|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Date** | **Method/Paper/Group** | **Features** | **Approach/Architecture** | **dataset** | **Seq/Structure-based** | **Labels/GO terms** | **Availability** | **Eva** |
| July 07, 2022  (**Recent)** | Integrating Self-Attention Transformer with Triplet Neural Networks for Protein Gene Ontology Prediction  Yi-Heng Zhu, Chengxin Zhang, Dong-Jun Yu, **Yang Zhang** | Input sequence->ESM-1b | transformer, neural network-based feature fusion, and triplet network-based function prediction | 123,774 (uniprot) for training,1089 and CAFA for testing | Seq | Go(MF, BP, and CC) multilabel (no hierarchical) | <https://zhanggroup.org/ATGO/>  (server works)  preprint: <https://www.biorxiv.org/>  content/10.1101/2022.07.07.499156v1.full | F1 (Fmax), AUPR |
| 27 July 2019 | **DeepGOPlus**: improved protein function prediction from sequence  By Maxat Kulmanov, Robert Hoehndorf | Input sequence->one hot encoding | 1D cnn+maxpooling | Train: MFO( 34488) BPO( 51716) CCO( 49346 )  Test: MFO( 679) BPO(1434)CCO(1148) | Seq | Go(MF:652, BP:3904, and CC:545) multilabel (no hierarchical)  [classes annotated with < 50 discarded | Tool: <http://deepgoplus.bio2vec.net/>  (server works)  Paper:  https://doi.org/10.1093/bioinformatics/btz595 | F1 (Fmax), AUPR  Smin |
| 15 July 2018 | **GOLabeler**: improving sequence-based large-scale protein function prediction by learning to rank | GO term frequenc, LR 3-mer, InterProScan, ProFET, blast KNN | Linear regression, Xgboost for GO notation ranking | Train: all data annotated before 2014,LTR1: No knowledge, LTR2: limited knowledge,  Test: jan-oct 2016 | Seq | Go(MF, BP, and CC) | Paper: <https://doi.org/10.1093/bioinformatics/bty130>  Tool: <http://datamining-iip.fudan.edu.cn/golabeler>  (failed to load) | F1 (Fmax), AUPR  Smin |
| Jan 2022, (**recent**) | Accurate protein function prediction via graph attention networks with predicted structure information  By Boqiao Lai, Jinbo Xu (**GAT-GO**) | Contact, esm 1b, 1d features aa, ss, sa,.. | 1D CNN, GAT | DEEPFRI dataset | Seq+Structural info | Go(MF, BP, and CC) | Paper: https://doi.org/10.1093/bib/bbab502 | F1 (Fmax), AUPR |
| 26 May 2021 (**recent**) | Structure-based protein function prediction using graph convolutional networks  By Vladimir Gligorijević et al  (**DEEPFRI**) | Structural info, contact map extracted from native | LSTM (1D features) + GCN (2D features) | combine ~30 k non-redundant experimental structures from the PDB and ~220 k non-redundant homology models from the SWISS-MODEL repository 80/10/10% | Seq/Structure | Go(MF, BP, and CC) and EC numbers, Residue-level site-specific annotation | Paper: <https://www.nature.com/articles/s41467-021-23303-9>  Webserver: <https://beta.deepfri.flatironinstitute.org/workspace/06D52Y>  Code: https://github.com/flatironinstitute/DeepFRI  (Server works) | F1 (Fmax),ROC |
| 23 mar 2021 (**recent**) | **TALE**: Transformer-based protein function Annotation with joint sequence–Label Embedding  By Yue Cao, Yang Shen | Seq embedding, Label embedding | Transformer encoder+cnn 1D | UniProtKB/Swiss-Prot dataset release-2015\_12  CAFA: before September 2016 and between September 2016 and November 2017 | Seq | Go(MF, BP, and CC) | Paper: <https://doi.org/10.1093/bioinformatics/btab198>  Code: <https://github.com/Shen-Lab/TALE> (locally installed) | F1 (Fmax), |
| 2 July 2021  (**recent**) | **NetGO 2.0**: improving large-scale protein function prediction with massive sequence, text, domain, family and network information  [[CAFA3 champion]] | Naive, BLAST-KNN, LR-3mer, LR-InterPro, NetKNN, LR-Text and Seq-RNN , PPI | Ranking, improved version of GOLABELER | All data annotated in December 2018 or before, LTR: January 2019 to January 2020,test: between February 2020 and October 2020 | Seq, PPI | Go(MF, BP, and CC) | Paper: <https://academic.oup.com/nar/article/49/W1/W469/6285266>  Server: <https://issubmission.sjtu.edu.cn/ng2/>  (says error in input fasta format) | F1 (Fmax), AUPR, Coverage |
| 12 July 2021 (**recent**) | **DeepGraphGO**: graph neural network for large-scale, multispecies protein function prediction by Ronghui You, Shuwei Yao, Hiroshi Mamitsuka, Shanfeng Zhu | N binary feature vector by InterProScan, protein network G | FCN, 2GCN, FCN | Train: before Jan 2018,  Val:jan-dec 2018  Test: jan 19-jan20 | Seq, higher order protein network | Go(MF, BP, and CC) | Paper: https://doi.org/10.1093/bioinformatics/btab270  code: <https://github.com/yourh/DeepGraphGO>  ((cannot be installed because of dependency issues)) | F1 (Fmax), AUPR |
| July 2022  (**recent**) | **DeepGOZero**: improving protein function prediction from sequence and zero-shot learning based on ontology axioms  By Maxat Kulmanov, Robert Hoehndorf | Protein embedded by MLP and GO terms embedded with ELEMbedding | Zero-shot prediction | Train: before dec 2018  Val: jan 2019-jan 2020  Test: feb 2020, oct 2020 | Seq | Go(MF, BP, and CC) | Paper: <https://doi.org/10.1093/bioinformatics/btac256>  Code: <http://github.com/bio-ontology-research-group/deepgozero>.  <<cannot be installed because requirements cannot be installed ( Could not find a version that satisfies the requirement pkg\_resources)>> | F1 (Fmax), AUPR  Smin, (for class-centric metric: AUROC) |
| May 2019 | **DEEPred:** Automated Protein Function Prediction with Multi-task Feed-forward Deep Neural Networks | Seq features vector by (i) Conjoint triad, (ii) Pseudo amino acid composition and (iii) Subsequence profile map (SPMap) | Multiple ( 1,101 different models) MLPs to predict 4-5 terms each, and hierarchical post processing | UniProtKB/Swiss-Prot database (version 2017\_08) | Seq | Go(MF, BP, and CC) | Code: <https://github.com/cansyl/DEEPred>  Paper: <https://www.nature.com/articles/s41598-019-43708-3> | F1, AUC |
| Jan 2021 | Unsupervised protein embeddings outperform hand-crafted sequence and structure features at predicting molecular function by Amelia et al | ElMo 1024, seq one hot (26), k-mer(1024), contact map, DeepFold | MLP, GCN |  | Seq | Go(**MF**, BP, and CC) | Paper: <https://doi.org/10.1093/bioinformatics/btaa701>  Code:  https://github.com/stamakro/GCN-for-Structure-and-Function  <<cannot be installed, dependency packages are conflicting to each other, features cannot be used>> | F1: shows adding structural information cannot improve prediction |
| June 2022 | **ContactPFP:** Protein Function Prediction Using Predicted Contact Information by kihara lab | Msa->contact->gr-align | Gr-align |  | Seq/msa | Go(**MF**, BP, and CC) | Code: <https://github.com/kiharalab/contactpfp>,  Paper: https://www.frontiersin.org/articles/10.3389/fbinf.2022.896295/full | Fmax,smin |