Protein Sequence Encoder		
LSTM [63]	LSTM	640-dim. token embedding (21 entries)	3 × bidirectional LSTM layers (hidden dim.: 640)	weighted sum over all residues + linear (output dim.: 640) + Tanh	26.7M
Transformer [63]	Transformer	512-dim. embedding (24 entries)	4 × Transformer blocks (hidden dim.: 512; #attn. heads: 8; activation: GELU)	linear (output dim.: 512) + Tanh upon [CLS] token	21.3M
CNN [74]	CNN	21-dim. one-hot residue type	2 × 1D conv. layers (hidden dim.: 1024; kernel size: 5; stride: 1; padding: 2)	max pooling over all residues	5.4M
ResNet [63]	CNN	512-dim. token embedding (21 entries) + 512-dim. positional embedding	8 × residual blocks (hidden dim.: 512; kernel size: 3; stride: 1; padding: 1)	attentive weighted sum over all residues	11.0M
			Pre-trained Protein Language Model		
ProtBert [19]	Transformer	1024-dim. token embedding (30 entries) + 1024-dim. positional embedding	30 × Transformer blocks (hidden dim.: 1024; #attn. heads: 16; activation: GELU)	linear (output dim.: 1024) + Tanh upon [CLS] token	419.9M
ESM-1b [65]	Transformer	1280-dim. token embedding (33 entries)	33 × Transformer blocks (hidden dim.: 1280; #attn. heads: 20; activation: GELU)	mean pooling over all residues	652.4M

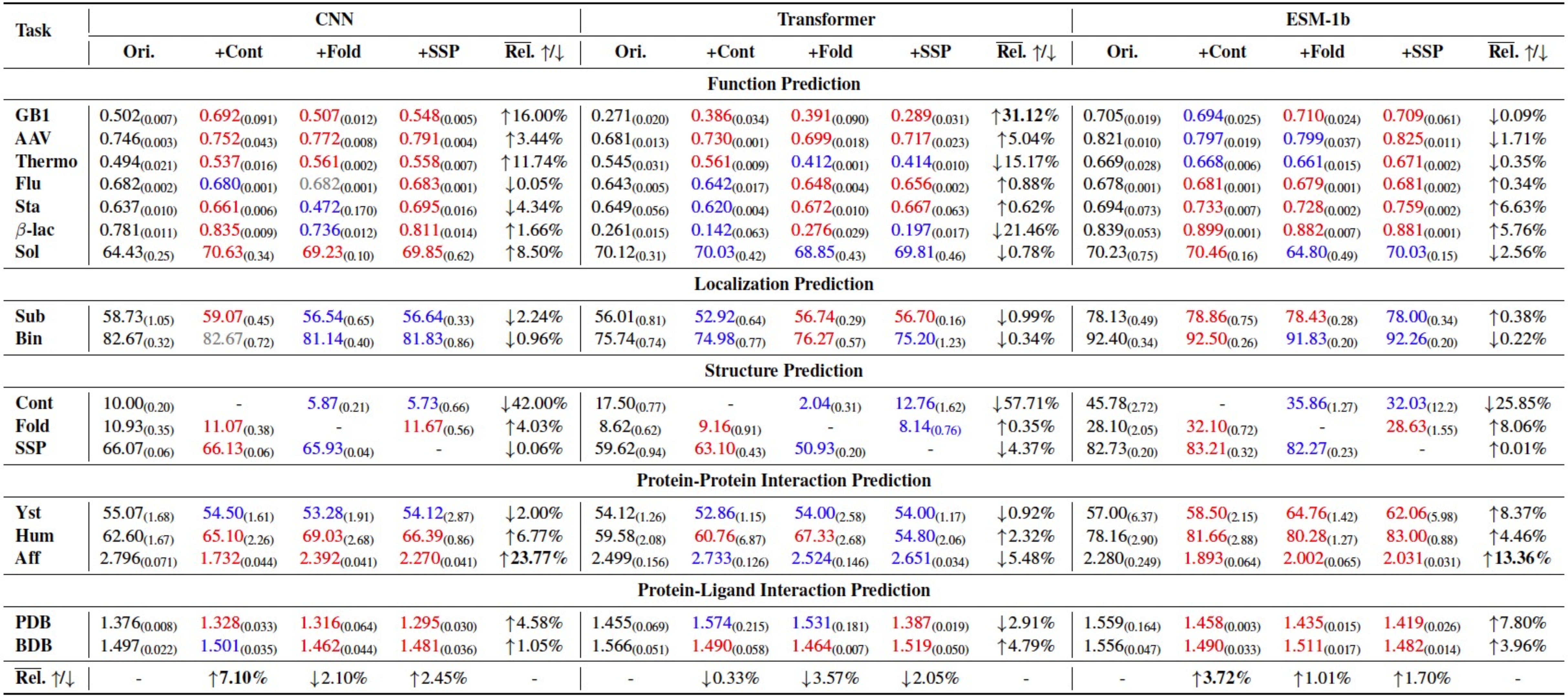
Benchmark Results on Single-Task Learning



Task	Feature Engineer		Protein Sequence Encoder				Pre-trained Protein Language Model				Literature SOTA
	DDE	Moran	LSTM	Transformer	CNN	ResNet	ProtBert	ProtBert*	ESM-1b	ESM-1b*	
Function Prediction											
GB1	0.445 (0.023)	0.069 (0.003)	-0.002 (0.003)	0.271 (0.020)	0.502 (0.007)	0.133 (0.095)	0.634 (0.047)	0.123 (0.012)	0.704 (0.018)	0.337 (0.013)	0.73 (CARP-640M [99])
AAV	0.649 (0.012)	0.437 (0.008)	0.125 (0.025)	0.681 (0.013)	0.746 (0.003)	0.739 (0.013)	0.794 (0.014)	0.209 (0.001)	0.821 (0.010)	0.454 (0.008)	0.81 (CARP-640M [99])
Thermo	0.349 (0.007)	0.331 (0.003)	0.564 (0.007)	0.545 (0.031)	0.494 (0.021)	0.528 (0.009)	0.660 (0.009)	0.562 (0.001)	0.669 (0.028)	0.674 (0.002)	0.78 (ESM-1v [16])
Flu	0.638 (0.003)	0.400 (0.001)	0.494 (0.071)	0.643 (0.005)	0.682 (0.002)	0.636 (0.021)	0.679 (0.001)	0.339 (0.003)	0.679 (0.002)	0.430 (0.002)	0.69 (Shallow CNN [74])
Sta	0.652 (0.033)	0.322 (0.011)	0.533 (0.101)	0.649 (0.056)	0.637 (0.010)	0.126 (0.094)	0.771 (0.020)	0.697 (0.013)	0.694 (0.073)	0.750 (0.010)	0.79 (Evoformer [32])
β-lac	0.623 (0.019)	0.375 (0.008)	0.139 (0.051)	0.261 (0.015)	0.781 (0.011)	0.152 (0.029)	0.731 (0.226)	0.616 (0.002)	0.839 (0.053)	0.528 (0.009)	0.89 (ESM-1b [74])
Sol	59.77 (1.21)	57.73 (1.33)	70.18 (0.63)	70.12 (0.31)	64.43 (0.25)	67.33 (1.46)	68.15 (0.92)	59.17 (0.21)	70.23 (0.75)	67.02 (0.40)	77.0 (DeepSol [39])
Localization Prediction											
Sub	49.17 (0.40)	31.13 (0.47)	62.98 (0.37)	56.02 (0.82)	58.73 (1.05)	52.30 (3.51)	76.53 (0.93)	59.44 (0.16)	78.13 (0.49)	79.82 (0.18)	86.0 (LA-ProtT5 [79])
Bin	77.43 (0.42)	55.63 (0.85)	88.11 (0.14)	75.74 (0.74)	82.67 (0.32)	78.99 (4.41)	91.32 (0.89)	81.54 (0.09)	92.40 (0.35)	91.61 (0.10)	92.34 (DeepLoc [2])
Structure Prediction											
Cont	-	-	26.34 (0.65)	17.50 (0.77)	10.00 (0.20)	20.43 (0.74)	39.66 (1.21)	24.35 (0.44)	45.78 (2.73)	40.37 (0.22)	82.1 (MSA Transformer [64])
Fold	9.57 (0.46)	7.10 (0.56)	8.24 (1.61)	8.52 (0.63)	10.93 (0.35)	8.89 (1.45)	16.94 (0.42)	10.74 (0.93)	28.17 (2.05)	29.95 (0.21)	56.5 (GearNet-Edge [104])
SSP	-	-	68.99 (0.76)	59.62 (0.94)	66.07 (0.06)	69.56 (0.20)	82.18 (0.05)	62.51 (0.06)	82.73 (0.21)	83.14 (0.10)	86.41 (DML_SS^{embed} [100])
Protein-Protein Interaction Prediction											
Yst	55.83 (3.13)	53.00 (0.50)	53.62 (2.72)	54.12 (1.27)	55.07 (0.02)	48.91 (1.78)	63.72 (2.80)	53.87 (0.38)	57.00 (6.38)	66.07 (0.58)	-
Hum	62.77 (2.30)	54.67 (4.43)	63.75 (5.12)	59.58 (2.09)	62.60 (1.67)	68.61 (3.78)	77.32 (1.10)	83.61 (1.34)	78.17 (2.91)	88.06 (0.24)	-
Aff	2.908 (0.043)	2.984 (0.026)	2.853 (0.124)	2.499 (0.156)	2.796 (0.071)	3.005 (0.244)	2.195 (0.073)	2.996 (0.462)	2.281 (0.250)	3.031 (0.014)	-
Protein-Ligand Interaction Prediction											
PDB	-	-	1.457 (0.131)	1.455 (0.070)	1.376 (0.008)	1.441 (0.064)	1.562 (0.072)	1.457 (0.024)	1.559 (0.164)	1.368 (0.076)	1.181 (SS-GNN [103])
BDB	-	-	1.572 (0.022)	1.566 (0.052)	1.497 (0.022)	1.565 (0.033)	1.549 (0.019)	1.649 (0.022)	1.556 (0.047)	1.571 (0.032)	1.34 (DeepAffinity [37])

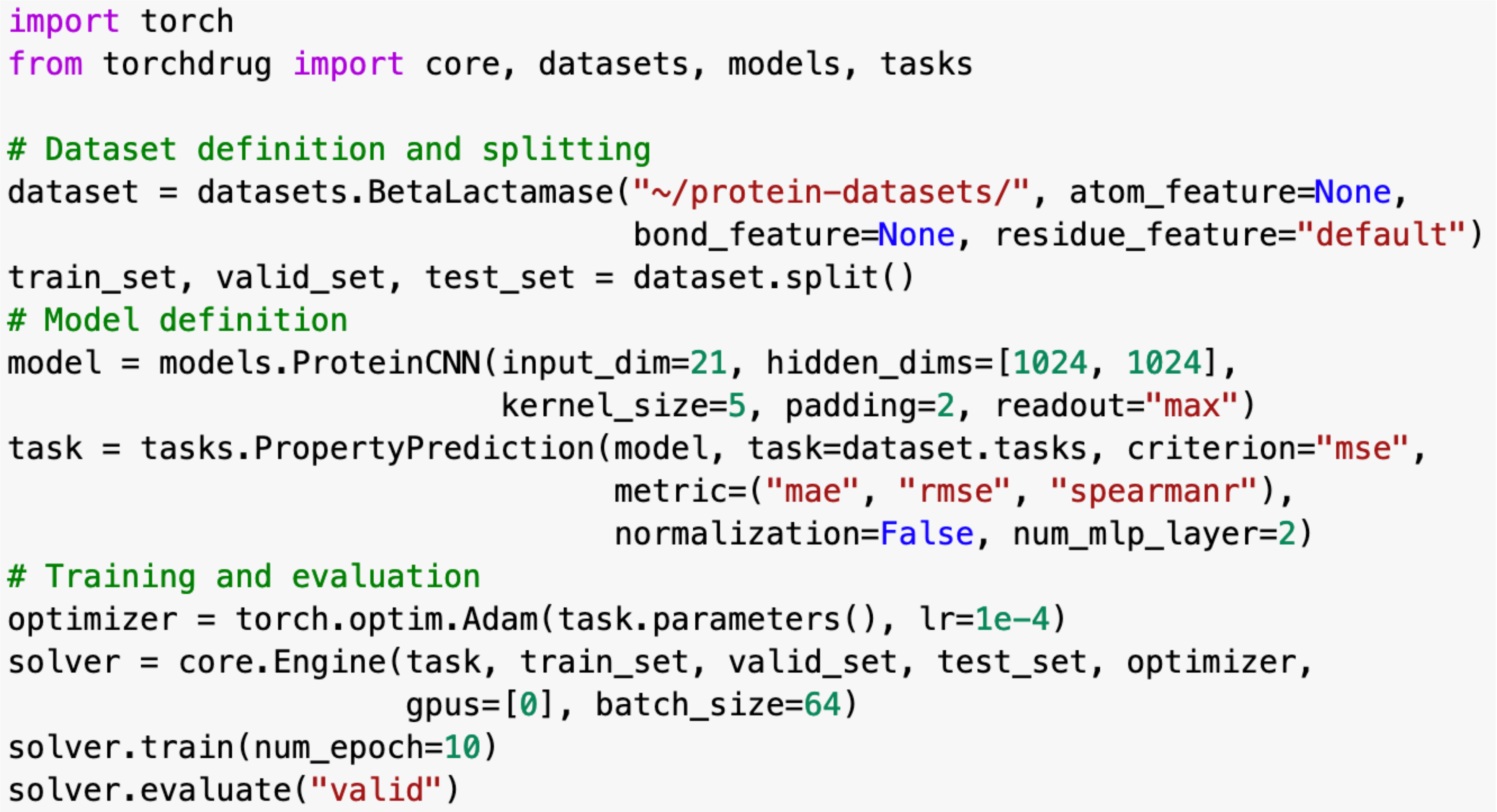
[*] Used as a feature extractor with pre-trained weights frozen.

Benchmark Results on Multi-Task Learning



Task	CNN					Transformer					ESM-1b				
	Ori.	+Cont	+Fold	+SSP	$\overline{\text{Rel.}}$ ↑/↓	Ori.	+Cont	+Fold	+SSP	$\overline{\text{Rel.}}$ ↑/↓	Ori.	+Cont	+Fold	+SSP	$\overline{\text{Rel.}}$ ↑/↓
Function Prediction															
GB1	$0.502_{(0.007)}$	$0.692_{(0.091)}$	$0.507_{(0.012)}$	$0.548_{(0.005)}$	↑16.00%	$0.271_{(0.020)}$	$0.386_{(0.034)}$	$0.391_{(0.090)}$	$0.289_{(0.031)}$	↑**31.12%**	$0.705_{(0.019)}$	$0.694_{(0.025)}$	$0.710_{(0.024)}$	$0.709_{(0.061)}$	↓0.09%
AAV	$0.746_{(0.003)}$	$0.752_{(0.043)}$	$0.772_{(0.008)}$	$0.791_{(0.004)}$	↑3.44%	$0.681_{(0.013)}$	$0.730_{(0.001)}$	$0.699_{(0.018)}$	$0.717_{(0.023)}$	↑5.04%	$0.821_{(0.010)}$	$0.797_{(0.019)}$	$0.799_{(0.037)}$	$0.825_{(0.011)}$	↓1.71%
Thermo	$0.494_{(0.021)}$	$0.537_{(0.016)}$	$0.561_{(0.002)}$	$0.558_{(0.007)}$	↑11.74%	$0.545_{(0.031)}$	$0.561_{(0.009)}$	$0.412_{(0.001)}$	$0.414_{(0.010)}$	↓15.17%	$0.669_{(0.028)}$	$0.668_{(0.006)}$	$0.661_{(0.015)}$	$0.671_{(0.002)}$	↓0.35%
Flu	$0.682_{(0.002)}$	$0.680_{(0.001)}$	$0.682_{(0.001)}$	$0.683_{(0.001)}$	↓0.05%	$0.643_{(0.005)}$	$0.642_{(0.017)}$	$0.648_{(0.004)}$	$0.656_{(0.002)}$	↑0.88%	$0.678_{(0.001)}$	$0.681_{(0.001)}$	$0.679_{(0.001)}$	$0.681_{(0.002)}$	↑0.34%
Sta	$0.637_{(0.010)}$	$0.661_{(0.006)}$	$0.472_{(0.170)}$	$0.695_{(0.016)}$	↓4.34%	$0.649_{(0.056)}$	$0.620_{(0.004)}$	$0.672_{(0.010)}$	$0.667_{(0.063)}$	↑0.62%	$0.694_{(0.073)}$	$0.733_{(0.007)}$	$0.728_{(0.002)}$	$0.759_{(0.002)}$	↑6.63%
β-lac	$0.781_{(0.011)}$	$0.835_{(0.009)}$	$0.736_{(0.012)}$	$0.811_{(0.014)}$	↑1.66%	$0.261_{(0.015)}$	$0.142_{(0.063)}$	$0.276_{(0.029)}$	$0.197_{(0.017)}$	↓21.46%	$0.839_{(0.053)}$	$0.899_{(0.001)}$	$0.882_{(0.007)}$	$0.881_{(0.001)}$	↑5.76%
Sol	$64.43_{(0.25)}$	$70.63_{(0.34)}$	$69.23_{(0.10)}$	$69.85_{(0.62)}$	↑8.50%	$70.12_{(0.31)}$	$70.03_{(0.42)}$	$68.85_{(0.43)}$	$69.81_{(0.46)}$	↓0.78%	$70.23_{(0.75)}$	$70.46_{(0.16)}$	$64.80_{(0.49)}$	$70.03_{(0.15)}$	↓2.56%
Localization Prediction															
Sub	$58.73_{(1.05)}$	$59.07_{(0.45)}$	$56.54_{(0.65)}$	$56.64_{(0.33)}$	↓2.24%	$56.01_{(0.81)}$	$52.92_{(0.64)}$	$56.74_{(0.29)}$	$56.70_{(0.16)}$	↓0.99%	$78.13_{(0.49)}$	$78.86_{(0.75)}$	$78.43_{(0.28)}$	$78.00_{(0.34)}$	↑0.38%
Bin	$82.67_{(0.32)}$	$82.67_{(0.72)}$	$81.14_{(0.40)}$	$81.83_{(0.86)}$	↓0.96%	$75.74_{(0.74)}$	$74.98_{(0.77)}$	$76.27_{(0.57)}$	$75.20_{(1.23)}$	↓0.34%	$92.40_{(0.34)}$	$92.50_{(0.26)}$	$91.83_{(0.20)}$	$92.26_{(0.20)}$	↓0.22%
Structure Prediction															
Cont	$10.00_{(0.20)}$	-	$5.87_{(0.21)}$	$5.73_{(0.66)}$	↓42.00%	$17.50_{(0.77)}$	-	$2.04_{(0.31)}$	$12.76_{(1.62)}$	↓57.71%	$45.78_{(2.72)}$	-	$35.86_{(1.27)}$	$32.03_{(12.2)}$	↓25.85%
Fold	$10.93_{(0.35)}$	$11.07_{(0.38)}$	-	$11.67_{(0.56)}$	↑4.03%	$8.62_{(0.62)}$	$9.16_{(0.91)}$	-	$8.14_{(0.76)}$	↑0.35%	$28.10_{(2.05)}$	$32.10_{(0.72)}$	-	$28.63_{(1.55)}$	↑8.06%
SSP	$66.07_{(0.06)}$	$66.13_{(0.06)}$	$65.93_{(0.04)}$	-	↓0.06%	$59.62_{(0.94)}$	$63.10_{(0.43)}$	$50.93_{(0.20)}$	-	↓4.37%	$82.73_{(0.20)}$	$83.21_{(0.32)}$	$82.27_{(0.23)}$	-	↑0.01%
Protein-Protein Interaction Prediction															
Yst	$55.07_{(1.68)}$	$54.50_{(1.61)}$	$53.28_{(1.91)}$	$54.12_{(2.87)}$	↓2.00%	$54.12_{(1.26)}$	$52.86_{(1.15)}$	$54.00_{(2.58)}$	$54.00_{(1.17)}$	↓0.92%	$57.00_{(6.37)}$	$58.50_{(2.15)}$	$64.76_{(1.42)}$	$62.06_{(5.98)}$	↑8.37%
Hum	$62.60_{(1.67)}$	$65.10_{(2.26)}$	$69.03_{(2.68)}$	$66.39_{(0.86)}$	↑6.77%	$59.58_{(2.08)}$	$60.76_{(6.87)}$	$67.33_{(2.68)}$	$54.80_{(2.06)}$	↑2.32%	$78.16_{(2.90)}$	$81.66_{(2.88)}$	$80.28_{(1.27)}$	$83.00_{(0.88)}$	↑4.46%
Aff	$2.796_{(0.071)}$	$1.732_{(0.044)}$	$2.392_{(0.041)}$	$2.270_{(0.041)}$	↑**23.77%**	$2.499_{(0.156)}$	$2.733_{(0.126)}$	$2.524_{(0.146)}$	$2.651_{(0.034)}$	↓5.48%	$2.280_{(0.249)}$	$1.893_{(0.064)}$	$2.002_{(0.065)}$	$2.031_{(0.031)}$	↑**13.36%**
Protein-Ligand Interaction Prediction															
PDB	$1.376_{(0.008)}$	$1.328_{(0.033)}$	$1.316_{(0.064)}$	$1.295_{(0.030)}$	↑4.58%	$1.455_{(0.069)}$	$1.574_{(0.215)}$	$1.531_{(0.181)}$	$1.387_{(0.019)}$	↓2.91%	$1.559_{(0.164)}$	$1.458_{(0.003)}$	$1.435_{(0.015)}$	$1.419_{(0.026)}$	↑7.80%
BDB	$1.497_{(0.022)}$	$1.501_{(0.035)}$	$1.462_{(0.044)}$	$1.481_{(0.036)}$	↑1.05%	$1.566_{(0.051)}$	$1.490_{(0.058)}$	$1.464_{(0.007)}$	$1.519_{(0.050)}$	↑4.79%	$1.556_{(0.047)}$	$1.490_{(0.033)}$	$1.511_{(0.017)}$	$1.482_{(0.014)}$	↑3.96%
$\overline{\text{Rel.}}$ ↑/↓	-	↑**7.10%**	↓2.10%	↑2.45%	-	-	↓0.33%	↓3.57%	↓2.05%	-	-	↑**3.72%**	↑1.01%	↑1.70%	-

Concise Benchmarking on TorchProtein



```python
import torch
from torchdrug import core, datasets, models, tasks

# Dataset definition and splitting
dataset = datasets.BetaLactamase("~/protein-datasets/", atom_feature=None,
                                 bond_feature=None, residue_feature="default")
train_set, valid_set, test_set = dataset.split()
# Model definition
model = models.ProteinCNN(input_dim=21, hidden_dims=[1024, 1024],
                          kernel_size=5, padding=2, readout="max")
task = tasks.PropertyPrediction(model, task=dataset.tasks, criterion="mse",
                                metric=("mae", "rmse", "spearmanr"),
                                normalization=False, num_mlp_layer=2)
# Training and evaluation
optimizer = torch.optim.Adam(task.parameters(), lr=1e-4)
solver = core.Engine(task, train_set, valid_set, test_set, optimizer,
                     gpus=[0], batch_size=64)
solver.train(num_epoch=10)
solver.evaluate("valid")
```

More Information

TorchProtein

PEER Benchmark

PEER Benchmark
GitHub Repo

Thanks for watching!