*2025 International Conference on Applied Intelligence*
*November 6-9, Nanning, China*
http://www.icai.org.cn/2025/index.php

# Prediction of Post-Translational Modification Sites of Paeonia lactiflora Proteins Based on Attention Mechanism

Yingyue Tang[1] and Wenzheng Bao[2,3]

[1] School of Information Engineering, Yancheng Institute of Technology, Yancheng, China 224051
[2] Institute for Regenerative Medicine, Medical Innovation Center and State Key Laboratory of Cardiology, Shanghai East Hospital, School of Medicine, Tongji University, Shanghai 200123, P. R. China

[3] Xuzhou University of Technology, Xuzhou, China, 221018

**Abstract.** The precise prediction of post-translational modification sites of proteins was an important aspect of elucidating the mechanisms of traditional Chinese medicine. Existing computational models struggled to meet the needs of herbal pharmacological research due to their neglect of the specificity of herbal materials and the complexity of modification sites. This study focused on the traditional Chinese medicinal material Paeonia lactiflora and constructed a deep learning model based on convolutional neural networks and attention mechanisms by integrating multi-dimensional features of amino acid sequences. A dataset for Paeonia lactiflora was built using the TCMSP database, which included 1080 positive samples and 1976 negative samples, and the input space was optimized through feature normalization and dimensionality reduction. Experimental results indicated that the model effectively captured the modification patterns of proteins from different herbal materials, and SHAP feature selection significantly improved the prediction accuracy of post-translational modification sites. Compared to traditional single algorithm models, the proposed integrated architecture demonstrated significant advantages in balancing sequence conservation and functional specificity, providing a new computational tool for elucidating the mechanisms of action of active components in traditional Chinese medicinal materials and guiding rational clinical medication.

**Keywords:** Post-translational Modification Sites, Attention Mechanism, Transfer Learning, Paeonia lactiflora

## 1 Introduction

Paeonia, as a traditional medicinal herb widely used in Chinese medicine, received significant attention due to its various pharmacological activities [1][2]. Recent studies

indicated that Paeonia exhibited remarkable efficacy in treating oral diseases, particularly positively influencing conditions such as xerostomia [3][4]. The active components contained not only possessed anti-inflammatory and analgesic properties but also demonstrated hepatoprotective and antioxidant functions, which made Paeonia play an important role in alleviating oral discomfort and improving quality of life [5].

Especially among patients with Sjögren's syndrome, Paeonia effectively promoted saliva secretion, thereby alleviating dry mouth symptoms and providing tangible assistance to patients [6-8]. This not only improved the patients' oral health but also further enhanced their overall quality of life. Additionally, research found that formulations related to white Paeonia could significantly enhance therapeutic effects when used in conjunction with chemical drugs while reducing drug toxicity [9]. This finding not only provided new ideas for the rational combination of clinical medications but also opened new avenues for the treatment of autoimmune diseases [10].

Therefore, in-depth exploration of the pharmacological mechanisms of Paeonia and its potential in clinical applications not only aided in better understanding its status in traditional Chinese medicine but also provided important references for the advancement of modern medicine [11][12]. As research on Paeonia deepened, it was anticipated that more innovative therapies based on its active components would emerge, bringing new hope for the treatment of various diseases [13][14].

Post-Translational Modifications (PTMs) of proteins referred to a series of chemical modification processes that occurred after protein synthesis [15][16]. These modifications involved various biochemical reactions, forming a complex and precise network, which not only influenced protein function and stability but also played key roles in biological processes such as cell signaling and metabolic regulation [17-18]. Thus, in-depth understanding of PTMs was crucial for revealing protein diversity and its biological functions. This indicated the unique complexity of PTMs in maintaining biological activity and mediating various biological interactions [19]. The biosynthetic mechanisms of tanshinones and phenolic acids in Salvia miltiorrhiza suggested that these modifications not only affected protein structure and function but also played significant regulatory roles in the mechanisms of action between white Paeonia-derived compounds and various diseases [20]. For instance, certain PTMs might enhance their antitumor or anti-inflammatory efficacy, providing new insights for exploring the modern applications of traditional Chinese medicine [21].

Although traditional experimental methods for studying PTMs, such as mass spectrometry, Western blotting, immunoprecipitation, and flow cytometry, had developed relatively maturely, these methods often proved time-consuming, costly in terms of equipment, and complex in process [22-23]. This made traditional experimental methods increasingly inadequate in the face of the growing amount of plant protein data. Therefore, to more efficiently process and analyze this data, AI-based predictive methods emerged [24].

In recent years, deep learning algorithms made significant contributions to bioinformatics. In the field of drug discovery, Hakime Öztürk et al. proposed a deep learning-based model, DeepDTA, which effectively predicted drug-target binding affinity using target and drug sequence information, overcoming limitations in previous methods [25]. Rohan Gupta et al. reviewed the extensive applications of AI and

machine learning in drug design, highlighting the modernizing impact of new technologies on the drug discovery process [26]. Ehsaneddin Asgari et al. introduced the BioVec biological sequence representation method, showcasing the potential of deep learning in proteomics and genomics research [27]. The P2Rank tool developed by Krivák and Hockx provided an efficient solution for rapid and accurate prediction of ligand binding sites using machine learning [28]. In the prediction of phosphorylation sites, the MusiteDeep framework proposed by Duolin Wang et al. significantly improved prediction accuracy through convolutional neural networks and attention mechanisms [29]. The DEEPre model by Li, Yu et al. enhanced enzyme function prediction performance through automated feature selection and classification training methods [30]. Schubach et al. developed the CADD v1.7 version, which integrated new annotation features to further improve the accuracy of genome-wide variant predictions, providing important tools for related research [31]. These studies collectively propelled advancements in the fields of drug discovery and bioinformatics.

Currently, attention mechanisms are widely used in deep learning due to their ability to reduce information redundancy, handle long sequence data, and possess high interpretability. In the field of drug-protein interaction prediction, Dayu Tan et al. constructed a dual-channel neural network model, DCA-DPI, which utilized drug molecular graphs and protein sequences as inputs, learning the feature representations of drugs and proteins through residual graph neural networks and residual convolutional networks, combined with efficient path attention mechanisms, thereby improving the accuracy of DPI predictions [32]. Ying Xu and Jinyong Cheng proposed a protein secondary structure prediction model based on multi-scale convolutional attention neural networks, aiming to effectively extract both local and long-range information from amino acid sequences, enhancing the reconstruction of feature maps through a multi-channel multi-scale parallel architecture, ultimately improving prediction outcomes [33]. The deep learning tool DLBWE-Cys developed by Zhengtao Luo et al. aimed to identify cysteine S-carboxyethylation sites, combining CNN, BiLSTM, Bahdanau attention mechanism, and fully connected neural networks, with experimental results showing its superior performance in cross-validation and independent testing compared to other models, validating the effectiveness of its encoding method [34]. Yuhao He et al. established a novel sound-based COVID-19 diagnostic framework, TFA-CLSTMNN, which significantly improved detection accuracy by extracting time-frequency domain features from cough recordings and integrating attention convolutional long short-term memory neural networks, achieving an accuracy of over 0.95 on public real-world datasets [35].

Although these emerging methods improved prediction efficiency to some extent, they still struggled to comprehensively meet all demands for PTM site prediction, necessitating the urgent development of new models and algorithms to enhance prediction accuracy and broad applicability [36].

To address the above issues, our experimental contributions included the following points:

(1) A deep learning model based on convolutional neural networks and attention mechanisms was constructed, where the convolutional neural network was responsible for extracting local information, and global attention could establish long-distance

interactions for each feature information; randomly generated orthogonal filters could adaptively adjust channel importance, and we preliminarily explored the impact of different attention mechanisms on encoding results.

(2) The prediction accuracy of PTM sites was significantly improved through SHAP feature selection. Compared to traditional single-algorithm models, our proposed ensemble architecture exhibited significant advantages in balancing sequence conservation and functional specificity.

(3) A dataset integrating various encoding features was created to verify the impact of encoding before and after feature selection on the model.

# 2 Feature input

## 2.1 Dataset Construction

The data used in this study was sourced from the Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform (TCMSP), which provided extensive information regarding the relationships between various herbal targets and diseases. We segmented the complete protein sequence of Paeonia lactiflora (Ps) according to Zhou's formula to obtain protein sequence fragments (Pss), which can be described by formula (1) [37]。

$$P_{ss} = [A_{-20}, A_{-19}, \ldots A_{-2}, A_{-1}, A_0, A_1, A_2, \ldots A_{19}, A_{20}], \quad P_{ss} \subseteq P_s \tag{1}$$

The amino acid at the center position A0 of Pss determined the positive or negative nature of the fragment. We used the amino acids at post-translational modification sites as A0, with the resulting protein sequence fragments cut from this position serving as positive examples. Specifically, we took 20 consecutive amino acids both forward and backward from A0, resulting in Pss having a length of 41 amino acids. Next, we identified the amino acids adjacent to the center position A0, namely A-1 and A1. If A-1and A1 were non-post-translational modification site amino acids, we redefined A-1 and A1 as the new center position A0 for the negative example protein fragment segmentation. During this process, if the center amino acid A0 of Pss was too close to the beginning or end of the protein sequence, it would lead to the problem of the protein sequence fragment being too short. To address this issue, we supplemented the sequence by adding the most homologous amino acid residues.

In the end, we obtained a dataset for Paeonia lactiflora consisting of 1,080 positive samples and 1,976 negative samples.

## 2.2 Amino acid encoding

### 2.2.1 Block Substitution Matrix

The Block Substitution Matrix (BLOSUM62) utilized a predefined amino acid substitution probability matrix to map individual amino acid residues to numerical

vectors[38]. Specifically, each residue was represented by an m×n element matrix, where n denoted the length of the peptide chain and m represented the number of amino acid types. Each residue was represented by a row in the matrix, allowing it to reflect the evolutionary information of proteins; however, this method also fixed it to encode equally long peptide segments.

### 2.2.2 Composition of k-spaced amino acid group pairs

The k-spaced amino acid group pair (CKSAAGP) was an improved version of the CKSAAP descriptor[39]. This metric quantified features by statistically measuring the frequency of occurrence of group pairs separated by any k amino acid residues, aiding in the capture of distant correlation information. The cksaagp feature vector was defined by equation (2).

$$\left(\frac{N_{g1g1}}{N_{total}},\frac{N_{g1g2}}{N_{total}},\frac{N_{g1g3}}{N_{total}},\ldots,\frac{N_{g5g5}}{N_{total}}\right)_{25} \tag{2}$$

The value of each descriptor represented the composition of the corresponding residue pairs in the protein or peptide sequence, where $N_{total}$ was the total number of spaced residue pairs in the protein. When k=0, there were 25 types of 0-spaced group pairs (i.e., g1g1, g1g2, g1g3, ..., g5g5). When the length of the protein was P and k=0, 1, 2, 3, 4, and 5, the corresponding $N_{total}$ values were P-1, P-2, P-3, P-4, P-5, and P-6, respectively.

### 2.2.3 Adaptive skip dinucleotide composition

Adaptive skip dinucleotide composition (ASDC) fully considered the correlation information between adjacent residues, including those between intervening residues, balancing the associations of both adjacent and non-adjacent residues[40]. Its feature vector was represented by equation (3).

$$\begin{cases} ASDC = (f_{v1}, f_{v1}, \ldots, f_{v400}), \\ f_{vi} = \dfrac{\sum_{g=1}^{L-1} O_i^g}{\sum_{i=1}^{400}\sum_{g=1}^{L-1} O_i^g} \end{cases} \tag{3}$$

where fvi denoted the frequency of occurrence of all possible dipeptides (with intervening nucleotide numbers ≤ L-1).

### 2.2.4 Enhanced amino acid composition

Enhanced amino acid composition (EAAC) employed a fixed-length sliding window that moved along the peptide chain, dynamically calculating the proportion of each amino acid within the window[41]. Specifically, EAAC slid continuously from the N-terminus to the C-terminus of each peptide chain to compute the amino acid composition encoding (AAC), where the ACC formula was as shown in equation (4), and the EAAC formula was as indicated in equation (5).

$$f(t) = \frac{N(t)}{N}, t \in \{A, C, D, \ldots, Y\} \tag{4}$$

AAC encoding calculated the occurrence frequency of the 20 natural amino acids ("ACDEFGHIKLMNPQRSTVWY") in the protein or peptide sequence. In this context, N(t) represented the quantity of amino acid type t, while N denoted the length of the protein or peptide sequence.

$$f(t,win) = \frac{N(t,win)}{N(win)}, t \in \{A, C, D, \ldots Y\}, \quad win \in \{window1, window2, window3, \ldots window17\} \tag{5}$$

In this context, $N(t,win)$ represented the quantity of amino acid type t within the sliding window win, while $N(win)$ denoted the size of the sliding window.

### 2.2.5 DPC (Di-Peptide Composition)

The Di-peptide Composition, as shown in equation (6), comprised 400 descriptors, where Nrs represented the quantity of dipeptides corresponding to amino acid types r and s. This method characterized the overall dipeptide abundance in protein sequences; however, its spatial order was disrupted and was often applied in scenarios requiring rapid global feature extraction.

$$D(r,s) = \frac{N_{rs}}{N-1}, \quad r,s \in \{A, C, D, \ldots Y\}, \tag{6}$$

## 3 Model construction

This study integrated various amino acid sequence-based feature extraction methods to create a classification model using Convolutional Neural Networks (CNN) and attention mechanisms, classifying the fused features. The flowchart is shown in Figure 1.

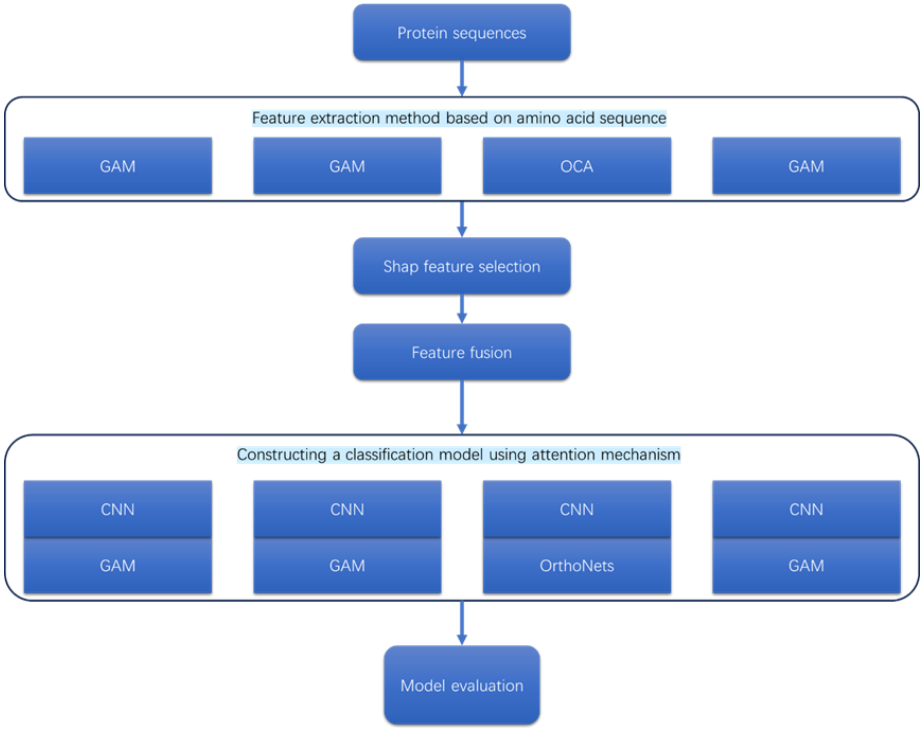


**Fig. 1.** Experimental flow chart

## 3.1 Attention Mechanisms

### 3.1.1 Global Attention Mechanism

Global Attention Mechanism (GAM) was a mechanism capable of reducing information loss and enhancing global interaction information[42], represented by equations (7) and (8). Originally designed for computer vision tasks, in the context of PTM site prediction, protein sequences could be encoded as multi-channel features. The channel attention sub-module of GAM dynamically allocated the importance of different feature channels, with the potential to automatically filter key biochemical signals. In contrast, traditional CNNs faced limitations in modeling long-range residue associations due to the constraints of convolution kernel size (such as interactions between enzyme active sites and substrate binding regions). The spatial attention sub-module of GAM, through global context awareness, could directly establish associations between any two residues to address this limitation.

$$F_2 = M_c(F_1) \otimes F_1 \tag{7}$$

$$F_3 = M_s(F_2) \otimes F_2 \tag{8}$$

In this context, $F_1 \in \mathbb{R}^{C\times H\times W}$ represented the given input feature map, $F_2$ denoted the intermediate state, and $F_3$ indicated the output result. $M_c$ and $M_s$ represented the channel and spatial attention maps, respectively, while $\otimes$ indicated element-wise multiplication.

GAM included two important sub-modules: the channel attention sub-module and the spatial attention sub-module. The channel attention sub-module utilized 3D permutation to retain information across three dimensions, and then amplified the inter-dimensional channel-space dependencies through a two-layer multi-layer perceptron (MLP). The spatial attention sub-module employed two convolutional layers to aggregate spatial information, with the reduction rate r being the same as that of the channel attention sub-module. Group convolution and channel shuffling were implemented to prevent a significant increase in parameters after de-pooling[43].

### 3.1.2 Polarized Self-Attention

The Polarized Self-Attention (PSA) block comprised two components: a pure channel module and a pure spatial module[44]. It accomplished filtering by fully collapsing features in one direction while maintaining high resolution in its orthogonal direction. The dynamic range of attention was increased by performing Softmax normalization at the bottleneck tensor (the minimal feature tensor within the attention block), using the Sigmoid function for tone mapping.

The pure channel module maintained high internal resolution during the calculations of channel and spatial attention while fully collapsing the corresponding dimensions of the input tensor $A^{ch}(X) \in \mathbb{R}^{C\times 1\times 1}$, which could be represented by the equation (9).

$$A^{ch}(X) = F_{SG}\left[W_{z|\theta_1}\left(\sigma_1\left(W_v(X)\right) \times Soft\,max\left(\sigma_2\left(W_q(X)\right)\right)\right)\right] \qquad (9)$$

In this context, $W_q$, $W_v$, and $W_z$ were each a 1 × 1 convolution layer, while $\sigma_1$ and $\sigma_2$ were two tensor reshaping operators. The symbol "×" represented the matrix dot product operation $Soft\,max(X) = \sum_{j=1}^{N_p} \frac{e^{xj}}{\sum_{m=1}^{N_p} e^{xm}} x_j$. The number of internal channels between $W_v$ and $W_q|W_z$ was C/2. The output of the channel branch alone was $Z^{ch} = A^{ch}(X) \odot^{ch} X \in \mathbb{R}^{C\times H\times W}$, where $\odot^{ch}$ was the channel multiplication operator.

The pure spatial module was capable of combining non-linearity to directly adapt to the output distribution of typical fine regression $A^{sp}(X) \in \mathbb{R}^{1\times H\times W}$, as represented by the following equation (10).

$$A^{sp}(X) = F_{SG}\left[\sigma_3\left(Soft\,max\left(\sigma_1\left(F_{GP}\left(W_q(X)\right)\right)\right) \times \sigma_2\left(W_v(X)\right)\right)\right] \qquad (10)$$

In this context, $W_q$ and $W_v$ were standard 1 x 1 convolution layers, while $\sigma_1$, $\sigma_2$, and $\sigma_3$ were three tensor reshaping operators. $F_{GP}(\cdot)$ was a global pooling operator $F_{GP}(X) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X(:,i,j)$, and "×" represented the matrix dot product operation. The output of the pure spatial branch was $Z^{sp} = A^{sp}(X) \odot^{sp} X \in \mathbb{R}^{C\times H\times W}$, where $\odot^{sp}$ was the spatial multiplication operator.

The outputs of the pure channel module and the pure spatial module could be arranged in either a parallel layout or a serial layout. In this case, a parallel layout method was adopted, as represented by the following equation (11).

$$PSA_p(X) = Z^{ch} + Z^{sp} = A^{ch}(X) \odot^{ch} X + A^{sp}(X) \odot^{sp} X \tag{11}$$

### 3.1.3 Multi-Attention

The MAB module was originally designed to address the single-image super-resolution (SR) problem, which is an important task in the field of computer vision aimed at reconstructing missing high-frequency information from low-quality inputs[45]. We could simply understand this attention module as a form of information completion. Since amino acid sequence information does not contain structural information about the amino acids, and the process of feature extraction may lead to the loss of inter-relationships between various amino acids, this completion of information is theoretically reasonable for the prediction of amino acid sequences.

The MAB module internally included Multi-Scale Large Kernel Attention (MLKA) and Gated Spatial Attention Units (GSAU), which could be represented by the following equation (12).

$$\begin{cases} N = LN(X) \\ X = X + \lambda_1 f_3\left(MLKA\left(f_1(N)\right) \otimes f_2(N)\right) \\ N = LN(X) \\ X = X + \lambda_2 f_6\left(GSAU\left(f_4(N), f_5(N)\right)\right) \end{cases} \tag{12}$$

Where $LN(\cdot)$ and $\lambda$ represented layer normalization and learnable scaling factors, respectively. $MLKA(\cdot)$ and $GSAU(\cdot)$ denoted the MLKA and GSAU modules, respectively. $\otimes$ indicated element-wise multiplication, and $f_i(\cdot)$ represented the i-th pointwise convolution that maintained dimensions. To preserve instance details and accelerate convergence, MAB utilized layer normalization instead of batch normalization or no normalization. MLKA combined large kernel attention (LKA) and a multi-scale mechanism, dynamically adjusting the attention maps through a gating mechanism to avoid potential block artifacts. This could be expressed using the following formula (13-15).

$$LKA(X) = f_{PW}\left(f_{DWD}\left(f_{DW}(X)\right)\right) \tag{13}$$

$$MLKA_i(X_i) = G_i(X_i) \otimes LKA_i(X_i) \tag{14}$$

$$GSAU(X, Y) = f_{DW}(X) \otimes Y \tag{15}$$

Given the input feature map $X \in \mathbb{R}^{C \times H \times W}$, LKA decomposed the convolution into three components to establish long-range relationships, including depthwise convolution, depthwise dilated convolution, and pointwise convolution, using K×K. Specifically, LKA achieved this by decomposing into one (2d-1)×(2d-1) depthwise

convolution, one $\left\lceil \frac{K}{d} \times \frac{K}{d} \right\rceil$ depthwise dilated convolution $f_{DWD}(\cdot)$, and one pointwise convolution $f_{DW}(\cdot)$. GSAU simplified the feedforward network through spatial attention and a gating mechanism, reducing computational costs while incorporating spatial information.

### 3.1.4 Orthogonal Channel Attention

Orthogonal Channel Attention (OrthoNets) randomly generated orthogonal filters and utilized an attention mechanism to weight the channels while incorporating residual connections, aiming to enhance the neural network's ability to process and represent input features[46]. This design integrated various techniques, including randomization, orthogonalization, attention mechanisms, and residual connections. The random generation of orthogonal filters introduced a certain degree of randomness and diversity, while the attention mechanism adaptively adjusted the importance of the channels. The residual connections helped alleviate issues such as gradient vanishing in deep networks, collectively enhancing the model's performance and generalization capability.

The filters were represented by $A(X) = E\left(F_{ortho}(X)\right)$, and the squeeze process was expressed using formula $F_{\mathrm{ortho}}(X)_c = \sum_{h=1}^{H} \sum_{w=1}^{W} K_{c,h,w} X_{c,h,w}$. The channel attention formula was denoted by $A(X) = E\left(F_{ortho}(X)\right)$.

In the first stage, random filters were initialized, with sizes matching those of the feature layer. The Gram-Schmidt process was used to orthogonalize these filters. Random numbers entered the Gram-Schmidt module, and through the Gram-Schmidt orthogonalization process, orthogonal filters were generated, with dimensions of H×W×C, where H×W×C represented height, width, and the number of channels, respectively.

In the second stage, the filters were utilized to extract compressed vectors, and the excitation proposed by SENet was employed to obtain attention vectors. By multiplying the attention vectors with the input features, the weighted output features were computed, and the residuals were added.

In the third stage, the input data was processed through convolution blocks, resulting in input feature maps with dimensions of H×W×C. The orthogonal filters from Stage 0 were operated on the input feature maps, and after a squeeze operation, the feature maps were compressed to dimensions of 1×1×C. The squeezed features underwent an excitation operation to generate attention weights. The attention weights were then multiplied by the input feature maps processed through the convolution blocks, resulting in weighted feature maps.

The residual connection from the original input X was added to the weighted feature maps, ultimately yielding the output with dimensions of H×W×C.

## 3.2 SHAP

Shapley Additive exPlanations (SHAP) was a model interpretation method based on game theory, aimed at providing clear explanations for the outputs of machine learning models[47-48]. It fairly distributed the contributions of features to the prediction results

using the classical Shapley values, helping us understand the decision-making process of the model.

When using SHAP for interpretation, it was first necessary to create an explainer that supported various model types, such as tree models and deep learning models. We utilized the default XGBoost model. The calculation of SHAP values helped us understand the impact of each feature on the model's predictions. SHAP provided several visualization tools, such as Summary Plot, Bar Plot, and Beeswarm Plot, which intuitively displayed the importance of features and their effects on the prediction results, as illustrated in Figure 2.

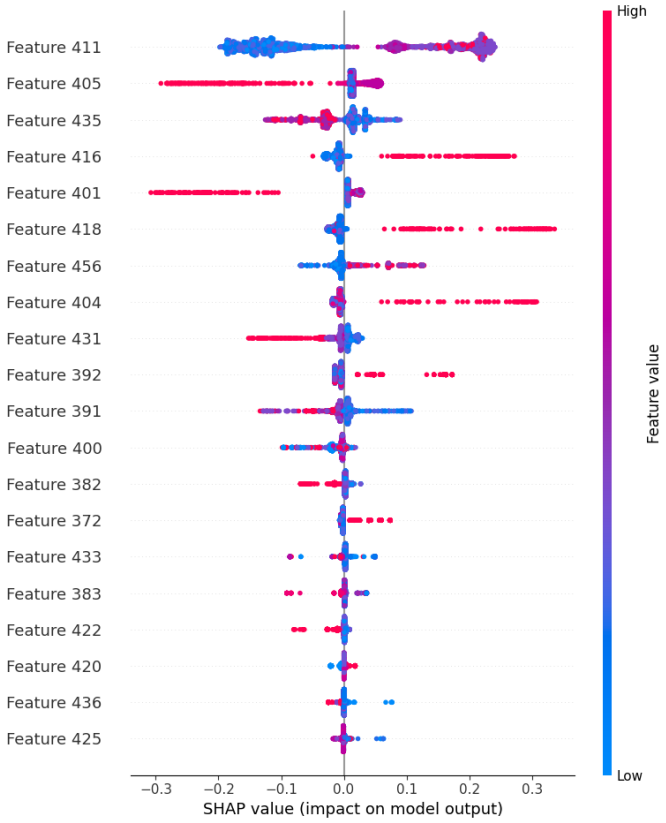


**Fig. 2.** SHAP Performance on BLOSUM62 Features

These visualizations not only helped identify important features but also revealed the interactions between features.

### 3.3 Model evaluation

We used accuracy (ACC), sensitivity (SN), specificity (SP), F1 score (F1), and Matthews correlation coefficient (MCC) to evaluate model performance[49] (see

formulas (16-20)). We took the average from ten-fold cross-validation to present the final results. In the prediction process, true positives (TP) referred to the positive samples correctly identified by the model, while true negatives (TN) referred to the negative samples accurately recognized by the model. False positives (FP) indicated negative samples incorrectly classified as positive, and false negatives (FN) indicated positive samples incorrectly classified as negative[50]. Accuracy (ACC) reflected the overall predictive accuracy of the samples.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

SN and SP represented the predictive accuracy of negative and positive samples, respectively.

$$SN = \frac{TP}{TP + TN} \tag{17}$$

$$SP = \frac{TN}{TN + FP} \tag{18}$$

The Matthews correlation coefficient (MCC) comprehensively considered true positives, true negatives, false positives, and false negatives, making it suitable for datasets with class imbalance. It effectively reduced the impact of class distribution on the results and summarized classification quality with a single value.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{19}$$

The F1 score achieved a balance between precision and recall by harmonizing the two. In cases of severe class imbalance, MCC was regarded as a more comprehensive performance evaluation metric, while F1 score could be prioritized as the evaluation criterion when focusing on the balance between precision and recall.

$$F1 = 2 \times \frac{TP}{2TP + FP + FN} \tag{20}$$

## 4 Model results and analysis

In this experiment, we concatenated convolutional neural networks with different attention mechanisms. The convolutional neural networks extracted rich local relational information in upstream tasks, while the attention mechanisms provided redundancy reduction and key information enhancement in downstream tasks, with important parameters required for each constructed model shown in Table 1.

**Table1.** Model Parameters

| Model | Key parameters | values |
|---|---|---|
| CNN+PSA | channels | 64 |
| CNN+GAM | channels, rate | 64, 4 |
| CNN+MAB | channels | 243 |
| CNN+OrthoNets | channels, height | 256, 64 |

After constructing the classification models, we input five different encoded features and their combined all_str feature into the model for evaluation, with results presented in Table 2. The BLOSUM62 feature performed the best in the model, followed by the combined all_str feature, while the ASDC, CKSAAGP, and DPC features were almost completely misclassified. In this round of experiments, the CNN+OrthoNets model achieved the highest accuracy, with an ACC value of 85.06%, followed by CNN+MAB, which reached an ACC value of 83.59%. The performances of the other two models, CNN+PSA and CNN+GAM, were relatively similar.

**Table2.** The performance of the CNN+attention model on various features

| Attention mechanism: PSA | | | | | |
|---|---|---|---|---|---|
| Feature | ACC | SP | SN | MCC | F1 |
| ASDC | 0.4989 | 0.9949 | 0.0028 | -0.0184 | 0.6650 |
| EAAC | 0.7281 | 0.8488 | 0.6074 | 0.4701 | 0.7574 |
| CKSAAGP | 0.4992 | 0.9985 | 0.0000 | -0.0276 | 0.6660 |
| BLOSUM62 | 0.8134 | 0.8777 | 0.7491 | 0.6320 | 0.8246 |
| DPC | 0.4922 | 0.9789 | 0.0056 | -0.0677 | 0.6585 |
| All_str | 0.8003 | 0.8701 | 0.7306 | 0.6066 | 0.8134 |
| Attention mechanism: GAM | | | | | |
| Feature | ACC | SP | SN | MCC | F1 |
| ASDC | 0.4989 | 0.9985 | 0.0000 | -0.0184 | 0.6650 |
| EAAC | 0.7281 | 0.7823 | 0.6065 | 0.4701 | 0.7574 |
| CKSAAGP | 0.4992 | 0.9574 | 0.0102 | -0.0276 | 0.6660 |
| BLOSUM62 | 0.8134 | 0.8684 | 0.8028 | 0.6320 | 0.8246 |
| DPC | 0.4922 | 0.9970 | 0.0000 | -0.0677 | 0.6585 |
| All_str | 0.8003 | 0.8902 | 0.7074 | 0.6066 | 0.8134 |
| Attention mechanism: MAB | | | | | |
| Feature | ACC | SP | SN | MCC | F1 |
| ASDC | 0.4986 | 0.9944 | 0.0028 | -0.0215 | 0.6648 |
| EAAC | 0.7333 | 0.8518 | 0.6148 | 0.4803 | 0.7616 |
| CKSAAGP | 0.4990 | 0.9980 | 0.0000 | -0.0319 | 0.6658 |

| | | | | | |
|---|---|---|---|---|---|
| BLOSUM62 | 0.8359 | 0.8792 | 0.7926 | 0.6743 | 0.8427 |
| DPC | 0.4845 | 0.9578 | 0.0111 | -0.0964 | 0.6501 |
| All_str | 0.8086 | 0.8736 | 0.7435 | 0.6224 | 0.8203 |
| Attention mechanism: OrthoNets | | | | | |
| Feature | ACC | SP | SN | MCC | F1 |
| ASDC | 0.4798 | 0.9254 | 0.0343 | -0.0889 | 0.6402 |
| EAAC | 0.7750 | 0.8853 | 0.6648 | 0.5640 | 0.7974 |
| CKSAAGP | 0.4784 | 0.9169 | 0.0398 | -0.0900 | 0.6374 |
| BLOSUM62 | 0.8506 | 0.8938 | 0.8074 | 0.7038 | 0.8568 |
| DPC | 0.4654 | 0.8901 | 0.0407 | -0.1310 | 0.6248 |
| All_str | 0.8117 | 0.8631 | 0.7602 | 0.6266 | 0.8209 |

Building on the above, in order to further investigate whether the ASDC, CKSAAGP, and DPC features truly lacked experimental value, we utilized the interpretable model SHAP to visualize the selected features. We then filtered out a maximum of 20 useful features from each reduced feature file and added the suffix "select" to the filtered feature files. To understand the impact of the ASDC, CKSAAGP, and DPC features on the overall feature set All_str, we created a new set that included the results from the filtered feature files, termed All_select_str, with a clear explanation provided in Figure 3.

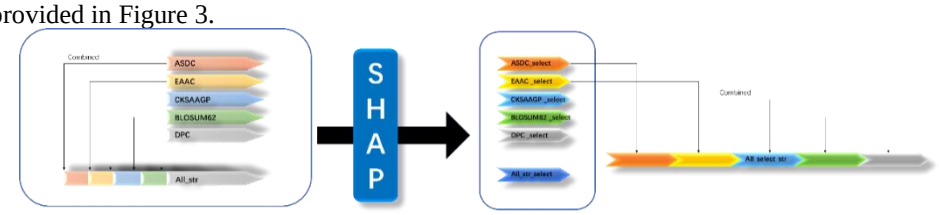


**Fig. 3.** feature selection

The benefit of this process was that it achieved dimensionality reduction for each feature, which could be interpreted as the pixel contribution of feature channels to the categories.

Subsequently, we input the selected features into the classification model again, and the performance results were presented in Table 3. The BLOSUM62 feature continued to perform the best across all models, but this time the model with the highest accuracy became CNN+MAB, with an ACC value increasing by 5.32% before and after feature selection, and an MCC value increasing by 0.1050. Additionally, an interesting phenomenon was observed in the regrouping results. Theoretically, the All_str_select features, as the selection results of All_str, should have drawn from the optimal parts of each feature; however, their performance on each model was generally lower than that of the All_select_str features. This indicated that the ASDC, CKSAAGP, and DPC feature sets also held certain research value.

**Table3.** The performance of the CNN + attention model on the selected features

| Attention mechanism: PSA | | | | | |
|---|---|---|---|---|---|
| Feature | ACC | SP | SN | MCC | F1 |
| ASDC_select | 0.5010 | 0.9919 | 0.0102 | 0.0108 | 0.6653 |
| EAAC_select | 0.7861 | 0.9185 | 0.6537 | 0.5934 | 0.8111 |
| CKSAAGP_select | 0.5000 | 1.0000 | 0.0000 | 0.0000 | 0.6667 |
| BLOSUM62_select | 0.8805 | 0.9074 | 0.8537 | 0.7622 | 0.8837 |
| DPC_select | 0.5000 | 1.0000 | 0.0000 | 0.0000 | 0.6667 |
| All_str_select | 0.8758 | 0.8988 | 0.8528 | 0.7524 | 0.8786 |
| All_select_str | 0.8829 | 0.9074 | 0.8583 | 0.7667 | 0.8857 |
| Attention mechanism: GAM | | | | | |
| Feature | ACC | SP | SN | MCC | F1 |
| ASDC_select | 0.5010 | 0.9959 | 0.0037 | 0.0108 | 0.6653 |
| EAAC_select | 0.7861 | 0.8685 | 0.6713 | 0.5934 | 0.8111 |
| CKSAAGP_select | 0.5000 | 0.9980 | 0.0009 | 0.0000 | 0.6667 |
| BLOSUM62_select | 0.8805 | 0.9090 | 0.8259 | 0.7622 | 0.8837 |
| DPC_select | 0.5000 | 1.0000 | 0.0000 | 0.0000 | 0.6667 |
| All_str_select | 0.8758 | 0.9054 | 0.8463 | 0.7524 | 0.8786 |
| All_select_str | 0.8829 | 0.9090 | 0.8389 | 0.7667 | 0.8857 |
| Attention mechanism: MAB | | | | | |
| Feature | ACC | SP | SN | MCC | F1 |
| ASDC_select | 0.5020 | 0.9939 | 0.0102 | 0.0228 | 0.6662 |
| EAAC_select | 0.7797 | 0.8928 | 0.6667 | 0.5743 | 0.8021 |
| CKSAAGP_select | 0.4990 | 0.9980 | 0.0000 | -0.0319 | 0.6658 |
| BLOSUM62_select | 0.8891 | 0.9170 | 0.8611 | 0.7793 | 0.8921 |
| DPC_select | 0.4923 | 0.9801 | 0.0046 | -0.0695 | 0.6588 |
| All_str_select | 0.8810 | 0.9130 | 0.8491 | 0.7636 | 0.8847 |
| All_select_str | 0.8865 | 0.9099 | 0.8630 | 0.7738 | 0.8891 |
| Attention mechanism: OrthoNets | | | | | |
| Feature | ACC | SP | SN | MCC | F1 |
| ASDC_select | 0.4998 | 0.9838 | 0.0157 | -0.0018 | 0.6629 |
| EAAC_select | 0.7688 | 0.8811 | 0.6565 | 0.5517 | 0.7922 |
| CKSAAGP_select | 0.4876 | 0.9668 | 0.0083 | -0.0873 | 0.6536 |
| BLOSUM62_select | 0.8768 | 0.9165 | 0.8370 | 0.7559 | 0.8815 |
| DPC_select | 0.5000 | 1.0000 | 0.0000 | 0.0000 | 0.6667 |
| All_str_select | 0.8750 | 0.9120 | 0.8380 | 0.7520 | 0.8794 |
| All_select_str | 0.8789 | 0.9079 | 0.8500 | 0.7591 | 0.8823 |

## 5 Conclusions

In recent years, with the deepening research into the functions of plant proteins, the precise prediction of post-translational modification sites of proteins became a critical bottleneck in elucidating the mechanisms of traditional Chinese medicine. However, traditional experimental methods faced limitations such as low efficiency, high costs, and difficulty in handling large-scale data. Therefore, developing efficient and reliable computational models emerged as an important research direction in this field.

For Paeonia lactiflora, a traditional Chinese medicinal material with significant pharmacological activity, this study innovatively integrated multi-dimensional features of amino acid sequences and successfully constructed a protein post-translational modification site prediction model based on a deep learning framework combining convolutional neural networks and attention mechanisms, suitable for the characteristics of different herbal materials. By utilizing a dedicated dataset for Paeonia lactiflora (1080 positive samples and 1976 negative samples) built from the TCMSP database, the feasibility of this method in predicting complex herbal protein modifications was systematically validated.

Although the current model demonstrated effective recognition capabilities for key modification sites, its classification accuracy still required further improvement, particularly in optimizing the distinction of boundary cases between adjacent non-modified sites. In the future, we can improve in the following aspects: (1) explore more encoding methods and perform pairwise combinations; (2) investigate the integration of three-dimensional structural information and multimodal data fusion strategies; (3) leverage advanced techniques such as transfer learning and federated learning for feature fusion, which are expected to significantly enhance the model's generalization ability and predictive accuracy. This provides important theoretical support for revealing the mechanisms of action of active components in traditional Chinese medicine and reducing clinical medication risks.

**Acknowledgement.** This work was supported by the National Natural Science Foundation of China (Grant No. 62333018), Xuzhou Science and Technology Plan Project (KC21047), Jiangsu Provincial Natural Science Foundation (No. SBK2019040953), Natural Science Fund for Colleges and Universities in Jiangsu Province (No. 19KJB520016) and Young Talents of Science and Technology in Jiangsu and ghfund202302026465.

## References

1. Yan, L., Wu, Y., Guan, R., Jin, C., Han, R., Ou, J., Tong, X.: Nontargeted Metabolomic Profiling of a Single Paeonia Lactiflora Plant and Its Quality Marker Identification. ChemistryOpen. 2400520 (2025)
2. Sun, S., Jimu, R. B., Lema, A. K., Elmamoune, H., Fan, Z., Jin, C., Han, R.: A Systematic Review on the Origin, Anti-Inflammatory Effect, Mechanism, Pharmacokinetics, and Toxicity of Albiflorin. Arabian Journal of Chemistry. 17(7), 105836 (2024)
3. Lim, J. W., Kang, M. K., Kim, H. E.: Antibacterial Effects of Paeonia Lactiflora Extract on Oral Microcosm Biofilms. Applied Sciences. 14(23), 11290 (2024)

4. Wu, F., Wu, G., Li, T., Lu, W., Fu, T., Zhang, Z.: Exploring the Target and Mechanism of Radix Paeoniae Alba on Sjogren's Syndrome. Combinatorial Chemistry & High Throughput Screening. 26(6), 1224--1232 (2023)
5. Huang, Y., He, M., Zhang, J., Cheng, S., Cheng, X., Chen, H., Zeng, S.: White Tea Aqueous Extract: A Potential Anti-Aging Agent Against High-Fat Diet-Induced Senescence in Drosophila melanogaster. Foods. 13(24), 4034 (2024)
6. Chen, L., Kan, J., Zheng, N., Li, B., Hong, Y., Yan, J., Li, H.: A Botanical Dietary Supplement from White Peony and Licorice Attenuates Nonalcoholic Fatty Liver Disease by Modulating Gut Microbiota and Reducing Inflammation. Phytomedicine. 91, 153693 (2021)
7. Mu, X., Luan, R., Gao, Y., Zhao, B., Wang, J., Ni, X., Gao, D.: The Traditional Applications, Phytochemistry, Pharmacology, Pharmacokinetics, Quality Control and Safety of Paeoniae Radix Alba: A Review. The American Journal of Chinese Medicine. 52(08), 2337--2376 (2024)
8. Fang, C., Xu, X., Lu, F., Liu, S.: Study on the Collaborative Protective Mechanism of Scutellariae Radix and Paeoniae Radix Alba Against Diabetic Cardiomyopathy Through the Gut-Heart Axis. Frontiers in Microbiology. 16, 1500935 (2025)
9. Jiang, H., Li, J., Wang, L., Wang, S., Nie, X., Chen, Y., He, Y.: Total Glucosides of Paeony: A Review of Its Phytochemistry, Role in Autoimmune Diseases, and Mechanisms of Action. Journal of Ethnopharmacology. 258, 112913 (2020)
10. Luan, X., Zhang, X., Nie, M., Zhao, Y.: Traditional Chinese Medicine Integrated Responsive Microneedles for Systemic Sclerosis Treatment. Research. 6, 0141 (2023)
11. Xu, M. Y., Zhang, J. B., Peng, Y. Z., Liu, M. C., Ma, S. Y., Zhou, Y., Ma, S.: Network Pharmacology and Experimental Validation Identify Paeoniflorin as a Novel SRC-Targeted Therapy for Castration-Resistant Prostate Cancer. Pharmaceuticals. 18(8), 1241 (2025)
12. Wang, Y., He, Q. Q., Zhu, Y. T., Zhang, Y., Yan, J., Liang, L. F., Ke, P. F.: Total Glucosides of Paeony Ameliorates Lupus Nephritis by Suppressing ZBP1-Mediated PANoptosis in Podocytes. Phytomedicine. 156996 (2025)
13. Liao, T., Kang, J., Ma, Z., Jie, L., Feng, M., Liu, D., Xing, R.: Total Glucosides of White Paeony Capsule Alleviate Articular Cartilage Degeneration and Aberrant Subchondral Bone Remodeling in Knee Osteoarthritis. Phytotherapy Research. 39(4), 1758--1775 (2025)
14. Yan, M., Wang, Q., Yang, H., Liu, D., Liang, W., Chen, H.: The Paeonol of Total Glucosides of White Paony Regulates the Differentiation of CD4+ Treg Cells Through the EP300/Foxp3 Axis to Relieve Pulmonary Fibrosis in Mice. Cell Biochemistry and Biophysics. 1--12 (2025)
15. Lee, J. M., Hammarén, H. M., Savitski, M. M., Baek, S. H.: Control of Protein Stability by Post-Translational Modifications. Nature Communications. 14(1), 201 (2023)
16. Agrata, R., Komander, D.: Ubiquitin—A Structural Perspective. Molecular Cell. 85(2), 323--346 (2025)
17. Wu, X., Xu, M., Geng, M., Chen, S., Little, P. J., Xu, S., Weng, J.: Targeting Protein Modifications in Metabolic Diseases: Molecular Mechanisms and Targeted Therapies. Signal Transduction and Targeted Therapy. 8(1), 220 (2023)
18. Wang, R., Li, Y., Ji, J., Kong, L., Huang, Y., Liu, Z., Lu, L.: The Emerging Role of Herbal Medicines in Cancer by Interfering with Posttranslational Modifications. Antioxidants & Redox Signaling. 42(1-3), 150--164 (2025)
19. Zhang, N., Wang, X., Li, Y., Lu, Y., Sheng, C., Sun, Y., Jiao, Y.: Mechanisms and Therapeutic Implications of Gene Expression Regulation by circRNA-Protein Interactions in Cancer. Communications Biology. 8(1), 77 (2025)

20. Shao, S., Lv, B., Wang, M., Zeng, S., Wang, S., Yang, Z., Ma, P.: Biosynthesis and Regulatory Mechanism of Tanshinones and Phenolic Acids in Salvia miltiorrhiza. The Plant Journal. 123(2), e70358 (2025)
21. Huang, Z., Zhu, J., Zhou, Y. L., Shi, J.: The cGAS-STING Pathway: A Dual Regulator of Immune Response in Cancer and Therapeutic Implications. Journal of Translational Medicine. 23(1), 766 (2025)
22. Geffen, Y., Anand, S., Akiyama, Y., Yaron, T. M., Song, Y., Johnson, J. L., Zhou, D. C.: Pan-Cancer Analysis of Post-Translational Modifications Reveals Shared Patterns of Protein Regulation. Cell. 186(18), 3945--3967 (2023)
23. van der Gaag, B. L., Deshayes, N. A., Breve, J. J., Bol, J. G., Jonker, A. J., Hoozemans, J. J., van de Berg, W. D.: Distinct tau and alpha-synuclein Molecular Signatures in Alzheimer's Disease with and without Lewy Bodies and Parkinson's Disease with Dementia. Acta Neuropathologica. 147(1), 14 (2024)
24. Ertelt, M., Mulligan, V. K., Maguire, J. B., Lyskov, S., Moretti, R., Schiffner, T., Schoeder, C. T.: Combining Machine Learning with Structure-Based Protein Design to Predict and Engineer Post-Translational Modifications of Proteins. PLOS Computational Biology. 20(3), e1011939 (2024)
25. Öztürk, H., Özgür, A., Ozkirimli, E.: DeepDTA: Deep Drug–Target Binding Affinity Prediction. Bioinformatics. 34(17), i821--i829 (2018)
26. Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R. K., Kumar, P.: Artificial Intelligence to Deep Learning: Machine Intelligence Approach for Drug Discovery. Molecular Diversity. 25(3), 1315--1360 (2021)
27. Asgari, E., Mofrad, M. R.: Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics. PloS One. 10(11), e0141287 (2015)
28. Krivák, R., Hoksza, D.: P2Rank: Machine Learning Based Tool for Rapid and Accurate Prediction of Ligand Binding Sites from Protein Structure. Journal of Cheminformatics. 10(1), 39 (2018)
29. Wang, D., Zeng, S., Xu, C., Qiu, W., Liang, Y., Joshi, T., Xu, D.: MusiteDeep: A Deep-Learning Framework for General and Kinase-Specific Phosphorylation Site Prediction. Bioinformatics. 33(24), 3909--3916 (2017)
30. Li, Y., Wang, S., Umarov, R., Xie, B., Fan, M., Li, L., Gao, X.: DEEPre: Sequence-Based Enzyme EC Number Prediction by Deep Learning. Bioinformatics. 34(5), 760--769 (2018)
31. Schubach, M., Maass, T., Nazaretyan, L., Röner, S., Kircher, M.: CADD v1.7: Using Protein Language Models, Regulatory CNNs and Other Nucleotide-Level Scores to Improve Genome-Wide Variant Predictions. Nucleic Acids Research. 52(D1), D1143--D1154 (2024)
32. Tan, D., Jiang, H., Li, H., Xie, Y., Su, Y.: Prediction of Drug–Protein Interaction Based on Dual Channel Neural Networks with Attention Mechanism. Briefings in Functional Genomics. 23(3), 286--294 (2024)
33. Xu, Y., Cheng, J.: Secondary Structure Prediction of Protein Based on Multi Scale Convolutional Attention Neural Networks. Mathematical Biosciences and Engineering: MBE. 184, 3404--3422 (2021)
34. Luo, Z., Wang, Q., Xia, Y., Zhu, X., Yang, S., Xu, Z., Gu, L.: DLBWE-Cys: A Deep-Learning-Based Tool for Identifying Cysteine S-Carboxyethylation Sites Using Binary-Weight Encoding. Frontiers in Genetics. 15, 1464976 (2025)
35. He, Y., Zheng, X., Miao, Q.: TFA-CLSTMNN: Novel Convolutional Network for Sound-Based Diagnosis of COVID-19. International Journal of Wavelets, Multiresolution and Information Processing. 21(03), 2250058 (2023)

36. Liang, B., Zhu, Y., Shi, W.: SARS-CoV-2 Spike Protein Post-Translational Modification Landscape and Its Impact on Protein Structure and Function via Computational Prediction. Res (Washington DC). 6, 0078 (2023)
37. Chou, K. C.: Prediction of Signal Peptides Using Scaled Window. Peptides. 22(12), 1973--1979 (2001)
38. Lee, T. Y., Chen, S. A., Hung, H. Y., Ou, Y. Y.: Incorporating Distant Sequence Features and Radial Basis Function Networks to Identify Ubiquitin Conjugation Sites. PloS One. 6(3), e17331 (2011)
39. Chen, Z., Zhao, P., Li, C., Li, F., Xiang, D., Chen, Y. Z., Song, J.: iLearnPlus: A Comprehensive and Automated Machine-Learning Platform for Nucleic Acid and Protein Sequence Analysis, Prediction and Visualization. Nucleic Acids Research. 49(10), e60--e60 (2021)
40. Wei, L., Zhou, C., Chen, H., Song, J.: Acpred-fl: A Sequence-Based Predictor Using Effective Feature Representation to Improve the Prediction of Anti-Cancer Peptides. Bioinformatics. 34, 4007--4016 (2018). DOI: https://doi.org/10.1093/bioinformatics/bty451
41. Chen, Z., Zhao, P., Li, F., Leier, A., Marquez-Lago, T. T., Wang, Y., Song, J.: iFeature: A Python Package and Web Server for Features Extraction and Selection from Protein and Peptide Sequences. Bioinformatics. 34(14), 2499--2502 (2018)
42. Liu, Y., Shao, Z., Hoffmann, N.: Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions. arXiv Preprint arXiv:2112.05561 (2021)
43. Zhang, X., Zhou, X., Lin, M., Sun, J.: Shufflenet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 6848--6856) (2018)
44. Liu, H., Liu, F., Fan, X., Huang, D.: Polarized Self-Attention: Towards High-Quality Pixel-Wise Regression. arXiv Preprint arXiv:2107.00782 (2021)
45. Wang, Y., Li, Y., Wang, G., Liu, X.: Multi-Scale Attention Network for Single Image Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5950--5960) (2024)
46. Salman, H., Parks, C., Swan, M., Gauch, J.: Orthonets: Orthogonal Channel Attention Networks. In 2023 IEEE International Conference on Big Data (BigData) (pp. 829--837). IEEE (2023)
47. Jabeur, S. B., Mefteh-Wali, S., Viviani, J. L.: Forecasting Gold Price with the XGBoost Algorithm and SHAP Interaction Values. Annals of Operations Research. 334(1), 679--699 (2024)
48. Lundberg, S. M., Lee, S. I.: A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems. 30 (2017)
49. Araf, I., Idri, A., Chairi, I.: Cost-Sensitive Learning for Imbalanced Medical Data: A Review. Artificial Intelligence Review. 57(4), 80 (2024)
50. Hassanzadeh, R., Farhadian, M., Rafieemehr, H.: Hospital Mortality Prediction in Traumatic Injuries Patients: Comparing Different SMOTE-Based Machine Learning Algorithms. BMC Medical Research Methodology. 23(1), 101 (2023)