frontiers
in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 10 May 2021
doi: 10.3389/fcell.2021.664669

4mCPred-MTL: Accurate Identification of DNA 4mC Sites in Multiple Species Using Multi-Task Deep Learning Based on Multi-Head Attention Mechanism

Rao Zeng[1], *Song Cheng*[2]* *and Minghong Liao*[1]*

[1] Department of Software Engineering, School of Informatics, Xiamen University, Xiamen, China, [2] Department of Thoracic Surgery, Heilongjiang Province Land Reclamation Headquarters General Hospital, Harbin, China

OPEN ACCESS

Edited by:
Yongchun Zuo,
Inner Mongolia University, China

Reviewed by:
Lin Ning,
University of Electronic Science and Technology of China, China
Yijie Ding,
Suzhou University of Science and Technology, China

***Correspondence:**
Song Cheng
csgsl01@163.com
Minghong Liao
liao@xmu.edu.cn

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Cell and Developmental Biology

Received: 05 February 2021
Accepted: 17 March 2021
Published: 10 May 2021

Citation:
Zeng R, Cheng S and Liao M (2021) 4mCPred-MTL: Accurate Identification of DNA 4mC Sites in Multiple Species Using Multi-Task Deep Learning Based on Multi-Head Attention Mechanism. Front. Cell Dev. Biol. 9:664669. doi: 10.3389/fcell.2021.664669

DNA methylation is one of the most extensive epigenetic modifications. DNA 4mC modification plays a key role in regulating chromatin structure and gene expression. In this study, we proposed a generic 4mC computational predictor, namely, 4mCPred-MTL using multi-task learning coupled with Transformer to predict 4mC sites in multiple species. In this predictor, we utilize a multi-task learning framework, in which each task is to train species-specific data based on Transformer. Extensive experimental results show that our multi-task predictive model can significantly improve the performance of the model based on single task and outperform existing methods on benchmarking comparison. Moreover, we found that our model can sufficiently capture better characteristics of 4mC sites as compared to existing commonly used feature descriptors, demonstrating the strong feature learning ability of our model. Therefore, based on the above results, it can be expected that our 4mCPred-MTL can be a useful tool for research communities of interest.

Keywords: multi-task learning, feature sharing, DNA 4mC modification, epigenetics, deep learning, transformer

INTRODUCTION

Epigenetics refers to the reversible and heritable changes in gene function when there is no change in the nuclear DNA sequence (Zuo et al., 2020). Epigenetic phenomena include DNA methylation, RNA interference, histone modification, etc. (Tang W. et al., 2018; Wang et al., 2018; Liu et al., 2019; Hong et al., 2020; Lv et al., 2020a; Zhang D. et al., 2020; Min et al., 2021). Among them, DNA methylation is one of the most extensive epigenetic modifications (Zhu et al., 2019). It is a form of DNA chemical modification that can change genetic performance without changing the DNA sequence. DNA methylation refers to the binding of a methyl group to the cytosine 5 carbon covalent bond of genomic CpG dinucleotides under the action of DNA methyltransferase (Jin et al., 2011; Lv et al., 2020b). A large number of studies have shown that DNA methylation can cause changes in chromatin structure, DNA conformation, DNA stability, and the way that DNA interacts with proteins, thereby controlling gene expression (Jin et al., 2011; Zeng et al., 2016; Zhang et al., 2019; Luo et al., 2020; Shen and Zou, 2020). DNA 4mC has been reported as an effective DNA modification, which can protect its own DNA from restriction enzyme-mediated degradation

(Chen et al., 2017; Wei et al., 2019b). Currently, we have relatively little knowledge regarding 4mC modifications. In order to further study its regulatory mechanism and its biological impact on the organism, it is critical to identify the distribution of 4mC sites in the whole genome.

With the development of high-throughput sequencing technology, 4mC sites can be effectively identified through web-lab biochemical experiments (Flusberg et al., 2010), but this kind of method is time-consuming and labor-intensive. Therefore, it is necessary to develop a computational model that can efficiently and accurately predict and identify 4mC sites. Chen et al. (2017) first developed a tool, namely, iDNA4mC for predicting 4mC sites by establishing a feature set based on chemical properties and occurrence frequency of nucleotides and training a support vector machine (SVM)-based predicting model. In order to take into account more of the physical and chemical properties of DNA, He et al. (2018) proposed 4mCPred, also an SVM-based predictor that used position-specific trinucleotide propensity (PSTNP) and electron–ion interaction potential (EIIP) for feature extraction. In particular, they further optimize the features based on F-score to enhance the generalization ability of the model. Similarly, through four feature coding schemes and using two-step feature optimization method, Wei et al. (2019a) constructed a prediction model called 4mCPred-SVM, which is shown to perform better than previous methods on benchmarking comparison. Later, Manavalan et al. (2019b) first proposed the meta-predictor Meta-4mCpred for predicting 4mC sites. It used a variety of feature extraction methods to convert DNA sequences into a total of 14 feature descriptors and trained four different classifiers. Particularly, meta-4mCpred exhibits good performance with independent test, demonstrating the excellent generalization ability. To make full use of the advantages of each prediction method mentioned above, Tang et al. (2020) developed DNA4mC-LIP, which for the first time linearly integrated all the previous methods for the 4mC site prediction. In recent years, deep learning has been widely used in the field of bioinformatics. Xu et al. (2020) developed the first deep learning Deep4mC, which converted sequences into digital vectors through binary, enhanced nucleic acid composition (ENAC), EIIP, and nucleotide chemical property (NCP) feature encoding schemes and inputted them into two convolutional layers without pooling layers and the attention layers. The average area under the ROC (receiver operating characteristic) curve (AUC) values of its prediction for multiple species were greater than 0.9 in multiple cross-validations. In our previous work, we proposed a two-layer deep learning model called Deep4mcPred, which utilizes a hybrid network of ResNet and long short-term memory (LSTM) (Zeng and Liao, 2020).

Although much progress has been made by the methods mentioned above, the performance is still not satisfactory. Moreover, most existing predictors are designed for one specific species. Although they provide a cross-species model and validation test, the performance is always not that good as compared to the original species-specific model. Therefore, to address this problem, we established a generic 4mC predictor, namely, 4mCPred-MTL using multi-task learning coupled with Transformer, which is a widely used NLP (natural language processing) technique, to predict 4mC sites in multiple species. In this predictor, we utilize a multi-task learning framework, in which each task is to train species-specific data based on Transformer. Extensive experimental results show that our multi-task predictive model can significantly improve the performance of the model based on a single task and outperform existing methods. Moreover, we found that the feature representations learned from our model can capture better characteristics of 4mC sites as compared to the existing commonly used feature descriptors, demonstrating the strong feature learning ability. Therefore, based on the above results, it can be expected that our 4mCPred-MTL can be a useful tool for research communities of interest.

MATERIALS AND METHODS

Datasets

Previous studies have demonstrated that a stringent dataset is essential for building a robust predictive model (Liang et al., 2017; Zeng and Liao, 2020; Su et al., 2021). In our previous work (Zeng and Liao, 2020), we constructed large-scale datasets for three species, including *Arabidopsis thaliana* (*A. thaliana*), *Caenorhabditis elegans* (*C. elegans*), and *Drosophila melanogaster* (*D. melanogaster*). As for the positive samples, there are 20,000 positive samples, and each sample is a 41-bp-long sequence centered with true 4mC sites. Similarly, the dataset contains the same number of negative samples, which are cytosine-centered sequences with lengths of 41 bp but are not recognized by the single-molecule, real-time (SMRT) sequencing technology.

Training Set and Independent Test Set

Considering the performance, most of the existing predictors are evaluated by cross validation test, which might produce performance bias; we here randomly split the datasets into (Zuo et al., 2020) training set for model training and evaluation and (Liu et al., 2019) independent test set for model robustness validation. Thus, we randomly divided the dataset into training set and testing set with the ratio of 8:2, resulting in 16,000 samples in the training set and 4,000 samples in the testing set. The details of the datasets are presented in **Table 1**. Notably, for fair comparison, all the existing methods are evaluated on the test set.

Architecture of 4mCPred-MTL

The network architecture of our model is illustrated in **Figure 1**. This network architecture consists of three main components: (i) sequence processing module, (ii) sharing module, and (iii)

TABLE 1 | Summary of benchmark datasets in three species.

Species	Training set		Testing set	
	Positives	Negatives	Positives	Negatives
A. thaliana	16,000	16,000	4,000	4,000
C. elegans	16,000	16,000	4,000	4,000
D. melanogaster	16,000	16,000	4,000	4,000

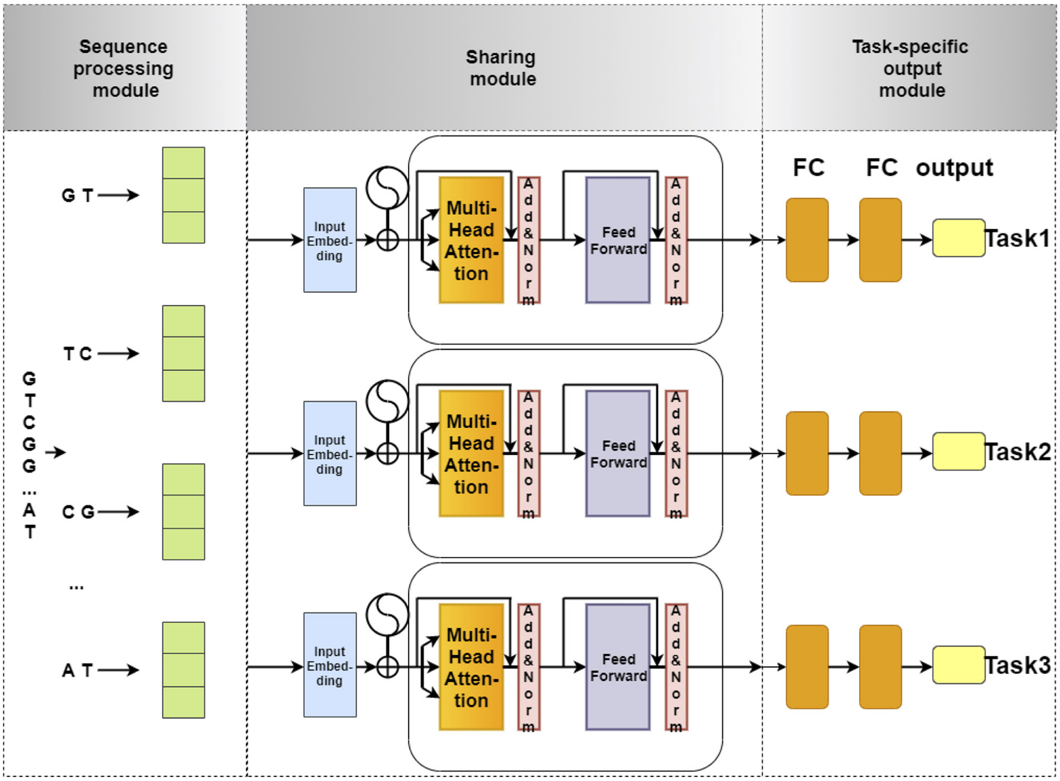

FIGURE 1 | The flowchart of 4mCPred-MTL. The sequence processing module uses 2-gram to split an original DNA sequence into overlapping subsequences and converts them into feature vectors by one-hot encoding. Next, the feature vectors of subsequences are fed into the sharing module, containing a Transformer encoder and a *max-pooling* layer, to capture the sharing information among different species. Finally, the output of the sharing module is fed into the task-specific output module to predict the 4mC site of a certain species.

task-specific output module. The sequence processing module is designed to encode the DNA sequences into feature matrices by one-hot encoding (Quang and Xie, 2016; Zou et al., 2019; Dao et al., 2020a). Next, the encoded matrix is passed through a Transformer, which is a popular technique for embedding different levels of dependency relationships between subsequences. Afterward, we used a max-pooling layer to automatically measure which feature plays a key role in the target task in each unit of the Transformer. Finally, the features derived from the max-pooling layer is fed to the task-specific output module to identify 4mC sites in three species, respectively. The task-specific output module contains three parts, and each part consists of fully connected layers that are designed in terms of the size of the training set for each species. The model is implemented using Pytorch. Each module of our model is described in detail as follows.

Sequence Processing Module

We first employed n-gram nucleobases to define “words” in DNA sequences (Dong et al., 2006; Zeng et al., 2018; Fu et al., 2020; Lin et al., 2020; Liu X. et al., 2020; Wang et al., 2020; Yang et al., 2020; Zhang Z. Y. et al., 2021). The n-grams are the set of all possible subsequences of nucleotides. Afterward, the DNA sequences are segmented into overlapping n-gram nucleotides. The number of possibilities is 4^n, since there are four types of nucleotides. To prevent the sparsity in the encoding, the n-gram number n is set to 2. For example, we split a DNA sequence into overlapping 2-gram nucleotide sequences as follows: *GTTGT. . .CTT*→ “*GT*,” “*TT*,” “*TG*,” “*GT*,” . . ., “*CT*,” “*TT*.”

For a given DNA sequence P with length L, it can be denoted as follows:

$$P = \ R_1, R_2, , R_L \tag{1}$$

where R_i is the ith word. These words are first randomly initialized and embedded by one-hot embedding, which is referred to as “word embeddings.” Here, we define the sequence of word embeddings as

$$\mathbf{x}_1, \mathbf{x}_2, , \mathbf{x}_L \ (2) \tag{2}$$

where $x_i \in \mathbb{R}^d$ is the d-dimensional embedding of the ith word.

Sharing Module

Attention Mechanism

The attention mechanism was proposed by Bahdanau et al. (2014) in the application of neural machine translation. The Attention mechanism is somewhat similar to the idea of human translating articles, that is, paying attention to the corresponding context of our translation part. For example, we can get the hidden states of the recurrent neural network (RNN) encoder: $(h_1, h_2, , h_t)$. By assuming the current decoder hidden state is s_{t-1}, we can calculate the correlation between each input position j and the current output position:

$$\overrightarrow{c_t} = \left(a\left(s_{t-1}, h_1\right), ?\cdots, a\left(s_{t-1}, h_T\right)\right) \tag{3}$$

where a is a correlation operator, such as dot product. We can get the attention distribution by normalizing the $\overrightarrow{c_t}$. The expanding form of the attention is

$$\mathbf{a}_{\mathbf{t}j} = \frac{\mathbf{e}xp\left(\mathbf{c}_{\mathbf{t}j}\right)}{\sum_{\mathbf{k=1}}^{\mathbf{T}} \mathbf{e}xp\left(\mathbf{c}_{\mathbf{t}k}\right)}. \tag{4}$$

Therefore, attention is a weight vector. These weights represent which tokens the machine focuses on. When the attention distribution is obtained, the weight of the more important input position for the current output position is obtained, which accounts for a larger proportion when predicting the output. By introducing the attention mechanism, we can only use the final single vector result of the encoder, so that the model can focus on all the input information that is important for the next target word, and the model effect is greatly improved.

Transformer With Multi-Head Attention

The development of deep learning (Dao et al., 2020b; Liu Y. et al., 2020; Long et al., 2020; Naseer et al., 2020; Zhang T. et al., 2020; Zhang Y. et al., 2020) in NLP is filled with RNN and LSTM. Transformer models completely abandon the RNN and LSTM layers and only use the attention mechanism for feature extraction. After the input has been embedded to matrix form, we first use the position encoding layer. Since the model has no recurrent or convolutional layers, there is no clear relative or absolute information about the position of the word in the source sentence. In order to let the model learn the position information better, position encoding is added and superimposed on the word embedding. An encoding method using trigonometric functions maintains its position invariance.

The position encoding function can be presented as

$$PE_{(pos,2i)} = \sin\left(pos/10,000^{2i/d_{model}}\right) \tag{5}$$

$$PE_{(pos,2i1)} = \cos\left(pos/10,000^{2i/d_{model}}\right) \tag{6}$$

where pos is the position of each token; $2i$ and $2i1$ are the even-numbered and odd-numbered dimensions of each token position vector of the cardinality, respectively, where all position subscripts start from 0; and d_{model} is the dimensionality of word vector, the same as the dimensionality of encoding.

Diving into the encoder of Transformer, we will first meet the multi-head attention module. The multi-head attention is actually a combination of multiple self-attention structures. Each head learns its characteristics in different representation spaces. The first step in calculating self-attention is to construct three vectors based on the input vector of the encoder. In our task, it is the embedding of each sequence. So for each embedding, we need to create a Query matrix, a Key matrix, and a Value matrix. These three matrices are created during the training process, all from the same input. The self-attention function can be written as

$$SA\left(Q, K, V\right) = softmax\left(\frac{QK}{\sqrt{d_k}}\right)V. \tag{7}$$

First, we need to calculate the dot product between Q and K. To prevent the result from being too large, we will divide it by a scale of $\sqrt{d_k}$, which is the dimension of query and key vectors. Then a Softmax operation is implemented to normalize the result to a probability distribution, and then it is multiplied by the matrix V to get the weighted summation. Multi-head attention means that we can have different Qs, Ks, and Vs representations and finally combine the results. For the encoder, these basic units are concatenated, where the keys, queries, and values are all from the output of the previous layer of encoder; that is, every position of the encoder can notice all the positions of the previous layer of encoder.

After the attention is achieved, we come to the Add-and-Norm module. The "Add" in it stands for residual connection (He et al., 2016), which is designed to solve the problem of difficult training of multi-layer neural networks. By passing the information of the last layer to the next layer without difference, it can effectively focus on only the difference part. On the other hand, "Norm" is short for the layer normalization (Ba et al., 2016). It can speed up the training process and make the model converge faster by normalizing the activation value of the layer.

Max-Pooling Layer

The feature vector **h** of each subsequence is fed into a *max-pooling* layer to capture the most significant feature in identifying the DNA modification to represent this subsequence. Then, all the most significant features of subsequences are concatenated into a vector to represent a DNA sequence, which is shown in the following equation:

$$\mathbf{y} = max_{i=1}^{n}\ \mathbf{h}_i \tag{8}$$

where i is the ith subsequence, n is the number of subsequences in a DNA sequence, and **y** is regarded as the feature vector of a

TABLE 2 | Performance comparison of the proposed method and existing single-task 4mC predictors.

Species	Method	SN (%)	SP (%)	ACC (%)	MCC
A. thaliana	4mcPred-IFL	70.4	**84.9**	77.7	0.559
	4mcPred_SVM	72.3	81.1	76.7	0.536
	Deep4mcPred	81.3	84.8	83.1	0.661
	Proposed	**89.7**	83.6	**86.5**	**0.728**
C. elegans	4mcPred-IFL	45.4	79.4	62.4	0.263
	4mcPred_SVM	43.7	75.4	59.5	0.201
	Deep4mcPred	75.6	**88.5**	82.0	0.646
	Proposed	**83.8**	83.2	**83.3**	**0.665**
D. melanogaster	4mcPred-IFL	65.5	**87.6**	76.5	0.544
	4mcPred_SVM	65.8	84.5	75.1	0.511
	Deep4mcPred	84.6	84.8	84.7	0.693
	Proposed	**88.0**	84.1	**86.0**	**0.722**

The bold denotes the best performance.

target sequence. The max-pooling layer attempts to find the most important dependencies in subsequences.

Task-Specific Output Module

This module consists of four sets of fully connected layers corresponding to each task. In each fully connected layer with a *relu* activation function, its output is calculated by the following equation:

$$\mathbf{f}_i^j = relu(\mathbf{W}_i^j \mathbf{f}_{i-1}^j \mathbf{b}_i^j) \quad (9)$$

where $\mathbf{f}_{i-1}^j$ is the output of the previous layer of jth task, $\mathbf{f}_i^j$ is the current layer output of jth task, $\mathbf{W}_i^j$ is the weight matrix, and $\mathbf{b}_i^j$ is the bias vector. In each layer, the "batch normalization" technique was used to improve generalization performance (Cheng and Baldi, 2006). Finally, a *softmax* layer is added on the top of final output $\mathbf{f}^j$ to perform the final prediction. Note that the parameters of different sets of the fully connected layer are designed differently in terms of the amount of data of the corresponding task.

TABLE 3 | Performance comparison with the model not using the multi-task learning.

Species	Method	SN (%)	SP (%)	ACC (%)	MCC
A. thaliana	Single-task	86.7	**84.2**	85.4	0.708
	Proposed	**89.7**	83.6	**86.5**	**0.728**
C. elegans	Single-task	**85.9**	82.8	**84.4**	**0.688**
	Proposed	83.8	**83.2**	83.3	0.665
D. melanogaster	Single-task	85.7	84.0	84.9	0.698
	Proposed	**88.0**	**84.1**	**86.0**	**0.722**

The bold denotes the best performance.

Training

The task-specific features, **y**, generated by the sharing module, are ultimately sent into one set of fully connected layers in terms of it belonging to which task. For classification tasks, we used binary cross-entropy loss function as the objective:

$$l = \frac{1}{N}\sum_i -[y_i log(p_i)(1 - y_i)log(1 - p_i)] \quad (10)$$

where N denotes the number of training samples, y_i denotes the label (i.e., 1 or 0) of sample i, and p_i denotes the probability that sample i is predicted to be positive. Our

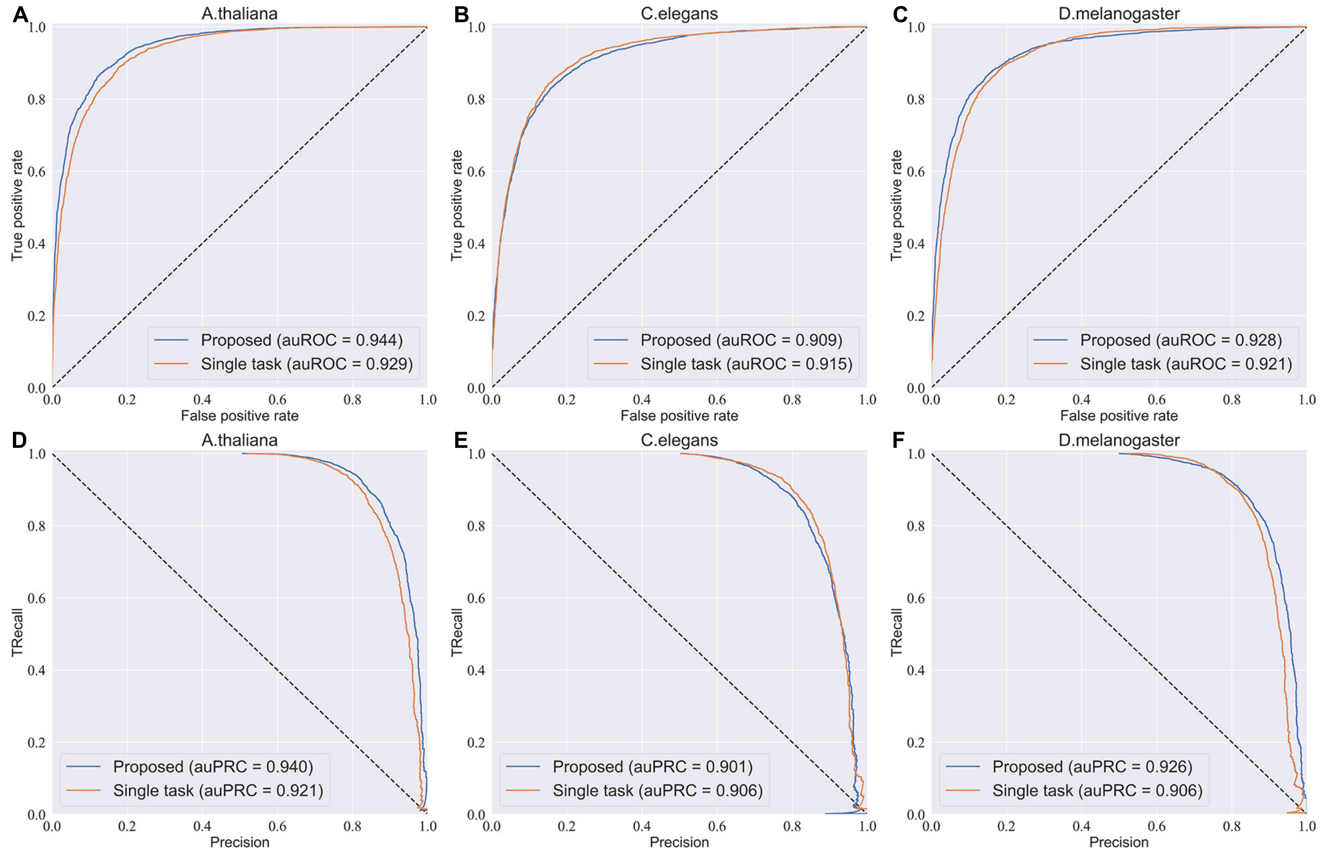

FIGURE 2 | ROC curves and PR curves of the model using multi-task learning and the model not using multi-task. **(A–C)** The ROC curves of the two models in three species. **(D–F)** The PR curves of the two models in three species.

global loss function is the linear combination of loss function for all tasks:

$$l_{all} = \sum_{k=1}^{k} \alpha_k l_k \tag{11}$$

where α_k is the weight for task k.

Evaluation Metrics

Here, we adopted four commonly used metrics to measure the performance of the proposed method and existing methods, including sensitivity (SN), specificity (SP), overall accuracy (ACC), and Matthew's correlation coefficient (MCC) (Wei et al., 2014, 2017a,c, 2018c, 2019a,c,d, 2020b; Feng et al., 2019; Jin et al., 2019; Zou et al., 2019; Hong et al., 2020; Qiang et al., 2020; Su et al., 2019a,b, 2020a; Zhao et al., 2020). They are formulated as follows:

$$SN = \frac{TP}{TP + FN}$$

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

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

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

where TP, TN, FP, and FN represent the numbers of true positives, true negatives, false positives, and false negatives, respectively. MCC and ACC are two metrics used to evaluate the overall prediction ability of a predictive model. In addition, we used the ROC curve to intuitively validate the overall performance. The AUC is to quantitatively evaluate the overall prediction performance of the model (Tang H. et al., 2018; Jin et al., 2020; Zeng et al., 2020; Cai et al., 2021; Zhang D. et al., 2021). The AUC ranges from 0.5 to 1. The higher the AUC score, the better the performance of the model.

RESULTS AND DISCUSSION

Performance Comparison With Other Single-Task State-of-the-Art Methods

To demonstrate the effectiveness of the proposed method, we compared its performance with four other existing single-task state-of-the-art methods on the benchmark dataset, including 4mcPred-IFL (Wei et al., 2019b), 4mcPred_SVM (Wei et al., 2019a), and Deep4mcPred (Zeng

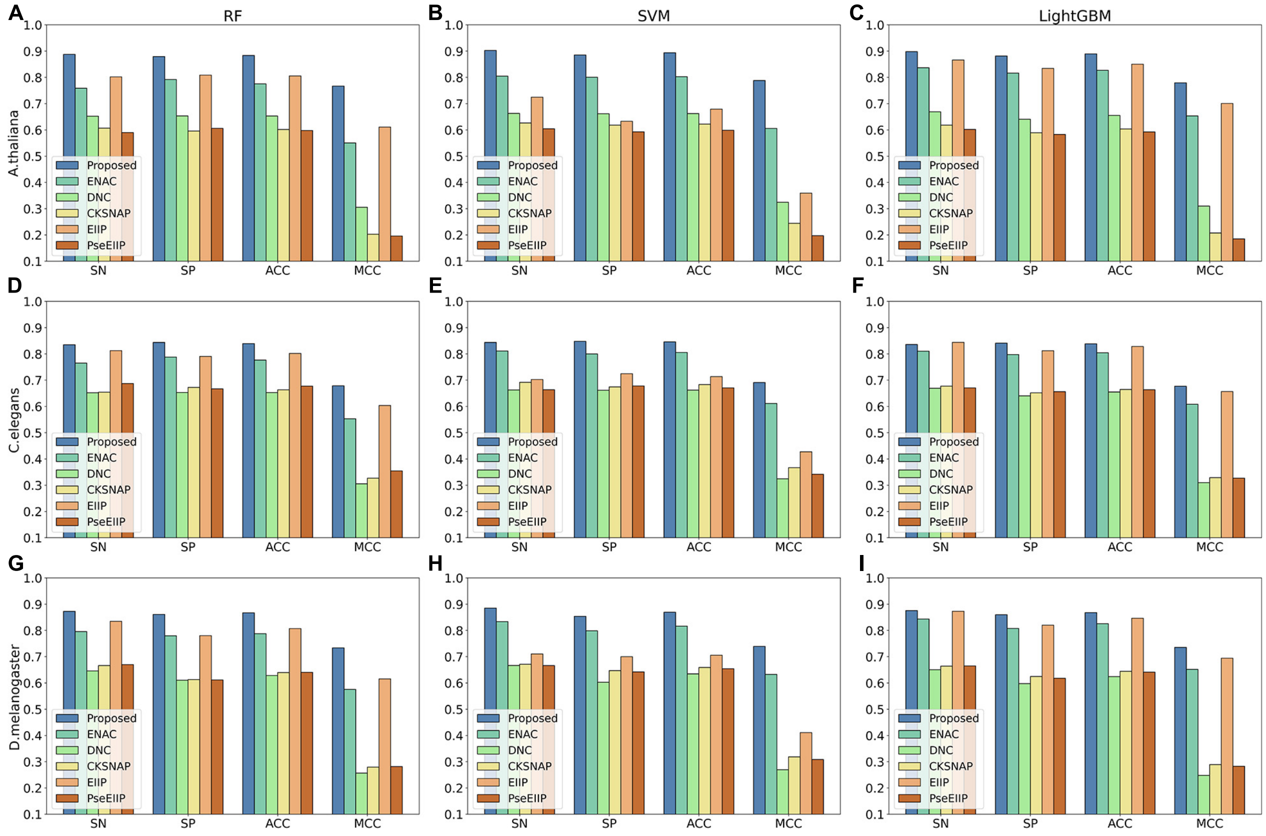

FIGURE 3 | The 10-fold cross-validation results of the Proposed, ENAC, DNC, CKSNAP, EIIP, and PreEIIP methods are based on the three basic classifiers for each species. **(A–C)** The results of the species *A. thaliana*. **(D–F)** The results of the species *C. elegans*. **(G–I)** The results of the species *D. melanogaster*.

and Liao, 2020). It is worth noting that among the three competing methods, except the method Deep4mcPred using deep learning technique, other methods all use traditional machine learning to train the respective models by hand-made features extracted from original DNA sequences. For a fair comparison, the source codes of these methods are used to carry out independent tests on our benchmark dataset.

The results of different methods are listed in **Table 2**. As shown in **Table 2**, we can see that for all species (i.e., *A. thaliana*, *C. elegans*, and *D. melanogaster*), our proposed method significantly outperform all other single-task competing methods in terms of SN, ACC, and MCC, with the only exception that the value of SP of our proposed method is lower than those of other methods. Specifically, for the species *A. thaliana*, when compared to the second-best method Deep4mcPred, our proposed method achieves an SN of 89.7%, an ACC of 86.5%, and an MCC of 0.728, yielding a relative improvement over Deep4mcPred of 10.33, 4.09, and 10.14%, respectively. However, Deep4mcPred does

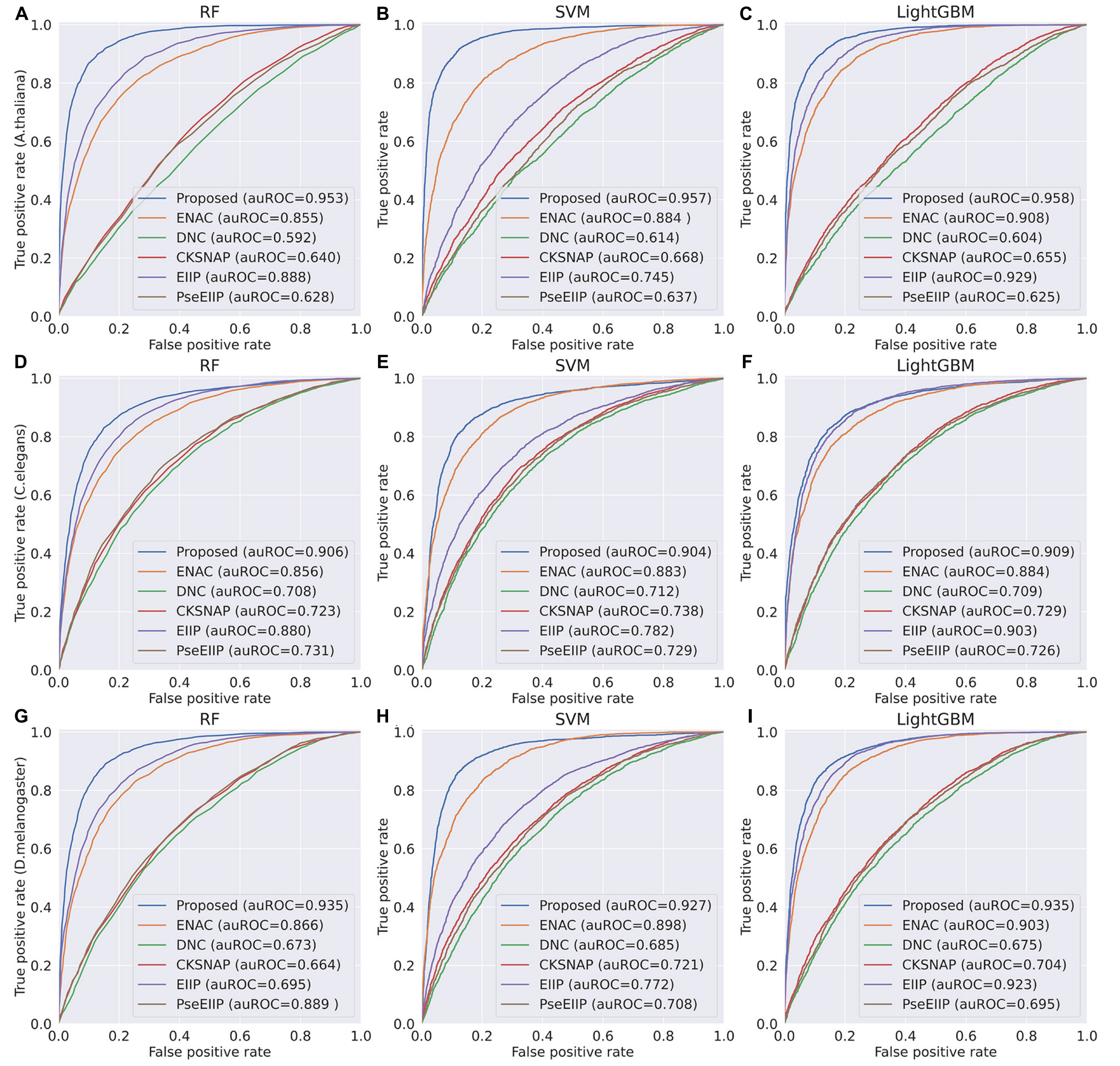

FIGURE 4 | ROC curves of the Proposed, ENAC, DNC, CKSNAP, EIIP, and PreEIIP methods are based on the three basic classifiers for each species. **(A–C)** The results of the species *A. thaliana*. **(D–F)** The results of the species *C. elegans*. **(G–I)** The results of the species *D. melanogaster*.

have a higher SP of 84.8, where our method only reaches an SP of 84.2. For the species *C. elegans*, compared to all competing methods, our proposed method achieves great improvement in terms of SN, ACC, and MCC, which are 6.06, 4.24, and 12.73% higher than that of the runner-up Deep4mcPred. For the species *D. melanogaster*, our proposed method also gets the best performance among all methods, achieving SN of 88.0%, ACC of 86.0%, and MCC of 0.722. Note that although the SP of our proposed methods is worse than those of other methods, the other three metrics are all higher than any competing single-task method. Therefore, we can conclude that our proposed method can achieve the best predictive performance for detecting 4mC sites in multiple species. The reason may be that in our method, we used the Transformer technique to learn more discriminative features based on multi-task learning that can leverage useful information among multiple related learning tasks to help learn a more accurate learner for each task, while the competing methods only use the information from one task. So the results are not surprising that our method achieves the best performance when using multi-task learning.

Effect of Multi-Task Learning

To investigate the efficiency of the multi-task learning technique, we compared the method using multi-task learning, namely,

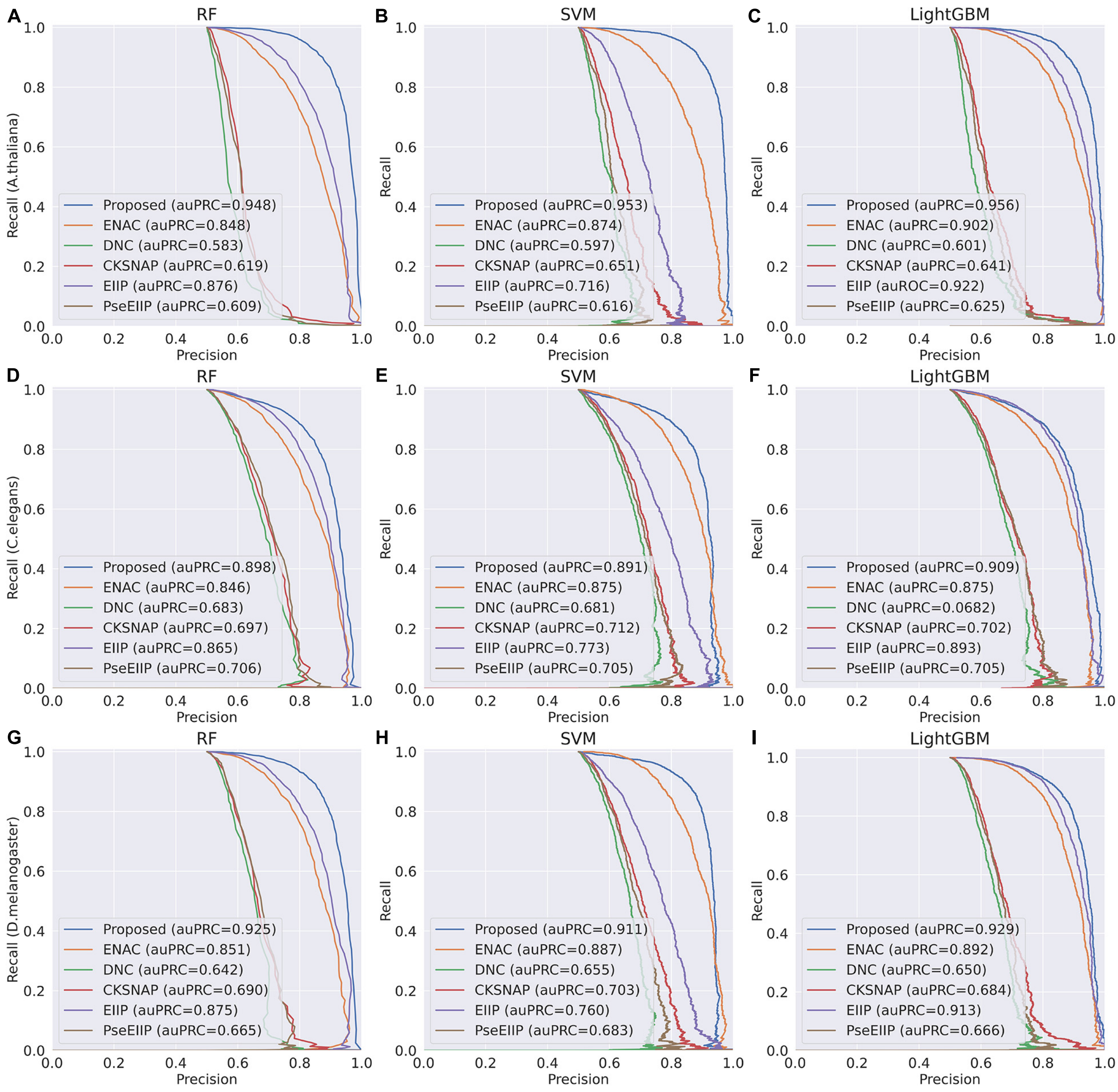

FIGURE 5 | PR curves of the Proposed, ENAC, DNC, CKSNAP, EIIP, and PreEIIP methods are based on the three basic classifiers for each species. **(A–C)** The results of the species *A. thaliana*. **(D–F)** The results of the species *C. elegans*. **(G–I)** The results of the species *D. melanogaster.*

our proposed method, with the method not using multi-task learning. The comparative results obtained are shown in **Table 3**. From **Table 3**, we can see that the method using multi-task learning outperforms the method not using multi-task learning in the species *A. thaliana* and *D. melanogaster*, with only one exception in the species *C. elegans*. in which the performance of the method using multi-task learning is slightly worse than the methods not using multi-task learning. To be specific, for the species *A. thaliana*, the SN, ACC, and MCC of the method using multi-task learning are 3.46, 1.29, and 2.82% higher than those of the method not using multi-task learning, while the SP of the method not using multi-task learning is lower. For *D. melanogaster*, the method using multi-task learning improves the performance from 85.7 to 88.0% in terms of SN, 84.0–84.1% in terms of SP, 84.9–86.0% in terms of ACC, and 69.8–72.2% in terms of MCC. For a more intuitive comparison, we further compared their ROC curve s and PR (precision-recall) curves, which are illustrated in **Figure 2**. We can observe that except in the species *C. elegans*, the method using multi-task learning achieves the best values of auROC and auPRC in the other species. When using multi-task learning, even if the performance of our method is not good in one species, the performance is improved in the other species. Therefore, we can conclude that employing the multi-task learning technique in a feature learning scheme can improve the feature representation ability and predictive performance because the multi-task learning technique aims to enhance the performance of each task by sharing information between related tasks so that they complement each other.

Analysis of Features Extracted From Multi-Task Learning Method on the Test Dataset

Discriminative features play a crucial role in developing a predictive tool with high accuracy. To investigate whether the features learning by our method is more discriminative, we compared them with five traditional hand-made feature descriptors, including ENAC, di-nucleotide composition (DNC), composition of k-spaced nucleic acid pairs (CKSNAP), electron–ion interaction pseudopotentials of trinucleotide (EIIP), and electron–ion interaction pseudopotentials of trinucleotide (PseEIIP). On the test dataset, all the features are evaluated with a 10-fold cross-validation technique by using three basic machine learning classifiers, including random forest (RF), SVM, and LightGBM.

The comparison results are illustrated in **Figure 3**. As shown in **Figure 3**, we can observe that for each species, the features extracted by our proposed method achieve the best performance among other traditional hand-made features in terms of the four metrics on every basic classifier, especially on the classifiers RF and SVM, indicating that the features generated by our proposed method are more effective for 4mC sites prediction in different species and are more suitable for most of the common classifiers.

In the feature learning scheme, we used the transformer network to learn the related information between DNA subsequences and added a max-pool layer to judge which feature plays a key role in detecting 4mC sites in each subsequence. Moreover, the multi-task learning technique was exploited to capture sharing information contained in multiple tasks to help learn a more discriminative and effective feature to represent DNA sequences for 4mC sites prediction. Therefore, the proposed method significantly outperforms other traditional handcraft features, which needs prior knowledge. **Figures 4**, **5** illustrate the ROC and PR curves of different features. It can be also seen that our learned features are more effective than existing handcraft features, further demonstrating that our model can capture more useful information than existing feature algorithms.

CONCLUSION

In this study, we have established a predictor called 4mcPred-MTL, using Transformer-based multi-task learning to predict DNA 4mC modifications in multiple species. To the best of our knowledge, this is the first 4mC predictor that can perform the prediction task for different species on a single run. Importantly, our predictor shows better performance as compared to state-of-the-art prediction tools on independent test, demonstrating the superiority of our model. In particular, via feature comparative analysis, we found that our model can sufficiently capture better characteristics of 4mC sites as compared to existing commonly used feature descriptors, demonstrating the strong feature learning ability of our model. We expect that our model can be a useful predictor for research communities of interest. In addition, we provide a new way to predict multi-species sequence prediction analysis, which can be extended to other bioinformatics fields (Ding et al., 2016a,b, 2019a,b,c,d, 2020a,b,c; Liu et al., 2017; Wei et al., 2017a,b,c, 2018a,b,c, 2020a; Jiang et al., 2018; Jin et al., 2019; Manavalan et al., 2019a,b; Su et al., 2019b, 2020b,c; Wang et al., 2019, 2021a,b; Dai et al., 2020; Guo et al., 2020a,b; Song et al., 2020; Zou et al., 2020; Yang et al., 2021).

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: http://server.malab.cn/Deep4mcPred/Download.html.

AUTHOR CONTRIBUTIONS

RZ surveyed the algorithms and implementations, preprocessed the datasets, and performed all the analyses. SC and ML designed the benchmarking test. All the authors have written, read, and approved the manuscript.

FUNDING

This work was supported in part by the Natural Science Foundation of China.

REFERENCES

Ba, J. L., Kiros, J. R., and Hinton, G. E. (2016). Layer normalization. *arXiv* [preprint] Available online at: https://arxiv.org/pdf/1607.06450.pdf (Accessed July 21, 2016) arXiv:160706450

Bahdanau, D., Cho, K., and Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv* [preprint] Available online at: http://arxiv.org/abs/1409.0473 (Accessed Sep 1, 2014) arXiv:14090473

Cai, L., Ren, X., Fu, X., Peng, L., Gao, M., and Zeng, X. J. B. (2021). iEnhancer-XG: interpretable sequence-based enhancers and their strength predictor. *Bioinformatics* btaa914. doi: 10.1093/bioinformatics/btaa914

Chen, W., Yang, H., Feng, P., Ding, H., and Lin, H. (2017). iDNA4mC: identifying DNA N4-methylcytosine sites based on nucleotide chemical properties. *Bioinformatics* 33, 3518–3523. doi: 10.1093/bioinformatics/btx479

Cheng, J., and Baldi, P. (2006). A machine learning information retrieval approach to protein fold recognition. *Bioinformatics* 22, 1456–1463. doi: 10.1093/bioinformatics/btl102

Dai, C., Feng, P., Cui, L., Su, R., Chen, W., and Wei, L. (2020). Iterative feature representation algorithm to improve the predictive performance of N7-methylguanosine sites. *Brief. Bioinfor.* doi: 10.1093/bib/bbaa278

Dao, F. Y., Lv, H., Yang, Y. H., Zulfiqar, H., Gao, H., and Lin, H. (2020a). Computational identification of N6-methyladenosine sites in multiple tissues of mammals. *Comput. Struct. Biotechnol. J.* 18, 1084–1091. doi: 10.1016/j.csbj.2020.04.015

Dao, F. Y., Lv, H., Zhang, D., Zhang, Z. M., Liu, L., and Lin, H. (2020b). DeepYY1: a deep learning approach to identify YY1-mediated chromatin loops. *Brief. Bioinform.* bbaa356.

Ding, Y., Jiang, L., Tang, J., and Guo, F. (2020a). Identification of human microRNA-disease association via hypergraph embedded bipartite local model. *Comput. Biol. Chem.* 89:107369. doi: 10.1016/j.compbiolchem.2020.107369

Ding, Y., Tang, J., and Guo, F. (2016b). Predicting protein-protein interactions via multivariate mutual information of protein sequences. *BMC Bioinformatics* 17:398. doi: 10.1186/s12859-016-1253-9

Ding, Y., Tang, J., and Guo, F. (2016a). Identification of protein–protein interactions via a novel matrix-based sequence representation model with amino acid contact information. *Int. J. Mol. Sci.* 17:1623. doi: 10.3390/ijms17101623

Ding, Y., Tang, J., and Guo, F. (2019c). Identification of drug-target interactions via fuzzy bipartite local model. *Neural Comput. Appl.* 32, 10303–10319. doi: 10.1007/s00521-019-04569-z

Ding, Y., Tang, J., and Guo, F. (2019b). Identification of drug-side effect association via semisupervised model and multiple kernel learning. *IEEE J. Biomed. Health Inform.* 23, 2619–2632. doi: 10.1109/jbhi.2018.2883834

Ding, Y., Tang, J., and Guo, F. (2019d). Protein crystallization identification via fuzzy model on linear neighborhood representation. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 1–1. doi: 10.1109/tcbb.2019.2954826

Ding, Y., Tang, J., and Guo, F. (2019a). Identification of drug-side effect association via multiple information integration with centered kernel alignment. *Neurocomputing* 325, 211–224. doi: 10.1016/j.neucom.2018.10.028

Ding, Y., Tang, J., and Guo, F. (2020b). Human protein subcellular localization identification via fuzzy model on Kernelized Neighborhood Representation. *Appl. Soft Comput.* 96:106596. doi: 10.1016/j.asoc.2020.106596

Ding, Y., Tang, J., and Guo, F. (2020c). Identification of drug–target interactions via Dual Laplacian regularized least squares with multiple kernel fusion. *Knowl. Based Syst.* 204:106254. doi: 10.1016/j.knosys.2020.106254

Dong, Q.-W., Wang, X.-L., and Lin, L. (2006). Application of latent semantic analysis to protein remote homology detection. *Bioinformatics* 22, 285–290. doi: 10.1093/bioinformatics/bti801

Feng, C. Q., Zhang, Z. Y., Zhu, X. J., Lin, Y., Chen, W., Tang, H., et al. (2019). iTerm-PseKNC: a sequence-based tool for predicting bacterial transcriptional terminators. *Bioinformatics* 35, 1469–1477. doi: 10.1093/bioinformatics/bty827

Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., et al. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7:461. doi: 10.1038/nmeth.1459

Fu, X., Cai, L., Zeng, X., and Zou, Q. J. B. (2020). StackCPPred: a stacking and pairwise energy content-based prediction of cell-penetrating peptides and their uptake efficiency. *Bioinformatics* 36, 3028–3034. doi: 10.1093/bioinformatics/btaa131

Guo, X. Y., Zhou, W., Shi, B., Wang, X. H., Du, A. Y., Ding, Y. J., et al. (2020a). An efficient multiple kernel support vector regression model for assessing dry weight of hemodialysis patients. *Curr. Bioinform.* 15, 466–469.

Guo, X. Y., Zhou, W., Yu, Y., Ding, Y. J., Tang, J. J., and Guo, F. (2020b). A novel triple matrix factorization method for detecting drug-side effect association based on kernel target alignment. *BioMed. Res. Int.* 2020:4675395.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition. Abs," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Las Vegas, NV: IEEE), 770–778.

He, W., Jia, C., and Zou, Q. (2018). 4mCPred: machine learning methods for DNA N4-methylcytosine sites prediction. *Bioinformatics* 35, 593–601. doi: 10.1093/bioinformatics/bty668

Hong, Z., Zeng, X., Wei, L., and Liu, X. (2020). Identifying enhancer–promoter interactions with neural network based on pre-trained DNA vectors and attention mechanism. *Bioinformatics* 36, 1037–1043.

Jiang, L., Xiao, Y., Ding, Y., Tang, J., and Guo, F. (2018). FKL-Spa-LapRLS: an accurate method for identifying human microRNA-disease association. *BMC Genomics* 19:911. doi: 10.1186/s12864-018-5273-x

Jin, B., Li, Y., and Robertson, K. D. (2011). DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes Cancer* 2, 607–617. doi: 10.1177/1947601910393957

Jin, Q., Meng, Z., Tuan, D. P., Chen, Q., Wei, L., and Su, R. (2019). DUNet: a deformable network for retinal vessel segmentation. *Knowl. Based Syst.* 178, 149–162. doi: 10.1016/j.knosys.2019.04.025

Jin, S., Zeng, X., Xia, F., Huang, W., and Liu, X. (2020). Application of deep learning methods in biological networks. *Brief. Bioinform.* 22, 1902–1917. doi: 10.1093/bib/bbaa043

Liang, Z. Y., Lai, H. Y., Yang, H., Zhang, C. J., Yang, H., Wei, H. H., et al. (2017). Pro54DB: a database for experimentally verified sigma-54 promoters. *Bioinformatics* 33, 467–469.

Lin, X., Quan, Z., Wang, Z. J., and Huang, H., and Zeng, X. (2020). A novel molecular representation with BiGRU neural networks for learning atom. *Brief. Bioinform.* 21, 2099–2111. doi: 10.1093/bib/bbz125

Liu, X., Hong, Z., Liu, J., Lin, Y., Rodríguez-Patón, A., Zou, Q., et al. (2020). Computational methods for identifying the critical nodes in biological networks. *Brief. Bioinform.* 21, 486–497. doi: 10.1093/bib/bbz011

Liu, Y., Huang, Y., Wang, G., and Wang, Y. (2020). A deep learning approach for filtering structural variants in short read sequencing data. *Brief. Bioinform.* bbaa370.

Liu, Y., Zeng, X., He, Z., and Zou, Q. (2017). Inferring MicroRNA-disease associations by random walk on a heterogeneous network with multiple data sources. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 14, 905–915. doi: 10.1109/tcbb.2016.2550432

Liu, Z.-Y., Xing, J.-F., Chen, W., Luan, M.-W., Xie, R., Huang, J., et al. (2019). MDR: an integrative DNA N6-methyladenine and N4-methylcytosine modification database for Rosaceae. *Hortic. Res.* 6:78.

Long, H., Sun, Z., Li, M., Fu, H. Y., and Lin, M. C. (2020). Predicting protein phosphorylation sites based on deep learning. *Curr. Bioinform.* 15, 300–308. doi: 10.2174/1574893614666190902154332

Luo, X., Wang, F., Wang, G., and Zhao, Y. (2020). Identification of methylation states of DNA regions for Illumina methylation BeadChip. *BMC Genomics* 21(Suppl 1):672. doi: 10.1186/s12864-019-6019-0

Lv, H., Dao, F.-Y., Guan, Z.-X., Yang, H., Li, Y.-W., and Lin, H. (2020a). Deep-Kcr: accurate detection of lysine crotonylation sites using deep learning method. *Brief. Bioinform.* bbaa255.

Lv, H., Dao, F. Y., Zhang, D., Guan, Z. X., Yang, H., Su, W., et al. (2020b). iDNA-MS: an integrated computational tool for detecting DNA modification sites in multiple genomes. *iScience* 23:100991. doi: 10.1016/j.isci.2020.100991

Manavalan, B., Basith, S., Shin, T. H., Wei, L., and Lee, G. (2019b). Meta-4mCpred: a sequence-based meta-predictor for accurate DNA 4mC site prediction using effective feature representation. *Mol. Ther. Nucleic Acids* 16, 733–744. doi: 10.1016/j.omtn.2019.04.019

Manavalan, B., Basith, S., Shin, T. H., Wei, L., and Lee, G. (2019a). mAHTPred: a sequence-based meta-predictor for improving the prediction of anti-hypertensive peptides using effective feature representation. *Bioinformatics* 35, 2757–2765. doi: 10.1093/bioinformatics/bty1047

Min, X., Ye, C., and Liu, X., Zeng, X. (2021). Predicting enhancer-promoter interactions by deep learning and matching heuristic. *Brief. Bioinform.* bbaa254. doi: 10.1093/bib/bbaa254

Naseer, S., Hussain, W., Khan, Y. D., and Rasool, N. (2020). Sequence-based identification of arginine amidation sites in proteins using deep representations of proteins and PseAAC. *Curr. Bioinform.* 15, 937–948. doi: 10.2174/1574893615666200129110450

Qiang, X., Zhou, C., Ye, X., Du, P-f, Su, R., and Wei, L. (2020). CPPred-FL: a sequence-based predictor for large-scale identification of cell-penetrating peptides by feature representation learning. *Brief. Bioinform.* 21, 11–23.

Quang, D., and Xie, X. (2016). DanQ: a hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences. *Nucleic Acids Res.* 44:e107. doi: 10.1093/nar/gkw226

Shen, Z., and Zou, Q. (2020). Basic polar and hydrophobic properties are the main characteristics that affect the binding of transcription factors to methylation sites. *Bioinformatics* 36, 4263–4268. doi: 10.1093/bioinformatics/btaa492

Song, B., Zeng, X., Jiang, M., and Pérez-Jiménez, M. J. (2020). Monodirectional tissue P systems with promoters. *IEEE Trans. Cybern.* 51, 438–450. doi: 10.1109/TCYB.2020.3003060 doi: 10.1109/tcyb.2020.3003060

Su, R., Hu, J., Zou, Q., Manavalan, B., and Wei, L. (2020a). Empirical comparison and analysis of web-based cell-penetrating peptide prediction tools. *Brief. Bioinform.* 21, 408–420. doi: 10.1093/bib/bby124

Su, R., Liu, X., and Wei, L. (2020b). MinE-RFE: determine the optimal subset from RFE by minimizing the subset-accuracy-defined energy. *Brief. Bioinform.* 21, 687–698. doi: 10.1093/bib/bbz021

Su, R., Liu, X., Wei, L., and Zou, Q. (2019a). Deep-Resp-Forest: a deep forest model to predict anti-cancer drug response. *Methods* 166, 91–102. doi: 10.1016/j.ymeth.2019.02.009

Su, R., Liu, X., Xiao, G., and Wei, L. (2020c). Meta-GDBP: a high-level stacked regression model to improve anticancer drug response prediction. *Brief. Bioinform.* 21, 996–1005. doi: 10.1093/bib/bbz022

Su, R., Wu, H., Xu, B., Liu, X., and Wei, L. (2019b). Developing a multi-dose computational model for drug-induced hepatotoxicity prediction based on toxicogenomics data. *IEEE ACM Trans. Comput. Biol. Bioinform.* 16, 1231–1239. doi: 10.1109/tcbb.2018.2858756

Su, W., Liu, M. L., Yang, Y. H., Wang, J. S., Li, S. H., Lv, H., et al. (2021). PPD: a manually curated database for experimentally verified prokaryotic promoters. *J. Mol. Biol.* 166860. doi: 10.1016/j.jmb.2021.166860

Tang, H., Zhao, Y. W., Zou, P., Zhang, C. M., Chen, R., Huang, P., et al. (2018). HBPred: a tool to identify growth hormone-binding proteins. *Int. J. Biol. Sci.* 14, 957–964. doi: 10.7150/ijbs.24174

Tang, Q., Kang, J., Yuan, J., Tang, H., Li, X., Lin, H., et al. (2020). DNA4mC-LIP: a linear integration method to identify N4-methylcytosine site in multiple species. *Bioinformatics* 36, 3327–3335. doi: 10.1093/bioinformatics/btaa143

Tang, W., Wan, S., Yang, Z., Teschendorff, A. E., and Zou, Q. (2018). Tumor origin detection with tissue-specific miRNA and DNA methylation markers. *Bioinformatics* 34, 398–406. doi: 10.1093/bioinformatics/btx622

Wang, G., Luo, X., Wang, J., Wan, J., Xia, S., Zhu, H., et al. (2018). MeDReaders: a database for transcription factors that bind to methylated DNA. *Nucleic Acids Res.* 46(D1), D146–D151.

Wang, H., Ding, Y., Tang, J., and Guo, F. (2019). Identification of membrane protein types via multivariate information fusion with Hilbert–Schmidt Independence Criterion. *Neurocomputing* 383, 257–269. doi: 10.1016/j.neucom.2019.11.103

Wang, H., Ding, Y., Tang, J., Zou, Q., and Guo, F. (2021a). Identify RNA-associated subcellular localizations based on multi-label learning using Chou's 5-steps rule. *BMC Genomics* 22:56. doi: 10.1186/s12864-020-07347-7

Wang, H., Tang, J., Ding, Y., and Guo, F. (2021b). Exploring associations of non-coding RNAs in human diseases via three-matrix factorization with hypergraph-regular terms on center kernel alignment. *Brief. Bioinform.* bbaa409.

Wang, J., Chen, S., Dong, L., and Wang, G. (2020). CHTKC: a robust and efficient k-mer counting algorithm based on a lock-free chaining hash table. *Brief. Bioinform.* bbaa063.

Wei, L., Chen, H., and Su, R. (2018a). M6APred-EL: a sequence-based predictor for identifying N6-methyladenosine sites using ensemble learning. *Mol. Ther. Nucleic Acids* 12, 635–644. doi: 10.1016/j.omtn.2018.07.004

Wei, L., Ding, Y., Su, R., Tang, J., and Zou, Q. (2018b). Prediction of human protein subcellular localization using deep learning. *J. Parallel Distrib. Comput.* 117, 212–217.

Wei, L., He, W., Malik, A., Su, R., Cui, L., and Manavalan, B. (2020a). Computational prediction and interpretation of cell-specific replication origin sites from multiple eukaryotes by exploiting stacking framework. *Brief. Bioinform.* doi: 10.1093/bib/bbaa275

Wei, L., Hu, J., Li, F., Song, J., Su, R., and Zou, Q. (2020b). Comparative analysis and prediction of quorum-sensing peptides using feature representation learning and machine learning algorithms. *Brief. Bioinform.* 21, 106–119.

Wei, L., Liao, M., Gao, Y., Ji, R., He, Z., and Zou, Q. (2014). Improved and promising identification of human microRNAs by incorporating a high-quality negative set. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 11, 192–201. doi: 10.1109/tcbb.2013.146

Wei, L., Luan, S., Nagai, L. A. E., Su, R., and Zou, Q. (2019a). Exploring sequence-based features for the improved prediction of DNA N4-methylcytosine sites in multiple species. *Bioinformatics* 35, 1326–1333. doi: 10.1093/bioinformatics/bty824

Wei, L., Su, R., Luan, S., Liao, Z., Manavalan, B., Zou, Q., et al. (2019b). Iterative feature representations improve N4-methylcytosine site prediction. *Bioinformatics* 35, 4930–4937. doi: 10.1093/bioinformatics/btz408

Wei, L., Su, R., Wang, B., Li, X., Zou, Q., and Gao, X. (2019c). Integration of deep feature representations and handcrafted features to improve the prediction of N-6-methyladenosine sites. *Neurocomputing* 324, 3–9. doi: 10.1016/j.neucom.2018.04.082

Wei, L., Tang, J., and Zou, Q. (2017a). Local-DPP: an improved DNA-binding protein prediction method by exploring local evolutionary information. *Inform. Sci.* 384, 135–144. doi: 10.1016/j.ins.2016.06.026

Wei, L., Wan, S., Guo, J., and Wong, K. K. L. (2017b). A novel hierarchical selective ensemble classifier with bioinformatics application. *Artif. Intell. Med.* 83, 82–90. doi: 10.1016/j.artmed.2017.02.005

Wei, L., Xing, P., Shi, G., Ji, Z., and Zou, Q. (2019d). Fast prediction of protein methylation sites using a sequence-based feature selection technique. *IEEE ACM Trans. Comput. Biol. Bioinform.* 16, 1264–1273. doi: 10.1109/tcbb.2017.2670558

Wei, L., Xing, P., Zeng, J., Chen, J., Su, R., and Guo, F. (2017c). Improved prediction of protein-protein interactions using novel negative samples, features, and an ensemble classifier. *Artif. Intell. Med.* 83, 67–74. doi: 10.1016/j.artmed.2017.03.001

Wei, L., Zhou, C., Chen, H., Song, J., and Su, R. (2018c). ACPred-FL: a sequence-based predictor using effective feature representation to improve the prediction of anti-cancer peptides. *Bioinformatics* 34, 4007–4016.

Xu, H., Jia, P., and Zhao, Z. (2020). Deep4mC: systematic assessment and computational prediction for DNA N4-methylcytosine sites by deep learning. *Brief. Bioinform.* bbaa099.

Yang, C., Ding, Y., Meng, Q., Tang, J., and Guo, F. (2021). Granular multiple kernel learning for identifying RNA-binding protein residues via integrating sequence and structure information. *Neural Comput. Appl.*

Yang, H., Yang, W., Dao, F. Y., Lv, H., Ding, H., Chen, W., et al. (2020). A comparison and assessment of computational method for identifying recombination hotspots in Saccharomyces cerevisiae. *Brief. Bioinform.* 21, 1568–1580. doi: 10.1093/bib/bbz123

Zeng, R., and Liao, M. (2020). Developing a multi-layer deep learning based predictive model to identify DNA N4-methylcytosine modifications. *Front. Bioeng. Biotechnol.* 8:274. doi: 10.3389/fbioe.2020.00274

Zeng, X., Liu, L., Lu, L., and Zou, Q. (2018). Prediction of potential disease-associated microRNAs using structural perturbation method. *Bioinformatics* 34, 2425–2432. doi: 10.1093/bioinformatics/bty112

Zeng, X., Wang, W., Chen, C., and Yen, G. (2020). A consensus community-based particle swarm optimization for dynamic community detection. *IEEE Trans. Cybern.* 50, 2502–2513. doi: 10.1109/tcyb.2019.2938895

Zeng, X., Zhang, X., and Zou, Q. (2016). Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks. *Brief. Bioinform.* 17, 193–203. doi: 10.1093/bib/bbv033

Zhang, D., Chen, H.-D., Zulfiqar, H., Yuan, S.-S., Huang, Q.-L., Zhang, Z.-Y., et al. (2021). iBLP: an XGBoost-based predictor for identifying bioluminescent proteins. *Comput. Math. Methods Med.* 2021:6664362.

Zhang, D., Xu, Z. C., Su, W., Yang, Y. H., Lv, H., Yang, H., et al. (2020). iCarPS: a computational tool for identifying protein carbonylation sites by novel encoded features. *Bioinformatics* btaa702.

Zhang, T., Wei, X., Li, Z., Shi, F., Xia, Z., Lian, M., et al. (2020). Natural scene nutrition information acquisition and analysis based on deep learning. *Curr. Bioinform.* 15, 662–670. doi: 10.2174/1574893614666190723121610

Zhang, Y., Kou, C., Wang, S., and Zhang, Y. (2019). Genome-wide differential-based analysis of the relationship between DNA methylation and gene expression in cancer. *Curr. Bioinform.* 14, 783–792. doi: 10.2174/1574893614666190424160046

Zhang, Y., Yan, J., Chen, S., Gong, M., Gao, D., Zhu, M., et al. (2020). Review of the applications of deep learning in bioinformatics. *Curr. Bioinform.* 15, 898–911. doi: 10.2174/1574893615999200711165743

Zhang, Z. Y., Yang, Y. H., Ding, H., Wang, D., Chen, W., and Lin, H. (2021). Design powerful predictor for mRNA subcellular location prediction in *Homo sapiens*. *Brief. Bioinform.* 22, 526–535. doi: 10.1093/bib/bbz177

Zhao, X., Jiao, Q., Li, H., Wu, Y., Wang, H., Huang, S., et al. (2020). ECFS-DEA: an ensemble classifier-based feature selection for differential expression analysis on expression profiles. *BMC Bioinformatics* 21:43. doi: 10.1186/s12859-020-3388-y

Zhu, T., Guan, J., Liu, H., and Zhou, S. (2019). RMDB: an integrated database of single-cytosine-resolution DNA methylation in *Oryza sativa*. *Curr. Bioinform.* 14, 524–531. doi: 10.2174/1574893614666190211161717

Zou, Q., Xing, P., Wei, L., and Liu, B. (2019). Gene2vec: gene subsequence embedding for prediction of mammalian N6-methyladenosine sites from mRNA. *RNA* 25, 205–218. doi: 10.1261/rna.069112.118

Zou, Y., Wu, H., Guo, X., Peng, L., Ding, Y., Tang, J., et al. (2020). MK-FSVM-SVDD: a multiple kernel-based Fuzzy SVM model for predicting DNA-binding proteins via support vector data description. *Curr. Bioinform.* 15, 1–1.

Zuo, Y., Song, M., Li, H., Chen, X., Cao, P., Zheng, L., et al. (2020). Analysis of the epigenetic signature of cell reprogramming by computational DNA methylation profiles. *Curr. Bioinform.* 15, 589–599. doi: 10.2174/1574893614666190919103752

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.