IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 22, 2019, accepted December 5, 2019, date of publication December 10, 2019,
date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2958618

iIM-CNN: Intelligent Identifier of 6mA Sites on Different Species by Using Convolution Neural Network

ABDUL WAHAB[1], SYED DANISH ALI[1,2], HILAL TAYARA[1], AND KIL TO CHONG[3]
[1]Department of Electronics and Information Engineering, Chonbuk National University, Jeonju 54896, South Korea
[2]Department of Electrical Engineering, The University of Azad Jammu and Kashmir, Muzaffarabad 13100, Pakistan
[3]Advanced Electronics and Information Research Center, Chonbuk National University, Jeonju 54896, South Korea

Corresponding authors: Hilal Tayara (hilaltayara@jbnu.ac.kr) and Kil To Chong (kitchong@jbnu.ac.kr)

This work was supported in part by the Brain Research Program of the National Research Foundation (NRF) funded by the Korean Government (MSIT) under Grant NRF-2017M3C7A1044815.

ABSTRACT DNA N6-methyladenine (6mA) is related to a vast range of biological progress like transcription, replication, and repair of DNA. The precise discrimination of the 6mA sites plays a vital role in the understanding of its biological functions. Even though biochemical experiments produced positive results, they were inefficient in terms of cost and time. Therefore, to facilitate the identification of 6mA sites it is important to develop a robust computational model. In this regard, we develop a deep learning-based computational model named as iIM-CNN for the identification of N6-methyladenine sites from DNA sequences. The iIM-CNN is capable of extracting important features using a convolution neural network (CNN). The proposed model achieves the Mathew correlation coefficient (MCC) of 0.651, 0.752 and 0.941 for cross-species, Rice, and M. musculus genome respectively. The comparison of the outcomes depicts that the proposed model outperforms the existing computational tools for the prediction of the 6mA sites. Finally, a publically available user-friendly web server is available at https://home.jbnu.ac.kr/NSCL/iIMCNN.htm

INDEX TERMS DNA N6-methyladenine, sequence analysis, cross-species, deep learning, convolution neural network.

I. INTRODUCTION

DNA N6-methyladenine (6mA) is non-canonical methylation on adenine by attaching a methyl group to the sixth location of the Adenine purine ring [1]. It has been spotted in three kingdoms of life namely bacteria, archaea, and eukaryotes out of six kingdoms [2]. Current research has established that 6mA modification is intimately related to several biological processes, for instance, DNA replication [3], transcription [4] and repair [5]. The uneven dissemination of 6mA positions through the genome suggests that, for consideration of its biological functions in more detail, it is essential to indicate its location in the genome.

Diverse experimental techniques have been proposed for the identification of 6mA modifications. The first method was proposed about the combination of ultraviolet absorption spectra, electrophoretic mobility, and paper chromatographic movement. But comparatively, this technique was not effectual due to which it cannot be utilized for the detection of 6mA modifications in animals [6]. Another method, restriction enzyme was introduced for the discovery of 6mA modification which was only able to identify the modified Adenosines that exist in the target motifs [7].

Also, various experimental procedures have been carried out for the detection of 6mA sites in both eukaryotes and prokaryotes, for instance, sequencing of methylated DNA immunoprecipitation [8], capillary electrophoresis with laser-induced fluorescence [9], single-molecule real-time sequencing [10], ultra-high-performance liquid chromatography and mass spectrometry [11]. After the experimental procedure such as 6mA immunoprecipitation sequencing (6mA-IP-Seq), 84% of 6mA modification were found in Chlamydomonas genes [12]. The identification of 6mA modification in vertebrates consists of the human, mouse, and

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaochun Cheng.

Xenopusdot using HPLC, blots, and subsequently, sequencing of methylated DNA Immunoprecipitation (MeDIP-seq) [13]. Through the single-molecule real-time SMRT sequencing, it was detected that 2.8% of initial-diverged fungi were belonging to all adenines of methylated sites [14]. Also, in the rice genome, it was found that 0.2% of adenines were 6mA methylated as a result of using 6mA immunoprecipitation, mass spectrometry, and SMRT [15]. Even though experimental techniques are time-consuming and costly to perform genome-wide detection for 6mA sites these methods show important roles by providing significant indications in promoting the progress of this valuable area.

To optimize the time and cost, many computational models were proposed by researchers for the identification of 6mA modifications. Recently, the prediction methods such as iDNA6mA-PseKNC [16] and csDMA [17] are freely available for the identification of DNA 6mA modification in cross-species, rice genome, and Mus musculus genome. They were based on machine learning algorithms. These computational models require the field knowledge for manual construction of the features which are built in such a way that should provide the information of a pattern in a sequence to be taken into consideration. While predictors based on deep learning can consequently extract the most significant features of 6mA from input sequences which enables us to design robust models using raw sequences and without using handy crafted features. Deep learning-based algorithms achieved flourishing outcomes in the field of image recognition [18]–[20], natural language processing [21] and speech recognition [22]. Presently, deep learning-based bioinformatics predictors such as iDeepS [23], branch point selection [24], Deep Splicing Code [25], iRNA-PseKNC(2methyl) prediction model [26], and DeePromoter [27] have been proposed.

In this regard, we propose a novel deep learning-based model to classify the DNA N6-methyladenine sites using convolutional neural networks (CNN). CNN is capable of extracting the most important features from the data to make an intelligent predictive model. We used the grid search for the hyperparameter selection method to choose the optimal parameters. The evaluation of the model's performance was based on the k-fold cross-validation method by using the value of k = 5. The achievements of the proposed model outperformed the state-of-the-art machine learning models [17]. A user-friendly web server was made freely accessible at https://home.jbnu.ac.kr/NSCL/iIMCNN.htm

II. MATERIALS AND METHODS

A. BENCHMARK DATASETS

The benchmark dataset of DNA 6mA for this study was downloaded from (https://github.com/liuze-nwafu/csDMA). It consists of benchmark datasets of rice genome [28], M. musculus genome [16], and using these two benchmark datasets a cross-species dataset [17] was created. For the reduction of sequence redundancy in the dataset, the threshold value was set to 0.8 using CD-HIT-EST software [29].

TABLE 1. Summary of dataset.

Species	Dataset	Number of Samples
Cross-species	Positive	2768
	Negative	2716
Rice	Positive	880
	Negative	880
M. musculus	Positive	1934
	Negative	1934

The benchmark dataset of rice genome [28] consists of 1760 sequences from which 880 sequences are regarded as the positive samples and 880 sequences are regarded as negative samples. The benchmark dataset of M. musculus genome [16] has 3868 sequences from which 1934 are the positive samples and 1934 are negative samples. The dataset of cross-species [17] has 5484 sequences from which 2768 sequences are positive samples and 2716 sequences are negative samples. In all of the benchmark datasets, the length of each sequence is 41nt. Details of the datasets are shown in Table 1.

In reference to literature, the benchmark dataset mostly consists of a training dataset and a testing dataset. The training dataset is typically for the learning of the model while testing data is used as a trial of the model. On the other hand, as stated in Chou and Shen [30], for a high-quality benchmark dataset, it would be appropriate if the model is tested by a jackknife or a subsampling (K-fold cross-validation) test [31], as a result we obtain a mixture of different independent test datasets.

B. THE PROPOSED MODEL

We proposed an efficient deep learning model based on a convolution neural network that identifies the DNA 6mA modification of different species. It is capable of learning the most significant features from raw sequences automatically while training the model. The input of iIM-CNN has a DNA sequence $Q = \{Q1, Q2, \ldots, Qn\}$ where $n = 41$ and it should be in vector form. For vectorization of input sequences one-hot encoding was used in which each nucleotides A, C, G, T of a sequence was represented as (1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1) respectively as a four-channel input vector. During the learning process of the model, different hyper-parameters were used which were tuned by the grid search algorithm. The tuned parameters consist of convolution layers, filters, filter size, pool-size, stride length, and dropout values. The scales of these hyper-parameters are enumerated in Table 2.

The most efficient parameters were chosen on the base of least validation loss that avoids overfitting and underfitting problems. We implemented a classical CNN model, which consists of two 1-D convolution layers with the number of filters are 32, having filter size 5 with the stride of 1.

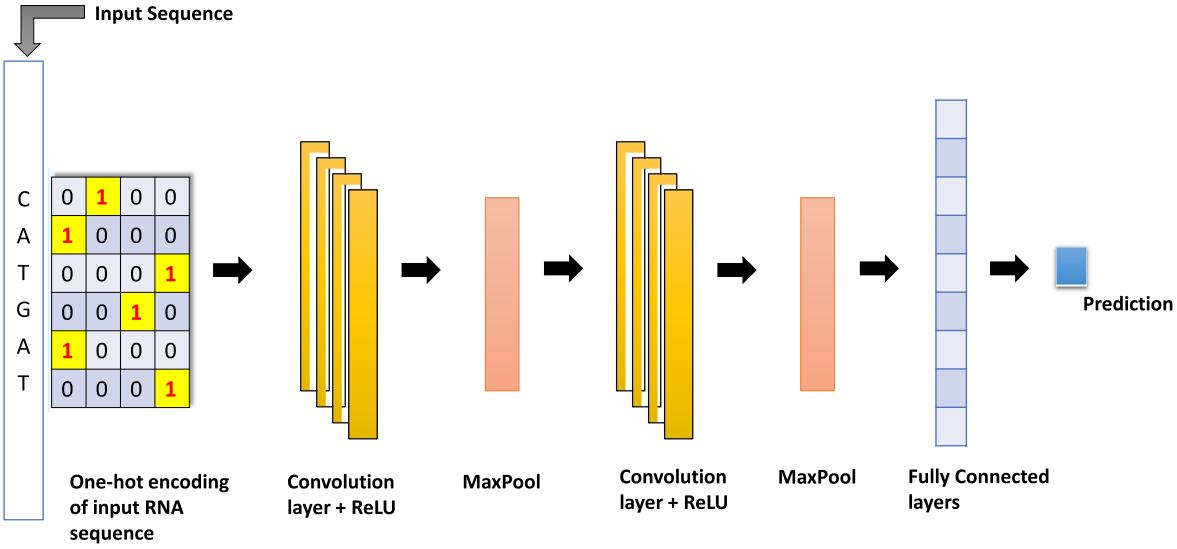


FIGURE 1. The architecture of the proposed model.

TABLE 2. Hypter-parameter preferences.

Parameters	Range
Convolution layers	[1,2,3]
Filters in each convolution Layer	[8,16,32,64,128]
Size of the filters	[3,5,7,9,12]
Maxpooling pool size	[1,2,3,4]
Maxpooling stride length	[1,2,3,4]
Dropout values	[0,2,0.3,0.4,0.5,0.6]

Each filter is responsible for finding the patterns in sequences to differentiate the positive and the negative samples of 6mA sites. Each convolution layer used the ReLU activation function as an argument of the layer which is known as a rectified linear unit. Max pooling layer with pool-size of 2 and stride of 2, is used in both layers to decrease the dimensionality of the features from the previous outputs, and a dropout layer with the probability of 0.4 after each convolution layer, which turn off the effect of some hidden neurons by setting the output of those neurons to zero at training. At the training time, the Maxpooling and dropout are regularization techniques used to avoid overfitting. Thus, some transitional features are eliminated which prevents overfitting and escalates the consistency of the model. We used flatten function to assimilate the intermediary features and to feed a fully connected layer, along with a sigmoid activation function for the prediction of 6mA sites. Sigmoid function squeezes the output results between 0 and 1, which assigns the probability values to the input data. If the probability is more than 0.5 then the model recognizes the sequence as a positive 6mA site, on the other hand, if it is less than 0.5 then the model distinguishes it as a negative 6mA site.

TABLE 3. The architecture of the proposed model.

Layer	Output shape
Input	(41,4)
Conv1D(32,5,1)	(41,32)
ReLU	(41,32)
MaxPool1D (2,2)	(20,32)
Dropout(0.4)	(20,32)
Conv1D(32,5,1)	(20,32)
ReLU	(20,32)
MaxPool1D(2,2)	(10,32)
Dropout(0.4)	(10,32)
Flatten	320
Dense(1)	1
Sigmoid	1

TABLE 4. The performance results of iIM-CNN.

Species	Sn	Sp	ACC	MCC	auROC	F1
Cross-species	0.869	0.780	0.824	0.651	0.892	0.831
Rice	0.841	0.914	0.875	0.752	0.934	0.870
M. musculus	0.938	1	0.969	0.941	0.971	0.968

Table 3, depicts the operations of the proposed model, where Conv1D (t, s, d) operator is a one-dimensional convolution layer where t is the number of filters, s is the sizes of the filters and d is the stride. The Maxpooling1D (p, e) operator is a max-pooling layer where p is the pool-size and e is the stride. The Dropout (r) represents a dropout layer

TABLE 5. Result comparison of state-of-the-art predictors with our model (iIM-CNN) by using three benchmark datasets.

Models	Species	Sn	Sp	ACC	MCC	auROC	F1
Cross-species	iIM-CNN	0.869	0.780	0.824	0.651	0.892	0.831
	csDMA	0.863	0.735	0.799	0.603	0.879	0.811
	iDNA6mA-PseKNC	0.762	0.769	0.765	0.531	0.844	0.764
Rice	iIM-CNN	0.841	0.914	0.875	0.752	0.934	0.870
	csDMA	0.842	0.880	0.861	0.723	0.923	0.858
	iDNA6mA-PseKNC	0.569	0.721	0.641	0.394	0.896	0.543
M. musculus	iIM-CNN	0.938	1	0.969	0.941	0.971	0.968
	csDMA	0.932	1	0.966	0.935	0.974	0.965
	iDNA6mA-PseKNC	0.869	1	0.935	0.877	0.974	0.930

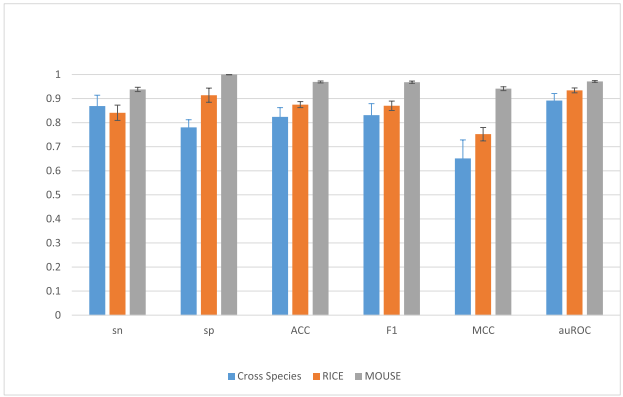


FIGURE 2. Graphical illustration of iIM-CNN results on different species with standard error.

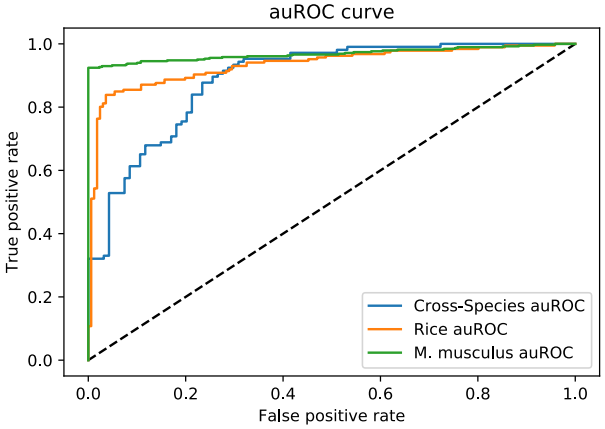


FIGURE 3. The auROC of different datasets in the proposed model.

with a probability of r. Dense (n) is a fully connected layer with n nodes. Finally, the Sigmoid () function is a nonlinear activation function that squeezes the output in the range [0-1] and represents the probability of having 6mA and non-6mA sites. Figure 1 shows the detailed architecture of the proposed model.

The iIM-CNN was implemented by using Keras framework [32]. In the proposed model Adam optimizer was utilized for optimization of the predictor with the learning rate of 0.006. The batch size was set to 32 and binary cross-entropy was used as a loss function [33]. The number of epochs was set to 80 and the early stopping method was used on validation loss, which means that training iterations will halt when the model performance stops improving the validation loss. Patience level for early stopping was set to 11, it means that after 11 iterations it would stop training if there would be no improvement in validation loss.

C. PERFORMANCE EVALUATION

For evaluating the performance of the proposed model, we used a 5-fold cross-validation method. Each subset was iteratively chosen as a test set in a separate cross-validation fold, while the remaining four subsets were used for the training of the model. The average results of the five trials were finally used as the performance estimation of the proposed model.

Several recent publications have used the following standard measures [34]–[39]. The definition of these measures,Accuracy (ACC),Sensitivity (SN), Specificity (SP), Matthews Correlation Coefficient (MCC), and F1 score, are described as:

$$ACC = 1 - \frac{M_-^+ + M_+^-}{M^+ + M^-} \tag{1}$$

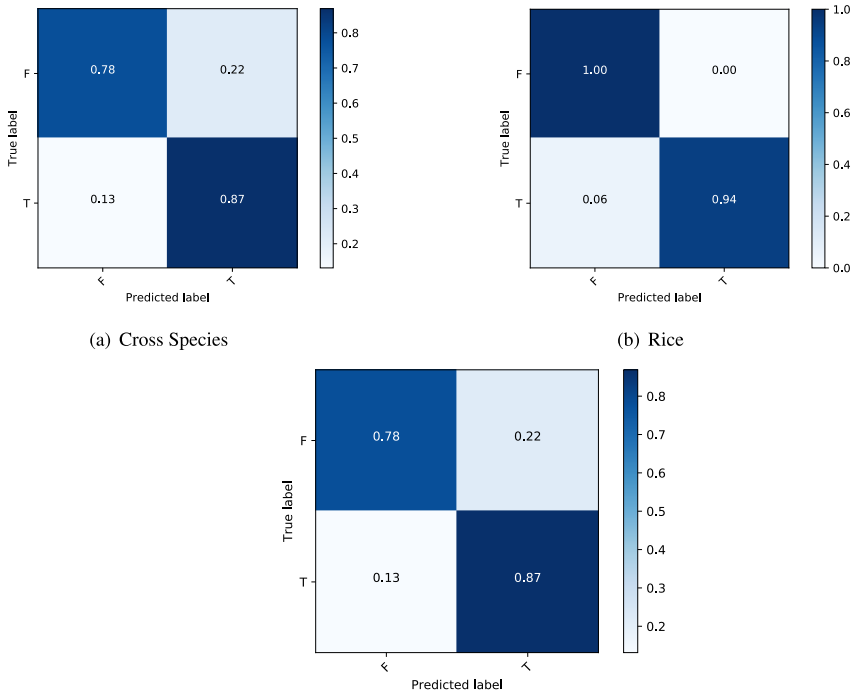
$$SN = 1 - \frac{M_-^+}{M^+} \tag{2}$$

$$SP = 1 - \frac{M_+^-}{M^-} \tag{3}$$

$$MCC = \frac{1 - \frac{M_-^+ + M_+^-}{M^+ + M^-}}{\sqrt{(1 + \frac{M_+^- - M_-^+}{M^+})(1 + \frac{M_-^+ - M_+^-}{M^-})}} \tag{4}$$

$$F1 = 2\frac{M^+ - M_-^+}{2M^+ - M_-^+ + N_+^-} \tag{5}$$

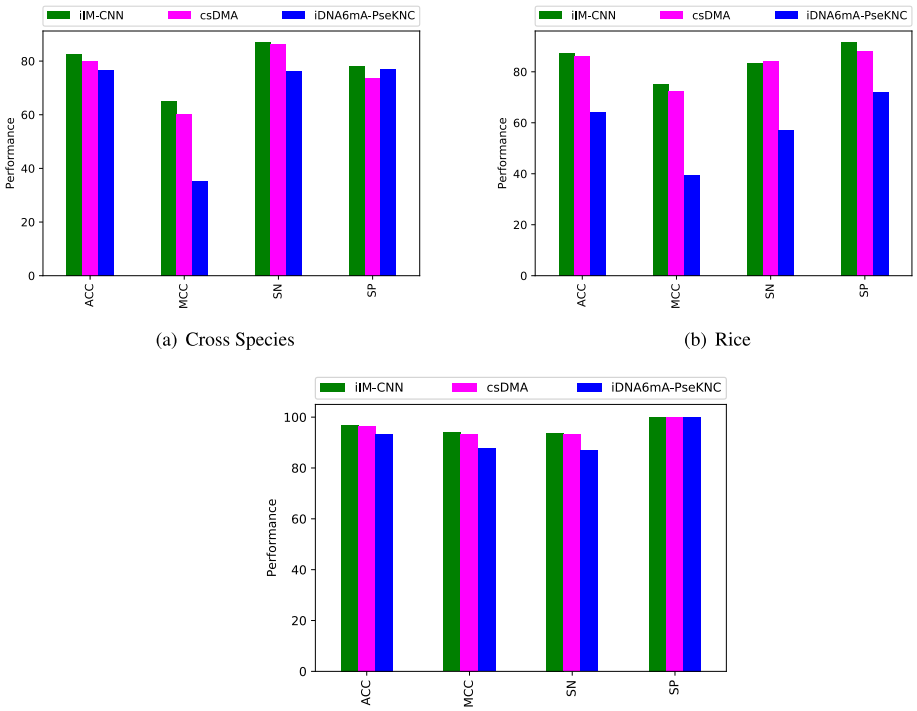
where M^+ and M^- represent the number of samples as positive or negative, respectively. M_-^+ is the number of



(a) Cross Species (b) Rice

(c) M. musculus

FIGURE 4. Confusion matrix of the proposed model iIM-CNN on 3 datasets. (a) Cross-species, (b) Rice, and (c) M. musculus.



(a) Cross Species (b) Rice

(c) M. musculus

FIGURE 5. Result comparisons of iIM-CNN model on three datasets with state-of-the-art models. (a) Cross Species. (b) Rice. (c) M. musculus.

positive examples that were identified as negatives, M_+^- states the number of negative samples that were predicted as positives samples. MCC depicts the prediction model performance for the skewed dataset. To calculate the success rate of the prediction model the receiver operating characteristic curve (ROC curve) was used. While the auROC (area

under the ROC curve) and F1 score are the significant measures for calculating a binary classifier's prediction quality and test accuracy respectively.

III. RESULT AND DISCUSSION

We evaluated iIM-CNN on three benchmark datasets containing 6mA sites sequences from the genomes of cross-species, rice, and Mus musculus respectively. Figure 2 and Table 4 depict the results of the proposed model, while Figure 3 and Figure 4 show the auROC curves of all species along with the visual representation of the confusion matrix, respectively.

To show the dominance of iIM-CNN, a thorough comparison with state-of-the-art-predictor csDMA [17] is shown in Table 5 and Figure 5 by using 5-fold cross-validation both of the predictors were evaluated on the same datasets. For the Cross-species, iIM-CNN enhanced the sensitivity, specificity, accuracy, MCC, auROC, F1 by 0.3%, 4.5%, 2.5%, 4.8%, 1.3%, 0.02%. respectively. For the rice genome, specificity, accuracy, MCC, auROC, F1 were improved by 3.4%, 1.4%, 2.9%, 1.1%, 1.2%. respectively. Finally, in the case of the Mus musculus genome, the sensitivity, accuracy, MCC, F1 were improved by 0.6%, 0.3%, 0.6%, 0.3% respectively. These results show that iIM-CNN outperforms the state-of-the-art csDMA [17] predictor which were achieved without handy crafted feature extraction from raw DNA sequences using CNN.

IV. WEBSERVER

As publicly accessible webservers have considerably increased the effects of bioinformatics on the research community and medical science [40] we made the (iIM-CNN) publically accessible at https://home.jbnu.ac.kr/NSCL/iIMCNN.htm. The webserver was built using Python and Flask library. It supports direct input sequence processing and uploading a FASTA file for processing. The allowed input sequence length is 41nt. The users can select the species types such as Mouse, Rice, and Cross-spices. The maximum number of the allowed sequence for processing is 1000 sequences.

V. CONCLUSION

In this study, we introduced an effective deep learning model called iIM-CNN for DNA N6-methyladenine (6mA) site prediction. The proposed model iIM-CNN used a convolution neural network for the automatic extraction of features from raw DNA sequences which is a major advantage in comparison with the state-of-the-art models. The achieved outcomes outperformed the current state-of-the-art models. The iIM-CNN is projected to be potentially effective in drug discovery and bioinformatics research. Finally, a web server has been established and made publicly and freely available at https://home.jbnu.ac.kr/NSCL/iIMCNN.htm

ACKNOWLEDGMENT

(Abdul Wahab and Syed Danish Ali contributed equally to this work.)

REFERENCES

[1] G.-Z. Luo and C. He, "Dna n6-methyladenine in metazoans: Functional epigenetic mark or bystander?" *Nature Struct. & Mol. Biol.*, vol. 24, no. 6, p. 503, 2017.
[2] Z. K. O'Brown and E. L. Greer, "N6-methyladenine: A conserved and dynamic dna mark," in *DNA Methyltransferases—Role and Function*. Switzerland: Springer, 2016, pp. 213–246.
[3] J. L. Campbell and N. Kleckner, "E. Coli oric and the dnaa gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork," *Cell*, vol. 62, no. 5, pp. 967–979, 1990.
[4] J. L. Robbins-Manke, Z. Z. Zdraveski, M. Marinus, and J. M. Essigmann, "Analysis of global gene expression and double-strand-break formation in dna adenine methyltransferase-and mismatch repair-deficient escherichia coli," *J. Bacteriol.*, vol. 187, no. 20, pp. 7027–7037, 2005.
[5] P. J. Pukkila, J. Peterson, G. Herman, P. Modrich, and M. Meselson, "Effects of high levels of dna adenine methylation on methyl-directed mismatch repair in *Escherichia Coli*," *Genetics*, vol. 104, no. 4, pp. 571–582, 1983.
[6] D. Dunn and J. Smith, "Occurrence of a new base in the deoxyribonucleic acid of a strain of bacterium coli," *Nature*, vol. 175, no. 4451, p. 336, 1955.
[7] A. P. Bird, "Use of restriction enzymes to study eukaryotic dna methylation: Ii. the symmetry of methylated sites supports semi-conservative copying of the methylation pattern," *J. Mol. Biol.*, vol. 118, no. 1, pp. 49–60, 1978.
[8] K. R. Pomraning, K. M. Smith, and M. Freitag, "Genome-wide high throughput analysis of dna methylation in eukaryotes," *Methods*, vol. 47, no. 3, pp. 142–150, 2009.
[9] A. M. Krais, M. G. Cornelius, and H. H. Schmeiser, "Genomic n^6-methyladenine determination by MEKC with LIF," *Electrophoresis*, vol. 31, no. 21, pp. 3548–3551, 2010.
[10] B. A. Flusberg, D. R. Webster, J. H. Lee, K. J. Travers, E. C. Olivares, T. A. Clark, J. Korlach, and S. W. Turner, "Direct detection of dna methylation during single-molecule, real-time sequencing," *Nature Methods*, vol. 7, no. 6, p. 461, 2010.
[11] E. L. Greer, M. A. Blanco, L. Gu, E. Sendinc, J. Liu, D. Aristizábal-Corrales , C.-H. Hsu, L. Aravind, C. He, and Y. Shi, "Dna methylation on n6-adenine in c. Elegans," *Cell*, vol. 161, no. 4, pp. 868–878, 2015.
[12] Y. Fu, G.-Z. Luo, K. Chen, X. Deng, M. Yu, D. Han, Z. Hao, J. Liu, X. Lu, L. C. Doré, X. Weng, Q. Ji, L. Mets, and C. He, "N6-methyldeoxyadenosine marks active transcription start sites in chlamydomonas," *Cell*, vol. 161, no. 4, pp. 879–892, 2015.
[13] M. J. Koziol, C. R. Bradshaw, G. E. Allen, A. S. Costa, C. Frezza, and J. B. Gurdon, "Identification of methylated deoxyadenosines in vertebrates reveals diversity in dna modifications," *Nature Struct. & Mol. Biol.*, vol. 23, no. 1, p. 24, 2016.
[14] S. J. Mondo, R. O. Dannebaum, R. C. Kuo, K. B. Louie, A. J. Bewick, K. LaButti, S. Haridas, A. Kuo, A. Salamov, and S. R. Ahrendt, "Widespread adenine n6-methylation of active genes in fungi," *Nature Genet.*, vol. 49, no. 6, p. 964, 2017.
[15] C. Zhou, C. Wang, H. Liu, Q. Zhou, Q. Liu, Y. Guo, T. Peng, J. Song, J. Zhang, and L. Chen, "Identification and analysis of adenine n6-methylation sites in the Rice genome," *Nature Plants*, vol. 4, no. 8, p. 554, 2018.
[16] P. Feng, H. Yang, H. Ding, H. Lin, W. Chen, and K.-C. Chou, "IDNA6mA-PseKNC: Identifying DNA N^6-methyladenosine sites by incorporating nucleotide physicochemical properties into PseKNC," *Genomics*, vol. 111, no. 1, pp. 96–102, 2019.
[17] Z. Liu, W. Dong, W. Jiang, and Z. He, "CSDMA: An improved bioinformatics tool for identifying dna 6 ma modifications via chou 5-step rule," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, 2019.
[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[19] H. Tayara and K. T. Chong, "Object detection in very high-resolution aerial images using one-stage densely connected feature pyramid network," *Sensors*, vol. 18, no. 10, p. 3341, 2018.
[20] H. Tayara, K. G. Soo, and K. T. Chong, "Vehicle detection and counting in high-resolution aerial images using convolutional regression neural network," *IEEE Access*, vol. 6, pp. 2220–2230, 2018.
[21] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, Aug. 2018.

[22] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, "Speech recognition using deep neural networks: A systematic review," *IEEE Access*, vol. 7, pp. 19143–19165, 2019.
[23] X. Pan, P. Rijnbeek, J. Yan, and H.-B. Shen, "Prediction of rna-protein sequence and structure binding preferences using deep convolutional and recurrent neural networks," *BMC Genomics*, vol. 19, no. 1, p. 511, 2018.
[24] I. Nazari, H. Tayara, and K. T. Chong, "Branch point selection in RNA splicing using deep learning," *IEEE Access*, vol. 7, pp. 1800–1807, 2018.
[25] Z. Louadi, M. Oubounyt, H. Tayara, and K. T. Chong, "Deep splicing code: Classifying alternative splicing events using deep learning," *Genes*, vol. 10, no. 8, p. 587, Aug. 2019.
[26] M. Tahir, H. Tayara, and K. T. Chong, "Irna-pseknc (2methyl): Identify rna 2'-o-methylation sites by convolution neural network and chou's pseudo components," *J. Theor. Biol.*, vol. 465, pp. 1–6, Mar. 2019.
[27] M. Oubounyt, Z. Louadi, H. Tayara, and K. T. Chong, "Deepromoter: Robust promoter predictor using deep learning," *Frontiers Genet.*, vol. 10, p. 286, Apr. 2019.
[28] W. Chen, H. Lv, F. Nie, and H. Lin, "I6ma-pred: Identifying dna n6-methyladenine sites in the rice genome," *Bioinformatics*, vol. 35, no. 16, pp. 2796–2800, Aug. 2019.
[29] L. Fu, B. Niu, Z. Zhu, S. Wu, and W. Li, "Cd-hit: Accelerated for clustering the next-generation sequencing data," *Bioinformatics*, vol. 28, no. 23, pp. 3150–3152, 2012.
[30] K.-C. Chou and H.-B. Shen, "Recent progress in protein subcellular location prediction," *Anal. Biochem.*, vol. 370, no. 1, p. 1, 2007.
[31] K.-C. Chou and C.-T. Zhang, "Prediction of protein structural classes," *Crit. Rev. Biochem. Mol. Biol.*, vol. 30, no. 4, pp. 275–349, 1995.
[32] F. Chollet. (2015). *Keras: Deep Learning Library for Theano and Tensorflow*. [Online]. Available: https://keras
[33] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Oper. Res.*, vol. 134, no. 1, pp. 19–67, 2005.
[34] M. Tahir, H. Tayara, and K. T. Chong, "Ipseu-cnn: Identifying rna pseudouridine sites using convolutional neural networks," *Mol. Therapy-Nucleic Acids*, vol. 16, pp. 463–470, Jun. 2019.
[35] I. Nazari, M. Tahir, H. Tayara, and K. T. Chong, "In6-methyl (5-step): Identifying rna n6-methyladenosine sites using deep learning mode via chou's 5-step rules and chou's general pseknc," *Chemometrics Intell. Lab. Syst.*, vol. 193, Oct. 2019, Art. no. 103811.
[36] M. Tahir, H. Tayara, and K. T. Chong, "Idna6ma (5-step rule): Identification of dna n6-methyladenine sites in the Rice genome by intelligent computational model via chou's 5-step rule," *Chemometrics Intell. Lab. Syst.*, vol. 189, pp. 96–101, Jun. 2019.
[37] H. Tayara, M. Tahir, and K. T. Chong, "Identification of prokaryotic promoters and their strength by integrating heterogeneous features," *Genomics*, 2019, doi: 10.1016/j.ygeno.2019.08.009.
[38] J. Khanal, I. Nazari, H. Tayara, and K. T. Chong, "4mccnn: Identification of n4-methylcytosine sites in prokaryotes using convolutional neural network," *IEEE Access*, vol. 7, pp. 145455–145461, 2019.
[39] B. Liu, S. Wang, R. Long, and K.-C. Chou, "IRSpot-EL: Identify recombination spots with an ensemble learning approach," *Bioinformatics*, vol. 33, no. 1, pp. 35–41, Jan. 2017.
[40] K. C. Chou, "Impacts of bioinformatics to medicinal chemistry," *Med. Chem.*, vol. 11, no. 3, pp. 218–234, 2015.

ABDUL WAHAB received the B.Sc. degree in computer science from the University of the Punjab, Lahore, Pakistan, in 2014. He is currently pursuing the master's degree with the Department of Electronics and Information Engineering, Chonbuk National University, Jeonju, South Korea. His research interests include bioinformatics, artificial intelligence, deep learning, machine learning, and image processing.

SYED DANISH ALI received the B.Sc. degree in electronics engineering from the Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Pakistan, in 2013, and the M.Sc. degree in electrical engineering from Abasyn University, Pakistan, in 2018. He is currently pursuing the Ph.D. degree in electronics and information engineering from Chonbuk National University, Jeonju, South Korea. He is currently working with the Department of Electrical Engineering, The University of Azad Jammu and Kashmir, Pakistan. His research interests include bioinformatics and machine learning.

HILAL TAYARA received the B.Sc. degree in computer engineering from the University of Aleppo, Aleppo, Syria, in 2008, and the M.S. and the Ph.D. degrees in electronics and information engineering from Chonbuk National University, Jeonju, South Korea, in 2015 and 2019, respectively. He is currently a Researcher with Chonbuk National University. His research interests include bioinformatics, machine learning, and image processing.

KIL TO CHONG received the Ph.D. degree in mechanical engineering from Texas A&M University, in 1995. He is currently a Professor with the School of Electronics and Information Engineering, Chonbuk National University, Jeonju, South Korea, and the Head of the Advanced Research Center of Electronics. His research interests include the areas of machine learning, signal processing, motor fault detection, network system control, and time-delay systems.