Protein-protein interactions prediction via multimodal deep polynomial network and regularized extreme learning machine

Lei, Haijun; Wen, Yuting; You, Zhuhong; Elazab, Ahmed; Tan, Ee-Leng; Zhao, Yujia; Lei, Baiying

2018

Lei, H., Wen, Y., You, Z., Elazab, A., Tan, E.-L., Zhao, Y., & Lei, B. (2018). Protein-protein interactions prediction via multimodal deep polynomial network and regularized extreme learning machine. *IEEE Journal of Biomedical and Health Informatics*, 23(3), 1290-1303.
doi:10.1109/JBHI.2018.2845866

<https://hdl.handle.net/10356/137162>

<https://doi.org/10.1109/JBHI.2018.2845866>

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at:
<https://doi.org/10.1109/JBHI.2018.2845866>

Downloaded on 26 Apr 2023 22:19:52 SGT

Protein-protein Interactions Prediction via Multimodal Deep Polynomial Network and Regularized Extreme Learning Machine

Haijun Lei, Yuting Wen, Zhuhong You, Ahmed Elazab, Ee-Leng Tan, Yujia Zhao, Baiying Lei*, *Member, IEEE*

Abstract—Predicting the protein-protein interactions (PPIs) has played an important role in many applications. Hence, a novel computational method for PPIs prediction is highly desirable. PPIs endow with protein amino acid mutation rate and two physicochemical properties of protein (e.g., hydrophobicity, and hydrophilicity). Deep polynomial network (DPN) is well-suited to integrate these modalities since it can represent any function on a finite sample dataset via the supervised deep learning algorithm. We propose a multimodal DPN (MDPN) algorithm to effectively integrate these modalities to enhance prediction performance. MDPN consists of a two-stage DPN, the first stage feeds multiple protein features into DPN encoding to obtain high-level feature representation while the second stage fuses and learns features by cascading three types of high-level features in the DPN encoding. We employ a regularized extreme learning machine (RELM) to predict PPIs. The proposed method is tested on the public dataset of *H. pylori*, *Human*, and *Yeast* and achieves average accuracies of 97.87%, 99.90%, and 98.11%, respectively. The proposed method also achieves good accuracies on other datasets. Furthermore, we test our method on three kinds of PPI networks and obtain superior prediction results.

Index Terms—Protein-protein interactions; Prediction; Multimodal deep polynomial network; Regularization extreme learning machine.

I. INTRODUCTION

PROTEIN-PROTEIN interaction (PPI) plays an important role to reveal molecular mechanisms in cellular processes [1], which takes part in nearly every aspect of the cell function in biological systems. In the recent years, many methods have

been proposed to discover PPIs such as the yeast two hybrid screens [2, 3] and the affinity purification [4], which belong to biological experimental methods via high-throughput with small scale and low scalability. Moreover, these methods are costly and time-consuming. Therefore, the development of efficient methods to predict PPIs is highly desired.

There are a myriad of computational methods to predict PPI developed in the past few years [5-18]. For example, in our previous works [6], we obtained higher prediction performance using similarity network fusion (SNF) to fuse features of multiple types based on the protein amino acid mutation rate and its hydrophobicity. However, the previous method still has some limitations, such as high time complexity, large memory consumption, and the performance needs to be improved. Recently, with the progress of algorithms, and ever-increasing computational demands and data volumes, deep learning has achieved an impressive performance in many fields due to the effectiveness of neural networks [19-25]. Similar to the principle of modern deep learning, deep polynomial network (DPN) utilizes a powerful layer-by-layer learning algorithm [26] where the output of each node is a quadratic function of its inputs. In this way, the features between different samples and dimensions are integrated through the network. Accordingly, this hierarchical features can represent data in a better way. Moreover, DPN runs in polynomial time since it is easy to define the number of layers and the width in any single layer. Namely, only the number of layers and the number of nodes per layer in DPN can be determined. As a feature encoding method, DPN compactly represents any function on a finite sample dataset based on its structure, which can further enhance the feature expression. As a result, DPN has achieved its dominance in machine learning domain [27-36]. For this reason, exploring DPN method to encode multimodal protein features is highly desirable.

Inspired by the efficiency and effectiveness of DPN in learning multimodal information [37-40], multimodal DPN (MDPN) algorithm for PPI prediction is devised to capture modalities with sufficient complementary information. Specifically, we consider three types of amino acids properties. The first is amino acid mutation rate extracted by BLOSUM62 matrix, where the feature dimension is reduced by two-dimensional linear discriminant analysis (2DLDA) method [6]. The second feature is the hydrophobicity extracted using the continuous wavelet transform (CWT) to obtain a feature

This work was supported partly by the integration project of production teaching and research by Guangdong Province and ministry of education (No. 2012B091100495), National Natural Science Foundation of Guangdong Province (Nos. 2017A030313377 and 2016A030313047), Shenzhen Key Basic Research Project (Nos. JCYJ20170818142347251, JCYJ20170818094109846, JCYJ20150930105133185, JCYJ20170302153337765 and JCYJ20120613113419607). Asterisk indicates corresponding author.

H. Lei, Y. Wen, and Y. Zhao are with School of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China, 518060.

Z. You is with Xinjiang Technical Institutes of Physics and Chemistry, Chinese Academy of Science, Urumqi 830011, China

B. Lei, and A. Elazab are with the National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China, 518060 (e-mail: leiby@szu.edu.cn)

E-L Tan is with School of Electronic and Electric Engineering, Nanyang Technological University, Singapore, 639798

matrix of uniform size, which is the same as that in our previous work [6]. Hydrophilicity index is an important factor in the study of protein structure [41], which affects PPIs. So, the third is the hydrophilicity obtained by the hydrophilicity index (AAindex) via discrete wavelet transform (DWT). MDPN integrates these three types of protein features by exploiting the complementary information. For the encoded feature, PPI prediction requires a classifier for final output. Inspired by promising classification performance achieved by extreme learning machine (ELM) [42-44], we adopt it for PPI classification. The key advantage of ELM is that, the calculation process is less manually intervened, and only the output weights can be determined by setting the number of hidden nodes. The leave-one-out (LOO) cross-validation approach [37] is one of the most effective methods of model selection and parameter optimization in machine learning. In order to enhance the effectiveness of the classifier, we employ the LOO cross validation approach to select an optimal regularization parameter of ELM (RELM) and achieve the best predictive performance. The predicted residual sum of squares statistic (PRESS) based on the mean square error (MSE) [37] of the LOO approach is used to select the optimal regularization term to reduce the computational complexity. Both MDPN and RELM are integrated to predict PPIs effectively. To the best of our knowledge, this is the first time to apply both MDPN and RELM for PPIs prediction.

The main contributions of this paper are: (1) We consider three types of properties of amino acids (e.g., amino acid mutation rate, hydrophobicity and hydrophilicity). In particular, the amino acid hydrophilicity index is first applied to extract protein feature. (2) The MDPN algorithm for PPI prediction is devised to capture modalities with sufficient complementary information. (3) We employ RELM approach to select an optimal regularization parameter of ELM and get good predictive performance.

The rest of this paper is structured as follows. Section II reviews the related work. Section III presents our proposed method including feature extraction, DPN, MDPN, and prediction. Section IV explores the experimental results of the proposed method on different datasets using different performance metrics. Section V discusses the properties of our proposed method and research directions. Finally, our conclusions are presented in Section VI.

II. RELATED WORK

A. PPI prediction

In the past few years, there are a myriad of computational methods to predict PPIs [5-17]. For instance, Shen *et al.* [7] considered protein amino acid sequences for information extraction assumed that amino acid sequences contain sufficient information to study protein and protein interactions. Later, many researchers employed machine learning techniques based on amino acid sequence to predict PPIs. Gou *et al.* [8] employed auto covariance (AC) to represent the protein sequence and combine support vector machine (SVM) to predict the PPIs of *S. cerevisiae* dataset. Zhou *et al.* [18] and You *et al.* [10] used protein local descriptors to capture multiple overlapping continuous and discontinuous binding patterns

within a protein sequence. Later, You *et al.* [11] obtained multi-scale local information to split the length of protein sequence to achieve a better result in *S. cerevisiae* dataset. Huang *et al.* [12] used global encoding (GE) of amino acid sequence to build a new PPIs prediction model via weighted sparse representation classifier (WSRC), which obtains high prediction accuracies of 96.82%, 97.66% and 92.83% in *Yeast*, *Human*, and *H. pylori* datasets, respectively. Ding *et al.* [13] constructed the amino acid contact matrix based on 6323 protein-protein complexes from protein database to represent protein sequence and used histogram of oriented gradient (HOG) and singular value decomposition (SVD) for feature extraction. This method was tested on *Yeast*, *Human*, and *H. pylori* datasets and achieved prediction accuracies of 94.83%, 97.60% and 89.06%, respectively. However, they did not consider amino acid mutation and two physicochemical properties of proteins (e.g., hydrophobicity, and hydrophobicity). In [6], we only used amino acid mutation and hydrophobicity properties of proteins.

B. Feature Representation

Generally, the prediction model of PPIs using protein sequences consists of three main steps: (i) feature extraction, (ii) feature processing, and (iii) feature classification. Feature extraction is a very important step and heavily affects the classification accuracy. Nanni *et al.* [14] evaluated several feature extraction methods to represent the protein sequence. Shen *et al.* [7] predicted PPIs using only sequence information with a kernel function and a conjoint triad feature for describing amino acids. However, the accuracy is still quite low because the protein information obtained by the string matching method is insufficient. Later, Yu *et al.* [15] described the differences between amino acids and partially incorporated the sequence-order information based on amino acid substitution matrix and auto covariance transformation. However, the accuracy is still not very high because they did not consider the protein amino acid mutation rate [15]. Accordingly, Huang *et al.* [16] further obtained the protein feature based on sequence alignment in database [45] based on BLOSUM62 and combined discrete cosine transformation. The accuracy is improved but the hydrophobicity of protein amino acids was not introduced. Huang *et al.* [17] demonstrated that hydrophobicity index (AAindex) can be effectively converted into the feature matrix of protein to represent protein sequences. However, they did not consider multiple proteins properties. Wen *et al.* [6] obtained higher prediction performance by SNF to fuse features of multiple types based on the protein amino acid mutation rate and its hydrophobicity. However, their method is time and memory consuming because of the network structure.

C. Deep polynomial network

Deep learning is well-known for representation ability to the detection and recognition task [46]. In general, deep learning consists of multiple processing layers to learn different level features [46, 47]. Hence, the deep learning model has achieved impressive practical performance in all fields [46, 48-51]. However, deep learning is strongly dependent on

heuristic methods with a large number of data sets. Despite decades of research, there is no clear guidance on how to select the architecture and size of the network. In 2013, Livni *et al.* proposed a new deep learning algorithm [26] with an efficient layer-by-layer learning. DPN integrates features among different samples and dimensions through the network, and finally represents the output with hierarchical features, which greatly improves the representation performance of the feature. In [27, 28], DPN and multiple kernel learning have been successfully applied to the classification of tumor image dataset. Later, Shi *et al.* [29] predicted tumor classification with small ultrasound image dataset based on stacked DPN. In [30], multimodal stacked DPN with multimodal neuroimaging feature were used for Alzheimer's disease diagnosis.

III. METHODOLOGY

Fig. 1 shows the flowchart of the proposed method. Specifically, we apply three physicochemical properties of the amino acids (mutation rate, hydrophobicity, hydrophilicity). The 20 amino acids of hydrophobicity and hydrophilicity index coefficient matrix are downloaded from the AAindex database (<http://www.genome.ad.jp/aaindex/>). Inspired by the promising performance of MDPN encoding [30], we create three original features of protein pairs and encode them via MDPN to explore feature complementarity. In MDPN encoding, the first stage is to put the features of the three protein matrices into DPN encoding to obtain high-level feature representation. The second stage is to cascade all three types of high-level features in the DPN encoding. Lastly, we employ the RELM method via LOO cross validation to get the optimal output weight matrix for the PPIs prediction.

A. Feature extraction

1) *Feature extraction method for amino acid mutation rate (BP)*: In [14], the protein amino acid sequence is represented by a $N \times 20$ block substitution matrix [15]. Here, we consider amino acid mutation rate based on BLOSUM62 matrix (HP) to transform protein $P = (p_1, p_2, \dots, p_N)$ transform into $N \times 20$ matrix which is defined as:

$$SMR(i, j) = HP(p_i, j) \quad i = 1, \dots, N, j = 1, \dots, 20, \quad (1)$$

where $HP(a, j)$ represents a 20×20 substitution matrix whose element denotes the probability of amino acid mutating to j in the evolution process. According to Eq. (1), each original protein sequence is replaced by a 400-dimensional feature. Thus, one protein pair has a total of 800-dimensional feature. The 2DLDA method is employed to reduce feature dimension, which can remove redundant information and concentrate energy as well. Let $HP_i (i = 1, \dots, N)$ denote the number of N protein pairs features, it can be reduced to the low dimension protein pair features $BP=49$ using 2DLDA method.

2) *Feature extraction method for hydrophobicity of amino acid (CP)*: To replace the corresponding amino acid sequence via AAindex, amino acid hydrophobic property [41] was introduced as a feature using CWT to transform protein sequences of different lengths into a feature matrix with sequences of different lengths into a feature matrix with uniform size (CP). In [17], Huang et al. demonstrated that, the CWT method can extract information from protein sequences. The CWT function is defined as:

$$CWT(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \bar{\phi}\left(\frac{t-b}{a}\right) dt, \quad (2)$$

where a is a controlling parameter, b is a translation parameter, $\phi(t)$ is the mother wavelet function, $f(t)$ is the protein amino acid sequence. We empirically set $a = 5$, $b = 5$ to represent protein sequences so that each original protein sequence of varying lengths is converted to a size of 25-dimensional feature. Finally, one protein pair consists of two protein sequences with 50-dimensional feature.

1) *Feature extraction method for hydrophilicity of amino acid (DP)*: To obtain the feature matrix of DP , we employ DWT to transform discrete time signal to discrete wavelet representation, which is defined as:

$$DWT(a, b) = \frac{1}{\sqrt{a}} \int f(t) \psi\left(\frac{t-b}{a}\right) dt, \quad (3)$$

where a is a scale parameter, b is a shift factor, $\psi(t)$ is the mother wavelet, and $f(t)$ is the protein amino acid sequence. DWT is a multiresolution and timescale representation using the approximation and detail coefficients. The approximation coefficients retain the large scale information of low frequency, while the detail coefficients retain high frequency small scale information. By experiments, we select 18-dimensional feature to represent the approximate coefficients, and 8 coefficients to represent the detail coefficients. Finally, one protein pair consists of two protein sequences with 52-dimensional feature.

B. Deep polynomial network

Fig. 2a is a basic DPN of 4 degrees, using training samples $\{(x_1, y_1), \dots, (x_m, y_m)\}$, coefficient vector $(w_1, w_2, \dots, w_m)$, and m polynomials $p_1, p_2, \dots, p_m$, we obtain $\sum_{i=1}^m w_i p_i(x_j) = y_j$ for any output values $(y_1, y_2, \dots, y_m)$ according to the Lemma in [26]. The degree-1 polynomials are

$$\{(< w, [1 \ x_1] >, \dots, < w, [1 \ x_m] >): w \in R^{m+1}\}. \quad (4)$$

We propose a SVD method to search a set of $d + 1$ -dimension vectors $w_1, w_2, \dots, w_{d+1}$. We finally map $[1 \ X]$ to the constructed basis via linear transformation. Let $F^1 \in \mathbb{R}^{m \times (d+1)} = < W_j, [1 \ X] >$ and its columns are linearly independent, the construction of the first layer F^1 and matrix $[F \ \tilde{F}^2]$ is

$$\tilde{F}^2 = [(F_1^1 \circ F_1^1) \dots (F_1^1 F_{|F_1|}^1) \dots (F_{|F_1|}^1 \circ F_1^1) \dots (F_{|F_1|}^1 \circ F_{|F_1|}^1)], \quad (5)$$

where $F_1^1 \circ F_1^1$ is the Hadamard product. $[F \ \tilde{F}^2]$ is obtained by degree-2 polynomials and its columns are linearly independent while the F^2 designates the 2nd-layer network.

We redefine F as the augmented matrix $[F \ F^2]$. The processes of constructing the 2nd layer of the network can be extended to construct $3^{rd}, \dots, z^{th}$ layer by repeating the same process. Any degree- z can be represented as the degree-1 polynomials and degree $-(z - 1)$ polynomials. Hence, $\tilde{F}^z$ matrix can be reformulated as

$$\tilde{F}^z = [(F_1^{z-1} \circ F_1^1) \dots (F_1^{z-1} \circ F_{|F_1|}^1) \dots (F_{|F_1|}^{z-1} \circ F_1^1) \dots (F_{|F_1|}^{z-1} \circ F_{|F_1|}^1)], \quad (6)$$

where $[F \ F^2]$ is from a basis $[F \ \tilde{F}^z]$. To maintain numerical stability, $\tilde{F}^z$ is converted to F^z via

$$F_r^z := W_{i(r), j(r)} F_{i(r)}^{z-1} \circ F_{j(r)}^1, \quad r = 1, \dots, |F^z|. \quad (7)$$

However, this DPN algorithm has two challenges: First, the network increases with m and becomes very large. Hence, this method is only suitable for small dataset. Second, the huge network is more likely to cause overfitting. To overcome these issues, the network width is set to be equal to the number of columns in F^t , which is explicitly constrained at each iteration

as in [52]. For the calculation of the first layer in the DPN, a linear transformation transforms the augmented data matrix $[1 X]$ to the top k leading singular vectors via principal component analysis (PCA). In the next layer, we employ a standard orthogonal least squares algorithm to iteratively select residuals (after projection on the existing base $\tilde{F}^z$), which is most relevant to the residuals of actual label.

C. Multimodal deep polynomial network

Fig. 2b is the basic DPN for multiple features, namely, linear combination of multiple features and further encoded by DPN. DPN encoding is utilized to learn high-level feature representation for each feature modality. Single DPN may ignore the complementary information between different modalities and fail to fuse and capture linear combination of multiple feature representation. The linear combination of multiple features neglects its own attributes and does not consider the relevant information among different modalities. To explore complementary information of multiple features effectively, each high-level feature representation is cascaded with other types of features in the DPN encoding. Fig. 2c shows the MDPN for multiple features. In the first stage, original feature of each protein is put into a basic DPN encoding. In the second stage, all three types of the high-level features are cascaded to the DPN encoding. The ultimate learned high-level features include both intrinsic attribute of each modality and the

relativities among all modalities. Therefore, the MDPN exploits the complementarity among multiple modalities and exhibits a highly non-linear correlation among multiple modalities. As shown in Fig. 2c, the features of *BP*, *CP*, and *DP* matrices are used to carry out the DPN encoding to obtain high-level feature representation. Compared with previous methods [6, 11-13], our method combines three modalities to include both common and complementary information of proteins. The iteration to build each layer by upper layer can reduce noise and obtain effective information. It is noteworthy that, if only one layer of network is constructed by the basic DPN encoding, the first layer computes a linear transformation that transforms the augmented data matrix $[1 X]$ into its first K leading singular vector, which is equivalent to the dimensionality reduction of PCA method. Taking into account that our protein datasets are small, the feature dimension of each of our modalities is relatively small. Therefore, we construct two layers in each basic DPN network and set the number of layers and width of each layer as in Section IV. Then, the features are obtained by the direct connection of the learned hierarchical features from different layers. Accordingly, MDPN can fuse different feature matrices to further enhance feature representation.

Fig. 1. Flowchart of the proposed method.

Fig. 2. Flowchart of DPN encoding, (a) Basic DPN of 4 degrees, (b) Basic DPN for multiple features, (c) Multimodal DPN (MDPN) for multiple features.

D. Prediction

The ELM is based on signal hidden layer feedforward neural network which randomly selects the hidden nodes and ensures the output weights. Here, we utilize ELM approach to predict the PPIs. Given training samples $\{(x_j, t_j)\}_{j=1}^N$ denote N samples and m classes, $x_j = [x_{j1}, x_{j2}, \dots, x_{jn}]^T \in R^n$ denotes the input, and $t_j = [t_{j1}, t_{j2}, \dots, t_{jn}]^T \in R^m$ denotes the desired output, the ELM with N hidden nodes and activation function $g(x)$ is defined as

$$\sum_{i=1}^L \beta_i g(w_i \cdot x_j + b_i) = o_j, j = 1, 2, \dots, N, \quad (8)$$

where b_i is the bias of i -th hidden node, and w_i denotes the weight linking the i -th hidden node to the output nodes, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ denotes the weight vector connecting the i -th hidden node and the output nodes, and o_j denotes the actual network output. For $\sum_{j=1}^N \|o_j - t_j\| = 0$, the ELM aims to minimize the following function

$$\min_{\beta} \|H\beta - T\|_F, \quad (9)$$

where

$$H(w_1, \dots, w_L, b_1, \dots, b_L) = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \dots & g(w_L \cdot x_1 + b_L) \\ \vdots & & \vdots \\ g(w_1 \cdot x_N + b_1) & \dots & g(w_L \cdot x_N + b_L) \end{bmatrix}, \beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}, T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}. \quad (10)$$

where H is the hidden layer output matrix [53, 54]. Generally, the regularized least square method [55] is proposed to solve the optimal solution of Eq.(10) as follows

$$\min_{\beta} \|H\beta - T\|_F + \frac{1}{\lambda} \|\beta\|_F, \quad (11)$$

where λ is the regularization parameter to balance between training error and regularization term. The value of λ is very crucial and the performance is heavily affected by it. Therefore, we employ LOO approach [55] to determine the optimal regularization λ . The LOO approach means that the dataset is divided into N different samples, $N-1$ as training samples, leaving one as a testing sample. Here, the PRESS statistic [37] is adopted to effectively calculate the mean square error (MSE) of the LOO approach for the linear model. Supposing $H = UDV^T$ is the SVD of H , where D is the diagonal matrix, the PRESS based on MSE is

$$MSE^{PRESS} = \frac{1}{N} \sum_{j=1}^N \left(\frac{t_j - o_j}{1 - (HAT_r)_{jj}} \right)^2. \quad (12)$$

If $L \leq N$, HAT_r can be expressed as

$$HAT_r = H(H^T H + \lambda I)^{-1} H^T = H(VDU^T UDV^T + \lambda I)^{-1} H^T = HV(D^2 + \lambda I)^{-1} V^T H^T. \quad (13)$$

As the optimal regularization parameter result λ_{opt} result from Eqs. (12) and (13), the corresponding optimal output weight matrix $\tilde{\beta}$ is given by

$$\tilde{\beta} = (H^T H + \lambda_{opt} I)^{-1} H^T T = V(D^2 + \lambda_{opt} I)^{-1} V^T H^T T. \quad (14)$$

If $L \geq N$, HAT_r can be expressed as

$$HAT_r = HH^T (HH^T + \lambda I)^{-1} = HH^T (UDV^T VDU^T + \lambda I)^{-1} = HH^T U(D^2 + \lambda I)^{-1} U^T. \quad (15)$$

As the optimal regularization parameter λ_{opt} result from Eqs. (12) and (15), the corresponding optimal output weight matrix $\tilde{\beta}$ is given by

$$\tilde{\beta} = H^T (HH^T + \lambda_{opt} I)^{-1} T = H^T U(D^2 + \lambda_{opt} I)^{-1} U^T T. \quad (16)$$

Thus, we complete the implementation steps of regularized ELM via the LOO cross validation approach to predict the PPIs.

IV. EXPERIMENTAL SETUP AND RESULTS

A. Evaluation measurements

In this paper, several evaluation metrics are selected to assess the proposed method, namely, accuracy (ACC), sensitivity (SN), precision (PE), and Matthews's correlation coefficient (MCC). The area under a receiver operating curve (ROC) (e.g., AUC) is used as performance evaluation. AUC is based on a series of a binary classifier system, the false positive rate and the true positive rate are the horizontal coordinate and the vertical coordinate, respectively.

B. PPIs dataset

Several datasets are used in our experiments. First, the *H. pylori* dataset comes from Martin *et al.* [56], which include 1458 Golden standard negative (GSN) and 1458 Golden standard positive (GSP). Second, the *Human* dataset is downloaded from the human protein reference database [57], and the protein pairs sequence identity of more than 25% in the original benchmark data is excluded [58]. For GPN dataset, the proteins in different subcellular compartments do not interact with each other and finally obtain 4262 protein pairs from 661 different human proteins as the negative dataset. Totally, 3899 PPIs are used as GSP and 4262 PPIs are used as GPN. Third, the *S. cerevisiae* dataset which was selected by Guo *et al.* [8] in a database of interacting protein (DIP) [59]. This dataset has 11188 PPIs, including 5594 as GSP and 5594 as GPN. It is noteworthy that the protein pairs which contain a protein with fewer than 50 residues or have $\geq 40\%$ sequence identity were removed in our PPIs dataset; the remaining 5,594 protein pairs comprise the final positive dataset. In this study, the negative PPIs dataset is derived from 5594 protein pairs occurring in two different subcellular locations. We also apply our method to other six independent datasets including *E. coli*, *C. elegans*, *H. sapien*, *H. pylori*, *M. musculus*, and *D. mela*, which are from the DIP dataset [59]. Moreover, we apply our method to three kinds of PPI network which are the one-core network of CD9, multiple-core network of Ras-Raf-Mek-Erk -Elk-Srf pathway, and crossover network of Wnt-related pathway, respectively. In the experimental verification stage, we use a popular 5-fold cross-validation method to evaluate performance. Namely, the dataset is divided into five parts, four of which are used as training sets, and one is used as a test set, which is repeated five times. Here, *H. pylori*, *Human*, and *Yeast* datasets yield the 5-cross-validation set.

C. Parameter selection

In RELM classifier, we experimentally evaluate different Hiddensizes to choose the best value. Fig. 3 shows the effects of different values of Hiddensize (2500-5500) on the accuracy of *H. pylori* dataset. We can see that, the average ACC reaches a peak of 97.87%, when Hiddensize=4500 and declining trend takes place afterwards. Therefore, we choose the value of Hiddensize=4500 in our experiments.

Fig. 3. Comparison of different values of Hiddensize on *H.pylori* dataset.

D. Performance of the proposed method

As mentioned before, the proposed MDPN algorithm is encoded to three different types of feature, *BP*, *CP*, and *DP*. We construct two layers in each basic DPN network. The first stage, the features of *BP*, *CP*, and *DP* matrices are constructed via a basic DPN encoding and we set the two layer widths=[20 30]. In the following, we use MP to denote the combination of *BP*, *CP* and *DP*. The second stage is to express all three types of feature network through cascading in the DPN encoding. We set two layer widths=[80 70]. Compared with MDPN, we also tested the following two cases: (1) features learned by only linear combination of three features, a protein pair sequence is represented as $(MP(BP + CP + DP) = 49 + 50 + 52 = 151$ features, and (2) features encoded by the basic DPN with 2-layer networks (MP-DPN) and we set the two layer widths=[80 70]. The RELM classifier is adopted to test *H. pylori*, *Human*, and *Yeast* datasets by a 5-fold cross-validation to evaluate our method. We set up Hiddensize=4500 in the RELM algorithm.

The prediction results are summarized in Tables I, II and III. From the comparison among MP-RELM, MP-DPN-RELM and MP-MDPN-RELM on the PPIs of *H.pylori*, *Human* and *Yeast* dataset, we found that the average ACC of MP-RELM, MP-DPN-RELM and MP-MDPN-RELM are *H.pylori*(90.12% vs. 97.70% vs. 97.87%), *Human*(94.62% vs. 99.84% vs. 99.90%), *Yeast*(90.67% vs. 98.17% vs. 98.11%) while the standard deviations are *H.pylori* (0.92% vs. 0.45% vs. 0.60%), *Human* (0.63% vs. 0.16% vs. 0.07%), *Yeast* (0.58% vs. 0.23% vs. 0.17%), respectively. The performance of MP-DPN-RELM and MP-MDPN-RELM are far higher than MP-RELM in *H.pylori*, *Human* and *Yeast* dataset. The performance of MP-MDPN-RELM is slightly higher than MP-DPN-RELM in *H.pylori* and *Human* dataset. Comparing MP-DPN-RELM and MP-RELM, MP-MDPN-RELM has the lowest standard deviation in *Human* and *Yeast* dataset, indicating that this method is the most stable and has better precision. Fig.6 (c) and (h) show that MP-MDPN-RELM also achieves very high AUC in *H.pylori* and *Yeast* dataset. Fig.4 intuitively shows the results using different encoding methods on three datasets with RELM.

These prediction results prove that, it is feasible to simultaneously combine the evolutionary information of BLUOSOM62, hydrophobicity, and hydrophilicity of amino acids in protein sequences. A basic DPN encoding is an effective method to capture each type of protein feature. The MDPN algorithm is concatenated after three features of DPN

TABLE I 5-FOLD CROSS VALIDATION RESULTS USING DIFFERENT ENCODING METHODS ON *H. PYLORI* DATASET AND CLASSIFIED WITH RELM (%).

Method	ACC	SN	PE	MCC	AUC
MP	90.12±0.92	81.68±1.49	100.00	80.25±1.55	95.37±1.08
MP-DPN	97.70±0.45	95.50±0.86	99.85±0.20	95.53±0.74	98.09±0.41
MP-MDPN	97.87±0.60	95.83±1.16	100.00	95.76±1.13	98.19±0.26

TABLE II 5-FOLD CROSS VALIDATION RESULTS USING DIFFERENT ENCODING METHODS ON *HUMAN* DATASET AND CLASSIFIED WITH RELM (%).

Method	ACC	SN	PE	MCC	AUC
MP	94.62±0.63	89.72±1.13	99.01±0.48	89.64±1.09	98.41±0.30
MP-DPN	99.84±0.16	99.68±0.32	100.00	99.67±0.32	100.00
MP-MDPN	99.90±0.07	99.80±0.07	100.00	99.79±0.07	100.00

TABLE III 5-FOLD CROSS VALIDATION RESULTS USING DIFFERENT ENCODING METHODS ON *YEAST* DATASET AND CLASSIFIED WITH RELM (%).

Method	ACC	SN	PE	MCC	AUC
MP	90.67±0.58	82.96±0.96	96.34±1.19	84.56±1.04	96.91±0.24
MP-DPN	98.17±0.23	96.40±0.44	100.00	96.34±0.43	99.35±0.22
MP-MDPN	98.11±0.17	96.28±0.33	100.00	96.21±0.37	99.27±0.08

TABLE IV 5-FOLD CROSS-VALIDATION RESULTS USING MP FEATURE MDPN ENCODING METHODS ON THREE DATASETS V.S. DIFFERENT CLASSIFIERS (%).

Dataset	Method	ACC	SN	PE	MCC	AUC
<i>H. pylori</i>	ELM	96.64±0.41	93.50±0.77	99.06±0.49	94.19±1.13	97.88±0.67
		97.63±0.22	95.37±0.42	95.53±0.56	99.94±0.14	98.32±0.28
	REL	97.87±0.60	95.83±1.16	100.00	95.76±1.13	98.19±0.26
<i>Human</i>	ELM	98.16±0.23	96.38±0.44	96.60±0.24	96.54±0.28	99.89±0.04
		99.51±0.11	99.02±0.23	99.55±0.28	99.55±0.18	99.99±0.01
	REL	99.90±0.07	99.80±0.07	100.00	99.79±0.07	100.00
<i>Yeast</i>	ELM	96.75±0.42	93.69±0.79	100.00	93.50±0.75	98.54±0.21
		98.02±0.51	96.13±0.97	96.26±0.93	99.93±0.16	98.73±0.50
	REL	98.11±0.17	96.28±0.33	100.00	96.21±0.37	99.27±0.08

encoding, which improves the complementarity and reduces the dimension of protein sequence feature. Lastly, we use the supervised learning based RELM classifier to predict PPIs due to its high speed and low error rate

E. Comparison with other classifiers

In this section, we validate the proposed MDPN algorithm with RELM classifier. The RELM algorithm is based on LOO approach to select an optimal regularization parameter. To further validate the performance of the RELM classifier, we compare it with two classifiers; ELM with Hiddensize=4500 and SVM classifier with radial basis function as kernel function [60]. In our work, the classification and recognition are implemented by the libsvm-3.18 (<http://www.csie.ntu.edu.tw/~cjlin/libsvm>) toolbox. The libsvm is a simple, fast and effective pattern recognition and regression software package. Through repeated experiments on *Yeast*, *Human*, and *H. pylori* dataset, we set $c=0.2$, $g=0.2$, respectively. The prediction results are shown in Table IV, Fig.5, and Fig.6 show the ROC curve (AUC) results of *H. pylori*, *Human*, and *Yeast* dataset.

Table IV illustrates the results of MP-MDPN with ELM, SVM, and RELM classifiers. The comparison between ELM, SVM, and RELM on the PPIs of *H. pylori*, *Human*, and *Yeast* dataset shows that, the averages of ACC and AUC on *H. pylori* dataset using ELM, SVM, and RELM classifiers are (96.64%,

and 97.88%), (97.63% and 98.32%), and (97.87% and 98.19%), respectively. For the *Human* dataset, the averages of the same performance measures are (98.16% and 99.89%), (99.51% and 99.99%), and (99.90% and 100%) for the ELM, SVM, and RELM classifiers, respectively. Finally, for the *Yeast* dataset, the averages of ACC and AUC using ELM, SVM, and RELM classifiers are (96.75% and 98.54%), (98.02% and 98.73%), and (98.11% and 99.27%), respectively. MP-MDPN-RELM has the lowest standard deviation, indicating that this method has a certain degree of robustness. The ROC curves for the three classifiers on the three different datasets using MDPN encoding are shown in Fig. 6.

Fig. 5 shows the results using MDPN encoding on three datasets with ELM, SVM, and RELM. The prediction performance of RELM classifier is higher than that of ELM classifier on all three datasets, which indicates that LOO is effective in selecting the optimal regularization parameter of ELM. Moreover, the prediction performance of the SVM classifier alone is also quite good, which shows that our feature coding method (MDPN algorithm) is effective. Generally, we use the supervised learning based RELM classifier with MDPN algorithm to predict PPIs, which outperforms the ELM and SVM classifier.

F. Comparison with other dimension reduction methods

In section III, we use BLOSUM62 matrix to replace the protein amino acid sequence and combine 2DLDA for feature extraction method using amino acid mutation rate. The 2DLDA method introduced by Ye *et al.* [61] is adopted to reduce the high dimension features after the data points to distinguish PPIs. In order to ensure the 2DLDA performance, we compare 2DLDA and PCA [62-65] for reducing feature dimension using our proposed method (MDPN-RELM). Table V lists the PPIs prediction performances of PCA and 2DLDA on *Yeast* dataset. It can be noticed that, the average ACC of 2DLDA is slightly higher than PCA (98.11% vs. 97.49%) with low standard deviations, which proves effectiveness of the proposed method in predicting PPIs using 2DLDA method.

G. Comparison with BLOSUM62 and PSSM for representing protein

In Section III, we use BLOSUM62 matrix to replace the protein amino acid sequence for feature extraction method using amino acid mutation rate. In fact, different types of protein representations are included in the position specific scoring matrix (PSSM) of the proteins and the 3D tertiary structure of the protein in Nanni *et al.* [14]. We can see that the excellent protein replacements are good for subcellular localization and PPIs. Therefore, we compare the performance of BLOSUM62 and PSSM for representing proteins and test them on *Yeast* dataset. For PSSM, the protein with a length of N is expressed as a $N \times 20$ matrix, which is calculated as $PSSM(i, j) = \sum_{k=1}^{20} w(i, k) \times Y(j, k)$, $i = 1, \dots, N, j = 1, \dots, 20$, where $w(i, k)$ represents the frequency ratio of the k -th amino acids at the position i of the probe and the total number of probes, $Y(j, k)$ is the j -th and k -th amino acids value of Dayhoff's mutation matrix. In Table VI, we compare the performance of BLOSUM62 and PSSM using our proposed MDPN-RELM method for the PPIs prediction on *Yeast* dataset.

TABLE V 5-FOLD CROSS-VALIDATION RESULTS USING 2DLDA FOR REDUCING FEATURE DIMENSION ON YEAST DATASET (%).

Method	ACC	SN	PE	MCC	AUC
PCA	97.49±0.36	95.22±0.66	100.00	95.10±0.69	98.27±0.33
2DLDA	98.11±0.17	96.28±0.33	100.00	96.21±0.37	99.27±0.08

TABLE VI 5-FOLD CROSS VALIDATION RESULTS USING PSSM FOR REPRESENTING PROTEIN ON YEAST DATASET (%).

Method	ACC	SN	PE	MCC	AUC
PSSM	98.03±0.35	96.30±0.59	99.91±0.3	96.14±0.67	99.24±0.27
BLOSUM 62	98.11±0.17	96.28±0.33	100.00	96.21±0.37	99.27±0.08

Fig. 4. Results using different encoding methods on three datasets with RELM.

Fig. 5. Results using MDPN encoding on three datasets with different classifiers.

It is found that, the average ACC of BLOSUM62 is slightly higher than PSSM (98.11% vs. 98.03%) with lower standard deviations on the *Yeast* dataset. This proves that, the proposed method is effective to predict PPIs using both PSSM and BLOSUM62.

H. Performance on independent datasets

The proposed method (MP-MDPN-RELM) exhibits superior performance for PPIs prediction on another six datasets including *E. coli*, *C. elegans*, *H. sapien*, *H. pylori*, *M. musculus*, and *D. mela* obtained from the DIP database. Here, we employ

11118 samples of the *Yeast* dataset as the training set and the other six datasets as testing sets. In MDPN encoding, we set widths=[20 30] and widths=[80 70] in *E. coli*, *C. elegans*, *H. sapien*, *H. pylori*, *M. musculus*, and *D. mela* datasets. In RELM, we set the number of hidden nodes Hiddensize=4500. The training testing methodology for these 6 datasets is also LOO as well as for the previous datasets. The average ACCs for the six independent datasets are 85.78%, 98.23%, 95.54%, 95.07%, 95.53% and 99.77%, respectively. The *D. mela* dataset achieves the highest ACC result. In addition, classification results of the *C. elegans*, *H. sapien*, *H. pylori*, *M. musculus*, and *D. mela* datasets are higher than 95.07%, while the *E. coli* dataset achieves 85.78%. Table VII illustrates that, our method is better than the methods of Ding *et al.* [13], Huang *et al.* [12], and You *et al.* [11]. However, our classification result on the *E. coli* dataset is lower than all competitors. The reason is that, Ding *et al.* [13] construct the matrix of Amino Acid Contact, based on the statistical analysis of residue-pairing frequencies in a data base of 6323 protein-protein complexes. It is not enough to construct amino acid contact matrix by relying solely on 6323 protein-protein complexes. Here, we consider the protein amino acid mutation rate and two physicochemical properties of protein to represent the protein sequence and achieve sufficient protein features. Moreover, we proposed MDPN method to encoding three features, and further improve the protein features. It is clear that, our method can be applied to other organisms for protein interactions prediction.

I. Performance on PPI network

Predicting the PPI network is an application to demonstrate the effectiveness of our proposed method. Therefore, we investigate the prediction of three different types of PPI network using our method. The proposed method exhibits superior performance for PPIs prediction on three different types of PPI network. Here, we also employ 11118 samples of the *Yeast* dataset as the training set and the other six datasets as testing sets. In MDPN encoding, we set widths=[20 30] and widths=[80 70] in MDPN encoding and set the number of hidden nodes Hiddensize=4500 in RELM classifier.

First, the one-core network which means one protein directly interacts with other proteins. The tetraspanin protein CD9 [66] is a typical one-core network. Fig. 7 shows that, 15 of all 16 PPIs can be discovered by our method achieving accuracy of 93.75%, where the red lines are wrong predictions and the gray lines are correct predictions. In addition, our method is better than Shen *et al.* [7] and Ding *et al.* [67] methods which achieve (13/16) and (14/16), respectively. Second, the Ras-Raf-Mek-Erk-Elk-Srf pathway [68] is a representative multiple-core network that is vital to participate in various cellular processes. Our method is able to successfully predict 187 of all 189 PPIs prediction in this dataset, which outperforms works in Shen *et al.* [7] (161/189) and Ding *et al.* [67] (174/189). The prediction results are shown in Fig.8. Among them, the core proteins are Ras, Raf, Mek, Erk, Elk, and Srf, which decide the signal transduction. The wrong predictions are distributed across each core protein network. This suggests that, our method can be applied to more complex network. Third, the Wnt-related pathway as the current

TABLE VII PREDICTION ACCURACIES ON SIX DATASETS BASED ON OUR METHOD.

Dataset	Testing Pairs	Proposed	Ding <i>et al.</i> [13]	Huan <i>g et al.</i> [12]	You <i>et al.</i> [11]
<i>E. coli</i>	6954	85.78	93.18	89.35	89.30
<i>C. elegans</i>	4013	98.23	90.28	72.92	87.71
<i>H. sapien</i>	1412	95.54	94.58	88.81	94.19
<i>H. pylori</i>	1420	95.07	92.03	85.77	90.99
<i>M. musculus</i>	313	95.53	92.25	83.39	91.96
<i>D. mela</i>	21975	99.77	N/A	89.35	N/A

crossover network is indispensable in signal transduction. Ulrich *et al* [69] proved the protein interaction topology with yeast two-hybrid experiments. The prediction results of our proposed method on Wnt-related pathway dataset reveal that 95 of all 96 PPIs are correctly predicted (Fig. 9). Obviously, our method is more accurate than Shen *et al.* [7] and Ding *et al.* [67] methods, which achieve (73/96) and (91/96), respectively. Generally, our approach achieves promising performance in predicting the complicate network of protein and protein interactions.

J. Comparison with other methods

We selected state-of-the-art methods to compare with our method using the *Yeast*, *Human*, and *H. pylori* datasets. Table VIII summarises the average accuracies of other six methods including K-local hyperplane distance nearest neighbour algorithm (HKNN) [70] signature products [56], ensemble of HKNN [71], PCA-ELM [10], Weighted sparse representation (WSRC)+Global encoding (GE) [12] and HOG+SVD+Random forest (RF) [13]. The average ACCs are 84.00%, 83.40%, 86.60%, 87.50%, 92.83%, 89.06%, respectively, which are lower than our method that achieves 97.87% on *H. pylori* dataset. Table IX summarizes the average accuracies of the methods: LDA+RF [58], LDA+Rotation Forest (RoF) [58] AC+RF [58], AC+RoF [58], WSRC+GE [12], and HOG+SVD+RF [13]. Their accuracies ranges between 95.10% and 97.60% on *Human* dataset which is still lower than our result (99.90%). Finally, Table X summarizes the average accuracies of other six methods including auto cross covariance (AutoCC) [8], SVM+LD [72], RF+Local phase quantization descriptor (PR-LPQ) [73], ensemble extreme learning machines, and PCA-ensemble extreme learning machines (EELM) [10], WSRC+ pseudo substitution matrix (PSM) [74] and HOG+SVD+RF [13]. The achieved accuracies are 89.33%, 88.56%, 93.92%, 87.00%, 97.09%, and 94.83%, respectively. On the other hand, the proposed method achieves the highest accuracy on *Yeast* dataset (98.11%). Similarly, MCC and SN of our method are higher than others. All the results show that our method achieves good PPIs prediction result. It can be observed that our model enhances the prediction performance for PPIs prediction. The main reason for the excellent performance of PPI is that MDPN encoding is useful to integrate features and RELM classifier is effective in prediction.

Fig. 6. ROC curves for three different datasets using MDPN encoding with different classifiers.

TABLE VIII PERFORMANCE COMPARISON OF STATE-OF-THE-ART METHODS ON *H. PYLORI* DATASET (%).

Method	ACC	SN	PE	MCC
HKNN [70]	84.00	86.00	84.00	N/A
Signature products [56]	83.40	79.90	85.70	N/A
Ensemble of HKNN [71]	86.60	86.70	85.00	N/A
PCA-ELM [10]	87.50	88.95	86.15	78.13
WSRC+GE [12]	92.83	89.32	96.13	86.43
HOG+SVD+RF [13]	89.06	88.15	89.79	78.15
Our method	97.87	95.83	100.00	95.76

TABLE IX PERFORMANCE COMPARISON OF STATE-OF-THE-ART METHODS ON *HUMAN* DATASET (%).

Method	ACC	SN	PE	MCC
LDA+RF [58]	96.40	94.20	N/A	92.80
LDA+RoF [58]	95.70	97.60	N/A	91.80
AC+RF [58]	95.50	94.00	N/A	91.40
AC+RoF [58]	95.10	93.30	N/A	91.00
WSRC+GE [12]	97.66	95.28	99.81	95.41
HOG+SVD+RF [13]	97.60	96.37	98.59	95.21
Our method	99.90	99.80	100.00	99.79

TABLE X PERFORMANCE COMPARISON OF STATE-OF-THE-ART METHODS ON *YEAST* DATASET (%).

Method	ACC	SN	PE	MCC
AutoCC [8]	89.33±2.67	89.93±3.68	88.87±6.16	N/A
SVM+LD [72]	88.56±0.33	87.37±0.22	89.50±0.60	77.15±0.68
RF+PR-LPQ [73]	93.92±0.36	91.10±0.31	96.45±0.45	88.56±0.63
PCA-EELM [10]	87.00±0.29	86.15±0.43	87.59±0.32	77.36±0.44
WSRC+PSM [74]	97.09±0.49	94.33±0.93	100.00±0.0	94.33±0.93
HOG+SVD+RF [13]	94.83±0.26	92.40±0.50	97.10±0.35	89.77±0.50
Our method	98.11±0.17	96.28±0.33	100.00	96.21±0.37

V. DISCUSSIONS

In this paper, we propose a MDPN method to encode multiple protein features and RELM classifier to predict PPIs. The experimental results demonstrate that, the proposed method outperforms state-of-the-art methods in predicting PPIs on *Yeast*, *Human*, and *H. pylori* datasets.

In our study, we compare with protein features learned by

only linear combination of BP, CP, and DP matrices, when three linear combinations of protein features are encoded by the DPN and MDPN, the average ACCs are improved with at least 7%, 5.2%, and 7.4% on the same datasets (Tables I, II, III). However, the MDPN has modest increase as compared to DPN on the same dataset (at most 0.7%). In addition, encoding the three protein features using MDPN improves the classification accuracy on *Yeast*, *Human*, and *H. pylori* datasets, particularly using RELM classifier, up to 99.9% (Table II). It is clear that MDPN algorithm can concatenate all modalities with the exploration of the complementary information. Each original protein feature is respectively encoded by the basic DPN in the first stage. All three types of the high level features are cascaded to DPN encoding in the second stage. The high level features ultimately learned include the intrinsic properties of each modality and the relativity among all modalities. Compared with method of encoding only a basic DPN, the MDPN exploits the complementarity among multiple modalities and exhibits a highly non-linear correlation among multiple modalities.

In this study, two other classifiers, namely, SVM and ELM, are employed to estimate the performance of MDPN. Compared with ELM classifier and SVM classifier, the average ACCs of RELM classifier are improved with at least 0.09% and 0.24% on the same datasets, even if three are very prominent (Table IV and Fig. 5). This not only indicates that RELM classifier via the LOO cross validation approach to ensure an optimal regularization parameter of ELM is available, but also indicates the feature encoding of the MDPN method is really effective.

The first is amino acid mutation rate extracted by BLOSUM62 matrix and the PSSM matrix are used to replace the protein amino acid sequence respectively. Table VI show that both the substitution matrices have excellent performance in this method model, and BLOSUM62 matrix is slightly better. In addition, when extracted by BLOSUM62 matrix, where the feature dimension is reduced by 2DLDA method, we also use PCA method to test. Compared with PCA method, the average ACC of 2DLDA method with MDPN-RELM increased by 0.64% (Table V). We applied the 2DLDA method to reduce noise interference and integrate the useful information, which is employed to perform discriminant feature analysis on a matrix instead of a vector.

The proposed method also achieves good performance when applied to *E. coli*, *C. elegans*, *H. sapien*, *H. pylori*, *M. musculus*, and *D. mela* datasets (Table VII). Particularly, on *C. elegans*, *H. pylori*, *M. musculus*, and *D. mela* where the classification accuracies are higher than the-state-of-the-art methods with at least 7.95%, 3.04%, 3.28%, and 10.42%, respectively. The main reason is that we consider the evolutionary and chemical information to represent the protein sequence and achieve the best PPIs prediction results.

Investigation of PPI network also proves the superiority of the proposed method when tested on the tetraspanin protein CD9, accuracies of the Ras-Raf-Mek-Erk-Elk-Srf pathway, and the Wnt-related pathway (Figs. 7,8,9) are superior to the Shen

et al. [7] and Ding *et al.* [67] with at least 6.25%, 6.88%, and 4.17%, respectively. Therefore, it is clear that our method can be applied to other organisms and complicate networks of PPIs.

Running time is also important aspect in the evaluation of our method. Therefore, we test the sum of training and testing running times for MP-MDPN-RELM, MP-MDPN-ELM and MP-MDPN-SVM on *Yeast* dataset where the average running times are 60.88 seconds, 29.35 seconds and 53.22 seconds, respectively. Obviously, RELM takes longer time than ELM due to selection process of the optimal parameters. However, there is not much difference compared to SVM. In general, our method takes about 60s for RELM training and prediction, which is still feasible.

Although extensive experiments demonstrate the good performance of the proposed method for PPIs prediction, it is not without limitations. One limitation is that, our dataset is relatively small, therefore, testing our method against larger and more datasets can effectively help to validate its robustness. Another limitation is that, our classification accuracy for some datasets, such as *E. coli*, is still lower than other competitive methods. However, as a newly developed deep learning algorithm, DPN will inevitably need further investigations and improvement. As a future work, we will try to further improve the performance of the DPN algorithm and test more datasets to address the above mentioned limitations.

VI. CONCLUSIONS

In this paper, we put forward a novel computational method based on MDPN to encode multiple data from protein properties for PPIs prediction. The MDPN approach consists of two stages of DPN encoding to integrate the three features of protein sequences and exploit feature complementarity. In addition, we utilize RELM to effectively classify the extracted features. Experimental results on different datasets and comparisons with state-of-the-art methods show that the proposed method has low standard deviation, high accuracy and high AUC value. These show that the proposed method is robust and superior for PPIs prediction. Moreover, it can be potentially used for complex protein networks as well.

Fig. 7. Predicted results of one-core network for CD9.

Fig. 8. Predicted results of multicore network for Ras-Raf-Mek-Erk-Elk-Srf pathway.

Fig. 9. Predicted results of crossover network for Wnt-related pathway.

REFERENCES

- [1] N. H. Andersen, "Protein structure, stability, and folding. methods in molecular biology," *Journal of the American Chemical Society*, vol. 123, pp. 12933-12934, 2001.
- [2] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki, "A comprehensive two-hybrid analysis to explore the yeast protein interactome," *Proc Natl Acad Sci U S A*, vol. 98, pp. 4569-4574, Apr 10 2001.
- [3] F. Pazos and A. Valencia, "In silico two-hybrid system for the selection of physically interacting protein pairs," *Proteins*, vol. 47, pp. 219-227, May 1 2002.
- [4] A. C. Gavin, M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, *et al.*, "Functional organization of the yeast proteome by systematic analysis of protein complexes," *Nature*, vol. 415, pp. 141-147, Jan 10 2002.
- [5] L. Skrabanek, H. K. Saini, G. D. Bader, and A. J. Enright, "Computational prediction of protein-protein interactions," *Mol Biotechnol*, vol. 38, pp. 1-17, Jan 1 2008.
- [6] Y. T. Wen, H. J. Lei, Z. H. You, B. Y. Lei, X. Chen, and L. P. Li, "Prediction of protein-protein interactions by label propagation with protein evolutionary and chemical information derived from heterogeneous network," *Journal of Theoretical Biology*, vol. 430, pp. 9-20, Oct 7 2017.
- [7] J. Shen, J. Zhang, X. Luo, W. Zhu, K. Yu, K. Chen, *et al.*, "Predicting protein-protein interactions based only on sequences information," *Proc Natl Acad Sci U S A*, vol. 104, pp. 4337-4341, 2007.
- [8] Y. Guo, L. Yu, Z. Wen, and M. Li, "Using support vector machine combined with auto covariance to predict protein-protein interactions from protein sequences," *Nucleic Acids Research*, vol. 36, pp. 3025-3030, 2008.
- [9] Y. Z. Zhou, Y. Gao, and Y. Y. Zheng, "Prediction of protein-protein interactions using local description of amino acid sequence," in *Advances in Computer Science and Education Applications: International Conference, CSE 2011, Qingdao, China, July 9-10, 2011. Proceedings, Part II*, M. Zhou and H. Tan, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 254-262.
- [10] Z. H. You, Y. K. Lei, L. Zhu, J. Xia, and B. Wang, "Prediction of protein-protein interactions from amino acid sequences with ensemble extreme learning machines and principal component analysis," *Bmc Bioinformatics*, vol. 14, pp. 1-11, 2013.
- [11] Z. H. You, K. C. Chan, and P. Hu, "Predicting protein-protein interactions from primary protein sequences using a novel multi-scale local feature representation scheme and the random forest," *PLoS One*, vol. 10, p. e0125811, 2015.
- [12] Y. A. Huang, Z. H. You, X. Chen, K. Chan, and X. Luo, "Sequence-based prediction of protein-protein interactions using weighted sparse representation model combined with global encoding," *BMC Bioinformatics*, vol. 17, p. 184, Apr 26 2016.
- [13] Y. Ding, J. Tang, and F. Guo, "Identification of protein-protein interactions via a novel matrix-based sequence representation model with amino acid contact information," *International Journal of Molecular Sciences*, vol. 17, Sep 24 2016.
- [14] L. Nanni, A. Lumini, and S. Brahnam, "An empirical study on the matrix-based protein representations and their combination with sequence-based approaches," *Amino Acids*, vol. 44, pp. 887-901, 2013.
- [15] X. Yu, X. Zheng, T. Liu, Y. Dou, and J. Wang, "Predicting subcellular location of apoptosis proteins with pseudo amino acid composition: approach from amino acid substitution matrix and auto covariance transformation," *Amino Acids*, vol. 42, pp. 1619-1625, 2012.
- [16] Y. A. Huang, Z. H. You, X. Gao, L. Wong, and L. Wang, "Using weighted sparse representation model combined with discrete cosine transformation to predict protein-protein interactions from protein sequence," *Biomed Research International*, vol. 2015, 2015.
- [17] Y.-A. Huang, Z.-H. You, X. Chen, and G.-Y. Yan, "Improved protein-protein interactions prediction via weighted sparse representation model combining continuous wavelet descriptor and PseAA composition," *BMC Systems Biology*, vol. 10, pp. 485-494, 2016.
- [18] Y. Z. Zhou, Y. Gao, and Y. Y. Zheng, "Prediction of protein-protein Interactions using local description of amino acid sequence," *Communications in Computer & Information Science*, vol. 202, pp. 254-262, 2011.
- [19] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, pp. 1-127, Jan 1 2009.
- [20] R. Collobert and J. Weston, "A unified architecture for natural language processing: deep neural networks with multitask learning," presented at the 25th international conference on Machine learning, 2008.
- [21] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput*, vol. 18, pp. 1527-1554, July 1 2006.
- [22] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *The handbook of brain theory and neural networks*, pp. 255-258, 1998.
- [23] M. Ranzato, F. J. Huang, Y. L. Boureau, and Y. LeCun, "Unsupervised learning of Invariant feature hierarchies with applications to object recognition," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2007.
- [24] C. Bian, R. Lee, Y.-H. Chou, and J.-Z. Cheng, "Boundary regularized convolutional neural network for layer parsing of breast anatomy in automated whole breast ultrasound," *Cham*, pp. 259-266, 2017.
- [25] J. Z. Cheng, D. Ni, Y. H. Chou, J. Qin, C. M. Tiu, Y. C. Chang, *et al.*, "Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans," *Scientific Reports*, vol. 6, 2016.
- [26] R. Livni, S. Shalev-Shwartz, and O. Shamir, "An Algorithm for Training Polynomial Networks," *Computer Science*, vol. 26, pp. 4748-4750, 2013.
- [27] X. Liu, J. Shi, and Q. Zhang, "Tumor classification by deep polynomial network and multiple kernel learning on small ultrasound image dataset," in *Machine Learning in Medical Imaging*, vol. 9352, pp. 313-320, 2015.
- [28] S. Jun, Q. Yiyi, W. Jinjie, Z. Shichong, C. Yehua, Z. Qi, *et al.*, "Ultrasound image based tumor classification via deep polynomial network and multiple kernel learning," *Current Medical Imaging Reviews*, vol. 13, pp. 1-8, 2017.
- [29] J. Shi, S. Zhou, X. Liu, Q. Zhang, M. Lu, and T. Wang, "Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset," *Neurocomputing*, vol. 194, pp. 87-94, 2016.
- [30] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, "Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease," *IEEE Journal of Biomedical & Health Informatics*, vol. 22, pp. 173-183, 2017.
- [31] R. Livni, D. Lehavi, S. Schein, H. Nachlieli, S. Shalev-Shwartz, and A. Globerson, "Vanishing component analysis," in *Proceedings of the 30th International Conference on Machine Learning*, pp. 1-9, 2013.
- [32] S. Ying, Z. Wen, J. Shi, Y. Peng, J. Peng, and H. Qiao, "Manifold preserving: an intrinsic approach for semisupervised distance metric learning," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1-12, 2018.
- [33] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient kNN classification with different numbers of nearest neighbors," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, pp. 1774-1785, 2018.
- [34] W. Zheng, X. Zhu, Y. Zhu, R. Hu, and C. Lei, "Dynamic graph learning for spectral feature selection," *Multimedia Tools and Applications*, vol. 2017, pp. 1-17, 2017.
- [35] X. Zhu, X. Li, and S. Zhang, "Block-row sparse multiview multilabel learning for image classification," *IEEE Transactions on Cybernetics*, vol. 46, pp. 450-461, 2016.
- [36] X. Zhu, S. Zhang, R. Hu, Y. Zhu, and s. j, "Local and global structure preservation for robust unsupervised spectral feature selection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, pp. 517-529, 2018.
- [37] J. A. Cornell, "Classical and modern regression with applications," *Technometrics*, vol. 29, pp. 377-378, 1987.
- [38] S. Chen, D. Ni, J. Qin, B. Lei, T. Wang, and J.-Z. Cheng, "Bridging computational features toward multiple semantic features with multi-task regression: a study of CT pulmonary nodules," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pp. 53-60, 2016.
- [39] S. Chen, J. Qin, X. Ji, B. Lei, T. Wang, D. Ni, *et al.*, "Automatic scoring of multiple semantic attributes with multi-task feature leverage: a study on pulmonary nodules in CT images," *IEEE Transactions on Medical Imaging*, vol. 36, pp. 802-814, 2017.
- [40] X. Zhu, H. I. Suk, S. W. Lee, and D. Shen, "Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification," *IEEE Transactions on Biomedical Engineering*, vol. 63, pp. 607-618, 2016.
- [41] R. S. Hodges, Y. Chen, E. Kopecky, and C. T. Mant, "Monitoring the hydrophilicity/hydrophobicity of amino acid side-chains in the non-polar

- and polar faces of amphipathic alpha-helices by reversed-phase and hydrophilic interaction/cation-exchange chromatography," *American Peptide Symposium*, vol. 1053, pp. 161-72, Oct 22 2004.
- [42] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, pp. 489-501, 2006.
- [43] J. Shi, Y. Cai, J. Zhu, J. Zhong, and F. Wang, "SEMG-based hand motion recognition using cumulative residual entropy and extreme learning machine," *Medical, Biological Engineering and Computing*, vol. 51, pp. 417-427, 2013.
- [44] X. Zhu, L. Zhang, and Z. Huang, "A sparse embedding and least variance encoding approach to hashing," *IEEE Transactions on Image Processing*, vol. 23, pp. 3737-3750, 2014.
- [45] S. Henikoff and J. G. Henikoff, "Amino acid substitution matrices from protein blocks," *Proceedings of the National Academy of Sciences*, vol. 89, pp. 10915-10919, 1992.
- [46] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, 2015.
- [47] X. Zhu, H. I. Suk, and D. Shen, "A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis," *Neuroimage*, vol. 100, pp. 91-105, 2014.
- [48] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," presented at the *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2014.
- [49] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," presented at the *International Conference on Neural Information Processing Systems*, vol. 60, pp. 1097-1105, 2012.
- [50] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," presented at the *IEEE Conference on Computer Vision & Pattern Recognition*, vol. 79, pp. 3431-3440, 2015.
- [51] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, pp. 1153-1159, May 1 2016.
- [52] S. CHEN, S. A. BILLINGS, and W. LUO, "Orthogonal least squares methods and their application to non-linear system identification," *International Journal of Control*, vol. 50, pp. 1873-1896, 1989.
- [53] H. Guang-Bin, "Learning capability and storage capacity of two-hidden-layer feedforward networks," *IEEE Transactions on Neural Networks*, vol. 14, pp. 274-281, 2003.
- [54] H. Guang-Bin and H. A. Babri, "Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions," *IEEE Transactions on Neural Networks*, vol. 9, pp. 224-229, 1998.
- [55] M. V. Heeswijk and Y. Miche, "Binary/ternary extreme learning machines," *Neurocomputing*, vol. 149, pp. 187-197, 2015.
- [56] S. Martin, D. Roe, and J.-L. Faulon, "Predicting protein-protein interactions using signature products," *Bioinformatics*, vol. 21, pp. 218-226, 2005.
- [57] G. R. Mishra, M. Suresh, K. Kumaran, N. Kannabiran, S. Suresh, P. Bala, et al., "Human protein reference database—2006 update," *Nucleic Acids Research*, vol. 34, pp. D411-D414, 2006.
- [58] X. Y. Pan, Y. N. Zhang, and H. B. Shen, "Large-scale prediction of human protein-protein interactions from amino acid sequence based on latent topic features," *J Proteome Res*, vol. 9, pp. 4992-5001, Oct 1 2010.
- [59] I. Xenarios, L. Salwinski, X. J. Duan, P. Higney, S.-M. Kim, and D. Eisenberg, "DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions," *Nucleic Acids Research*, vol. 30, pp. 303-305, 2002.
- [60] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211-244, 2001.
- [61] J. Ye, R. Janardan, and Q. Li, "Two-dimensional linear discriminant analysis," *Photogrammetric Engineering & Remote Sensing*, vol. 5, pp. 1431-1441, 2009.
- [62] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, pp. 559-572, 1901.
- [63] J. Shi, J. Wu, Y. Li, Q. Zhang, and S. Ying, "Histopathological image classification With color pattern random binary hashing-based PCANet and matrix-form classifier," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, pp. 1327-1337, 2017.
- [64] X. Zhu, Z. Huang, Y. Yang, H. T. Shen, C. Xu, and J. Luo, "Self-taught dimensionality reduction on the high-dimensional small-sized data," *Pattern Recogn.*, vol. 46, pp. 215-229, 2013.
- [65] X. Zhu, X. Li, S. Zhang, Z. Xu, L. Yu, and C. Wang, "Graph PCA hashing for similarity search," *IEEE Transactions on Multimedia*, vol. 19, pp. 2033-2044, 2017.
- [66] X. H. Yang, O. V. Kovalenko, T. V. Kolesnikova, M. M. Andzelm, E. Rubinstein, J. L. Strominger, et al., "Contrasting effects of EWI proteins, integrins, and protein palmitoylation on cell surface CD9 organization," *J Biol Chem*, vol. 281, pp. 12976-12985, 2006.
- [67] Y. Ding, J. Tang, and F. Guo, "Predicting protein-protein interactions via multivariate mutual information of protein sequences," *BMC Bioinformatics*, vol. 17, 2016.
- [68] R. J. Davis, "Transcriptional regulation by map kinases," *Mol Reprod Dev*, vol. 42, pp. 459-467, 1995.
- [69] U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F. H. Brembeck, H. Goehler, et al., "A human protein-protein interaction network: a resource for annotating the proteome," *Cell*, vol. 122, pp. 957-968, 2005.
- [70] L. Nanni, "Letters: Hyperplanes for predicting protein-protein interactions," *Neurocomputing*, vol. 69, pp. 257-263, 2005.
- [71] L. Nanni and A. Lumini, "An ensemble of K-local hyperplanes for predicting protein-protein interactions," *Oxford University Press*, vol. 22, pp. 1207-1210, 2006.
- [72] F. Guo, Y. Ding, Z. Li, and J. Tang, "Identification of protein-protein Interactions by detecting correlated mutation at the interface," *Journal of Chemical Information & Modeling*, vol. 55, pp. 2042-2049, 2015.
- [73] L. Wong, Z. H. You, S. Li, Y. A. Huang, and G. Liu, "Detection of protein-protein interactions from amino acid sequences using a rotation forest model with a novel PR-LPQ descriptor," in *International Conference on Intelligent Computing*, pp. 713-720, 2015.
- [74] Y. A. Huang, Z. H. You, P. Hu, S. Li, X. Luo, and L. Wong, "Construction of reliable protein-protein Interaction networks using weighted sparse representation based classifier with pseudo substitution matrix representation features," *Neurocomputing*, vol. 218, pp. 131-138, 2016.