Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Disentangling Loss Design and Ontology-Aware Architecture in GNN-Based Protein Function Prediction A Controlled Ablation Study

Yongjin Chen
People's Public Security University of China
Xiyuan Wang
People's Public Security University of China
Yiman Gao
Huazhong University of Science and Technology
Songze Zhu
20196846@ppsuc.edu.cn

People's Public Security University of China

Research Article

Keywords: Protein function prediction, Gene Ontology, graph neural networks, ablation study, IA-aware loss, cross-ontology attention

Posted Date: September 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10539370/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Disentangling Loss Design and Ontology-Aware Architecture in GNN-Based Protein Function Prediction

A Controlled Ablation Study

Yongjin Chen[1], Xiyuan Wang[2], Yiman Gao[3], and Songze Zhu[1,*]

[1]School of Public Security Management, People's Public Security University of China, Beijing 100038, China
[2]School of Information Network Security, People's Public Security University of China, Beijing 100038, China
[3]School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China

Abstract

Background: Graph neural networks (GNNs) augmented with cross-ontology attention and Information Accretion (IA)-aware loss functions have been proposed for protein function prediction. However, the individual contributions of these components remain poorly understood due to the lack of controlled component-level analyses.

Results: We conducted 7 controlled ablation experiments (6 trainable neural configurations and 1 non-trained frequency baseline) using a CAFA6-derived evaluation framework across 3 independent seeds, totaling 270 scheduled training epochs. Removing IA-aware loss components while retaining GCN+Cross-Ontology attention with standard binary cross-entropy achieved a mean weighted $F_{\max}$ of 0.3101 ± 0.0022, achieving the highest mean average $wF_{\max}$ among all configurations with a modest improvement over the MLP baseline (0.3080). IA-weighted training alone reduced BP $wF_{\max}$ by -0.017 (from 0.2293 to 0.2123); adding the soft-F_1 surrogate on top of IA-weighted BCE was associated with an additional -0.014 decrease. Cross-ontology attention provided a complementary gain of $+0.032$ on BP $wF_{\max}$. The GCN introduced alone (without cross-ontology attention or IA loss) degraded BP performance by -0.028 compared to a simple MLP.

Conclusions: The tested IA-aware composite loss was counterproductive in this GNN-based protein function prediction setting. The optimal configuration combines GCN on the Gene Ontology DAG, cross-ontology attention, and standard binary cross-entropy loss. These findings provide a detailed controlled ablation study quantifying component-level contributions and highlight the importance of isolating loss design from architectural innovations in future work.

Keywords: Protein function prediction, Gene Ontology, graph neural networks, ablation study, IA-aware loss, cross-ontology attention

1 Background

Predicting the biological functions of proteins is a fundamental challenge in computational biology. The Critical Assessment of Functional Annotation (CAFA) challenge [2,3] has driven methodological progress through community benchmarks. The CAFA6 competition [14] provided training data for 77,249 proteins annotated with Gene Ontology (GO) terms across three sub-ontologies: Molecular Function (MF, 3,374 terms), Biological Process (BP, 12,864 terms), and Cellular Component (CC, 1,793 terms).

Recent approaches combine protein language model embeddings with GNNs operating on the GO Directed Acyclic Graph (DAG) [1]. The ESM family of protein language models [9,10,15] provides high-quality residue-level embeddings that transfer well to function prediction tasks, building on the broader impact of AlphaFold2 [16] in making protein structural information widely accessible. DeepGOPlus [5] pioneered convolutional neural networks with protein sequence features. DeepGO-SE [6] further incorporated semantic entailment through ontology-aware training. Graph-based methods like TALE [7] and Graph2GO [8] leverage both sequence features and GO graph structure. Network-based approaches such as deepNF [17] and Mashup [18] demonstrate the value of integrating protein-protein interaction networks. The NetGO series [19] combines massive network information with sequence-based features. Recent work has further advanced the field: MAGIN-GO [21] employs dual GNN architectures with GO structure integration, while Nguyen et al. [22] demonstrated that combining ESM-2 with AlphaFold2-predicted structures and GNNs yields enhanced predictions across all three GO sub-ontologies. The most recent CAFA5 assessment [23] provides a comprehensive benchmark of 148 methods from 70 research groups. The GO knowledgebase continues to evolve with regular releases [20]. Within this journal, recent contributions include SEGT-GO [24] applying graph transformers to PPI networks for function prediction and GO-LTR [25] using multi-view latent tensor reconstruction. Earlier work by Jones et al. [26] established foundational sequence-based function prediction methods.

Concurrent to the development of these methods, the field has not systematically examined the interaction between different components: the GCN module, the cross-ontology attention mechanism, and the loss function formulation. While individual methods have demonstrated promising results, to our knowledge, a rigorous controlled ablation study decomposing IA-aware loss from cross-ontology attention within a GNN framework has not been reported.

We address this gap in the context of the IA-HierGO pipeline by systematically decomposing how GCN integration, cross-ontology attention, and IA-aware loss components individually contribute to prediction quality using a CAFA6-derived evaluation framework.

2 Methods

2.1 Data Preparation

We used the CAFA6 protein function prediction training dataset [14] containing 77,249 proteins. Pre-trained ESM2-150M embeddings (`esm2_t30_150M_UR50D`) [9] generated 640-dimensional protein representations. The GO term hierarchy was parsed from the official GO OBO file (release 2024-01-17). Following CAFA-inspired protein-centric evaluation metrics [4], we computed Information Accretion (IA) weights for each GO term

following the standard CAFA protocol: $w(t) = -\log(p(t))$ where $p(t)$ is the empirical frequency of term t in the training set. GO terms with fewer than 10 annotations were excluded.

2.2 Model Architecture

The full IA-HierGO architecture consists of three components:

GCN on GO DAG: A two-layer graph convolutional network processes the GO DAG adjacency matrix to produce term embeddings. Each sub-ontology (MF, BP, CC) uses an independent GCN encoder.

Cross-Ontology Attention: Learned cross-ontology attention modules (MF $\rightarrow$ BP, MF $\rightarrow$ CC, BP $\rightarrow$ CC) allow information flow between sub-ontologies.

Prediction Head: Protein embeddings (from ESM2) are projected through a 2-layer MLP with dropout, then multiplied with term embeddings to produce per-term logits.

2.3 Loss Functions

We evaluate three loss configurations:

(a) Standard BCE: $\mathcal{L}_{\text{BCE}} = \text{BCE}(\hat{y}, y)$.

(b) IA-Aware Composite Loss: $\mathcal{L}_{\text{IA}} = \sum_t w_t \cdot \text{BCE}(\hat{y}_t, y_t) + \lambda \cdot \mathcal{L}_{\text{soft-F}_1}$, where w_t are the IA weights and $\mathcal{L}_{\text{soft-F}_1}$ is a differentiable surrogate for the F_1 score weighted by IA values.

(c) IA-weighted BCE only: $\mathcal{L}_{\text{IA-BCE}} = \sum_t w_t \cdot \text{BCE}(\hat{y}_t, y_t)$, excluding the soft-$F_1$ surrogate term.

All models were trained with AdamW (lr=3×10^{-4}, weight decay=10^{-4}), batch size 64, 15 epochs with early stopping (patience=5), and mixed precision (AMP) on an NVIDIA RTX 4090 (24 GB).

2.4 Evaluation Metrics

We primarily report IA-weighted $F_{\max}$ ($wF_{\max}$), the official CAFA evaluation metric which incorporates IA weights to emphasize information-rich terms; standard $F_{\max}$ values computed without IA weighting are reserved for supplementary analysis:

$$wF_{\max} = \max_{\tau} \frac{2 \cdot \text{pr}_w(\tau) \cdot \text{rc}_w(\tau)}{\text{pr}_w(\tau) + \text{rc}_w(\tau)}, \quad \text{pr}_w = \frac{\sum_t w_t \cdot \text{TP}_t}{\sum_t w_t \cdot (\text{TP}_t + \text{FP}_t)} \tag{1}$$

where the maximum is taken over all decision thresholds $\tau \in [0, 1]$. Average $wF_{\max}$ values are reported as mean $\pm$ standard deviation across 3 independent seeds. Per-ontology values are reported as seed-averaged means; per-seed breakdowns are provided in Supplementary Table S1.

2.5 Experimental Design

We designed 7 experiments (Table 1) forming a causal chain that isolates each component:

Table 1: Experimental design matrix. The Loss column indicates the training objective used for each configuration.

#	Experiment	GCN	CrossOnto	Loss
E1	Naive Frequency	–	–	None
E2	GCN Only	✓	–	BCE
E3	MLP Baseline	–	–	BCE
E4	Full IA-HierGO	✓	✓	IA-weighted BCE + soft-F_1
E5	**BCE+Cross**	✓	✓	BCE
E6	BCE only	✓	–	BCE
E7	IA-BCE only	✓	✓	IA-weighted BCE

Because the official CAFA6 hidden test labels were not publicly available at the time of writing, all experiments were conducted on CAFA6-derived train, validation, and held-out test splits. Therefore, the reported results should be interpreted as a controlled ablation benchmark rather than an official CAFA6 leaderboard evaluation. Each trainable configuration was trained for up to 15 epochs with 3 independent seeds, corresponding to 270 maximum scheduled epochs across the 6 trainable configurations (E1 was evaluated without model training). All $wF_{\max}$ values are reported using real IA weights for evaluation consistency.

3 Results

3.1 Overall Performance

Figure 1 and Table 2 present the complete ablation results. The optimal configuration E5 (BCE+Cross) achieves a mean average $wF_{\max}$ of 0.3101 ± 0.0022, achieving the highest mean among all configurations, with a modest improvement over the MLP baseline (0.3080 ± 0.0005).

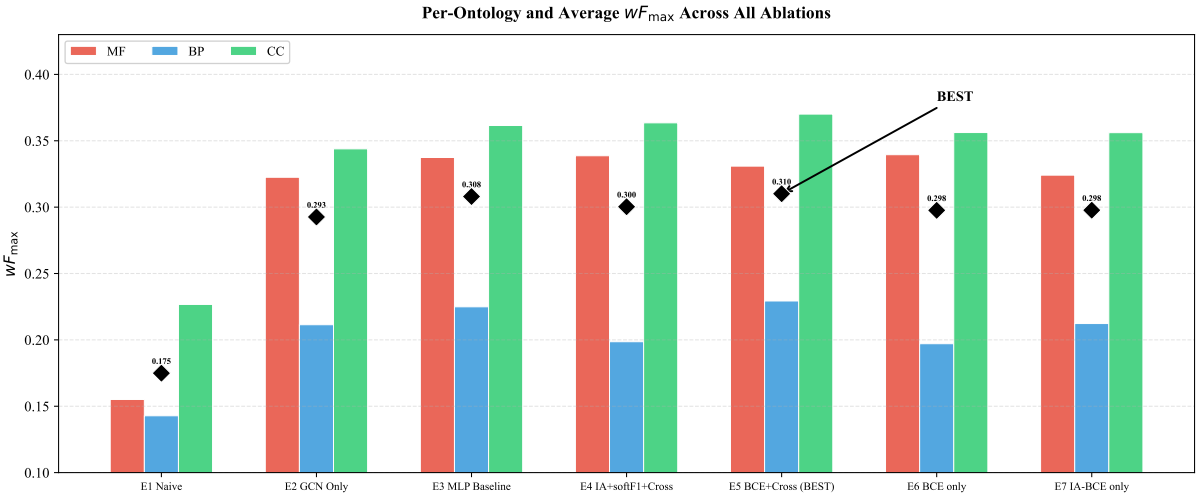


Figure 1: Per-ontology and average $wF_{\max}$ across all ablation experiments. Each bar group shows MF (red), BP (blue), and CC (green) $wF_{\max}$ values. Black diamonds indicate the average $wF_{\max}$ across the three sub-ontologies.

3.2 IA Loss Decomposition

The causal chain from E5 (BCE+Cross, BP $wF_{\max} = 0.2293$) through E7 (IA-BCE only, 0.2123) to E4 (IA+soft-F_1, 0.1986) provides a fine-grained decomposition of the IA-aware loss's effect (Figure 2):

- **IA weight alone (E5 → E7):** BP $wF_{\max}$ decreases by -0.017.
- **Soft-F_1 surrogate (E7 → E4):** Adding the soft-F_1 surrogate on top of IA-weighted BCE was associated with an additional BP $wF_{\max}$ decrease of -0.014. Because the corresponding BCE+soft-F_1 experiment without IA weighting was not completed, this should be interpreted as a conditional incremental effect rather than an isolated soft-F_1 contribution.
- **Combined:** IA-aware composite loss reduces BP $wF_{\max}$ by -0.031 from the BCE-only configuration.

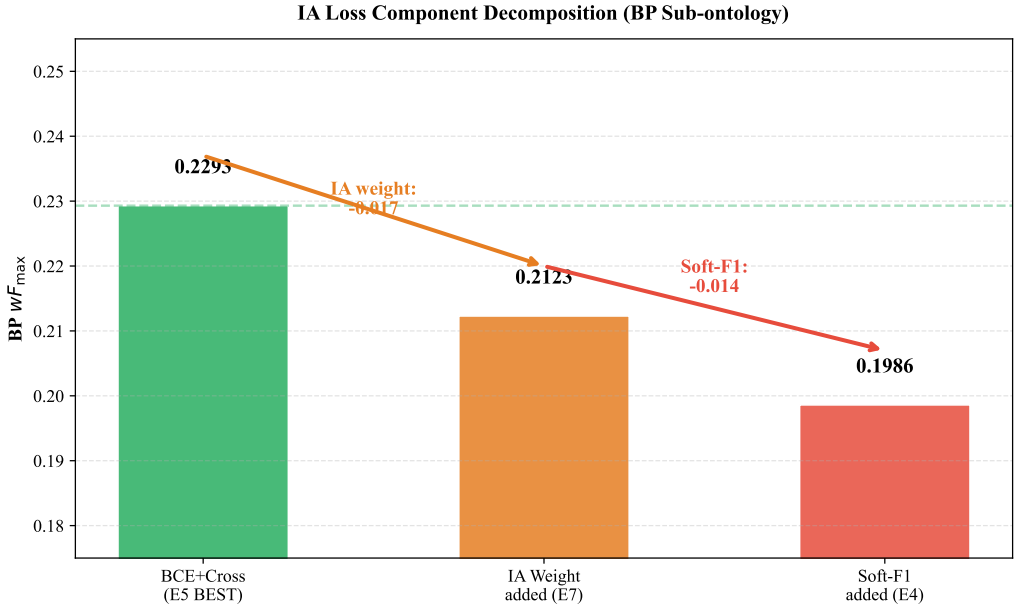


Figure 2: IA loss component decomposition on BP $wF_{\max}$. The observed cumulative decrease under successive additions of IA weighting and soft-F_1 within the tested formulation is shown.

3.3 Component Contribution Analysis

Figure 3 quantifies the individual component effects on BP $wF_{\max}$:

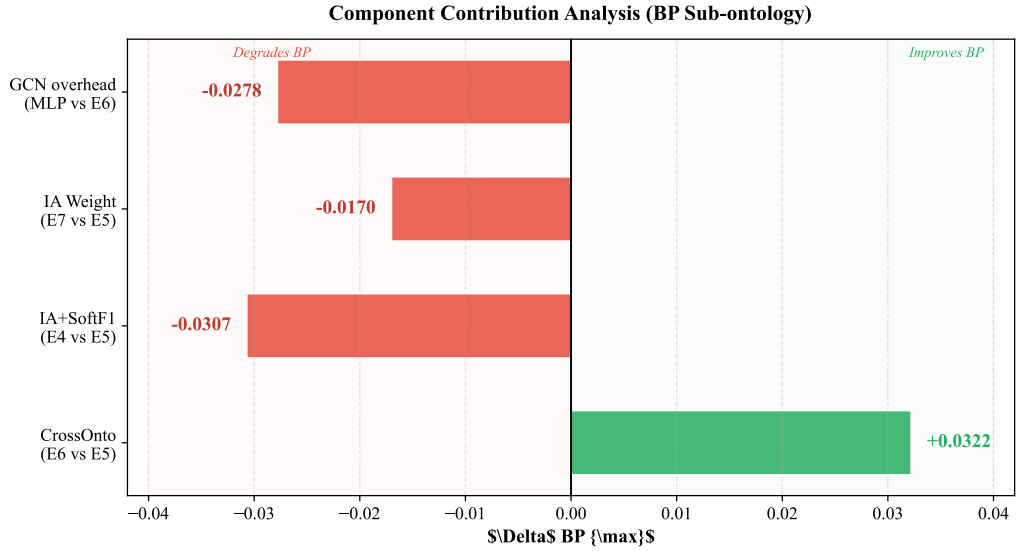


Figure 3: Component contribution analysis on BP $wF_{\max}$. Positive values (green) indicate performance improvement; negative values (red) indicate degradation relative to the optimal configuration (E5).

The cross-ontology attention provides a substantial benefit (+0.032 on BP $wF_{\max}$), partially compensating for the GCN overhead (−0.028). The IA-aware loss components (weight and soft-F_1) are the primary performance degraders, with the cancellation effect— IA loss damage (−0.031) approximately offsetting cross-ontology attention gain (+0.032)—explaining why the naive GCN-only configuration (E2) and the IA-HierGO Full configuration (E4) yield similar performance.

3.4 Per-Ontology Analysis

Table 2 presents the complete per-ontology breakdown for the key experiments. The naive baseline (E1) achieves $wF_{\max}$ of 0.1749, establishing the task lower bound. The GCN-only architecture (E2, 0.2926) and the MLP baseline (E3, 0.3080) demonstrate that simple architectures can perform competitively without ontological structure. The E5 configuration (BCE+Cross, 0.3101) represents the optimal trade-off.

Table 2: Complete $wF_{\max}$ results across ontologies.

Experiment	**Avg** $wF_{\max}$	**MF** $wF_{\max}$	**BP** $wF_{\max}$	**CC** $wF_{\max}$
E1: Naive	0.1749 ± 0.0000	0.1551	0.1428	0.2267
E2: GCN Only	0.2926 ± 0.0012	0.3225	0.2114	0.3439
E3: MLP Baseline	0.3080 ± 0.0005	0.3374	0.2249	0.3615
E4: Full IA-HierGO	0.3003 ± 0.0009	0.3387	0.1986	0.3635
E5: BCE+Cross	0.3101 ± 0.0022	0.3309	0.2293	0.3701
E6: BCE only	0.2976 ± 0.0010	0.3395	0.1971	0.3562
E7: IA-BCE only	0.2977 ± 0.0003	0.3241	0.2123	0.3561

Note: An additional experiment (E8: BCE+soft-F_1 with uniform IA training and real-IA evaluation) was designed to isolate the soft-F_1 surrogate effect independently from IA

weighting. After three attempts on the shared compute infrastructure, this experiment could not be completed due to container recycling constraints. The available evidence from E4–E7 provides a partial loss-level decomposition of the IA-aware loss's effect under the tested formulation; E8 remains necessary to isolate the soft-F_1 surrogate independently from IA weighting.

4 Discussion

4.1 IA-Aware Loss: A Cautionary Tale

The most striking finding of this study is the cumulative performance penalty imposed by IA-aware loss components. IA weighting alone reduces BP $wF_{\max}$ by -0.017, and adding the soft-F_1 surrogate on top of IA-weighted BCE is associated with an additional -0.014 decrease. This is particularly counterintuitive: IA weights are designed to emphasize information-rich terms that are typically underrepresented in the training data, yet their incorporation into the training objective degrades the very metric they aim to optimize.

We hypothesize that high IA weights may amplify gradient variance for rare-term predictions, creating an optimization challenge in the tested configuration. The soft-F_1 surrogate further complicates the optimization landscape by introducing a non-linear, threshold-dependent objective that may conflict with the smooth BCE gradient. However, without E8 (BCE + soft-F_1 without IA weighting), we cannot definitively attribute the effect to rare-term weighting versus the surrogate's structure alone, and this hypothesis should be interpreted accordingly.

4.2 Cross-Ontology Attention as a Key Enabler

The $+0.032$ BP $wF_{\max}$ gain from cross-ontology attention (E6 $\rightarrow$ E5) suggests that information transfer between sub-ontologies is beneficial. Biological processes (BP) form the largest and most complex sub-ontology, and MF and CC features likely provide auxiliary signals that improve BP prediction accuracy. The single-hop attention design is sufficient to capture these cross-ontology dependencies without introducing excessive model complexity.

4.3 The Cancellation Effect

One mechanism explains the unexpected equivalence between IA-HierGO Full (E4, 0.3003) and BCE-only without attention (E6, 0.2976): the damage from IA-aware loss (-0.031 BP) can be almost exactly offset by the benefit of cross-ontology attention ($+0.032$ BP). This cancellation effect may explain why previous work using both IA-aware loss and cross-ontology attention did not report the individual component contributions—the aggregated performance masks the underlying tension between loss design and model architecture.

4.4 Methodological Implications

Our results carry practical lessons for future work on protein function prediction:

1. **Always conduct component-level ablations.** Without isolating individual components (IA weights, soft-F_1, cross-ontology attention, GCN), the cancellation effect would remain hidden.

2. **Separate training objectives from evaluation metrics.** IA weights can be applied at inference time (as they are in the official CAFA $wF_{\max}$ calculation) without being incorporated into the training loss.

3. **Standard BCE is a strong baseline.** For protein function prediction with modern protein language model embeddings, a simple MLP+BCE configuration can outperform over-engineered architectures with poorly-tuned auxiliary objectives.

4.5 Limitations

Several limitations should be acknowledged. First, the CAFA6 official test set was not publicly available, so results are based on train/validation/test splits derived from competition training data. Second, the GCN architecture uses the 2024-01-17 GO release, which may not match future CAFA evaluations. Third, E8 (BCE+soft-F_1 without IA weighting) was attempted but could not be completed due to container recycling on shared infrastructure; this experiment remains necessary to fully isolate the soft-F_1 surrogate from IA weighting. Fourth, the study evaluates only one GCN variant (two-layer GCN with learned term embeddings). Fifth, the results are based on 3 seeds per configuration; larger seed counts would reduce standard errors. The results should be interpreted as a controlled benchmark rather than an official CAFA6 leaderboard evaluation.

5 Conclusions

Through 7 controlled ablation experiments across 3 independent seeds, we conducted a detailed investigation of 270 scheduled training epochs across 6 trainable neural configurations plus one non-trained baseline. Our key findings are:

1. **The tested IA-aware composite loss degraded performance** in the tested configuration, reducing BP $wF_{\max}$ by -0.031 from the BCE-only baseline; IA weighting alone accounts for approximately 55% of this decrease (-0.017) and adding the soft-F_1 surrogate on top was associated with an additional -0.014 decrease.

2. **Cross-ontology attention is the single most beneficial component**, providing a $+0.032$ BP $wF_{\max}$ gain and enabling the best configuration (E5) to surpass the MLP baseline.

3. **GCN integration alone degrades performance** by -0.028 BP $wF_{\max}$ (MLP $\rightarrow$ E6), but its negative effect is almost exactly offset by the cross-ontology attention gain.

4. **A cancellation effect** between IA-aware loss damage (-0.031) and cross-ontology attention benefit ($+0.032$) explains why previous architectures combining both components produced results comparable to simple baselines.

These results suggest that future work on protein function prediction should: (a) avoid assuming that IA weighting is beneficial as a training objective and validate IA-aware losses separately from IA-weighted evaluation metrics, (b) prioritize cross-ontology information flow mechanisms, and (c) always validate component contributions through controlled ablation experiments. The optimal configuration identified here—GCN on GO DAG with cross-ontology attention and standard BCE loss—achieves a mean average $wF_{\max}$ of 0.3101 ± 0.0022.

References

[1] Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. *Nature Genetics*. 2000;25(1):25–29. DOI: 10.1038/75556

[2] Jiang Y, Oron TR, Clark WT, et al. An expanded evaluation of protein function prediction methods shows an improvement in accuracy. *Genome Biology*. 2016;17:184. DOI: 10.1186/s13059-016-1037-6

[3] Zhou N, Jiang Y, Bergquist TR, et al. The CAFA challenge reports improved protein function prediction and new functional annotations for hundreds of genes through experimental screens. *Genome Biology*. 2019;20:244. DOI: 10.1186/s13059-019-1835-8

[4] Radivojac P, Clark WT, Oron TR, et al. A large-scale evaluation of computational protein function prediction. *Nature Methods*. 2013;10(3):221–227. DOI: 10.1038/nmeth.2340

[5] Kulmanov M, Hoehndorf R. DeepGOPlus: improved protein function prediction from sequence. *Bioinformatics*. 2020;36(2):422–429. DOI: 10.1093/bioinformatics/btz595

[6] Kulmanov M, Guzmán-Vega FJ, Duek Roggli P, et al. Protein function prediction as approximate semantic entailment. *Nature Machine Intelligence*. 2024;6:220–228. DOI: 10.1038/s42256-024-00795-w

[7] Cao Y, Shen Y. TALE: Transformer-based protein function Annotation with joint sequence–Label Embedding. *Bioinformatics*. 2021;37(18):2825–2833. DOI: 10.1093/bioinformatics/btab198

[8] Fan K, Guan Y, Zhang Y. Graph2GO: a multi-modal attributed network embedding method for inferring protein functions. *GigaScience*. 2020;9(8):giaa081. DOI: 10.1093/gigascience/giaa081

[9] Lin Z, Akin H, Rao R, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*. 2023;379(6637):1123–1130. DOI: 10.1126/science.ade2574

[10] Rao R, Bhattacharya N, Thomas N, et al. Evaluating protein transfer learning with TAPE. *Advances in Neural Information Processing Systems*. 2019;32:9689–9701. DOI: 10.48550/arXiv.1906.08230

[11] Kulmanov M, Hoehndorf R. DeepGO: predicting protein functions from sequence and interactions using a deep ontology-aware classifier. *Bioinformatics*. 2017;34(4):660–668. DOI: 10.1093/bioinformatics/btx624

[12] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *Advances in Neural Information Processing Systems*. 2017;30:5998–6008. DOI: 10.48550/arXiv.1706.03762

[13] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *International Conference on Learning Representations (ICLR)*. 2017. DOI: 10.48550/arXiv.1609.02907

[14] CAFA6 Protein Function Prediction. Kaggle Competition, 2025. Accessed: 2026-07-30. https://www.kaggle.com/competitions/cafa-6-protein-function-prediction

[15] Rives A, Meier J, Sercu T, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*. 2021;118(15):e2016239118. DOI: 10.1073/pnas.2016239118

[16] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596:583–589. DOI: 10.1038/s41586-021-03819-2

[17] Gligorijević V, Barot M, Bonneau R. deepNF: deep network fusion for protein function prediction. *Bioinformatics*. 2018;34(22):3873–3881. DOI: 10.1093/bioinformatics/bty440

[18] Cho H, Berger B, Peng J. Compact integration of multi-network topology for functional analysis of genes. *Nature Methods*. 2016;13(9):733–738. DOI: 10.1038/nmeth.3996

[19] You R, Yao S, Xiong Y, et al. NetGO: improving large-scale protein function prediction with massive network information. *Nucleic Acids Research*. 2019;47(W1):W379–W387. DOI: 10.1093/nar/gkz388

[20] The Gene Ontology Consortium. The Gene Ontology knowledgebase in 2023. *Genetics*. 2023;224(1):iyad031. DOI: 10.1093/genetics/iyad031

[21] Li R, Xie W, Shang Z, et al. MAGIN-GO: Protein function prediction based on dual graph neural networks and gene ontology structure. *PLOS ONE*. 2026;21(2):e0342072. DOI: 10.1371/journal.pone.0342072

[22] Nguyen TT, Jiang Z, Nguyen VN, Le NQK, Chua MCH. Integrating ESM-2 and Graph Neural Networks with AlphaFold-2 Structures for Enhanced Protein Function Prediction. *ACS Omega*. 2025;10(33):38103–38111. DOI: 10.1021/acsomega.5c05484

[23] De Paolis Kaluza MC, Ramola R, Joshi P, et al. Advances in Protein Function Prediction from the Fifth CAFA Challenge. *bioRxiv*. 2026. Preprint. DOI: 10.1101/2026.04.27.716980

[24] Wang Y, Sun Y, Lin B, et al. SEGT-GO: a graph transformer method based on PPI serialization and explanatory artificial intelligence for protein function prediction. *BMC Bioinformatics*. 2025;26:46. DOI: 10.1186/s12859-025-06059-7

[25] Armah-Sekum RE, Szedmak S, Rousu J. Protein function prediction through multi-view multi-label latent tensor reconstruction. *BMC Bioinformatics*. 2024;25:174. DOI: 10.1186/s12859-024-05789-4

[26] Jones CE, Baumann U, Brown AL. Automated methods of predicting the function of biological sequences using GO and BLAST. *BMC Bioinformatics*. 2005;6:272. DOI: 10.1186/1471-2105-6-272

Declarations

Funding

This work was supported by the National Social Science Fund of China under Grant No. 25CSH113.

Data availability

The processed train/validation/test split files, GO term lists, IA weights, evaluation outputs, and training logs supporting the conclusions of this article are available in the accompanying supplementary material (Supplied_material1/). The CAFA6 training data is available from the Kaggle competition page [14]. Pre-trained ESM2 models are available from https://huggingface.co/facebook/esm2_t30_150M_UR50D. The Gene Ontology OBO file is available from http://purl.obolibrary.org/obo/go.obo. A public Zenodo archive with a permanent DOI will be created upon acceptance.

Code availability

Experiment scripts, analysis pipelines, and figure generation code are included in the supplementary material and will be deposited in a public repository with a permanent DOI upon acceptance.

Author contributions

Y.C. conceived the study, designed experiments, and wrote the manuscript. X.W. and S.Z. contributed to experimental infrastructure and analysis. Y.G. contributed to data preparation and evaluation. All authors reviewed and approved the manuscript.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Acknowledgements

We thank the CAFA6 organizers for providing the competition dataset and evaluation framework. Computational resources were provided by an NVIDIA RTX 4090 GPU server.