


MMSMAPlus: a multi-view multi-scale multi-attention embedding model for protein function prediction

Zhongyu Wang, Zhaohong Deng , Wei Zhang, Qiongdan Lou, Kup-Sze Choi, Zhisheng Wei, Lei Wang and Jing Wu

Corresponding author: Zhaohong Deng, Jiangnan University, Wuxi, Jiangsu 214012, China. Tel: +86-13771571629; Fax: 0510-85327307;

E-mail: dengzhaohong@jiangnan.edu.cn

Abstract

Protein is the most important component in organisms and plays an indispensable role in life activities. In recent years, a large number of intelligent methods have been proposed to predict protein function. These methods obtain different types of protein information, including sequence, structure and interaction network. Among them, protein sequences have gained significant attention where methods are investigated to extract the information from different views of features. However, how to fully exploit the views for effective protein sequence analysis remains a challenge. In this regard, we propose a multi-view, multi-scale and multi-attention deep neural model (MMSMA) for protein function prediction. First, MMSMA extracts multi-view features from protein sequences, including one-hot encoding features, evolutionary information features, deep semantic features and overlapping property features based on physiochemistry. Second, a specific multi-scale multi-attention deep network model (MSMA) is built for each view to realize the deep feature learning and preliminary classification. In MSMA, both multi-scale local patterns and long-range dependence from protein sequences can be captured. Third, a multi-view adaptive decision mechanism is developed to make a comprehensive decision based on the classification results of all the views. To further improve the prediction performance, an extended version of MMSMA, MMSMAPlus, is proposed to integrate homology-based protein prediction under the framework of multi-view deep neural model. Experimental results show that the MMSMAPlus has promising performance and is significantly superior to the state-of-the-art methods. The source code can be found at <https://github.com/wzy-2020/MMSMAPlus>.

Keywords: protein function prediction, multi-view deep feature learning, multi-scale, attention mechanism

INTRODUCTION

Protein is the most important building block in organisms and plays an essential role in various life activities. Protein function describes the role of protein in biochemical reactions, cellular activities, biological expressions and other life activities [1]. The study of protein function is helpful to understand the molecular mechanisms of various life activities in organisms and is of great importance to the research of physiology, pathology and pharmaceutical science.

In the post-genomic era, many protein databases have been available. However, there are still a great number of proteins that have no functional annotations [2]. For these proteins with unknown functions, homology-based transfer is often adopted for function prediction in early methods [3], such as BLAST [4],

HHblits [5] and Diamond [6]. These methods perform homology alignment in databases with known functional sequences by evaluating the similarity between unknown and known functional proteins, so that the function of a known protein is transferred to an unknown but highly similar protein. Homology-based transfer is based on the biological principle that if two protein sequences are highly similar, they are likely to have evolved from a common ancestor, and thus they have similar functions. However, the literature indicates that homologous transfer based on the similarity between protein sequences is not very reliable [1].

Benefiting from the development of intelligent modeling with machine learning, new protein function prediction methods have been proposed in recent years. For example, Loblely *et al.* [7] proposed the FFPred method, by using support vector machine

Zhongyu Wang is a master student in the School of Artificial Intelligence and Computer Science, Jiangnan University. His research interests include text data mining, natural language processing and their applications in bioinformatics.

Zhaohong Deng is a full-time professor in the School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China. His research interests include bioinformatics and artificial intelligence.

Wei Zhang is currently pursuing the Ph.D. degree in the School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China. His research interests include artificial intelligence and data mining.

Qiongdan Lou is currently pursuing the Ph.D. degree in the School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China. Her research interests include artificial intelligence and data mining.

Kup-Sze Choi is a full-time professor at the Hong Kong Polytechnic University, Hongkong. His research interests include pattern recognition, data mining and smart health.

Zhisheng Wei is currently pursuing the Ph.D. degree in the National Key Laboratory of Food Science and Resource Mining, Jiangnan University, Wuxi, China. His research interests include enzyme engineering and intelligent mining of enzyme data.

Lei Wang is an assistant researcher in the National Key Laboratory of Food Science and Resource Mining, Jiangnan University, Wuxi, China. His research interests include enzyme fermentation engineering and food science.

Jing Wu is a full-time professor in the National Key Laboratory of Food Science and Resource Mining, Jiangnan University, Wuxi, China. Her research interests include the protein engineering of enzyme preparations and the genetic engineering of metabolic engineering strains.

Received: December 29, 2022. Revised: April 16, 2023. Accepted: May 8, 2023

© The Author(s) 2023. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

to predict amino acid sequences, which is a potential substitute for homology-based annotation method. Cozzetto et al. [8] further proposed the FFPred3 method to predict protein function by using feature information such as secondary structure, transmembrane helices, intrinsic disordered regions and signal peptides. Kulmanov et al. [9] proposed the DeepGO method, which combines sequence motifs with protein–protein interaction network to achieve function prediction. You et al. [10] proposed the DeepText2GO method to improve the automated function prediction (AFP) performance by incorporating text mining information alongside protein sequence information. In addition, the consensus mechanism in DeepText2GO is used to combine the text-based classifier with the sequence-based classifier. Gligorijevic et al. [11] proposed the DeepFRI method to learn the sequence and structure binding preferences from experimental data. The method uses the deep language model LSTM-LM [12] to extract protein sequence information and inputs the protein structure information and sequence information into the graph convolutional network to obtain complex structure–function relationships. You et al. [13] proposed the DeepGraphGO, which combines sequence information and protein network information and uses graph convolutional neural networks to obtain higher-order network information.

The above intelligent methods have demonstrated effectiveness in protein function prediction, but most of them use experimental information (such as interaction networks) other than protein sequence, which has a limited ability to express large-scale data and is more expensive to obtain than sequencing. Furthermore, for proteins whose function is unknown, it is easiest to obtain sequence features. Therefore, sequence-based prediction method is still a major research in AFP. Hence, the protein function prediction discussed in this paper focuses on sequence features.

Research has shown that combining different feature sets extracted from the same data source can achieve information complementarity and performance improvement [14, 15]. That is, it is reasonable and effective to extract multiple feature views from amino acid sequences and use these views to construct prediction models. Yet, it is critical to determine the importance of the views.

In response to the above challenges, we conduct in-depth research and propose a multi-view, multi-scale and multi-attention-based deep neural model (MMSMA) to predict the protein function. The main idea is as follows: (1) we construct multi-view features for protein sequences from the view of one-hot encoding information, evolutionary information, deep semantic information and overlapping property based on physiochemistry. (2) Based on the multi-view data, we design a specific multi-scale multi-attention (MSMA) deep network model for each view to learn deep features and conduct the preliminary prediction for protein function. In each MSMA, a multi-scale deep feature extractor (MSFE) with a feature pyramid structure is designed to capture multi-scale local features, and a multi-head attention (MHA) mechanism is adopted to capture long-range dependence between multi-scale local features. (3) Based on the preliminary predictions of multiple views, we introduce a multi-view adaptive decision mechanism to balance the impact of the multiple views on the final prediction results. (4) We extend MMSMA by combining it with homology-based method to propose the MMSMAPlus to further improve the prediction performance of protein function. The contributions of this study are summarized as follows:

- (i) We investigate feature extraction techniques suitable for protein function prediction from four views, including one-hot encoding information, evolutionary information, deep semantic information and overlapping property information.
- (ii) We design MSMA to extract deep features from different views and obtain the preliminary protein function predictions. For each MSMA, an MSFE with a feature pyramid structure is designed to capture local features, and an MHA mechanism is adopted to capture the long-range dependence between local features.
- (iii) We present a multi-view adaptive decision mechanism to make a comprehensive decision based on the classification results of all the views.
- (iv) We further propose an extended version of MMSMA, MMSMAPlus, to integrate homology-based protein prediction under the framework of multi-view deep neural model.
- (v) We conduct comprehensive experimental evaluations and show that the methods proposed in this study can achieve excellent performance in protein function prediction.

MATERIALS AND METHODS

Overview

The framework of the proposed MMSMA is shown in Figure 1. MMSMA mainly consists of three modules: the multi-view feature extraction module, the deep feature extractor and sub-classifier learning module, and the adaptive decision module, which are briefly described as follows: for the multi-view feature extraction module, one-hot encoding features are extracted from the amino acid sequence, evolutionary information features are represented through position-specific scoring matrices (PSSM), deep semantic features are obtained through language model and overlapping property features are captured from the amino acid sequence. For the deep feature extractor and sub-classifier learning module, we propose an MSMA deep neural network for each single view. The proposed MSMA performs deep feature extraction and preliminary classification for four views, respectively. Each view can generate a specific preliminary prediction for protein function. Finally, for the adaptive decision module, we utilize the multi-view adaptive decision mechanism to realize joint decision of the four views and obtain the final prediction result of protein function.

Benchmark data set

We obtained the gene ontology (GO) data (February 2021) from the GO official website (<http://geneontology.org/>). The data have 44 085 terms in three branches, including 11 153 terms in MFO, 28 748 terms in BPO and 4184 terms in CCO. In this paper, we collected the reviewed and manually annotated human proteome sequences from SwissProt (<http://www.uniprot.org/uniprot/>) [16], which contains 18 673 protein sequences. For comparison with other function prediction methods, we also obtained data sets of the training sequences, experimental annotations (published before September 2016) and the test benchmarks (published on 15 November 2017; <https://github.com/bio-ontology-research-group/deepgoplus>) from the Critical Assessment of Functional Annotation Challenge Three (CAFA3). The GO glossary released on 1 June 2016 was used in this paper to evaluate methods.

For each sub-ontology in GO, we first learn the structure knowledge of GO. In particular, we follow the true path rule [17] to propagate annotations. For example, if a protein *P* is annotated with the GO term *C*, then *P* will be annotated by the ancestor

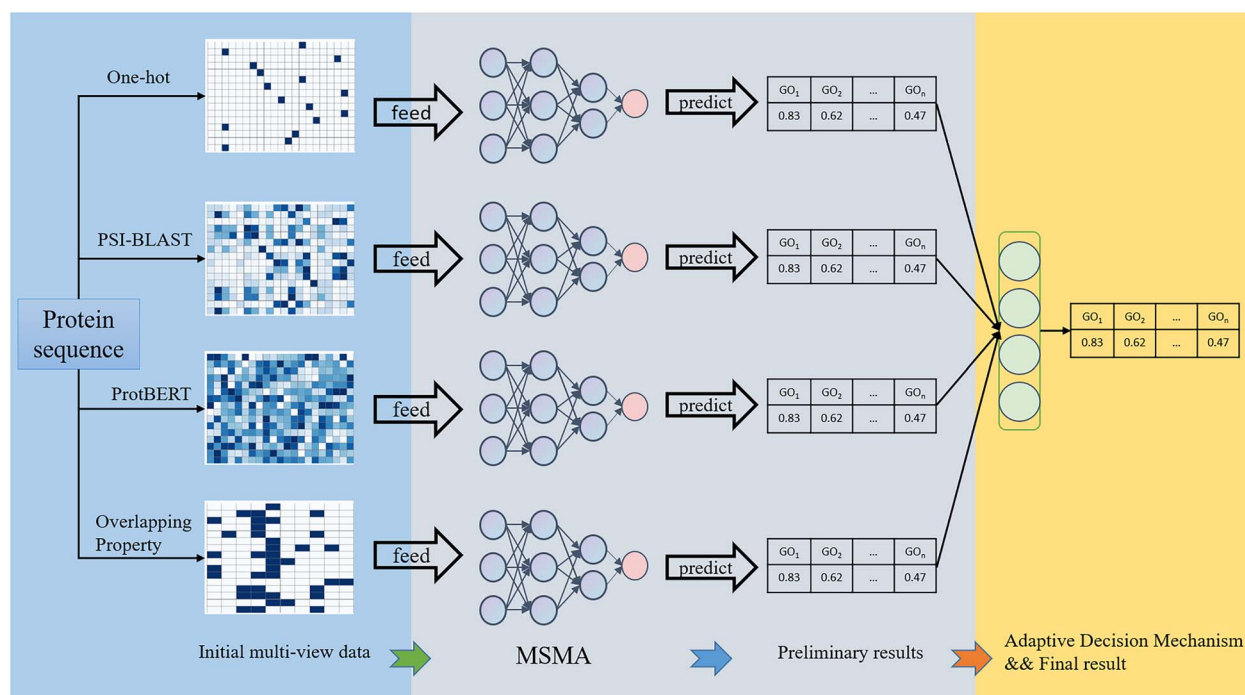


Figure 1. Overview of MMSMA for predicting protein functions. The model consists of three modules: (i) the initial multi-view data construction module is used to generate four kinds of sequence encoding features. (ii) The deep feature learning and sub-classifier construction module uses MSMA for each view to obtain deep features and preliminary classification results. (iii) The adaptive decision module implements integrated decisions to produce final prediction scores for each GO term.

terms of C . Then, we rank the GO terms according to the number of annotations and select the terms with 50 or more annotations for the proposed prediction model. The adopted cutoff values are the same as that in DeepGOPlus [18]. We convert the protein annotations into a binary label vector. If a protein sequence is annotated with a GO term, we will assign 1 to the term position in the binary vector. Otherwise, we will assign 0.

Multi-view features of protein sequences

In this section, protein sequences are extracted from four views, i.e. one-hot encoding information, evolutionary information, deep semantic information and overlapping property information. The feature extraction process of these four views is as follows.

One-hot encoding features

One-hot encoding has been widely used in protein function prediction [18, 19]. A protein sequence can be represented as:

$$P = [p_1, \dots, p_l, \dots, p_L], p_l \in \mathbb{R}^{21 \times 1} \quad (1 < l < L) \quad (1)$$

where p_l represents the l th residue in protein P . L is the length of protein P . We encode the residues in protein P one by one. Specifically, the amino acid sequence is listed and corresponds to a 21D vector. In each residue, if an amino acid appears, its corresponding position is assigned 1, and 0 otherwise. Therefore, a one-hot encoding matrix with the size of $21 \times L$ can be obtained.

Evolutionary information features

PSSMs are commonly used to represent patterns in proteins [4, 20, 21]. In this paper, the PSI-BLAST algorithm is used to align each target protein with the SwissProt database. The number of iterations is set to 3, and the inclusion e -value is set to 0.001. And

then 20 scores can be obtained for each amino acid (corresponding to 20 outputs of PSI-BLAST). Therefore, for each sequence of length L , the size of PSSMs is $20 \times L$. Furthermore, the sigmoid function is used to normalize each element in PSSMs into the interval $[0, 1]$ [22], that is,

$$\tilde{x} = \frac{1}{1 + e^{-x}} \quad (2)$$

Deep semantic features

Proteins are composed of different types of amino acids, and different amino acid fragments usually have different biological functions [23]. A simple way to numerically represent amino acid sequences is one-hot encoding. However, because of the sparsity, one-hot encoding features cannot reflect the relationship between amino acids. Distributed representation, using dense vectors to represent sentences, can describe the semantic distance between words more effectively and has therefore gained momentum in the field of natural language processing [24].

According to the above analysis, we use the deep language model ProtBERT [25] to extract deep semantic information from amino acid sequences. Based on the bidirectional encoder representations from transformers (BERT) model [26], ProtBERT increases the number of layers and completes the pre-training on the UniRef100 data set. Unlike the CNN-based model and the LSTM-based model SeqVec [27], ProtBERT uses the self-attention mechanism to compare each character in the current sequence with other parallel sequence characters; therefore, it has a global receptive field to capture global context information more effectively.

The specific extraction process of deep semantic features is as follows: we fine-tune the ProtBERT model on the CAFA3 data set and use the fine-tuned model to extract deep semantic

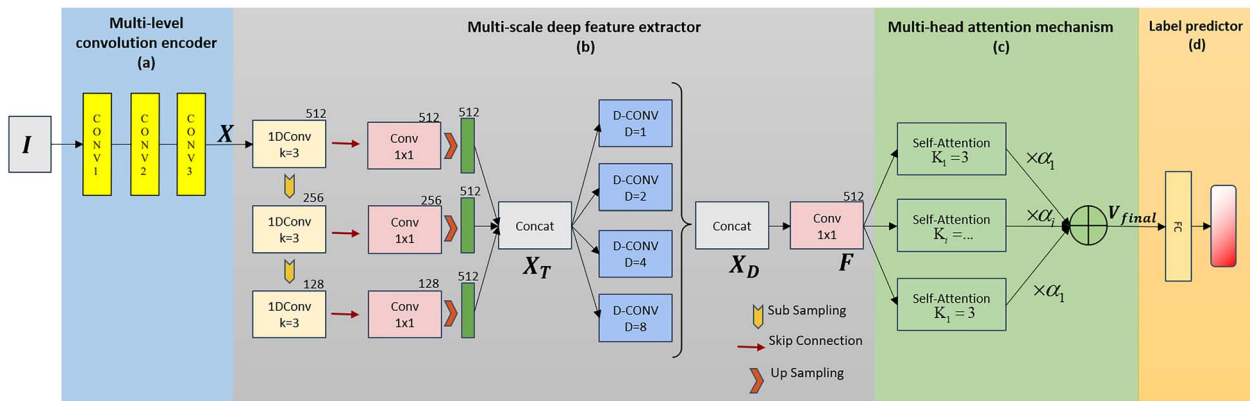


Figure 2. The architecture of MSMA. An initial sequence feature I is first sent to a multi-level convolution encoder to get a feature matrix X . Based on the feature matrix X , an MSFE is adopted to construct a feature pyramid structure to obtain multi-scale feature matrix X_T . Then four dilation convolution layers are built to extract high-order multi-scale information from matrix X_T . A 1×1 convolution layer is applied to fuse multi-scale information and output feature matrix F . Furthermore, the MHA mechanism is used to extract the long-range dependence for the multi-scale feature F , and the compact embeddings V_{final} is obtained. Finally, a label predictor is adopted to produce prediction scores for each GO term.

features that are valuable for downstream tasks. For each sequence with length L , fine-tuned ProtBERT is first used to extract semantic-level features. Principal component analysis (PCA) [28] is employed to reduce the dimension of semantic-level features. The experimental setting of the percentage of PCA is analyzed in part D of the Supplementary Material section. Finally, a feature matrix with the size of $292 \times L$ is obtained by keeping 95% of the principal components.

Overlapping property features

According to the physicochemical property, amino acids with common properties can be grouped into 10 groups, including 'Polar' (NQSDCKTRHYW), 'Positive' (KHR), 'Negative' (DE), 'Charged' (KHRDE), 'Hydrophobic' (AGCTIVLKHFYWM), 'Aliphatic' (IVL), 'Aromatic' (FYWH), 'Small' (PNDTGAGSV), 'Tiny' (ASGC) and 'Proline' (P) [29, 30]. In general, an amino acid may have several physicochemical properties at the same time. For example, a residue can be related to 'Hydrophobic', 'Aliphatic', 'Small' and 'Tiny' simultaneously.

According to the above analysis, we design a 10D vector consisting of 0 or 1 to represent the physicochemical properties of each amino acid. The 10 physicochemical properties correspond to the elements in the 10D vector. If an amino acid has a certain property, the corresponding position in the vector is set to 1, and 0 otherwise. For the amino acid sequence with the length L , the feature matrix with the size of $10 \times L$ can be obtained.

Deep sub-classifier learning with the multi-scale and multi-attention network

Model structure

In MMSMA, we use MSMA to obtain the preliminary predictions of the multiple views. The MSMA structure is shown in Figure 2. First, a multi-level convolutional encoder (MLCE) is applied, where multiple convolutional layers are cascaded one after another. Because of the small convolution scale of MLCE, local pattern information can be obtained. Second, an MSFE is proposed by combining MLCE block with the feature pyramid structure. The proposed MSFE can maintain the scale invariance. Third, since protein function may be influenced by long-range information [31], local patterns cannot capture enough long-range dependence between protein sequences. Therefore, an MHA mechanism

is introduced to capture not only sufficient local patterns, but also long-range dependence in protein sequences. More details of MSMA are described as follows.

Multi-level convolution encoder

In order to generate efficient representations for functional fragments, we utilize a multi-layer convolutional neural network to capture the local correlation between amino acids from each view. Specifically, the structure of MLCE is a three-layer 1D convolution, which is based on the sequential correlation to establish the local patterns. It should be noted that the number of channels in 1D convolution is equal to the number of units in hidden layer, so that the information in each dimension of the representation vector will not be destroyed [32]. In addition, convolution padding is not used because we capture functional fragments in amino acid sequences rather than long-range dependence.

Based on the initial embeddings for each view, we design MLCE to obtain more discriminative feature representations by extracting local features. First, these embeddings are horizontally concatenated into the matrix $I = [x_1, \dots, x_i, \dots, x_L] \in \mathbb{R}^{d_e \times L}$, where L is the length of amino acid sequence and $x_i \in \mathbb{R}^{d_e \times 1}$ ($1 < i < L$) represents the feature vector of the i th amino acid in the sequence. Adjacent amino acid fragments are combined through a convolutional filter $W_c \in \mathbb{R}^{k \times d_e \times d_c}$, where k is the filter width, d_e is the size of the initial input and d_c is the size of the filter output. For the n th step, we have

$$h_n = g(W_c * x_{n:n+k-1} + b_c) \quad (3)$$

where $*$ is the convolution operator, g is an element-wise nonlinear transformation, $b_c \in \mathbb{R}^{1 \times d_c}$ is the bias and h_n is the output of the n th convolution step. Therefore, the output embeddings of the encoding layer are expressed as $X \in \mathbb{R}^{d_c \times L}$.

Multi-scale deep feature extractor

Learning local features of protein sequences is helpful to improve the performance of protein function prediction [23]. Stacking multiple convolutional filters can obtain more local features [9, 18, 23], but they increase the computational complexity and generate redundancy. Inspired by FPN [33], AugFPN [34] and FastFCN [35], we extract multi-scale features by constructing a feature pyramid structure, which has both lower computation cost and

better feature extraction ability. Different from stacking convolutional filters [9, 18, 23], progressively finer semantic features can be obtained by downsampling based on pyramid structure. The above work has a common advantage, i.e. the scale invariance of the semantic information, which means that the semantic information is consistent but has different degrees of detailed information at different scales [36]. Motivated by the above characteristics, we propose an MSFE that continuously enhances the semantic information related to the functional prediction task through a series of scale transformations. Finally, we merge the deep features at different scales so that the final multi-scale features contain as much semantic information as possible.

The learning process of the proposed MSFE is described below.

For the amino acid feature matrix $X \in \mathbb{R}^{d_c \times L}$ obtained by MLCE, we adopt a top-down path to construct pyramid features and obtain the outputs C_1, C_2 and C_3 successively. C_3 from the deepest layer has the strongest semantic information.

To fuse the multi-scale context information, a simple and crude way is to summarize the features under different scales in the top-down path. However, the features under different scales are different in semantics, and reducing the channel dimension leads to information loss [34].

Hence, we propose an MSFE. In order to reduce the semantic difference between different scale features, the proposed MSFE performs the feature processing on multi-scale features before feature fusion. First, for multi-scale features $\{C_1, C_2, C_3\}, \dots$, we upsample them to the same scale through 1×1 convolution to obtain the feature pyramid $\{M_1, M_2, M_3\}$. Second, the matrix $X_T \in \mathbb{R}^{3d_c \times L}$ can be obtained by concatenating $M_1 M_2$ and M_3 . To widen the receptive field, four separable convolutions with different dilation rates are employed in parallel to extract features from X_T and concatenate these features to obtain $X_D \in \mathbb{R}^{4d_c \times L}$. Different dilation rates have different functions. Third, another regular 1×1 convolution block is used to transform X_D into the final multi-scale deep feature $F \in \mathbb{R}^{d_c \times L}$.

MHA mechanism

In a protein sequence, long-range dependence between residues affects protein function [31]. Therefore, we use a self-attention mechanism to establish long-range dependence between protein sequences after obtaining multi-scale deep features $F \in \mathbb{R}^{d_c \times L}$. As shown in Figure 3, the self-attention is designed to reweight each channel according to the interaction of local cross-channel [37, 38]. First, the average-pooling and max-pooling operations are used to collect the global information of a feature map, generating two different spatial context descriptors F_{avg}^c and F_{max}^c . F_{avg}^c and F_{max}^c represent the average-pooled features and max-pooled features, respectively. Second, a shared network is utilized to generate the weight vectors $M_\ell \in \mathbb{R}^{d_c \times 1}$. The shared network is implemented by a 1D convolution with kernel size k , where k is the number of neighbors of the current channel. The specific formula of self-attention is as follows:

$$M_\ell = \sigma \left(C1D_k \left(F_{avg}^c \right) + C1D_k \left(F_{max}^c \right) \right) \quad (4)$$

where σ is the sigmoid function and C1D represents the 1D convolution. In general, it is difficult to adjust k because different GO terms require different numbers of neighbors. Therefore, we further extend this attention with a multi-head trick [39], that is, multiple self-attention branches.

Specifically, each head utilizes a different kernel size k . The number of attention heads is denoted as h . To avoid adjusting the kernel size k , the strategy we adopt is either to choose a single

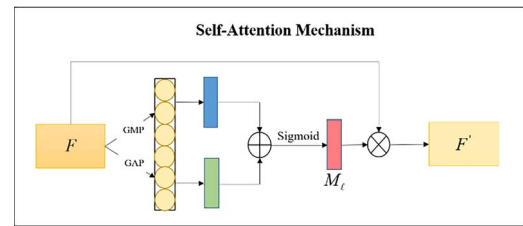


Figure 3. The proposed self-attention mechanism. First, global information from the multi-scale feature F is aggregated using global average (GAP) and global maximum (GMP) pooling. Then, a shared convolutional layer is applied to excite the global information and fuse them by summation. Furthermore, the attention score matrix M is obtained via the sigmoid function. Finally, the dot-product of F and M is calculated to obtain the attention-weighted features F' .

head ($h = 1$) with fixed $k = 3$, or to use MHA ($h > 1$) with fixed sequences of kernel sizes k_1, k_2, \dots, k_n . In addition to $h = 1$, we also set $h = 2, 4, 6, 8$. The details are as follows:

When $h = 2, k = 3, 5$.

When $h = 4, k = 3, 5, 7, 9$.

When $h = 6, k = 3, 5, \dots, 13$.

When $h = 8, k = 3, 5, \dots, 17$.

When $h > 1$, the kernel size k is set in an increasing order. Diversity is introduced into the branches with different k values to generate better long-range dependence information.

Because of the overlapping effect caused by the similarity between different heads, we use adaptive weighted fusion (AWF) to adaptively combine the long-range information, instead of simple summation. For each head, the max-pooling is executed first to obtain long-range information representations $V_{k_1}, V_{k_2}, \dots, V_{k_n}$ ($V_{k_i} \in \mathbb{R}^{d_c \times 1}$). These long-range representations are then fed into the AWF to generate the weight map, which can be used to aggregate the context features $V_{final} \in \mathbb{R}^{d_c \times 1}$. Context features V_{final} containing long-range dependence are calculated as follows:

$$V_{final} = \sum_{i=1}^h \alpha_i \cdot V_{k_i} \quad (5)$$

where α_i is the weight of the i th head. k_i is the number of neighbors channels of the i th head. Equation (5) indicates that V_{final} is a global context view that performs a weighted sum of different context information. It is worth noting that Equation (5) can be used to adaptively aggregate contexts based on different neighbors. In fact, the information gain is achieved through similar semantic features to improve the compactness and consistency of features in the channel.

Label (go term) predictor

Based on the comprehensive feature representation, the GO term classifier is constructed through the multilayer perceptron with one hidden layer. The predicted probability of each label is estimated as follows:

$$\hat{y} = \text{Sigmoid} \left(f \left(W \left(V_{final} \right) \right) \right) \quad (6)$$

where W is the weight of the fully connected layer. f is the nonlinear activation function $ReLU$. The sigmoid function is used to convert the output value into the probability. Cross-entropy loss is used in this paper, because it has been proved to be suitable for protein function prediction [9, 18, 19]. Therefore, the loss function

is defined as follows [40]:

$$L = - \sum_{i=1}^N \sum_{j=1}^l (y_{ij} \log(\hat{y}_{ij}) + (1 - y_{ij}) \log(1 - \hat{y}_{ij})) \quad (7)$$

where N is the number of training sequences, l is the number of GO terms, $\hat{y}_{ij} \in [0, 1]$ is the predicted probability and $y_{ij} \in \{0, 1\}$ indicates the ground truth of the i th sequence along the j th GO term.

Multi-view adaptive decision mechanism

With the four views extracted through different representation theories, it is necessary to combine their preliminary predictions to make a comprehensive decision.

According to the above analysis, a multi-view adaptive loss-weighted fusion network is constructed based on the proposed AWF. The final prediction results of the proposed MMSMA can be obtained by the joint decision of multiple views. The following equation defines the comprehensive decision of multiple views, that is,

$$\hat{f} = \sum_{v=1}^M w_v \hat{y}_v \text{ s.t. } w^T \mathbf{1} = 1, w \geq 0 \quad (8)$$

where M is the number of views, $w_v \in \mathbb{R}$ is the weight of the v th view, \hat{y}_v is the preliminary prediction results of the v th view and \hat{f} is the comprehensive prediction result. Finally, we use the cross-entropy loss function (Equation (7)) to optimize the comprehensive prediction result, which is given by,

$$\Phi = L(\hat{f}, y) \quad (9)$$

MMSMAPlus: the extension of MMSMA

The proposed MMSMA is built based on data-driven learning. To further improve the performance of protein function prediction, we combine MMSMA with the homology-based prediction method and propose an extended version of MMSMA, i.e. MMSMAPlus. The specific process is as follows:

First, the Diamond prediction method is utilized to perform homologous alignment for test set in the training set database. The e -value of Diamond is set to 0.001, and a bitscore is calculated for each similar sequence. All annotations of similar sequences are transferred to the query sequence, where bitscores are used to calculate the prediction scores. Second, the two prediction scores S_{MMSMA} and $S_{\text{DiamondScore}}$ are combined to calculate the final prediction score of MMSMAPlus as follows:

$$S_{\text{MMSMAPlus}} = \alpha * S_{\text{MMSMA}} + (1 - \alpha) * S_{\text{DiamondScore}} \quad (0 \leq \alpha \leq 1) \quad (10)$$

where $0 \leq \alpha \leq 1$ is a hyperparameter, which balances the influence of two terms.

Experiments and results

Experimental setting

The proposed methods are validated on the Pytorch platform with GO annotations and amino acids of the Human and CAFA3 protein data sets.

First, we consider proteins with sequence length in the range [0, 2000]. For sequence longer than 2000, we take the first 2000 amino acids. On the contrary, the corresponding protein vectors are zero-padded.

For the Human protein data set, 5-fold cross-validation is used. In each fold, 20% of the training data is randomly selected as the validation set. After annotation propagation, we select 475, 2933 and 446 GO terms in MFO, BPO and CCO for experiments, respectively.

For the CAFA3 data set, the experimental setup follows literature [9, 18]. That is, the training set, testing set and the number of GO terms are all fixed. The training and testing set contain 66 841 and 3328 protein sequences, respectively. The number of terms for MFO, BPO and CCO is 677, 3992 and 551, respectively. In the subsequent experiment, we randomly select 10% of the training data as the validation set.

In the experiments, the proposed neural network model has many hyperparameters, such as the number of outputs for the pyramid structure in MSFE, the number of attention heads in MHA, optimizers and learning rates. In general, all parameters are determined by performance on the validation set. Specifically, we set the number of outputs for the pyramid structure in MSFE to 3, the number of attention heads in MHA to 4 and the initial learning rate of Adam optimizer [41] at 0.0005. Further details about the setting of the number of outputs for the pyramid structure can be found in part E of the Supplementary Material section. In MMSMAPlus, the parameter α is involved to combine MMSMA and homology-based method. The setting of α is analyzed in Part G of the Supplementary Material section. We tune the values of α using the validation sets of MFO, BPO and CCO, and final 0.7, 0.7 and 0.9 are used for the three tasks, respectively.

All the baseline methods are downloaded from the websites provided by authors. The details of the baseline methods are described in part A of the Supplementary Material section. For DeepGOCNN and DeepGOPlus, the codes published on GitHub are used to train the models on the Human data set and the results on the CAFA3 data set in [18] are referred to for comparison.

Evaluation metrics

To evaluate the effectiveness of the proposed method and to compare it with the existing baseline methods, we use CAFA evaluation metrics F_{max} and S_{min} [42, 43], and the area under the precision-recall curve (AUPR) [44] for performance evaluation. The details of each metric are described in part B of the Supplementary Material.

Evaluation and comparison

First, the proposed methods are compared with the baselines Naive [42], DeepGOCNN [18], DiamondBLAST [4], DiamondScore [18], TALE+ [45] and DeepGOPlus [18] on the human protein data set. The results are shown in Table 1. It can be seen from Table 1 that MMSMAPlus achieves the optimal F_{max} , S_{min} and AUPR values in three sub-ontologies. Compared with DeepGOCNN, MMSMAPlus performs multi-view deep learning in addition to homology and deep network. Therefore, the comparison results between DeepGOCNN and MMSMAPlus indicate that the multi-view deep learning is able to make fuller use of the protein sequence information. In addition, we note that TALE+ also achieves much better performance, which uses features extracted from label in addition to sequence-based features.

Second, the large-scale CAFA3 data set is used to demonstrate the universality of the proposed MMSMAPlus. We use the same training set and test set as in the literature [18], so that the performance of the proposed methods can be directly compared with that of the relevant methods. The results in Table 2 show that MMSMAPlus achieves the optimal performance on MFO and BPO in terms of F_{max} and AUPR, and the suboptimal performance on

Table 1. The performance comparison of eight methods on the Human data sets

Method	F_{\max}			S_{\min}			AUPR		
	MFO	BPO	CCO	MFO	BPO	CCO	MFO	BPO	CCO
Naive	0.345	0.378	0.549	19.450	72.932	16.583	0.227	0.301	0.483
DiamondBLAST	0.673	0.548	0.622	12.567	66.522	13.731	0.029	0.041	0.041
DiamondScore	0.681	0.556	0.628	12.165	61.723	13.232	0.137	0.154	0.159
DeepGOCNN	0.494	0.468	0.674	17.108	68.665	14.235	0.481	0.441	0.650
MMSMA	0.680	0.497	0.728	13.165	61.237	12.489	0.678	0.488	0.682
TALE+	0.712	0.609	0.729	11.405	58.512	12.315	0.689	0.587	0.714
DeepGOPlus	0.691	0.588	0.698	11.942	59.729	12.861	0.675	0.572	0.684
MMSMAPlus	0.740	0.612	0.742	11.095	58.087	11.884	0.740	0.626	0.719

Note: Best performance in bold F_{\max} and AUPR, highest; S_{\min} , lowest.

Table 2. The performance comparison of eight methods on the CAFA3 data sets

Method	F_{\max}			S_{\min}			AUPR		
	MFO	BPO	CCO	MFO	BPO	CCO	MFO	BPO	CCO
Naive	0.290	0.357	0.562	10.733	25.026	8.465	0.130	0.254	0.456
DiamondBLAST	0.431	0.399	0.506	10.233	25.320	8.800	0.178	0.116	0.142
DiamondScore	0.509	0.427	0.557	9.031	22.860	8.198	0.340	0.267	0.335
DeepGOCNN ^a	0.420	0.378	0.607	9.711	24.234	8.153	0.355	0.323	0.616
MMSMA	0.583	0.518	0.620	7.914	21.785	7.693	0.541	0.457	0.590
TALE+	0.558	0.480	0.622	8.360	22.549	7.822	0.539	0.427	0.595
DeepGOPlus ^a	0.544	0.469	0.623	8.724	22.573	7.823	0.487	0.404	0.627
MMSMAPlus	0.595	0.535	0.622	7.922	22.202	7.631	0.559	0.470	0.601

Note: The performance of the models with an alphabet (DeepGOCNN and DeepGOPlus) was taken from the related literature (Ref. [18] in main text). Best performance in bold F_{\max} and AUPR, highest; S_{\min} , lowest.

CCO in terms of F_{\max} . The proposed deep network model MMSMA of MMSMAPlus achieves the best performance in terms of S_{\min} on MFO and BPO. The S_{\min} evaluation depends on the number of false negatives, false positives and the information content of GO classes [18]. It indicates that MMSMA method is more specific in false-positive predictions.

In addition, we also evaluate the AuROC of the proposed methods on the Human and CAFA3 data sets, and the results are detailed in part C of the Supplementary Material section. It can be seen that MMSMAPlus achieves the best class-centric average AuROC in MFO and CCO evaluation and is ranked second in BPO on the Human data set. In summary, the experimental results show that the proposed method significantly improves the prediction performance of Human and CAFA3 data sets.

Performance analysis of a multi-view adaptive decision mechanism

We verify the effectiveness of the multi-view adaptive decision mechanism employed in this paper by comparing it with four single-view versions of MSMA, namely MSMA_Onehot, MSMA_PSSM, MSMA_BERT and MSMA_OPF, referring to MSMA relying only on the view of Onehot, PSSM, BERT and OPF, respectively, to predict protein function. The results are shown in Figure 4. It can be seen that MMSMA, which is based on all the four views, significantly outperforms the four single-view counterparts, demonstrating the benefits of the multi-view adaptive decision mechanism.

Note that MSMA_BERT achieves the best performance among all single-view versions, suggesting that the BERT view is an effective feature. To further verify this finding, we conduct four experiments to analyze the performance of MMSMA using three views only, by discarding one of the four views each time, respectively.

The results in Figure 5 show that the performance of MMSMA decreases the most when BERT view is removed, which indicates that the BERT view is an essential one that contains the most discriminative features. This result also shows that the unsupervised language model-based encoding features have great potential to capture the functional features of proteins, which agrees with the finding reported in the literature [46, 47].

Ablation analysis

Ablation experiments are conducted to demonstrate the effectiveness of the three modules MLCE, MSFE and MHA in the proposed MSMA. We take the one-hot encoding view as an example to conduct the ablation analysis on the large-scale CAFA3 data set. Experiments are conducted with four different combinations of the three modules, i.e. MLCE only, MLCE and MSFE, MLCE and MHA, and all three modules combined. The following findings are obtained from the results in Table 3. First, compared with MLCE, the introduction of MSFE obviously improves the prediction performance of the three sub-ontologies in terms of F_{\max} , S_{\min} and AUPR. For example, the incorporation of MSFE into MLCE is able to increase by 10.4, 4.8 and 3.2% on MFO, BPO and CCO, respectively. That is, MSFE can expand the receptive field and enhance the semantic information representation. Second, the incorporation of MHA into MLCE improves the performance in terms of F_{\max} , S_{\min} and AUPR and shows that the long-range information generated by MHA is beneficial for protein function prediction. Third, the incorporation of both MSFE and MHA into MLCE yields the best performance in terms of F_{\max} , S_{\min} and AUPR in the three sub-ontologies, which indicates that the combination of MLCE, MSFE and MHA can effectively extract deep features that are conducive to protein function prediction.

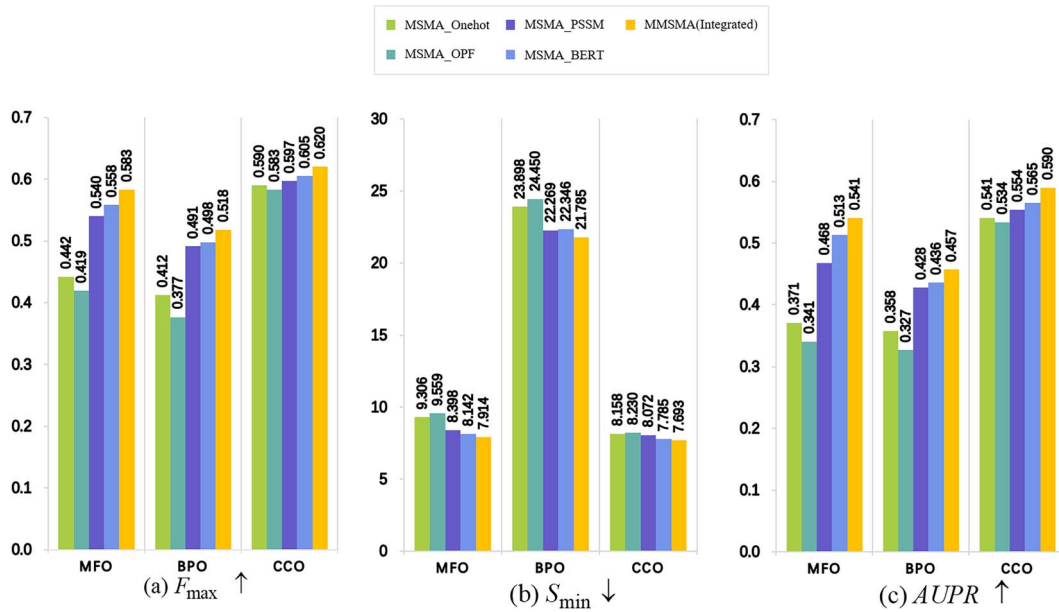


Figure 4. Performance of multi-view and single-view MSMA in terms of F_{max} , S_{min} and AUPR, where MSMA_onehot is trained using the Onehot view, and MMSMA (Integrated) refers to MSMA trained with multiple views. The arrows denote achievements of better performance (i.e. lower values of S_{min} , or higher values of F_{max} and AUPR).

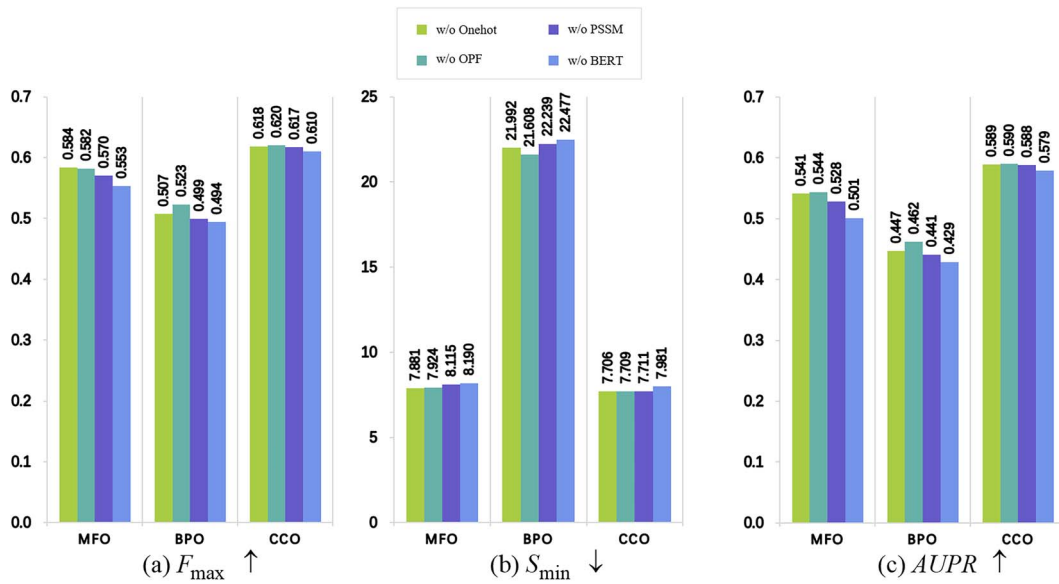


Figure 5. Performance of MMSMA with different views removed. (w/o Onehot, w/o OFF, w/o PSSM and w/o BERT refer to the removal of the views of Onehot, OFF, PSSM and BERT, respectively, from MMSMA.)

Table 3. Effect evaluation of different components in MSMA

MLCE	MSFE	MHA	F_{max}			S_{min}			AUPR		
			MFO	BPO	CCO	MFO	BPO	CCO	MFO	BPO	CCO
✓			0.291	0.315	0.552	10.544	24.987	8.631	0.204	0.270	0.433
✓	✓		0.395	0.363	0.584	9.745	24.863	8.320	0.320	0.309	0.541
✓		✓	0.357	0.351	0.556	10.086	24.842	8.570	0.284	0.292	0.484
✓	✓	✓	0.442	0.412	0.595	9.350	24.115	8.150	0.363	0.356	0.550

Note: Best performance in bold F_{max} and AUPR, highest; S_{min} , lowest.

The influence of the number of attention heads on prediction performance

In this section, we analyze the influence of the number of attention heads (h) on the prediction performance. Experiments

are conducted by setting h to 1, 2, 4, 6 and 8, respectively. The experimental results are shown in Table 4. It can be seen that when h ranges from 1 to 6, the overall trend of F_{max} increases steadily. When h is increased to 8, F_{max} on MFO, BPO and CCO

Table 4. The performance of MSMA with different h on the CAFA3 data set

Number of attention heads	F_{\max}			S_{\min}			AUPR		
	MFO	BPO	CCO	MFO	BPO	CCO	MFO	BPO	CCO
$h=1$	0.405	0.385	0.587	9.742	24.301	8.256	0.329	0.324	0.542
$h=2$	0.425	0.386	0.580	9.502	24.414	8.255	0.347	0.338	0.522
$h=4$	0.442	0.412	0.590	9.306	23.898	8.158	0.371	0.358	0.541
$h=6$	0.447	0.400	0.586	9.335	24.214	8.210	0.358	0.348	0.535
$h=8$	0.420	0.391	0.592	9.527	24.526	8.126	0.329	0.338	0.548

Note: Best performance in bold F_{\max} and AUPR, highest; S_{\min} , lowest.

Table 5. Prediction results of LYPA2_MOUSE (Uniprot Symbol: Q9WTL7) in BPO [the root GO term (GO:0002084 biological process) is omitted]

Method	Annotations	F_{\max}
DiamondScore	GO:0043170 , GO:0007155, GO:0098734 , GO:0098732 , GO:0043412 , GO:0098609, GO:0071704 , GO:0065007 , GO:0008152, GO:0022610, GO:0044699, GO:0050789	0.417
DeepGOCNN	GO:0009987 , GO:0008152 , GO:0044237	0.378
DeepGOPlus	GO:0098609, GO:0043170 , GO:0065007, GO:0043412 , GO:0071704 , GO:0008152 , GO:0044699, GO:0007155, GO:0009987 , GO:0050789, GO:0022610, GO:0098732 , GO:0098734 , GO:0044237	0.469
MMSMA	GO:0006082, GO:0006464 , GO:0006793, GO:0008152 , GO:0009056 , GO:0009987 , GO:0019538 , GO:0032502, GO:0036211 , GO:0042221, GO:0043170 , GO:0043412 , GO:0043436, GO:0044237 , GO:0044238 , GO:0044248, GO:0044260 , GO:0044267 , GO:0044699, GO:0071704 , GO:1901564, GO:1901575	0.467
MMSMAPlus	GO:0006464 , GO:0007155, GO:0008152 , GO:0009056 , GO:0009987 , GO:0019538 , GO:0022610, GO:0036211 , GO:0043170 , GO:0043412 , GO:0044237 , GO:0044238 , GO:0044260 , GO:0044267 , GO:0044699, GO:0048523, GO:0050789, GO:0065007, GO:0071704 , GO:0098609, GO:0098732 , GO:0098734 , GO:1901575	0.667
Ground truth	GO:0042159, GO:0042157, GO:0009056, GO:1901575, GO:0002084 , GO:0098734, GO:0044237, GO:0009057, GO:0019538, GO:0044238, GO:0043412, GO:0071704, GO:0044260, GO:0036211, GO:0035601, GO:0008152, GO:0006464, GO:0043170, GO:0009987, GO:0044267, GO:0030163, GO:0098732	

shows a downward trend. Therefore, setting an appropriate h value is important to improve the prediction performance of protein function. If h is too large, redundant information and large neighborhoods may be generated, which can lead to misleading long-range dependence and degrade prediction performance.

Case analysis

In this section, we use a protein that does not appear in the training set as an example to illustrate the performance difference in GO annotation between MMSMAPlus and the comparison methods. Table 5 shows the predicted results for the protein LYPA2_MOUSE (Uniprot Symbol: Q9WTL7) in BPO. The predicted results of the methods are shown in the upper part of the table, whereas the last row shows the ground truth for LYPA2_MOUSE based on the propagation of the BPO experimental annotation (GO:0002084), containing 22 true GO terms. The GO terms with correct predictions are bold-faced.

The models predicted different numbers of GO terms for LYPA2_MOUSE. Compared with the ground truth, DiamondScore gives correct results for six out of 12 predicted GO terms with a 50% accuracy, which indicates the existence of homologous proteins in training set of LYPA2_MOUSE. DeepGOCNN correctly predicts all three GO terms, but its F_{\max} is the lowest. DeepGOPlus gives correct results for eight of the 14 GO terms with 57.1% accuracy, and the F_{\max} is better than that of DiamondScore and DeepGOCNN. The proposed MMSMA correctly predicts 14 out of the 22 GO terms with 63.6% accuracy. The incorporation of homology information enables the proposed MMSMAPlus to achieve a higher accuracy of 69.5%, giving correct results for 16 out of the 23 terms and the highest F_{\max} . In addition, MMSMAPlus method successfully predicts all the GO terms that

DeepGOPlus can predict and is able to predict eight more GO terms.

To further analyze the proposed methods in terms of their practical significance in life science studies, we have highlighted the predictions of the five methods in the directed acyclic graph (DAG) of the protein LYPA2_MOUSE at BPO as shown in Figure 6. It is clear that MMSMAPlus can make the same correct predictions of the comparison methods and surpass them by making more correct predictions, it also provides more annotations at a deeper layer, demonstrating that MMSMAPlus is a more practical and powerful method for biological research.

In summary, the real-case analysis demonstrates that the proposed method MMSMAPlus is superior to the comparison methods.

DISCUSSION

As a sequence-based deep learning method for predicting protein function, MMSMAPlus is trained on multi-view sequence features of proteins, which can rapidly predict GO terms and improves the performance on the majority of function terms over state-of-the-art sequence-based methods. Although the high-quality homology-based methods are not always effective, nevertheless, they remain a useful approach to inferring protein function [42]. Thus, one important advantage of MMSMAPlus is its ability to integrate homology-based methods.

Comparing with the benchmark models, the proposed MMSMAPlus has shown higher prediction accuracy and more reliable term annotations ability for target proteins. In addition, the ablation studies illustrate the ability of MMSMAPlus to extract both local pattern features and long-range dependencies features,

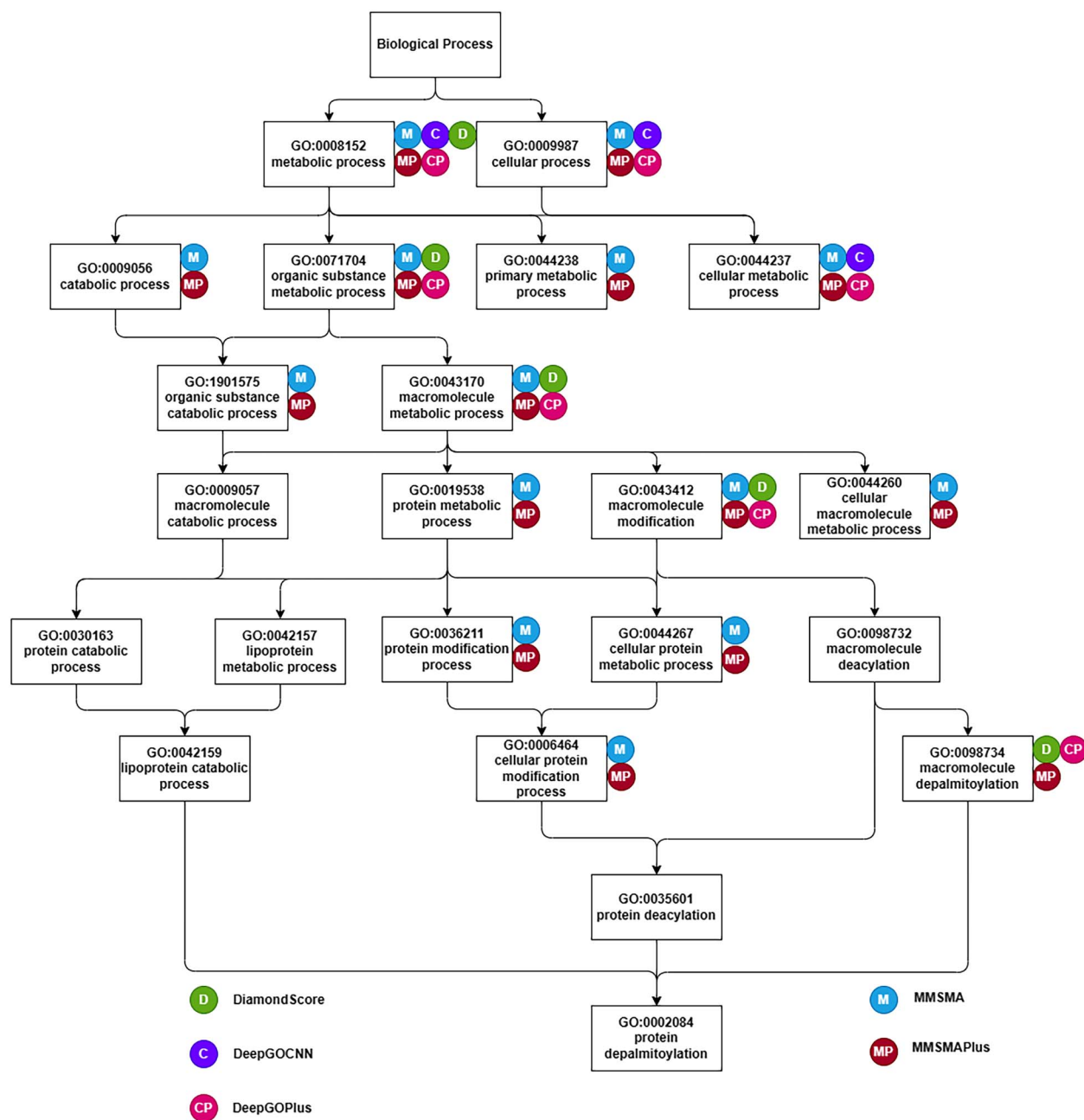


Figure 6. Predicted GO terms of LYPA2_MOUSE (Uniprot Symbol: Q9WTL7) in DAG of BPO by different methods. Ground truth is obtained by annotation propagation based on the BPO experimental annotation (GO:0002084).

thus enabling complementary predictions beyond homology-based transfer.

In summary, our method has both the comprehensive feature learning ability and the homology-based transfer ability. Thus, this method has the potential to address the challenges in annotation because of the increasing number of genome sequence data.

CONCLUSION

With the development of high-throughput sequencing technologies, automated protein function prediction has become one of the fundamental challenges in the post-genomic era. In this study, we only focus on sequence-based protein function prediction. In order to fully explore the information in protein sequences, we

extract amino acid sequences from four views, that is, one-hot encoding information, evolutionary information, deep semantic information and overlapping physicochemical property. Based on these four views, we build a multi-view deep network model MMSMA with MSFE, MHA mechanism and multi-view adaptive decision mechanism. Furthermore, the homology-based extension MMSMAPlus is proposed. Experimental results show that the design of these modules makes MMSMAPlus superior to existing methods.

Although the MMSMA and MMSMAPlus achieved promising performance, there is still room for further improvement. For example, the proposed model learns the four views independently to obtain the respective preliminary prediction results, which are then integrated by the multi-view adaptive decision mechanism. Alternatively, we can use multi-view learning techniques [48–50]

to jointly learn multiple views. In the future, more views will be extracted from sequence features [51, 52], and some multi-view joint learning techniques will be employed to further improve the performance of protein function prediction.

Key Points

- We investigate feature extraction techniques suitable for protein function prediction from four views, including one-hot encoding information, evolutionary information, deep semantic information, and overlapping property information.
- We design MSMA to extract deep features from different views and obtain the preliminary protein function predictions. For each MSMA, a multi-scale deep feature extractor with a feature pyramid structure is designed to capture local features, and a multi-head attention mechanism is adopted to capture the long-range dependence between local features.
- We present a multi-view adaptive decision mechanism to make a comprehensive decision based on the classification results of all the views.
- We further propose an extended version of MMSMA, MMSMAPlus, to integrate homology-based protein prediction under the framework of multi-view deep neural model.
- We conduct comprehensive experimental evaluations and show that the methods proposed in this study can achieve excellent performance in protein function prediction.

SUPPLEMENTARY DATA

Supplementary data are available online at <https://academic.oup.com/bib>.

FUNDING

National Key Research and Development Program of China (2021YFE010178); National Natural Science Foundation of China (62176105); Hong Kong Research Grants Council (PolyU 152006/19E).

DATA AVAILABILITY

The data and source code are available for research and non-commercial use at <https://github.com/wzy-2020/MMSMAPlus>.

REFERENCES

1. Friedberg I. Automated protein function prediction—the genomic challenge. *Brief Bioinform* 2006;**7**:225–42.
2. Shumilin IA, Cymborowski M, Chertihin O, et al. Identification of unknown protein function using metabolite cocktail screening. *Structure* 2012;**20**:1715–25.
3. Clark WT, Radivojac P. Analysis of protein function and its prediction from amino acid sequence. *Proteins* 2011;**79**:2086–96.
4. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;**25**:3389–402.
5. Remmert M, Biegert A, Hauser A, et al. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* 2011;**9**:173–5.
6. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 2015;**12**:59–60.
7. Lobley AE, Nugent T, Orengo CA, et al. FFPred: an integrated feature-based function prediction server for vertebrate proteomes. *Nucleic Acids Res* 2008;**36**:W297–302.
8. Cozzetto D, Minneci F, Currant H, et al. FFPred 3: feature-based function prediction for all gene ontology domains. *Sci Rep* 2016;**6**:31865.
9. Kulmanov M, Khan MA, Hoehndorf R. DeepGO: predicting protein functions from sequence and interactions using a deep ontology-aware classifier. *Bioinformatics* 2018;**34**:660–8.
10. You R, Huang X, Zhu S. DeepText2GO: improving large-scale protein function prediction with deep semantic text representation. *Methods* 2018;**145**:82–90.
11. Gligorišević V, Renfrew PD, Kosciółek T, et al. Structure-based protein function prediction using graph convolutional networks. *Nat Commun* 2021;**12**:3168.
12. Graves A. “Generating sequences with recurrent neural networks,” CoRR, [abs/1308.0850](https://arxiv.org/abs/1308.0850), 2013. [Online]. Available: <http://arxiv.org/abs/1308.0850>.
13. You R, Yao S, Mamitsuka H, et al. DeepGraphGO: graph neural network for large-scale, multispecies protein function prediction. *Bioinformatics* 2021;**37**:i262–71.
14. Zhang Z, Zhu Q, Xie G-S, et al. Discriminative margin-sensitive autoencoder for collective multi-view disease analysis. *Neural Netw* 2020;**123**:94–107.
15. Liu C-L, Xiao B, Hsiao W-H, et al. Epileptic seizure prediction with multi-view convolutional neural networks. *IEEE Access* 2019;**7**:170352–61.
16. Marcotte EM, Pellegrini M, Ng H-L, et al. Detecting protein function and protein-protein interactions from genome sequences. *Science* 1999;**285**:751–3.
17. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. *Nat Genet* 2000;**25**:25–9.
18. Kulmanov M, Hoehndorf R. DeepGOPlus: improved protein function prediction from sequence. *Bioinformatics* 2020;**36**:422–9.
19. Zhou G, Wang J, Zhang X, et al. Predicting functions of maize proteins using graph convolutional network. *BMC Bioinformatics* 2020;**21**.
20. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 1999;**292**:195–202.
21. Li Z, Yu Y. Protein secondary structure prediction using cascaded convolutional and recurrent neural networks [C]. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. New York, USA: IJCAI, 2016, 2560–67.
22. Xia Y, Xia C-Q, Pan X, et al. GraphBind: protein structural context embedded rules learned by hierarchical graph neural networks for recognizing nucleic-acid-binding residues. *Nucleic Acids Res* 2021;**49**:e51–1.
23. Zhang F, Song H, Zeng M, et al. A deep learning framework for gene ontology annotations with sequence- and network-based information. *IEEE/ACM Trans Comput Biol Bioinform* 2021;**18**:2208–17.
24. Han X, Zhang Z, Ding N, et al. Pre-trained models: past, present and future. *AI Open* 2021;**2**:225–50.
25. Elnaggar A, Heinzinger M, Dallago C, et al. ProtTrans: towards cracking the language of life's code through self-supervised deep learning and high performance computing. *IEEE Trans Pattern Anal Mach Intell* 2021.

26. Kenton JDMWC, Toutanova LK. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL-HLT*. Minneapolis, Minnesota USA: NAACL-HIT, 2019, 4171–86.
27. Heinzinger M, Elnaggar A, Wang Y, et al. Modeling aspects of the language of life through transfer-learning protein sequences. *BMC Bioinformatics* 2019;20.
28. Zare A, Ozdemir A, Iwen MA, et al. Extension of PCA to higher order data structures: an introduction to tensors, tensor decompositions, and tensor PCA. *Proc IEEE* 2018;106:1341–58.
29. Dou Y, Yao B, Zhang C. PhosphoSVM: prediction of phosphorylation sites by integrating various protein sequence attributes with a support vector machine. *Amino Acids* 2014;46:1459–69.
30. Wei L, Zhou C, Chen H, et al. ACPred-FL: a sequence-based predictor using effective feature representation to improve the prediction of anti-cancer peptides. *Bioinformatics* 2018;34:4007–16.
31. Yang H, Wang M, Liu X, et al. PhosIDN: an integrated deep neural network for improving protein phosphorylation site prediction by combining sequence and protein–protein interaction information. *Bioinformatics* 2021;37:4668–76.
32. Lin J, Su Q, Yang P, Ma S, Sun X. Semantic-Unit-Based Dilated Convolution for Multi-Label Text Classification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, BE: EMNLP, 2018, 4554–64.
33. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. Honolulu, HI, USA: CVPR, 2017, 2117–25.
34. Guo C, Fan B, Zhang Q, Xiang S, Pan C. Augfpn: Improving multi-scale feature learning for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Seattle, WA, USA: CVPR, 2020, 12595–604.
35. Huikai W, Zhang J, Huang K, Liang K, Yizhou Y. FastFCN: Rethinking dilated convolution in the backbone for semantic segmentation, 2019. CoRR,abs/1903.11816.
36. Bdaneshvar M. Scale invariant feature transform plus hue feature, the international archives of photogrammetry, remote sensing and spatial. *Inform Sci* 2017;42:27.
37. Woo S, et al. “Cbam: Convolutional block attention module.” *Proceedings of the European conference on computer vision*. Munich, Germany: ECCV, 2018, 3–19.
38. Wang Q, et al. “ECA-Net: Efficient channel attention for deep convolutional neural networks.” *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Seattle, WA, USA: CVPR, 2020, 11534–542.
39. Zhu K, Wu J. Residual attention: A simple but effective method for multi-label recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Montreal, Canada: ICCV, 2021, 184–93.
40. Xiao L, Huang X, Chen B, Jing L. Label-specific document representation for multi-label text classification. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing*. Hong Kong, China: EMNLP-IJCNLP, 2019, 466–75.
41. Kingma DP, Ba J. Adam: A method for stochastic optimization. In: *Published as a Conference Paper at the 3rd International Conference for Learning Representations*, San Diego, 2015.
42. Radivojac P, Clark WT, Oron TR, et al. A large-scale evaluation of computational protein function prediction. *Nat Methods* 2013;10:221–7.
43. Clark WT, Radivojac P. Information-theoretic evaluation of predicted ontological annotations. *Bioinformatics* 2013;29:i53–61.
44. Davis J, Goadrich M. The relationship between precision–recall and roc curves. In: *Proceedings of the 23rd International Conference on Machine Learning*. ACM, New York, NY, USA: ICML, 2006, 233–40.
45. Cao Y, Shen Y. TALE: transformer-based protein function annotation with joint sequence–label embedding. *Bioinformatics* 2021;37:2825–33.
46. Villegas-Morcillo A, Makrodimitris S, van Ham RCHJ, et al. Unsupervised protein embeddings outperform hand-crafted sequence and structure features at predicting molecular function. *Bioinformatics* 2021;37:162–70.
47. Littmann M, Heinzinger M, Dallago C, et al. Embeddings from deep learning transfer GO annotations beyond homology. *Sci Rep* 2021;11:1–14.
48. He J, Du C, Zhuang F, et al. Online Bayesian max-margin subspace learning for multi-view classification and regression. *Mach Learn* 2020;109:219–49.
49. Huang RB, Zhang H, Chang MS. “Multi-view face detection based on multi-features AdaBoost collaborative learning algorithm.” *Advanced Materials Research*. Vol. 998. Trans Tech Publications Ltd, 2014.
50. Zhang J, Zhang P, Liu L, et al. Collaborative weighted multi-view feature extraction. *Eng Appl Artif Intel* 2020;90:103527.
51. Chen Z, Zhao P, Li F, et al. iLearn: an integrated platform and meta-learner for feature engineering, machine-learning analysis and modeling of DNA, RNA and protein sequence data. *Brief Bioinform* 2020;21:1047–57.
52. Chen Z, Zhao P, Li C, et al. iLearnPlus: a comprehensive and automated machine-learning platform for nucleic acid and protein sequence analysis, prediction and visualization. *Nucleic Acids Res* 2021;49:e60.