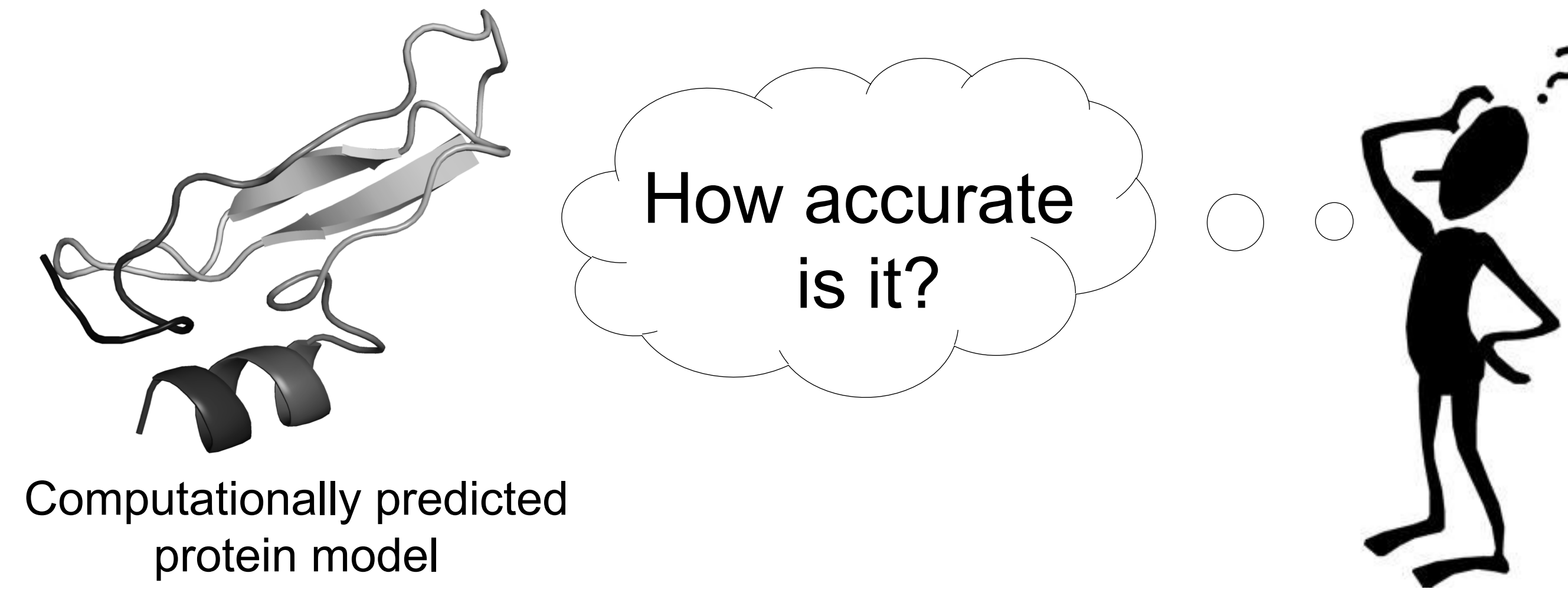# QDeep: distance-based protein model quality estimation by residue-level ensemble error classifications using stacked deep residual neural networks

Md Hossain Shuvo[1], Sutanu Bhattacharya[1], Debswapna Bhattacharya[1,2,*]

[1]Department of Computer Science and Software Engineering and [2]Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA

*bhattacharyad@auburn.edu

1. Protein model quality estimation problem

Computationally predicted protein model — How accurate is it?

2. Our contributions

- We introduced **Distance-based** protein single-model quality estimation method

- We employed **deep residual neural network** in protein scoring problem

- We utilized **ensemble of ResNet classifiers** and their combination for estimating the quality of protein models
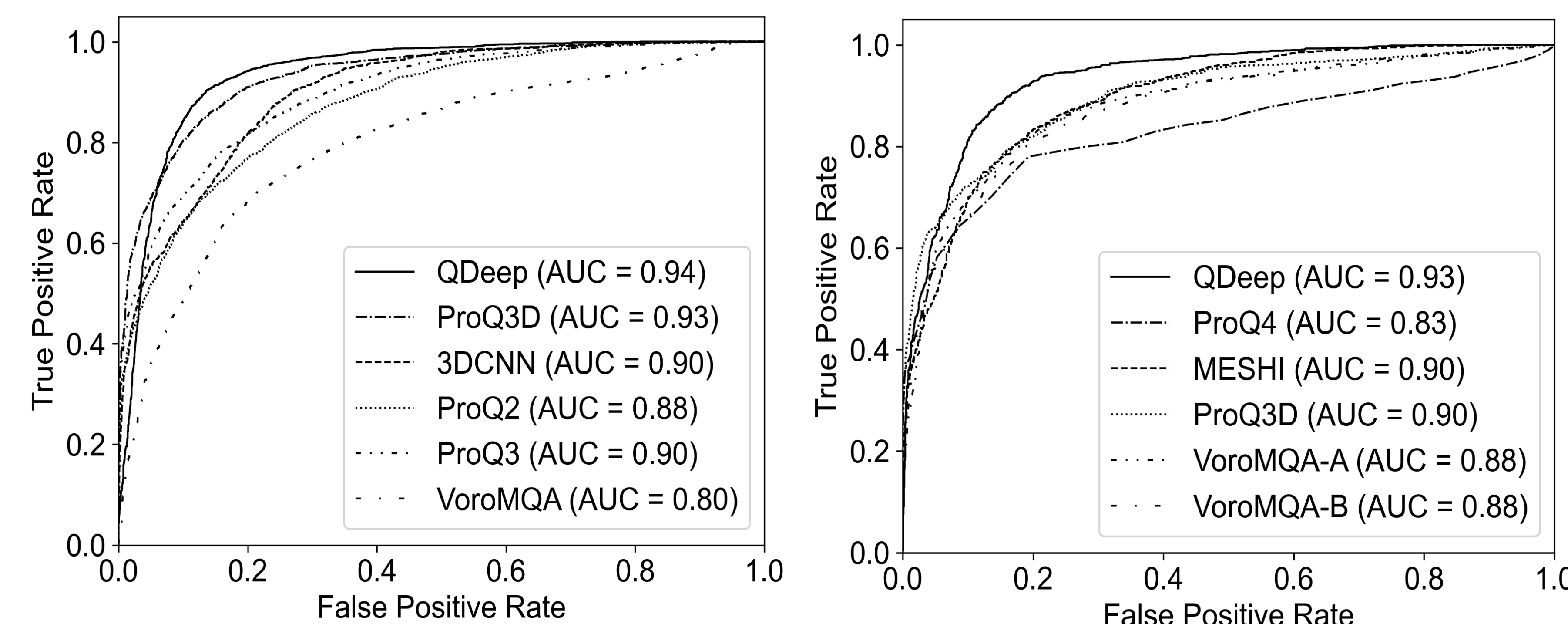
3. QDeep architecture

QDeep-score $= \dfrac{N1 + N2 + N4 + N8}{4L}$
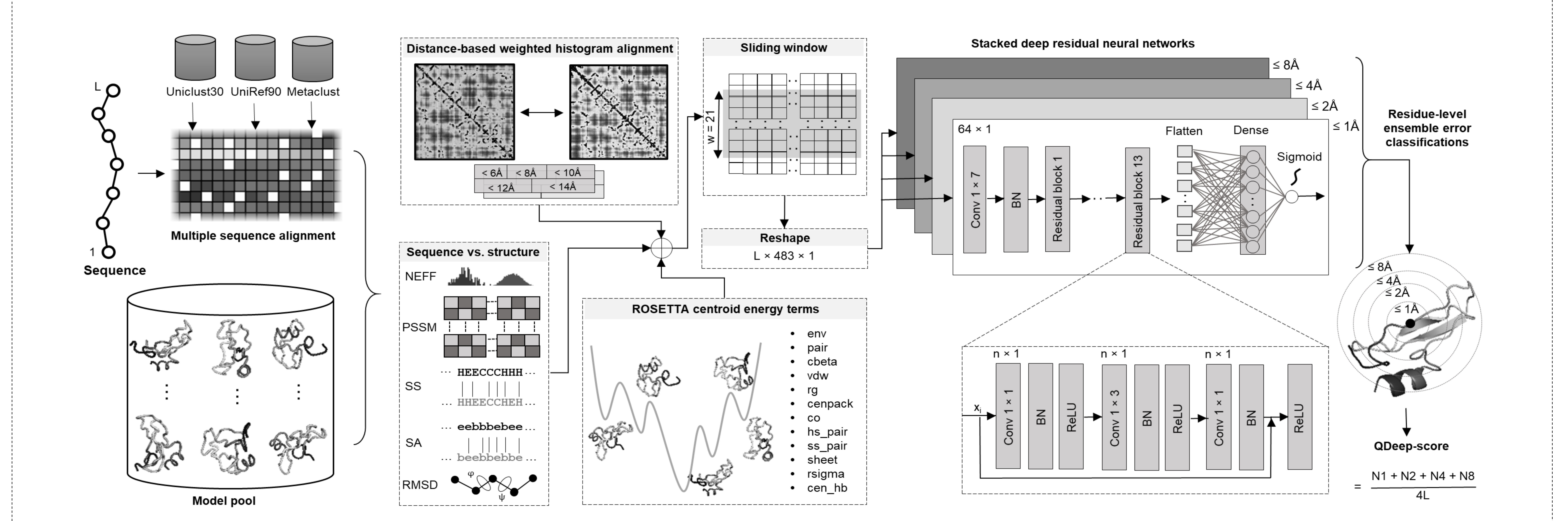
4. Results

Performance comparison of single-model quality estimation methods on CASP12 and CASP13 stage 2 datasets

Dataset	Method	Avg. r	Avg. ρ	Avg. τ	Avg. loss	Global r	Global ρ	Global τ
CASP12 (stage 2)	QDeep	**0.740**	**0.657**	**0.492**	**0.051**	**0.863**	**0.871**	**0.678**
	ProQ3D	0.688	0.631	0.467	0.086	0.851	0.847	0.660
	3DCNN	0.661	0.585	0.427	0.081	0.834	0.818	0.620
	ProQ2	0.624	0.556	0.404	0.091	0.784	0.770	0.577
	ProQ3	0.604	0.536	0.390	0.071	0.806	0.793	0.600
	VoroMQA	0.560	0.502	0.362	0.105	0.604	0.603	0.444
CASP13 (stage 2)	QDeep	**0.752**	**0.692**	**0.512**	0.088	**0.866**	**0.868**	**0.678**
	ProQ4	0.733	0.667	0.507	0.089	0.667	0.642	0.491
	MESHI	0.713	0.663	0.492	**0.070**	0.833	0.845	0.659
	ProQ3D	0.671	0.619	0.457	0.084	0.849	0.811	0.626
	VoroMQA-A	0.665	0.606	0.442	0.092	0.769	0.767	0.574
	VoroMQA-B	0.651	0.592	0.429	0.072	0.754	0.750	0.554

The ability of single-model quality estimation methods to distinguish good and bad models in CASP12 and CASP13 stage 2 datasets

QDeep (AUC = 0.94)
ProQ3D (AUC = 0.93)
3DCNN (AUC = 0.90)
ProQ2 (AUC = 0.88)
ProQ3 (AUC = 0.90)
VoroMQA (AUC = 0.80)

QDeep (AUC = 0.93)
ProQ4 (AUC = 0.83)
MESHI (AUC = 0.90)
ProQ3D (AUC = 0.90)
VoroMQA-A (AUC = 0.88)
VoroMQA-B (AUC = 0.88)

Performance comparison of deep ResNets models used in QDeep with other deep learning architectures

	CASP12 (stage 2)				CASP13 (stage 2)			
	Avg. r	Avg. ρ	Avg. τ	Avg. loss	Avg. r	Avg. ρ	Avg. τ	Avg. loss
ResNet	0.740	0.657	**0.492**	**0.051**	0.752	0.692	0.512	0.088
LSTM	0.716	0.596	0.452	0.059	0.735	0.668	0.500	0.116
CNN	0.657	0.581	0.433	0.097	0.735	0.660	0.487	0.116

Impact of deeper sequence alignment and the contribution of distance information

	CASP12 (stage 2)				CASP13 (stage 2)			
	Avg. r	Avg. ρ	Avg. τ	Avg. loss	Avg. r	Avg. ρ	Avg. τ	Avg. loss
QDeep	0.740	0.657	0.492	0.051	0.752	0.692	0.512	0.088
QDeep DeepMSA	**0.741**	**0.667**	**0.505**	0.062	**0.777**	**0.720**	**0.538**	**0.084**
QDeep NoDistance	0.677	0.601	0.442	0.065	0.668	0.613	0.445	0.091

5. Conclusions

- Utilizing the **ensemble of deep residual neural networks** potentially improves the performance of model quality estimation

- Effective integration of **distance information** showed performance boost across all accuracy measures

- The performance is further improved in part by incorporating **deeper sequence alignments**

6. Reference

M.H. Shuvo, S. Bhattacharya, D. Bhattacharya, QDeep: distance-based protein model quality estimation by residue-level ensemble error classifications using stacked deep residual neural networks, **ISMB Proceedings**, (2020)