MDC-Kace: A Model for Predicting Lysine Acetylation Sites Based on Modular Densely Connected Convolutional Networks

HUIQING WANG, ZHILIANG YAN, DAN LIU, HONG ZHAO, AND JIAN ZHAO

College of Information and Computer, Taiyuan University of Technology, Taiyuan 030024, China
Corresponding author: Zhiliang Yan (1437202210@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61976150, and in part by the Key Research and Development Plan of Shanxi Province under Grant 201903D121151.

ABSTRACT Lysine acetylation (Kace) is a conservative protein posttranslational modification (PTM) closely related to various metabolic diseases. Therefore, Kace sites identification is of great significance for investigating metabolic disease treatments. Existing studies have shown that protein structural properties contain highly useful structural information, which provides a strong basis for identifying PTMs. During the feature learning process, features at different levels are complementary, and taking them into consideration can effectively improve the quality of the features. However, existing deep-learning methods used protein sequence-level information as input without considering the protein structural properties. Furthermore, only high-level features were focused on, resulting in considerable information loss and weakening the prediction results. Therefore, we propose a novel deep-learning model based on modular densely connected convolutional networks (MDC) for Kace sites prediction, called MDC-Kace. MDC-Kace introduces the protein structural properties and combines them with the original protein sequence and the amino acid physicochemical properties to construct the site’s feature space. Then, modular densely connected convolutional networks are used to capture the information of features at different levels and reduce information loss and crosstalk during the feature learning step. We add a squeeze-excitation layer to evaluate the importance of different features and improve the network abstraction ability to identify potential Kace sites. The experimental results of ten-fold cross-validation and independent testing in human, Mus musculus and Escherichia coli datasets showed that our MDC-Kace model outperforms the existing Kace site predictors and can predict potential Kace sites effectively. MDC-Kace can be available at https://github.com/lianglianggg/MDC-Kace.

INDEX TERMS Lysine acetylation, protein structure properties, densely connected convolutional networks, modular network structure, feature learning.

I. INTRODUCTION

Protein posttranslational modifications (PTMs) are chemical modifications of proteins following translation. PTMs play an important regulatory role in the protein maturation process [1]–[3] and participate in nearly all cellular activities [4]. Common types of PTMs are acetylation, phosphorylation and ubiquitination. Acetylation is a conservative PTM that mainly occurs in ε-amino groups of protein lysine residues [5], [6]. It plays a major role in various aspects of mammalian cell physiology, such as cell metabolism, differentiation, aging and signal transduction and is closely related to metabolic diseases such as diabetes, cancer, and cardiovascular disease [3], [6]–[8]. Some studies have shown that metabolic process can be controlled through using drugs to modify lysine acetylation (Kace), which can alter the activities of a series of related enzymes in human body, thereby effectively preventing and treating metabolic diseases [8]–[10]. Thus, the identification of Kace sites and their substrates is of great significance to understand the potential molecular mechanisms of acetylation modification [14], and it also provides useful site modification information for the design of drugs for the treatments of related metabolic diseases [11], [12], [17].

As a conservative PTM, Kace has been studied by many researchers [3], [5], [12]–[20]. Li et al. [12] developed the PAIL tool based on original protein sequence...
information to successfully predict Kace sites. To address the uniqueness of original protein sequence information, Lee et al. [13] increased the solvent accessibility and amino acid physicochemical property information and achieved a higher accuracy than Li et al. [12]. Chen et al. [17] used the original protein sequence, amino acid physicochemical properties and protein evolution information to predict Kace sites of prokaryotes and achieved better results than previous studies. Suo et al. [18] combined the amino acid composition, protein evolutionary similarity, and amino acid physicochemical properties, which greatly improved the prediction accuracy of Kace sites. However, these methods considered only sequence-level information, such as the original protein sequence, amino acid physicochemical properties and protein evolution information. They are limited to the protein sequence-level and do not take protein structural properties into account, so the information is single. With the development of related technologies that can predict the structural properties of proteins [21], [22], researchers have included protein structural properties in PTM site prediction, and the inclusion of this information can enrich the feature space of sites [23]–[26]. Reddy et al. [24] introduced protein structural characteristics into the prediction of lysine glycosylation, which solved the problem of insufficient information at protein sequence level. The authors achieved a substantial improvement in prediction performance. Chandra et al. [25] input protein structural properties into a multilayer perceptron and developed a method called PhoglyStruct, which demonstrated that protein structural properties were important in distinguishing phosphorylated and non-phosphorylated lysine residues. López et al. [26] achieved good results in the recognition of lysine succinylation sites by combining protein structural properties and protein evolution characteristics. Therefore, protein structural properties include highly useful local and global structural features of proteins, which provide a strong basis for PTM identification.

At present, most Kace site prediction methods are based on traditional machine learning methods, such as the support vector machine (SVM) [3], [13]–[18], naive Bayes (NB) [12], logistic regression (LR) [19] and random forest (RF) [5], [20] algorithms. Although these methods have achieved some good results, they rely on the manual selection of features, which is a time-consuming, labor-intensive, subjective task that is difficult to be used for mining potential information [27]. Deep learning has an ability to automatically learn advanced representations from raw data. It is a good way to solve the above problems [27], [28] and has been successfully applied to PTM site prediction [27], [29]–[31], protein function prediction [32], object detection [33], [36], natural language processing [34], [35] and other fields. Zhao et al. [33] used residual connections to move details of low-level features, such as position and distance, to a higher level, making a substantial breakthrough in object detection tasks. Luo et al. [30] adopted dense jump connections to introduce low-level features to high-level features, which improved the flow of phosphorylation information in the network. Considering the crosstalk between different information strands, He et al. [29] used a modular deep convolutional neural network (CNN) to learn the advanced features of the original protein sequence, the amino acid physicochemical properties and the position-specific scoring matrix of lysine ubiquitination sites. The authors solved the problem of crosstalk between different types of information and successfully predicted lysine ubiquitination sites at a large-scale. Huang et al. [36] proposed densely connected convolutional networks based on the idea of dense jump connections, which connect each layer to every other layer in a feed-forward fashion. The authors realized that the high-level convolution layers contain the complementary information passed from the lower levels. And they verified the effectiveness of their structure on object recognition tasks. In contrast to CNNs, modular densely connected convolutional networks focus on low-level visual features and high-level semantic features at the same time, which can effectively avoid information crosstalk and reduce information loss. In addition, some studies have shown that the importance levels of features are also different [32], [37]. The direct fusion of features from different sources will weaken the quality of advanced features and affect the classification outcome.

Recently, several deep-learning methods for Kace site prediction have been developed. Kiemer et al. [38] used a neural network to develop the NetAcet model for site prediction, but due to limited data, the advantages of the neural network were not considerable. Wang et al. [4] connected 5-dimensional comprehensive amino acid physicochemical properties with 1-dimensional position information in series and trained a capsule network to predict protein Kace sites. Wu et al. [39] fused six encoding methods, such as one-hot encoding, AAindex and the position-specific scoring matrix, and trained a deep neural network (DNN) to predict Kace sites. The authors verified their method on an independent test dataset. However, these methods consider only sequence level information, such as the original protein sequence and amino acid physicochemical properties, and do not consider the protein structural properties. Additionally, they focus only on the high-level features during the feature learning period and do not pay attention to the information complementation between features at different levels. Thus, there is considerable information loss, which lowers the accuracy of Kace site prediction.

Based on the above problems, we introduced modular densely connected convolutional networks (MDC) and proposed a novel deep-learning model for Kace site prediction, called MDC-Kace. MDC-Kace considers the protein structural properties, the original protein sequence and the amino acid physicochemical properties. It describes Kace sites from three aspects, and constructs the initial feature space of sites. Then, modular densely connected convolutional networks are introduced to extract the advanced features of the protein structural properties, the original protein sequence and the amino acid physicochemical properties from the initial
feature space. The networks simultaneously focus on low-level and high-level features through dense jump connections, avoiding information crosstalk and reducing information loss. In addition, a squeeze-excitation (SE) layer is introduced to evaluate the importance of the features, weight each feature map, and then realize the adaptive dynamic fusion of the three types of features. The fused feature is then matched with the softmax layer to construct a Kace site classifier to predict potential Kace sites. To verify the performance of MDC-Kace, we performed ten-fold cross-validation and independent testing based on data from human, *Mus musculus* and *Escherichia coli*. The experimental results show that our model is superior to the existing methods and can effectively learn the abstract pattern of Kace sites. The prediction and analysis of potential Kace sites further illustrate that MDC-Kace is a powerful tool for identifying unknown Kace sites.

### II. MATERIALS AND METHODS

The prediction of Kace sites can be abstracted as a binary classification problem; i.e., each potential site can be classified as an acetylated site or non-acetylated site [4]. With lysine (K) as the center, we extracted a peptide chain with length \( L = 2n + 1 \) \( (n \) residues on each side) as a motif. The motifs were converted into numerical vectors by coding, which were used as the input of MDC-Kace. Then, we trained MDC-Kace based on the training dataset to enable it to learn the deep abstractions and motif patterns of Kace sites. Finally, the trained model was used to predict potential Kace sites.

#### A. DATA COLLECTION AND REFINEMENT

We collected and downloaded 6078, 3645 and 1860 experimentally verified human, *Mus musculus* and *Escherichia coli* lysine acetylated proteins from the Protein Lysine Modification Database (PLMD) [40], respectively. Considering that the SPIDER3 server [22] cannot handle protein sequences containing non-standard amino acids, we manually deleted these protein sequences. For details, please refer to Supplementary Materials Table S1. Here, we took the human data as an example. To avoid model deviations caused by large sequence homology, the CD-HIT tool [41] was used to remove redundant sequences, and the threshold was set to 0.4 [3]. Thus, 4977 acetylated proteins were retained. To conveniently compare with other predictors, we randomly selected 10% (498) of the filtered 4977 acetylated proteins as independent test data, and the remaining were training data. The statistical information of the dataset is shown in Table 1. For the information of the other two species datasets, please refer to Supplementary Materials Table S2. All the data are available at https://github.com/lianglianggg/MDC-Kace/tree/master/dataset.

#### B. INFORMATION CODING

In this work, we used one-of-21 coding [4] to encode the original protein sequence information of sites. For a motif with length \( L \), an \( L \times 21 \) dimensional vector representation of the

| Dataset type | Number of proteins | Positive samples | Negative samples |
|--------------|--------------------|------------------|------------------|
| Training     | 4479               | 21992            | 196040           |
| Independent test | 498            | 2294             | 19941            |

original protein sequence information was obtained. Atchley factors [42] were used to encode the amino acid physicochemical property information. Each amino acid residue was represented by 5 Atchley factors. Atchley factors were chosen because they statistically summarize 500 amino acid attributes that reflect polarity, secondary structure, molecular volume, codon diversity and electrostatic charge. Thus, an \( L \times 5 \) dimensional vector representation of the amino acid physicochemical property information was obtained. In addition, we obtained the protein structural property information from SPIDER3 [22], including 8 indexes, i.e., secondary structures: the \( \alpha \)-helix (ph), \( \beta \)-strand (pe), \( \gamma \)-coil (pc), local backbone torsion angles (\( \varphi \), \( \psi \), \( \theta \) and \( \tau \)), and accessible surface area (ASA). Therefore, an \( L \times 8 \) dimensional vector representation of the protein structural property information was obtained. The information coding diagram is described in Supplementary Materials Figure S1, and a summary of the properties of human acetylated proteins is described in Supplementary Materials Figure S2. The MDC-Kace model takes the vector representations of the three information types of each sample as input for training. Since the Atchley factors and 8 protein structural property indexes have some information overlap, the joint training strategy was used to train the model to reduce information redundancy.

#### C. MDC-KACE MODEL ARCHITECTURE

In this work, we aimed to build a deep-learning model that can more efficiently and accurately learn the deep hidden characteristics of Kace sites. To avoid information crosstalk among the original protein sequence, amino acid physicochemical properties and protein structural properties of the site motifs, we adopted the modular network structure design [29], [34], [35] to construct three feature extraction submodules: a sequence module, a physicochemical module and a structure module. The parameter spaces between each submodule are independent. In each module, stacked densely connected convolutional blocks [36] were used to extract the advanced features. Through dense jump connections, both low-level and high-level features are simultaneously considered to achieve information complementation between features at different levels and to reduce information loss. Then, we added SE layers [37] to evaluate the importance of the features to enhance the flow of acetylation information in the network, further achieving the adaptive dynamic fusion of the above three types of feature information. Finally, we input the fused advanced feature into the softmax layer for classification to efficiently predict potential Kace sites. The architecture of MDC-Kace is described in Figure 1.
1) EXTRACTING ADVANCED FEATURES BY MODULAR DENSELY CONNECTED CONVOLUTIONAL NETWORKS

To extract the features of the three types of information (i.e., the original protein sequence, the amino acid physicochemical properties and the protein structural properties) of Kace sites automatically and efficiently, we combined a modular network structure design with densely connected convolutional networks. We designed three submodules, namely, a sequence module, a physicochemical module and a structure module, to learn their advanced features. Densely connected convolutional networks have been shown to perform better than traditional CNNs, which use dense jump connections to focus on low-level and high-level information, simultaneously, to achieve information complementation [36]. Their feature propagation processes are shown in Figure 2 (A) and (B). The modular network structure can avoid information crosstalk [29], [34], [35]. Therefore, combining the two designs enables the network to learn better feature representations.

As shown in Figure 1, because the network structures of the sequence module, physicochemical module and structure module are the same, here, we describe only the sequence module.

First, the sequence module received one-of-21 coding vectors of site motifs with length L as input. Then the low-level feature maps of the original protein sequence information were generated through the one-dimensional convolutional layer, as shown in formula (1):

\[
X^0 = \sigma(I \ast W + b) \tag{1}
\]

where \( I \) refers to the one-of-21 coding vector, \( W \in \mathbb{R}^{21 \times S \times D} \) refers to the weight matrix with \( S = 3 \) (the filter size) and \( D = 96 \) (the number of filters, for relevant selection procedures, please refer to Supplementary Materials Figure S3 (C)), \( b \) is the bias term, and \( \sigma \) is the exponential linear unit (ELU) [43] activation function (for relevant experimental results, please refer to Supplementary Materials Figure S3 (A) and (B)). \( X^0 \in \mathbb{R}^{L \times D} \) refers to the output of the one-dimensional convolutional layer.

Then, the advanced feature representation of the original protein sequence information was extracted by the densely connected convolutional block, which is shown in formula (2):

\[
X^l = \sigma([X^0; X^1; \ldots; X^{l-1}] \ast W' + b') \tag{2}
\]
dimension. \( W' \in \mathbb{R}^{D' \times S' \times D'} \) refers to the weight matrix where \( D' \) is determined by \( l \), and \( D' = 32 \). \( b' \) is the bias term, and \( \sigma \) is the ELU activation function [43]. The output of a densely connected convolutional block is the concatenation of \( X'^0 \) and \( X'^1, X'^2, \ldots, X'^l \) (i.e., \( \{ X'^0, X'^1, \ldots, X'^l \} \)).

Next, the transition layer was used to perform the convolution and activation operations on the feature map of the original protein sequence information obtained by formula (2). The transition layer process is shown in formula (3):

\[
X = \sigma ( X'^0; X'^1; \ldots; X'^l ) \ast W'' + b''
\]

(3)

where \( W'' \in \mathbb{R}^{(D'' + D') \times S' \times (D'' + D')} \) refers to the weight matrix with \( S' = 1 \) (the filter size), \( b'' \) is the bias term, \( \sigma \) is the ELU activation function [43], and \( X \) represents the output of the transition layer. Then, the average pooling operation was performed on \( X \) to decrease its dimension and reduce the risk of model overfitting.

Formulas (2) and (3) were repeated to construct stacked densely connected convolutional blocks. We set the number of densely connected convolutional blocks to 4; for relevant experimental results, please refer to Supplementary Materials Figure S3 (A) and (B)). Importantly, formula (3) is not performed after the fourth implementation of formula (2), but uses global average pooling for replacement. Therefore, the advanced feature \( X^{(seq)} \) of the original protein sequence of a site was extracted through a sequence module after the above process. Similarly, the physicochemical module and structure module can also extract the advanced features of the amino acid physicochemical properties and protein structural properties, respectively, of a site through the above process.

2) LEARNING FEATURE IMPORTANCE SCORE

Considering that different feature maps have different contributions [37], we introduced a SE layer into each module to learn the importance scores of the feature maps, weight each feature map and enhance the acetylation information flow in the network. In this way, the adaptive dynamic fusion of the above three types of features can be achieved and the discrimination ability of the network can be improved. The SE layer is implemented based on global average pooling and two fully connected (FC) layers, as shown in Figure 3.

For the advanced feature \( X^{(seq)} \), which was extracted by the sequence module, the SE layer used global average pooling to squeeze the global space information of \( X^{(seq)} \) into a channel descriptor \( z \in \mathbb{R}^C \), where \( C \) is the number of channels, which is shown in formula (4):

\[
z_c = F_{sq}( X^{(seq)} ) = \frac{1}{W \times H} \sum_{i=1}^{H} \sum_{j=1}^{W} X^{(seq)}(i, j)
\]

(4)

where \( W \) and \( H \) are the width and height, respectively, of the feature map \( X^{(seq)} \).

Then, two FC layers were used to capture the channel dependence of \( X^{(seq)} \) and learn the specific weight of each feature map of \( X^{(seq)} \), as shown in formula (5):

\[
s = F_{ex}( z, W ) = \sigma ( W_2 \ast \delta( W_1 \ast z ) )
\]

(5)

where \( s \in \mathbb{R}^C \) refers to the specific weight of each feature map of \( X^{(seq)} \), \( \delta \) and \( \sigma \) are the activation functions of two FC layers. \( \delta \) is a rectified linear unit (ReLU) function with the parameter \( W_1 \in \mathbb{R}^{\frac{C}{r} \times C} \), where \( r \) is the reduction ratio, and we set \( r = 16 \) according to reference [37]. \( \sigma \) is a sigmoid function with the parameter \( W_2 \in \mathbb{R}^{C \times \frac{C}{r}} \), which ensures that the dimension of \( s \) is the same as the number of channels of \( X^{(seq)} \).

Next, the output of the SE layer, \( \tilde{X}^{(seq)} = [ \tilde{X}_1^{(seq)}, \tilde{X}_2^{(seq)}, \ldots, \tilde{X}_C^{(seq)} ] \), is obtained by scaling \( X^{(seq)} \) with the following activation shown in formula (6):

\[
\tilde{X}_c^{(seq)} = F_{scale} ( X^{(seq)} , s_c ) = s_c \cdot X^{(seq)}
\]

(6)

where \( F_{scale}( X^{(seq)} , s_c ) \) means that each value of the feature map \( X^{(seq)} \in \mathbb{R}^{H \times W} \) is multiplied by the weight \( s_c \).

Similarly, the physicochemical and structural modules can also obtain the weighted advanced features through the above process. Thus, the output of each sub-module is concatenated to generate the fused feature \( \tilde{X} \) for classification. The SE layer weights each feature map, which makes the feature fusion process adaptive and dynamic.

3) SOFTMAX LAYER PREDICTION OF THE KACE SITES

Based on the fused advanced feature \( \tilde{X} \), we trained the softmax classifier to predict Kace sites. The softmax layer received the fused advanced feature \( \tilde{X} \) as input, and obtained the prediction category of the sample after the weighted summation and activation operations. The forward propagation process of the softmax layer is shown in formula (7):

\[
P( y = i | x ) = \frac{ \exp( W_i^s \ast \tilde{X} + b_i^s ) }{ \sum_{j=1}^{2} \exp( W_j^s \ast \tilde{X} + b_j^s ) }
\]

(7)

where \( W_i^s \) and \( W_j^s \) refer to weight matrices, \( b_i^s \) and \( b_j^s \) are bias terms. \( P( y = i | x ) \) represents the probability that sample \( x \) is predicted to be in class \( i \). As mentioned above, the prediction of Kace sites can be considered as a binary classification problem, thus \( i \in \{ 0, 1 \} \). The prediction category of the softmax classifier is the category with the highest probability value.
D. MODEL TRAINING

The standard cross-entropy was used as the cost function to minimize the training errors in our MDC-Kace model:

\[
L_C = -\frac{1}{N} \sum_{j=1}^{N} y^j \ln P(y^j = 1 | x^j) + (1 - y^j) \ln P(y^j = 0 | x^j) \tag{8}
\]

where \( N \) refers to the total number of training samples, \( y^j \) is the true label of the \( j \)-th input motif, and \( x^j \) is the \( j \)-th input motif.

In addition, to reduce the impact of overfitting, L2 regularization was used in the training process; thus, the final objective function of MDC-Kace is defined as:

\[
\min_W L_C + \lambda \sum (\|W\|_2^2) \tag{9}
\]

where \( \lambda \) refers to the regularization coefficient and \( \|W\|_2 \) refers to the L2 norm of the weight matrix. Here, we adopted the Adam optimizer \([44]\) to optimize the objective function. Since a large learning rate and a small batch size would make the model training unstable, we set them to 0.0001 and 1000, respectively, with comprehensive consideration.

During the model training process, dropout and the early stopping strategy were used to further prevent the model from overfitting. To reduce the negative impact of data imbalance, we adopted class reweighting to increase the influence of the positive samples and force the model to learn the abstraction mechanism of the minority (taking the human motif). The detail definitions are as follows:

\[
\text{Sn} = \frac{TP}{TP + FN} \tag{10}
\]
\[
\text{Sp} = \frac{TN}{TN + FP} \tag{11}
\]
\[
\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}
\]
\[
\text{Pre} = \frac{TP}{TP + FP} \tag{13}
\]
\[
\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FN) \times (TN + FP)}} \tag{14}
\]
\[
G - \text{mean} = \sqrt{\frac{TP \times TN}{(TP + FN) \times (TN + FP)}} \tag{15}
\]

Here, \( TP, TN, FP \) and \( FN \) are the number of true positives, true negatives, false positives and false negatives, respectively. When the positive and negative samples are imbalanced, the MCC and G-mean indicators can better reflect the quality of the model than the other metrics \([25], [45], [46]\). Furthermore, we used the area under the receiver operating characteristic (ROC) curve (AUC) and the area under the precision-recall (PR) curve (AUPR) to measure the overall performance of the model. The higher the AUC and AUPR, the better the overall model performance.

III. RESULTS AND DISCUSSIONS

A. SELECTION OF THE KACE SITE WINDOW SIZE

In previous studies, different models had their own window sizes to maximize the efficiency of the model. Wang et al. \([4]\) predicted Kace sites with a window size of \( L = 33 \), Wu et al. \([39]\) used \( L = 31 \), Suo et al. \([18]\) used \( L = 21 \), and other researchers used 19 \([15]\), 21 \([17]\), etc. To select the optimal Kace sites window size for MDC-Kace and ensure that the amount of input information is sufficient, especially to make full use of the characteristics of the modular densely connected convolutional networks can extract features efficiently. We applied the training dataset described in Table 1 as a benchmark and set \( L \) to values from 21 to 61 in increments of 2. Therefore, 21 values were tested by performing ten-fold cross-validation, and the experimental results were averaged. As the window size increases, the performance of MDC-Kace improves. As shown in Figure 4, when the window size was \( L > 51 \), the AUC value of the model rose very slowly and had small fluctuations. Therefore, by comprehensively considering the calculation cost and model complexity, we selected 51 as the optimal Kace site window size for MDC-Kace. This result implied that our deep architecture requires longer sequence fragments to offer potential long distance information \([29]\).

B. EVALUATION OF MDC-KACE PREDICTION ABILITY

To assess the performance of the proposed model, MDC-Kace, we compared it with several existing Kace site prediction models through ten-fold cross-validation and independent testing by means of several statistical metrics: the sensitivity (Sn), specificity (Sp), accuracy (Acc), precision (Pre), Matthew’s correlation coefficient (MCC), and geometric mean (G-mean). The detailed definitions are as follows:

\[
\text{Sn} = \frac{TP}{TP + FN} \tag{10}
\]
\[
\text{Sp} = \frac{TN}{TN + FP} \tag{11}
\]
\[
\text{Acc} = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}
\]
\[
\text{Pre} = \frac{TP}{TP + FP} \tag{13}
\]
\[
\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FN) \times (TN + FP)}} \tag{14}
\]
\[
G - \text{mean} = \sqrt{\frac{TP \times TN}{(TP + FN) \times (TN + FP)}} \tag{15}
\]
attention mechanisms. However, in their follow-up work, the developers showed that MusiteDeep can be retrained as a Kace site prediction model [4]. EnsemblPail used an integrated support vector machine to predict Kace sites [15]. GPS-PAIL2.0 adopted an algorithm of Group-Based Prediction System (GPS) to predict specific acetylation sites of lysine acetyltransferases [7]. ProAcePred predicted nine species of prokaryote lysine acetylation sites based on the combination of sequence-based, physicochemical property and evolutionary information features via elastic net [17]. Since ProAcePred is a prokaryote-specific lysine acetylation site predictor [17], we evaluated it on only the *Escherichia coli* dataset. In our work, MusiteDeep was used as a typical deep-learning model. CapsNet was a successful exploration of the capsule network in PTM prediction, which has important guiding significance. DeepAcet was a pioneering application of deep neural network (DNN) for Kace site prediction. PSKAcePred, EnsemblPail, GPS-PAIL2.0 and ProAcePred were regarded as representatives of traditional machine learning methods. About the seven models, the former four provided independent tools for model retraining; while the latter three only provided web servers and we evaluated them only on the independent test datasets.

The results of ten-fold cross-validation of the above models on the human training dataset (described in Table 1) are shown in Table 2. Our model, MDC-Kace, achieved the highest G-mean and AUC indicators values. We noticed that both DeepAcet and PSKAcePred had high AUPRs, because they were trained on a balanced subset built by downsampling on the original training dataset, and the AUPR is sensitive to dataset imbalance [46]. As shown in column 6 of Table 2, PSKAcePred achieved the highest MCC (38.54%), because it was calculated based on a balanced dataset, and the negative samples were considered insufficient, causing biased results. In addition, we calculated Sn, Sp, Acc, and Pre for all the predictors, which are shown in columns 2 to 5 in Table 2. We found that the Sn and Pre values of the proposed model MDC-Kace were relatively low, because Sn is antagonistic to Sp and Pre according to their detailed definitions, and Pre is also sensitive to the data distribution [46]. To further verify the performance of MDC-Kace under the stricter redundancy removal threshold, we changed the threshold of the CD-HIT tool in Section 2.1 to 0.3, and the subsequent experimental process was the same as before. For the related experimental results, please refer to Supplementary Materials Tables S4. From Table S4, it can be concluded that the performance of our model, MDC-Kace, under the stricter redundancy removal threshold does not decrease, indicating that it has good robustness. In addition, we conducted experiments on the datasets of *Mus musculus* and *Escherichia coli*. For details, please refer to Supplementary Materials Tables S5-S8. The results indicated a similar conclusion as above. After comparisons, the performance of our model is better than the other predictors, indicating that the design of the deep-learning architecture of MDC-Kace is reasonable. The MDC-Kace model fully extracts the advanced features of the three types of information, the original protein sequence, the amino acid physicochemical properties and the protein structural properties, of potential Kace sites through modular stacked dense convolutional blocks. These blocks simultaneously focus on low-level and high-level features. The SE layer is introduced to weight features to achieve feature adaptive dynamic fusion. Compared to the other models, these factors allow MDC-Kace to better characterize the hierarchical relationships between the simple and complex abstraction of potential sites.

To determine whether MDC-Kace was overfit, we drew the training/validation loss/accuracy curve of the model in a ten-fold cross-validation process on the training dataset described in Table 1. For details, please refer to Supplementary Materials Figure S4. According to the curve trend, we concluded that our model MDC-Kace was not overfit. This finding indicates that the model can be effectively prevented from overtraining by using the early stopping strategy, L2 regularization and dropout, to improve its prediction ability.

Additionally, we further compared the prediction ability of the proposed model MDC-Kace with the other models. For the models with independent tools, we trained them based on the training dataset described in Table 1. Potential Kace site prediction was then performed on the independent test dataset described in Table 1. For the models that provide web servers, we tested their prediction performance based on only the independent test dataset. The histogram of the experimental results is shown in Figure 5.

MDC-Kace had the highest MCC, G-mean, AUC, and AUPR values and performed best on the independent test dataset, demonstrating that our model has better Kace site prediction ability than others. We noticed that PSKAcePred gave large differences between the independent test dataset and the training dataset in the above four indicators, as did DeepAcet. Here, the independent test dataset was imbalanced. However, the corresponding MDC-Kace results between the independent test dataset and the training dataset were roughly consistent. This finding indicates that MDC-Kace can deal with imbalanced data effectively by using class reweighting in model training process, without

![AUC value vs Kace site window size](image_url)
TABLE 2. The ten-fold cross-validation performance of the different methods on the human training dataset described in Table 1 (the following are %s). Since EnsemblePail and GPS-PAIL2.0 do not provide independent tools for model training, the results of their ten-fold cross-validation cannot be calculated. The largest value for each indicator is highlighted in bold.

| Model      | Sn   | Sp   | Acc  | Pre  | MCC  | G-mean | AUC  | AUPR         |
|------------|------|------|------|------|------|--------|------|--------------|
| MustieDeep | 72.73| 69.97| 70.25| 21.39| 27.10| 71.32  | 78.35| 32.80        |
| CapsNet    | 74.90| 68.51| 69.15| 21.10| 27.29| 71.60  | 78.41| 32.85        |
| DeepAcet   | 67.55| 62.88| 65.21| 64.53| 30.47| 65.16  | 70.41| 63.30        |
| PSKAccPred | 70.97| 67.55| 69.26| 68.63| 38.54| 69.24  | 76.09| 74.99        |
| MDC-Kace   | 69.85| 76.27| 75.62| 24.92| 30.88| 72.90  | 81.19| 33.48        |

FIGURE 5. The prediction performance of the different predictors on the human independent test dataset described in Table 1. Since the prediction results of the EnsemblePail and GPS-PAIL2.0 predictors do not include the probability score of the category, their AUC and AUPR cannot be calculated.

To the above comparisons, our model, MDC-Kace, has a stronger induction and abstraction ability, with exceptional prediction ability.

To further prove the generalization ability of our model MDC-Kace, we showed the prediction results of Kace sites for the 0.3 redundancy removal threshold for the human dataset, the 0.4 and 0.3 redundancy removal thresholds for the Mus musculus dataset and the Escherichia coli 0.4 and 0.3 redundancy removal thresholds in Supplementary Materials Tables S10-S14, respectively. According to the results, MDC-Kace has good generalization ability and can be applied to datasets of other species. This wider application ability provides an available reference for the prediction of Kace sites in other species. MDC-Kace adopts modular densely connected convolutional networks to mine the deep features of the three types of information of potential Kace sites, namely, the original protein sequence, the amino acid physicochemical properties and the protein structural properties. The networks capture the information of different levels at the same time by using dense jump connections. The SE layer is used to evaluate feature importance to construct a more precise advanced representation. Therefore, MDC-Kace classifies acetylated and non-acetylated lysine residues more accurately than the other models.

C. MDC-KACE ABLATION EXPERIMENTS

To verify the important role of the protein structural property information in predicting Kace sites and the necessity of each component in MDC-Kace, we conducted ablation experiments by ten-fold cross-validation using the training dataset described in Table 1. The experiments included the use of densely connected convolutional networks, the adoption of the modular network structure design idea, the addition of protein structural property information as the model’s input, and the introduction of the SE layer in the model. The experimental results are shown in Table 3. Here, the baseline model adopted the concatenation of the original protein sequence and amino acid physicochemical property information of the site motifs as its input and extracted advanced features by CNNs.

The densely connected convolutional networks improved the performance of the baseline model, as shown in columns 2 and 3 of Table 3. This finding indicates that the densely

FIGURE 5. The prediction performance of the different predictors on the human independent test dataset described in Table 1. Since the prediction results of the EnsemblePail and GPS-PAIL2.0 predictors do not include the probability score of the category, their AUC and AUPR cannot be calculated.
connected convolutional networks use the information of features at different levels simultaneously through dense jump connections, and achieve information complementation between the low-level and high-level features. Compared with CNNs, these networks can make full use of the Kace information flow in the network during feature extraction, which reduces information loss and enables the model to learn higher-quality abstract representations [36]. Then, we modified the network structure to a modular network. The MCC, G-mean, AUC, and AUPR also improved, as shown in columns 3 and 4 of Table 3. This improvement indicates that it is reasonable to introduce the modular network structure design idea into MDC-Kace. The construction of three submodules, i.e., the sequence, physicochemical and structure modules, can effectively avoid information crosstalk and further reduce information loss during the feature learning process. These submodules enable the network to better learn the advanced representations. The results are consistent with the findings of previous studies [29], [34], [35], [47]. Protein structural property information has been proven to play an important role in predicting PTM sites [23]–[26]. Thus, we added protein structural property information as input into the model. The MCC, G-mean, AUC, and AUPR increased by 2.15%, 0.47%, 1.41% and 3.03% respectively, compared to the model from column 4 of Table 3. These results indicate that the protein local and global structural features in the protein structural properties are important parts of the motif pattern of Kace sites. Introducing protein structural properties can enrich the feature space of sites, and our structure module can be a good structure with which to extract them. The function of the SE layer was also analyzed (column 6 of Table 3). We found that the model performance was further improved when the SE layers were introduced after the stacked densely connected convolutional blocks. This finding implied that the SE layer can force the network to learn more important features and enhance the network representation ability while learning feature weights to achieve feature adaptive dynamic fusion.

D. VISUALIZATION OF LEARNED FEATURES
To analyze the difference between lysine acetylated and non-acetylated sites effectively, we adopted the t-distributed stochastic neighbor embedding (t-SNE) algorithm [48] to project the coding vectors and the advanced abstract features extracted by MDC-Kace into a 2-dimensional space based on all the positive samples and negative samples in the independent test dataset described in Table 1. Then, we normalized the values to $[-1, 1]$, as shown in Figure 7. Figure 7 shows that the original coding vectors of the lysine acetylated and non-acetylated site motifs are mixed, and it is difficult to separate them. In contrast, the advanced abstract representations from MDC-Kace show a relatively good distinction. Through the t-SNE visualization tool, we demonstrated that the original coding vector of a motif can be transformed into a well-discriminated advanced representation through MDC-Kace, which is helpful for further analysis of Kace sites.

E. PREDICTING POTENTIAL KACE SITES
In this section, to evaluate the ability of MDC-Kace to identify unknown Kace sites, we performed an analysis based on the results of the independent test dataset described in Table 1. Table 4 lists the top 20 candidate sites predicted by MDC-Kace to be acetylated. Similar to reference [29], we manually checked the top 20 candidate sites in PLMD and the protein database UniProt (https://www.uniprot.org).
Overall, 13 out of 20 candidate sites (65%) were truly acetylated. This result shows that the proposed model, MDC-Kace, makes full use of the three types of information, i.e., the original protein sequence, the amino acid physicochemical properties and the protein structural properties. The deep hidden characteristics of Kace sites can be extracted by considering the information complementation between low-level and high-level features and the feature importance evaluation. Therefore, MDC-Kace can detect potential Kace sites effectively, which is helpful for the investigations of related pathogenic process mechanisms.

### TABLE 4. The top 20 candidate sites from the human independent test acetylated proteins described in Table 1. The sites where acetylation modification actually occurs are highlighted in bold.

| Rank | Protein | Position | Score  | Rank | Protein | Position | Score  |
|------|---------|----------|--------|------|---------|----------|--------|
| 1    | P23588  | 586      | 0.9969 | 11   | Q04837  | 103      | 0.9915 |
| 2    | Q9BXV9  | 94       | 0.9960 | 12   | P68036  | 64       | 0.9913 |
| 3    | O60884  | 223      | 0.9954 | 13   | P49750  | 646      | 0.9912 |
| 4    | Q13029  | 39       | 0.9952 | 14   | Q9P275  | 469      | 0.9909 |
| 5    | Q9NZ01  | 22       | 0.9947 | 15   | Q14289  | 457      | 0.9906 |
| 6    | Q16526  | 565      | 0.9932 | 16   | P10911  | 755      | 0.9903 |
| 7    | Q5T5U3  | 690      | 0.9924 | 17   | Q13029  | 48       | 0.9901 |
| 8    | A8MYA2  | 20       | 0.9924 | 18   | Q96KR1  | 1065     | 0.9900 |
| 9    | Q9NW08  | 150      | 0.9920 | 19   | P22307  | 479      | 0.9899 |
| 10   | P20290  | 93       | 0.9917 | 20   | Q13126  | 32       | 0.9896 |

### IV. CONCLUSION AND FUTURE WORK

Kace modification involves various cell physiology processes and is closely related to metabolic diseases such as diabetes and cancer. Therefore, as the first step in the study of Kace modification, identifying Kace sites is of great significance. Due to the high cost of experimental techniques to verify Kace sites and the low feature learning efficiency of current calculation methods, it is necessary to develop more efficient methods.

In this work, we considered that using only protein sequence level information would be too narrow; thus, protein structural properties were introduced to enrich the feature space of sites. Then, we learned the advanced features of the original protein sequence, the amino acid physicochemical properties and the protein structural properties of the sites through modular subnets, which avoided information crosstalk. During the feature learning process, taking the complementation of the information from features at different levels into account, dense jump connections were used to focus on low-level and high-level features simultaneously to reduce information loss. The SE layer evaluated the importance of the features, weighted them, and realized their adaptive dynamic fusion. Finally, the fused advanced feature was used to predict potential Kace sites. The results of ten-fold cross-validation and independent testing on the three species datasets (human, Mus musculus and Escherichia coli) showed that our model MDC-Kace achieves a comparable or better performance than existing Kace site predictors. Thus, MDC-Kace can help researchers to better identify potential Kace sites. The results of ablation experiments showed that the protein structural properties, the network structure design of the modular densely connected convolutional networks and the SE layer all contribute to improving the model’s prediction ability. The feature visualization result also showed that the architecture of MDC-Kace can convert the three types of information, i.e., the original protein sequence, the amino acid physicochemical properties and the protein structural properties of a Kace site motif into a meaningful abstract representation. In this way, the task of site prediction is completed. Additionally, the prediction and analysis of potential Kace sites further indicated that MDC-Kace is helpful for the discovery and research of new sites. The model can generate useful site modification information for the development of drugs to treat metabolic diseases. Moreover, MDC-Kace is a successful exploration of deep-learning in predicting Kace sites, which provides a reference for the study of PTMs.

Although MDC-Kace has shown promising performance in Kace site prediction, it still has some limitations. The deep-learning method is still a black box, and it lacks meaningful explanations of biological processes [49]. Our future work will concentrate on biological interpretation and will consider some effective structures, such as spatial attention [50] and modular attention [35], to improve the framework. Owing to the limited number of known Kace sites, in the next work, we will consider additional methods to process the unbalanced datasets during the model training period. Protein functional features are also important for PTM site prediction [20], therefore, we will incorporate this information in future studies.

### REFERENCES

1. 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, Feb. 2013, Art. no. e55844.

2. M. Audagnotto and M. D. Peraro, “Protein post-translational modifications: In silico prediction tools and molecular modeling,” Comput. Struct. Biotechnol. J., vol. 15, pp. 307–319, Jan. 2017.

3. Y. Liu, M. Wang, J. Xi, F. Luo, and A. Li, “PTM-ssMP: A Web server for predicting different types of post-translational modification sites using novel site-specific modification profile,” Int. J. Biol. Sci., vol. 14, no. 8, pp. 946–956, 2018.

4. D. Wang, Y. Liang, and D. Xu, “Capsule network for protein post-translational modification site prediction,” Bioinformatics, vol. 35, no. 14, pp. 2386–2394, Jul. 2019.
[5] W.-R. Qiu, B.-Q. Sun, X. Xiao, Z.-C. Xu, and K.-C. Chou, “IPTM-nLys: Identifying multiple lysine PTM sites and their different types,” Bioinformatics, vol. 32, no. 20, pp. 3116–3123, Jun. 2016.

[6] C. Choudhary, B. T. Weinert, Y. Nishida, E. Verdin, and M. Mann, “The growing landscape of lysine acetylation links metabolism and cell signaling,” Nature Rev. Mol. Cell Biol., vol. 15, no. 8, pp. 536–550, Aug. 2014.

[7] W. Deng, C. Wang, Y. Zhang, Y. Xu, S. Zhang, Z. Liu, and Y. Xue, “GPS-PAIL: Prediction of lysine acetylation and deacetylation sites from protein sequences,” Sci. Rep., vol. 6, no. 1, Dec. 2016, Art. no. 39787.

[8] C. A. Butler, P. D. Veith, M. F. Nieto, S. G. Dashper, and E. C. Reynolds, “Lysine acetylation is a common post-translational modification of key metabolic pathway enzymes of the anaerobe porphyromonas gingivalis,” J. Proteomics, vol. 128, pp. 352–364, Oct. 2015.

[9] S.-M. Zhao et al., “Regulation of cellular metabolism by protein lysine acetylation,” Science, vol. 327, no. 5968, pp. 1000–1004, Feb. 2010.

[10] R. Heffernan, Y.-D. Yang, K. Paliwal, and Y.-Q. Zhou, “DeepPhos: Prediction of lysine acetylation sites with deep learning,” Bioinformatics, vol. 35, no. 16, pp. 2766–2773, Jan. 2019.

[11] R. Heffernan, N. He, Y. Huang, W. T. Qiu, X. Liu, and L. Li, “Integration of a deep learning classifier with a random forest approach for predicting lysine acetylation,” Genomics, Proteomics Bioinf., vol. 16, no. 6, pp. 451–459, Dec. 2018.

[12] G. Vligoritićević, M. Barot, and R. Bonneau, “DeepNF: Deep network fusion for protein function prediction,” Bioinformatics, vol. 34, no. 22, pp. 3873–3881, Jun. 2018.

[13] Q.-Z. Zhao, T. Shao, Y.-T. Wang, Z. Tang, Y. Chen, L. Cai, and H.-B. Li, “M2Det: A single-shot object detector based on multi-level feature pyramid network,” in Proc. AAAI, Honolulu, HI, USA, Jan. 2019, pp. 9259–9266.

[14] R.-H. Hu, J. Andrews, M. Rohrbach, T. Darrell, and K. Saenko, “Learning to reason: End-to-end module networks for visual question answering,” in Proc. ICCV, Venice, Italy, Oct. 2017, pp. 804–813.

[15] L.-C. Yu, Z. Lin, X.-H. Shen, J.-M. Yang, X. Lu, M. Bansal, and T. L. Berg, “J-PINNet: Modular attention network for referring expression comprehension,” in Proc. CVPR, Salt Lake City, UT, USA, Jun. 2018, pp. 1307–1315.

[16] G. Wang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proc. CVPR, Honolulu, HI, USA, Jun. 2017, pp. 4700–4708.

[17] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in Proc. CVPR, Salt Lake City, UT, USA, Jun. 2018, pp. 7132–7141.

[18] L. Kienzer, J. D. Bendtsen, and N. Blom, “NetAcet: Prediction of N-terminal acetylation sites,” Bioinformatics, vol. 21, no. 7, pp. 1269–1270, Nov. 2004.

[19] M. Wu, Y. Yang, H. Wang, and Y. Xu, “A deep learning method to more accurately recall known lysine acetylation sites,” BMC Bioinf., vol. 20, no. 1, Art. no. 99.

[20] H. Xu, J. Zhou, S. Lin, W. Deng, Y. Zhang, and Y. Xue, “PLMD: An updated data resource of protein lysine modifications,” J. Genet. Genomics, vol. 44, no. 5, pp. 243–250, May 2017.

[21] Y. Huang, B. Niu, Y. Gao, L. Fu, and W. Li, “CD-HIT suite: A Web server for clustering and comparing biological sequences,” Bioinformatics, vol. 26, no. 5, pp. 680–682, May 2010.

[22] W. R. Atchley, J. Zhao, A. D. Fernandez, and T. Druke, “Solving the protein sequence metric problem,” Proc. Nat. Acad. Sci. USA, vol. 102, no. 21, pp. 7395–7400, May 2005.

[23] A. C. Clevert, D. Unterthiner, and S. Hochreiter, “Fast and accurate deep network learning by exponential linear units (ELUs),” in Proc. ICLR, San Juan, PR, USA, May 2016.

[24] D. A. Clevert, D. Unterthiner, and S. Hochreiter, “Fast and accurate deep network learning by exponential linear units (ELUs),” in Proc. ICLR, San Juan, PR, USA, May 2016.

[25] D. A. Clevert, D. Unterthiner, and S. Hochreiter, “Fast and accurate deep network learning by exponential linear units (ELUs),” in Proc. ICLR, San Juan, PR, USA, May 2016.
[49] J. Ma, M. K. Yu, S. Fong, K. Ono, E. Sage, B. Demchak, R. Sharan, and T. Ideker, “Using deep learning to model the hierarchical structure and function of a cell,” Nature Methods, vol. 15, no. 4, pp. 290–298, Mar. 2018.
[50] L. Chen, H.-W. Zhang, J. Xiao, L.-Q. Nie, J. Shao, W. Liu, and T. S. Chua, “SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning,” in Proc. CVPR, Honolulu, HI, USA, Jul. 2017, pp. 5659–5667.

HUIQING WANG received the Ph.D. degree from the Taiyuan University of Technology, China, in 2011. Since 2003, she has been working with the Taiyuan University of Technology, where she is currently an Assistant Professor with the College of Information and Computer. Her research interests include proteomics, deep learning, and data analysis in bioinformatics.

ZHILIANG YAN received the B.Eng. degree from the Taiyuan University of Technology, China, in 2017, where he is currently pursuing the degree with the College of Information and Computer, where he is currently pursuing the degree with the College of Information and Computer. His research interests include post-translational modifications and deep learning.

DAN LIU received the B.Eng. degree from Harbin Normal University, China, in 2017. She is currently pursuing the degree with the College of Information and Computer, Taiyuan University of Technology. Her research interests include bioinformatics and deep learning.

HONG ZHAO received the B.Eng. degree from Dalian Minzu University, China, in 2018. She is currently pursuing the degree with the College of Information and Computer, Taiyuan University of Technology. Her research interests include protein post-translational modifications and deep learning.

ZHILIANG YAN received the B.Eng. degree from the Taiyuan University of Technology, China, in 2017, where he is currently pursuing the degree with the College of Information and Computer, where he is currently pursuing the degree with the College of Information and Computer. His research interests include post-translational modifications and deep learning.

JIAN ZHAO received the B.Eng. degree from the Inner Mongolia University of Technology, China, in 2018. He is currently pursuing the degree with the College of Information and Computer, Taiyuan University of Technology. His research interests include sequence analysis and deep learning.