



Research article

iPro2L-DG: Hybrid network based on improved densenet and global attention mechanism for identifying promoter sequences

Rufeng Lei^a, Jianhua Jia^{a,*}, Lulu Qin^a, Xin Wei^b^a School of Information Engineering, Jingdezhen Ceramic University, Jingdezhen, 333403, China^b Business School, Jiangxi Institute of Fashion Technology, Nanchang, 330044, China

ARTICLE INFO

Keywords:

Promoter
Promoter strength
DenseNet
Global attention mechanism
Encoding

ABSTRACT

The promoter is a key DNA sequence whose primary function is to control the initiation time and the degree of expression of gene transcription. Accurate identification of promoters is essential for understanding gene expression studies. Traditional sequencing techniques for identifying promoters are costly and time-consuming. Therefore, the development of computational methods to identify promoters has become critical. Since deep learning methods show great potential in identifying promoters, this study proposes a new promoter prediction model, called iPro2L-DG. The iPro2L-DG predictor, based on an improved Densely Connected Convolutional Network (DenseNet) and a Global Attention Mechanism (GAM), is constructed to achieve the prediction of promoters. The promoter sequences are combined feature encoding using C2 encoding and nucleotide chemical property (NCP) encoding. An improved DenseNet extracts advanced feature information from the combined feature encoding. GAM evaluates the importance of advanced feature information in terms of channel and spatial dimensions, and finally uses a Full Connect Neural Network (FNN) to derive prediction probabilities. The experimental results showed that the accuracy of iPro2L-DG in the first layer (promoter identification) was 94.10% with Matthews correlation coefficient value of 0.8833. In the second layer (promoter strength prediction), the accuracy was 89.42% with Matthews correlation coefficient value of 0.7915. The iPro2L-DG predictor significantly outperforms other existing predictors in promoter identification and promoter strength prediction. Therefore, our proposed model iPro2L-DG is the most advanced promoter prediction tool. The source code of the iPro2L-DG model can be found in <https://github.com/leirufeng/iPro2L-DG>.

1. Introduction

Genes [1] are the genetic code of life, which direct the synthesis of various proteins [2] through transcription [3] and translation [4] processes to maintain the normal activities of life. The promoter is a key DNA fragment for gene transcription, located in the upstream region of the gene usually [5]. The promoter is responsible for the initiation of gene transcription and under what conditions

Abbreviations: DenseNet, Densely Connected Convolutional Network; GAM, Global Attention Mechanism; NCP, nucleotide chemical property; FNN, Full Connect Neural Network; SVM, support vector machine; PseKNC, pseudo-K element nucleotide composition; CNN, Convolutional Neural Network; BiLSTM, Bi-directional Long Short-Term Memory; ResNet, Residual Neural Network; RNN, Recurrent Neural Network; CBAM, Convolutional Block Attention Module; Sn, sensitivity; Sp, specificity; Acc, accuracy; MCC, Matthew correlation coefficient.

* Corresponding author.

E-mail addresses: rufeng_lei@163.com (R. Lei), jjh163yx@163.com (J. Jia).

<https://doi.org/10.1016/j.heliyon.2024.e27364>

Received 26 October 2023; Received in revised form 24 February 2024; Accepted 28 February 2024

Available online 6 March 2024

2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

it begins [6]. The promoter sequence recognizes and amplifies RNA sequences by directing RNA polymerase [7] but does not transcribe the gene by itself. The promoter can also interact with other regulatory elements (e.g., enhancers) to regulate the rate of gene expression. Considering the level of transcription, promoters can be classified as strong promoters and weak promoters. A strong promoter enhances gene expression, while a weak promoter attenuates gene expression. An increasing number of medical studies have shown that mutations in promoter sequences may lead to dysregulation of gene expression, resulting in various malignant diseases [8–10]. Therefore, accurate identification of promoters is essential for understanding gene expression studies.

Previously, biologists may use traditional sequencing techniques to identify promoters, which were costly and time consuming. Therefore, it became critical for develop computational methods to identify promoters, which provided biologists with more tools and resources to help them better understand the role of promoters in gene expression. In recent years, an increasing number of computational methods have been developed to identify promoters. In 2005, Florquin et al. [11] proposed an approach to identify promoters based on the structural features of promoter sequences, which achieved good results. Subsequently, more and more predictors based on machine learning methods and complex feature engineering [12–14] have been widely used to identify promoters. For example, Liu et al. [15] developed a predictor called Ipro54-PSEKNC based on the support vector machine (SVM) algorithm and the sequence pseudo-K element nucleotide composition (PseKNC) method [16] to identify promoter sequences. Meanwhile, more excellent predictors based on SVM algorithms [17–19] were applied to identify promoters. These predictors can effectively identify promoters and non-promoters, however, have some limitations in predicting the class of promoters. In 2018, Liu et al. [20] developed a two-layer promoter predictor named iPromoter-2L. The first layer of this predictor is used to identify promoters, and the second layer is used to predict six classes of promoter. In the same year, Xiao et al. [21] developed a predictor named iPSW(2L)-pseKNC. They used multiple complex feature encoding methods to encode promoter sequences. The predictor achieved 83.13% and 71.20% accuracy in predicting promoters and their strengths, respectively. Meanwhile, some excellent and more sophisticated predictors [22–26] have been developed, allowing a deeper and more comprehensive study of promoter identification and function. However, the performance of machine learning models is very dependent on complex feature encodings, and too many complex combinations of feature encodings can lead to dimensionality disasters. Therefore, it is crucial to develop a series of predictors that require only simple feature information.

In recent years, deep learning has been widely used in the field of bioinformatics [27] and has achieved remarkable results. Deep learning can automatically extract meaningful feature information from promoter sequences without the need for complex feature engineering [28]. For example, in 2020, Habibi et al. [29] developed a predictor called iPSW (PseDNC-DL), which uses One-hot feature encoding way to encode promoter sequences and then uses a Convolutional Neural Network (CNN) [30] to extract local feature information. Combining the local features extracted by CNN with the features encoded by PseKNC and finally using a FNN for classification prediction, the predictor achieved an accuracy of 85.10% and 72.35% in predicting promoters and their strengths, respectively. In 2019, Nguyen et al. [31] used FastText N-gram model to represent DNA sequences as text vectors and this feature encoding method is more efficient. Therefore, the accuracy is slightly higher than iPSW (PseDNC-DL) predictor. Although the predictors developed by Habibi et al. and Nguyen et al. have high accuracy, they still use complex feature encoding to accomplish the prediction task. In 2022, Zhang et al. [32] developed a predictor called iPromoter-CLA, which uses only a One-hot encoding to encode promoter sequences and uses CNN and deep Capsule Networks [33] to learn local features. In addition, the predictor uses a Bi-directional Long Short-Term Memory (BiLSTM) [34] neural network and a Self-Attention mechanism to extract global features. The predictor achieves an accuracy of 86% and 73.46% in predicting promoters and their strengths, respectively. The models developed by Habibi et al. and Nguyen et al. only consider the local feature extraction. However, the iPromoter-CLA model developed by Zhang et al. considers a combination of local and global features and is superior in performance.

In 2022, Nguyen et al. [35] created four datasets of human and mouse TATA promoters. They proposed a computational model called iPromoter-Seqvec, the model used word embedding technique to encode promoter sequences and Bi-LSTM network for prediction with remarkable results. Meanwhile, Zhang et al. [36] developed a promoter predictor called iPro-WAEL using random forest and convolutional neural networks. This predictor combined word embedding techniques and multiple sophisticated artificial feature coding methods to express promoter sequences and achieved remarkable experimental results in promoter datasets of multiple species. It is worth pointing out that they are using relatively simple deep learning methods, which means that there is still much space for exploration of deep learning methods. Therefore, using a better deep learning network model can significantly improve the accuracy of promoter prediction.

In deep learning, using CNN to extract local features from original data is a common approach. In fact, increasing the number of Convolutional layers can extract more advanced local features, but may lead to problems such as gradient disappearance and network degradation, which affect the training efficiency and accuracy of the model. To solve these problems, the Residual Neural Network (ResNet) [37] overcomes this problem by using Residual blocks. ResNet replaces the Convolutional layer with a Residual layer, which allows the network to learn the residuals of the previous layer of data by short-circuiting the connections, thus avoiding the gradient disappearance problem. The DenseNet [38] is an improvement on ResNet. DenseNet uses a densely connected mechanism, which combines the outputs of all previous layers by channel dimension and feeds them into the next convolutional layer. This approach can extract higher-level feature information and the model performance and robustness are better than ResNet results. The use of Attention mechanism is also increasing. The Attention mechanism can be applied to different types of neural networks, such as Convolutional Neural Network and Recurrent Neural Network (RNN) [39]. It is a special neural network layer that improves the performance of a neural network by weighting the output feature maps. For example, the Convolutional Block Attention Module (CBAM) [40] can perform adaptive weighting of feature maps from both channel and spatial dimensions. In addition, the Self-Attention mechanism [41] can help neural network model to better capture the contextual information in the input feature maps, thus improving the performance and robustness of the model.

The current promoter predictors do have some drawbacks, such as poor performance and complex features. For this purpose, we developed a predictive model called iPro2L-DG. The structure of the iPro2L-DG prediction model is shown in Fig. 1(B). The iPro2L-DG model is based on DenseNet to extract high-level local feature information. Immediately after, a novel Convolutional Block Attention Module, called Global Attention Mechanism [42], is used to evaluate the high-level local feature information. Experimentally, iPro2L-DG predictor proved to be far superior to previous predictors in the two-layer promoter prediction results. In addition, iPro2L-DG predictor outperforms other predictors on the *E. coli* [43] promoter dataset. Therefore, iPro2L-DG predictor is currently the most representative choice for predicting promoters and their strengths.

2. Materials and methods

2.1. Benchmark dataset

When developing a predictor, it is crucial to establish a reliable benchmark dataset. In this study, we used the benchmark dataset created by Xiao et al. [21]. They extracted the promoter fragment from RegulonDB [44] and experimentally verified that a reasonable promoter sequence length is 81 bp. In addition, they used CD-HIT [45] to remove promoter sequences with more than 85% similarity. A higher CD-HIT threshold ensures the uniqueness of each sample and better tests the performance of the model. The benchmark dataset consists of 3382 promoter samples (including 1591 strong promoter samples and 1791 weak promoter samples) and 3382 non-promoter samples. The promoter sequences of different classes are independent of each other, ensuring the uniqueness of each

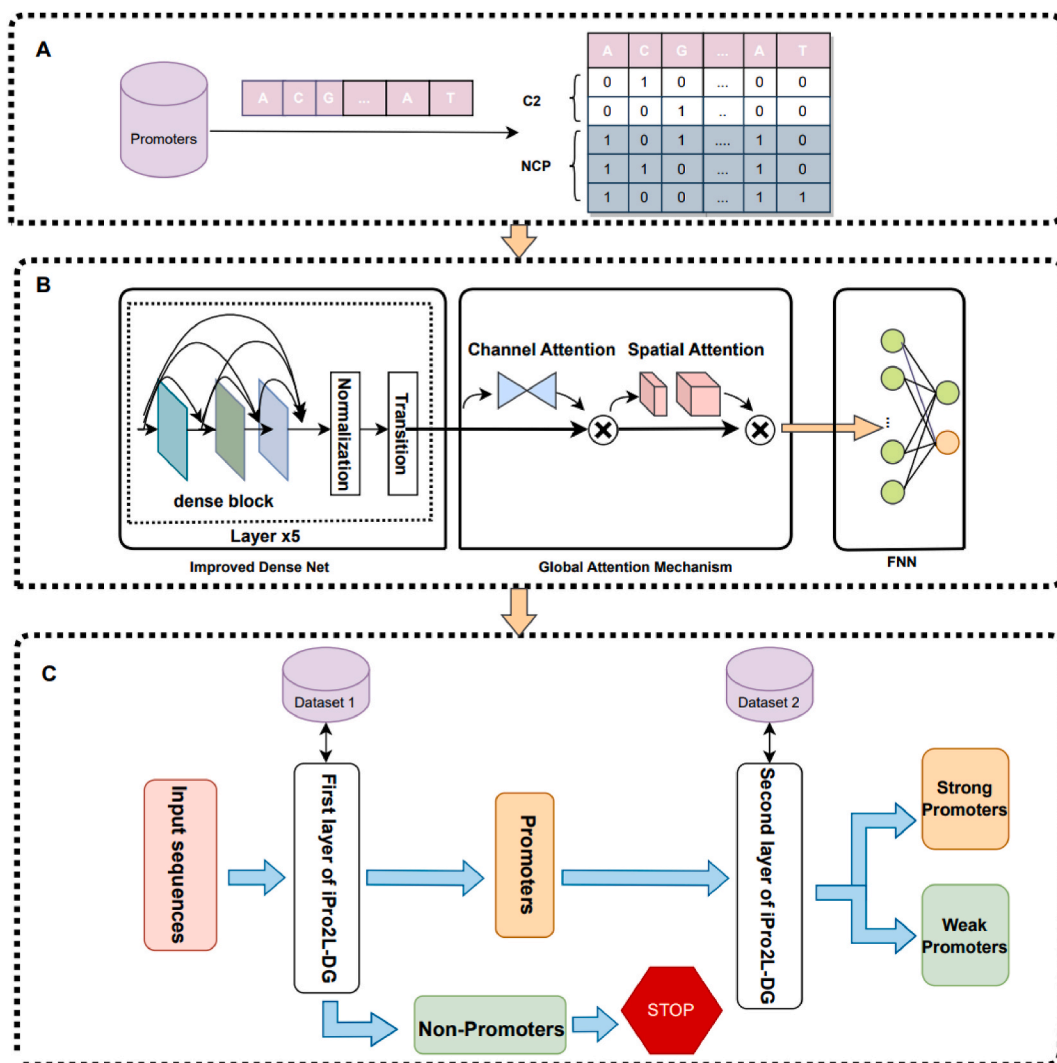


Fig. 1. An overview of iPro2L-DG model. (A) Feature encoding. We used two feature encoding methods, namely C2 and NCP. (B) iPro2L-DG model framework. The model framework consists of three modules. They are the improved DenseNet, the GAM attention mechanism and the FNN. (C) Two-layer classification framework.

sequence. In order to further validate the performance of our model, we performed ten cycles of five-fold cross-validation on the benchmark dataset. In this way, we were able to evaluate and validate the performance of the model more comprehensively. The benchmark dataset is shown in [Table 1](#).

2.2. Feature coding schemes

2.2.1. C2 encoding

One-hot encoding [46] is the most frequently used encoding method in bioinformatics and is favored by researchers for its simplicity and efficiency. One-hot encoding is the most primitive method of encoding gene sequences and protein sequences, which ensures that the alphabets of each gene and protein are coded independently. One-hot encoding encodes the nucleotides of a DNA sequence in "ACGT" order as a four-dimensional binary vector, where the nucleotides are encoded as 1 at their positions and 0 at other positions. For example, A is represented as (1, 0, 0, 0), C is represented as (0, 1, 0, 0), G is represented as (0, 0, 1, 0), and T is represented as (0, 0, 0, 1). However, One-hot encoding suffers from a sparsity drawback, which makes the model computation lack some robustness. In contrast, C2 encoding [47] is another commonly used binary vector encoding that retains the advantages of simplicity and efficiency of one-hot encoding. C2 encoding generates a two-dimensional vector and has less sparsity, so storage and computation are also better than One-hot encoding. C2 encoding represents each nucleotide in a DNA sequence as a two-dimensional vector, where the vector of A bases is (0,0), the vector of C bases is (1,1), the vector of G bases is (1,0), and the vector of T bases is (0,1). The C2 encoding is shown in [Fig. 1\(A\)](#).

2.2.2. NCP encoding

NCP encoding [48] is a feature encoding approach based on the chemical nature of nucleotides, which uses the chemical molecular structure between the four nucleotides for rational encoding. For example, in terms of ring structure, C and T each contain one ring, while A and G each contain two rings. In terms of hydrogen bond strength, A and T have two hydrogen bonds, while C and G have three hydrogen bonds. In terms of chemical structure, G and T belong to ketone groups, while A and C belong to amino groups. The specific chemical properties between the nucleotides are shown in [Table 2](#).

Based on the chemical properties in [Table 2](#), each nucleotide of the DNA sequence encoded by NCP can be represented as: A (1, 1, 1), C (0, 1, 0), G (1, 0, 0), and T (0, 0, 1). The NCP encoding is shown in [Fig. 1\(A\)](#).

2.3. Model construction

In this study, we constructed a deep learning network framework to automatically extract meaningful feature representations from promoter sequences, called iPro2L-DG. iPro2L-DG's framework is divided into two parts: (A) Feature encoding. We used two simple and efficient encoding methods, namely C2 and nucleotide chemical property. For each promoter sequence, we use these two encoding methods to generate a 2×81 feature encoding matrix and a 3×81 feature encoding matrix, respectively. These two feature encoding matrices are then combined to form a 5×81 feature encoding matrix. In this study, we treat the feature encoding matrix as a grayscale image and input it directly into the network framework of iPro2L-DG model. (B) iPro2L-DG model framework. The model framework consists of three modules, which are the improved DenseNet, GAM attention mechanism, and FNN. First, a 5×81 feature encoding matrix is directly input to the improved DenseNet (including five Dense blocks, Batch Normalization layer and Transition layer), and the improved DenseNet can extract high-level features. We used five layers of Dense blocks to enhance the feature extraction capability. Then, we use the GAM attention mechanism to evaluate the high-level feature information in terms of channel and spatial dimensions. Finally, we input the evaluated high-level feature maps into a FNN and obtain the predicted probabilities using the SoftMax activation function. (C) Two-layer classification framework. In this study, we divided promoters into two prediction tasks. The first task is to identify promoters and non-promoters, while the second task is to predict the strength of promoters. These two tasks are independent, and the results of the first task are not used as input for the second task. Nevertheless, However, during training, we use the same network model structure and hyperparameters for both prediction tasks. Detailed information of iPro2L-DG model is shown in [Fig. 1](#).

2.3.1. DenseNet

Compared to the traditional convolutional neural network structure, DenseNet utilizes a densely connected design. This innovative design allows DenseNet to generate more efficient feature representations and achieve higher performance with fewer parameters. As a result, DenseNet has become a very useful network structure in the field of deep learning. The network structure of DenseNet consists of one Convolutional layer, one or more Dense block layers, and one or more Transition layers. The feature encoding first passes through a Convolutional layer, then into a Dense block layer, and finally into a Transition layer. We improved the network structure of

Table 1
The specifics of the benchmark dataset.

Layers	Original dataset	Promoters	Non-promoters
First layer	Training dataset	3382	3382
	Original dataset	Strong enhancers	Weak enhancers
Second layer	Training dataset	1591	1791

Table 2
Nucleotide chemical property.

Chemical property	Category	Nucleotide
Ring structure	Purine	A, G
	Pyrimidine	C, T
Functional group	Amino	A, C
	Keto	G, T
Hydrogen bonding	Strong	C, G
	Weak	A, T

DenseNet by removing the first Convolutional layer and feeding the feature encoding directly to the Dense block layer. In addition, we added a Batch Normalization layer between the Dense block layer and the Transition layer. This improved DenseNet can extract higher-level feature information and enhance the robustness of the model.

2.3.1.1. Dense block. The Dense block [49] uses a dense connection mechanism to concatenate the outputs of all the previous Convolutional layers as inputs to the next layer, thus enabling reuse of features. The Dense block structure improves the efficiency of the model and enhances the expressiveness of the model. The structure of Dense block is shown in Fig. 2.

In a Dense block, the input of the X_i layer is related to the output of all previous layers, and X_i is represented as follows:

$$X_i = H_i ([X_0, X_1, X_2, \dots, X_{i-1}]) \quad (1)$$

where $[]$ denotes the "concat" operation; $H_i(\cdot)$ stands for a series of nonlinear transformation functions with a combination of Batch Normalization, ReLU activation function and Convolution (3×3).

2.3.1.2. Transition layer. The Transition layer [50] is the key layer connecting two adjacent Dense blocks in the DenseNet network, and its main role is to reduce the size of the Dense block output feature map. The Transition layer uses a 1×1 Convolution to reduce the number of channels in the final feature map, and a 2×2 Averaging Pooling layer to compress the size of the receptive field. This reduces the model complexity and improves the generalization ability of the network.

In the Transition layer, we insert a Batch Normalization layer in front of the 1×1 Convolution layer. The Batch Normalization layer can normalize the input data, which can effectively reduce the gradient disappearance and gradient explosion as well as avoid overfitting.

2.3.2. Global attention mechanism

Global Attention Mechanism(GAM) [42] is a new Convolutional Block Attention Module designed to evaluate the channel and spatial information of feature maps. As shown in Fig. 3, the GAM uses the same tandem structure of Channel Attention and Spatial Attention mechanisms as CBAM, but it redesigns the structure of Channel Attention and Spatial Attention mechanisms. A drawback of CBAM is the use of Pooling layers to compress dimensions and represent weight values with limited feature information, which leads to reduced information and ignores the interaction of channel with spatial. For example, the Pooling layer of CBAM picks out the largest or average pixel values in the image in the channel and spatial dimensions, which are trained by the network as weight values.

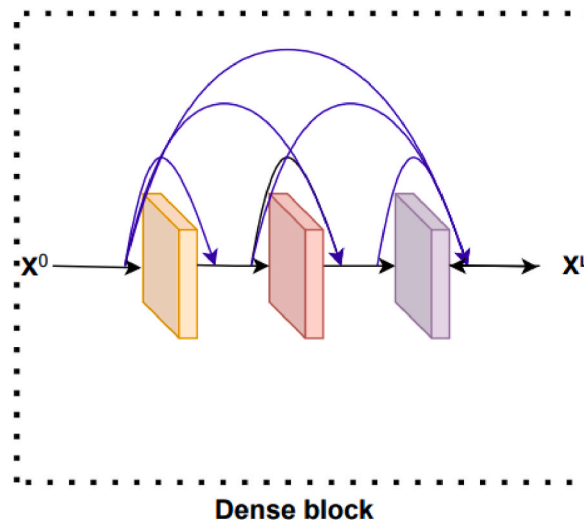


Fig. 2. Structure of the Dense block.

However, this approach ignores the true weights of other pixel values. To solve these problems, GAM abandons the Pooling layer and adopts a special computational structure that can directly calculate all the weight values of the three dimensions of the feature map. As a result, GAM is able to focus more comprehensively on useful feature information in the feature map and suppress useless feature information in the feature map.

The Channel Attention module changes the position of the three dimensions (channel, spatial width, and spatial height) of the original feature map and compresses the spatial feature information for each channel into vectors. Next, all spatial feature information in each channel is trained using two layers of FNN to amplify the channel and spatial dimensional dependencies. By training of the FNN, the weight values of all the spatial feature information in each channel were successfully obtained. These weight values are compressed by the Sigmoid activation function and reshaped into a feature map with the same size of the original feature map. Finally, the original feature map and these weights are multiplied to generate the Channel Attention weighted feature map. The Channel Attention module is shown in Fig. 4(a).

In the spatial attention module, firstly, the original feature map with channel number C is compressed to C/r by using $C/r \times 7 \times 7$ convolution kernel to convolve the feature map, which can make the spatial information more concentrated. Then, the output features are reshaped into the shape of the original feature map using a 7×7 convolution kernel, and these weights are compressed into weight values using the Sigmoid activation function. The convolution kernel can automatically compress the spatial location information of all channels for the purpose of focusing spatial information. All channel feature information at each spatial location of the feature map can be automatically trained using a two-layer convolution kernel. Finally, the original feature map and these weights are multiplied to generate the Spatial Attention weighted feature map. The Spatial Attention module is shown in Fig. 4(b).

2.4. Performance evaluation

Assessing the performance of a model requires the use of scientific evaluation criteria. These evaluation criteria include sensitivity (S_n), specificity (S_p), accuracy (Acc), and Matthew correlation coefficient (MCC) [51]. These metrics allow a comprehensive assessment of model performance and can guide model optimization and improvement. The calculation formula of the evaluation index is shown as.

$$\left\{ \begin{array}{l} S_p = \frac{TN}{TN + FP} \\ S_n = \frac{TP}{TP + FN} \\ Acc = \frac{TP + TN}{TP + TN + FP + FN} \\ MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \end{array} \right. \quad (2)$$

The confusion matrix is one of the important metrics to assess the performance of the model, which includes four metrics: true positive (TP), true negative (TN), false positive (FP) and false negative (FN) [52]. These metrics represent the accuracy and recall of the model for sample classification, respectively. To describe the model performance more accurately, we also used AUC metrics [53] for evaluation. These evaluation metrics have higher values indicating better model performance.

3. Results and discussion

3.1. Results of model

In this study, we divided the promoters into two prediction tasks. The two-layer classification framework is shown in Fig. 1(C). To further validate the performance and stability of the iPro2L-DG predictor, we executed 10 five-fold cross-validations and listed the

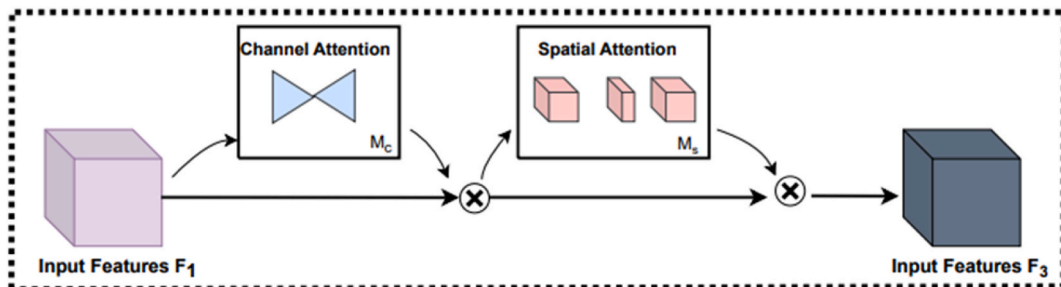


Fig. 3. The overview of a GAM.

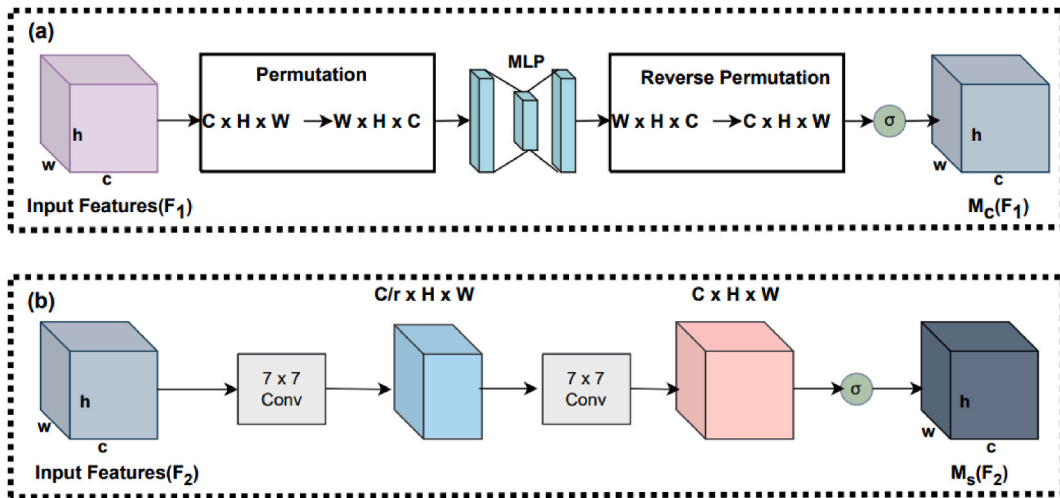


Fig. 4. Submodule structure of the GAM attention mechanism (a) Structure of the channel attention submodule of the GAM. (b) Structure of the spatial attention submodule of the GAM.

experimental results in Fig. 5. Finally, we take the average of the evaluation metrics from 10 5-fold cross-validations to ensure the reliability of the model. In the first layer task, the average results of the 10 experiments for Sn, Sp, Acc, and MCC are 93.89%, 94.71%, 94.29% and 88.65%, respectively. In the second layer task, the average results of the 10 experiments for Sn, Sp, Acc and MCC are 87.11%, 92.11%, 89.67% and 79.47%, respectively. In addition, the experimental results of AUC metrics for the first layer task (promoter identification) and the second layer task (promoter strength prediction) were 97.74% and 93.53%, respectively. The results show that our proposed iPro2L-DG predictor has good robustness.

3.2. Compare different combinations of dense blocks and convolutional layers

In this study, DenseNet is the main high-level feature extraction network. The iPro2L-DG model performance depends on the depth of the DenseNet network. However, overly complex DenseNet networks are prone to overfitting, while overly simple networks are prone to underfitting. The complexity of the DenseNet network is mainly in the number of Dense blocks and the number of Convolutional layers in the Dense blocks. To explore different combinations of Dense blocks and Convolutional layers, we tried various combinations in the two layers tasks. We use accuracy (Acc) as the evaluation metric because it is more reflective of the model's performance on a balanced sample. According to Fig. 6, the accuracy of the first and second layers is highest when the number of Dense blocks is five and the number of Convolutional layers is three. Therefore, we decided to use DenseNet, where the number of Dense blocks is 5 and the number of Convolutional layers for each Dense block is 3.

3.3. Comparison of different coding methods

In this research, we have been working on developing predictors that require only simple feature encoding, since deep learning can automatically extract meaningful feature information from simple feature encoding. The simpler the feature encoding used, which is more indicative of the superiority of the model. We used six simple combinations of features: One-hot encoding, C2 encoding, NCP encoding, Word2Vec encoding [34], One-hot + NCP encoding, C2 + NCP + ND encoding, and C2 + NCP encoding. We put these seven

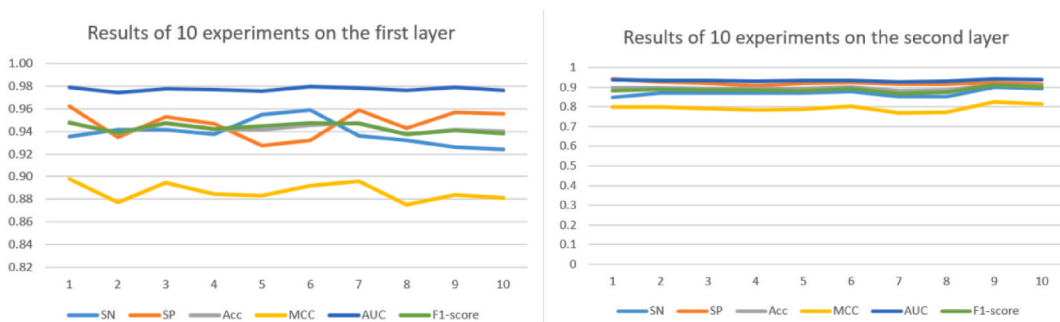


Fig. 5. Degree of volatility of the evaluation metrics for 10 cycles of 5-fold cross-validation.

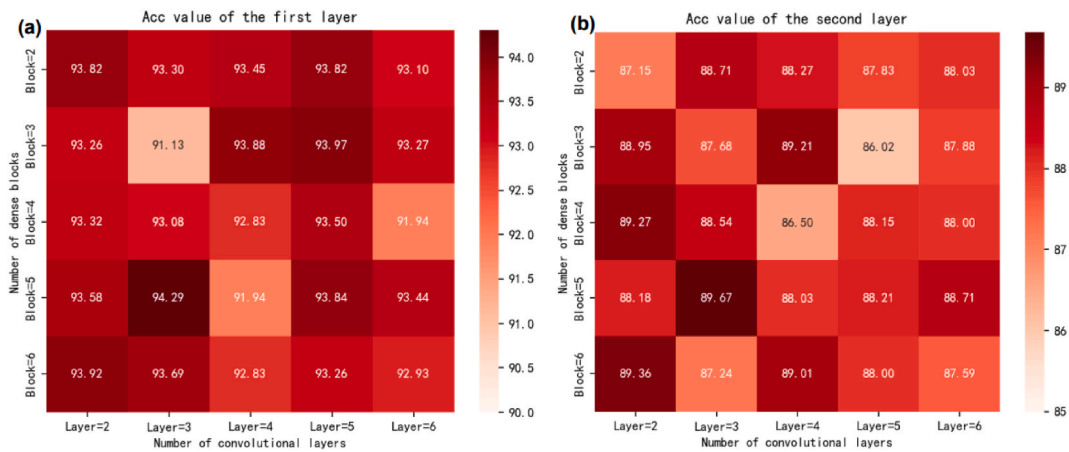


Fig. 6. Comparative results for different combinations of dense blocks and convolutional layers. (a) Acc values (%) for promoter identification. (b) Acc values (%) for promoter strength prediction.

encoding combinations into the two layers network models, respectively. According to Figs. 7 and 8, In the experimental results of the first layer task, C2 and one-hot encoding are comparable, but in the second layer C2 encoding is slightly better. From a computational point of view, compared to the one-hot coding method, the feature encoding size of the C2 coding method is reduced by half, so the corresponding number of parameters is also reduced by half. This shows that C2 encoding not only has the advantages of simplicity and efficiency of one-hot encoding, but also outperforms one-hot encoding in terms of storage and computation. Word2Vec encoding is a word vector coding method obtained through network training, which is based on contextual context information. Compared to other encoding methods, complex Word2Vec encoding is the least effective. This shows that the model is able to automatically learn abstract and complex feature representations from simple sequence coding without relying on complex sequence coding methods. The hybrid encoding has better classification performance than the single encoding approach, indicating that better performance is achieved through the combined encoding approach. In the two layers prediction tasks, C2+NCP coding outperforms other coding combinations. Therefore, we identified C2+NCP encoding as the preferred scheme for feature engineering.

3.4. Comparison of different model frameworks

In this study, we investigated eight network framework strategies and tested these network frameworks in the layer 1 and layer 2 prediction tasks, respectively. According to the experimental results shown in Figs. 9 and 10, all evaluation metrics of the DenseNet network is significantly higher than those of ResNet, indicating that the DenseNet network has better prediction performance. In addition, we found in our experimental results that excellent results can be achieved using a combination of DenseNet networks and arbitrary Attention mechanisms. Meanwhile, we find that the evaluation metrics of CBAM and GAM are better than single Attention mechanism, indicating a significant performance improvement with simultaneous use of Channel Attention and Spatial Attention. Comparing the experimental results of Pri DenseNet + GAM and DenseNet + GAM, we can see that the improved DenseNet network better results than the primitive DenseNet network, and thus the improved DenseNet network extracts more advanced feature information. From the experimental results, we can see that the performance of GAM attention mechanism is better than CBAM attention

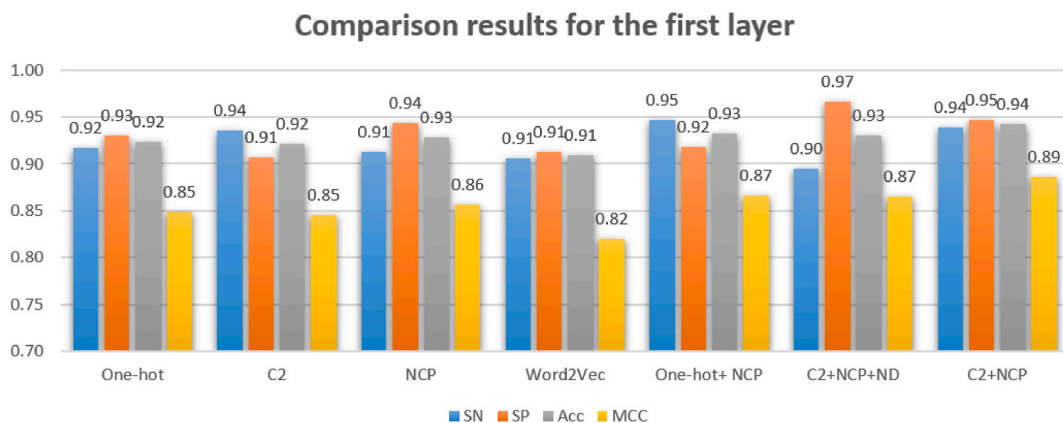


Fig. 7. Comparison results of different encoding schemes in the first layer (promoter recognition).

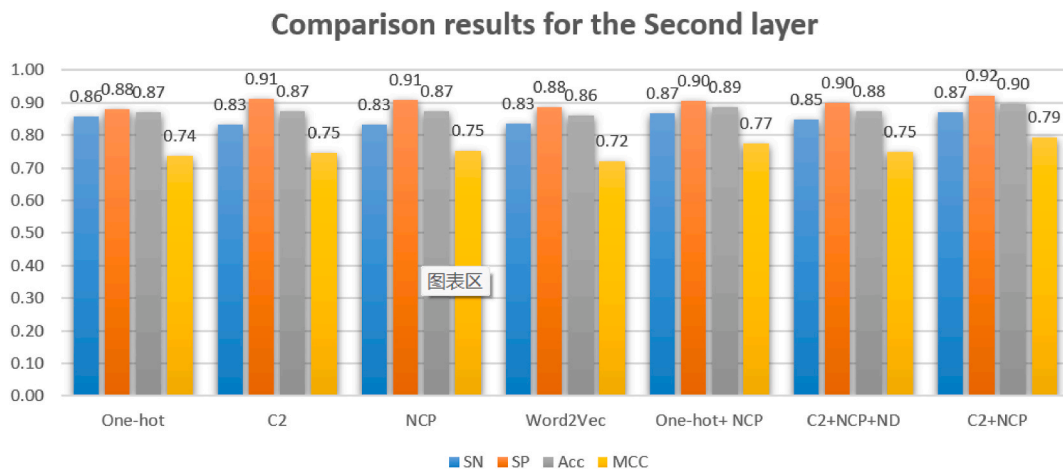


Fig. 8. Comparison results of different encoding schemes in the second layer (promoter strength prediction).

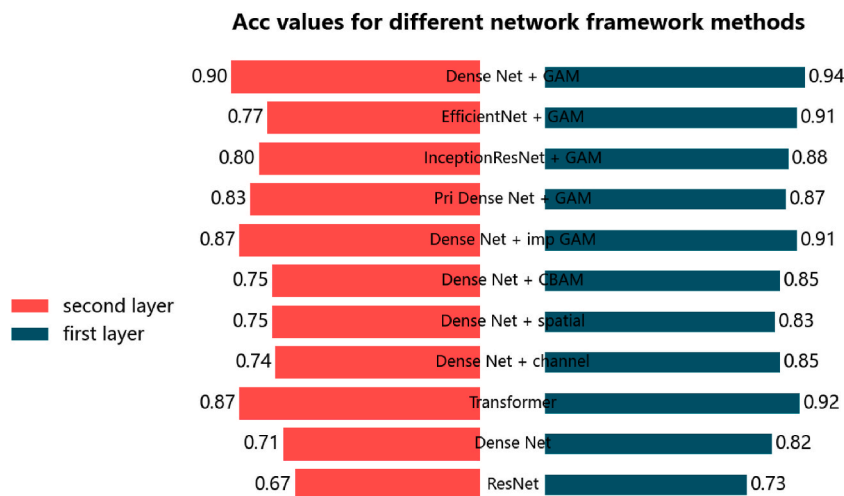


Fig. 9. Comparison of accuracy (Acc) values for different network frameworks. The results for layer 1 (promoter identification) are shown in green and the results for layer 2 (promoter strength prediction) are shown in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

mechanism. We have tried to change the serial sequential approach in the GAM attention mechanism to the parallel approach [54], but the results are not satisfactory. By comparing the experimental results of Transformer, InceptionResNet + GAM and EfficientNet + GAM, we can observe that these complex advanced neural network models do not perform well in the experiments. This may be due to the small amount of data, which does not allow the performance of these large models to be fully utilized. Comprehensively taking the experimental results into consideration, we finally chose the network framework of DenseNet + GAM.

3.5. Comparison of iPro2L-DG with existing predictors

In recent years, an increasing number of computational methods have been applied to the field of promoter research. To evaluate the performance of the iPro2L-DG predictor, we compared it with four other excellent predictors. The comparison results obtained using the five-fold cross-validation method on the same dataset is shown in Table 3. The iPro2L-DG predictor far outperforms other predictors in four evaluation metrics (Sn, Sp, Acc and MCC), which indicates that the iPro2L-DG predictor has good generalization ability. Analyzing and comparing the other predictors, iPSW(2L)-PseKNC is a predictor based on machine learning methods, however, it has relatively low predictive performance. iPSW(PseDNC-DL) and FastText N-grams are predictors based on CNN networks. Traditional CNNs may have problems of insufficient feature extraction, which may be the reason for the low performance of the models. In addition, iPro2L-CLA is a predictor based on capsule networks. Although capsule networks can improve some of the drawbacks of CNN networks, there are some limitations to their performance improvement. DenseNet networks can extract more advanced local feature information by iteratively utilizing previous features, which is the key reason for the excellent performance of

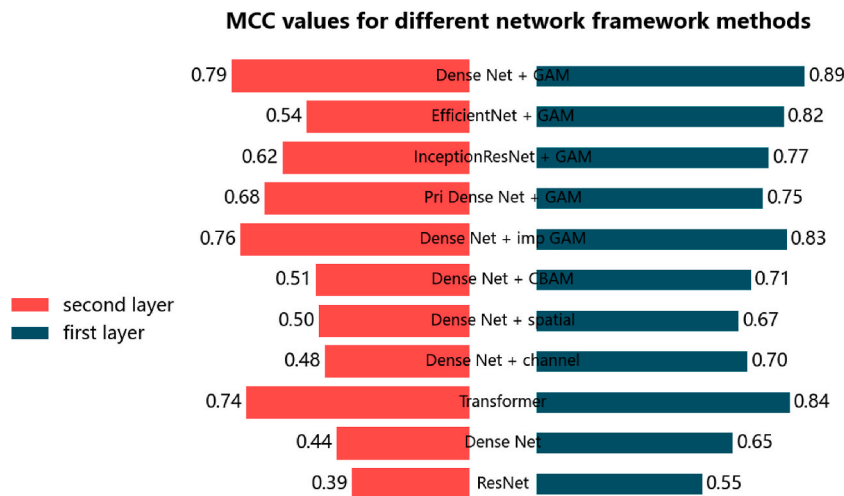


Fig. 10. Comparison of MCC values for different network frameworks. The results for layer 1 (promoter identification) are shown in green and the results for layer 2 (promoter strength prediction) are shown in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Comparison with existing predictors on the same benchmark dataset.

Layer	Predictors	Sn	Sp	Acc	MCC	AUC	F1-score
First layer	iPSW(2L)-PseKNC	0.8137	0.8489	0.8313	0.6630	-	-
	iPSW(PseDNC-DL)	0.8334	0.8683	0.8510	0.7024	-	-
	FastText N-grams	0.8276	0.8805	0.8541	0.7090	-	-
	iPro2L-CLA	0.8687	0.8513	0.8600	0.7211	-	-
	iPro2L-DG	0.9389 ± 0.0106	0.9471 ± 0.0117	0.9429 ± 0.0038	0.8865 ± 0.0076	0.9774 ± 0.0017	0.9432 ± 0.0040
Second layer	iPSW(2L)-PseKNC	0.6223	0.7917	0.7120	0.4213	-	-
	iPSW(PseDNC-DL)	0.6581	0.7816	0.7235	0.4440	-	-
	FastText N-grams	0.6940	0.7640	0.7310	0.4600	-	-
	iPro2L-CLA	0.7763	0.6878	0.7346	0.4700	-	-
	iPro2L-DG	0.8711 ± 0.0168	0.9211 ± 0.0088	0.8967 ± 0.0086	0.7947 ± 0.0165	0.9353 ± 0.0042	0.8880 ± 0.0117

our model.

In the first layer of the task (promoter identification), the iPro2L-DG predictor had a Sn improvement of 6.94%–12.44%, Sp improvement of 6.37%–9.53%, Acc and MCC improvement of 8.10%–10.97% and 16.22%–22.03%, respectively. In the second layer task (promoter strength prediction), the iPro2L-DG predictor has a Sn improvement of 6.94%–12.44%, Sp improvement of 6.37%–9.53%, and Acc and MCC improvement of 8.10%–10.97% and 16.22%–22.03%, respectively. The improvement of the iPro2L-DG predictor in the promoter two-layer prediction task is tremendous. In particular, the prediction accuracy of the iPro2L-DG predictor remains high in the second layer task when the amount of data on promoter strength is small. This indicates that the iPro2L-DG predictor performs more consistently and superiorly when processing small amounts of data. Unfortunately, the predictive performance of the iPro2L-DG predictor would have been better if the promoter dataset was larger. Thus, the iPro2L-DG predictor becomes a

Table 4
Comparison with existing predictors on the benchmark dataset of Liu et al.

Method	SN	SP	Acc	MCC
PCSF [13]	0.7890	0.7070	0.7480	0.4980
vw Z-curve [14]	0.7776	0.8280	0.8028	0.6100
Stability [12]	0.7660	0.7950	0.7800	0.5620
iPro54 [15]	0.7776	0.8315	0.8045	0.6100
iPromoter-2L [20]	0.7920	0.8416	0.8168	0.6343
iPSW(2L)-PseKNC [21]	0.8378	0.8434	0.8406	0.6811
iPro2L-CLA [32]	0.8627	0.8480	0.8553	0.7114
iPro2L- DG	0.9286	0.9490	0.9386	0.8779

most advanced prediction tool for promoters and their strengths.

3.6. Performance of iPro2L-DG on other datasets

In addition, we tested the iPro2L-DG predictor again on the *E. coli* promoter dataset proposed by Liu et al. [20], since the promoter dataset proposed by Xiao et al. [21] did not have an independent test set. As shown in Table 4, our predictor also outperforms the other seven predictors on the *E. coli* promoter dataset. The improvement range of Sn was 6.59%–16.26%, that of Sp was 10.10%–24.20%, and that of ACC and MCC was 8.33%–19.06% and 32.15%–37.02%, respectively. This indicates that the iPro2L-DG predictor has good robustness and generalization in predicting other promoter tasks and is the most representative predictor for predicting promoter tasks.

4. Conclusions

This study proposes a new promoter prediction model called iPro2L-DG. The model is based on an improved DenseNet and GAM attention mechanism implementation. The experimental results showed that the iPro2L-DG model achieved a significant MCC value of 0.8833 for promoter identification and an MCC value of 0.7915 for promoter strength prediction in the five-fold cross-validation. Compared with other existing predictors, the iPro2L-DG predictor performs well on all evaluation metrics (Sn, Sp, Acc, and MCC). Meanwhile, the iPro2L-DG predictor outperformed the other seven predictors in all metrics on the *E. coli* promoter dataset. This indicates that the iPro2L-DG predictor has excellent generalization ability and expressive power. We are committed to exploring and innovating deep learning methods in combination with the promoter investigation to advance the research in this field. In the future, the iPro2L-DG predictor will be applied to prediction tasks not only in the promoter field but also in other bioinformatics fields, which will be able to provide researchers with more convenient tools.

Of course, our proposed iPro2L-DG model, although excellent in some respects, still has some shortcomings. Due to the small size of the promoter dataset, the performance of the model had a lot of space for improvement. In addition, we did not employ a data enhancement strategy to increase the sample data. The predictive power of the iPro2L-DG model is limited when dealing with imbalanced data. These are problems will be considered and addressed in our future study. However, as promoter research continues, we believe that these problems will be gradually resolved. Meanwhile, we expect that the iPro2L-DG model will continue to be optimized and improved to provide better prediction tools for promoter studies. Moreover, we expect that more advanced deep learning methods will emerge in the future to bring more possibilities for promoter and their strength prediction.

Funding

This work was partially supported by the National Natural Science Foundation of China (Nos. 61761023, 62162032, and 31760315), the Natural Science Foundation of Jiangxi Province, China (Nos. 20202BABL202004 and 20202BAB202007), the Scientific Research Plan of the Department of Education of Jiangxi Province, China (GJJ190695 and GJJ2202814). These funders had no role in the study design, data collection and analysis, decision to publish or preparation of manuscript.

Data availability statement

The data set and source code used in this study can be easily derived from <https://github.com/leirufeng/iPro2L-DG>.

CRedit authorship contribution statement

Rufeng Lei: Writing – original draft, Software, Methodology, Data curation, Conceptualization. **Jianhua Jia:** Funding acquisition, Conceptualization. **Lulu Qin:** Writing – review & editing. **Xin Wei:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgements

The authors are grateful for the constructive comments and suggestions made by the reviewers.

References

- [1] Q. Dai, C. Bao, Y. Hai, S. Ma, T. Zhou, C. Wang, Y. Wang, W. Huo, X. Liu, Y. Yao, et al., MTGpick allows robust identification of genomic islands from a single genome, *Briefings Bioinf.* 19 (2016) 361–373, [10.1093/bib/bbw118](https://doi.org/10.1093/bib/bbw118) %J Briefings in Bioinformatics.
- [2] C. Buccitelli, M. Selbach, mRNAs, proteins and the emerging principles of gene expression control, *Nat. Rev. Genet.* 21 (2020) 630–644. [https://10.1038/s41576-020-0258-4](https://doi.org/10.1038/s41576-020-0258-4).
- [3] G.A. Maston, S.K. Evans, M.R. Green, Transcriptional regulatory elements in the human genome, *Annu. Rev. Genom. Hum. Genet.* 7 (2006) 29–59. [https://10.1146/annurev.genom.7.080505.115623](https://doi.org/10.1146/annurev.genom.7.080505.115623).

- [4] Z. Xiao, Q. Zou, Y. Liu, X. Yang, Genome-wide assessment of differential translations with ribosome profiling data, *Nat. Commun.* 7 (2016) 11194. <https://doi.org/10.1038/ncomms11194>.
- [5] N. Cvetesic, B. Lenhard, Core promoters across the genome, *Nat. Biotechnol.* 35 (2017) 123–124. <https://doi.org/10.1038/nbt.3788>.
- [6] H. Lin, Z.Y. Liang, H. Tang, W. Chen, Identifying Sigma70 promoters with novel pseudo nucleotide composition, *IEEE ACM Trans. Comput. Biol. Bioinf* 16 (2019) 1316–1321. <https://doi.org/10.1109/tcbb.2017.2666141>.
- [7] R.M. Saecker, J. Chen, C.E. Chiu, B. Malone, J. Sotiris, M. Ebrahim, L.Y. Yen, E.T. Eng, S.A. Darst, Structural origins of Escherichia coli RNA polymerase open promoter complex stability, *Proc. Natl. Acad. Sci. U.S.A.* 118 (2021). <https://doi.org/10.1073/pnas.2112877118>.
- [8] R. Coles, R. Caswell, D.C. Rubinsztein, Functional analysis of the Huntington's disease (HD) gene promoter, *Hum. Mol. Genet.* 7 (1998) 791–800. <https://doi.org/10.1093/hmg/7.5.791>.
- [9] S. Döhr, A. Klingshoff, H. Maier, M. Hrabé de Angelis, T. Werner, R. Schneider, Linking disease-associated genes to regulatory networks via promoter organization, *Nucleic Acids Res.* 33 (2005) 864–872. <https://doi.org/10.1093/nar/gki230>.
- [10] C. Ionescu-Tîrgoviște, P.A. Gagniuc, C. Guja, Structural properties of gene promoters highlight more than two phenotypes of diabetes, *PLoS One* 10 (2015) e0137950. <https://doi.org/10.1371/journal.pone.0137950>.
- [11] K. Florquin, Y. Saeys, S. Degroove, P. Rouzé, Y. Van de Peer, Large-scale structural analysis of the core promoter in mammalian and plant genomes, *Nucleic Acids Res.* 33 (2005) 4255–4264. <https://doi.org/10.1093/nar/gki737>.
- [12] S. de Avila e Silva, F. Forte, I.T.S.S. Andrighetti T, G.J. L. G, A.P. Longaray Delamare, S. Echeverrigaray, DNA duplex stability as discriminative characteristic for Escherichia coli $\sigma(54)$ - and $\sigma(28)$ - dependent promoter sequences, *Biologicals* 42 (2014) 22–28. <https://doi.org/10.1016/j.biologicals.2013.10.001>.
- [13] Q.Z. Li, H. Lin, The recognition and prediction of sigma70 promoters in Escherichia coli K-12, *J. Theor. Biol.* 242 (2006) 135–141. <https://doi.org/10.1016/j.jtbi.2006.02.007>.
- [14] K. Song, Recognition of prokaryotic promoters based on a novel variable-window Z-curve method, *Nucleic Acids Res.* 40 (2012) 963–971. <https://doi.org/10.1093/nar/gkr795>.
- [15] H. Lin, E.Z. Deng, H. Ding, W. Chen, K.C. Chou, iPro54-PseKNC: a sequence-based predictor for identifying sigma-54 promoters in prokaryote with pseudo k-tuple nucleotide composition, *Nucleic Acids Res.* 42 (2014) 12961–12972. <https://doi.org/10.1093/nar/gku1019>.
- [16] W. Chen, T.Y. Lei, D.C. Jin, H. Lin, K.C. Chou, PseKNC: a flexible web server for generating pseudo K-tuple nucleotide composition, *Anal. Biochem.* 456 (2014) 53–60. <https://doi.org/10.1016/j.ab.2014.04.001>.
- [17] R.V. Coelho, E.S.S. de Avila, S. Echeverrigaray, A.P.L. Delamare, Bacillus subtilis promoter sequences data set for promoter prediction in Gram-positive bacteria, *Data Brief* 19 (2018) 264–270. <https://doi.org/10.1016/j.dib.2018.05.025>.
- [18] H.Y. Lai, Z.Y. Zhang, Z.D. Su, W. Su, H. Ding, W. Chen, H. Lin, iProEP: a computational predictor for predicting promoter, *Mol. Ther. Nucleic Acids* 17 (2019) 337–346. <https://doi.org/10.1016/j.omtn.2019.05.028>.
- [19] M.S. Rahman, U. Aktar, M.R. Jani, S. Shatabda, iPromoter-FSEn: identification of bacterial $\sigma(70)$ promoter sequences using feature subspace based ensemble classifier, *Genomics* 111 (2019) 1160–1166. <https://doi.org/10.1016/j.ygeno.2018.07.011>.
- [20] B. Liu, F. Yang, D.S. Huang, K.C. Chou, iPromoter-2L: a two-layer predictor for identifying promoters and their types by multi-window-based PseKNC, *Bioinformatics* 34 (2018) 33–40. <https://doi.org/10.1093/bioinformatics/btx579>.
- [21] X. Xiao, Z.C. Xu, W.R. Qiu, P. Wang, H.T. Ge, K.C. Chou, iPSW(2L)-PseKNC: a two-layer predictor for identifying promoters and their strength by hybrid features via pseudo K-tuple nucleotide composition, *Genomics* 111 (2019) 1785–1793. <https://doi.org/10.1016/j.ygeno.2018.12.001>.
- [22] S. Burden, Y.X. Lin, R. Zhang, Improving promoter prediction for the NNPP2.2 algorithm: a case study using Escherichia coli DNA sequences, *Bioinformatics* 21 (2005) 601–607. <https://doi.org/10.1093/bioinformatics/bti047>.
- [23] V. Rangannan, M. Bansal, Relative stability of DNA as a generic criterion for promoter prediction: whole genome annotation of microbial genomes with varying nucleotide base composition, *Mol. Biosyst.* 5 (2009) 1758–1769. <https://doi.org/10.1039/b906535k>.
- [24] M.G. Reese, Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome, *Comput. Chem.* 26 (2001) 51–56. [https://doi.org/10.1016/s0097-8485\(01\)00099-7](https://doi.org/10.1016/s0097-8485(01)00099-7).
- [25] I.A. Shahmuradov, R. Mohamad Razali, S. Bougouffa, A. Radovanovic, V.B. Bajic, bTSSfinder: a novel tool for the prediction of promoters in cyanobacteria and Escherichia coli, *Bioinformatics* 33 (2017) 334–340. <https://doi.org/10.1093/bioinformatics/btw629>.
- [26] M. Zhang, F. Li, T.T. Marquez-Lago, A. Leier, C. Fan, C.K. Kwok, K.C. Chou, J. Song, C. Jia, MULTIPLY: a novel multi-layer predictor for discovering general and specific types of promoters, *Bioinformatics* 35 (2019) 2957–2965. <https://doi.org/10.1093/bioinformatics/btz016>.
- [27] E. Seo, Y.N. Choi, Y.R. Shin, D. Kim, J.W. Lee, Design of synthetic promoters for cyanobacteria with generative deep-learning model, *Nucleic Acids Res.* (2023). <https://doi.org/10.1093/nar/gkad451>.
- [28] D.K. Hawley, W.R. McClure, Compilation and analysis of Escherichia coli promoter DNA sequences, *Nucleic Acids Res.* 11 (1983) 2237–2255. <https://doi.org/10.1093/nar/11.8.2237>.
- [29] H. Tayara, M. Tahir, K.T. Chong, Identification of prokaryotic promoters and their strength by integrating heterogeneous features, *Genomics* 112 (2020) 1396–1403. <https://doi.org/10.1016/j.ygeno.2019.08.009>.
- [30] S. Patiyal, A. Dhall, K. Bajaj, H. Sahu, G.P.S. Raghava, Prediction of RNA-interacting residues in a protein using CNN and evolutionary profile, *Briefings Bioinf.* 24 (2023). <https://doi.org/10.1093/bib/bbab538>.
- [31] N.Q.K. Le, E.K.Y. Yapp, N. Nagasundaram, H.Y. Yeh, Classifying promoters by interpreting the hidden information of DNA sequences via deep learning and combination of continuous FastText N-grams, *Front. Bioeng. Biotechnol.* 7 (2019) 305. <https://doi.org/10.3389/fbioe.2019.00305>.
- [32] Z.M. Zhang, J.P. Zhao, P.J. Wei, C.H. Zheng, iPromoter-CLA: identifying promoters and their strength by deep capsule networks with bidirectional long short-term memory, *Comput. Methods Progr. Biomed.* 226 (2022) 107087. <https://doi.org/10.1016/j.cmpb.2022.107087>.
- [33] J. Khanal, H. Tayara, Q. Zou, K. To Chong, DeepCap-Kcr: Accurate Identification and Investigation of Protein Lysine Crotonylation Sites Based on Capsule Network, *Briefings in Bioinformatics*, vol. 23, 2021. <https://doi.org/10.1093/bib/bbab492>.
- [34] Z.-Y. Zhang, L. Ning, X. Ye, Y.-H. Yang, Y. Futamura, T. Sakurai, H. Lin, iLoc-miRNA: Extracellular/intracellular miRNA Prediction Using Deep BiLSTM with Attention Mechanism, *Briefings in Bioinformatics*, vol. 23, 2022. <https://doi.org/10.1093/bib/bbac395>.
- [35] P. Zhang, H. Zhang, H. Wu, iPro-WAEL: a comprehensive and robust framework for identifying promoters in multiple species, *Nucleic Acids Res.* 50 (2022) 10278–10289. <https://doi.org/10.1093/nar/gkac824>.
- [36] T.H. Nguyen-Vo, Q.H. Trinh, L. Nguyen, P.U. Nguyen-Hoang, S. Rahardja, B.P. Nguyen, iPromoter-Seqvec: identifying promoters using bidirectional long short-term memory and sequence-embedded features, *BMC Genom.* 23 (2022) 681. <https://doi.org/10.1186/s12864-022-08829-6>.
- [37] X. Li, P. Han, W. Chen, C. Gao, S. Wang, T. Song, M. Niu, A. Rodriguez-Patón, MARPPi: boosting prediction of protein-protein interactions with multi-scale architecture residual network, *Brief Bioinform.* vol. 24 (2023). <https://doi.org/10.1093/bib/bbac524>.
- [38] P.J. Wei, Z.Z. Pang, L.J. Jiang, D.Y. Tan, Y.S. Su, C.H. Zheng, Promoter prediction in nanochloropsis based on densely connected convolutional neural networks, *Methods* 204 (2022) 38–46. <https://doi.org/10.1016/j.ymeth.2022.03.017>.
- [39] S. Min, B. Lee, S. Yoon, Deep learning in bioinformatics, *Briefings Bioinf.* 18 (2016) 851–869. [10.1093/bib/bbw068](https://doi.org/10.1093/bib/bbw068) %J *Briefings in Bioinformatics*.
- [40] Y. Zhang, Z. Wang, Y. Zeng, Y. Liu, S. Xiong, M. Wang, J. Zhou, Q. Zou, A novel convolution attention model for predicting transcription factor binding sites by combination of sequence and shape, *Briefings Bioinf.* 23 (2022). <https://doi.org/10.1093/bib/bbab525>.
- [41] L. Zeng, Y. Liu, Z.G. Yu, Y. Liu, iEnhancer-DLRA: identification of enhancers and their strengths by a self-attention fusion strategy for local and global features, *Brief Funct Genomics* 21 (2022) 399–407. <https://doi.org/10.1093/bfpg/elac023>.
- [42] Y. Huang, J. Yan, Y. Zhang, W. Ye, C. Zhang, P. Gao, X. Lv, Automatic segmentation of cotton roots in high-resolution minirhizotron images based on improved OCRNet, *Front. Plant Sci.* 14 (2023) 1147034. <https://doi.org/10.3389/fpls.2023.1147034>.
- [43] I.D. Olaru, B. Walther, F. Schaumburg, Zoonotic sources and the spread of antimicrobial resistance from the perspective of low and middle-income countries, *Infect Dis Poverty* 12 (2023) 59. <https://doi.org/10.1186/s40249-023-01113-z>.

- [44] S. Gama-Castro, H. Salgado, A. Santos-Zavaleta, D. Ledezma-Tejeda, L. Muñoz-Rascado, J.S. García-Sotelo, K. Alquicira-Hernández, I. Martínez-Flores, L. Pannier, J.A. Castro-Mondragón, et al., RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond, *Nucleic Acids Res.* 44 (2016) D133–D143. <https://10.1093/nar/gkv1156>.
- [45] Y. Huang, B. Niu, Y. Gao, L. Fu, W. Li, CD-HIT Suite: a web server for clustering and comparing biological sequences, *Bioinformatics* 26 (2010) 680–682. <https://10.1093/bioinformatics/btq003>.
- [46] Z. Abbas, H. Tayara, K.T. Chong, ENet-6mA: identification of 6mA modification sites in plant genomes using ElasticNet and neural networks, *Int. J. Mol. Sci.* 23 (2022). <https://10.3390/ijms23158314>.
- [47] S.B. Arniker, H.K. Kwan, N.-F. Law, D.P.-K. Lun, Promoter prediction using DNA numerical representation and neural network: case study with three organisms, in: *2011 Annual IEEE India Conference*, 2011, pp. 1–4.
- [48] T. Zhang, Q. Tang, F. Nie, Q. Zhao, W. Chen, DeepLncPro: an interpretable convolutional neural network model for identifying long non-coding RNA promoters, *Briefings Bioinf.* 23 (2022). <https://10.1093/bib/bbac447>.
- [49] T. Yang, T. Wu, L. Li, C. Zhu, SUD-GAN: deep convolution generative adversarial network combined with short connection and dense block for retinal vessel segmentation, *J. Digit. Imag.* 33 (2020) 946–957. <https://10.1007/s10278-020-00339-9>.
- [50] J. Jia, Z. Wei, X. Cao, EMDL-ac4C: identifying N4-acetylcytidine based on ensemble two-branch residual connection DenseNet and attention, *Front. Genet.* 14 (2023) 1232038. <https://10.3389/fgene.2023.1232038>.
- [51] J. Jia, M. Sun, G. Wu, W. Qiu, DeepDN_iGlu: prediction of lysine glutarylation sites based on attention residual learning method and DenseNet, *Math. Biosci. Eng.* 20 (2023) 2815–2830. <https://10.3934/mbe.2023132>.
- [52] K. Niu, X. Luo, S. Zhang, Z. Teng, T. Zhang, Y. Zhao, iEnhancer-EBLSTM: identifying enhancers and strengths by ensembles of bidirectional long short-term memory, *Front. Genet.* 12 (2021) 665498. <https://10.3389/fgene.2021.665498>.
- [53] J. Jia, G. Wu, M. Li, W. Qiu, pSuc-EDBAM: predicting lysine succinylation sites in proteins based on ensemble dense blocks and an attention module, *BMC Bioinf.* 23 (2022) 450. <https://10.1186/s12859-022-05001-5>.
- [54] J. Jia, R. Lei, L. Qin, G. Wu, X. Wei, iEnhancer-DCSV: predicting enhancers and their strength based on DenseNet and improved convolutional block attention module, *Front. Genet.* 14 (2023) 1132018. <https://10.3389/fgene.2023.1132018>.