

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

TS-m6A-DL: Tissue-specific identification of N6-methyladenosine sites using a universal deep learning model



Zeeshan Abbas^{a,b,1}, Hilal Tayara^{c,1}, Quan Zou^{d,*}, Kil To Chong^{a,e,*}

^aDepartment of Electronics and Information Engineering, Jeonbuk National University, Jeonju 54896, South Korea

^bInstitute of Avionics and Aeronautics (IAA), Air University, Islamabad 44000, Pakistan

^cSchool of International Engineering and Science, Jeonbuk National University, Jeonju 54896, South Korea

^dInstitute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610054, China

^eAdvances Electronics and Information Research Center, Jeonbuk National University, Jeonju 54896, South Korea

ARTICLE INFO

Article history:

Received 24 May 2021

Received in revised form 8 August 2021

Accepted 9 August 2021

Available online 10 August 2021

Keywords:

Binary-encoding

Deep neural network

N6-methyladenosine (m6A)

Motif

Tissue-specific

ABSTRACT

The most communal post-transcriptional modification, N6-methyladenosine (m6A), is associated with a number of crucial biological processes. The precise detection of m6A sites around the genome is critical for revealing its regulatory function and providing new insights into drug design. Although both experimental and computational models for detecting m6A sites have been introduced, but these conventional methods are laborious and expensive. Furthermore, only a handful of these models are capable of detecting m6A sites in various tissues. Therefore, a more generic and optimized computational method for detecting m6A sites in different tissues is required. In this paper, we proposed a universal model using a deep neural network (DNN) and named it TS-m6A-DL, which can classify m6A sites in several tissues of humans (*Homo sapiens*), mice (*Mus musculus*), and rats (*Rattus norvegicus*). To extract RNA sequence features and to convert the input into numerical format for the network, we utilized one-hot-encoding method. The model was tested using fivefold cross-validation and its stability was measured using independent datasets. The proposed model, TS-m6A-DL, achieved accuracies in the range of 75–85% using the fivefold cross-validation method and 72–84% on the independent datasets. Finally, to authenticate the generalization of the model, we performed cross-species testing and proved the generalization ability by achieving state-of-the-art results.

© 2021 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The post-transcriptional RNA modification is an evolutionarily conserved process of RNAs that exists in all biological organisms [1]. The sophistication of biological knowledge and the evenness of regulation increases as RNA undergoes a post-transcriptional transformation. Over and above 150 different types of RNA post-translational modifications have been discovered, with methylation accounting for two-third of them [2]. N6-methyladenosine (m6A) modification is a typical and generous post-transcriptional RNA modification that affects nearly all cell cycle processes, includ-

ing translation efficiency, cell growth, and cell viability [3–8]. This indicates the methylation of the adenosine base at the nitrogen-6 location. Furthermore, m6A modification is a reversible process that can be triggered by methyltransferases and demethylases [9–11]. Research has shown that m6A is related to the occurrence of diseases including thyroid tumors [12], prostate cancer [13], obesity [14], and acute myelogenous leukemia [15]. m6A is a common transcriptional modification, which may occur in various species such as mammals, plants, and bacteria [16]. Research has shown that m6A acts as a regulator at each stage of mRNA metabolism [17]. This warrants extensive research on m6A modification, however, our current knowledge is still limited regarding m6A modification. Therefore, it is crucial to extensively study m6A, and correctly determine m6A modification sites in the transcriptome.

The m6A sites can be identified using two primary approaches. The first is experimental techniques including Methylated RNA Immunoprecipitation (MeRIP) [18], photo-crosslinking-assisted

* Corresponding authors at: Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610054, China (Q. Zou). Department of Electronics and Information Engineering, Jeonbuk National University, Jeonju 54896, South Korea (K.T. Chong).

E-mail addresses: zouquan@nclab.net (Q. Zou), kitchong@jbnu.ac.kr (K.T. Chong).

¹ Zeeshan Abbas and Hilal Tayara contributed equally.

(PA)-m6A-seq [19], m6A sequencing (m6A-seq) [20], m6A-crosslinking immunoprecipitation (CLIP) [21], individual-nucleotide-resolution cross-linking and immunoprecipitation (miCLIP) [22], DART-seq [23], MAZTER-Seq [24] and m6A-eCLIP (meCLIP) [25]. These experimental approaches provide a layout for the identification of potential m6A modification sites. Moreover, a few bioinformatics techniques, which are capable of discovering m6A sites directly from the data gathered by experimental techniques have been suggested [26–28]. However, the sequencing data gathered is already too large to be catered by basic bioinformatics techniques; therefore, there is a requirement for more efficient techniques to identify m6A sites in the transcriptome. The second approach to identify m6A sites is the application of computational techniques to sequencing data. In recent studies using spatial specificity of gene expression, it was discovered that the site of m6A modification varies in different tissues and species. Doa et al., [29] suggested a technique called iRNA-m6A that can detect m6A modification sites in various tissues of human, mouse, and rat by using a support vector machine (SVM), which was applied to the dataset provided by Zhang et al., [30]. This approach significantly enhances the accuracy of identifying m6A modification sites. However, there is still much room for improvement in techniques for detecting m6A sites.

In recent years, deep learning models have insignificantly impacted the field of bioinformatics. Various computational methodologies comprising deep-learning techniques, including Gene2Vec [31], DeepM6ASeq [32], BERMP [33], DNA6mA-MINT [34], pcPromoter-CNN [35], 4mCPred-CNN [36], im6A-TS-CNN [37], and iPseU-CNN [38] have been suggested. Liu et al., proposed a CNN-based solution in im6A-TS-CNN, which utilizes one-hot encoding technique to encode the data samples and then classify the data using CNN architecture using the datasets of *Homo sapiens*, *Rattus norvegicus*, and *Mus musculus*. Dao et al., [29] proposed an SVM-based machine-learning model to classify m6A sites on the same datasets. They used various encoding schemes including mono-nucleotide binary encoding, physical-chemical property matrix, and nucleotide chemical property.

In light of the above studies, the current research is focused on building a computational method that can detect m6A modification sites in human, mouse, and rat tissues. To create a benchmark dataset based on the experimental analysis performed by Zhang et al., [30], we first compiled experimentally validated m6A and non-m6A sequences. To formulate the samples, only a single encoding scheme, one-hot-encoding, was used. Motivated by the increasing implementation of deep-learning algorithms, we proposed a universal CNN-based approach called, TS-m6A-DL. Performing 5-fold cross-validation and independent testing revealed that, TS-m6A-DL, outperforms the current state-of-the-art approaches. Additionally, cross-species validation testing proved the stability of our model achieving state-of-the-art results.

2. Materials and methods

2.1. Benchmark datasets

A major step in training an efficient computational model is the construction of high-quality datasets. Zhang et al., [30] produced an efficient antibody-independent m6A detection tool, m6A-REF-seq, to classify the sites of alteration in various tissues including brain, liver, kidney, heart, and testis of *Homo sapiens*, *Mus musculus*, and *Rattus norvegicus* species. Considering the excellent caliber of these data, benchmark dataset was developed using them. To further enhance the data integrity, we only used sequences with a length of 41 nucleotides (nt) having the m6A site in the middle for positive sequences. Sample sequences exceeding 80% resem-

blance were eliminated using the CD-HIT software [39] to prevent duplication and eliminate homology bias. The negative sequences, which were experimentally proved as non-methylated, for the abovementioned tissues were extracted by fulfilling the 41nt length requirement containing Adenine in the center. They show m6A consensus motif but were not enriched in the analysis of m6A. Since an unbalanced dataset replicates biasness towards one class [40], random negative sequences equal to the number of positive sequences were retrieved. The datasets are divided into training and independent datasets to critically analyze the developed model's performance and generalization. There is minimal overfitting when the proposed technique is appropriate for the independent dataset [29]. Table 1 lists the specifics of the positive and negative sequences in each dataset.

2.2. Sequence encoding

In neural networks, making the sequence readable for the network using an efficient encoding scheme is the foremost prerequisite that directly affects the model's performance. One-hot-encoding is a widely used encoding scheme that can accurately represent the nucleotides as a four-dimensional binary vector. We can represent the nucleotides as follows:

$$A \mapsto [1, 0, 0, 0]$$

$$U \mapsto [0, 1, 0, 0]$$

$$C \mapsto [0, 0, 1, 0]$$

$$G \mapsto [0, 0, 0, 1]$$

2.3. Network architecture

An input layer, several intermediary hidden layers, and an output layer comprise a neural network. After converting the sequence into a numerical format to make it readable for the network, the input layer receives the 41 x 4 matrix as an input. The proposed network architecture is illustrated in Fig. 1. The input matrix is fed to the first convolution layer with 64 filters with a kernel size of 3 and a stride of 1. The number of feature maps in DNNs multiplies with the network's depth, resulting in a drastic increase in the number of parameters and computational requirements when larger kernel sizes are used. Therefore, this convolution layer is followed by another convolution layer with 16 filters having unit kernel size and the same stride. The concept was first introduced by Lin et al., in their Network in Network study [41]. This particular strategy has been used in this study, to reduce dimensionality by reducing the number of feature maps while keeping the most important features and neglecting the features which are not contributing. The convolution layer can be represented mathematically as follows:

$$\text{conv}(X)_{ik} = \text{ReLU} \left(\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} W_{mn}^k X_i + m, n \right) \quad (1)$$

where 'X' represents the input, 'i' represents the output position index, and 'k' determine the kernel index. W^k is a $M \times N$ weight matrix, where 'M' is the window size and 'N' is the number of input channels.

This layer is then followed by a max-pooling layer with a pool size of 2, to downsample the content of feature maps, shrinking their height and width whilst preserving their salient features. It can be calculated mathematically by the following equation:

$$\text{pooling}(X)_{ik} = \max(\{X_{iM,k}, X_{(iM+1),k}, \dots, X_{(iM+M-1),k}\}) \quad (2)$$

Table 1

Benchmark datasets of different tissues of Human, Mouse, and Rat. Here, the methylated sequences are referred as positive, and non-methylated sequences are referred as negative. The negative sequences show an m6A consensus motif but were not enriched in the analysis for m6A.

Species	Tissues	Positive		Negative	
		Training	Testing	Training	Testing
Human	Brain	4605	4604	4605	4604
	Liver	2634	2634	2634	2634
	Kidney	4574	4573	4574	4573
Mouse	Brain	8025	8025	8025	8025
	Liver	4133	4133	4133	4133
	Kidney	3953	3952	3953	3952
	Heart	2201	2201	2200	2200
	Testis	4707	4706	4707	4706
Rat	Brain	2352	2351	2352	2351
	Liver	1762	1762	1762	1762
	Kidney	3433	3432	3433	3432

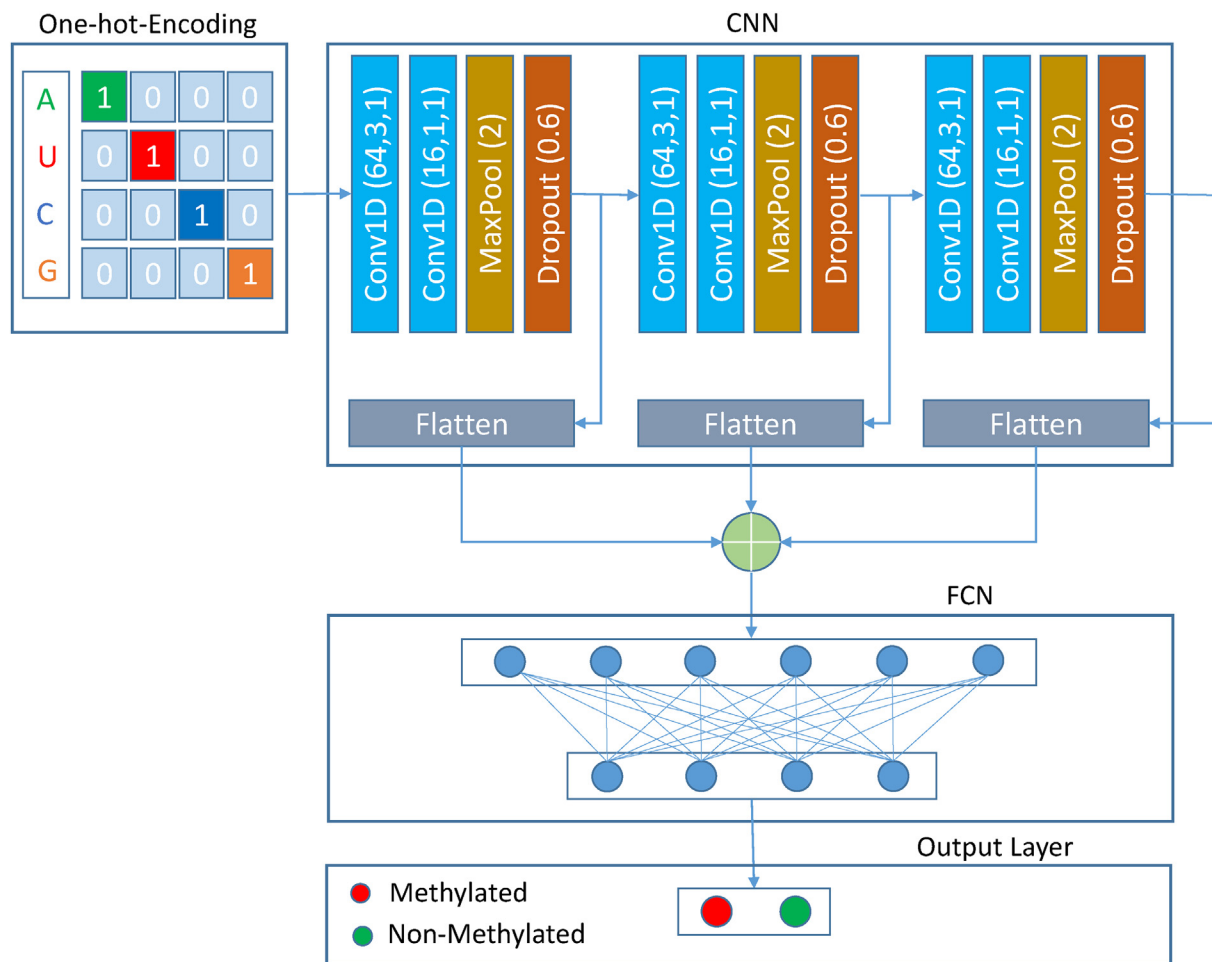


Fig. 1. Architecture of the proposed network, TS-m6A-DL. First step is to encode the sequence data to make it readable for the network. Then the encoded data is fed to the convolution layer; Conv1D(f,k,s), where f is the number of filters, k is the kernel size, and s is the stride. The convolution layer is followed by the pooling layer MaxPool(2), where 2 is the pooling size; which is then followed by the dropout layer Dropout(0.6), where 0.6 is the dropout rate.

where 'X' represents for the input feature map, 'i' represents the output position index, 'k' denotes the kernel index, and 'M' denotes the pool window size.

To avoid the overfitting bias, we used a dropout function with a rate of 0.6. The output after the dropout was flattened and simultaneously fed as input to another convolution layer, repeating the same procedure as above, two more times. The three flattened outputs after every block were concatenated and fed as input to the dense layer containing 16 nodes, followed by the output layer,

which determines whether the sequence is methylated or non-methylated. Each convolution layer including the dense layer uses a rectified linear function (ReLU) as the activation function, except for the output layer which uses the sigmoid function for classification. ReLU and sigmoid can be expressed mathematically as follows:

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

$$\text{Sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

We further used the L2 regularization approach for the kernel and bias within the convolution layer to prevent overfitting, setting the rate to 0.001. The Nadam (Nesterov accelerated Adaptive Moment Estimation) optimizer with a learning rate of 0.0021 was used to construct the model accompanying binary cross-entropy as the loss function. The maximum training epoch was set to 50, and the training batch size was set to 32. Furthermore, during the training phase, we also used the early stop strategy through which the training process would halt if the prediction accuracy on the validation set stopped improving for 10 epochs.

3. Evaluation metrics

K-fold cross-validation and independent testing are widely used to assess the performance of a model. To analyze the effectiveness and robustness of the model, we performed both of the above-mentioned tests on our model, defining K equals to 5 in our model. To objectively assess the results from the analysis and for an unbiased comparison with the previous state-of-the-art methods, we used the same widely used metrics including sensitivity (Sn), specificity (Sp), accuracy (ACC), and Matthews correlation coefficient (MCC).

$$\text{Sensitivity}(Sn) = \frac{TP}{TP + FN} \quad (5)$$

$$\text{Specificity}(Sp) = \frac{TN}{TN + FP} \quad (6)$$

$$\text{Accuracy}(ACC) = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (8)$$

where TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively. Furthermore, we also used the area under the curve (AUC) which is a useful metric for assessing predictive performance of a model.

4. Results and discussion

We developed a generalized model for the classification of m6A sites based on the data defined in the Materials and methods. To prove the model's stability in order to classify tissue-specific m6A modifications in humans, mice, and rats; we performed both 5-fold cross-validation and independent testing, and formulated

the results in Table 2. The proposed approach was found to be robust when the findings of the 5-fold cross-validation and independent tests were compared with existing state-of-the-art models.

4.1. Comparison with existing methods

To validate our model, we used the same validation approach as that used for the validation of iRNA-m6A [29] and m6A-TS-CNN [37], retaining the same number of folds, to obtain a better comparative study. Figs. 2a and 2b show the comparison results in terms of accuracy, as determined by 5-fold cross-validation and independent testing. High accuracy on independent dataset implies that the model has the ability to identify m6A sites for an unknown sequence. A complete comparison using all five evaluation matrices; ACC, Sn, Sp, MCC, and AUC is shown in Table S1 and S2 for the 5-fold cross-validation and independent testing, respectively.

4.2. Cross-species testing

Since the datasets come from various organisms and tissues, it is useful to see how a model trained on samples from a particular tissue of one species can recognize m6A in the same tissues of another species. Therefore, for the same tissues in different species, we applied cross-species testing and provided the accuracies in Figs. 3a, 3b, and 3v for the brain, kidney, and liver, respectively. The x-axis shows the datasets on which we trained our model and the y-axis shows the datasets on which we tested our model.

The results indicate that the proposed tool, TS-m6A-DL, is efficient for cross-species testing for the classification of m6A sites, proving its universality. The complete results using the same five evaluation matrices are illustrated in Table S3.

4.3. Motif analysis

Motifs were calculated from the first activation layer [42]. Each input sequence produced a corresponding activation map in the first layer from which we selected the maximum activation. This maximum activation was mapped back to the input sequence to select a subsequence with the filter-size length. The selected subsequences from each filter were aligned and used to find the motif using position weight matrix (PWM) technique [43]. Multiple motifs were generated using 64 filters in the first layer. Then, we compared these motifs with biologically reported motifs [30]. We found a strong match of ACA motif detected by our model with that reported by Zhang et al., [30]. Fig. 4 shows the motif detected by our model in which the first A is the putative m6A.

Table 2

Performance of TS-m6A-DL; Before the underscore, characters h, m, and r stand for human, mouse, and rat, respectively; and the b,h,k,l, and t after underscore denotes brain, heart, kidney, liver, and testis, respectively.

Species	5-fold Cross Validation					Independent Testing				
	ACC	Sn	Sp	MCC	AUC	ACC	Sn	Sp	MCC	AUC
h_b	0.7507	0.8191	0.6823	0.5068	0.8262	0.7384	0.8123	0.6646	0.4822	0.8097
h_k	0.8099	0.8393	0.7804	0.6211	0.8904	0.8020	0.8045	0.7996	0.6042	0.8802
h_l	0.8335	0.8595	0.8075	0.6684	0.9135	0.8056	0.8204	0.7908	0.6115	0.8784
m_b	0.7985	0.8134	0.7836	0.5974	0.8831	0.7877	0.8291	0.7462	0.5774	0.8725
m_h	0.7823	0.8150	0.7496	0.5664	0.8504	0.7502	0.7931	0.7072	0.5023	0.8234
m_k	0.8222	0.8259	0.8186	0.6451	0.9079	0.8076	0.8421	0.7732	0.6168	0.8892
m_l	0.7506	0.7953	0.7060	0.5044	0.8288	0.7203	0.7805	0.6600	0.4438	0.7913
m_t	0.7756	0.8179	0.7333	0.5544	0.8630	0.7644	0.8425	0.6863	0.5354	0.8432
r_b	0.7955	0.8227	0.7682	0.5922	0.8758	0.7728	0.8132	0.7324	0.5475	0.8543
r_k	0.8341	0.8400	0.8281	0.6683	0.9066	0.8327	0.8522	0.8132	0.6660	0.9083
r_l	0.8348	0.8347	0.8348	0.6706	0.9025	0.8101	0.8547	0.7656	0.6227	0.8853

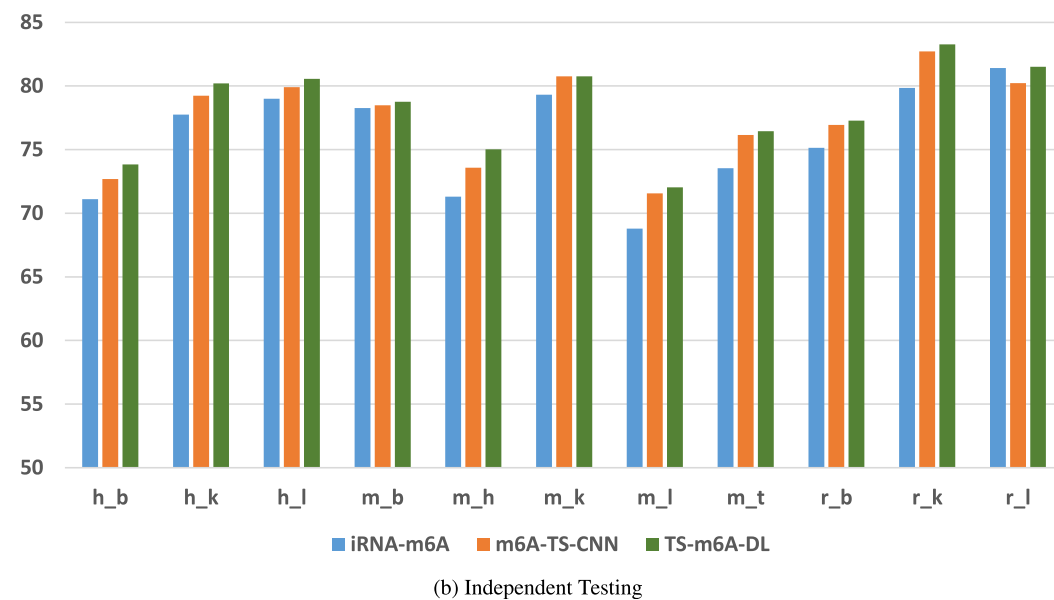
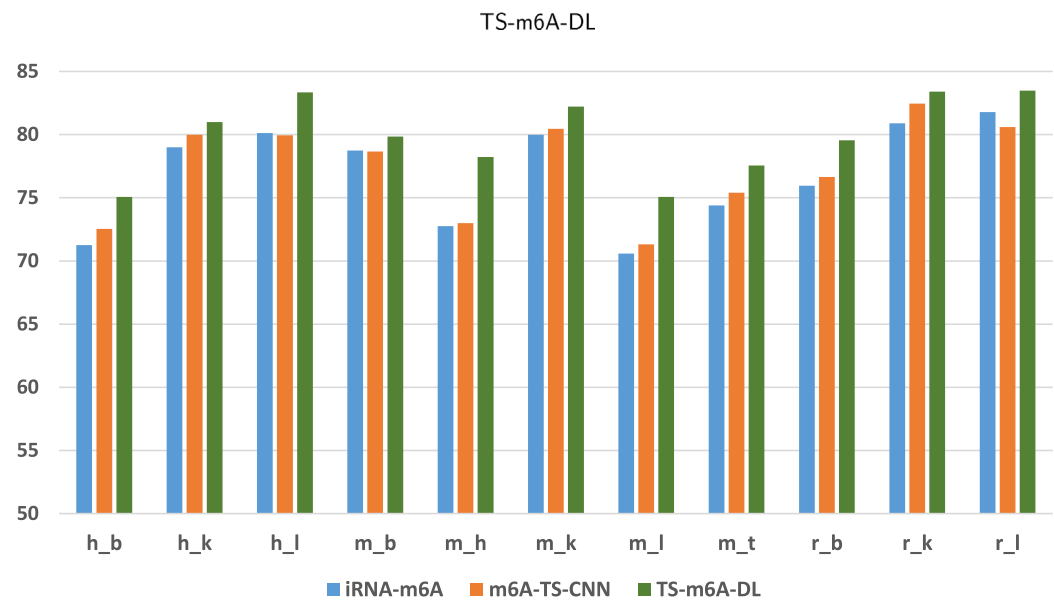


Fig. 2. Comparison of the models; iRNA-m6A, m6A-TS-CNN, and TS-m6A-DNN in term of accuracy.

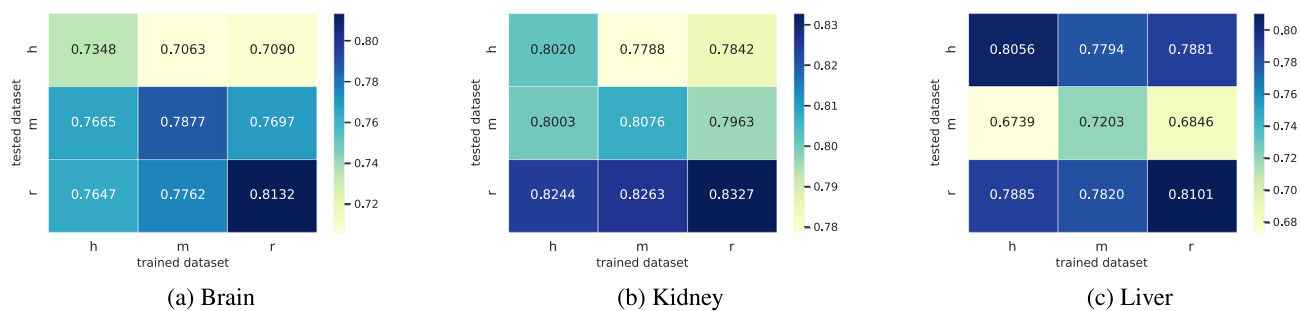


Fig. 3. Cross-species testing. Here, h,m, and r on x and y-axis represents the species human, mouse, and rat.

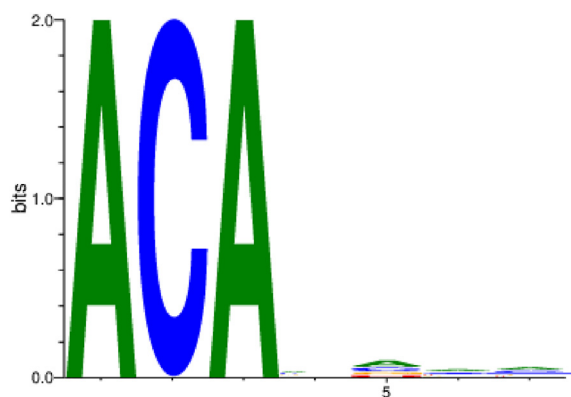


Fig. 4. The ACA-motif detected by our model, TS-m6A-DL.

5. Conclusion

Since m6A is involved in a variety of biological processes, precise detection of m6A sites is critical for scientific investigations to understand its regulatory function and to obtain various insights into drug design. In this study, we present TS-m6A-DL, a deep-learning-based universal model for detecting m6A sites in various tissues of humans, mice, and rats. To validate our model, we performed 5-fold cross-validation and independent testing, and the achieved results exemplified that the TS-m6A-DL tool performs better than the previous state-of-the-art tools. A web-based server was developed and made available at <http://nscbio.jbnu.ac.kr/tools/TS-m6A-DL/> for the benefit of the research community. We expect that the new framework will be useful for detecting m6A-sites and facilitate drug development.

CRedit authorship contribution statement

Zeeshan Abbas: Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing. **Hilal Tayara:** Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing. **Quan Zou:** Conceptualization, Validation, Supervision, Writing - review & editing. **Kil To Chong:** Conceptualization, Validation, Supervision, Writing - review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2020R1A2C2005612) and in part by the Brain Research Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. NRF-2017M3C7A1044816).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.csbj.2021.08.014>.

References

- [1] Sun Xi Hong, Zhao Ling Ping, Zou Quan, Wang Zhan Bin. Identification of microRNA genes and their mRNA targets in *Festuca arundinacea*. *Appl Biochem Biotechnol* 2014;172(8):3875–87.
- [2] Toone Eric J. *Advances in enzymology and related areas of molecular biology*, volume 240. John Wiley & Sons 2011.
- [3] Wang Xiao, Zhao Boxuan Simen, Roundtree Ian A, Lu Zhike, Han Dali, Ma Honghui, Weng Xiaocheng, Chen Kai, Shi Hailing, He Chuan, et al. N6-methyladenosine modulates messenger RNA translation efficiency. *Cell* 2015;161(6):1388–99.
- [4] Wang Yang, Li Yue, Toth Julia I, Petroski Matthew D, Zhang Zhaolei, Zhao Jing Crystal. N 6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat Cell Biol* 2014;16(2):191–8.
- [5] Bokar Joseph A. The biosynthesis and functional roles of methylated nucleosides in eukaryotic mRNA. In *Fine-tuning of RNA functions by modification and editing*. Springer; 2005. p. 141–77.
- [6] Rehman Mobeen Ur, Hong Kim Jee, Tayara Hilal, Chong Kil to. m6a-neuraltool: convolution neural tool for RNA N6-methyladenosine site identification in different species. *IEEE Access* 2021;9:17779–86.
- [7] Alam Waleed, Ali Syed Danish, Tayara Hilal, Chong Kil to. A CNN-based RNA N6-methyladenosine site predictor for multiple species using heterogeneous features representation. *IEEE Access* 2020;8:138203–9.
- [8] Ali Syed Danish, Alam Waleed, Tayara Hilal, Chong Kil. Identification of functional piRNAs using a convolutional neural network. *IEEE/ACM Trans Comput Biol Bioinf* 2020.
- [9] Jia Guifang, Ye Fu, Zhao Xu, Dai Qing, Zheng Guanqun, Yang Ying, Yi Chengqi, Lindahl Tomas, Pan Tao, Yang Yun-Gui, et al. N 6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated fat. *Nat Chem Biol* 2011;7(12):885–7.
- [10] Jia Guifang, Ye Fu, He Chuan. Reversible RNA adenosine methylation in biological regulation. *Trends Genet* 2013;29(2):108–15.
- [11] Xu Yamei Niu, Zhao Yong-Sheng Wu, Li Ming-Ming, Wang Xiu-Jie, Yang Yun-Gui. N6-methyl-adenosine (m6A) in RNA: an old modification with a novel epigenetic function. *Genomics Proteomics Bioinf* 2013;11(1):8–17.
- [12] Yu Zhi-Hao, Feng Shao-Ting, Zhang Di, Cao Xu-Chen, Yu Yue, Wang Xin. The functions and prognostic values of m6A RNA methylation regulators in thyroid carcinoma. *Cancer Cell Int* 2021;21(1):1–15.
- [13] Cai Jiarong, Yang Fei, Zhan Hailun, Situ Jie, Li Wenbiao, Mao Yunhua, Luo Yun. RNA m6A methyltransferase METTL3 promotes the growth of prostate cancer by regulating hedgehog pathway. *Oncotargets Ther* 2019;12:9143.
- [14] Machiela Mitchell J, Lindström Sara, Allen Naomi E, Haiman Christopher A, Albanes Demetrius, Barricarte Aurelio, Berndt Sonja I, Bas Bueno-de Mesquita H, Chanock Stephen, Michael Gaziano J, et al. Association of type 2 diabetes susceptibility variants with advanced prostate cancer risk in the breast and prostate cancer cohort consortium. *Am J Epidemiol* 2012;176(12):1121–9.
- [15] Bansal H, Yihua Q, Iyer Swaminathan Padmanabhan, Ganapathy S, Proia D, Penalva LO, Uren PJ, Suresh U, Carew JS, Karnad AB, et al. Wtap is a novel oncogenic protein in acute myeloid leukemia. *Leukemia* 2014;28(5):1171–4.
- [16] Bodi Zsuzsanna, Button James D, Grierson Donald, Fray Rupert G. Yeast targets for mRNA methylation. *Nucl Acids Res* 2010;38(16):5327–35.
- [17] Zhao Boxuan Simen, Roundtree Ian A, He Chuan. Post-transcriptional gene regulation by mRNA modifications. *Nat Rev Mol Cell Biol* 2017;18(1):31–42.
- [18] Meyer Kate D, Saletore Yogesh, Zumbo Paul, Elemento Olivier, Mason Christopher E, Jaffrey Samie R. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 2012;149(7):1635–46.
- [19] Chen Kai, Zhike Lu, Wang Xiao, Fu Ye, Luo Guan-Zheng, Liu Nian, Han Dali, Dominissini Dan, Dai Qing, Pan Tao, et al. High-resolution N6-methyladenosine (m6A) map using photo-crosslinking-assisted m6A sequencing. *Angew Chem* 2015;127(5):1607–10.
- [20] Dominissini Dan, Moshitch-Moshkovitz Sharon, Schwartz Schraga, Salmon-Divon Mali, Ungar Lior, Osenberg Sivan, Cesarkas Karen, Jacob-Hirsch Jasmine, Amariglio Ninette, Kupiec Martin, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 2012;485(7397):201–6.
- [21] Ke Shengdong, Alemu Endalkachew A, Mertens Claudia, Gantman Emily Conn, Fak John J, Mele Aldo, Haripal Bhagwattie, Zucker-Scharff Ilana, Moore Michael J, Park Christopher Y, et al. allowing the potential for 3' UTR regulation. *Genes Develop* 2015;29(19):2037–53.
- [22] Linder Bastian, Grozhik Anya V, Olarerin-George Anthony O, Meydan Cem, Mason Christopher E, Jaffrey Samie R. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods* 2015;12(8):767–72.
- [23] Meyer Kate D. Dart-seq: an antibody-free method for global m6A detection. *Nat Methods* 2019;16(12):1275–80.
- [24] Pandey Radha Raman, Pillai Ramesh S. Counting the cuts: Mazter-seq quantifies m6A levels using a methylation-sensitive ribonuclease. *Cell* 2019;178(3):515–7.
- [25] Roberts Justin T, Porman Allison M, Johnson Aaron M. Identification of m6A residues at single-nucleotide resolution using eclip and an accessible custom analysis pipeline. *RNA* 2021;27(4):527–41.
- [26] Meng Jia, Lu Zhiliang, Liu Hui, Zhang Lin, Zhang Shaowu, Chen Yidong, Rao Manjeet K, Huang Yufei, et al. A protocol for RNA methylation differential analysis with merip-seq data and exomepeak R/bioconductor package. *Methods* 2014;69(3):274–81.

- [27] Chen Zhen, Zhao Pei, Li Fuyi, Wang Yanan, Ian Smith A, Webb Geoffrey I, Akutsu Tatsuya, Baggag Abdelkader, Bensmail Halima, Song Jiangning, et al. Comprehensive review and assessment of computational methods for predicting rna post-transcriptional modification sites from rna sequences. *Brief Bioinf* 2020;21(5):1676–96.
- [28] Lorenz Daniel A, Sathe Shashank, Einstein Jaclyn M, Yeo Gene W. Direct rna sequencing enables m6a detection in endogenous transcript isoforms at base-specific resolution. *Rna* 2020;26(1):19–28.
- [29] Dao Fu-Ying, Lv Hao, Yang Yu-He, Zulficar Hasan, Gao Hui, Lin Hao. Computational identification of n6-methyladenosine sites in multiple tissues of mammals. *Comput Struct Biotechnol J* 2020;18:1084–91.
- [30] Zhang Zhang, Chen Li-Qian, Zhao Yu-Li, Yang Cai-Guang, Roundtree Ian A, Zhang Zijie, Ren Jian, Xie Wei, He Chuan, Luo Guan-Zheng, et al. Single-base mapping of m6a by an antibody-independent method. *Sci Adv* 2019;5(7):eaax0250.
- [31] Zou Quan, Xing Pengwei, Wei Leyi, Liu Bin. Gene2vec: gene subsequence embedding for prediction of mammalian n6-methyladenosine sites from mrna. *Rna* 2019;25(2):205–18.
- [32] Zhang Yiqian, Hamada Michiaki. Deepm6aseq: prediction and characterization of m6a-containing sequences using deep learning. *BMC Bioinf* 2018;19(19):1–11.
- [33] Yu Huang, Yu Ningning He, Chen Zhen Chen, Li Lei. Bermpp: a cross-species classifier for predicting m6a sites by integrating a deep learning algorithm and a random forest approach. *Int J Biol Sci* 2018;14(12):1669.
- [34] Rehman Mobeen Ur, Chong Kil To. Dna6ma-mint: Dna-6ma modification identification neural tool. *Genes* 2020;11(8):898.
- [35] Shujaat Muhammad, Wahab Abdul, Tayara Hilal, Chong Kil To. pcpromoter-cnn: a cnn-based prediction and classification of promoters. *Genes* 2020;11(12):1529.
- [36] Abbas Zeeshan, Tayara Hilal, Chong Kil To. 4mcpred-cnn-prediction of dna n4-methylcytosine in the mouse genome using a convolutional neural network. *Genes* 2021;12(2):296.
- [37] Liu Kewei, Cao Lei, Du Pufeng, Chen Wei. im6a-ts-cnn: identifying the n6-methyladenine site in multiple tissues by using the convolutional neural network. *Mol Ther-Nucl Acids* 2020;21:1044–9.
- [38] Tahir Muhammad, Tayara Hilal, Chong Kil To. ipseu-cnn: identifying rna pseudouridine sites using convolutional neural networks. *Mol Ther-Nucl Acids* 2019;16:463–70.
- [39] Li Weizhong, Godzik Adam. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006;22(13):1658–9.
- [40] Wang Ruohan, Wang Zishuai, Wang Jianping, Li Shuaicheng. Splicefinder: ab initio prediction of splice sites using convolutional neural network. *BMC Bioinf* 2019;20(23):1–13.
- [41] Lin Min, Chen Qiang, Yan Shuicheng. Network in network. *arXiv preprint arXiv:1312.4400*; 2013..
- [42] Quang Daniel, Xie Xiaohui. Factornet: a deep learning framework for predicting cell type specific transcription factor binding from nucleotide-resolution sequential data. *Methods* 2019;166:40–7.
- [43] Xia Xuhua. Position weight matrix, gibbs sampler, and the associated significance tests in motif characterization and prediction. *Scientifica* 2012;2012.