

Supplementary Information for: Improved Enzyme Functional Annotation Prediction Using Contrastive Learning with Structural Inference

Yuxin Yang, Abby Jerger, Song Feng, Zixu Wang, Christina
Brasfield, Margaret S. Cheung, Jeremy Zucker, and Qiang Guan

1 Supplementary Text

1.1 Model performance when combining representations in different ways

To explore the impact of different combination methods of sequence and structure representations on the performance of CLEAN-Contact, we investigated two additional combination methods beyond the one detailed in the main text (See Section 4). The combination method outlined in the main text, where the two representations are projected to the same dimension and then added together, is denoted as *Addition*.

For the first additional combination method, we initially concatenate the sequence representation and structure representation. This concatenated representation is then passed through the projector:

$$\begin{aligned} z &= \mathcal{F}_\Psi(a, c), \\ &= W_3 (\text{LN}_2 (W_2 (\text{LN}_1 (W'_1 \text{Concat}(a, c))))), \end{aligned} \tag{1}$$

where $W'_1 \in \mathbb{R}^{d_1 \times (d_a + d_c)}$ is the weight of the first linear layer. This combination method is denoted as *Concat*₁, as the two representations are concatenated before the first linear layer.

For the second additional combination method, we initially pass both representations through the first level linear layers $W_{1,a}$ and $W_{1,c}$ as in the ‘Addition’ method. Subsequently, we concatenate the outputs of $W_{1,a}$ and $W_{1,c}$ before passing the combined representations through the rest of the neural network:

$$\begin{aligned} z &= \mathcal{F}_\Psi(a, c), \\ &= W_3 (\text{LN}_2 (W'_2 (\text{LN}'_1 (\text{Concat}(W_{1,a}a, W_{1,c}c))))), \end{aligned} \tag{2}$$

where $W'_2 \in \mathbb{R}^{d_2 \times (d_1 + d_1)}$ is the weight of the second linear layer. This combination method is denoted as *Concat*₂, reflecting the concatenation of the two representations before the second linear layer.

We proceed to evaluate all three combination methods alongside CLEAN¹ on both the New-392 and Price-149 test datasets. It is evident that, regardless of the approach used to combine sequence and structure representations, they consistently outperformed CLEAN, which lacks structural information (Supplementary Fig. 1). Notably, among the three combination methods, the *Addition* consistently demonstrated superior performance. We attribute this outcome to the principles of residual learning, known to facilitate optimization processes².

1.2 Cross-validation evaluation

To rigorously evaluate model performance and address potential overfitting concerns, we conducted a 5-fold cross-validation for both CLEAN-Contact and CLEAN, where 80% of the training dataset (179,793 amino acid sequences) are used for training. We then assessed each of these models on our test datasets (New-392 and Price-149), calculating the mean and standard deviations of performance metrics across the five runs. Notably, even when trained only on 80% of the training dataset, CLEAN-Contact consistently outperformed CLEAN across both test datasets (Supplementary Fig. 2).

2 Supplementary Figures

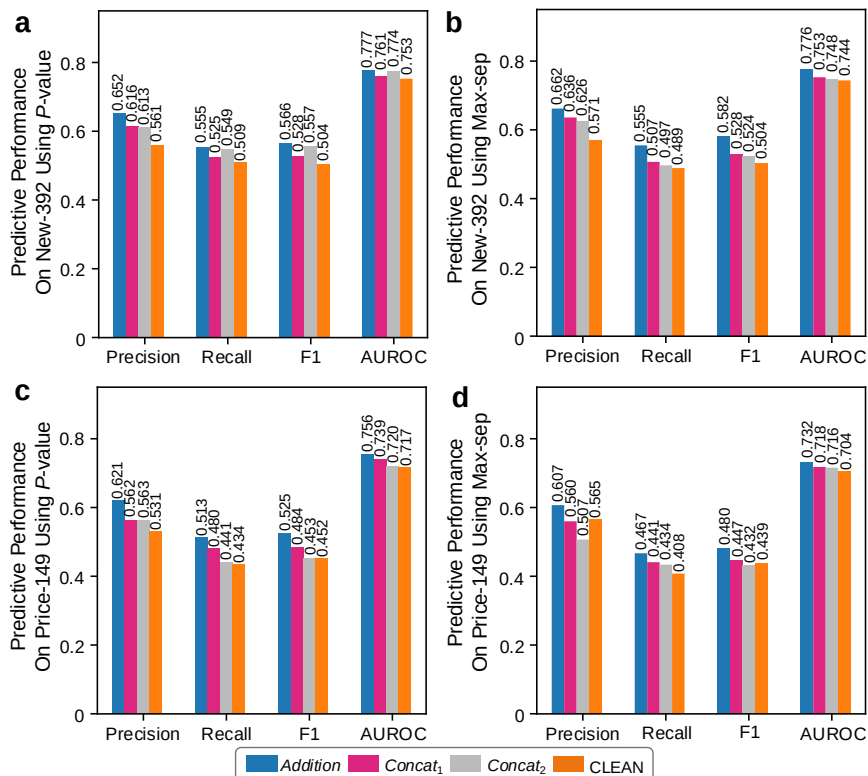


Fig. 1 Evaluation of predictive performance among various combinations of sequence and structure representations alongside CLEAN on **a**, the New-392 test dataset using the *P*-value EC number selection algorithm, **b**, the New-392 test dataset using the Max-separation EC number selection algorithm, **c**, the Price-149 test dataset using the *P*-value EC number selection algorithm, and **d**, the Price-149 test dataset using the Max-separation EC number selection algorithm.

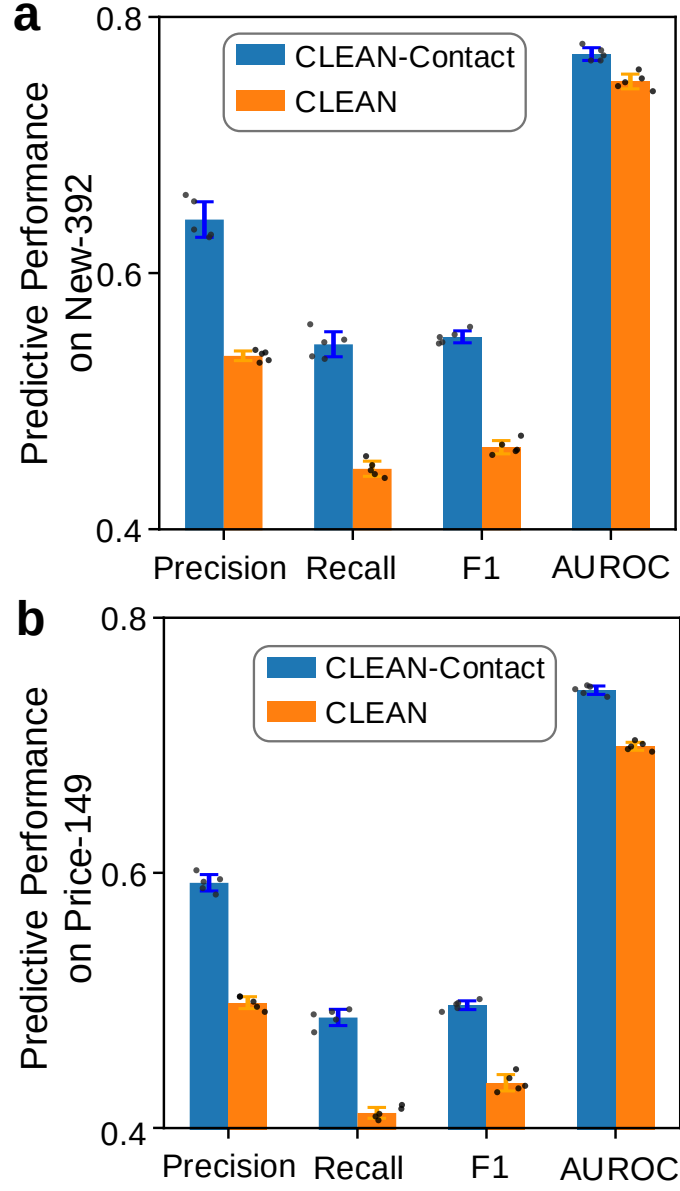


Fig. 2 Comparison of $n=5$ fold cross-validation results between CLEAN-Contact and CLEAN on **a**, the New-392 test dataset and **b**, the Price-149 test dataset. Bar plot elements: the bars represent the mean predictive performance and the error bars represent standard deviations of predictive performance. Bar chart represents mean of the metric from 5-fold cross-validation and error bar represents standard deviation of the metric from 5-fold cross-validation.

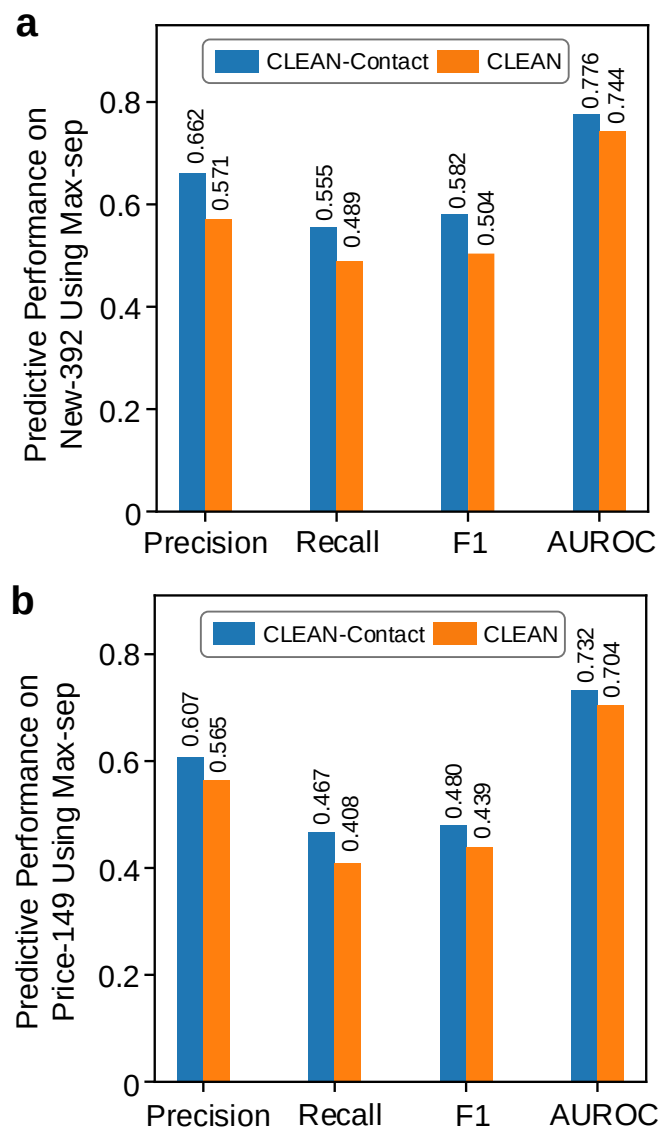


Fig. 3 Evaluation of predictive performance between CLEAN-Contact and CLEAN on **a**, the New-392 test dataset, and **b**, the Price-149 test dataset, evaluated using Precision, Recall, F1-score, and AUROC metrics. The Max-separation EC number selection algorithm was employed for selecting potential EC numbers.

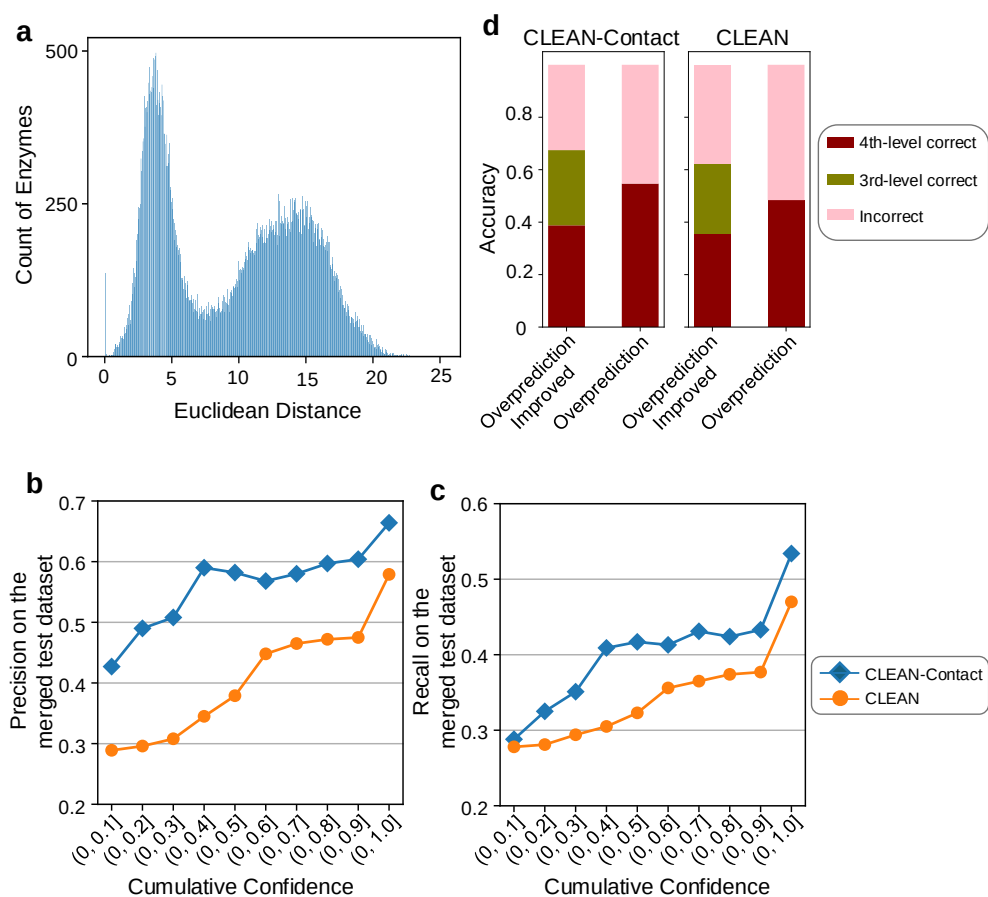


Fig. 4 Quantification of CLEAN-Contact's prediction confidence. **a**, Histogram showing the distribution of randomly sampled enzymes plotted against the Euclidean distance between the enzyme's combined representation and the cluster centers of EC numbers. The left peak illustrates the distribution when sampled enzymes correspond to their respective EC numbers while the right peak illustrates the distribution when sampled enzymes do not match their corresponding EC numbers. **b**, **c**, Evaluation of predictive performance comparing CLEAN-Contact and CLEAN on the merged test dataset, evaluated using the Precision and Recall metrics. This evaluation correlates with the ascending cumulative confidence derived from the fitted Gaussian Mixture Model. **d**, Mitigating overprediction issue through an adaptive prediction strategy, evaluated via Accuracy. Left: predictive performance of CLEAN-Contact is depicted when employing two strategies: predicting 4th-level EC numbers solely at high prediction confidence and predicting 3rd-level EC numbers otherwise (left bar: "Overprediction Improved"), versus predicting 4th-level EC numbers at all times (right bar: "Overprediction"). Right: predictive performance of CLEAN under the same two strategies: predicting 4th-level EC numbers solely at high confidence and predicting 3rd-level EC numbers otherwise (left bar: "Overprediction Improved"), versus predicting 4th-level EC numbers at all times (right bar: "Overprediction").

3 Supplementary Tables

	CLEAN-Contact		CLEAN	
	GFLOPs	Inference time (Nvidia A5000)	GFLOPs	Inference time (Nvidia A5000)
ESM-2	46,035	0.2379 sec	N/A	N/A
ESM-1b	N/A	N/A	10,846	0.0682 sec
ResNet-50	16.164	0.0078 sec	N/A	N/A
Prediction	0.003	0.0009 sec	0.001	0.0008 sec
Total	46,051.167	0.2466 sec	10,846.001	0.0690 sec

Table 1 Comparison of computational cost for inference of CLEAN-Contact and CLEAN on one protein using Giga Floating Operations (GFLOPs) and inference time on Nvidia A5000 GPU. The computational costs of the pretrained models, which are ESM-2, ESM-1b, and ResNet-50, are calculated using the average length of amino acid sequences in the test datasets, which is 439 amino acids.

	Accuracy (Top-1)	Parameter #
ResNet-18	69.758%	11.7M
ResNet-50	80.858%	25.6M
ResNet-101	81.886%	44.5M
SwinV2-B	84.112%	87.9M

Table 2 Comparison of benchmark performance on computer vision classification task and computational cost between commonly used convolutional neural networks (CNNs).

Model	Precision	Recall	F1-score	AUROC
ResNet-50 & ESM-2	0.634	0.562	0.563	0.780
ResNet-18 & ESM-2	0.593	0.515	0.516	0.757
ResNet-101 & ESM-2	0.639	0.566	0.560	0.778
SwinV2-B & ESM-2	0.370	0.246	0.269	0.622

Table 3 Performance evaluation of CLEAN-Contact variants on the merged dataset. Different pretrained computer vision models (ResNet-50, ResNet-18, ResNet-101, and SwinV2-B) were used to extract structure representations from protein contact maps.

# of Anchor Samples	# of Positive (Negative) Samples	Precision	Recall	F1-score	AUROC
1	1	0.634	0.562	0.563	0.78
1	2	0.626	0.542	0.552	0.77
1	4	0.601	0.513	0.521	0.757
2	1	0.632	0.51	0.515	0.754
2	2	0.631	0.527	0.542	0.763
4	1	0.632	0.53	0.536	0.764
4	4	0.62	0.51	0.528	0.754

Table 4 Evaluation of performance of CLEAN-Contact on the merged test dataset with different number of anchor, positive, and negative samples. The performance was measured using Precision, Recall, F1-score, and AUROC.

References

- [1] Yu, T., Cui, H., Li, J.C., Luo, Y., Jiang, G., Zhao, H.: Enzyme function prediction using contrastive learning. *Science* **379**(6639), 1358–1363 (2023)
- [2] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)