

## Review Article

# Recent Advances in Predicting Protein S-Nitrosylation Sites

Qian Zhao,<sup>1</sup> Jiaqi Ma ,<sup>1</sup> Fang Xie,<sup>1</sup> Yu Wang,<sup>1</sup> Yu Zhang,<sup>1</sup> Hui Li,<sup>1</sup> Yuan Sun,<sup>2</sup> Liqi Wang,<sup>1</sup> Mian Guo,<sup>3</sup> and Ke Han <sup>1</sup>

<sup>1</sup>School of Computer and Information Engineering, Heilongjiang Provincial Key Laboratory of Electronic Commerce and Information Processing, Harbin University of Commerce, Harbin 150028, China

<sup>2</sup>School of Pharmacy, Center of Pharmaceutical Engineering and Technology, Harbin University of Commerce, Harbin 150076, China

<sup>3</sup>Department of Neurosurgery, The Second Affiliated Hospital of Harbin Medical University, Harbin 150086, China

Correspondence should be addressed to Ke Han; thruster@163.com

Received 7 January 2021; Revised 24 January 2021; Accepted 25 January 2021; Published 11 February 2021

Academic Editor: Quan Zou

Copyright © 2021 Qian Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Protein S-nitrosylation (SNO) is a process of covalent modification of nitric oxide (NO) and its derivatives and cysteine residues. SNO plays an essential role in reversible posttranslational modifications of proteins. The accurate prediction of SNO sites is crucial in revealing a certain biological mechanism of NO regulation and related drug development. Identification of the sites of SNO in proteins is currently a very hot topic. In this review, we briefly summarize recent advances in computationally identifying SNO sites. The challenges and future perspectives for identifying SNO sites are also discussed. We anticipate that this review will provide insights into research on SNO site prediction.

## 1. Introduction

Protein S-nitrosylation (SNO) is one of the most important and common posttranslational modifications (PTMs), as shown in Figure 1, incorporating the covalent modification of nitric oxide (NO) and its derivatives and cysteine residues [1]. Numerous studies have shown that S-nitrosylation regulates multiple physiological and pathological processes, such as the immune response [2], cellular senescence [3], transcription, and posttranslational regulation [4]. In addition, abnormalities in protein S-nitrosylation and other posttranslational modifications can also lead to many diseases, such as Alzheimer's disease [5–7] and breast cancer [8]. In recent years, through molecular recognition and labelling of SNO sites in proteins, many large-scale proteomics experimental screenings have been completed, and the number of SNO proteins verified by experiments is also increasing [9, 10]. As to other protein posttranslational modification sites [11–18], the predicted SNO sites are time-wasting, strenuous, and extortionate through large-scale experimental screening methods. With continuous breakthroughs in sequence and

structural biology, computational biology using machine learning has become an indispensable part of drug development [19–36].

As an alternative to biochemical experiments, identifying SNO sites in biological sequences with the least cost and efficiency in recent years is a focus of current research. To help researchers understand the development of this field, this review will use Chou's five-step rule as the literature selection criteria [37]: (1) how to select or construct an effective fiducial marker dataset subcellular location to train and test predictors, (2) how to express the sample with an effective formula that can truly reflect the intrinsic correlation between the sample and predicted target, (3) how to introduce or develop powerful algorithms to make predictions, (4) how to correctly conduct cross-validation tests to objectively evaluate the expected prediction accuracy, and (5) how to build a user-friendly web server for forecasters. In addition, to help researchers overcome the overall development of this field, this review briefly introduces early research on the identification of SNO sites using biochemical methods.

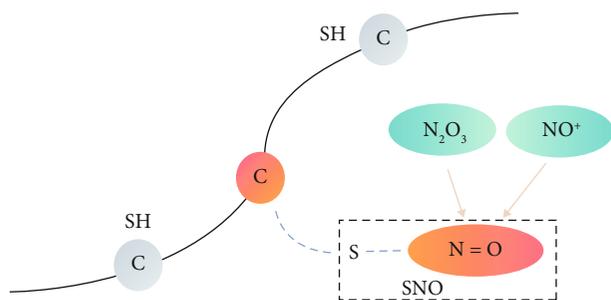


FIGURE 1: A schematic diagram of protein S-nitrosylation sites. Protein fragments have many residues, of which C (cysteine) is depicted as a circle. When NO and cysteine residues are covalently modified, SNO is formed, which is represented by a warm color, and the rest is gray.

## 2. Materials

High-quality datasets are the cornerstone of scientific research [38, 39]. With the development of proteomics and the advancement of research by scientists, the number of experimentally identified SNO sites is also increasing. In the process of predicting S-nitrosylation sites, the dynamic changes of the database and the dataset are sorted in this part.

**2.1. Database.** UniProt [40] (Universal Protein Resource) is a high-quality, extensive, and open-access database of protein sequences and functional annotations created and maintained by the UniProt Consortium, namely, EBI, SIB (Swiss Institute of Bioinformatics), and PIR (Protein Information Resource), an association of three institutions. It mainly includes three parts: the UniProtKB knowledge base, UniParc archive library, and UniRef reference sequence set. The UniProt database collects cysteine SNO sites from different species. With the continuous addition of a large number of experimentally verified SNO sites, the dataset used by scientists to predict SNO sites is also updated accordingly [41, 42].

dbSNO [43] (database of cysteine S-nitrosylation) is the first database specifically designed to integrate experimentally determined SNO sites and their structure or function information. SNO peptide sequences collected from different sources are heterogeneous, so dbSNO maps the identity of these sequences to UniProtKB protein entry. In addition, the dbSNO database also provides powerful structural and functional analysis functions to help researchers better understand the structural correlation and shared motifs of these SNO peptide sequences. The dbSNO database is divided into two versions: the first version ended in April 2012, and this version contains 43,000 experimentally verified SNO peptide sequences collected in numerous published studies using text mining methods; the second version is dbSNO2.0. In this version, dbSNO2.0 is also expanded to explore the structural environment of low SNO sites and the regulatory network resources of S-nitrosylation proteins. In SNO site prediction experiments, many scientists have also used the S-nitrosylation peptide sequence of the dbSNO database [41, 44].

PRISMOID [45] is a newly established database focusing on posttranslational modification and mutations with functional impact. Compared with traditional databases that focus on protein sequences, PRISMOID has added the real 3D structure of proteins and is equipped with various friendly operation interactions for information visualization. This database is the first version and contains 37 kinds of PTM annotation data (323 nitrosylation sites) manually compiled and is expected to be updated at least every 6 months. In addition, PRISMOID also integrates information such as the protein secondary structure and protein disordered regions to facilitate researchers to carry out scientific research.

**2.2. Datasets.** With the continuous in-depth understanding of the characteristics of S-nitrosylation, an increasing number of SNO peptide sequences have been identified, and the datasets used to predict SNO sites are based on previous studies. Dynamic changes are taking place. Therefore, the datasets commonly used by researchers for the detection of S-nitrosylation sites are chronologically explained in this section.

SNOSID, the first bioinformatics tool for predicting S-nitrosylation sites, was developed by Hao et al. [46]. In this study, they used S-nitrosoglutathione-treated rat cerebellar lysates. In 56 of the proteins, 68 cysteine sites were designated, and the initial limited 65 positive and negative samples were selected in the random sampling process. Xue et al. [47] also developed a predictor GPS-SNO for predicting SNO sites. They collected 363 experimentally verified S-nitrosylation sites published on PubMed using nitrosylated or nitrosylation as keywords and then integrated the public database SysPTM [48] and two large-scale S-nitrosylation site surveys [46, 49]. Finally, 504 positive sites and 2581 negative sites were obtained through sequence identity threshold setting [50] and protein sequence alignment [51]. A year later, Li et al. [52] used GPS-SNO datasets to develop a method for predicting SNO sites using SVM. Before long, Lee et al. [53] also developed a tool for predicting SNO sites, SNOsite. In this study, the training set and test set are from Chen et al. [54] and GPS-SNO data, respectively. Chen et al. used a high-throughput S-alkylating biotin conversion method in SNAP/L cysteine-stimulated mouse endothelial cells to obtain 586 positive sites and 2728 negative sites. In addition, since the data came from different datasets, the test set and training set may have the same homology. Therefore, they first defined SNO sequences with more than 30% identity as homologous sequences and then used BLAST 2 [55] to compare the fragment sequences. A test set containing 479 positive sites and 2501 negative sites was finally obtained.

In 2012, Li et al. [56] developed a method to predict and analyse SNO sites using minimal-redundancy-maximal-relevance and incremental feature selection. The dataset used in this experiment had three sources. The first source of SNO sites was the UniProt database [57] (version 2011\_07) and the second from GPS-SNO and the third from large-scale S-nitrosylation site surveys [58–61] at that time to obtain the remaining two datasets. Finally, a training set (784 positive sites and 1568 negative sites) and a test set (43 positive

sites and 121 negative sites) were obtained. In 2013, Xu et al. [41] developed iSNO-PseAAC, a tool for predicting S-nitrosylation sites. They randomly selected 438 proteins from the dbSNO database, and the sequence identity of these proteins was less than 40%. After comparison with the annotations in the dbSNO database, 731 positive sites and 810 negative sites verified by experiments were collected in the UniProt database [62] (version 2012\_08). Xu et al. [63] improved on the basis of iSNO-PseAAC and developed iSNO-AAPair. This experiment used the original data of S-nitrosylation sites from dbSNO (version 1.0) and the UniProt database (version 2012\_08). By using Chou's peptide formula [64–67], sequence identity setting, and random selection, 2300 SNO-positive and SNO-negative sites verified by experiments were obtained as training sets, and 81 positive and 100 negative sites were obtained as test sets.

In 2014, Jia et al. [68] developed a bioinformatics tool named iSNO-ANBPB used to predict SNO sites. In this experiment, they used the dataset constructed by Li et al. [52] and iSNO-PseAAC and obtained 1229 positive sites and 1223 negative sites by sequence identity setting and clustering. Soon, Zhang et al. [44] also developed the experimental tool PSNO. To reach a consensus assessment with previous experiments, they first constructed a training set containing 731 positive and 810 negative loci and a test set containing 53 positive and 103 negative sites from the dbSNO database. In addition, the 2302 positive sites selected from the GPS-SNO dataset and the 81 positive sites and 100 negative sites selected from the iSNO-PseAAC dataset were used as the test set.

After a brief stagnation, Xie et al. [69] used deep learning technology to develop a bioinformatics tool, DeepNitro, to predict SNO sites. They searched the relevant literature published before June 30, 2015, from PubMed and obtained a training set containing 20862 sites (3409 positive sites and 17453 negative sites) through residue modification and sequence clustering. To reach a consensus with previous research, they collected the latest data and eliminated the repeated sequences in previous work. Finally, an independent test set was built (485 positive sites and 4947 negative sites).

In 2019, Li et al. [70] predicted S-nitrosylation sites by multifeature fusion. In this study, they used 731 positive sites and 810 negative sites of iSNO-PseAAC and iSNO-AAPair as the training set and 43 positive sites and 121 negative sites of Li et al. as the test set. At the same time, Hasan et al. [71] developed PreSNO and used the DeepNitro dataset. To avoid overestimation of the prediction model, CD-HIT was used to screen homology and eliminate SNO sequences with the same window. Furthermore, to avoid prediction bias, they adopted the method of randomly taking and merging the sequences to balance the number of SNO-positive and SNO-negative sites.

### 3. Research Review

For protein S-nitrosylation site prediction, the traditional method is based on biochemical methods, but the SNO sites predicted are time-wasting, strenuous, and extortionate. With continuous breakthroughs in sequence and structural

biology, computing methods have gradually become the mainstream of current research. This method is low cost and efficient. This section focuses on computational methods based on machine learning or deep learning to provide researchers with a systematic understanding of the development of this field. Traditional biochemical methods are also briefly introduced.

*3.1. Biochemical Methods.* Jaffrey and Snyder [72] invented biotin switch assay (BSA) technology. This method first converts nitrosylated cysteine residues into biotinylated cysteine residues and then detects biotin or specific proteins by Western blot [73] to detect the proteins labelled by biotin. BSA not only greatly improves the feasibility of SNO protein identification but also promotes the improvement of high-throughput identification of SNO sites. In 2005, Gao et al. [74] proposed using BSA and protein sequencing technology to identify endogenous SNO sites. The method is simple and rapid and can meet the needs of separation, purification, and identification of SNO proteins.

In 2006, Hao et al. [46] extended the original biotin method and proposed a new improved method, SNOSID. SNOSID introduced a protein hydrolysis and digestion step before capturing the antibiotin protein. This step was not like the previous complete separation of the peptide fragment of SNO protein but the selective separation of the residues containing the SNO site before. SNOSID also introduced the machine learning algorithm SVM for the first time. In addition, the original limited 65 positive samples and 65 negative samples as training data, but the prediction results were not ideal.

Although SNOSID technology can identify the target proteins and target sites of S-nitrosylation, the degree of protein nitrosamine cannot be accurately measured. With the advancement of proteomics technology, Wu et al. [75] and Fares (2014) developed a technology combining BSA with an isotope-coded affinity tag (ICAT). This technique was the first to achieve large-scale identification of S-nitrosylation residues but is disadvantaged by its use of isotopes.

*3.2. Computational Biology Methods.* With the continuous emergence of massive biological sequences in the postgene era, traditional biochemical sequencing methods are far from being able to meet the needs of development. However, machine learning algorithms cannot directly deal with biological sequence data. Therefore, how to use discrete models or a certain way to express biological sequences and fully express their sequence information or key pattern features has become the focus and content of research in computational biology [76–84]. Since Chou proposed the pseudoamino acid composition [85, 86] or PseAAC [87], computational biology based on machine learning or deep learning has also developed rapidly. The following introduces the software and server based on Chou's five-step rule to predict protein S-nitrosylation sites through algorithms. See Table 1 for details.

*3.2.1. GPS-SNO.* Xue et al. [47] developed GPS-SNO1.0, a tool for predicting protein S-nitrosylation sites using the

TABLE 1: List of 13 predictors for predicting the SNO sites in protein sequences.

No.	Name	Link	Time	Refs
1	SNOSID	Not provided	2006	[46]
2	GPS-SNO	<a href="http://sno.biocuckoo.org/">http://sno.biocuckoo.org/</a>	2010	[47]
3	CPR-SNO	<a href="http://math.cau.edu.cn/CPR-SNO">http://math.cau.edu.cn/CPR-SNO</a>	2011	[52]
4	SNOSite	<a href="http://csb.cse.yzu.edu.tw/SNOSite">http://csb.cse.yzu.edu.tw/SNOSite</a>	2011	[53]
5	Li et al.	Not provided	2012	[56]
6	iSNO-PseAAC	<a href="http://app.aporc.org/iSNO-PseAAC">http://app.aporc.org/iSNO-PseAAC</a>	2013	[41]
7	iSNO-AAPair	<a href="http://app.aporc.org/iSNO-AAPair">http://app.aporc.org/iSNO-AAPair</a>	2013	[63]
8	iSNO-ANBPB	Not provided	2014	[68]
9	PSNO	<a href="http://59.73.198.144:8088/PSNO">http://59.73.198.144:8088/PSNO</a>	2014	[44]
10	DeepNitro	<a href="http://deepnitro.renlab.org">http://deepnitro.renlab.org</a>	2018	[69]
11	PreSNO	<a href="http://kurata14.bio.kyutech.ac.jp/PreSNO">http://kurata14.bio.kyutech.ac.jp/PreSNO</a>	2019	[71]
12	Li et al.	Not provided	2019	[70]

GPS3.0 algorithm. The software is developed on the basis of the GPS2.0 algorithm [88] previously proposed. In this study, they first used the amino acid substitution matrix to calculate the nitrosylation peptide sequence and obtain the corresponding score. Then,  $k$ -means clustering, peptide selection (PS), weight training (WT), and matrix mutation (MaM) were used to improve the performance. The accuracy of the experiment under low threshold conditions was 75.80%, the sensitivity was 53.57%, and the specificity was 80.14%. In addition, the prediction ability of GPS-SNO on 485 potential S-nitrosylation low positions was also tested, and 371 positions of these targets were successfully predicted. GPS-SNO can be obtained for free from the website <http://sno.biocuckoo.org/>.

**3.2.2. CPR-SNO.** GPS-SNO has initially explored its ability on S-nitrosylated substrates. Although good results have been achieved, there is still room for improvement. Li et al. [52] developed CPR-SNO. In this study, they used SVM as a classifier and used the coding scheme based on coupling mode to realize the prediction system. In the performance evaluation, the  $F$ -score is used to identify the effective coding scheme, and referencing the work of Xue et al. [47], tenfold cross-validation is used for verification. In addition, this research solves the problem of existing coding schemes not being able to provide enough information to predict SNO sites. By using the  $F$ -score to identify effective coupling modes, they proved that some coupling modes are not related to S-nitrosylation. The CPR-SNO server is no longer in use.

**3.2.3. SNOSite.** Although traditional research on the characteristics and mechanism of S-nitrosylation has made great progress, the understanding of its substrate specificity is still insufficient. In 2011, Lee et al. [53] made a breakthrough on this issue and developed a new bioinformatics tool, SNOSite, for predicting SNO sites. In this study, they used maximal dependence decomposition (MDD) to serialize the nitrosylation sites into different subgroups and used SVM to generate a prediction model for each MDD cluster motif. By using fivefold cross-validation, the SVM using MDD clustering

achieves 90% accuracy. SNOSite can be used for free on the website <http://csb.cse.yzu.edu.tw/SNOSite/>.

**3.2.4. mRMR and IFS Method.** Feature selection is useful for machine learning-based biosequence analysis [22, 89–105], including SNO prediction. Li et al. [56] developed a predictor based on the nearest neighbour algorithm [106] (NNA), which uses maximum relevance minimum redundancy [107] (mRMR) for incremental feature selection [108–110] (IFS). In this work, they generated 666 features from the peptide sequences used in the experiment and then used mRMR to rank the relevance and redundancy of these features in order of importance. For the obtained feature rankings, the best features are determined through IFS, and then, these features are constructed into different feature sets. Finally, the predictive evaluation performance of each feature set is generated by NNA. The best feature combination composed of 67 features is selected through the above method, and an accuracy of 0.61607 is obtained in the test set. In addition, this experiment also shows that the characteristics of the site far from the central cysteine can help determine the S-nitrosylation site. There is no online server for this predictor.

**3.2.5. iSNO-PseAAC and iSNO-AAPair.** Xu et al. [41] proposed a new SNO site predictor iSNO-PseAAC. In this study, they used PseAAC to represent protein sequence information, constructed as a  $21 \times 20$  position-specific amino acid propensity (PSAAP) matrix, and finally used the conditional random field (CRF) algorithm to construct a predictor for predicting SNO sites. The cross-validation test of iSNO-PseAAC on an independent dataset also achieved a success rate of over 90%. iSNO-PseAAC can be obtained for free on the website <http://app.aporc.org/iSNO-PseAAC/>. However, iSNO-PseAAC simply considers the positional orientation of each group of amino acids when predicting variables but does not consider any correlation between them. The amino acids in all proteins are processed individually. However, there must be some connection between them in physiology or mechanism. To solve this problem, Xu et al. [63] made improvements on the basis of iSNO-PseAAC, added related influences when predicting protein SNO sites, and

released a new SNO site prediction tool iSNO-AAPair. It considers the coupling effects of all pairs formed by the closest residues along the protein chain and the pairs formed by the closest residues. The predictor was cross-validated on the latest benchmark test set and achieved good performance. iSNO-AAPair can be obtained for free on the website <http://app.aporc.org/iSNO-AAPair/>.

**3.2.6. iSNO-ANBPB.** Jia et al. [68] proposed an iSNO-ANBPB predictor based on support vector machines. In this study, they constructed four feature extraction schemes and combined Chou's pseudoamino acid composition for model evaluation. The cross-validation of the basic SVM showed that the combination scheme using ANBPB for feature extraction obtained the best test results. In addition, studies [56] have shown that examples of the static charge of amino acids in cysteine residues and the secondary structure of amino acids play a key role in the prediction of SNO sites. Therefore, in addition to feature extraction, this study also considered the physical and chemical information in the peptide sequence. There is no online server for this predictor.

**3.2.7. PSNO.** In 2014, Zhang et al. [44] proposed a new bioinformatics tool, PSNO, for predicting SNO sites. In this study, they studied various derived features of the experimental sequence and integrated them into PseAAC to represent the experimental sample. In addition, to prevent the increase in the amount of information from increasing the difficulty of feature dimensions and predictors [111], they used relative entropy to discard noisy features from the high-level space and then optimize the optimal feature subset. However, the features of the optimal subset are different, so IFS is used here to rank these features, and a classifier based on 10-fold cross-validation is constructed for each of the optimal feature subsets. Finally, the *k*-nearest neighbour algorithm is used to predict the input sample and discriminate the prediction samples. In 10-fold cross-validation, the accuracy of PSNO was 75.67%, and the accuracy of MCC was 0.5119. With the completion of the whole-genome sequencing project, the gap in the sequence structure is rapidly expanding. In the absence of a protein structure, sequence-based prediction represented by PSNO can become a powerful supplement to replace structure-based prediction. The server provided by the software is now invalid.

**3.2.8. DeepNitro.** Since Hinton et al. [112] proposed the hierarchical training strategy to solve the gradient diffusion problem in 2006, deep learning technology has also been widely used in computational biology [113–126] and drug discovery [127–134]. In 2018, Xie et al. [69] used the deep learning algorithm for the first time to develop the S-nitrosylation site prediction bioinformatics tool DeepNitro. DeepNitro is an eight-layer neural network. The first layer is the data input layer, which is used to assign prediction and training values to neurons; the second to seventh layers are fully connected layers, of which the second to fourth layers use the dropout algorithm to improve the generalization ability of unknown data.

In the process of neural network design, to solve the problem of gradient diffusion in the training process, the ReLU function was used as the activation function, and the log-likelihood probability was used as the loss function to optimize the weights and other parameters in the neural network. In the process of backpropagation, a minibatch gradient descent algorithm is used to update the network parameters. Compared with traditional optimization algorithms, the momentum method is superior in optimizing parameters such as weights, so the momentum method was selected as the optimization function. In addition, L1 and L2 regular terms are introduced as hyperparameters to prevent overfitting. For the last layer, the softmax algorithm is used to obtain the probability distribution of the prediction results. Finally, through principal component analysis (PCA), DeepNitro obtained an AUC value of 0.7437 on the test set. DeepNitro uses deep learning algorithms, new encoding algorithms, and a position-specific scoring matrix [135] (PSSM) to greatly improve the accuracy of nitrosation site prediction and provides a free website server (<http://deepnitro.renlab.org/>) for academic research.

**3.2.9. PreSNO.** In 2019, Hasan et al. [71] proposed a prediction tool, PreSNO, for predicting protein SNO sites by an ensemble algorithm. The focus of the study was the use of four different coding schemes, including the composition of profile-based amino acids (CPA), *K*-space spectral amino acid composition (SAC), tripeptide composition from the PSSM (TCP), and physical-chemical properties of amino acids (PPA). The four coding schemes use SVM and random forest to calculate the probability score and then multiply it by weight to calculate the prediction effect of PreSNO. Through 5-fold cross-validation, PreSNO also achieved excellent performance. The predictor can be obtained for free on the website (<http://kurata14.bio.kyutech.ac.jp/PreSNO/>).

**3.2.10. Multiple Features Combination Method.** Soon, Li et al. [70] proposed a method to predict protein S-nitrosylation sites using multifeature mixing. This work improves prediction performance by extracting nine sequence features, such as parallel correlation pseudoamino acid composition (PC-PseAAC), general parallel correlation pseudoamino acid composition [136], and ANBPB. Then, the importance of amino acids is evaluated by subtracting the given amino acids from the information gain [137] (IG), and finally, the max-relevance-max-distance [138] (MRMD) generates a feature subset with lower redundancy and strong correlation with the target category. In the cross-validation of the test set, the ACC and MCC of this method were 73.17% and 0.3788, respectively, which becomes a useful supplement to the existing SNO identification tools.

## 4. Concluding Remarks and Perspectives

Many physiological and pathological studies of SNO have been reported in recent years. Therefore, accurate prediction of SNO sites will pave the way to speed up related drug development.

Several exciting computational methods have been proposed to predict SNO. Although these works promoted research on SNO and facilitated the prediction of SNO sites, the following challenges should be considered in future works.

Although many predictors have been developed to predict SNO sites, some corresponding indicators have greatly improved the space. This is because existing methods were trained on the basis of an imbalanced dataset. To solve this problem, it is necessary to collect many more positive SNO sites to enlarge the number of SNO sites in the dataset and balance it. In addition, the focus of future research in this field is to use these new technologies and methods to predict more nitrosylated target proteins and sites to reveal the mechanism by which nitrosylation regulates various physiological processes.

### Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

### Acknowledgments

The work was supported by the Natural Science Foundation of China (32072259, 31901114, and 81773161), the Natural Science Foundation of Heilongjiang Province (LH2020F009), the Support Program for Young Creative Talents of Harbin University of Commerce (2019CX01 and 2019CX22), the University Nursing Program for Young Scholar with Creative Talents in Heilongjiang Province (UNPYSCT-2020212), the Chinese Postdoctoral Science Foundation (2017M621294), and the first batch of industry-university cooperative education project of the Ministry of Education in 2019 (201901240006).

### References

- [1] M. W. Foster, D. T. Hess, and J. S. Stamler, "Protein S-nitrosylation in health and disease: a current perspective," *Trends in Molecular Medicine*, vol. 15, no. 9, pp. 391–404, 2009.
- [2] B. Cui, Q. Pan, D. Clarke et al., "S-nitrosylation of the zinc finger protein SRG1 regulates plant immunity," *Nature Communications*, vol. 9, no. 1, article 4226, 2018.
- [3] S. Rizza, S. Cardaci, C. Montagna et al., "S-Nitrosylation drives cell senescence and aging in mammals by controlling mitochondrial dynamics and mitophagy," *Proceedings of the National Academy of Sciences*, vol. 115, no. 15, pp. E3388–E3397, 2018.
- [4] R. González, A. Cruz, G. Ferrín et al., "Nitric oxide mimics transcriptional and post-translational regulation during  $\alpha$ -tocopherol cytoprotection against glycochenodeoxycholate-induced cell death in hepatocytes," *Journal of Hepatology*, vol. 55, no. 1, pp. 133–144, 2011.
- [5] G. Liu and Q. Jiang, "Alzheimer's disease CD33rs3865444 variant does not contribute to cognitive performance," *Proceedings of the National Academy of Sciences*, vol. 113, no. 12, pp. E1589–E1590, 2016.
- [6] Q. Jiang, S. Jin, Y. Jiang et al., "Alzheimer's disease variants with the genome-wide significance are significantly enriched in immune pathways and active in immune cells," *Molecular Neurobiology*, vol. 54, no. 1, pp. 594–600, 2017.
- [7] T. S. Wijasa, M. Sylvester, N. Brocke-Ahmadinejad et al., "Quantitative proteomics of synaptosome S-nitrosylation in Alzheimer's disease," *Journal of Neurochemistry*, vol. 152, no. 6, pp. 710–726, 2020.
- [8] D. Mishra, V. Patel, and D. Banerjee, "Nitric oxide and S-nitrosylation in cancers: emphasis on breast cancer," *Breast Cancer: Basic and Clinical Research*, vol. 14, article 117822341988268, 2020.
- [9] S. L. Cook and G. P. Jackson, "Characterization of tyrosine nitration and cysteine nitrosylation modifications by metastable atom-activation dissociation mass spectrometry," *Journal of the American Society for Mass Spectrometry*, vol. 22, no. 2, pp. 221–232, 2011.
- [10] B. Li, J. Tang, Q. Yang et al., "NOREVA: normalization and evaluation of MS-based metabolomics data," *Nucleic Acids Research*, vol. 45, no. W1, pp. W162–W170, 2017.
- [11] K. Chen, Z. Wei, Q. Zhang et al., "WHISTLE: a high-accuracy map of the human N6-methyladenosine (m6A) epitranscriptome predicted using a machine learning approach," *Nucleic Acids Research*, vol. 47, no. 7, article e41, 2019.
- [12] S. Amanat, A. Ashraf, W. Hussain, N. Rasool, and Y. D. Khan, "Identification of lysine carboxylation sites in proteins by integrating statistical moments and position relative features via general PseAAC," *Current Bioinformatics*, vol. 15, no. 5, pp. 396–407, 2020.
- [13] W. Bao, D.-S. Huang, and Y.-H. Chen, "MSIT: malonylation sites identification tree," *Current Bioinformatics*, vol. 15, no. 1, pp. 59–67, 2020.
- [14] W. Chen and K. Liu, "Analysis and comparison of RNA pseudouridine site prediction tools," *Current Bioinformatics*, vol. 15, no. 4, pp. 279–286, 2020.
- [15] M. A. M. Hasan, M. K. Ben Islam, J. Rahman, and S. Ahmad, "Citrullination site prediction by incorporating sequence coupled effects into PseAAC and resolving data imbalance issue," *Current Bioinformatics*, vol. 15, no. 3, pp. 235–245, 2020.
- [16] H. Long, Z. Sun, M. Li, H. Y. Fu, and M. C. Lin, "Predicting protein phosphorylation sites based on deep learning," *Current Bioinformatics*, vol. 15, no. 4, pp. 300–308, 2020.
- [17] H. Zhu, X. Du, and Y. Yao, "ConvsPPIS: identifying protein-protein interaction sites by an ensemble convolutional neural network with feature graph," *Current Bioinformatics*, vol. 15, no. 4, pp. 368–378, 2020.
- [18] G. Wang, X. Luo, J. Wang et al., "MeDReaders: a database for transcription factors that bind to methylated DNA," *Nucleic Acids Research*, vol. 46, no. D1, pp. D146–D151, 2018.
- [19] W.-Z. Zhong and S.-F. Zhou, "Molecular science for drug development and biomedicine," *International Journal of Molecular Sciences*, vol. 15, no. 11, pp. 20072–20078, 2014.
- [20] W. Xue, F. Yang, P. Wang et al., "What contributes to serotonin-norepinephrine reuptake inhibitors' dual-targeting mechanism? The key role of transmembrane domain 6 in human serotonin and norepinephrine transporters revealed by molecular dynamics simulation," *ACS Chemical Neuroscience*, vol. 9, no. 5, pp. 1128–1140, 2018.
- [21] H. Xu, J. Zhou, S. Lin, W. Deng, Y. Zhang, and Y. Xue, "PLMD: an updated data resource of protein lysine modifications," *Journal of Genetics and Genomics*, vol. 44, no. 5, pp. 243–250, 2017.

- [22] H. Lv, F.-Y. Dao, D. Zhang et al., “iDNA-MS: an integrated computational tool for detecting DNA modification sites in multiple genomes,” *iScience*, vol. 23, no. 4, article 100991, 2020.
- [23] Q. Zou, G. Lin, X. Jiang, X. Liu, and X. Zeng, “Sequence clustering in bioinformatics: an empirical study,” *Briefings in Bioinformatics*, vol. 21, no. 1, pp. 1–10, 2018.
- [24] J. Shao and B. Liu, “ProtFold-DFG: protein fold recognition by combining Directed Fusion Graph and PageRank algorithm,” *Briefings in Bioinformatics*, 2020.
- [25] B. Liu, C. Li, and K. Yan, “DeepSVM-fold: protein fold recognition by combining support vector machines and pairwise sequence similarity scores generated by deep learning networks,” *Briefings in Bioinformatics*, vol. 21, no. 5, pp. 1733–1741, 2020.
- [26] Y.-d. Liao and Z.-r. Jiang, “MoABank: an integrated database for drug mode of action knowledge,” *Current Bioinformatics*, vol. 14, no. 5, pp. 446–449, 2019.
- [27] A. Munir, S. I. Malik, and K. A. Malik, “Proteome mining for the identification of putative drug targets for human pathogen *Clostridium tetani*,” *Current Bioinformatics*, vol. 14, no. 6, pp. 532–540, 2019.
- [28] N. Srivastava, B. N. Mishra, and P. Srivastava, “In-silico identification of drug lead molecule against pesticide exposed-neurodevelopmental disorders through network-based computational model approach,” *Current Bioinformatics*, vol. 14, no. 5, pp. 460–467, 2019.
- [29] X. Zhao, L. Chen, Z.-H. Guo, and T. Liu, “Predicting drug side effects with compact integration of heterogeneous networks,” *Current Bioinformatics*, vol. 14, no. 8, pp. 709–720, 2019.
- [30] R. Jakhar, M. Dangi, A. Khichi, and A. K. Chhillar, “Relevance of molecular docking studies in drug designing,” *Current Bioinformatics*, vol. 15, no. 4, pp. 270–278, 2020.
- [31] Z. Li, T. Zhang, H. Lei et al., “Research on gastric cancer’s drug-resistant gene regulatory network model,” *Current Bioinformatics*, vol. 15, no. 3, pp. 225–234, 2020.
- [32] J. Liu, X. Lian, F. Liu et al., “Identification of novel key targets and candidate drugs in oral squamous cell carcinoma,” *Current Bioinformatics*, vol. 15, no. 4, pp. 328–337, 2020.
- [33] J. Wang, H. Wang, X. Wang, and H. Chang, “Predicting drug-target interactions via FM-DNN learning,” *Current Bioinformatics*, vol. 15, no. 1, pp. 68–76, 2020.
- [34] Q. Jiang, G. Wang, S. Jin, Y. Li, and Y. Wang, “Predicting human microRNA-disease associations based on support vector machine,” *International Journal of Data Mining and Bioinformatics*, vol. 8, no. 3, pp. 282–293, 2013.
- [35] G. Liu, Y. Hu, S. Jin, and Q. Jiang, “Genetic variant rs763361 regulates multiple sclerosis CD226 gene expression,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 6, pp. E906–E907, 2017.
- [36] W. Chen, P. Feng, T. Liu, and D. Jin, “Recent advances in machine learning methods for predicting heat shock proteins,” *Current Drug Metabolism*, vol. 20, no. 3, pp. 224–228, 2019.
- [37] K.-C. Chou, “Some remarks on protein attribute prediction and pseudo amino acid composition,” *Journal of Theoretical Biology*, vol. 273, no. 1, pp. 236–247, 2011.
- [38] Y. W. Zhao, Z. D. Su, W. Yang, H. Lin, W. Chen, and H. Tang, “IonchanPred 2.0: a tool to predict ion channels and their types,” *International Journal of Molecular Sciences*, vol. 18, no. 9, article 1838, 2017.
- [39] Z. Y. Liang, H. Y. Lai, H. Yang et al., “Pro54DB: a database for experimentally verified sigma-54 promoters,” *Bioinformatics*, vol. 33, no. 3, pp. 467–469, 2017.
- [40] U Consortium, “UniProt: a worldwide hub of protein knowledge,” *Nucleic Acids Research*, vol. 47, no. D1, pp. D506–D515, 2019.
- [41] Y. Xu, J. Ding, L.-Y. Wu, and K. C. Chou, “iSNO-PseAAC: predict cysteine S-nitrosylation sites in proteins by incorporating position specific amino acid propensity into pseudo amino acid composition,” *PLoS One*, vol. 8, no. 2, article e55844, 2013.
- [42] Y. Wang, S. Zhang, F. Li et al., “Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics,” *Nucleic Acids Research*, vol. 48, no. D1, pp. D1031–D1041, 2020.
- [43] Y.-J. Chen, C.-T. Lu, M.-G. Su et al., “dbSNO 2.0: a resource for exploring structural environment, functional and disease association and regulatory network of protein S-nitrosylation,” *Nucleic Acids Research*, vol. 43, no. D1, pp. D503–D511, 2015.
- [44] J. Zhang, X. Zhao, P. Sun, and Z. Ma, “PSNO: predicting cysteine S-nitrosylation sites by incorporating various sequence-derived features into the general form of Chou’s PseAAC,” *International Journal of Molecular Sciences*, vol. 15, no. 7, pp. 11204–11219, 2014.
- [45] F. Li, C. Fan, T. T. Marquez-Lago et al., “PRISMOID: a comprehensive 3D structure database for post-translational modifications and mutations with functional impact,” *Briefings in Bioinformatics*, vol. 21, no. 3, pp. 1069–1079, 2020.
- [46] G. Hao, B. Derakhshan, L. Shi, F. Campagne, and S. S. Gross, “SNOSID, a proteomic method for identification of cysteine S-nitrosylation sites in complex protein mixtures,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 4, pp. 1012–1017, 2006.
- [47] Y. Xue, Z. Liu, X. Gao et al., “GPS-SNO: computational prediction of protein S-nitrosylation sites with a modified GPS algorithm,” *PLoS One*, vol. 5, no. 6, article e11290, 2010.
- [48] H. Li, X. Xing, G. Ding et al., “SysPTM: a systematic resource for proteomic research on post-translational modifications,” *Molecular & Cellular Proteomics*, vol. 8, no. 8, pp. 1839–1849, 2009.
- [49] P. Han and C. Chen, “Detergent-free biotin switch combined with liquid chromatography/tandem mass spectrometry in the analysis of S-nitrosylated proteins,” *Rapid Communications in Mass Spectrometry*, vol. 22, no. 8, pp. 1137–1145, 2008.
- [50] W. Li and A. Godzik, “Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences,” *Bioinformatics*, vol. 22, no. 13, pp. 1658–1659, 2006.
- [51] S. F. Altschul, T. L. Madden, A. A. Schäffer et al., “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs,” *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389–3402, 1997.
- [52] Y.-X. Li, Y.-H. Shao, L. Jing, and N. Y. Deng, “An efficient support vector machine approach for identifying protein S-nitrosylation sites,” *Protein and Peptide Letters*, vol. 18, no. 6, pp. 573–587, 2011.
- [53] T.-Y. Lee, Y.-J. Chen, T.-C. Lu, H. D. Huang, and Y. J. Chen, “SNOSite: exploiting maximal dependence decomposition to

- identify cysteine S-nitrosylation with substrate site specificity,” *PLoS One*, vol. 6, no. 7, article e21849, 2011.
- [54] Y.-J. Chen, W.-C. Ku, P.-Y. Lin, H. C. Chou, K. H. Khoo, and Y. J. Chen, “S-Alkylating labeling strategy for site-specific identification of the S-nitrosoproteome,” *Journal of Proteome Research*, vol. 9, no. 12, pp. 6417–6439, 2010.
- [55] T. A. Tatusova and T. L. Madden, “BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences,” *FEMS Microbiology Letters*, vol. 174, no. 2, pp. 247–250, 1999.
- [56] B.-Q. Li, L.-L. Hu, S. Niu, Y. D. Cai, and K. C. Chou, “Predict and analyze S-nitrosylation modification sites with the mRMR and IFS approaches,” *Journal of Proteomics*, vol. 75, no. 5, pp. 1654–1665, 2012.
- [57] U Consortium, “The universal protein resource (UniProt) in 2010,” *Nucleic Acids Research*, vol. 38, suppl\_1, pp. D142–D148, 2010.
- [58] P.-T. Doulias, J. L. Greene, T. M. Greco et al., “Structural profiling of endogenous S-nitrosocysteine residues reveals unique features that accommodate diverse mechanisms for protein S-nitrosylation,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 39, pp. 16958–16963, 2010.
- [59] Q. Yang, Y. Wang, Y. Zhang et al., “NOREVA: enhanced normalization and evaluation of time-course and multi-class metabolomic data,” *Nucleic Acids Research*, vol. 48, no. W1, pp. W436–W448, 2020.
- [60] M. Liu, J. Hou, L. Huang et al., “Site-specific proteomics approach for study protein S-nitrosylation,” *Analytical Chemistry*, vol. 82, no. 17, pp. 7160–7168, 2010.
- [61] Q. Yang, B. Li, S. Chen et al., “MMEASE: online meta-analysis of metabolomic data by enhanced metabolite annotation, marker selection and enrichment analysis,” *Journal of Proteomics*, vol. 232, article 104023, 2021.
- [62] U Consortium, “Reorganizing the protein space at the Universal Protein Resource (UniProt),” *Nucleic Acids Research*, vol. 40, no. D1, pp. D71–D75, 2012.
- [63] Y. Xu, X.-J. Shao, L.-Y. Wu, N. Y. Deng, and K. C. Chou, “iSNO-AApair: incorporating amino acid pairwise coupling into PseAAC for predicting cysteine S-nitrosylation sites in proteins,” *PeerJ*, vol. 1, article e171, 2013.
- [64] K.-C. Chou, “Prediction of signal peptides using scaled window,” *Peptides*, vol. 22, no. 12, pp. 1973–1979, 2001.
- [65] J. Tang, J. Fu, Y. Wang et al., “ANPELA: analysis and performance assessment of the label-free quantification workflow for metaproteomic studies,” *Briefings in Bioinformatics*, vol. 21, no. 2, pp. 621–636, 2020.
- [66] Q. Yang, J. Hong, Y. Li et al., “A novel bioinformatics approach to identify the consistently well-performing normalization strategy for current metabolomic studies,” *Briefings in Bioinformatics*, vol. 21, no. 6, pp. 2142–2152, 2020.
- [67] B. Liu, X. Gao, and H. Zhang, “BioSeq-Analysis2.0: an updated platform for analyzing DNA, RNA, and protein sequences at sequence level and residue level based on machine learning approaches,” *Nucleic Acids Research*, vol. 47, no. 20, article e127, 2019.
- [68] C. Jia, X. Lin, and Z. Wang, “Prediction of protein S-nitrosylation sites based on adapted normal distribution bi-profile Bayes and Chou’s pseudo amino acid composition,” *International Journal of Molecular Sciences*, vol. 15, no. 6, pp. 10410–10423, 2014.
- [69] Y. Xie, X. Luo, Y. Li et al., “DeepNitro: prediction of protein nitration and nitrosylation sites by deep learning,” *Genomics, Proteomics & Bioinformatics*, vol. 16, no. 4, pp. 294–306, 2018.
- [70] T. Li, R. Song, Q. Yin, M. Gao, and Y. Chen, “Identification of S-nitrosylation sites based on multiple features combination,” *Scientific Reports*, vol. 9, no. 1, article 3098, 2019.
- [71] M. M. Hasan, B. Manavalan, M. S. Khatun, and H. Kurata, “Prediction of S-nitrosylation sites by integrating support vector machines and random forest,” *Mol Omics*, vol. 15, no. 6, pp. 451–458, 2019.
- [72] S. R. Jaffrey and S. H. Snyder, “The biotin switch method for the detection of S-nitrosylated proteins,” *Science’s STKE*, vol. 2001, no. 86, p. pl1, 2001.
- [73] C. Lindermayr, G. Saalbach, and J. Durner, “Proteomic identification of S-nitrosylated proteins in Arabidopsis,” *Plant Physiology*, vol. 137, no. 3, pp. 921–930, 2005.
- [74] C. Gao, H. Guo, J. Wei, Z. Mi, P. Y. Wai, and P. C. Kuo, “Identification of S-nitrosylated proteins in endotoxin-stimulated RAW264.7 murine macrophages,” *Nitric Oxide*, vol. 12, no. 2, pp. 121–126, 2005.
- [75] C. Wu, T. Liu, W. Chen et al., “Redox regulatory mechanism of transnitrosylation by thioredoxin,” *Molecular & Cellular Proteomics*, vol. 9, no. 10, pp. 2262–2275, 2010.
- [76] K.-C. Chou, “Impacts of bioinformatics to medicinal chemistry,” *Medicinal Chemistry*, vol. 11, no. 3, pp. 218–234, 2015.
- [77] X. Liang, W. Zhu, Z. Lv, and Q. Zou, “Molecular computing and bioinformatics,” *Multidisciplinary Digital Publishing Institute*, vol. 24, no. 13, article 2358, 2019.
- [78] Z. Lv, C. Ao, and Q. Zou, “Protein function prediction: from traditional classifier to deep learning,” *Proteomics*, vol. 19, no. 14, article 1900119, 2019.
- [79] Z. Lv, P. Wang, Q. Zou, and Q. Jiang, “Identification of sub-Golgi protein localization by use of deep representation learning features,” *Bioinformatics*, 2020.
- [80] G. Liu, S. Jin, Y. Hu, and Q. Jiang, “Disease status affects the association between rs4813620 and the expression of Alzheimer’s disease susceptibility gene TRIB3,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 45, pp. E10519–E10520, 2018.
- [81] X. Zhao, Q. Jiao, H. Li et al., “ECFS-DEA: an ensemble classifier-based feature selection for differential expression analysis on expression profiles,” *BMC Bioinformatics*, vol. 21, no. 1, pp. 43–43, 2020.
- [82] K. Liu and W. Chen, “iMRM: a platform for simultaneously identifying multiple kinds of RNA modifications,” *Bioinformatics*, vol. 36, no. 11, pp. 3336–3342, 2020.
- [83] D. Zhang, Z. C. Xu, W. Su et al., “iCarPS: a computational tool for identifying protein carbonylation sites by novel encoded features,” *Bioinformatics*, 2020.
- [84] H. Lv, F.-Y. Dao, Z.-X. Guan, H. Yang, Y. W. Li, and H. Lin, “Deep-Kcr: accurate detection of lysine crotonylation sites using deep learning method,” *Briefings in Bioinformatics*, 2020.
- [85] K. C. Chou, “Prediction of protein cellular attributes using pseudo-amino acid composition,” *Proteins*, vol. 43, no. 3, pp. 246–255, 2001.
- [86] J. Hong, Y. Luo, Y. Zhang et al., “Protein functional annotation of simultaneously improved stability, accuracy and false discovery rate achieved by a sequence-based deep learning,” *Briefings in Bioinformatics*, vol. 21, no. 4, pp. 1437–1447, 2020.

- [87] J. Hong, Y. Luo, M. Mou et al., "Convolutional neural network-based annotation of bacterial type IV secretion system effectors with enhanced accuracy and reduced false discovery," *Briefings in Bioinformatics*, vol. 21, no. 5, pp. 1825–1836, 2020.
- [88] Y. Xue, J. Ren, X. Gao, C. Jin, L. Wen, and X. Yao, "GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy," *Molecular & Cellular Proteomics*, vol. 7, no. 9, pp. 1598–1608, 2008.
- [89] Z. Lv, D. Wang, H. Ding, B. Zhong, and L. Xu, "Escherichia coli DNA N-4-methylcytosine site prediction accuracy improved by light gradient boosting machine feature selection technology," *Ieee Access*, vol. 8, pp. 14851–14859, 2020.
- [90] Z. Lv, J. Zhang, H. Ding, and Q. Zou, "RF-PseU: a random forest predictor for RNA pseudouridine sites," *Frontiers in Bioengineering and Biotechnology*, vol. 8, p. 134, 2020.
- [91] R. Su, J. Hu, Q. Zou, B. Manavalan, and L. Wei, "Empirical comparison and analysis of web-based cell-penetrating peptide prediction tools," *Briefings in Bioinformatics*, vol. 21, no. 2, pp. 408–420, 2020.
- [92] R. Su, X. Liu, L. Wei, and Q. Zou, "Deep-Resp-Forest: a deep forest model to predict anti-cancer drug response," *Methods*, vol. 166, pp. 91–102, 2019.
- [93] R. Su, X. Liu, G. Xiao, and L. Wei, "Meta-GDBP: a high-level stacked regression model to improve anticancer drug response prediction," *Briefings in Bioinformatics*, vol. 21, no. 3, pp. 996–1005, 2020.
- [94] R. Su, H. Wu, B. Xu, X. Liu, and L. Wei, "Developing a multi-dose computational model for drug-induced hepatotoxicity prediction based on toxicogenomics data," *Ieee-Acm Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 4, pp. 1231–1239, 2019.
- [95] L. Wei, J. Hu, F. Li, J. Song, R. Su, and Q. Zou, "Comparative analysis and prediction of quorum-sensing peptides using feature representation learning and machine learning algorithms," *Briefings in Bioinformatics*, vol. 21, no. 1, pp. 106–119, 2018.
- [96] L. Wei, M. Liao, Y. Gao, R. Ji, Z. He, and Q. Zou, "Improved and promising identification of human microRNAs by incorporating a high-quality negative set," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 11, no. 1, pp. 192–201, 2014.
- [97] L. Wei, S. Wan, J. Guo, and K. K. L. Wong, "A novel hierarchical selective ensemble classifier with bioinformatics application," *Artificial Intelligence in Medicine*, vol. 83, pp. 82–90, 2017.
- [98] L. Wei, P. Xing, J. Zeng, J. X. Chen, R. Su, and F. Guo, "Improved prediction of protein-protein interactions using novel negative samples, features, and an ensemble classifier," *Artificial Intelligence in Medicine*, vol. 83, pp. 67–74, 2017.
- [99] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: a deformable network for retinal vessel segmentation," *Knowledge-Based Systems*, vol. 178, pp. 149–162, 2019.
- [100] L. Wei, C. Zhou, H. Chen, J. Song, and R. Su, "ACPred-FL: a sequence-based predictor using effective feature representation to improve the prediction of anti-cancer peptides," *Bioinformatics*, vol. 34, no. 23, pp. 4007–4016, 2018.
- [101] G. Liu, Y. Hu, S. Jin, F. Zhang, Q. Jiang, and J. Hao, "Cis-eQTLs regulate reduced LST1 gene and NCR3 gene expression and contribute to increased autoimmune disease risk," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 42, pp. E6321–E6322, 2016.
- [102] G. Wang, Y. Wang, W. Feng et al., "Transcription factor and microRNA regulation in androgen-dependent and -independent prostate cancer cells," *BMC Genomics*, vol. 9, article S22, Supplement 2, 2008.
- [103] Y. Zhao, F. Wang, and L. Juan, "MicroRNA promoter identification in Arabidopsis using multiple histone markers," *BioMed Research International*, vol. 2015, Article ID 861402, 10 pages, 2015.
- [104] Y. Zhao, F. Wang, S. Chen, J. Wan, and G. Wang, "Methods of microRNA promoter prediction and transcription factor mediated regulatory network," *BioMed Research International*, vol. 2017, Article ID 7049406, 8 pages, 2017.
- [105] W. Chen, P. Feng, and F. Nie, "iATP: a sequence based method for identifying anti-tubercular peptides," *Medicinal Chemistry*, vol. 16, no. 5, pp. 620–625, 2020.
- [106] T. Denoeux, "A k-nearest neighbor classification rule based on Dempster-Shafer theory," in *Classic works of the Dempster-Shafer theory of belief functions*, pp. 737–760, Springer, 2008.
- [107] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [108] Z. He, J. Zhang, X.-H. Shi et al., "Predicting drug-target interaction networks based on functional groups and biological features," *PLoS One*, vol. 5, no. 3, article e9603, 2010.
- [109] F. Li, C. Li, M. Wang et al., "GlycoMine: a machine learning-based approach for predicting N-, C-and O-linked glycosylation in the human proteome," *Bioinformatics*, vol. 31, no. 9, pp. 1411–1419, 2015.
- [110] F. Li, C. Li, J. Revote et al., "GlycoMine struct: a new bioinformatics tool for highly accurate mapping of the human N-linked and O-linked glycoproteomes by incorporating structural features," *Scientific Reports*, vol. 6, no. 1, pp. 1–16, 2016.
- [111] J. Qian, D. Q. Miao, Z. Zhang, and W. Li, "Hybrid approaches to attribute reduction based on indiscernibility and discernibility relation," *International Journal of Approximate Reasoning*, vol. 52, no. 2, pp. 212–230, 2011.
- [112] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [113] D. Quang, Y. Chen, and X. Xie, "DANN: a deep learning approach for annotating the pathogenicity of genetic variants," *Bioinformatics*, vol. 31, no. 5, pp. 761–763, 2015.
- [114] Q. Yang, B. Li, J. Tang et al., "Consistent gene signature of schizophrenia identified by a novel feature selection strategy from comprehensive sets of transcriptomic data," *Briefings in Bioinformatics*, vol. 21, no. 3, pp. 1058–1068, 2020.
- [115] Z. Hong, X. Zeng, L. Wei, and X. Liu, "Identifying enhancer-promoter interactions with neural network based on pre-trained DNA vectors and attention mechanism," *Bioinformatics*, vol. 36, no. 4, pp. 1037–1043, 2020.
- [116] J. Tang, M. Mou, Y. Wang, Y. Luo, and F. Zhu, "MetaFS: performance assessment of biomarker discovery in metaproteomics," *Briefings in Bioinformatics*, 2020.
- [117] S. Jin, X. Zeng, F. Xia, W. Huang, and X. Liu, "Application of deep learning methods in biological networks," *Briefings in Bioinformatics*, 2020.

- [118] J. Tang, J. Fu, Y. Wang et al., "Simultaneous improvement in the precision, accuracy, and robustness of label-free proteome quantification by optimizing data manipulation chains\*," *Molecular & Cellular Proteomics*, vol. 18, no. 8, pp. 1683–1699, 2019.
- [119] Y.-J. Tang, Y.-H. Pang, and B. Liu, "IDP-Seq2Seq: identification of intrinsically disordered regions based on sequence to sequence learning," *Bioinformatics*.
- [120] C.-C. Li and B. Liu, "MotifCNN-fold: protein fold recognition based on fold-specific features extracted by motif-based convolutional neural networks," *Briefings in Bioinformatics*, vol. 21, no. 6, pp. 2133–2141, 2020.
- [121] Z. Lv, H. Ding, L. Wang, and Q. Zou, "A convolutional neural network using dinucleotide one-hot encoder for identifying DNA N6-methyladenine sites in the rice genome," *Neurocomputing*, vol. 422, pp. 214–221, 2021.
- [122] Z. Wang, W. He, J. Tang, and F. Guo, "Identification of highest-affinity binding sites of yeast transcription factor families," *Journal of Chemical Information and Modeling*, vol. 60, no. 3, pp. 1876–1883, 2020.
- [123] H. Wang, Y. Ding, J. Tang, and F. Guo, "Identification of membrane protein types via multivariate information fusion with Hilbert-Schmidt Independence criterion," *Neurocomputing*, vol. 383, pp. 257–269, 2020.
- [124] J. Li, Y. Pu, J. Tang, Q. Zou, and F. Guo, "DeepAVP: a dual-channel deep neural network for identifying variable-length antiviral peptides," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 3012–3019, 2020.
- [125] Y. Shen, J. Tang, and F. Guo, "Identification of protein sub-cellular localization via integrating evolutionary and physicochemical information into Chou's general PseAAC," *Journal of Theoretical Biology*, vol. 462, pp. 230–239, 2019.
- [126] G. Liu, F. Zhang, Y. Jiang et al., "Integrating genome-wide association studies and gene expression data highlights dys-regulated multiple sclerosis risk pathways," *Multiple Sclerosis*, vol. 23, no. 2, pp. 205–212, 2017.
- [127] X. Zeng, S. Zhu, Y. Hou et al., "Network-based prediction of drug-target interactions using an arbitrary-order proximity embedded deep forest," *Bioinformatics*, vol. 36, no. 9, pp. 2805–2812, 2020.
- [128] Y. H. Li, X. X. Li, J. J. Hong et al., "Clinical trials, progression-speed differentiating features and swiftness rule of the innovative targets of first-in-class drugs," *Briefings in Bioinformatics*, vol. 21, no. 2, pp. 649–662, 2020.
- [129] X. Fu, L. Cai, X. Zeng, and Q. Zou, "StackCPPred: a stacking and pairwise energy content-based prediction of cell-penetrating peptides and their uptake efficiency," *Bioinformatics*, vol. 36, no. 10, pp. 3028–3034, 2020.
- [130] J. Yin, W. Sun, F. Li et al., "VARIDT 1.0: variability of drug transporter database," *Nucleic Acids Research*, vol. 48, no. D1, pp. D1042–D1050, 2020.
- [131] M. Wang, X. Cui, B. Yu, C. Chen, Q. Ma, and H. Zhou, "Sul-Site-GTB: identification of protein S-sulfonylation sites by fusing multiple feature information and gradient tree boosting," *Neural Computing and Applications*, vol. 32, no. 17, pp. 13843–13862, 2020.
- [132] Y. Ding, J. Tang, and F. Guo, "Identification of drug-side effect association via multiple information integration with centered kernel alignment," *Neurocomputing*, vol. 325, pp. 211–224, 2019.
- [133] Y. Ding, J. Tang, and F. Guo, "The computational models of drug-target interaction prediction," *Protein and Peptide Letters*, vol. 27, no. 5, pp. 348–358, 2020.
- [134] Y. Ding, J. Tang, and F. Guo, "Identification of drug-target interactions via fuzzy bipartite local model," *Neural Computing & Applications*, vol. 32, no. 14, pp. 10303–10319, 2020.
- [135] V. Vacic, L. M. Iakoucheva, and P. Radivojac, "Two Sample Logo: a graphical representation of the differences between two sets of sequence alignments," *Bioinformatics*, vol. 22, no. 12, pp. 1536–1537, 2006.
- [136] D.-S. Cao, Q.-S. Xu, and Y.-Z. Liang, "propy: a tool to generate various modes of Chou's PseAAC," *Bioinformatics*, vol. 29, no. 7, pp. 960–962, 2013.
- [137] B. Liu, X. Wang, L. Lin, Q. Dong, and X. Wang, "A discriminative method for protein remote homology detection and fold recognition combining Top-n-grams and latent semantic analysis," *BMC Bioinformatics*, vol. 9, no. 1, p. 510, 2008.
- [138] Q. Zou, J. Zeng, L. Cao, and R. Ji, "A novel features ranking metric with application to scalable visual and bioinformatics data classification," *Neurocomputing*, vol. 173, pp. 346–354, 2016.