applied sciences

MDPI

Article

6mAPred-MSFF: A Deep Learning Model for Predicting DNA N6-Methyladenine Sites across Species Based on a Multi-Scale Feature Fusion Mechanism

Rao Zeng and Minghong Liao *

Department of Software Engineering, School of Informatics, Xiamen University, Xiamen 361005, China; rao.zeng@hotmail.com
* Correspondence: liao@xmu.edu.cn

Abstract: DNA methylation is one of the most extensive epigenetic modifications. DNA N6-methyladenine (6mA) plays a key role in many biology regulation processes. An accurate and reliable genome-wide identification of 6mA sites is crucial for systematically understanding its biological functions. Some machine learning tools can identify 6mA sites, but their limited prediction accuracy and lack of robustness limit their usability in epigenetic studies, which implies the great need of developing new computational methods for this problem. In this paper, we developed a novel computational predictor, namely the 6mAPred-MSFF, which is a deep learning framework based on a multi-scale feature fusion mechanism to identify 6mA sites across different species. In the predictor, we integrate the inverted residual block and multi-scale attention mechanism to build lightweight and deep neural networks. As compared to existing predictors using traditional machine learning, our deep learning framework needs no prior knowledge of 6mA or manually crafted sequence features and sufficiently capture better characteristics of 6mA sites. By benchmarking comparison, our deep learning method outperforms the state-of-the-art methods on the 5-fold cross-validation test on the seven datasets of six species, demonstrating that the proposed 6mAPred-MSFF is more effective and generic. Specifically, our proposed 6mAPred-MSFF gives the sensitivity and specificity of the 5-fold cross-validation on the 6mA-rice-Lv dataset as 97.88% and 94.64%, respectively. Our model trained with the rice data predicts well the 6mA sites of other five species: *Arabidopsis thaliana*, *Fragaria vesca*, *Rosa chinensis*, *Homo sapiens*, and *Drosophila melanogaster* with a prediction accuracy 98.51%, 93.02%, and 91.53%, respectively. Moreover, via experimental comparison, we explored performance impact by training and testing our proposed model under different encoding schemes and feature descriptors.

Keywords: DNA N6-methyladenine; deep learning; site prediction; depthwise separable convolution; inverted residual structure; attention mechanism; feature fusion

check for updates

Citation: Zeng, R.; Liao, M. 6mAPred-MSFF: A Deep Learning Model for Predicting DNA N6-Methyladenine Sites across Species Based on a Multi-Scale Feature Fusion Mechanism. *Appl. Sci.* **2021**, *11*, 7731. https://doi.org/10.3390/app11167731

Academic Editor: Leyi Wei

Received: 9 June 2021
Accepted: 18 August 2021
Published: 22 August 2021

1. Introduction

Epigenetics refers to the reversible and heritable changes in gene function when there is no change in the nuclear DNA sequence [1]. DNA methylation modifications play important roles in epigenetic regulation of gene expression without altering the sequence, and it is widely distributed in the genome of different species [2]. It can be divided into three categories according to the position of methylation modification: N6-methyladenine (6mA), 5-Methylcytosine (5mC), and N4-methylcytosine (4mC) [3,4]. DNA methylations at the 5th position of the pyrimidine ring of cytosine 5-Methylcytosine(5mC) and at the 6th position of the purine ring of adenine N6-methyladenine (6mA) are the most common DNA modifications in eukaryotes and prokaryotes, respectively [5]. Previous studies have shown that DNA N6-methyladenine (6mA) is associated with germ cell differentiation, stress response, embryonic development, nervous system, and other processes [6–9]. N6-Methyladenine (6mA) DNA methylation has recently been implicated as a potential new epigenetic marker in eukaryotes, including the *Arabidopsis thaliana*, *Rice*, *Drosophila melanogaster*, and so on [5,8,10]. Zhang et al. reveal that 6mA is a conserved DNA modification that is positively associated with gene expression and contributes to key agronomic traits in plants [11]. Some studies have found that N6-methyladenine DNA modification is also widely distributed in the Human Genome and plays important biological functions. Xiao et al. demonstrate that 6mA DNA modification is extensively present in human cells, and the decrease of genomic DNA 6mA promotes human tumorigenesis [12]. Zhou et al. proposed a 6mA DNA modification area as a new mechanism for the epigenetic regulation of stem cell differentiation [13]. Xie et al. report that N6-methyladenine DNA modifications are enriched in human glioblastoma, and targeting regulators of this modification can inhibit cancer growth by altering heterochromatin landscapes and downregulating oncogenic programs [14]. Therefore, how to quickly and accurately detect the DNA 6mA modification sites is also an important research topic in these epigenetic studies.

Due to the rapid development of high-throughput sequence technology, various experimental techniques have been proposed to detect DNA 6mA modifications and study protein function. Pormraning et al. developed a protocol using bisulfite sequencing and a methyl-DNA immunoprecipitation technique to analyze genome-wide DNA methylation in eukaryotes [15]. Krais et al. reported a fast and sensitive method for the quantification of global adenine methylation in DNA, using laser-induced fluorescence and capillary electrophoresis [16]. Flusberg et al. proposed the single-molecule real-time sequencing (SMRT) technology to detect the 4mC and 6mA sites from the whole genome [17]. Greer et al. used ultra-high performance liquid chromatography coupled with mass spectrometry technique to access DNA 6mA levels in *Caenorhabditis elegans* [18]. By performing mass spectrometry analysis and 6mA immunoprecipitation followed by sequencing (IP-seq), Zhou and his colleagues obtained the 6mA profile of the rice genome [10].

Although experimental methods indeed yielded encouraging results, the technology cannot detect m6A sites from the whole genome and the cost of the technique is high. Therefore, it is necessary to develop a computational model that can efficiently and accurately predict and identify 6mA sites. Recent studies focus more on the recognition of 6mA sites using machine learning [19], which is capable of predicting 6mA sites based on genome sequences, without any prior experimental knowledge. Chen et al. developed the first ML-based method named i6mA-Pred for identifying DNA 6mA sites and provided a benchmark 6mA dataset containing 880 6mA sites and 880 non-6mA sites in the rice genome. Their method used a Support Vector Machine (SVM) classifier based on chemical features of nucleotides and position-specific nucleotide frequencies [20]. The i6mA-Pred shows a good classification performance in rice 6mA data. However, the association information among nucleotides near 6mA sites is ignored. Pian et al. proposed a new classification method called MM-6mAPred based on a Markov model which makes use of the transition probability between adjacent nucleotides to identify 6mA sites [21]. Pian et al. built and evaluated their MM-6mAPred based on the 6mA-rice-Chen benchmark dataset. Their results show that MM-6mAPred outperformed i6mA-Pred in prediction of 6mA sites. Basith et al. developed a novel computational predictor, called SDM6A, which explores various features and five encoding methods to identify the DNA 6mA sites [22]. Basith et al. also trained and evaluated their SDM6A based on the 6mA-rice-Chen benchmark dataset, and they found that SDM6A outperformed i6mA-Pred on the 6mA-rice-Chen benchmark dataset. The above three prediction models are trained on the 6mA-rice-Chen benchmark dataset including only the 880 rice 6mA sites and 880 non-6mA sites. Even though the above methods have improved the performance for identifying 6mA sites, too few data sets have been adopted to fully reflect the whole genome and to build robust models. Lv et al. developed a machine learning method of predicting 6mA sites named iDNA6mA-rice which was trained and evaluated on the dataset 6mA-rice-Lv containing 154,000 6mA sites and 154,000 non-6mA sites in the rice genome [23]. iDNA6mA-rice utilized Random Forest to perform the classification for identifying 6mA sites after using mono-nucleotide binary encoding to formulate positive and negative samples.

In recent years, deep learning has not only been developed as a new research direction in machine learning, but has also made a lot of achievements in data mining, machine translation, natural language processing, and other related fields. In the field of computational biology [24–28], deep learning has been widely applied [29–31], especially in solving the problems of genome sequence-based by convolutional neural networks (CNN). Tahir et al. proposed an intelligent computational model called iDNA 6mA (5-step rule) which extracts the key features from DNA input sequences via the convolution neural network (CNN) to identify 6mA sites in the rice genome [32]. Yu et al. developed a simple and lightweight deep learning model named SNNRice6mA to identify DNA 6mA sites in the rice genome and showed its advantages over other methods [33]. Li et al. developed a deep learning framework named Deep6mA to identify DNA 6mA sites. Deep6mA, composed of a CNN and a bidirectional LSTM (BLSTM) module, is shown to have a better performance than other methods above on 6mA prediction [34]. Although the above methods have made great progress, their performance is still not satisfactory. Furthermore, most of the methods are designed for one specific species such as the rice genome. Although some methods provide the across species validation test, the results are not as good as the original species-specific model.

In this paper, we proposed a novel deep learning method called 6mAPred-MSFF to identify DNA 6mA sites. In this predictor, we establish the model by integrating the inverted residual block, the multi-scale channel attention module (MS-CAM), Bi-directional Long Short Term Memory Network (Bi-LSTM), and attentional feature fusion (AFF). To capture more feature information, we use the inverted residual block to expand the input features to a high dimension and filter them with a lightweight depthwise convolution. After that, we project them back to a low-dimensional representation as the input of the multi-scale channel attention module (MS-CAM). Bi-LSTM is designed to capture the long dependent information in sequences. The features generated from the Bi-LSTM are fed into the AFF to be fused. In the experimental part, we evaluate our predictor and other 6mA predictors with 5-fold cross validation on seven benchmark datasets. We also evaluate these 6mA predictors trained by the 6mA-rice-Lv dataset to predict the 6mA sites of other species. Finally, we explore the performance impact under the different encoding methods and feature descriptors.

2. Materials and Methods

2.1. Dataset Collection

Previous studies have demonstrated that a stringent dataset is essential for building a robust predictive model. We collected seven 6mA site datasets of six species including *Rice*, *Arabidopsis thaliana* (*A. thaliana*), *Rosa chinensis* (*R. chinensis*), *Fragaria vesca* (*F. vesca*), *Homo sapiens* (*H. sapiens*) and *Drosophila melanogaster* (*D. melanogaster*). There are two benchmark datasets 6mA-rice-Lv and 6mA-rice-Chen for *Rice*. The 6mA-rice-Lv dataset include 15,400 positive samples and 15,400 negative samples which are obtained from NCBI Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/genome/10 (accessed on 20 August 2021)) following the steps proposed by Lv et al. [23]. The 6mA-rice-Chen benchmark dataset includes the 880 rice 6mA sites and 880 non-6mA sites which are obtained from GEO (https://www.ncbi.nlm.nih.gov/geo/ (accessed on 20 August 2021)) under the accession number GSE103145 following the steps proposed by Chen et al. [20].

To demonstrate that our model has the ability to predict the 6mA sites of other species and 6mA shares similar patterns across different species, we collected DNA 6mA sequences of *Arabidopsis thaliana* and *Drosophila melanogaster* from the MethSMRT database (http://sysbio.gzzoc.com/methsmrt/ (accessed on 20 August 2021)) [35] and also obtained the DNA 6mA sequences of *Fragaria vesca* and *Rosa chinensis* from the MDR database [36]. Preliminary trials indicated that, when the length of the segments is 41 bp with the 6mA in the center, the highest predictive results could be obtained. Thus, the sequences of all positive samples are 41 bp. In order to construct a high-quality benchmark dataset, the following two steps were performed. Firstly, according to the the Methylome Analysis

Technical Note [35], the sequences with modQV no less than 30 are left for the subsequent analysis. It should be noted that, in order to obtain statistically significant results, if the raw data are too small, this step was ignored to get more samples. Secondly, to avoid redundancy and reduce homology bias, sequences with more than 80% sequence similarity were removed using the CD-HIT program [37]. After the above two steps, the objective and strict positive datasets for the above species were obtained. Negative samples for *Arabidopsis thaliana*, *Drosophila melanogaster*, *Fragaria vesca*, and *Rosa chinensis* were also collected. These samples are 41 nt long sequences with Adenine in the center and were proved to be unmethylated by experiments. For convenience, we used the *Homo sapiens* dataset (http://lin-group.cn/server/iDNA-MS/download.html (accessed on 20 August 2021)) [38] directly, including 18,335 positive samples and 18,335 negative samples. The details of the datasets are presented in Table 1.

Table 1. Summary of seven benchmark datasets of six species.

Dataset	Positives	Negatives	Total
6mA-rice-Lv	154,000	154,000	300,800
6mA-rice-Chen	880	880	1760
A. thaliana	15,937	15,937	31,874
R. chinensis	11,815	11,815	23,630
F. vesca	22,700	22,700	45,400
H. sapiens	18,335	18,335	36,670
D. melanogaster	11,191	11,191	22,382

2.2. Architecture of 6mAPred-MSFF

In this work, we proposed a new predictive model based on a multi-scale feature fusion mechanism. An overview of our model architecture is illustrated in Figure 1. There are four modules in our model including the sequence embedding module, the feature extraction module, feature fusion module, and prediction module. First, in the sequence embedding module, we use four different encoding schemes (1-gram, *NAC*, 2-grams, *DNC*) for the input of the embedding layer. Each nucleotide is represented as a dense vector of float point values. As a result, the genomic sequences can be represented by four feature matrices, respectively. We concatenate the feature matrices of the 1-gram and *NAC* and the feature matrices of the 2-grams and *DNC* matrices, respectively. Second, the resulting two feature matrices are fed into the feature extraction module which consists of the inverted residual block and the multi-scale channel attention mechanism module (MS-CAM). To extract more information from the global features and filter the features as the source of the MS-CAM, we use the inverted residual block including an expansion point-wise layer, a convolution layer, and a projection point-wise layer. The MS-CAM extract and combine the global and local features of the two feature matrices, respectively. Before fusing the two feature matrices, we use the bidirectional LSTM layer to learn the information about the long-distance dependence. Afterward, in the feature fusion module, we combine the features extracted from the bidirectional LSTM layer by adding operation and feed them into the MS-CAM module to calculate the fusion weight of the two feature matrices. We multiply the features matrices with the corresponding fusion weights and combine the results by adding operation. In this way, we obtain the features by aggregating the local and global contexts of the four different encoding schemes. Third, in the prediction module, the outputs of the features are passed through dropout layers for the prevention of overfitting during the training. After the dropout layer, to unstack the output, a flatten function was used to squash the feature vectors from the previous layers. Right after a flattened layer, a fully connected (FC) dense layer used as Dense(n), with the number of n neurons set as 32. At last, an FC layer was applied and used a sigmoid function for the binary classification. Sigmoid is used to squeeze the values between the range of 0 and 1 to represent the probability of having 6mA and non-6mA sites.

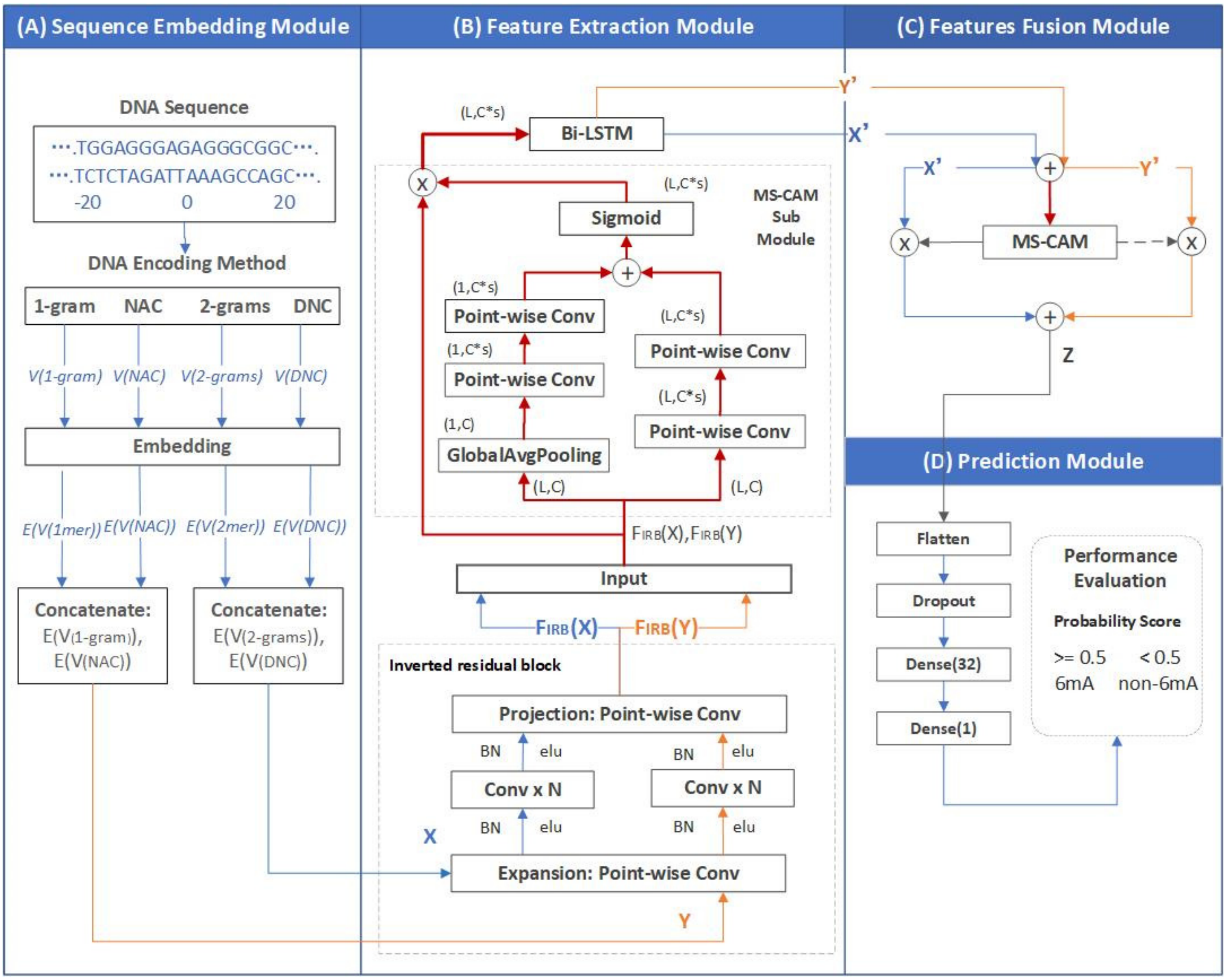

Figure 1. The flowchart of 6mAPred-MSFF. The sequence embedding module uses four different encoding schemes (1-gram, *NAC*, 2-grams, *DNC*) for the input of the embedding layer. Next, the feature matrices are fed into the feature extraction module to capture and combine the global and local features, respectively. Afterwards, the local and global contexts of the four different encoding schemes are aggregated into the feature fusion module. Finally, the output of the feature fusion module is fed into the prediction module to predict the 6mA site of a certain species.

The 6mAPred-MSFF is trained by the Adam optimizer and the binary cross-entropy loss function. Via giving a prediction score, the 6mAPred-MSFF can determine whether the 6mA sites are detectable in DNA sequences. It is detectable if the prediction score is >0.5; otherwise, it is not. The model is implemented using Keras 2.4.3. Below, each of ourmodules will be described in detail.

2.2.1. Sequence Embedding Module

DNA sequences are consisting of four nucleotides: "A" (adenine), "G" (guanine), "C" (cytosine), and "T" (thymine). Undetermined bases are annotated as "N". The DNA sequences can be expressed as $S = s_1, s_2, \ldots s_L$ where s_i is the i-th word and the L is the length of the sequence. We use four encoding methods to define 'words' in DNA sequences including 1-gram, *NAC*, 2-grams, and *DNC*.

1-gram encoding method. The n-grams are the set of all possible subsequences of nucleobases [39]. One-gram encoding sets the n-grams number n to 1. As a result, the 1-gram nucleobase sequences include 'A', 'T', 'C', 'G'. We define a dictionary that maps the nucleotides to numbers, such as A (adenine) to 1, T (thymine) is 2, C (cytosine) to 3, and G (guanine) to 4. In this way, we map a DNA sequence into a real number vector of length 41, which is denoted as $V_{i,1-gram}$:

$$V_{i,1-gram} = [v_1 v_2 v_3 \ldots v_{39} v_{40} v_{41}]_i, \quad v \in [1,2,3,4] \tag{1}$$

NAC encoding method. The Nucleic Acid Composition (*NAC*) encoding calculates the frequency of each nucleic acid type in a nucleotide sequence. The frequencies of all four natural nucleic acids (i.e., *"ATCG"*) can be calculated as:

$$f_{NAC}(t) = \frac{N(t)}{N}, \quad t \in \{A,T,C,G\} \tag{2}$$

for a DNA sequence S of length L, and it can be converted into a vector of length 4, which is denoted as $V_{i,NAC}$:

$$V_{i,NAC} = [f_{NAC}(A), f_{NAC}(T), f_{NAC}(C), f_{NAC}(G)]_i \tag{3}$$

2-gram encoding method. A 2-gram encoding sets the n-grams number n to 2. We split the DNA sequences into overlapping n-gram nucleobases. The total possible number is 16, since there are four types of nucleobases. For example, we split a DNA sequence into overlapping 2-gram nucleobase sequences as follows: *GTTGT* … *CTT* → *'GT'*, *'TT'*, *'TG'*, *'GT'*, … , *'CT'*, *'TT'*. We define a dictionary that maps the 2-gram nucleobase sequences to the 1–16 numbers set. In this way, we map a DNA sequence into a real number vector of length 40, which donates $V_{i,2-gram}$:

$$V_{i,2-gram} = [v_1 v_2 v_3 \ldots v_{39} v_{40}]_i, \quad v \in [1,2,3,\ldots 14,15,16] \tag{4}$$

DNC encoding method. The Di-Nucleotide Composition (*DNC*) encoding calculates the frequency of every two nucleic acid types in a nucleotide sequence which gives 16 descriptors. It is defined as:

$$f_{DNC}(r,s) = \frac{N_{rs}}{N-1}, \quad r,s \in \{A,T,C,G\} \tag{5}$$

where N_{rs} is the number of di-nucleotide represented by nucleic acid types r and s. For a DNA sequence S of length L, it can be converted into a vector of length 16, which is donated as $V_{i,DNC}$:

$$V_{i,DNC} = [f_{DNC}(AA), f_{DNC}(AC), \ldots, f_{DNC}(TC), f_{DNC}(TG)]_i \tag{6}$$

Next, the vectors $V_{i,1-gram}$, $V_{i,NAC}$, $V_{i,2-gram}$, and $V_{i,DNC}$ are fed to an embedding layer, transforming into learnable embedding vectors. To establish the relationship between $V_{i,1-gram}$ and $V_{i,NAC}$, we concatenate the two embedding vectors of $V_{i,1-gram}$ and $V_{i,NAC}$ to a new feature vector that is denoted as X_i:

$$X_i = Concatenate\big(Embedding\big(V_{i,1-gram}\big), Embedding(V_{i,NAC})\big) \tag{7}$$

We do the same operation on the embedding vectors of $V_{i,2-gram}$ and $V_{i,DNC}$, generating a new feature vector which is denoted as Y_i:

$$Y_i = Concatenate\big(Embedding\big(V_{i,2-gram}\big), Embedding(V_{i,DNC})\big) \tag{8}$$

To extract more feature information, the vectors X_i and Y_i are fed into the feature extraction module.

2.2.2. Feature Extraction Module

The inverted residual block. To reduce the computation of the network, we always design the network with a low-dimensional space. However, the network can not capture enough information in low-dimensional space. To solve these problems, the inverted residual with linear bottleneck was proposed by Sandler et al. [40] in the MobileNetV2. This module takes an input as a low-dimensional compressed representation, which is first expanded to high dimensions and filtered with a lightweight depthwise convolution. Features are subsequently projected back to a low-dimensional representation with a linear convolution which can prevent the destruction of information in low-dimensional space. Our proposed method follows the idea to build an inverted residual block that consists of an expansion layer, a depthwise layer, and a projection layer. The expansion layer and the projection layer are the 1×1 convolution layer, called the pointwise convolution with a different expansion ratio. The depthwise layer is a depthwise separable convolution layer that is a key building block for many efficient neural network architectures [40–44]. The basic idea is to replace a full convolutional operator with a factorized version that splits convolution into two separate layers. The first layer is called a depthwise convolution, and it performs lightweight filtering by applying a single convolutional filter per input channel. The second layer is a point-wise convolution, which is responsible for building new features through computing linear combinations of the input channels. The inverted residual block operator can be expressed as a composition of three operators which is denoted as $F_{IRB}(X)$:

$$F_{IRB}(X) = [A \circ N \circ B]X \tag{9}$$

where A is the expansion layer that is a linear transformation: $A : R^{s \times s \times k} \to R^{s \times s \times n}$, N is the depthwise layer which contains several non-linear per-channel transformation: $N : R^{s \times s \times n} \to \ldots \to R^{s' \times s' \times n}$, and B is the projection layer which is again a linear transformation to the output domain: $B : R^{s \times s \times n} \to R^{s' \times s' \times k'}$. For our inverted residual block $F_{IRB}(X)$, A and B are the linear transformation without any nonlinear activations. We use the ELU activation function in the operator $N = ELU \circ dwise \circ ELU$. which accelerates the learning rate of the network. The ELU function is shown in the following equation:

$$f(x) = \begin{cases} a(e^x - 1) & if(x < 0) \\ x & if(0 \leq x) \end{cases} \tag{10}$$

Thus, we can obtain k' feature maps containing the 6mA site features captured by the inverted residual block.

Multi-scale attention mechanism block. The attention mechanism in deep learning, which mimics the human visual attention mechanism [45,46], is originally developed on a global scale. For example, the matrix multiplication in self-attention draws global dependencies of each word in a sentence [47] or each pixel in an image [48–51]. The Squeeze-and-Excitation Networks (SENet) squeeze global spatial information into a channel descriptor to capture channel-wise dependencies [52]. However, merely aggregating contextual information on a global scale is too biased and weakens the features of small objects. To solve the problem, the multi-scale channel attention module (MS-CAM) proposed by Yimian Dai et al. [53] aggregates global and local channel contexts given an intermediate feature $X \in R^{H \times W \times C}$ with C channels and feature maps of size $H \times W$. The attention channel weights $M(X) \in R^C$ can be computed as:

$$M(X) = L(X) \oplus G(X) \tag{11}$$

where $L(X)$ denotes the local channel context, and $G(X)$ denotes the global channel context. The stacked architecture of MS-CAM is shown in Figure 1. $L(X) \in R^{C \times H \times W}$ is computed as follows:

$$L(X) = \beta(PWConv_2(\beta(PWConv_1(X)))) \tag{12}$$

where β denotes the Batch Normalization (BN) [54]. The pointwise convolution (PWConv) is used as the local channel context aggregator, which only exploits pointwise channel interactions for each spatial position. The expansion ratios of $PWConv_1$ and $PWConv_2$ are r and $1/r$, respectively. Different from the original MS-CAM, we remove the nonlinear activation function between the convolution layer to prevent the destruction of feature information. $G(X) \in R^C$ is computed as follows:

$$G(X) = \beta(PWConv_2(\beta(PWConv_1(g(X))))) \tag{13}$$

where $g(X)$ is global average pooling (GAP) and computed as follows:

$$g(X) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{[:,i,j]} \tag{14}$$

The expansion ratios of $PWConv_1$ and $PWConv_2$ are r and $1/r$, respectively. β denotes the Batch Normalization (BN). The refined $X' \in R^{H \times W \times C}$ can be obtained as follows:

$$X' = X \otimes \sigma(M(X)) = X \otimes \sigma(L(X) \oplus G(X)) \tag{15}$$

where σ is the sigmoid function. $\oplus$ denotes the broadcasting addition and $\otimes$ denotes the element-wise multiplication.

Bi-directional Long Short Term Memory Network. Recurrent Neural Network (RNN) is a powerful neural network for processing sequential data [55]. As a special type of RNN, Long Short Term Memory Network (LSTM) is not only designed to capture the long dependent information in sequence but also overcome the training difficulty of RNN due to gradient explosion and disappearance [56]; thus, it is the most widely used RNN in real applications. In LSTM, a storage mechanism is used to replace the hidden function in traditional RNN, with a purpose to enhance the learning ability of LSTM for long-distance dependency. Bi-directional LSTM (BLSTM), compared with unidirectional LSTM, better captures the information of sequence context. The LSTM components are updated by the following formulations:

$$\begin{gathered} f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \\ i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ C'_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\ C_t = i_t C' + f_t C_{t-1} \\ o_t = \sigma(W_x h_{t-1} + W_h x_t + W_c C_t + b_0) \\ h_t = o_t \tanh(C_t) \end{gathered} \tag{16}$$

where i_t,f_t, and o_t represent the input, forget, and output gate, respectively; C'_t is temp value for calculating C_t; t denotes the recurrent time step; W_f, W_i, W_C and b are the weights of each equation. h_t is the output of LSTM at the t time step. The bi-directional LSTM consists of two direction LSTM networks. Therefore, the h_t is computed as follows:

$$h_t = \left[\overrightarrow{h}_t \oplus \overleftarrow{h}_t\right] \tag{17}$$

2.2.3. Feature Fusion Module

Feature fusion, the combination of features from different layers or branches, is an omnipresent part of modern network architectures. It is often implemented via simple operations, such as summation or concatenation, but this might not be the best choice. We use a uniform and general scheme, namely attentional feature fusion (AFF) which is proposed by Yimian Dai et al. [53]. The architecture of AFF is shown in Figure 1. Based on the multi-scale channel attention module $M(X)$, the AFF can be expressed as follows:

$$Z = M(X \oplus Y) \otimes X + (1 - M(X \oplus Y)) \otimes Y \tag{18}$$

where $Z \in R^{H \times W \times C}$ is the fused feature, and the element-wise summation of $X \oplus Y$ is used as the initial integration for MS-CAM. $M(X \oplus Y)$ and $1 - M(X \oplus Y)$ are the fusion weights for X and Y, respectively. Please note that the fusion weights consist of real numbers between 0 and 1, which enable the network to conduct a soft selection or weighted averaging between X and Y.

2.2.4. Prediction Module

The fused feature Z generated from the AFF module is fed into the dropout layer which prevents overfitting and can improve the reliability and robustness of the model by discarding some intermediate features. We use a flatten function to integrate the feature vectors that are generated from the dropout layer. Afterwards, the flattened feature vectors are fed to the fully connected (FC) layer that has 32 hidden units. The activation function used in the FC layer is the ELU activation function. Finally, a sigmoid activation function is used to combine the outputs from the FC layer to make the final decision.

6mAPred-MSFF is trained by the Adam optimizer [57], which is computationally efficient, has little memory requirements, is invariant to diagonal rescaling of gradients, and is well suited for problems that are large in terms of data/parameters. For binary classification tasks, the loss function for training 6mAPred-MSFF is set as the binary cross-entropy for the measuring the difference between the target and the predicted output:

$$L(w) = -\sum_{i=1}^{N} y_i \log(y'_i) + (1 - y_i)\log(1 - y'_i) + \alpha\|w_2\| \tag{19}$$

where y_i is the true label, y'_i is the corresponding predicted value from 6mAPred-MSFF, and $\alpha\|w_2\|$ is a regularization term to avoid overfitting.

2.2.5. Evaluation Metrics

In our experiment, we used the following four indicators to evaluate the predictive performance of our proposed model, including Accuracy (ACC), Sensitivity (SN), Specificity (SP), and Mathew's Correlation Coefficient (MCC). They are the four commonly used indicators for classifier performance evaluation in other bioinformatics fields [58–87]. They are formulated as follows:

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

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

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

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

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

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. The receiver operating characteristic curve (ROC), the area under the curve (AUC), and precision recall curves (PRC) are used to show the detailed performance of different methods [88]. The AUC ranges from 0.5 to 1. The higher the AUC score, the better the performance of the model [89–91].

3. Results and Discussion

3.1. Performance Comparison on Seven Benchmark Datasets

To demonstrate the effectiveness of the proposed method, we compared its performance with three other existing state-of-the-art methods, including MM-6mAPred [21], SNNRice6mA [33], and Deep6mA [34]. MM-6mAPred is traditional machine learning method training the model by hand-made features extracted from original DNA sequences.

SNNRice6mA and Deep6mA are the deep learning methods. All the predictors are evaluated on the seven benchmark datasets by using 5-fold cross validation.

Table 2 lists the performances of the proposed method, namely 6mAPred-MSFF, and four existing predictors on the seven datasets. For the 6mA-rice-Lv dataset, when compared to the second-best method Deep6mA, our proposed method achieves an SN of 97.88%, an SP of 96.64%, an ACC of 96.26%, and an MCC of 0.926, yielding a relative improvement over Deep6mA of 3.15%, 0.92%, 2.03%, and 4.63%, respectively. Compared with the machine learning method MM-6mAPred, our proposed method improves the performance of SN, SP, ACC, and MCC by 4.61%, 5.8%, 5.12%, and 12.65%, respectively. For the 6mA-rice-Chen dataset, compared with the runner-up predictor Deep6mA, with the only exception that the value of SP is better than ours, our proposed method improves the performance of SN, ACC, and MCC by 9.89%, 2.79%, and 6.81%, respectively. For the *A. thaliana* and *R. chinensis* datasets, with the exception that the predictor Deep6mA achieves the best value of SP, and our proposed method improves the performance of SN, ACC, and MCC by 0.28–5.2%. Moreover, we can see that the performance of 6mAPred-MSFF is also better than all other methods on the *F. vesca*, *H. sapiens*, and *D. melanogaster* datasets.

Table 2. Performance comparison of the proposed method and 6mA predictors based on seven datasets.

Benchmark Dataset	Method	SN (%)	SP (%)	ACC (%)	MCC	AUC
6mA-rice-Lv	6mAPred-MSFF	97.88	94.64	96.26	0.926	0.989
	Deep6mA	94.73	93.72	94.23	0.885	0.981
	SNNRice6mA-large	94.97	92.22	93.59	0.872	0.978
	MM-6mAPred	93.27	88.84	91.05	0.822	0.955
6mA-rice-Chen	6mAPred-MSFF	96.59	89.43	93.01	0.862	0.976
	Deep6mA	86.70	93.75	90.22	0.807	0.958
	SNNRice6mA-large	87.61	86.70	87.16	0.743	0.942
	MM-6mAPred	87.50	88.86	88.18	0.764	0.943
A. thaliana	6mAPred-MSFF	89.92	87.85	88.88	0.778	0.954
	Deep6mA	84.72	89.52	87.12	0.743	0.942
	SNNRice6mA-large	85.28	88.15	86.71	0.735	0.939
	MM-6mAPred	82.62	88.82	83.82	0.664	0.904
R. chinensis	6mAPred-MSFF	95.61	96.56	96.09	0.922	0.988
	Deep6mA	94.97	96.66	95.81	0.917	0.987
	SNNRice6mA-large	91.96	94.55	93.25	0.865	0.977
	MM-6mAPred	92.27	90.72	91.50	0.830	0.959
F. vesca	6mAPred-MSFF	98.79	98.44	98.62	0.972	0.998
	Deep6mA	98.35	97.10	98.48	0.970	0.998
	SNNRice6mA-large	97.72	97.65	97.72	0.955	0.996
	MM-6mAPred	95.13	94.72	94.93	0.899	0.980
H. sapiens	6mAPred-MSFF	94.65	94.91	94.78	0.896	0.985
	Deep6mA	90.16	95.18	92.67	0.854	0.985
	SNNRice6mA-large	88.62	93.56	91.09	0.823	0.972
	MM-6mAPred	88.08	87.22	87.65	0.753	0.937
D. melanogaster	6mAPred-MSFF	97.08	96.72	96.90	0.938	0.990
	Deep6mA	95.12	95.04	95.08	0.902	0.984
	SNNRice6mA-large	94.64	93.36	94.00	0.880	0.982
	MM-6mAPred	90.65	89.94	90.30	0.806	0.955

For a more intuitive comparison, we further compared the ROC (receiver operating characteristic) curves and PR (precision-recall) curves of the four predictors, which are illustrated in Figures 2 and 3, respectively. We can observe that our method achieves the best values for an auROC of 0.989 and an auPRC of 0.981 on the 6mA-rice-Lv dataset. We also achieve the best values for an auROC of 0.976 and auPRC of 0.973 on the 6mA-rice-Chen dataset. For the other species, our proposed method also achieves the best values of auROC and auPRC. Therefore, we can conclude that our proposed method can achieve the best predictive performance for detecting 6mA sites in rice species.

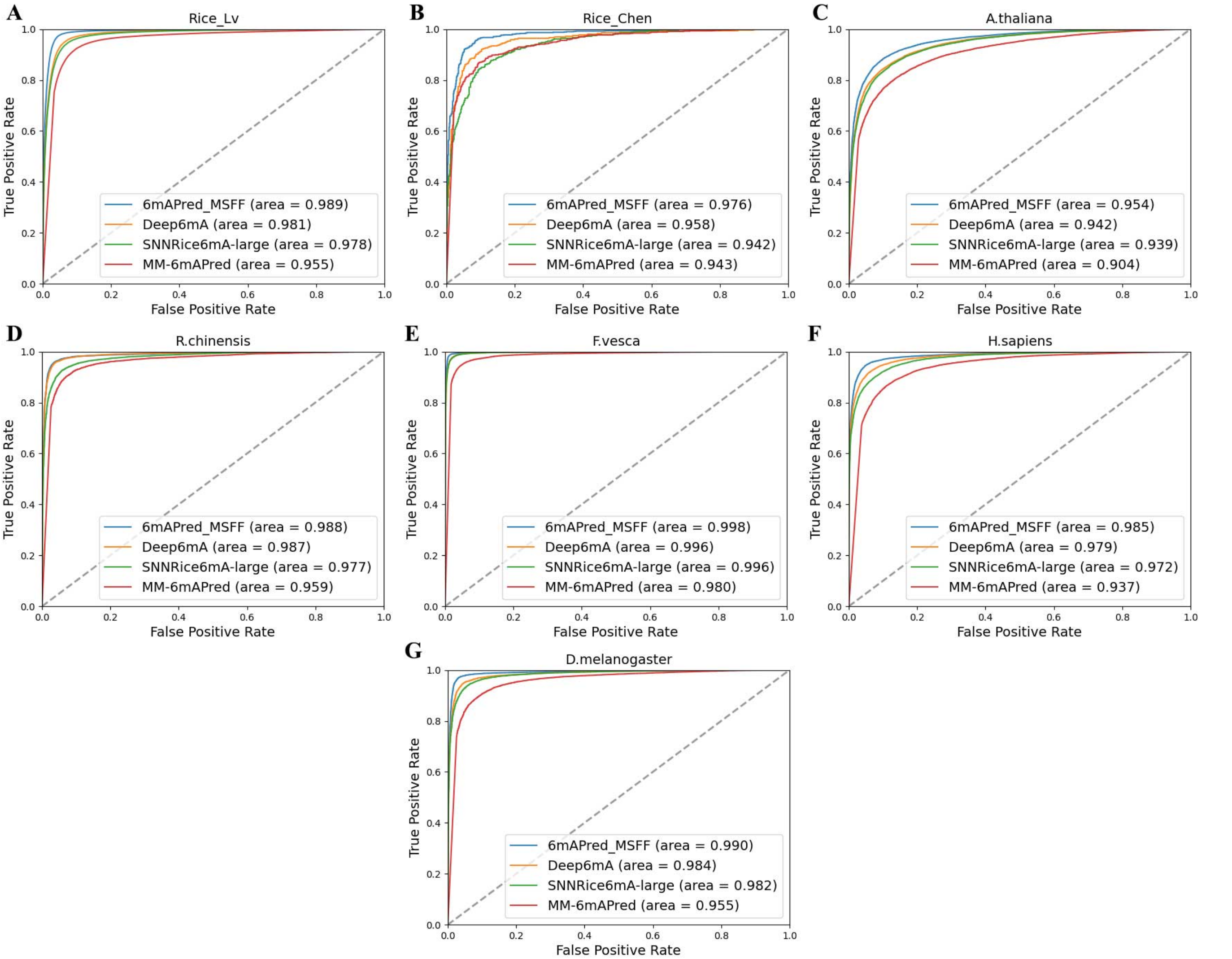

Figure 2. The ROC curves (**A**–**G**) of four models (6mAPred-MSFF, Deep6mA, SNNRice6mA-large, and MM-6mAPred) based on the seven datasets (6mA-rice-Lv, 6mA-rice-Chen, *A. thaliana*, *R. chinensis*, *F. vesca*, *H. sapiens*, and *D. melanogaster*).

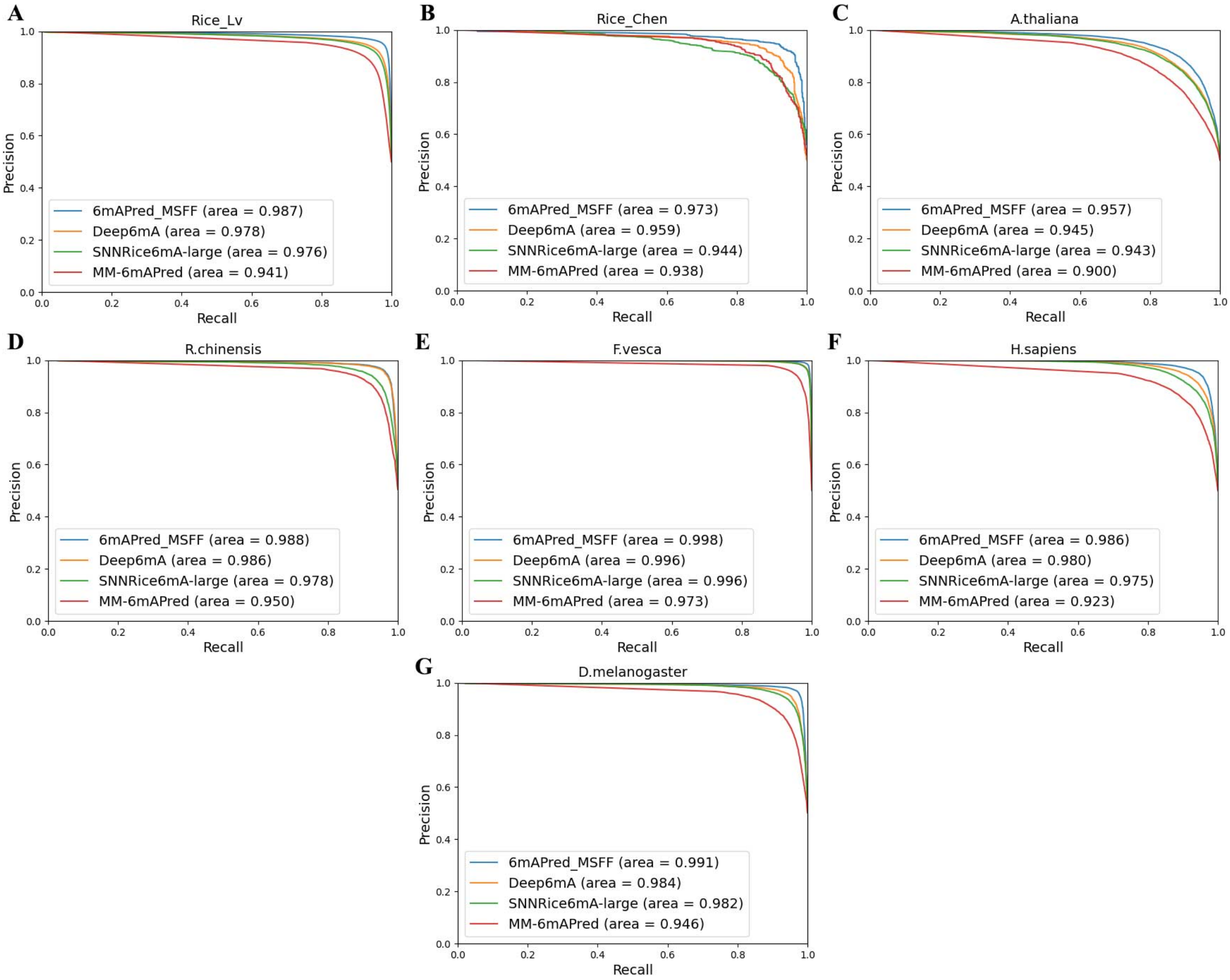

Figure 3. The PRC curves (**A**–**G**) of four models (6mAPred-MSFF, Deep6mA, SNNRice6mA-large, and MM-6mAPred) based on the seven datasets (6mA-rice-Lv, 6mA-rice-Chen, *A. thaliana, R. chinensis, F. vesca, H. sapiens*, and *D. melanogaster*).

3.2. Validation on Other Species

To validate whether the model trained on rice data is applicable to predict the 6mA sites of other species. We take the datasets including *A. thaliana*, *R. chinensis*, *F. vesca*, *H. sapiens*, *and D. melanogaster* as the independent test datasets. We evaluate the performance of 6mAPred-MSFF, which was trained on the rice 6mA-rice-Lv dataset and performed the same evaluation for other methods including Deep6mA, SNNRice6mA-large, and MM-6mAPred.

The prediction results are listed in Table 3. For the two plant species (*F. vesca* and *R. chinensis*), with the only exception that the value of SP of our proposed method is lower than Deep6mA, and our proposed method outperforms all other methods in terms of SN, SP, ACC, and MCC. Specifically, for *R. chinensis*, our proposed method achieves an SN of 90.66%, an SP of 94.26%, an ACC of 92.46%, and an MCC of 0.850, where Deep6mA does have a higher SP of 95.29%. For *F. vesca*, our proposed method achieves an SN of 93.95%, an SP of 95.15%, an ACC of 94.55%, and an MCC of 0.891. However, Deep6mA does have a higher SP of 96.22%. Compared with the two plant species (*F. vesca* and *R. chinensis*), the performance of predicting 6mA sites of the two species (*H. sapiens* and *D. melanogaster)* is not very good. For the two animal species (*H. sapiens* and *D. melanogaster)*, our proposed method outperforms all other methods in terms of SN, ACC, and MCC;

however, the Deep6mA achieves the best value of SP. Surprisingly, all models trained on the 6mA-rice-Lv dataset fail in predicting the 6mA sites of *A. thaliana*, since the best value of SN is 58.66%; although there is a better value of SP, the performance of predicting 6mA sites of *A. thaliana* can not meet the requirement in practice. Therefore, we can conclude that the models trained with an 6mA-rice-Lv dataset can be used to predict the 6mA sites of other species and the features of 6mA sites also have certain similarities among different species.

Table 3. Performance comparison of methods trained with the 6mA-rice-Lv dataset on cross species.

Species	Method	SN (%)	SP (%)	ACC (%)	MCC	AUC
R. chinensis	6mAPred-MSFF	90.66	94.26	92.46	0.850	0.966
	Deep6mA	88.87	95.29	92.07	0.843	0.966
	SNNRice6mA-large	89.79	94.13	91.96	0.840	0.964
	MM-6mAPred	88.51	91.35	89.93	0.799	0.949
F. vesca	6mAPred-MSFF	93.95	95.15	94.55	0.891	0.982
	Deep6mA	92.52	96.22	94.37	0.888	0.982
	SNNRice6mA-large	93.26	95.15	94.20	0.884	0.980
	MM-6mAPred	90.33	93.48	91.91	0.839	0.965
H. sapiens	6mAPred-MSFF	72.51	93.33	82.92	0.673	0.884
	Deep6mA	68.82	94.44	81.63	0.655	0.881
	SNNRice6mA-large	70.93	93.22	82.07	0.658	0.894
	MM-6mAPred	71.85	88.67	80.26	0.614	0.873
D. melanogaster	6mAPred-MSFF	74.30	92.62	83.46	0.680	0.903
	Deep6mA	68.81	93.78	81.30	0.646	0.901
	SNNRice6mA-large	70.73	92.27	81.50	0.645	0.906
	MM-6mAPred	64.96	88.84	76.90	0.554	0.868
A. thaliana	6mAPred-MSFF	58.66	94.28	76.47	0.567	0.837
	Deep6mA	54.56	95.26	74.91	0.545	0.834
	SNNRice6mA-large	57.92	94.11	76.01	0.558	0.853
	MM-6mAPred	57.14	91.44	74.30	0.517	0.845

3.3. Performance Impact by Different Encoding Schemes and Feature Descriptors

Our proposed model uses four different encoding schemes (1-gram, *NAC*, 2-grams, *DNC*) as the input of the embedding layer. The initial features will have an important impact on the training results of deep learning models. To explore the impact of different encoding methods, we will evaluate our proposed model with N-grams encoding scheme and different feature descriptors as the initial input.

N-grams encoding. To validate the impact of the N-grams encoding, we compare our proposed model with the encoding schemes (1-gram, 2-grams, 3-grams 1-gram and 2-grams, 1-gram and 3-grams, 2-grams and 3-grams) based on the 6mA-rice-Lv dataset. The prediction results are listed in Table 4. From Table 4, we can see that the 1-gram encoding scheme achieves an SN of 97.15% second only to our proposed method. The 1-gram and 2-gram encoding scheme achieves an SN of 94.47%, an SP of 94.67%, an ACC of 95.86, and an MCC of 0.917. Therefore, our proposed method uses the 1-gram and 2-gram encoding scheme integrating *NAC* and *DNC* to encode the input sequences.

Table 4. Performance comparison of the proposed method on different encoding schemes based on the 6mA-rice-Lv dataset.

Method	SN (%)	SP (%)	ACC (%)	MCC	AUC
6mAPred-MSFF	97.88	94.64	96.26	0.926	0.989
1-gram	97.15	93.68	95.42	0.909	0.987
2-grams	95.63	93.56	94.59	0.892	0.983
3-grams	96.10	93.60	94.85	0.897	0.984
1-gram and 2-grams	97.05	94.67	95.86	0.917	0.988
1-gram and 3-grams	95.97	93.49	94.73	0.895	0.983
2-grams and 3-grams	95.77	94.51	95.14	0.903	0.985

Feature descriptors. There are several feature descriptors including Kmer, NCP, EIIP, and ENAC which are employed in machine learning methods. To validate the impact of the different feature descriptors, we take the feature descriptors as the initial input of our proposed model trained on 6mA-rice-Lv and perform the same evaluation on the Deep6mA, Random Forest (RF), and Artificial Neural Network (ANN). The prediction results are listed in Table 5. From Table 5, we can see that our proposed method trained by the NCP achieves the best values of an SN of 96.05%, an SP of 94.56%, an ACC of 95.31%, and an MCC of 0.906. The models trained by Kmer which concatenate 1mer, 2mer, and 3mer do not achieve satisfactory results. Meanwhile, we can observe that our proposed method outperforms all other methods in terms of SN, SP, ACC, and MCC under different feature descriptors.

Table 5. Performance comparison of methods based on the 6mA-rice-Lv dataset under different feature descriptors.

Feature	Method	SN (%)	SP (%)	ACC (%)	MCC	AUC
Kmer (1mer + 2mer + 3mer)	6mAPred-MSFF	67.17	64.94	66.08	0.321	0.719
	Deep6mA	69.00	60.35	64.67	0.294	0.700
	RF	66.96	62.33	64.65	0.293	0.697
	ANN	66.87	62.35	64.61	0.293	0.697
NCP	6mAPred-MSFF	96.05	94.56	95.31	0.906	0.986
	Deep6mA	95.82	92.95	94.38	0.888	0.983
	RF	92.90	91.51	92.20	0.844	0.966
	ANN	92.92	91.53	92.22	0.845	0.966
EIIP	6mAPred-MSFF	95.35	91.10	93.22	0.865	0.975
	Deep6mA	94.20	88.77	91.48	0.831	0.966
	RF	92.56	91.21	91.89	0.838	0.964
	ANN	92.61	91.15	91.88	0.838	0.964
ENAC	6mAPred-MSFF	95.34	91.89	93.62	0.872	0.979
	Deep6mA	95.35	90.29	92.82	0.858	0.972
	RF	90.82	91.16	90.99	0.820	0.959
	ANN	90.80	91.17	90.99	0.820	0.959

For a more intuitive comparison, we further compared the ROC (receiver operating characteristic) curves and PR (precision-recall) curves of the four feature descriptors, which are illustrated in Figures 4 and 5, respectively. We can observe that our proposed method achieves the best values of an auROC of 0.986 and an auPRC of 0.984 under the NCP. Our proposed method also achieves the best values of an auROC and auPRC under other feature descriptors. Therefore, we can conclude that our proposed method has a better ability to capture feature information.

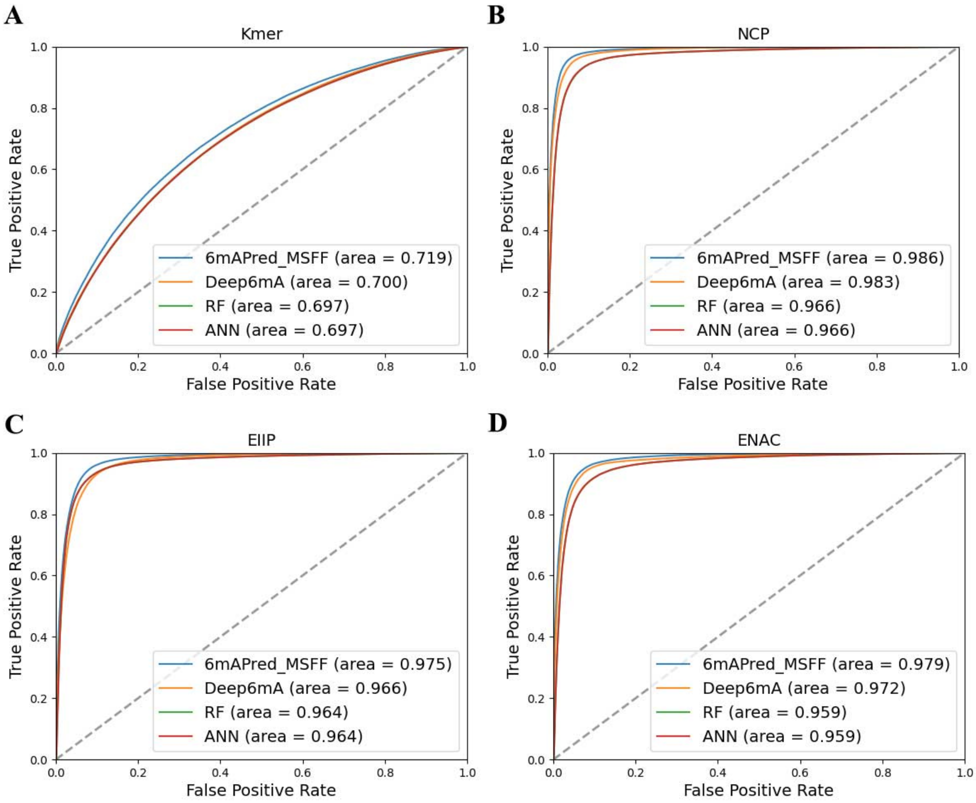

Figure 4. The ROC curves (**A**–**D**) of four models (6mAPred-MSFF, Deep6mA, SNNRice6mA-large, and MM-6mAPred) based on four feature descriptors (Kmer, NCP, EIIP, ENAC).

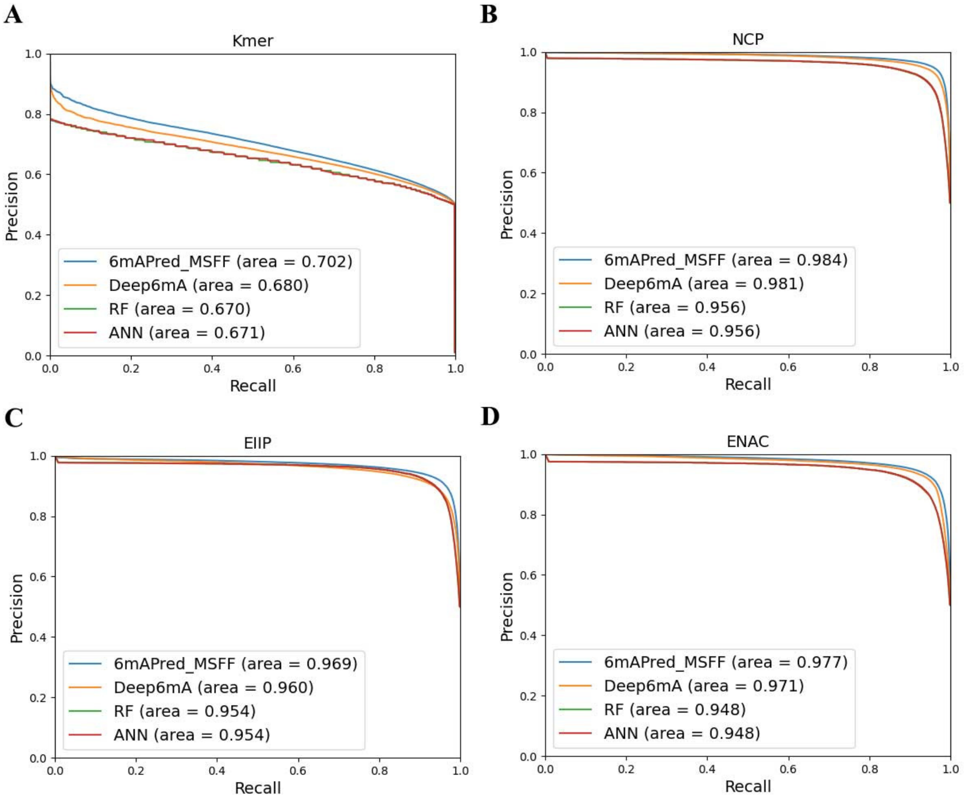

Figure 5. The PR curves (**A**–**D**) of four models (6mAPred-MSFF, Deep6mA, SNNRice6mA-large, and MM-6mAPred) based on four feature descriptors (Kmer, NCP, EIIP, ENAC).

4. Conclusions

In this study, we have proposed a novel predictor called 6mAPred-MSFF to predict the DNA 6mA sites. 6mAPred-MSFF is the first deep learning predictor, in which we integrate the global and local context by the inverted residual block and multi-scale channel attention module (MS-CAM). To fuse different feature vectors by fusion weights, 6mAPred-MSFF uses an attentional feature fusion (AFF) module based on the MS-CAM. As compared to existing predictors, our proposed method can automatically learn the global and local features and capture the characteristic specificity of 6mA sites. On the other hand, the experimental results demonstrate that our proposed method can effectively increase the accuracy and the generalization ability of the DNA 6mA sites prediction. Our proposed deep learning method shows better performance on the rice benchmark and the independent test of the other species.

Author Contributions: R.Z. and M.L.: conceptualization, investigation, methodology, formal analysis, supervision, project administration. R.Z.: software, validation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization. M.L.: funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by "The National Natural Science Foundation of China, grant number 61702450".

Institutional Review Board Statement: Not available.

Informed Consent Statement: Not available.

Data Availability Statement: Publicly available datasets were analyzed in this study. The data and source code can be found here: https://github.com/raozeng/6mAPred-MSFF (accessed on 20 August 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zuo, Y.; Song, M.; Li, H.; Chen, X.; Cao, P.; Zheng, L.; Cao, G. Analysis of the Epigenetic Signature of Cell Reprogramming by Computational DNA Methylation Profiles. *Curr. Bioinform.* **2020**, *15*, 589–599. [CrossRef]
2. Ratel, D.; Ravanat, J.-L.; Berger, F.; Wion, D. N6-methyladenine: The other methylated base of DNA. *BioEssays* **2006**, *28*, 309–315. [CrossRef]
3. Chen, W.; Yang, H.; Feng, P.; Ding, H.; Lin, H. iDNA4mC: Identifying DNA N^4-methylcytosine sites based on nucleotide chemical properties. *Bioinformatics* **2017**, *33*, 3518–3523. [CrossRef] [PubMed]
4. Wei, L.; Su, R.; Luan, S.; Liao, Z.; Manavalan, B.; Zou, Q.; Shi, X. Iterative feature representations improve N4-methylcytosine site prediction. *Bioinformatics* **2019**, *35*, 4930–4937. [CrossRef] [PubMed]
5. Liang, Z.; Shen, L.; Cui, X.; Bao, S.; Geng, Y.; Yu, G.; Liang, F.; Xie, S.; Lu, T.; Gu, X.; et al. DNA N^6-adenine methylation in *Arabidopsis thaliana*. *Dev. Cell* **2018**, *45*, 406–416. [CrossRef] [PubMed]
6. Liu, J.; Zhu, Y.; Luo, G.-Z.; Wang, X.; Yue, Y.; Wang, X.; Zong, X.; Chen, K.; Yin, H.; Fu, Y.; et al. Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. *Nat. Commun.* **2016**, *7*, 13052. [CrossRef] [PubMed]
7. Yao, B.; Cheng, Y.; Wang, Z.; Li, Y.; Chen, L.; Huang, L.; Zhang, W.; Chen, D.; Wu, H.; Tang, B.; et al. DNA N6-methyladenine is dynamically regulated in the mouse brain following environmental stress. *Nat. Commun.* **2015**, *8*, 1122. [CrossRef] [PubMed]
8. Zhang, G.; Huang, H.; Liu, D.; Cheng, Y.; Liu, X.; Zhang, W.; Yin, R.; Zhang, D.; Zhang, P.; Liu, J.; et al. N^6-Methyladenine DNA Modification in *Drosophila*. *Cell* **2015**, *161*, 893–906. [CrossRef] [PubMed]
9. Zhang, Y.; Kou, C.; Wang, S.; Zhang, Y. Genome-wide Differential-based Analysis of the Relationship between DNA Methylation and Gene Expression in Cancer. *Curr. Bioinform.* **2019**, *14*, 783–792. [CrossRef]
10. Zhou, C.; Wang, C.; Liu, H.; Zhou, Q.; Liu, Q.; Guo, Y.; Peng, T.; Song, J.; Zhang, J.; Chen, L.; et al. Identification and analysis of adenine N^6-methylation sites in the rice genome. *Nat. Plants* **2018**, *4*, 554–563. [CrossRef]
11. Zhang, Q.; Liang, Z.; Cui, X.; Ji, C.; Li, Y.; Zhang, P.; Liu, J.; Riaz, A.; Yao, P.; Liu, M.; et al. N^6-Methyladenine DNA Methylation in *Japonica* and *Indica* Rice Genomes and Its Association with Gene Expression, Plant Development, and Stress Responses. *Mol. Plant* **2018**, *11*, 1492–1508. [CrossRef]
12. Xiao, C.-L.; Zhu, S.; He, M.; Chen, D.; Zhang, Q.; Chen, Y.; Yu, G.; Liu, J.; Xie, S.-Q.; Luo, F.; et al. N6-Methyladenine DNA Modification in the Human Genome. *Mol. Cell* **2018**, *71*, 1–13. [CrossRef] [PubMed]
13. Zhou, C.; Liu, Y.; Li, X.; Zou, J.; Zou, S. DNA N6-methyladenine demethylase ALKBH1 enhances osteogenic differentiation of human MSCs. *Bone Res.* **2016**, *4*, 16033. [CrossRef] [PubMed]

14. Xie, Q.; Wu, T.P.; Gimple, R.C.; Li, Z.; Prager, B.C.; Wu, Q.; Yu, Y.; Wang, P.; Wang, Y.; Gorkin, D.U.; et al. N6-methyladenine DNA Modification in Glioblastoma. *Cell* **2018**, *175*, 306–318. [CrossRef]
15. Pomraning, K.R.; Smith, K.M.; Freitag, M. Genome-wide high throughput analysis of DNA methylation in eukaryotes. *Methods* **2009**, *47*, 142–150. [CrossRef]
16. Krais, A.M.; Cornelius, M.G.; Schmeiser, H.H. Genomic N^6-methyladenine determination by MEKC with LIF. *Electrophoresis* **2010**, *31*, 3548–3551. [CrossRef]
17. Flusberg, B.A.; Webster, D.R.; Lee, J.H.; Travers, K.J.; Olivares, E.C.; Clark, T.A.; Korlach, J.; Turner, S.W. Direct detection of dnA methylation during single-molecule, real-time sequencing. *Nat. Methods* **2010**, *7*, 461–465. [CrossRef] [PubMed]
18. Greer, E.L.; Blanco, M.A.; Gu, L.; Sendinc, E.; Liu, J.; Aristizábal-Corrales, D.; Hsu, C.-H.; Aravind, L.; He, C.; Shi, Y. DNA Methylation on N^6 Adenine in *C. elegans*. *Cell* **2015**, *161*, 868–878. [CrossRef] [PubMed]
19. Ao, C.; Yu, L.; Zou, Q. Prediction of bio-sequence modifications and the associations with diseases. *Brief. Funct. Genom.* **2021**, *20*, 1–18. [CrossRef] [PubMed]
20. Chen, W.; Lv, H.; Nie, F.; Lin, H. i6mA-Pred: Identifying DNA N^6-methyladenine sites in the rice genome. *Bioinformatics* **2019**, *35*, 2796–2800. [CrossRef]
21. Pian, C.; Zhang, G.; Li, F.; Fan, X. MM-6mAPred: Identifying DNA N6-methyladenine sites based on Markov Model. *Bioinformatics* **2019**, *36*, 388–392. [CrossRef] [PubMed]
22. Basith, S.; Manavalan, B.; Shin, T.H.; Lee, G. SDM6A: A Web-Based Integrative Machine-Learning Framework for Predicting 6mA Sites in the Rice Genome. *Mol. Ther. Nucleic Acids* **2019**, *18*, 131–141. [CrossRef]
23. Lv, H.; Dao, F.-Y.; Guan, Z.-X.; Zhang, D.; Tan, J.-X.; Zhang, Y.; Chen, W.; Lin, H. iDNA6mA-Rice: A Computational Tool for Detecting N6-Methyladenine Sites in Rice. *Front. Genet.* **2019**, *10*, 793. [CrossRef] [PubMed]
24. Chen, Y.; Ma, T.; Yang, X.; Wang, J.; Song, B.; Zeng, X. MUFFIN: Multi-scale feature fusion for drug–drug interaction prediction. *Bioinformatics* **2021**, *10*, 793.
25. Jin, S.; Zeng, X.; Xia, F.; Huang, W.; Liu, X. Application of deep learning methods in biological networks. *Brief. Bioinform.* **2021**, *22*, 1902–1917. [CrossRef]
26. Min, X.; Ye, C.; Liu, X.; Zeng, X. Predicting enhancer-promoter interactions by deep learning and matching heuristic. *Brief. Bioinform.* **2021**, *22*, bbaa254. [CrossRef] [PubMed]
27. Zeng, X.; Zhu, S.; Lu, W.; Liu, Z.; Huang, J.; Zhou, Y.; Fang, J.; Huang, Y.; Guo, H.; Li, L.; et al. Target identification among known drugs by deep learning from heterogeneous networks. *Chem. Sci.* **2020**, *11*, 1775–1797. [CrossRef]
28. Zeng, X.; Song, X.; Ma, T.; Pan, X.; Zhou, Y.; Hou, Y.; Zhang, Z.; Li, K.; Karypis, G.; Cheng, F. Repurpose open data to discover therapeutics for COVID-19 using deep learning. *J. Proteome Res.* **2020**, *19*, 4624–4636. [CrossRef]
29. Zhang, Y.; Yan, J.; Chen, S.; Gong, M.; Gao, D.; Zhu, M.; Gan, W. Review of the Applications of Deep Learning in Bioinformatics. *Curr. Bioinform.* **2020**, *15*, 898–911. [CrossRef]
30. Zeng, X.; Lin, Y.; He, Y.; Lv, L.; Min, X. Deep collaborative filtering for prediction of disease genes. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2020**, *17*, 1639–1647. [CrossRef]
31. Du, Z.; Xiao, X.; Uversky, V.N. Classification of Chromosomal DNA Sequences Using Hybrid Deep Learning Architectures. *Curr. Bioinform.* **2020**, *15*, 1130–1136. [CrossRef]
32. Tahir, M.; Tayara, H.; Chong, K.T. iDNA6mA (5-step rule): Identification of DNA N6-methyladenine sites in the rice genome by intelligent computational model via Chou's 5-step rule. *Chemom. Intell. Lab. Syst.* **2019**, *189*, 96–101. [CrossRef]
33. Yu, H.; Dai, Z. SNNRice6mA: A Deep Learning Method for Predicting DNA N6-Methyladenine Sites in Rice Genome. *Front. Genet.* **2019**, *10*, 1071. [CrossRef] [PubMed]
34. Li, Z.; Jiang, H.; Kong, L.; Chen, Y.; Lang, K.; Fan, X.; Zhang, L.; Pian, C. Deep6mA: A deep learning framework for exploring similar patterns in DNA N6-methyladenine sites across different species. *Plos Comput. Biol.* **2021**, *17*, e1008767. [CrossRef] [PubMed]
35. Ye, P.; Luan, Y.; Chen, K.; Liu, Y.; Xiao, C.; Xie, Z. MethSMRT: An integrative database for DNA N6-methyladenine and N4-methylcytosine generated by single-molecular real-time sequencing. *Nucleic Acids Res.* **2017**, *45*, D85–D89. [CrossRef]
36. Liu, Z.-Y.; Xing, J.-F.; Chen, W.; Luan, M.-W.; Xie, R.; Huang, J.; Xie, S.-Q.; Xiao, C.-L. MDR: An integrative DNA N6-methyladenine and N4-methylcytosine modification database for Rosaceae. *Hortic. Res.* **2019**, *6*, 1–7. [CrossRef] [PubMed]
37. Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [CrossRef] [PubMed]
38. Lv, H.; Dao, F.-Y.; Zhang, D. iDNA-MS: An Integrated Computational Tool for Detecting DNA Modification Sites in Multiple Genomes. *iScience* **2020**, *23*, 100991. [CrossRef] [PubMed]
39. Sharma, A.K.; Srivastava, R. Protein Secondary Structure Prediction Using Character bi-gram Embedding and Bi-LSTM. *Curr. Bioinform.* **2021**, *16*, 333–338. [CrossRef]
40. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
41. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.

42. Yang, Q.; Wu, J.; Zhao, J.; Xu, T.; Han, P.; Song, X. The Expression Profiles of lncRNAs and Their Regulatory Network During Smek1/2 Knockout Mouse Neural Stem Cells Differentiation. *Curr. Bioinform.* **2020**, *15*, 77–88. [CrossRef]
43. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. Shufflenet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018.
44. Geete, K.; Pandey, M. Robust Transcription Factor Binding Site Prediction Using Deep Neural Networks. *Curr. Bioinform.* **2020**, *15*, 1137–1152. [CrossRef]
45. Fu, K.; Fan, D.-P.; Ji, G.-P.; Zhao, Q. JLDCF: Joint Learning and Densely-Cooperative Fusion Framework for RGB-D Salient Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 3049–3059.
46. Fan, D.-P.; Wang, W.; Cheng, M.-M.; Shen, J. Shifting More Attention to Video Salient Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 8554–8564.
47. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Llion Jones, A.N.G.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
48. Bello, I.; Zoph, B.; Vaswani, A.; Shlens, J.; Le, Q.V. Attention Augmented Convolutional Networks. In Proceedings of the 2019 IEEE International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–3 November 2019; pp. 3286–3295.
49. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual Attention Network for Scene Segmentation. In Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 3146–3154.
50. Wang, X.; Girshick, R.B.; Gupta, A.; He, K. Non-Local Neural Networks. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 19–23 June 2018; pp. 7794–7803.
51. Ma, X.S.; Xi, B.H.; Zhang, Y.; Zhu, L.J.; Sui, X.; Tian, G.; Yang, J.L. A Machine Learning-based Diagnosis of Thyroid Cancer Using Thyroid Nodules Ultrasound Images. *Curr. Bioinform.* **2020**, *15*, 349–358. [CrossRef]
52. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 19–23 June 2018; pp. 7132–7141.
53. Dai, Y.; Gieseke, F.; Oehmcke, S.; Wu, Y.; Barnard, K. Attentional Features Fusion. In Proceedings of the 2021 Winter Conference on Applications of Computer Vision, Waikola, HI, USA, 5–9 January 2021.
54. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the 32nd International Conference on Machine Learning (ICML), Lille, France, 6–11 July 2015; pp. 448–456.
55. Naseer, S.; Hussain, W.; Khan, Y.D.; Rasool, N. NPalmitoylDeep-pseaac: A predictor of N-Palmitoylation Sites in Proteins Using Deep Representations of Proteins and PseAAC via Modified 5-Steps Rule. *Curr. Bioinform.* **2021**, *16*, 294–305. [CrossRef]
56. Greff, K.; Srivastava, R.K.; Koutnik, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A Search Space Odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *28*, 2222–2232. [CrossRef]
57. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
58. Nasir, M.A.; Nawaz, S.; Huang, J. A Mini-review of Computational Approaches to Predict Functions and Findings of Novel Micro Peptides. *Curr. Bioinform.* **2020**, *15*, 1027–1035. [CrossRef]
59. Wang, X.-F.; Gao, P.; Liu, Y.-F.; Li, H.-F.; Lu, F. Predicting Thermophilic Proteins by Machine Learning. *Curr. Bioinform.* **2020**, *15*, 493–502. [CrossRef]
60. Guo, Z.; Wang, P.; Liu, Z.; Zhao, Y. Discrimination of Thermophilic Proteins and Non-thermophilic Proteins Using Feature Dimension Reduction. *Front. Bioeng. Biotechnol.* **2020**, *8*, 584807. [CrossRef]
61. Zhao, X.; Wang, H.; Li, H.; Wu, Y.; Wang, G. Identifying Plant Pentatricopeptide Repeat Proteins Using a Variable Selection Method. *Front. Plant Sci.* **2021**, *12*, 298.
62. Tao, Z.; Li, Y.; Teng, Z.; Zhao, Y. A Method for Identifying Vesicle Transport Proteins Based on LibSVM and MRMD. *Comput. Math. Methods Med.* **2020**, *2020*, 8926750. [CrossRef]
63. Zhai, Y.; Chen, Y.; Teng, Z.; Zhao, Y. Identifying Antioxidant Proteins by Using Amino Acid Composition and Protein-Protein Interactions. *Front. Cell Dev. Biol.* **2020**, *8*, 591487. [CrossRef] [PubMed]
64. Wei, L.; Liao, M.; Gao, Y.; Ji, R.; He, Z.; Zou, Q. Improved and Promising Identification of Human MicroRNAs by Incorporating a High-Quality Negative Set. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2014**, *11*, 192–201. [CrossRef]
65. Wei, L.; Tang, J.; Zou, Q. Local-DPP: An improved DNA-binding protein prediction method by exploring local evolutionary information. *Inf. Sci.* **2017**, *384*, 135–144. [CrossRef]
66. Wei, L.; Wan, S.; Guo, J.; Wong, K.K.L. A novel hierarchical selective ensemble classifier with bioinformatics application. *Artif. Intell. Med.* **2017**, *83*, 82–90. [CrossRef]
67. Wei, L.; Xing, P.; Shi, G.; Ji, Z.; Zou, Q. Fast Prediction of Protein Methylation Sites Using a Sequence-Based Feature Selection Technique. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *16*, 1264–1273. [CrossRef]
68. Wei, L.; Xing, P.; Zeng, J.; Chen, J.; Su, R.; Guo, F. Improved prediction of protein-protein interactions using novel negative samples, features, and an ensemble classifier. *Artif. Intell. Med.* **2017**, *83*, 67–74. [CrossRef] [PubMed]
69. Wei, L.; Zhou, C.; Chen, H.; Song, J.; Su, R. ACPred-FL: A sequence-based predictor using effective feature representation to improve the prediction of anti-cancer peptides. *Bioinformatics* **2018**, *34*, 4007–4016. [CrossRef] [PubMed]

70. Hong, Z.; Zeng, X.; Wei, L.; Liu, X. Identifying enhancer-promoter interactions with neural network based on pre-trained DNA vectors and attention mechanism. *Bioinformatics* **2020**, *36*, 1037–1043. [CrossRef] [PubMed]
71. Jin, Q.; Meng, Z.; Tuan, D.P.; Chen, Q.; Wei, L.; Su, R. DUNet: A deformable network for retinal vessel segmentation. *Knowl.-Based Syst.* **2019**, *178*, 149–162. [CrossRef]
72. Manavalan, B.; Basith, S.; Shin, T.H.; Wei, L.; Lee, G. Meta-4mCpred: A Sequence-Based Meta-Predictor for Accurate DNA 4mC Site Prediction Using Effective Feature Representation. *Mol. Ther. Nucleic Acids* **2019**, *16*, 733–744. [CrossRef]
73. Manayalan, B.; Basith, S.; Shin, T.H.; Wei, L.; Lee, G. mAHTPred: A sequence-based meta-predictor for improving the prediction of anti-hypertensive peptides using effective feature representation. *Bioinformatics* **2019**, *35*, 2757–2765. [CrossRef]
74. Qiang, X.; Zhou, C.; Ye, X.; Du, P.-f.; Su, R.; Wei, L. CPPred-FL: A sequence-based predictor for large-scale identification of cell-penetrating peptides by feature representation learning. *Brief. Bioinform.* **2020**, *21*, 11–23. [CrossRef]
75. Su, R.; Hu, J.; Zou, Q.; Manavalan, B.; Wei, L. Empirical comparison and analysis of web-based cell-penetrating peptide prediction tools. *Brief. Bioinform.* **2020**, *21*, 408–420. [CrossRef]
76. Su, R.; Liu, X.; Wei, L. MinE-RFE: Determine the optimal subset from RFE by minimizing the subset-accuracy-defined energy. *Brief. Bioinform.* **2020**, *21*, 687–698. [CrossRef]
77. Su, R.; Liu, X.; Wei, L.; Zou, Q. Deep-Resp-Forest: A deep forest model to predict anti-cancer drug response. *Methods* **2019**, *166*, 91–102. [CrossRef] [PubMed]
78. Su, R.; Liu, X.; Xiao, G.; Wei, L. Meta-GDBP: A high-level stacked regression model to improve anticancer drug response prediction. *Brief. Bioinform.* **2020**, *21*, 996–1005. [CrossRef] [PubMed]
79. Su, R.; Wu, H.; Xu, B.; Liu, X.; Wei, L. Developing a Multi-Dose Computational Model for Drug-Induced Hepatotoxicity Prediction Based on Toxicogenomics Data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *16*, 1231–1239. [CrossRef] [PubMed]
80. Wei, L.; Chen, H.; Su, R. M6APred-EL: A Sequence-Based Predictor for Identifying N6-methyladenosine Sites Using Ensemble Learning. *Mol. Ther.-Nucleic Acids* **2018**, *12*, 635–644. [CrossRef] [PubMed]
81. Wei, L.; Ding, Y.; Su, R.; Tang, J.; Zou, Q. Prediction of human protein subcellular localization using deep learning. *J. Parallel Distrib. Comput.* **2018**, *117*, 212–217. [CrossRef]
82. Wei, L.; Hu, J.; Li, F.; Song, J.; Su, R.; Zou, Q. Comparative analysis and prediction of quorum-sensing peptides using feature representation learning and machine learning algorithms. *Brief. Bioinform.* **2020**, *21*, 106–119. [CrossRef] [PubMed]
83. Jin, S.; Zeng, X.; Fang, J.; Lin, J.; Chan, S.Y.; Erzurum, S.C.; Cheng, F. A network-based approach to uncover microRNA-mediated disease comorbidities and potential pathobiological implications. *NPJ Syst. Biol. Appl.* **2019**, *5*, 41. [CrossRef]
84. Wei, L.; Su, R.; Wang, B.; Li, X.; Zou, Q.; Gao, X. Integration of deep feature representations and handcrafted features to improve the prediction of N-6-methyladenosine sites. *Neurocomputing* **2019**, *324*, 3–9. [CrossRef]
85. Zou, Q.; Xing, P.; Wei, L.; Liu, B. Gene2vec: Gene subsequence embedding for prediction of mammalian N^6-methyladenosine sites from mRNA. *RNA* **2019**, *25*, 205–218. [CrossRef]
86. Dai, C.; Feng, P.; Cui, L.; Su, R.; Chen, W.; Wei, L. Iterative feature representation algorithm to improve the predictive performance of N^7-methylguanosine sites. *Brief. Bioinform.* **2020**, *22*, bbaa278. [CrossRef] [PubMed]
87. Wei, L.; He, W.; Malik, A.; Su, R.; Cui, L.; Manavalan, B. Computational prediction and interpretation of cell-specific replication origin sites from multiple eukaryotes by exploiting stacking framework. *Brief. Bioinform.* **2020**, *22*, bbaa275. [CrossRef] [PubMed]
88. Zhao, X.; Jiao, Q.; Li, H.; Wu, Y.; Wang, H.; Huang, S.; Wang, G. ECFS-DEA: An ensemble classifier-based feature selection for differential expression analysis on expression profiles. *BMC Bioinform.* **2020**, *21*, 43. [CrossRef]
89. Zeng, X.; Zhu, S.; Liu, X.; Zhou, Y.; Nussinov, R.; Cheng, F.J.B. deepDR: A network-based deep learning approach to in silico drug repositioning. *Bioinformatics* **2019**, *35*, 5191–5198. [CrossRef]
90. Fu, X.; Cai, L.; Zeng, X.; Zou, Q. StackCPPred: A stacking and pairwise energy content-based prediction of cell-penetrating peptides and their uptake efficiency. *Bioinformatics* **2020**, *36*, 3028–3034. [CrossRef]
91. Liu, Y.; Zhang, X.; Zou, Q.; Zeng, X.J.B. Minirmd: Accurate and fast duplicate removal tool for short reads via multiple minimizers. *Bioinformatics* **2020**, *37*, 1604–1606. [CrossRef]