Methods 204 (2022) 142–150

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Deep6mAPred: A CNN and Bi-LSTM-based deep learning method for predicting DNA N6-methyladenosine sites across plant species

Xingyu Tang [a], Peijie Zheng [a], Xueyong Li [a], Hongyan Wu [b], Dong-Qing Wei [b,c,*], Yuewu Liu [d], Guohua Huang [a,*]

[a] School of Electrical Engineering, Shaoyang University, Shaoyang, Hunan 422000, China
[b] The Joint Engineering Research Center for Health Big Data Intelligent Analysis Technology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[c] State Key Laboratory of Microbial Metabolism, and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China
[d] College of Information and Intelligence, Hunan Agricultural University, Changsha, Hunan 410081, China

ARTICLE INFO

Keywords:
6mA
DNA modification
Convolution neural network
Long-short term memory
Feed-forward attention
Deep learning

ABSTRACT

DNA N6-methyladenine (6mA) is a key DNA modification, which plays versatile roles in the cellular processes, including regulation of gene expression, DNA repair, and DNA replication. DNA 6mA is closely associated with many diseases in the mammals and with growth as well as development of plants. Precisely detecting DNA 6mA sites is of great importance to exploration of 6mA functions. Although many computational methods have been presented for DNA 6mA prediction, there is still a wide gap in the practical application. We presented a convolution neural network (CNN) and bi-directional long-short term memory (Bi-LSTM)-based deep learning method (Deep6mAPred) for predicting DNA 6mA sites across plant species. The Deep6mAPred stacked the CNNs and the Bi-LSTMs in a paralleling manner instead of a series-connection manner. The Deep6mAPred also employed the attention mechanism for improving the representations of sequences. The Deep6mAPred reached an accuracy of 0.9556 over the independent rice dataset, far outperforming the state-of-the-art methods. The tests across plant species showed that the Deep6mAPred is of a remarkable advantage over the state of the art methods. We developed a user-friendly web application for DNA 6mA prediction, which is freely available at http://106.13.196.152:7001/ for all the scientific researchers. The Deep6mAPred would enrich tools to predict DNA 6mA sites and speed up the exploration of DNA modification.

1. Introduction

DNA methylation is defined as a biological process where the methyl groups are transferred to residues of the DNA molecules [1]. So far, DNA methylation has been found to occur only in two types of nucleotide (adenine and cytosine). The methylation of cytosine includes 5-methylcytosine (5mC) and N4-methylcytosine (4mC), while the methylation of adenine includes N6-methyladenine (6mA) [2]. DNA methylation is widely distributed both in prokaryote and in eukaryote [3–6], but the proportion of methylated residues differs greatly with species. For instance, the 5mc accounted for no less than 8% of the total cytosine in Physarum polycephalum [7,8]. Up to 75% of CpG dinucleotides were methylated in somatic cells of the mammals [9], 14% of cytosines were methylated in Arabidopsis thaliana, 2.3% in Escherichia coli, 0.03% in Drosophila, and 0.006% in Dictyostelium [10]. DNA methylation alters activities of DNA segments without changing the sequence, which thus yields a wide variety of roles in the cellular processes across organisms or tissues [1]. DNA methylation in the promoter regions typically serves as a repressor of the gene transcription [11,12], while the methylation of the transposable elements correlates positively with genome expansion [13]. The DNA methylation appears essential for normal development. For example, the DNA methylation is associated with many cellular processes such as genomic imprinting, X-chromosome inactivation [14,15], also is implied in many diseases including cancer [16], and acts as an accurate epigenetic clock to estimate the age of tissues as well as cell types in human or mammals [17]. Due to its importance to cellular mechanisms, the DNA methylation is increasingly attracting attentions from biologists.

* Corresponding authors at: The Joint Engineering Research Center for Health Big Data Intelligent Analysis Technology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China.
E-mail addresses: dqwei@sjtu.edu.cn (D.-Q. Wei), guohuahhn@163.com (G. Huang).

https://doi.org/10.1016/j.ymeth.2022.04.011
Received 18 March 2022; Received in revised form 16 April 2022; Accepted 20 April 2022
Available online 25 April 2022

Table 1
Recently published methods for DNA 6 mA prediction.

order	Name	Deep Learning	Webserver/source code
1	MM-6mAPred [45]	no	https://www.insect-genome.com/MM-6mAPred/
2	p6mA [75]	no	https://github.com/Kongla b404/p6mA
3	6mAPred-FO [76]	no	https://server.malab.cn/6mAPred-FO
4	i6mA-DNCP [51]	no	https://ww2.mathworks.cn/matla bcentral/fileexchange/72549-i6mA-dncp
5	iDNA6mA-PseKNC [52]	no	https://lin-group.cn/server/iDNA6 mA-PseKNC
6	i6mA-Pred [47]	no	https://lin-group.cn/server/ i6mA-Pred
7	iDNA6mA-Rice [1]	no	https://lin-group.cn/server/iDN A6mA-Rice
8	6mA-RicePred [53]	no	https://github.com/huangqianfe i0916/6ma-rice/tree/master/dataset
9	i6mA-stack [77]	no	https://nsclbio.jbnu.ac.kr/tools/i6 mA-stack/
10	i6mA-Fuse [57]	no	https://kurata14.bio.kyutech.ac.jp/i 6mA-Fuse/
11	6mA-Finder [55]	no	https://bioinfo.uth.edu/6mA_Finder
12	csDMA [54]	no	https://github.com/liuze-nwafu/ csDMA
13	i6mA-VC [63]	no	https://www.zhanglab.site/
14	Meta-i6mA [62]	no	https://kurata14.bio.kyutech.ac. jp/Meta-i6mA/
15	iDNA6mA-Rice-DL [78]	yes	https://hub.docker.com/r/his1server /idna6ma-rice-dl
16	SNNRice6mA [79]	yes	https://github.com/yuht4/SNNRic e6mA
17	iIM-CNN [64]	yes	https://home.jbnu.ac.kr/NSCL/i IMCNN.htm
18	SICD6mA [80]	yes	https://github.com/lwzyb/SICD6mA
19	SDM6A [58]	no	https://thegleelab.org/SDM6A
20	6mA-Pred [72]	yes	https://39.100.246.211:5004/6 mA_Pred/
21	i6mA-CNN [81]	yes	https://cutt.ly/Co6KuWG
22	DNA6mA-MINT [74]	yes	https://home.jbnu.ac.kr/NSCL/DNA 6mA-MINT.htm
23	SpineNet-6mA [82]	yes	https://nsclbio.jbnu.ac.kr/tools/Spi neNet6mA/.
24	Deep6mA [83]	yes	https://www.pianlab.cn/deep6ma/
25	BERT-DNA [73]	yes	https://github.com/khanhlee/be rt-dna
26	iDNA-MT [84]	yes	none
27	iDNA-ABT [85]	yes	https://server.wei-group.net/i DNA_ABT

The 6mA refers to a biological process where the methyl group is attached to the 6-th nitrogen atom of adenine by the enzyme of DNA methyltransferase. The 6mA is a type of non-canonical DNA modification because it might occur in other nucleotide molecules including mRNA, tRNA, rRNA, small nuclear RNA (snRNA) as well as long non-coding RNA, such as Xist [18–20]. The 6mA was found in a wide range of tissues or species such as viruses [21], plants [22], yeast [23] as well as insects [24]. The DNA 6mA was reported to be more prevalent in the prokaryote than in the eukaryotes [3,5,6,25–30]. A large volume of evidence suggested that the DNA 6mA would play vital roles in many key biological processes. The DNA 6mA participated in regulation of gene expression both in prokaryotes and in some eukaryotes [31–34], and was responsible for DNA repair [35,36] as well as DNA replication [37]. The 6mA distinguished invading foreign DNAs from host DNA in prokaryotes [38], was closely associated with many disease including tumor in human genome [39], and was involved in regulation of drug resistance in triple negative breast cancer [40]. Even Luo et al. [41] proposed the DNA 6mA as a new epigenetic mark in eukaryotes.

Identification of DNA 6mA sites is fundamental for investigating its functions in the biological processes. There are two major ways to detect DNA 6mA sites: the wet and the dry experiments. The former includes mass spectrometry [42], methylation-specific polymerase chain reaction [43], and single-molecule real-time sequencing [44]. These wet experiments are generally time-consuming, expensive, and labor-intensive. The dry experiment refers to computational methods, which automatically annotates 6mA sites by learning a classifier from the known 6mA data (generally labelled as positive) and non-6mA data (labelled as negative) [45]. For example, Pian et al. employed transition probability between adjacent nucleotides to construct a Markov model to identify 6mA site [45]. To date, there are at least 27 methods developed for 6mA prediction [46], as listed in Table 1. These computational methods are grouped into the feature-based and the deep learning-based methods [46,47]. The features (called also representations) and the chosen learning algorithms would jointly determine the performance of the feature-based methods to a great extent. Features are generally extracted from sequences and structures. The learning algorithms includes support machine vector (SVM) [48], random forest [49], and extremely randomized tree [50]. The i6mA-Pred proposed by Chen et al. is a SVM-based method for rice 6mA prediction, which employed only the nucleotide frequency and the nucleotide chemical property as its representation [47]. The i6mA-DNCP [51] improved the representation with optimized dinucleotide-based features and thus promoted the performance of the i6mA-Pred. Similarly, the iDNA6mA-PseKNC incorporated nucleotide physicochemical properties into pseudo k-tuple nucleotide composition (PseKNC) for predicting 6mA sites in Mus musculus genome [52]. Different representations would contribute differently to detection of 6mA sites. Some representations contributed much, while some little. It is ideal to combine all the representations to generate the best performance. However, the combination of all the representation is not able to ensure the promotion of predictive performance. What's more, it increases computation burden. The combination of parts of representations is a more practical way. The same is true for the learning algorithms. Therefore, the feature-based method is to search for the best combination of different representations and various algorithms. There are many published works in this respect. For example, the iDNA6mA-Rice utilized the multiple encoding schemes (PseKNC, mono-nucleotide binary encoding, and natural vector) as representations, and used random forest as the learning algorithm [1]. The 6mA-RicePred fused multiple features including Markov feature for rice 6mA prediction [53]. The csDMA utilized three representations and explored performances of different algorithms on 6mA prediction, and finally chose the algorithm which performed best over experiments to construct the predictor [54]. The 6mA-Finder combined seven sequence-based coding schemes such as accumulated nucleotide frequency, the composition of K-spaced nucleic acid pairs, three types of physicochemical features, and seven conventional learning algorithms [55]. Recursive feature elimination strategy [56] was employed to select discriminable features. The 6mA-Finder found that the combination of optimal features improved performance but the combination of models didn't outperform the single best model [55]. The i6mA-Fuse was the first computational method to predict 6mA sites in Rosaceae genomes, which used linear regression model to ensemble the predicted probabilities of five single-encoding-based random forest models [57]. The SDM6A explored nine features as well as four learning algorithms, and chose five types of optimal features as well as two learning algorithms (SVM and extremely randomized tree). The classifiers trained by respectively optimal features and algorithms were combined for 6mA prediction [58]. The Meta-i6mA explored combinations between ten representations and six commonly used machine learning methods (random forest [49], SVM [48], extremely randomized tree [50], logistic regression [59], naïve Bayes [60] and AdaBoost [61]), and finally integrated 30 optimal models generated by five key informative features for DNA 6mA prediction [62]. Similar to the Meta-i6mA, the i6mA-VC also is an ensemble learning method which ensembles multi-classifier for voting [63].

The feature-based method had difficulty in achieving the best performance in practice because the combination was infinite. On the other

Table 2
The numbers of positive and negative samples in the four experimental datasets.

Dataset	Positives	Negatives	Total samples
6mA-rice-Lv	154,000	154,000	300,800
6mA-rice-Chen	880	880	1760
i6mA-Fuse-R	5733	5733	11,466
i6mA-Fuse-F	1417	1417	2834

hand, the deep learning method requires no feature design, and is an end-to-end method. The iIM-CNN [64] employed the convolution neural network (CNN), a widely used structure of the neural network. The iDNA6mA is a deep learning-based method, which simply utilized the CNN [65]. The CNN was initially created for image recognitions [66,67], and later extensively applied to other fields including natural language processing [68,69]. The major advantage of the CNN lies in its ability to characterize the local trait of an image. The CNN didn't learn relationships hidden in characters. The features learned by CNN alone are not sufficient to represent 6mA sequences. Therefore, other structures of neural networks are applied to 6mA prediction, such as long short-term memory (LSTM). The LSTM is capable of capturing semantics hidden within sequences, and thus is utilized in a wide range of fields including bioinformatics [70], financial market forecasting [70], and energy consumption prediction [71]. Huang et al. [72] proposed a LSTM-based method called 6mA-Pred. The empirical evidence validated the effectiveness of the LSTM in detecting 6mA sites. Le et al. employed transformer-based BERT pre-trained model to extract deep representations of sequences [73]. The DNA6mA-MINT is a deep learning-based method for detecting DNA 6mA sites, which stacked CNN and LSTM simply [74]. Similarly, the Deep6mA stacked CNN and LSTM, and achieved quite a success in 6mA prediction. In fact, the way of stacking is not a good alternative because it is not beneficial to accumulate representations learned by CNN and LSTM respectively. We improved ways of combining CNN and LSTM, and presented an improved computational method called Deep6mAPred for predicting 6mA site across different plant species.

2. Data and methods

2.1. Datasets

Four DNA 6mA datasets from three diverse plants were employed to test the presented Deep6mAPred for the predictive performance. The first dataset was 6mA-rice-lv [1,38]. The 6mA-rice-lv was originated from rice genome by zhou et al. [38], further collected and processed by Lv et al. [1], and was used by Lv et al. [1] and Li et al. [83] for examining performance of methods. The 6mA-rice-lv contained 154,000 positive and 154,000 negative samples, with similarities between sequences being equal to or less than 0.8. The second dataset was the 6mA-rice-chen [47], which is also from the rice genome. The 6mA-rice-chen consisted of 880 positive and 880 negative samples, and was widely used as the training data or testing data [45,47,58,65,79,83]. To test the ability to detect 6mA sites across plant species, we used the other two datasets: i6mA-Fuse-F and i6mA-Fuse-R, both of which are subsets of the i6mA-Fuse dataset [57]. The i6mA-Fuse dataset was created from the MDR database [86]. The i6mA-Fuse-R was from Rosa chinensis (R. chinensis), containing 5733 positive and 5733 negative samples. The i6mA-Fuse-F was from Fragaria vesca (F. vesca), containing 1417 positive and 1417 negative samples. All the positive and negative samples are 41 nt with adenine at the center. The samples with 6mA site are positive and otherwise are negative. The numbers of positive and negative samples in four datasets were listed in Table 2.

2.2. Methods

As shown in Fig. 1, the schematic of the presented Deep6mAPred consisted of three modules: input, feature extraction and classification. The input was to map a DNA sequence of 41 nt into a binary vector with the one-hot encoding scheme. The feature extraction contained two

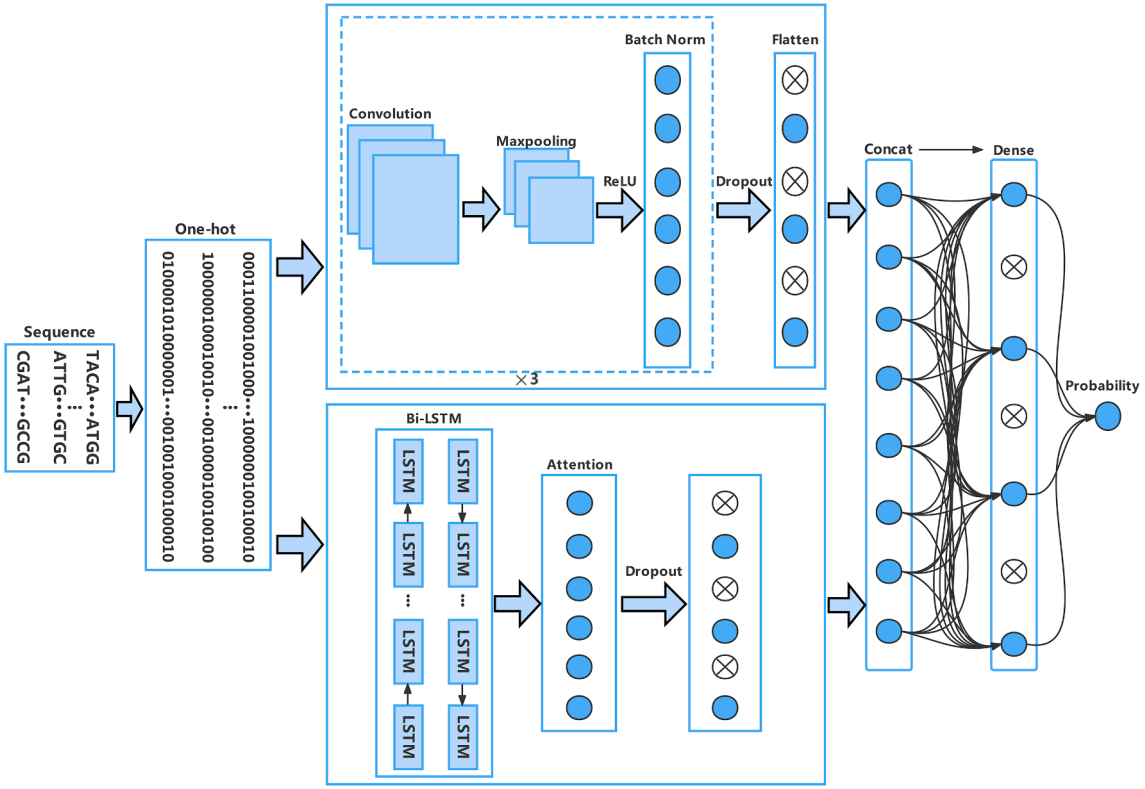

Fig. 1. The flowchart of the Deep6mAPred.

paralleling parts. One consisted mainly of one-dimensional convolution (1D CNN) layer and the batch normalization. The maximum pooling followed each 1D CNN to reduce the dimension and to refine abstract features. The other part was comprised of two bidirectional LSTM (Bi-LSTM) layers and the attention mechanism. The Bi-LSTM was intended to extract contextual semantics of the sequences, while the attention mechanism was to catch the key information. Both parts used the dropout to reduce or avoid the over-fitting issue. The classification module was actually a multilayer perceptron, which concatenated outputs of the feature extraction module as inputs and used a fully-connected layer as a hidden layer.

2.2.1. One-hot encoding

One-hot encoding is an intuitive but effective scheme to encode DNA, RNA and protein sequences, and thus has been extensively applied in the field of bioinformatics. Here, A, C, G, and T were encoded into, and respectively. Because the sequences contained non-sequenced nucleotide, namely the letter N, an additional vector was utilized to represent it. A sequence of 41 nt was mapped into a 41 by 4 matrix, which worked as an input to the feature extraction module.

2.2.2. CNN

CNN is a type of feed-forward neural network [87]. Unlike the common neural network such as multilayer perceptron, the CNN shared weights (called the convolution kernels or filters) by sliding along input features. Due to its fault tolerance, the parallel processing, and the self-learning ability, the CNN has attracted comprehensive attentions over recent decades. Some CNN-based deep learning methods such as LeNet [67], AlexNet [88], VGG [89], GoogLeNet [90], and ResNet [91] have been proposed to achieve great advance in the recognition of image [92,93].

The CNN generally includes convolution and pooling operations. The convolution operation is to compute the dot product between the convolution kernel and receptive field of the input. Sliding the convolution kernel along the input is to obtain the different feature maps. The convolution kernel is the learnable parameters shared by the input. Assume that the convolution kernel was and the input was. The convolution is computed by.

$$f(a,b) = \sum_i \sum_j C(i,j) I(a+i, b+j) \tag{1}$$

This sliding size is called the stride, which controls the allocation of depth columns around the width and height. If the stride was set to 1, the convolution kernel moved one pixel a time. The less the stride, the more over-lapping the receptive fields and the larger the size of output. The size of convolutional kernel is far less than size of the input. To obtain features from various dimensions, the multiple convolutional kernels (also called multiple channels) can be used. An activation function is generally added to the end of the convolutional kernel to enhance non-linear characteristic of the CNN. The commonly used activation functions includes sigmoid function, the ReLU (rectified linear units), and tanh, which are defined respectively by.

$$\sigma(x) = (1 + e^{-x})^{-1} \tag{2}$$

$$ReLU(x) = \begin{cases} 0, & if\ x < 0 \\ x, & if\ x \geq 0 \end{cases} \tag{3}$$

$$\tau(x) = tanh(x) \tag{4}$$

We adopted the ReLU function because it is faster than other activation functions when training the CNN [88].

Another important operation of the CNN, pooling, is in fact a type of non-linear down-sampling. The pooling includes max pooling and mean pooling. The max pooling is to return the maximum of the pooling region, which is defined by.

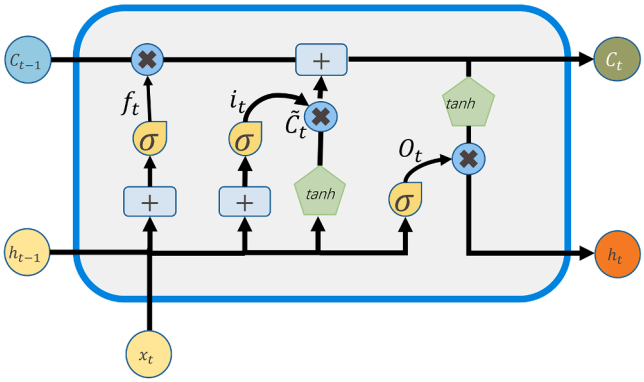

Fig. 2. The architecture diagram of LSTM.

$$p(a,b) = \max_{i,j}\{f(a+i, b+j)\} \tag{5}$$

The pooling reduces the spatial size of the representation, the number of the trained parameters, memory, as well as computation, and also reduces or avoids over-fitting issue.

2.2.3. LSTM

The LSTM [86,87] is an advanced recurrent neural network (RNN) [87,94–96], which is a model to deal with time series data. The advantage of the RNN is that it is able to capture semantics of words in the context of sequences, but the vanishing or exploding gradient issue causes the RNN to fail to remember long term dependencies in the long sequences. The LSTM is intended to avoid or solve this issue. As shown in Fig. 2, the processing unit of the LSTM at the time step t generally contains three inputs, two outputs, three operation gates, and one candidate cell. Three inputs are the input x_t at the time step t, the cell state C_{t-1}, and the hidden state h_{t-1} of the time step t-1, respectively. The cell state is like a conveyor to allow information to be passed from front to back. The gates are respectively the input, the forget gate, and the output gate, optionally allowing information of the cell state to be passed. These gates use the sigmoid function, whose output ranges from 0 to 1. The forget gate decides what information of the cell state is preserved. If the forget gate outputted 1, which implied that all the information of the cell state was reserved and 0 implied that all the information was forgotten. The forget gate is computed by

$$f_t = \sigma\left(w_f \bullet [h_{t-1}, x_t] + b_f\right) \tag{6}$$

where denotes weight between the input and the hidden state, and is the bias. The input gate decides what information is updated, which is computed by

$$i_t = \sigma(w_i \bullet [h_{t-1}, x_t] + b_i) \tag{7}$$

where denotes the weights between the input and the hidden state, and is the bias. The candidate cell is a tanh function, which is computed by

$$\tilde{C}_t = tanh(w_c \bullet [h_{t-1}, x_t] + b_c) \tag{8}$$

where w_c denotes the weights between the input and the hidden state, and b_c is the bias. The input gate and the candidate cell are combined to update the cell state. The update is computed by.

$$C_t = f_t \otimes \tilde{C}_t + f_t \otimes C_{t-1} \tag{9}$$

where denotes element-wise multiplication. The output gate is computed by

$$O_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \tag{10}$$

where w_o denotes the weights between the input and the hidden state,

and b_o is the bias. The hidden state is updated with the cell state and the output gate by

$$h_t = O_t \bullet tanh(C_t) \tag{11}$$

In the context of sentences, the latter words might be related to the previous ones and vice versa. Therefore, we used the Bi-LSTM to capture the semantic relationship among the words. The forward and the backward hidden states are concatenated as the output at the time step t, which is computed by

$$h_t = \left[\overrightarrow{h_t} \otimes \overleftarrow{h_t} \right] \tag{12}$$

2.2.4. Feed-forward attentions

Attention [97] is a mechanism to allocate different weights to different positions of a sequence, as one views an image with different attentions on different parts. The attention mechanism is increasingly becoming a popular trick of deep learning since it was proposed. Many attention mechanisms have been proposed over the past five years, such as well-known self-attention [97], feed-forward attention [98], external attention [99], and double attentions [100]. Here we employed the feed-forward attention [98]. The feed-forward attention was initially intended to solve the long-term dependency issue in the LSTM. At the heart of the feed-forward attention is to compute weight of hidden state at each time step by the following formula:

$$e_t = a(h_t) \tag{13}$$

where a is a learnable parameter, and h_t is the hidden state at time step t. The softmax function was used to normalize weights over time steps, namely,

$$\theta_t = \frac{\exp(e_t)}{\sum_{i=1}^{T}\exp(e_i)} \tag{14}$$

The final output is a sum over the normalized weights multiplying the hidden state, which is computed by

$$\mathrm{c} = \sum_{t=1}^{T} \theta_t h_t \tag{15}$$

The feed-forward attention integrates information over various time step. The average over all the hidden state is viewed as the simplest feed-forward attention, which is computed by

$$\mathrm{c} = \frac{1}{T}\sum_{t=1}^{T} h_t \tag{16}$$

2.2.5. Loss function

We used the binary cross entropy as the loss function, which is computed by

$$L(w) = -\sum_{i=1}^{N} y_i log\left(y_i^{'}\right) + (1-y_i) log\left(1-y_i^{'}\right) \tag{17}$$

where y_i denotes the label of the sample i, and $y_i^{'}$ is the predicted value of the sample being positive. The label of positive sample is 1 and the label of negative sample is 0.

3. Metrics

We used the sensitivity (S_n), specificity (S_P), accuracy (ACC), and Matthews correlation coefficient (MCC) as assessment metrics, which are respectively defined as

$$S_n = \frac{T_P}{T_P + F_N} \tag{18}$$

Table 3
Performances by 5-fold cross validation over the 6mA-rice-Lv.

Method	S_n	S_P	ACC	MCC	AUC
Deep6mAPred	**0.9538**	0.9255	0.9397	0.8798	0.9793
Deep6mA*	0.9506	**0.9296**	**0.9401**	**0.8800**	**0.9800**
SNNRice6mA-large*	0.9347	0.8975	0.9204	0.8400	0.9700
MM-6mApred *	0.9347	0.8951	0.9149	0.8300	0.9600

The asterisk (*) indicated that the results were from the literature [83].

Table 4
Performances over the 6mA-rice-chen dataset.

Method	S_n	S_P	ACC	MCC	AUC
Deep6mAPred	**0.9545**	0.9568	**0.9556**	**0.9136**	**0.9837**
Deep6mA*	0.7973	**0.9640**	0.8806	0.7700	0.9600
SNNRice6mAlarge*	0.7790	0.8742	0.8267	0.6500	0.8900
MM-6mApred*	0.7682	0.9170	0.8426	0.6800	0.9100

The asterisk (*) indicated that the results were from the literature [83].

$$S_P = \frac{T_N}{T_N + F_P} \tag{19}$$

$$ACC = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{20}$$

$$MCC = \frac{T_P \times T_N - F_P \times F_N}{\sqrt{(T_P + F_N)(T_P + F_P)(T_N + F_P)(T_N + F_N)}} \tag{21}$$

where T_P and T_N denote the numbers of the true positive and the true negative samples respectively, F_P and F_N denote the numbers of the false positive and the false negative samples respectively.

We also used the receiver operating characteristic (ROC) curve and the precision-recall (PR) curve to visually describe predictive performances. The ROC curve is created by plotting the true positive rate (y-axis) against the false positive rate (x-axis) under various thresholds. The ROC curve located nearer the left-top corner implied better performance. A PR curve is a plot of the precision (y-axis) against the recall (x-axis) under different thresholds. The recall is identical to S_n, and the precision is defined by

$$Pre = \frac{T_P}{T_P + F_P} \tag{22}$$

The areas under the ROC curve and the PR curve (AUC) were used to quantitatively measure the predictive performance. The AUC ranges from 0 to 1, 1 meaning perfect prediction, 0.5 random prediction, and 0 completely opposite prediction. The larger the AUC, the better the predictive performances.

4. Results and discussion

We used the 6mA-rice-Lv [1] as the training dataset, and the 6mA-rice-chen as the testing dataset. To test the predictive performance across different plant species, both datasets from R. chinensis and F. vesca were also used as the testing datasets.

4.1. Comparison with the state-of-the-art methods

As mentioned previously, a dozens of computational methods have been proposed to detect 6mA sites over the recent ten years. The performances of methods varied with datasets and evaluating ways. We compared our Deep6mAPred with three state-of-the-art methods. Table 3 listed the performances by 5-fold cross validation over the 6mA-rice-Lv. The Deep6mAPred reached better S_n than three baseline methods (Deep6mA [83], SNNRice6mA-large [79] and Deep6mAPred [45]), and achieved competitive S_P, ACC and MCC in contrast with the

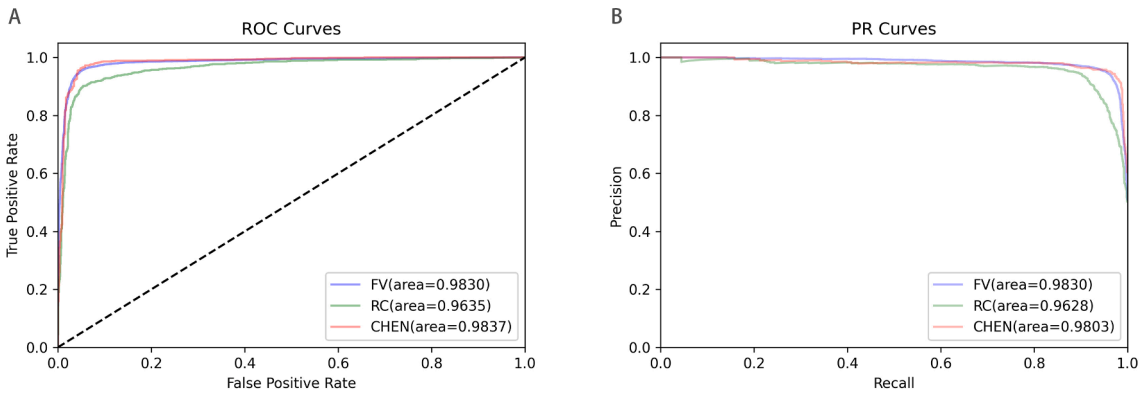

Fig. 3. The Performance by independent test over different plants. (A) ROC Curves. (B) PR Curves.

Table 5
Performance over the F.vesca and the R.chinensis dataset.

	i6mA-Fuse-F		*i6mA-Fuse-R*	
Model	S_n	S_P	S_n	S_P
Deep6mAPred	**0.9534**	**0.9505**	**0.8920**	**0.9506**
Deep6mA*	0.9400	0.9500	0.8700	0.9500
SNNRice6mA-large*	0.9200	0.8400	0.8400	0.8500
MM-6mApred*	0.9600	0.7600	0.8700	0.9500

The asterisk (*) indicated that the results were from the literature [83].

Deep6mA, which completely outperformed the SNNRice6mA-large and Deep6mAPred.

The performance of the independent test over the 6mA-rice-chen was listed in Table 4. The Deep6mAPred increased S_n by 0.1572, *ACC* by 0.0750, *MCC* by 0.1436, and AUC of ROC curve by 0.0237 over the Deep6mA, which completely outperformed the other two methods (the SNNRice6mA-large and the MM-6mAPred). The S_P of the Deep6mAPred was slightly less than that of the Deep6mA, but was much more than those of SNNRice6mA-large and the MM-6mAPred.

The previous cross validation and the independent test indicated that the Deep6mAPred is a better computational method for detecting 6mA sites in rice. To test the Deep6mAPred for the ability to predict 6mA sites across plant species, two additional independent tests were performed, where the Deep6mAPred was trained by the 6mA-rice-Lv. As shown in Fig. 3, both the AUCs of the ROC and PR curve reached 0.9830 on the 6mA-Fuse-F, which were larger than those of the three state-of-the-art methods (AUCs for Deep6mA were 0.9820 and 0.9820, for SNNRice6mA-large were 0.9640 and 0.9630, and for the MM-6mApred were 0.9600 and 0.9590, respectively). As for the 6mA-Fuse-R, the Deep6mAPred outperformed three baseline methods in terms of the AUCs of ROC curves, while in terms of the AUCs of the PR curves it was equivalent to the Deep6mA but superior to the SNNRice6mA-large and MM-6mAPred a bit. As for S_n and S_P, the Deep6mAPred completely exceeded three baseline methods on the two datasets, as shown in Table 5. These results indicated that the Deep6mAPred is not only applicable to detect rice 6mA sites, but also suitable to other plants.

4.2. Discussion

Over the recent 10 years, many methods have employed the CNN and the LSTM to predict 6mA sites, such as the Deep6mA [83]. These methods stacked CNN and LSTM in a series-connection manner. The CNN is good at characterizing local properties, while the LSTM does well in long-term dependency of words in sequences. Stacking them in a series-connection manner would lose their respective merits. Instead, the Deep6mAPred uses the CNN and the LSTM in parallel, which makes full use of their advantages, but avoid their limitations. This might be a reason that the Deep6mAPred outperformed the Deep6mA. In addition, the Deep6mAPred uses the attention mechanism. We further investigate whether the attention mechanism improves predictive performances or not. The ROC curves and the PR curves after removing the attentions from the Deep6mAPred are shown in Fig. 4. Obviously, the inclusion of attentions promotes the predictive performance. The deep learning algorithm is generally iterative, and the iterating number might be influential in the predictive performance of it. We investigated the effects of the iterating number on the Deep6mAPred. We used 50, 100, 150, 200, 250, and 300 epochs to respectively train the Deep6mApred, and used three independent datasets to examine it. As shown in Fig. 5, ROC curves and PR curves across different epochs are nearly identical. The standard deviations of the AUCs are less than 0.0030, implying that the Deep6mAPred is robust to the iterative number.

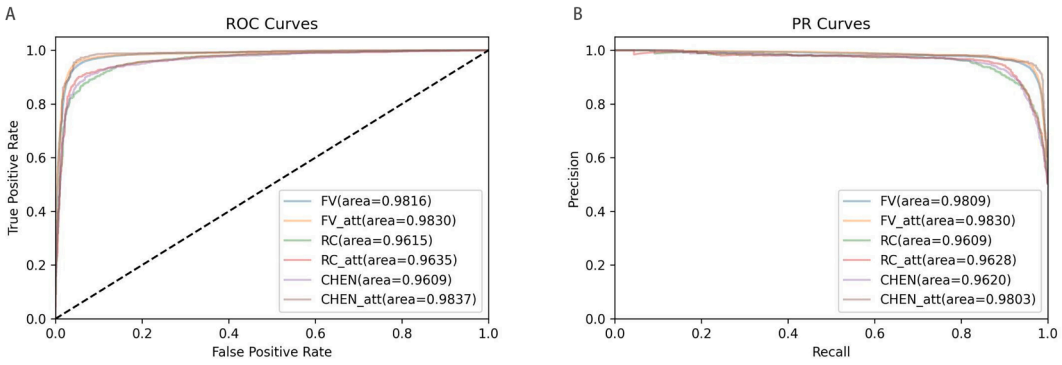

Fig. 4. The Performances with attention mechanism and without attention mechanism over different plant species (A) ROC Curves. (B) PR Curves. FV, RC, and CHEN represent the performances without attention mechanism, and FV_att, RC_att, CHEN_att represent the performance with attention mechanism.

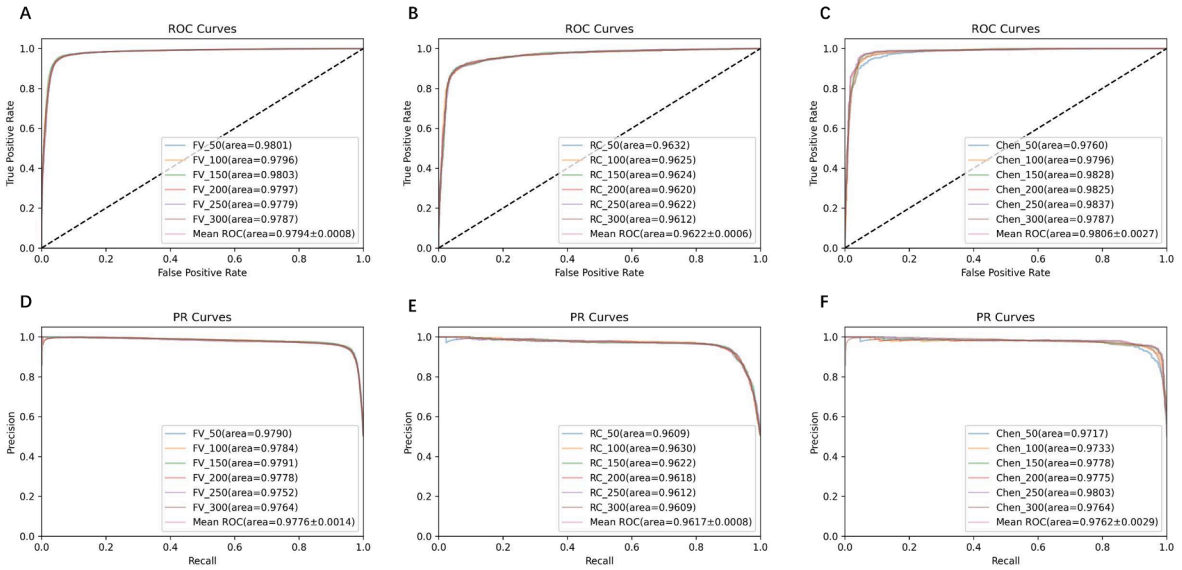

Fig. 5. The ROC curves and PR curves under variously iterating numbers. (A)-(C) is ROC curves for F.vesca, R. chinensis, and 6mA-rice-Chen respectively. (D)-(F) is PR curves for F.vesca, R. chinensis, and 6mA-rice-Chen respectively.

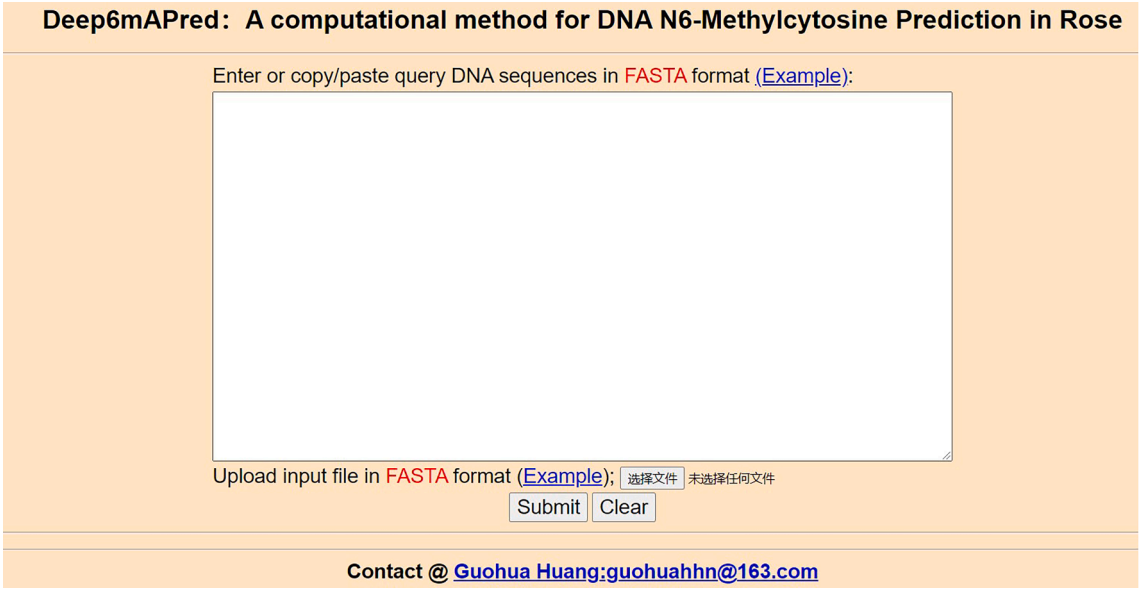

Fig. 6. The web server page of the Deep6mAPred.

4.3. Deep6mAPred webserver

To help scientific researchers use the Deep6mAPred more conveniently, we developed a user-friendly webserver, which is freely available at http://106.13.196.152:7001/. The webserver is easy to use, as shown in Fig. 6. Users firstly choose species, and then submit the sequences in the fasta format. There are two ways to submit sequences: paste sequence to textbox, and upload the file. Clicking the submit button, users would obtain predictive results in several minutes if the number of sequences was less than 100. The time the computation cost is positively proportional to the number of submitted sequences.

5. Conclusion

The 6mA is a key mechanism of regulation in the cellular processes. It is still a challenging task to precisely detect 6mA sites due to the limitation of knowledge about it. We presented a CNN and LSTM-based method (Deep6mAPred) to address this issue. The Deep6mAPred stacked the CNN and the LSTM in a paralleling manner, which makes full use of the strength of both CNN and LSTM. The attention mechanism improved the representation of the 6mA sequences. We also developed a user-friendly webserver to automatically detect 6mA sites. The Deep6mAPred is a competitive method or tool not only for predicting 6mA in rice, but also in other plants.

6. Data availability statements

Data are available in a repository and can be accessed at http://106.13.196.152:7001/.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

Acknowledgements

This work is supported by Scientific Research Fund of Hunan Provincial Education Department (21A0466, 19A215), by the Natural Science Foundation of Hunan Province (2020JJ4034).

References

[1] H. Lv, F.Y. Dao, Z.X. Guan, D. Zhang, J.X. Tan, Y. Zhang, W. Chen, H. Lin, iDNA6mA-Rice: a computational tool for detecting N6-methyladenine sites in rice, Front. Genet. 10 (2019) 793, https://doi.org/10.3389/fgene.2019.00793.
[2] M. Ehrlich, M.A. Gama-Sosa, L.H. Carreira, L.G. Ljungdahl, K.C. Kuo, C. W. Gehrke, DNA methylation in thermophilic bacteria: N 4-methylcytosine, 5-methylcytosine, and N 5 methyladenine, Nucleic Acids Res. 13 (4) (1985) 1399–1412, https://doi.org/10.1093/nar/13.4.1399.
[3] K.J. Wu, The epigenetic roles of DNA N(6)-Methyladenine (6mA) modification in eukaryotes, Cancer Lett. 494 (2020) 40–46, https://doi.org/10.1016/j.canlet.2020.08.025.
[4] G. Zhang, H. Huang, D. Liu, Y. Cheng, X. Liu, W. Zhang, R. Yin, D. Zhang, P. Zhang, J. Liu, C. Li, B. Liu, Y. Luo, Y. Zhu, N. Zhang, S. He, C. He, H. Wang, D. Chen, N6-methyladenine DNA modification in Drosophila, Cell 161 (4) (2015) 893–906, https://doi.org/10.1016/j.cell.2015.04.018.
[5] S.J. Mondo, R.O. Dannebaum, R.C. Kuo, K.B. Louie, A.J. Bewick, K. LaButti, S. Haridas, A. Kuo, A. Salamov, S.R. Ahrendt, R. Lau, B.P. Bowen, A. Lipzen, W. Sullivan, B.B. Andreopoulos, A. Clum, E. Lindquist, C. Daum, T.R. Northen, G. Kunde-Ramamoorthy, R.J. Schmitz, A. Gryganskyi, D. Culley, J. Magnuson, T. Y. James, M.A. O'Malley, J.E. Stajich, J.W. Spatafora, A. Visel, I.V. Grigoriev, Widespread adenine N6-methylation of active genes in fungi, Nat. Genet. 49 (6) (2017) 964–968, https://doi.org/10.1038/ng.3859.
[6] Y. Fu, G.Z. Luo, K. Chen, X. Deng, M. Yu, D. Han, Z. Hao, J. Liu, X. Lu, L.C. Dore, X. Weng, Q. Ji, L. Mets, C. He, N6-methyldeoxyadenosine marks active transcription start sites in Chlamydomonas, Cell 161 (4) (2015) 879–892, https://doi.org/10.1016/j.cell.2015.04.010.
[7] H.H. Evans, T.E. Evans, Methylation of the deoxyribonucleic acid of Physarum polycephalum at various periods during the mitotic cycle, J. Biol. Chem. 245 (23) (1970) 6436–6441, https://doi.org/10.1016/S0021-9258(18)62627-4.
[8] J.G. Reilly, R. Braun, C. Thomas, Methylation in Physarum DNA, FEBS Lett. 116 (2) (1980) 181–184, https://doi.org/10.1016/0014-5793(80)80638-7.
[9] J. Tost, DNA methylation: an introduction to the biology and the disease-associated changes of a promising biomarker, Mol. Biotechnol. 44 (1) (2010) 71–81, https://doi.org/10.1007/s12033-009-9216-2.
[10] J.L. Steenwyk, J.S. Denis, J.M. Dresch, D.A. Larochelle, R.A. Drewell, Whole genome bisulfite sequencing reveals a sparse, but robust pattern of DNA methylation in the Dictyostelium discoideum genome, bioRxiv (2017) 166033. doi:10.1101/166033.
[11] A. Zemach, I.E. McDaniel, P. Silva, D. Zilberman, Genome-wide evolutionary analysis of eukaryotic DNA methylation, Science 328 (5980) (2010) 916–919, https://doi.org/10.1126/science.1186366.
[12] S. Feng, S.J. Cokus, X. Zhang, P.-Y. Chen, M. Bostick, M.G. Goll, J. Hetzel, J. Jain, S.H. Strauss, M.E. Halpern, Conservation and divergence of methylation patterning in plants and animals, Proc. Natl. Acad. Sci. 107 (19) (2010) 8689–8694, https://doi.org/10.1073/pnas.1002720107.
[13] W. Zhou, G. Liang, P.L. Molloy, P.A. Jones, DNA methylation enables transposable element-driven genome expansion, Proc. Natl. Acad. Sci. 117 (32) (2020) 19359–19366, https://doi.org/10.1073/pnas.1921719117.
[14] E. Li, C. Beard, R. Jaenisch, Role for DNA methylation in genomic imprinting, Nature 366 (6453) (1993) 362–365, https://doi.org/10.1038/366362a0.
[15] C. Beard, E. Li, R. Jaenisch, Loss of methylation activates Xist in somatic but not in embryonic cells, Genes Dev. 9 (19) (1995) 2325–2334, https://doi.org/10.1101/gad.9.19.2325.
[16] S. Gonzalo, Epigenetic alterations in aging, J. Appl. Physiol. 109 (2) (2010) 586–597, https://doi.org/10.1152/japplphysiol.00238.2010.
[17] S. Horvath, DNA methylation age of human tissues and cell types, Genome Biol. 14 (10) (2013) 1–20, https://doi.org/10.1186/gb-2013-14-10-r115.
[18] P. Ji, X. Wang, N. Xie, Y. Li, N6-Methyladenosine in RNA and DNA: an epitranscriptomic and epigenetic player implicated in determination of stem cell fate, Stem Cells Int. 2018 (2018) 3256524, https://doi.org/10.1155/2018/3256524.
[19] K.D. Meyer, Y. Saletore, P. Zumbo, O. Elemento, C.E. Mason, S.R. Jaffrey, Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons, Cell 149 (7) (2012) 1635–1646, https://doi.org/10.1016/j.cell.2012.05.003.
[20] D. Dominissini, Moshitch-Moshkovitz s, schwartz s, salmon-Divon M, Ungar L, Osenberg L, Cesarkas K, Jacob-Hirsch J, Amariglio N, Kupiec M, et al: Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq, Nature 485 (2012) 201-206. doi:10.1038/nature11112.
[21] K. Beemon, J. Keith, Localization of N6-methyladenosine in the Rous sarcoma virus genome, J. Mol. Biol. 113 (1) (1977) 165–179, https://doi.org/10.1016/0022-2836(77)90047-x.
[22] S. Zhong, H. Li, Z. Bodi, J. Button, L. Vespa, M. Herzog, R.G. Fray, MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor, Plant Cell 20 (5) (2008) 1278–1288, https://doi.org/10.1105/tpc.108.058883.
[23] Z. Bodi, J.D. Button, D. Grierson, R.G. Fray, Yeast targets for mRNA methylation, Nucleic Acids Res. 38 (16) (2010) 5327–5335, https://doi.org/10.1093/nar/gkq266.
[24] R. Levis, S. Penman, 5′-Terminal structures of poly (A)+ cytoplasmic messenger RNA and of poly (A)+ and poly (A)− heterogeneous nuclear RNA of cells of the dipteran Drosophila melanogaster, J. Mol. Biol. 120 (4) (1978) 487–515, https://doi.org/10.1016/0022-2836(78)90350-9.
[25] M.G. Marinus, A. Lobner-Olesen, DNA methylation, EcoSal Plus 6 (2014) 1, https://doi.org/10.1128/ecosalplus.ESP-0003-2013.
[26] E.L. Greer, M.A. Blanco, L. Gu, E. Sendinc, J. Liu, D. Aristizabal-Corrales, C.H. Hsu, L. Aravind, C. He, Y. Shi, DNA Methylation on N6-Adenine in C. elegans, Cell 161(4) (2015) 868-878. doi:10.1016/j.cell.2015.04.005.
[27] T. Zhang, P. Tan, L. Wang, N. Jin, Y. Li, L. Zhang, H. Yang, Z. Hu, L. Zhang, C. Hu, C. Li, K. Qian, C. Zhang, Y. Huang, K. Li, H. Lin, D. Wang, RNALocate: a resource for RNA subcellular localizations, Nucleic Acids Res. 45 (D1) (2017) D135–D138, https://doi.org/10.1093/nar/gkw728.
[28] M.J. Koziol, C.R. Bradshaw, G.E. Allen, A.S.H. Costa, C. Frezza, J.B. Gurdon, Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications, Nat. Struct. Mol. Biol. 23 (1) (2016) 24–30, https://doi.org/10.1038/nsmb.3145.
[29] J. Liu, Y. Zhu, G.Z. Luo, X. Wang, Y. Yue, X. Wang, X. Zong, K. Chen, H. Yin, Y. Fu, D. Han, Y. Wang, D. Chen, C. He, Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig, Nat. Commun. 7 (1) (2016) 13052, https://doi.org/10.1038/ncomms13052.
[30] Y. Wang, X. Chen, Y. Sheng, Y. Liu, S. Gao, N6-adenine DNA methylation is associated with the linker DNA of H2A. Z-containing well-positioned nucleosomes in Pol II-transcribed genes in Tetrahymena, Nucleic Acids Res. 45 (20) (2017) 11594–11606, https://doi.org/10.1093/nar/gkx883.
[31] J. Willson, DNA 6mA in times of mitochondrial stress, Nat. Rev. Mol. Cell Biol. 21 (5) (2020) 252–253, https://doi.org/10.1038/s41580-020-0240-1.
[32] D. Ratel, J.L. Ravanat, F. Berger, D. Wion, N6-methyladenine: the other methylated base of DNA, BioEssays 28 (3) (2006) 309–315, https://doi.org/10.1002/bies.20342.
[33] D.A. Low, N.J. Weyand, M.J. Mahan, Roles of DNA adenine methylation in regulating bacterial gene expression and virulence, Infect. Immun. 69 (12) (2001) 7197–7204, https://doi.org/10.1128/IAI.69.12.7197-7204.2001.
[34] Y. Fu, D. Dominissini, G. Rechavi, C. He, Gene expression regulation mediated through reversible m6A RNA methylation, Nat. Rev. Genet. 15 (5) (2014) 293–306, https://doi.org/10.1038/nrg3724.
[35] K.G. Au, K. Welsh, P. Modrich, Initiation of methyl-directed mismatch repair, J. Biol. Chem. 267 (17) (1992) 12142–12148, https://doi.org/10.1016/S0021-9258(19)49816-5.
[36] P.J. Pukkila, J. Peterson, G. Herman, P. Modrich, M. Meselson, Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in Escherichia coli, Genetics 104 (4) (1983) 571–582, https://doi.org/10.1093/genetics/104.4.571.
[37] J.L. Campbell, N. Kleckner, E. coli oriC and the dnaA gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork, Cell 62 (5) (1990) 967–979, https://doi.org/10.1016/0092-8674(90)90271-f.
[38] C. Zhou, C. Wang, H. Liu, Q. Zhou, Q. Liu, Y. Guo, T. Peng, J. Song, J. Zhang, L. Chen, Y.u. Zhao, Z. Zeng, D.-X. Zhou, Identification and analysis of adenine N6-methylation sites in the rice genome, Nat. Plants 4 (8) (2018) 554–563, https://doi.org/10.1038/s41477-018-0214-x.
[39] C.L. Xiao, S. Zhu, M. He, Chen, Q. Zhang, Y. Chen, G. Yu, J. Liu, S.Q. Xie, F. Luo, Z. Liang, D.P. Wang, X.C. Bo, X.F. Gu, K. Wang, G.R. Yan, N(6)-Methyladenine DNA Modification in the Human Genome, Molecular cell 71(2) (2018) 306-318. doi:10.1016/j.molcel.2018.06.015.
[40] X. Sheng, J. Wang, Y. Guo, J. Zhang, J. Luo, DNA N6-Methyladenine (6mA) modification regulates drug resistance in triple negative breast cancer, Front. Oncol. 10 (2021) 3241, https://doi.org/10.3389/fonc.2020.616098.
[41] G.Z. Luo, M.A. Blanco, E.L. Greer, C. He, Y. Shi, DNA N(6)-methyladenine: a new epigenetic mark in eukaryotes? Nat. Rev. Mol. Cell Biol. 16 (12) (2015) 705–710, https://doi.org/10.1038/nrm4076.
[42] A.K. Rana, Crime investigation through DNA methylation analysis: methods and applications in forensics, Egypt. J. Forensic Sci. 8 (1) (2018) 1–17, https://doi.org/10.1186/s41935-018-0042-1.
[43] H.G. Hernández, M.Y. Tse, S.C. Pang, H. Arboleda, D.A. Forero, Optimizing methodologies for PCR-based DNA methylation analysis, Biotechniques 55 (4) (2013) 181–197, https://doi.org/10.2144/000114087.
[44] B.A. Flusberg, D.R. Webster, J.H. Lee, K.J. Travers, E.C. Olivares, T.A. Clark, J. Korlach, S.W. Turner, Direct detection of DNA methylation during single-molecule, real-time sequencing, Nat. Methods 7 (6) (2010) 461–465, https://doi.org/10.1038/nmeth.1459.
[45] C. Pian, G. Zhang, F. Li, X. Fan, MM-6mAPred: identifying DNA N6-methyladenine sites based on Markov model, Bioinformatics 36 (2) (2020) 388–392, https://doi.org/10.1093/bioinformatics/btz556.
[46] M.M. Hasan, W. Shoombuatong, H. Kurata, B. Manavalan, Critical evaluation of web-based DNA N6-methyladenine site prediction tools, Briefings Funct. Genomics 20 (4) (2021) 258–272, https://doi.org/10.1093/bfgp/elaa028.
[47] W. Chen, H. Lv, F. Nie, H. Lin, i6mA-Pred: identifying DNA N6-methyladenine sites in the rice genome, Bioinformatics 35 (16) (2019) 2796–2800, https://doi.org/10.1093/bioinformatics/btz015.

[48] J.A. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neural Process. Lett. 9 (3) (1999) 293–300, https://doi.org/10.1023/A:1018628609742.

[49] L. Breiman, Random forests, Machine Learning 45 (1) (2001) 5–32, https://doi.org/10.1023/A:1010933404324.

[50] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Machine Learning 63 (1) (2006) 3–42, https://doi.org/10.1007/s10994-006-6226-1.

[51] L. Kong, L. Zhang, i6mA-DNCP: computational identification of DNA N6-methyladenine sites in the rice genome using optimized dinucleotide-based features, Genes 10 (10) (2019) 828, https://doi.org/10.3390/genes10100828.

[52] P. Feng, H. Yang, H. Ding, H. Lin, W. Chen, K.-C. Chou, iDNA6mA-PseKNC: Identifying DNA N6-methyladenosine sites by incorporating nucleotide physicochemical properties into PseKNC, Genomics 111 (1) (2019) 96–102, https://doi.org/10.1016/j.ygeno.2018.01.005.

[53] Q. Huang, J. Zhang, L. Wei, F. Guo, Q. Zou, 6mA-RicePred: a method for identifying DNA N6-methyladenine sites in the rice genome based on feature fusion, Front. Plant Sci. 11 (2020) 4, https://doi.org/10.3389/fpls.2020.00004.

[54] Z. Liu, W. Dong, W. Jiang, Z. He, csDMA: an improved bioinformatics tool for identifying DNA 6 mA modifications via Chou's 5-step rule, Sci. Rep. 9 (1) (2019) 1–9, https://doi.org/10.1038/s41598-019-49430-4.

[55] H. Xu, R. Hu, P. Jia, Z. Zhao, 6mA-Finder: a novel online tool for predicting DNA N6-methyladenine sites in genomes, Bioinformatics 36 (10) (2020) 3257–3259, https://doi.org/10.1093/bioinformatics/btaa113.

[56] P.M. Granitto, C. Furlanello, F. Biasioli, F. Gasperi, Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products, Chemometrics Intelligent Lab. Syst. 83 (2) (2006) 83–90, https://doi.org/10.1016/j.chemolab.2006.01.007.

[57] M.M. Hasan, B. Manavalan, W. Shoombuatong, M.S. Khatun, H. Kurata, i6mA-Fuse: improved and robust prediction of DNA 6mA sites in the Rosaceae genome by fusing multiple feature representation, Plant Mol. Biol. 103 (1) (2020) 225–234, https://doi.org/10.1007/s11103-020-00988-y.

[58] S. Basith, B. Manavalan, T.H. Shin, G. Lee, SDM6A: a web-based integrative machine-learning framework for predicting 6mA sites in the rice genome, Mol. Ther.-Nucleic Acids 18 (2019) 131–141, https://doi.org/10.1016/j.omtn.2019.08.011.

[59] C.-Y.-J. Peng, K.L. Lee, G.M. Ingersoll, An introduction to logistic regression analysis and reporting, J. Educ. Res. 96 (1) (2002) 3–14, https://doi.org/10.1080/00220670209598786.

[60] D.J. Hand, K. Yu, Idiot's Bayes—not so stupid after all? Int. Statistical Rev. 69 (3) (2001) 385–398, https://doi.org/10.2307/1403452.

[61] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. Syst. Sci. 55 (1) (1997) 119–139, https://doi.org/10.1006/jcss.1997.1504.

[62] M.M. Hasan, S. Basith, M.S. Khatun, G. Lee, B. Manavalan, H. Kurata, Meta-i6mA: an interspecies predictor for identifying DNA N 6-methyladenine sites of plant genomes by exploiting informative features in an integrative machine-learning framework, Briefings Bioinf. 22 (3) (2021) bbaa202, https://doi.org/10.1093/bib/bbaa202.

[63] T. Xue, S. Zhang, H. Qiao, i6mA-VC: a multi-classifier voting method for the computational identification of DNA N6-methyladenine sites, Interdiscip. Sci: Comput. Life Sci. 13 (3) (2021) 413–425, https://doi.org/10.1007/s12539-021-00429-4.

[64] A. Wahab, S.D. Ali, H. Tayara, K.T. Chong, iIM-CNN: Intelligent identifier of 6mA sites on different species by using convolution neural network, IEEE Access 7 (2019) 178577–178583, https://doi.org/10.1109/ACCESS.2019.2958618.

[65] M. Tahir, H. Tayara, K.T. Chong, iDNA6mA (5-step rule): Identification of DNA N6-methyladenine sites in the rice genome by intelligent computational model via Chou's 5-step rule, Chemometrics Intelligent Lab. Syst. 189 (2019) 96–101, https://doi.org/10.1016/j.chemolab.2019.04.007.

[66] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, L. Jackel, Handwritten digit recognition with a back-propagation network, Adv. Neural Inf. Process. Syst. 2 (1989).

[67] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324, https://doi.org/10.1109/5.726791.

[68] T. Young, D. Hazarika, S. Poria, E. Cambria, Recent trends in deep learning based natural language processing, ieee Computational intelligenCe magazine 13(3) (2018) 55-75. doi:10.1109/MCI.2018.2840738.

[69] A.U. Rehman, A.K. Malik, B. Raza, W. Ali, A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis, Multimedia Tools Appl. 78 (18) (2019) 26597–26613, https://doi.org/10.1007/s11042-019-07788-7.

[70] A.H. Bukhari, M.A.Z. Raja, M. Sulaiman, S. Islam, M. Shoaib, P. Kumam, Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting, IEEE Access 8 (2020) 71326–71338, https://doi.org/10.1109/ACCESS.2020.2985763.

[71] T.-Y. Kim, S.-B. Cho, Predicting residential energy consumption using CNN-LSTM neural networks, Energy 182 (2019) 72–81, https://doi.org/10.1016/j.energy.2019.05.230.

[72] Q. Huang, W. Zhou, F. Guo, L. Xu, L. Zhang, 6mA-Pred: identifying DNA N6-methyladenine sites based on deep learning, PeerJ 9 (2021), e10813, https://doi.org/10.7717/peerj.10813.

[73] N.Q.K. Le, Q.T. Ho, Deep transformers and convolutional neural network in identifying DNA N6-methyladenine sites in cross-species genomes, Methods (2021), https://doi.org/10.1016/j.ymeth.2021.12.004.

[74] M.U. Rehman, K.T. Chong, DNA6mA-MINT: DNA-6mA Modification Identification Neural Tool, Genes 11 (8) (2020) 898, https://doi.org/10.3390/genes11080898.

[75] H.-T. Wang, F.-H. Xiao, G.-H. Li, Q.-P. Kong, Identification of DNA N 6-methyladenine sites by integration of sequence features, Epigenet. Chromatin 13 (1) (2020) 1–10, https://doi.org/10.1186/s13072-020-00330-2.

[76] J. Cai, D. Wang, R. Chen, Y. Niu, X. Ye, R. Su, G. Xiao, L. Wei, A Bioinformatics tool for the prediction of DNA N6-methyladenine modifications based on feature fusion and optimization protocol, Front. Bioeng. Biotechnol. 8 (2020) 502, https://doi.org/10.3389/fbioe.2020.00502.

[77] J. Khanal, D.Y. Lim, H. Tayara, K.T. Chong, i6mA-stack: A stacking ensemble-based computational prediction of DNA N6-methyladenine (6mA) sites in the Rosaceae genome, Genomics 113 (1) (2021) 582–592, https://doi.org/10.1016/j.ygeno.2020.09.054.

[78] S. He, L. Kong, J. Chen, iDNA6mA-Rice-DL: A local web server for identifying DNA N6-methyladenine sites in rice genome by deep learning method, J. Bioinf. Comput. Biol. 19 (5) (2021) 2150019, https://doi.org/10.1142/S0219720021500190.

[79] H. Yu, Z. Dai, SNNRice6mA: a deep learning method for predicting DNA N6-methyladenine sites in rice genome, Front. Genet. 10 (2019) 1071, https://doi.org/10.3389/fgene.2019.01071.

[80] W. Liu, H. Li, SICD6mA: identifying 6mA sites using deep memory network, BioRxiv (2020), https://doi.org/10.1101/2020.02.02.930776.

[81] R. Amin, C.R. Rahman, M. Toaha, S. Islam, S. Shatabda, i6mA-CNN: a convolution based computational approach towards identification of DNA N6-Methyladenine sites in rice genome, arXiv preprint arXiv:2007.10458 (2020). doi:10.48550/arXiv.2007.10458.

[82] Z. Abbas, H. Tayara, K. to Chong, SpineNet-6mA: A Novel Deep Learning Tool for Predicting DNA N6-Methyladenine Sites in Genomes, IEEE Access 8 (2020) 201450-201457. doi:10.1109/ACCESS.2020.3036090.

[83] Z. Li, H. Jiang, L. Kong, Y. Chen, K. Lang, X. Fan, L. Zhang, C. Pian, Deep6mA: A deep learning framework for exploring similar patterns in DNA N6-methyladenine sites across different species, PLoS Comput. Biol. 17 (2) (2021), e1008767, https://doi.org/10.1371/journal.pcbi.1008767.

[84] X. Yang, X. Ye, X. Li, L. Wei, iDNA-MT: identification DNA modification sites in multiple species by using multi-task learning based a neural network tool, Front. Genet. 12 (2021) 411, https://doi.org/10.3389/fgene.2021.663572.

[85] Y. Yu, W. He, J. Jin, L. Cui, R. Zeng, L. Wei, iDNA-ABT: advanced deep learning model for detecting DNA methylation with adaptive features and transductive information maximization, Bioinformatics 37 (24) (2021) 4603–4610, https://doi.org/10.1093/bioinformatics/btab677.

[86] Z.Y. Liu, J.F. Xing, W. Chen, M.W. Luan, R. Xie, J. Huang, S.Q. Xie, C.L. Xiao, MDR: an integrative DNA N6-methyladenine and N4-methylcytosine modification database for Rosaceae, Hortic. Res. 6 (1) (2019) 78, https://doi.org/10.1038/s41438-019-0160-4.

[87] W. Yin, K. Kann, M. Yu, H. Schütze, Comparative study of CNN and RNN for natural language processing 10.48550/arXiv (2017) 1702.01923, arXiv preprint arXiv:1702.01923.

[88] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012) 1097–1105, https://doi.org/10.1145/3065386.

[89] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014). doi:10.48550/arXiv.1409.1556.

[90] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions. CoRR, arXiv preprint arXiv:1409.4842 (2014). doi:10.1109/cvpr.2015.7298594.

[91] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, 770-778. doi: 10.1109/CVPR.2016.90.

[92] Z. Wu, C. Shen, A. Van Den Hengel, Wider or deeper: revisiting the resnet model for visual recognition, Pattern Recogn. 90 (2019) 119–133, https://doi.org/10.1016/j.patcog.2019.01.006.

[93] Z. Lu, X. Jiang, A. Kot, Deep coupled resnet for low-resolution face recognition, IEEE Signal Process Lett. 25 (4) (2018) 526–530, https://doi.org/10.1109/LSP.2018.2810121.

[94] B.A. Pearlmutter, Learning state space trajectories in recurrent neural networks, Neural Comput. 1 (2) (1989) 263–269, https://doi.org/10.1162/neco.1989.1.2.263.

[95] B.A. Pearlmutter, Dynamic recurrent neural networks, (1990).

[96] W. Zaremba, I. Sutskever, O. Vinyals, Recurrent neural network regularization, arXiv preprint arXiv:1409.2329 (2014). doi:10.48550/arXiv.1409.2329.

[97] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. (2017) 5998–6008, https://doi.org/10.1109/CVPR.2016.90.

[98] C. Raffel, D.P. Ellis, Feed-forward networks with attention can solve some long-term memory problems, arXiv preprint arXiv:1512.08756 (2015). doi:10.48550/arXiv.1512.08756.

[99] M.-H. Guo, Z.-N. Liu, T.-J. Mu, S.-M. Hu, Beyond self-attention: External attention using two linear layers for visual tasks, arXiv preprint arXiv:2105.02358 (2021). doi:10.48550/arXiv.2105.02358.

[100] Y. Chen, Y. Kalantidis, J. Li, S. Yan, J. Feng, Aˆ 2-nets: Double attention networks, Adv. Neural Inf. Process. Syst. 31 (2018).