

Original Research

DT-m6A: A DenseNet–Transformer Hybrid Framework for Accurate Prediction of m6A Modification Sites across Diverse Cell Lines and Tissues

Qiyu Tao¹ , Jianhua Jia^{1,*} 

¹School of Information Engineering, Jingdezhen Ceramic University, 333403 Jingdezhen, Jiangxi, China

*Correspondence: jjh163yx@163.com (Jianhua Jia)

Academic Editor: Amancio Carnero Moya

Submitted: 6 November 2025 | Revised: 21 December 2025 | Accepted: 31 December 2025 | Published: 22 January 2026

Abstract

Background: N6-methyladenosine (m6A) RNA methylation is a crucial epigenetic modification that plays an essential role in regulating diverse biological processes. Accurate identification of m6A sites is therefore fundamental to understanding its regulatory mechanisms. In this study, we proposed DT-m6A, a novel deep learning framework that integrates DenseNet and Transformer architectures for accurate m6A site identification across diverse cell lines and tissues. **Methods:** RNA sequences are first encoded using nucleotide chemical properties (NCP) for initial features extraction, after which DenseNet captures and reuses local sequence features through dense connections. The Transformer module then models long-range dependencies and extracts nonlinear representations, in which Batch Normalization replaces the conventional Layer Normalization in both sublayers to enhance training stability. Finally, a fully connected layer predicts m6A modification sites. **Results:** Evaluated on 11 independent test sets spanning eight cell lines and three tissue types, DT-m6A demonstrated robust performance, achieving average accuracy (ACC) of 76.97%, Matthews correlation coefficient (MCC) of 54.27%, precision (PRE) of 75.18%, recall (REC) of 79.76%, and F1 score of 77.26%. **Conclusions:** DT-m6A surpassed the state-of-the-art method MST-m6A by 0.63% in average accuracy ($p = 0.0023$) and 1.4% in mean MCC ($p = 0.0012$) across 11 independent test sets. Although its performance on the CD8T and MOLM13 cell lines was comparable to MST-m6A, DT-m6A consistently achieved superior results across all other cell lines and tissues. Overall, DT-m6A effectively captures both local patterns and global dependencies in RNA sequences, improving prediction performance across diverse biological contexts.

Keywords: N6-methyladenosine; m6A site identification; deep learning; DenseNet; transformer

1. Introduction

Transcription gives rise to RNA molecules that are extensively regulated by more than 170 chemically distinct modifications, collectively referred to as the epitranscriptome [1]. These modifications modulate RNA metabolism and gene expression and are widely distributed across diverse RNA species, including rRNA, tRNA, snRNA, mRNA, and long non-coding RNAs, in organisms ranging from viruses and yeast to plants and animals [2]. Although RNA nucleotide modifications have been recognized for decades, recent advances in high-throughput sequencing and analytical technologies have revealed their dynamic regulatory mechanisms and biological significance, driving rapid growth in epitranscriptome research [3]. Among these modifications, N6-methyladenosine (m6A) is the most prevalent, abundant, and evolutionarily conserved internal modification in eukaryotic messenger RNAs. It is especially enriched in mammalian mRNAs, where it occurs at tens of thousands of sites and accounts for approximately 0.15%–0.6% of total adenosines [4,5]. As the most abundant internal mRNA modification, m6A preferentially appears within consensus motifs such as DRACH and RRACH (D=A/G/U; R=A/G; H=A/C/U), while ex-

hibiting considerable site-specific variability in methylation levels [6]. In addition to sequence motifs, m6A is enriched in long exons, near stop codons, and within 3' untranslated regions (3' UTRs) [7,8], suggesting its strategic role in regulating mRNA metabolism and function [9,10]. Importantly, m6A is a dynamic and reversible epitranscriptomic modification whose deposition, recognition, and removal are orchestrated by writer, reader, and eraser proteins [11,12]. Through this coordinated regulatory network, m6A participates in diverse biological processes, including gene transcription [13], cell signal transduction [14], and DNA damage response [15]. As many of these functions depend on the precise positioning and dynamic modulation of m6A, accurate identification of m6A sites across the transcriptome is essential for elucidating RNA regulatory mechanisms.

Before 2012, transcriptome-wide m6A distribution was poorly understood. High-throughput methods such as MeRIP-seq [16] and m6A-seq [6] enabled transcriptome-wide profiling but suffered from low resolution. Higher-resolution techniques, including miCLIP [17] (integrating CLIP [18] with m6A-specific antibodies), m6A-REF-seq [19] and DART-seq [20], improved site-level mapping and reduced antibody dependence. Despite these



Copyright: © 2026 The Author(s). Published by IMR Press.
This is an open access article under the CC BY 4.0 license.

Publisher's Note: IMR Press stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

advances, experimental approaches remain time- and resource-intensive.

With advances in artificial intelligence and the increasing availability of experimentally validated m6A datasets, computational prediction of RNA N⁶-methyladenosine (m6A) sites has become an active research area. Existing methods can generally be classified into three categories: (i) machine learning-based methods, such as iRNAMethy [21], iRNA-PseColl [22], iRNA(m6A)-PseDNC [23], M6AMRFS [24], and WHISTLE [25]; (ii) deep learning-based methods, including DL-m6A [26], DeepM6ASeq [27], MASS [28], TS-m6A-DL [29], MultiRM [30], iMethyl-Deep [31], and MTDeepM6A-2S [32]; and (iii) ensemble learning methods that integrate machine learning or deep learning models, such as SRAMP [33], M6APred-EL [34], EMDLP [35], and DeepM6ASeq-EL [36].

The recently developed MST-m6A [37] leverages the transfer learning framework DNA-BERT [38] to extract sequence features, followed by a three-layer convolutional neural network and a multilayer perceptron for prediction. Although this strategy achieves competitive performance, the large number of parameters in DNA-BERT leads to substantial computational overhead, making inference and fine-tuning memory- and time-intensive, particularly for long sequences or large-scale datasets.

To address these limitations, we propose DT-m6A, a computationally efficient DenseNet-Transformer hybrid architecture specifically designed for m6A site prediction. Unlike prior CNN-Transformer or CNN-attention hybrid methods that primarily rely on shallow convolutional feature extraction, DT-m6A introduces a DenseNet-inspired convolutional module that promotes dense feature propagation and enables the reuse of hierarchical local representations across layers. This design provides the Transformer with richer and more discriminative convolutional features, improving sequence modeling while avoiding excessive parameter growth. In addition, we design a lightweight Transformer tailored to short RNA sequences with relatively fixed lengths by replacing standard layer normalization with batch normalization, which improves optimization stability under small-batch training conditions. Parameter sharing and hierarchical processing are further incorporated to reduce redundancy and computational cost. Combined with NCP encoding, which provides biochemically meaningful nucleotide representations, these components form a task-driven and efficient hybrid architecture that balances predictive accuracy and computational efficiency. Based on this design, DT-m6A enables accurate identification of m6A sites across multiple cell lines and tissues.

2. Materials and Methods

2.1 Benchmark Dataset

The dataset used in this study was derived from the MST-m6A method, with raw data originally generated by CLSM6A [39]. CLSM6A integrates m6A RNA modification sites from the same cell lines and tissues in *Homo sapiens*, drawing from m6A-Atlas [40]—a comprehensive, high-confidence knowledgebase of experimentally validated m6A sites identified by base-resolution analysis—and maps these sites to the reference genome. Importantly, the resulting RNA sequences are represented in DNA coding format, where uracil (U) is replaced by thymine (T), since all sites are anchored to genomic DNA coordinates. The DRACH motif was then applied to filter candidate sequences. To construct a balanced dataset, the authors generated reliable negative samples for each cell line and tissue according to the following criteria: (i) no negative sample contained any known m6A site; (ii) each negative site was located at least 200 nucleotides away from all positive sites; and (iii) negative sequences did not contain DRACH motifs. Each positive or negative site was represented by a 201-nucleotide sequence centered on adenine (A). Redundant sequences with >80% similarity were removed using CD-HIT [41]. In total, 11 datasets were established, comprising 8 cell lines (A549, CD8T, HCT116, HeLa, HEK293, HEK293T, HepG2, and MOLM13) and 3 tissues (brain, kidney, and liver). Each dataset was randomly split at a 9:1 ratio, with the larger portion used for training and the smaller portion reserved for evaluation. A detailed breakdown of these datasets is provided in Table 1. The table summarizes all 11 datasets, each of which is balanced with an equal number of positive and negative samples. Model training follows a five-fold cross-validation strategy, in which one-fifth of the training portion is used as the validation set in each fold. All sequences are fixed at 201 nucleotides in length; accordingly, a window size of 201 is used during preprocessing and encoding to ensure consistent input representation.

2.2 Construction of DT-m6A Framework

This study proposes a novel deep learning-based model, DT-m6A, designed to accurately predict RNA m6A modification sites across specific cell lines and tissues. As illustrated in Fig. 1, DT-m6A adopts an integrated multi-module architecture consisting of four components: (i) a sequence encoding and initial convolutional module that extracts local features from raw sequences; (ii) a densely connected network (DenseNet) module that enhances feature representation through dense skip connections, enabling efficient feature reuse; (iii) a lightweight Transformer module that captures long-range dependencies; and (iv) a classification module that generates the final predictions. The model combines local and global feature extraction in a computationally efficient manner and leverages conserved motif

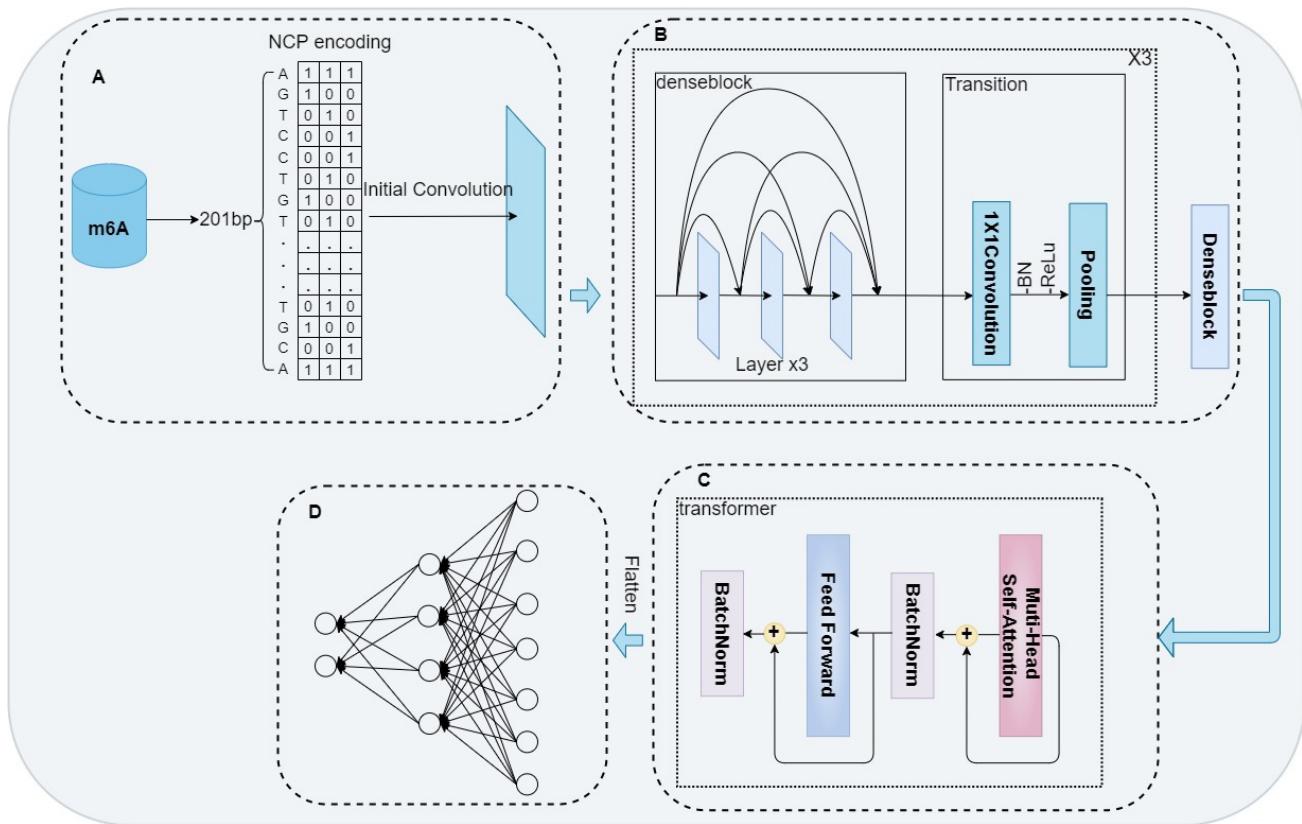


Fig. 1. Overview of the DT-m6A model. (A) RNA sequences are encoded using NCP and fed into the model's initial convolutional layer. (B) DenseNet module: composed of four DenseBlocks and three Transition Layers, responsible for local feature extraction. (C) Transformer module: composed of multi-head self-attention, a feed-forward network, and batch normalization layers. (D) Classification layer: a fully connected layer that outputs the final prediction results. Created using draw.io (<https://app.diagrams.net/>). NCP, nucleotide chemical property; m6A, N6-methyladenosine.

Table 1. Distribution of the benchmark data set.

Dataset	Train		Independent		Total
	Positive	Negative	Positive	Negative	
A549	8418	8400	926	944	18,688
Brain	3521	3532	398	387	7838
CD8T	11,147	11,086	1205	1266	24,704
HCT116	3650	3673	419	396	8138
HEK293	7402	7395	819	826	16,442
HEK293T	39,845	39,925	4472	4392	88,634
HeLa	12,894	12,877	1424	1441	28,636
HepG2	6366	6345	696	717	14,124
Kidney	3044	3079	358	323	6804
Liver	1411	1421	163	153	3148
MOLM13	13,746	13,723	1515	1538	30,522

structures to enable effective transfer of learned features across datasets.

2.2.1 Sequence Encoding and Initial Feature Extraction Module

The input length is fixed at 201 nt because all samples share the same length; this ensures precise alignment and

avoids artificial padding or truncation that may introduce noise. To convert RNA sequences into numerical feature vectors, we employ the nucleotide chemical property (NCP) encoding proposed by Bari *et al.* [42]. Each nucleotide is represented by a three-dimensional vector capturing intrinsic chemical properties: ring structure, hydrogen bond strength, and functional group type. RNA sequences are

expressed using DNA symbols (A, T, C, G), with T replacing U without altering chemical properties. Specifically, pyrimidines (C, T) are encoded as 0 and purines (A, G) as 1 for ring structure; strong (C, G) and weak (A, T) hydrogen bonds are encoded as 0 and 1, respectively; keto (G, T) and amino (A, C) groups are encoded as 0 and 1. These encoding rules can be expressed as:

$$\begin{cases} R(B) \begin{cases} 0, & \text{if } B \in \{C, T\} \\ 1, & \text{if } B \in \{A, G\} \end{cases} \\ H(B) \begin{cases} 0, & \text{if } B \in \{C, G\} \\ 1, & \text{if } B \in \{A, T\} \end{cases} \\ F(B) \begin{cases} 0, & \text{if } B \in \{G, T\} \\ 1, & \text{if } B \in \{A, C\} \end{cases} \end{cases} \quad (1)$$

In this way, each nucleotide is mapped to a three-dimensional feature vector, denoted as:

$$V(B) = [R(B) \ H(B) \ F(B)] \quad (2)$$

where the three dimensions correspond to the nucleotide's ring structure, hydrogen bond strength, and functional group, respectively. Accordingly, the four bases can be represented as:

$$\begin{cases} V(A) = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \\ V(G) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \\ V(T) = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \\ V(C) = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \end{cases} \quad (3)$$

The NCP encoding represents nucleotides A, G, T, and C as dense three-dimensional vectors ($A = [1, 1, 1]$, $G = [1, 0, 0]$, $T = [0, 1, 0]$, and $C = [0, 0, 1]$). Compared with one-hot encoding ($A = [1, 0, 0, 0]$, $G = [0, 1, 0, 0]$, $T = [0, 0, 1, 0]$, and $C = [0, 0, 0, 1]$), NCP provides a more compact and less sparse representation while preserving chemically meaningful similarities between nucleotides. This dense encoding facilitates the learning of sequence motifs relevant to m6A modification and improves feature interactions in downstream models.

Given that each sequence contains 201 nucleotides and each nucleotide is encoded as a three-dimensional vector, an RNA sequence is represented as a 3×201 feature matrix. The dimension "3" corresponds to the three chemical property features and is therefore used as the input channel number of the initial convolutional layer. This feature matrix is then passed through the initial convolutional layer, which employs 32 convolution kernels of size 1×3 to extract preliminary local sequence features, providing the foundation for subsequent deep feature learning.

2.2.2 DenseNet Model

DenseNet (Densely Connected Convolutional Networks) [43] is a deep convolutional architecture with dense connectivity, promoting feature reuse and efficient feature

extraction. In lysine succinylation site prediction, Wang *et al.* [44] proposed MDCAN-Lys, which combines DenseNet with CBAM modules to capture local sequence patterns and emphasize key regions. Similarly, Jia *et al.* [45] developed i5mC-DCGA, integrating an improved DenseNet, Bi-GRU, and self-attention to progressively extract local features, model sequence dependencies, and focus on key sites.

The classical Residual Network (ResNet) [46] mitigates the vanishing gradient problem through skip connections implemented via element-wise addition, where the output of a residual block is defined as:

$$y = F(x) + x \quad (4)$$

However, such additive skip connections may lead to the gradual attenuation of shallow features in deeper layers. To enable more effective feature reuse, DenseNet extends this concept by introducing dense connections, in which the input to the i -th layer is the concatenation of feature maps from all preceding layers:

$$x_i = H_i(x_0, x_1, \dots, x_{i-1}) \quad (5)$$

where H_i denotes a composite transformation consisting of convolution, normalization, and nonlinear activation.

Architecturally, DenseNet comprises DenseBlocks and Transition Layers. DenseBlocks facilitate extensive feature reuse through dense connections, while Transition Layers, placed between adjacent DenseBlocks, reduce feature-map dimensionality and compress channel numbers. The detailed structures of DenseBlocks and Transition Layers are illustrated in Figs. 2,3.

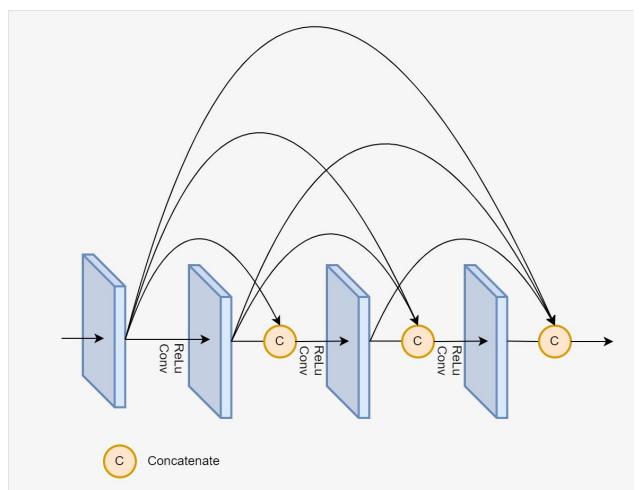


Fig. 2. The structure of DenseBlock. The input to each convolutional layer is constructed by concatenating all feature maps from the previous layers. Created using draw.io (<https://app.diagrams.net>).

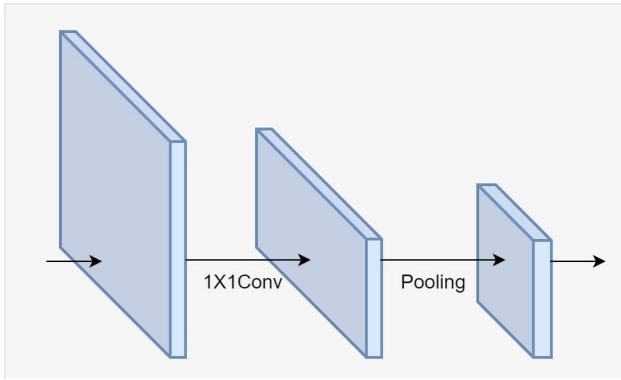


Fig. 3. The structure of Transition Layer. The feature map is compressed in both channel number and spatial dimensions through 1×1 convolution followed by pooling. Created using draw.io (<https://app.diagrams.net/>).

Five-fold cross-validation experiments across all datasets showed a consistent performance improvement (ACC and MCC) when increasing the number of DenseBlock from 1 to 4. This trend suggests that deeper dense feature propagation enhances the model’s ability to capture motif-related patterns, and thus we selected four DenseBlocks as the default configuration. The hyperparameter settings of the DenseBlock are summarized in Table 2. Each dense block comprises 3 convolutional layers and each convolutional layer applied a kernel size of 1×3 to capture local motif patterns and growth rate is 32, while each transition layer consists of a 1×1 convolutional layer (compression rate $\theta = 0.5$) followed by a 1×2 max-pooling layer with a stride of 2 which can compress channel dimensions and reduce redundancy while maintaining computational efficiency.

Table 2. Description of the hyperparameters of the DenseBlock.

Parameters	Number
Dense block	4
Convolution layer number of a dense block	3
Convolution kernel size	3
Growth rate	32
Dropout ratio	0.2

2.2.3 Transformer Module

The Transformer model [47], based on self-attention, efficiently captures long-range dependencies and has been applied in nucleotide modification prediction. In epigenetic research, Fu *et al.* [48] introduced it for m5C site recognition in the trans-m5C framework, using multi-head self-attention and feedforward networks to model complex positional interactions. Building on this, our study incorporates a Transformer module—comprising multi-head self-

attention and a feedforward network (Fig. 1C)—to enhance feature extraction for m6A site prediction.

The multi-head self-attention mechanism, illustrated in Fig. 4, is formulated as:

$$\begin{aligned} \text{MultiHead}(M) &= \text{Concatenate}(\text{Head}_1, \text{Head}_2, \dots, \text{Head}_i) \\ W^o, i &= 1, 2, \dots, h \end{aligned} \quad (6)$$

where $M \in R^{d_{ten} \times d_{model}}$ denotes the input feature map, h is the number of attention heads, and $W^o \in R^{d_{model} \times d_{model}}$ is a learnable projection matrix. Each attention head is computed as:

$$\text{Head}_i = \text{Attention}_i V_i = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right), \text{dim} = -1 V_i \quad (7)$$

where $d_k = d_{model}/h$, and

$$Q_i = M W_i^q, K_i = M W_i^K, V_i = M W_i^v \quad (8)$$

with $W_i^q, W_i^K, W_i^v \in R^{d_{model} \times d_k}$.

Following the attention module, a feedforward neural network (FFN) is applied to further transform the attention-weighted features:

$$FFN(x) = \text{ReLU}(x W_1 + b_1) W_2 + b_2 \quad (9)$$

where $W_1 \in R^{d_{model} \times d_{ff}}, W_2 \in R^{d_{ff} \times d_{model}}$.

Each sublayer is wrapped with a residual connection followed by normalization to improve training stability and preserve original feature information:

$$\text{BatchNorm}(x + \text{Sublayer}(x)) \quad (10)$$

where $\text{Sublayer}(\cdot)$ represents either the multi-head self-attention or the FFN module. The detailed hyperparameter settings of the Transformer are summarized in Table 3.

Table 3. Description of the hyperparameters of the Transformer.

Parameters	Number
Encoder layer	1
Attention heads	5
Dimensions of input	25
Hidden Dimensions of Feedforward Networks	184

2.2.4 Classification Module

The classification module consists of a three-layer multilayer perceptron (MLP), as illustrated in Fig. 1D. Features extracted by the DenseNet and Transformer modules are flattened into a one-dimensional vector, which serves as the input to the first layer of the MLP. The first, second, and

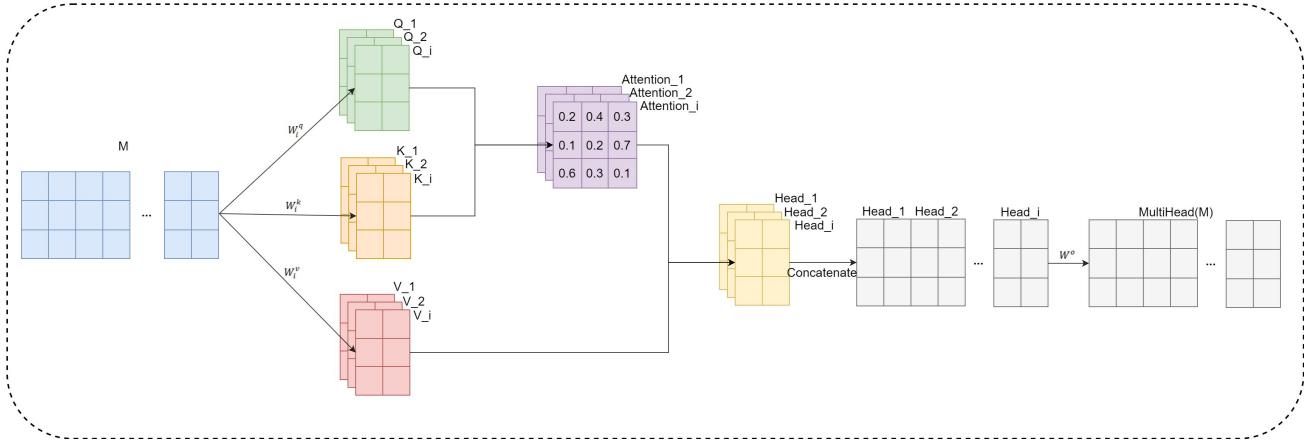


Fig. 4. Multi-head self-attention mechanism. The feature map is first split into multiple sub-feature maps of equal size, which are processed in parallel through self-attention. The outputs are then concatenated to reconstruct a new feature map, which is finally transformed linearly before being output. Created using draw.io (<https://app.diagrams.net/>).

third layers contain 4600, 184, and 2 neurons, respectively. The 4600-dimensional feature vector results from flattening the DenseNet-Trans output tensor of size [batch_size, 184, 25], where 25 is obtained after three pooling operations applied to the 201-nt input ($201 \rightarrow 100 \rightarrow 50 \rightarrow 25$). The MLP first reduces this representation and then projects it back to the 184-channel latent space before outputting class probabilities, ensuring efficient compression without losing dense feature interactions encoded by the CNN [49]. To mitigate overfitting, a dropout layer is applied between the first and second layers. Model training is performed using the AdamW optimizer in conjunction with the cross-entropy loss function, which quantifies the discrepancy between the predicted class probabilities and the ground truth labels. The cross-entropy loss is defined as:

$$Loss = -\frac{1}{N} \sum_i^N \sum_j^M y_{ij} \log(p_{ij}) \quad (11)$$

where N is the number of sequence samples; M is the number of classes; $y_{ij} \in \{0, 1\}$ denotes whether the sample i belongs to class j (with $y_{ij} = 1$ if the sample i belongs to category j , otherwise $y_{ij} = 0$); $p_{ij} \in (0, 1)$ represents the predicted probability of sample i being assigned to class j , and $\log(\cdot)$ is the natural logarithm.

2.3 Performance Evaluation Metrics

To evaluate the performance of the proposed model, several metrics were employed, including accuracy (ACC) [50], precision (PRE), recall (REC), F1 score (F1), and Matthews correlation coefficient (MCC) [51]. These metrics are formally defined as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

$$PRE = \frac{TP}{TP + FP} \quad (13)$$

$$REC = \frac{TP}{TP + FN} \quad (14)$$

$$F1 = 2 \times \frac{PRE \times REC}{PRE + REC} \quad (15)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \quad (16)$$

where TP denotes true positives (the number of m6A sites correctly predicted as positive), TN true negatives (the number of non-m6A sites correctly predicted as negative), FP false positives (the number of non-m6A sites incorrectly predicted as positive), and FN false negatives (the number of m6A sites incorrectly predicted as negative). In addition, for the 5-fold cross-validation experiments, the performance of each fold was further assessed using the area under the receiver operating characteristic curve (AUC), which serves as a standard metric for quantifying the model's classification ability across different folds.

2.4 Training Details

All models were implemented in PyTorch 2.0.0+cu11.8 (Meta AI, Menlo Park, CA, USA) and trained on a single NVIDIA RTX 4080 GPU. The AdamW optimizer was used for all experiments. In five-fold cross-validation, the random seed was fixed to 42. Hyperparameters were selected via randomized grid search, with



Fig. 5. Five-fold cross-validation analysis of different sequence encoding schemes across diverse datasets. Each panel presents the mean performance and corresponding standard deviation for a specific cell line or tissue, including Liver, Brain, Kidney, HEK293, HeLa, CD8T, A549, MOLM13, HEK293T, HCT116, and HepG2. Blue, orange, and green bars represent NCP + one-hot, one-hot, and NCP encoding schemes, respectively. Generated using Python 3.10 (Python Software Foundation, Wilmington, DE, USA).

batch size $\in [16, 32, 64]$, learning rate $\in [0.0001, 0.0002, 0.0003]$, and weight decay $\in [0.01, 0.001, 0.0001]$. The hyperparameter combination yielding the best validation performance was selected. No learning rate scheduler was applied. Training was conducted for up to 50 epochs with early stopping based on validation performance (patience = 10). A dropout rate of 0.2 was applied throughout the network.

3. Results and Discussion

3.1 Analysis of Model Structure

Before being fed into a neural network, nucleotide sequences must be transformed into a computable numerical representation. To identify a suitable encoding scheme, we compared three commonly used representations—one-hot, NCP, and combined NCP + one-hot encoding. Each strategy was evaluated using five-fold cross-validation to assess its impact on model performance (Fig. 5). The results show that NCP encoding achieves slightly higher ACC and

MCC scores across most datasets compared with one-hot and combined NCP + one-hot encoding. Therefore, NCP was adopted in this study as an empirically effective and computationally efficient representation.

In addition, NCP encoding provides a more compact and dense representation than one-hot encoding, enabling the model to capture intrinsic physicochemical properties of nucleotides while reducing input sparsity.

To systematically evaluate the impact of the number of Dense Blocks in the DenseNet architecture on model performance, we conducted a series of comparative experiments by constructing networks with 1, 2, 3, and 4 Dense Blocks. For each configuration, we performed five-fold cross-validation on every cell line or tissue and computed the mean and standard deviation of the performance metrics across all folds. As shown in Fig. 6, the model achieves superior performance across multiple datasets when the number of Dense Blocks is set to 4, with both accuracy (ACC) and Matthews correlation coefficient (MCC) reaching op-

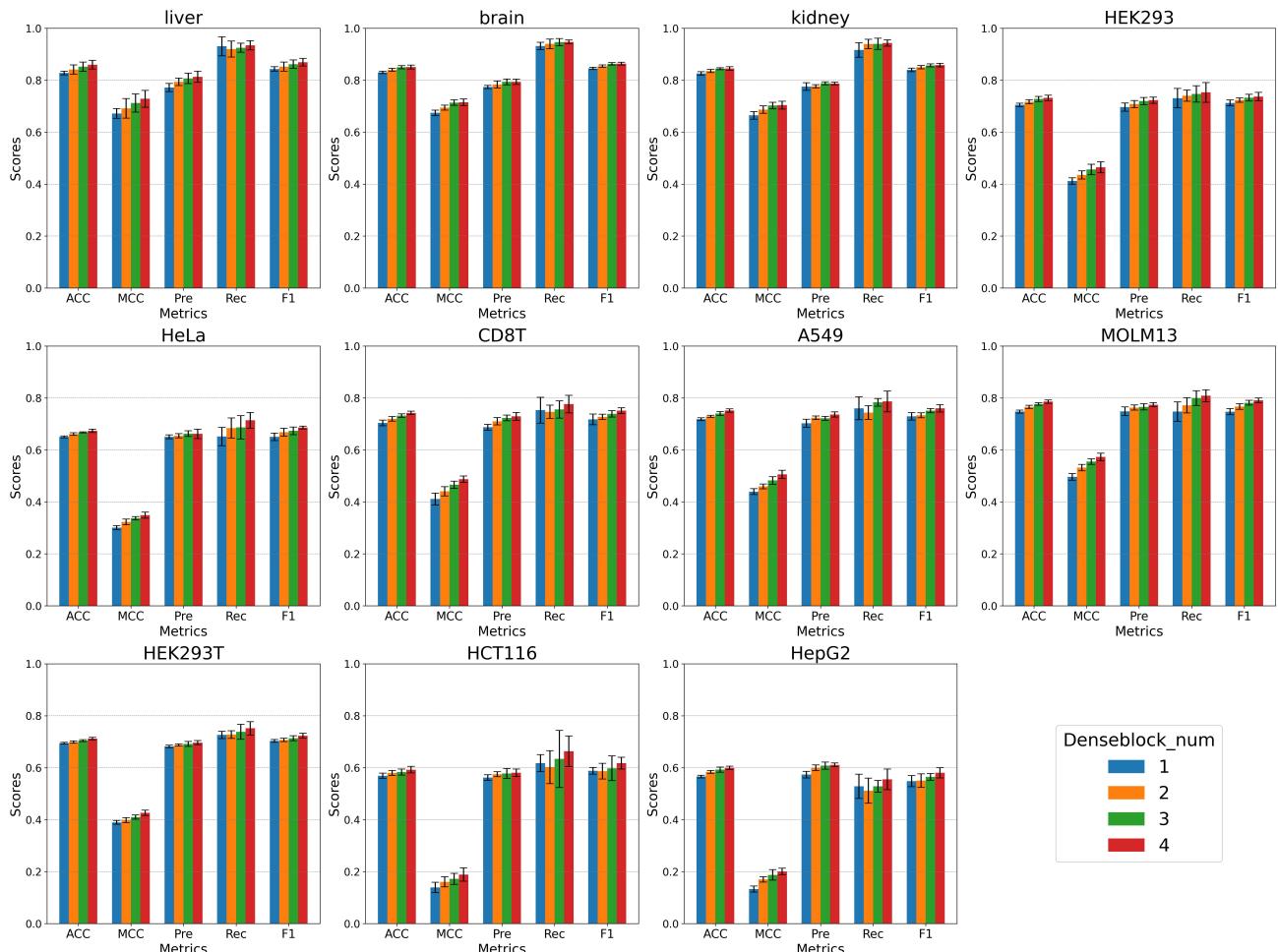


Fig. 6. Five-fold cross-validation analysis of DenseNet architectures with varying numbers of Dense Blocks across diverse cell lines and tissues. Each panel presents the mean performance and corresponding standard deviation for a specific cell line or tissue, including Liver, Brain, Kidney, HEK293, HeLa, CD8T, A549, MOLM13, HEK293T, HCT116, and HepG2. Blue, orange, green, and red bars represent architectures with 1, 2, 3, and 4 Dense Blocks, respectively. Generated using Python.

timal values. Notably, the model exhibits a monotonic improvement in overall performance as the number of Dense Blocks increases from 1 to 4. This trend is reflected in consistent gains in both accuracy (ACC) and the Matthews correlation coefficient (MCC), with particularly pronounced enhancements observed in specific datasets such as liver, HeLa, CD8T, A549, MOLM13, HCT116T, HCT116, and HepG2. Although the improvement in ACC is relatively modest for a few datasets (e.g., brain, kidney and HEK293) when the number of Dense Blocks is set to four, the model overall maintains strong robustness. These results demonstrate that deeper DenseNet architectures with four Dense Blocks are more effective in capturing complex feature representations, thereby enabling more robust and reliable predictions across diverse datasets. However, model complexity must also be considered: as the number of Dense Blocks increases, the parameter size expands substantially, leading not only to higher computational costs but also to a greater risk of overfitting. In addition, an excessive num-

ber of dense blocks will lead to more transition layers, causing the feature maps to be progressively compressed from the initial sequence length of 201 to a single-digit dimension, thereby limiting the amount of information the model can learn. To address this, we set the maximum number of Dense Blocks to four. This configuration preserves strong feature extraction capabilities while mitigating potential drawbacks such as overfitting caused by excessive complexity.

In addition, we compared two commonly used normalization strategies within the Transformer module—LayerNorm and BatchNorm—to assess their suitability for m6A site prediction. As shown in Fig. 7, BatchNorm consistently outperformed LayerNorm across all 11 datasets, with average improvements of 1.9% in ACC and 3.9% in MCC. In our ablation experiments, models using BatchNorm exhibit better performance, providing empirical support for its suitability in this task-specific Transformer design.

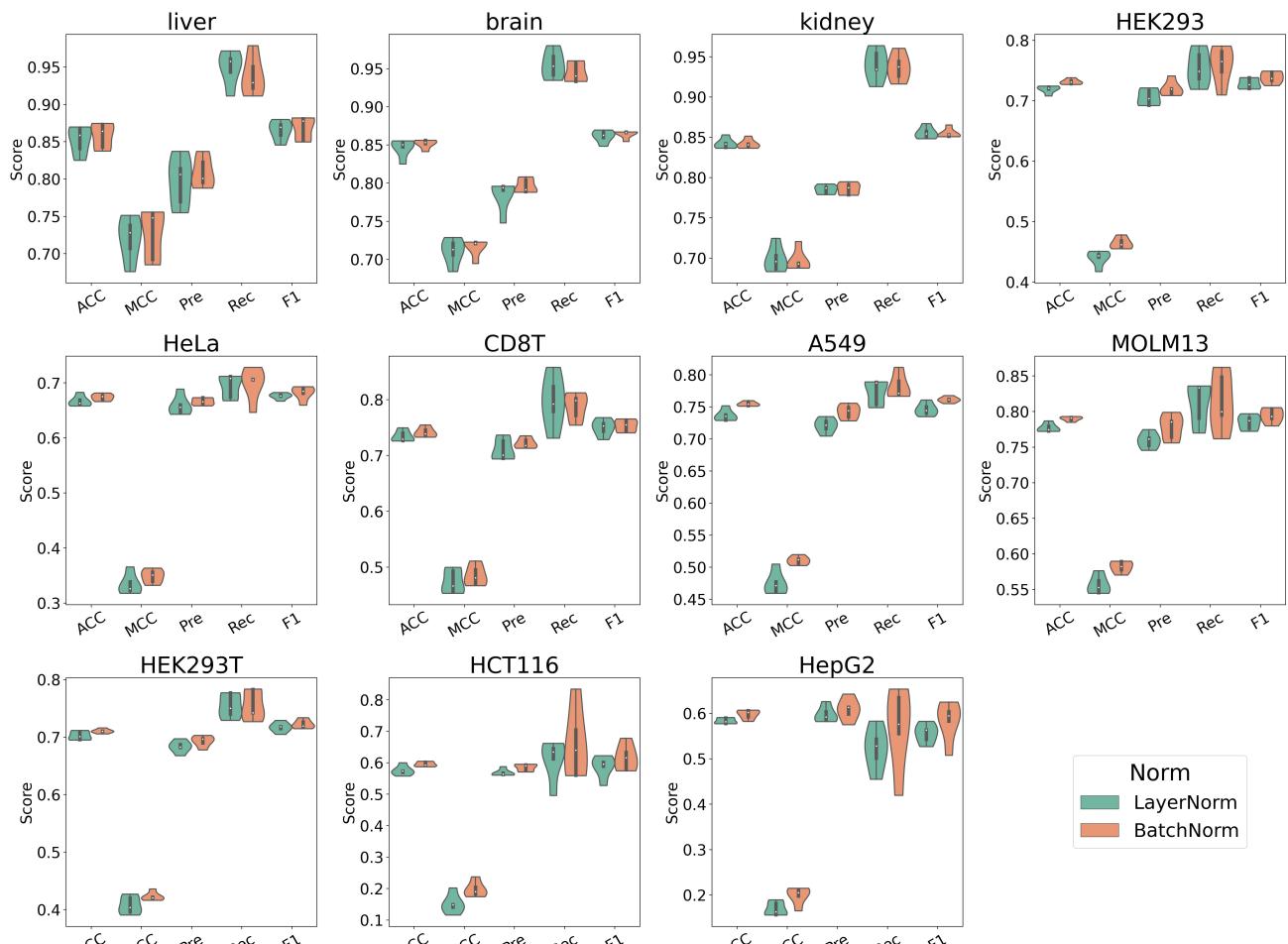


Fig. 7. Five-fold cross-validation analysis of Transformer architectures with different normalization strategies across diverse cell lines and tissues. Each panel reports the distribution of performance metrics for models employing different normalization strategies in a specific cell line or tissue, including Liver, Brain, Kidney, HEK293, HeLa, CD8T, A549, MOLM13, HEK293T, HCT116, and HepG2. Green and orange violins represent LayerNorm and BatchNorm, respectively. Generated using Python.

Our decision to replace LayerNorm with BatchNorm is primarily motivated by the specific characteristics of the m6A prediction task and the overall design of our model. First, all input sequences have a fixed and relatively short length (201 nucleotides). Compared with variable-length natural language inputs, this leads to more consistent feature distributions across samples. Second, our Transformer module is intentionally designed to be lightweight, consisting of only a single encoder layer, which reduces the depth-related instabilities that typically require LayerNorm in deeper Transformer models. Third, the convolutional backbone produces relatively stable intermediate representations, allowing BatchNorm to effectively improve gradient flow and enhance training stability under this setting.

To ensure robustness, five-fold cross-validation was conducted with randomized grid search for hyperparameter tuning, resulting in different batch sizes across folds. Multiple random seeds were used, and training samples were randomly shuffled in each run. Despite these variations in

batch-related settings, the BatchNorm-based Transformer consistently outperformed its LayerNorm-based counterpart across all folds, indicating that the observed improvement is not an artifact of batch effects or sample ordering.

These findings suggest that, under fixed-length and balanced datasets, BatchNorm is a robust and effective normalization strategy for shallow Transformer modules in m6A prediction tasks.

Finally, we also added different attention mechanisms after DenseNet to compare the performance of different attention mechanisms for m6A site prediction. Experimental results (Fig. 8) reveal that Transformer-based attention consistently outperforms both self-attention and CBAM across the majority of datasets, particularly in terms of Recall and F1 score, indicating a stronger capability in capturing the complex contextual dependencies of m6A sequence patterns. While self-attention achieves competitive performance in some datasets (e.g., CD8T and HCT116), its overall stability is inferior to that of Transformer. In contrast,

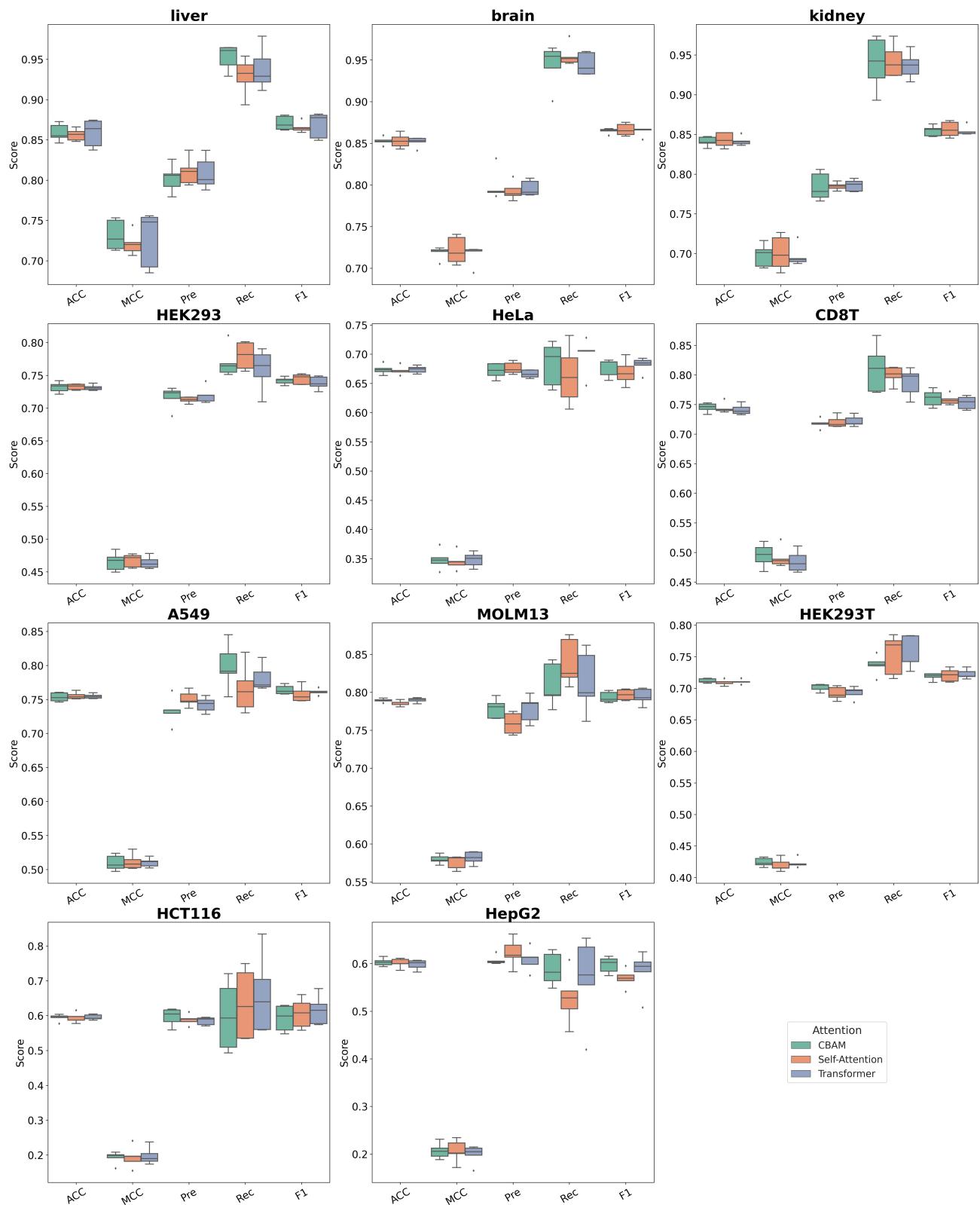


Fig. 8. Five-fold cross-validation analysis of different attention mechanisms integrated with DenseNet across diverse cell lines and tissues. Each panel shows the distribution of performance metrics for models incorporating different attention mechanisms in a specific cell line or tissue, including Liver, Brain, Kidney, HEK293, HeLa, CD8T, A549, MOLM13, HEK293T, HCT116, and HepG2. Green, orange, and blue boxes represent CBAM, self-attention, and Transformer-based attention, respectively. Generated using Python.

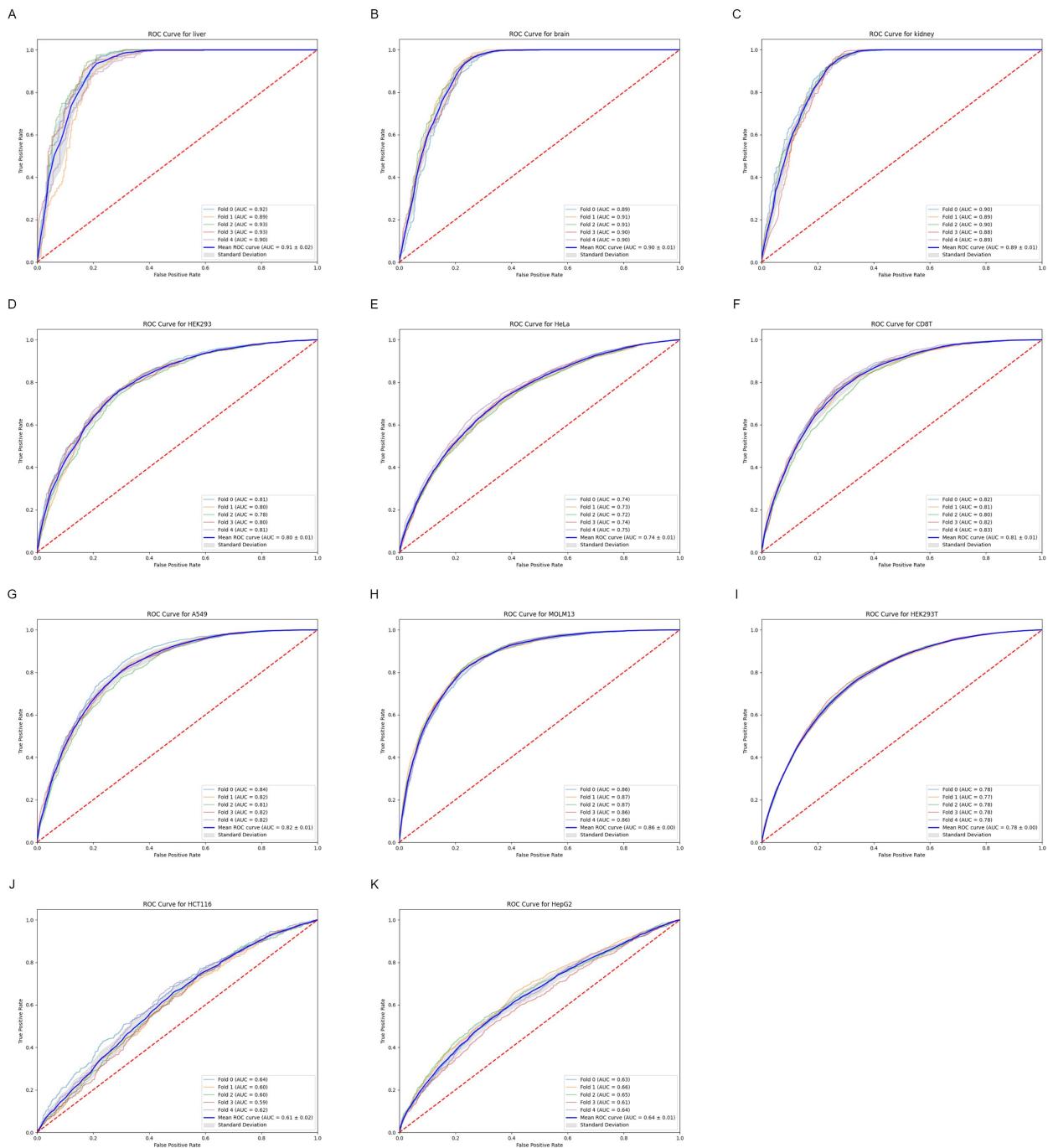


Fig. 9. Five-fold cross-validation was performed on the training datasets across diverse cell lines and tissues, with receiver operating characteristic (ROC) curves evaluated on the validation set of each fold. Each panel displays the ROC curves from the five folds together with the mean ROC curve for a specific cell line or tissue: (A) Liver, (B) Brain, (C) Kidney, (D) HEK293, (E) HeLa, (F) CD8T, (G) A549, (H) MOLM13, (I) HEK293T, (J) HCT116, and (K) HepG2. Generated using Python.

CBAM yields the weakest results, especially in MCC, suggesting its limited ability to discriminate positive and negative sites. These findings highlight that advanced multi-head Transformer attention provides a more suitable inductive bias for biological sequence modeling, as it can better represent long-range dependencies and heterogeneous sequence contexts inherent in m6A modification prediction.

The DenseNet-Transformer combination provides complementary advantages for m6A site prediction. Specifically, the DenseNet module facilitates dense feature reuse and progressive enrichment of local motif representations across layers, enabling the extraction of more informative and hierarchically refined local sequence patterns. In contrast, the Transformer excels at modeling

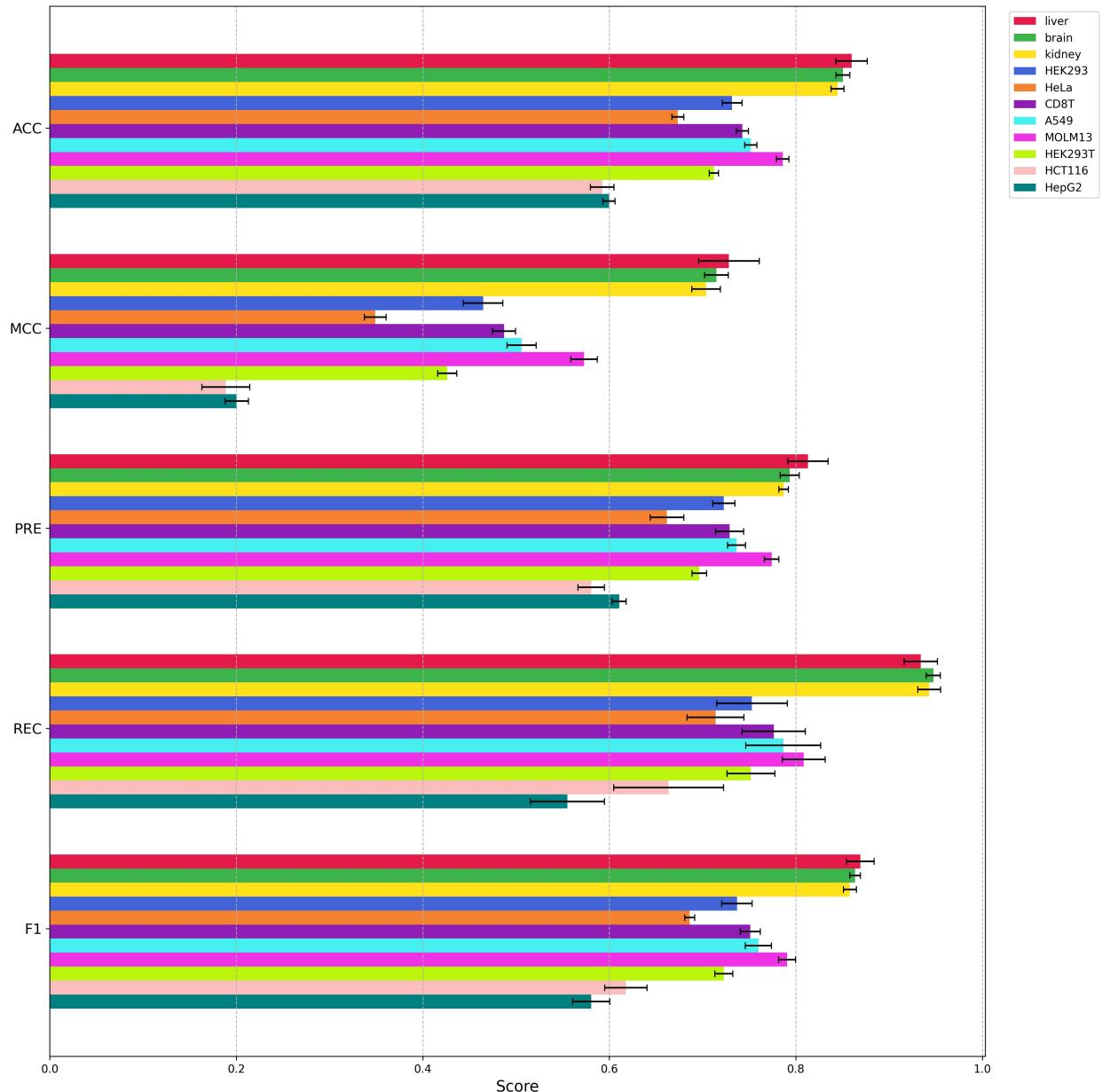


Fig. 10. Comparison of average performance metrics (ACC, MCC, PRE, REC, and F1) with corresponding standard deviations obtained from five-fold cross-validation of the DT-m6A model across eleven datasets. Each colored bar represents a specific tissue or cell line, and the error bars denote the standard deviation across folds. Generated using Python.

long-range dependencies and global contextual interactions that influence methylation outcomes. By providing the Transformer with these deeply fused convolutional representations rather than shallow low-level features, DT-m6A effectively integrates both fine-grained local motif information and broader sequence context. This complementary interaction helps capture the distributed regulatory signals underlying m6A modification.

3.2 Performance of DT-m6A on 5-Fold Cross Validation

To comprehensively evaluate the prediction performance of the DT-m6A model, we conducted five-fold cross-

validation on each dataset. Specifically, each dataset was randomly partitioned into five equal, mutually exclusive subsets (F1–F5). In each iteration, one subset was used as the validation set, while the remaining four subsets were used for training, ensuring that every sample was included in both training and validation at least once. During this process, we recorded the prediction results for each validation subset and plotted the corresponding ROC curves (Fig. 9). The mean and standard deviation of the area under the ROC curve (AUC) across all folds were then calculated to quantify the model's stability and consistency under different data splits. Experimental results demonstrate that

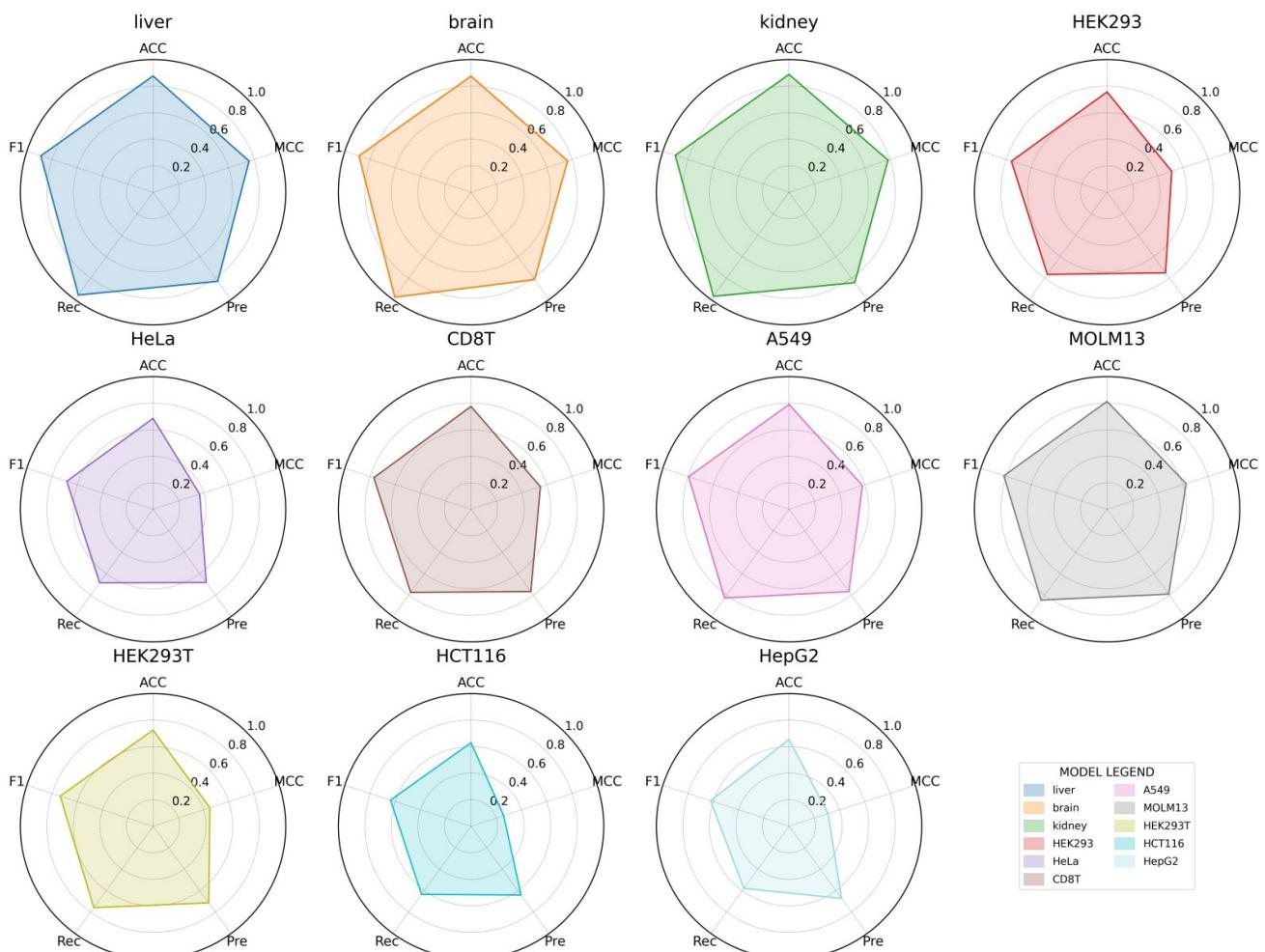


Fig. 11. Performance of DT-m6A across diverse cell lines and tissues, as evaluated using independent datasets. Each panel presents the performance metrics of DT-m6A on a specific cell line or tissue: Liver, Brain, Kidney, HEK293, HeLa, CD8T, A549, MOLM13, HEK293T, HCT116, and HepG2. Generated using Python.

DT-m6A exhibits significant performance variability across different cell types and tissues. Based on AUC values, its performance can be categorized into three tiers: high (AUC $>85\%$, encompassing Liver, Brain, Kidney, MOLM13), moderate ($70\% < \text{AUC} < 85\%$, encompassing HEK293, HeLa, CD8T, A549, HEK293T), and low ($\text{AUC} < 70\%$, encompassing HCT116, HepG2). Notably, DT-m6A demonstrates strong performance on tissue-specific datasets, particularly excelling on the Liver datasets, with the highest AUC among tissue datasets (0.91 ± 0.02). Among all cell lines, the model achieved its best performance on the MOLM13 dataset, with an AUC of 0.86 ± 0.00 . Furthermore, we calculated the mean and standard deviation of ACC, MCC, PRE, ROC and F1 indicators for each data set under five-fold cross validation (Fig. 10). These results collectively indicate that the performance of m6A prediction models is strongly tissue- and cell-type-specific.

3.3 Performance of DT-m6A on the Independent Test Datasets

To rigorously evaluate the robustness capability of the DT-m6A framework, we assessed its performance across multiple independent datasets (Fig. 11). Using a five-fold cross-validation strategy, the model trained in each fold was treated as an individual base learner. Feature representations extracted by each base learner were projected using UMAP (**Supplementary Figs. 1–5**). In addition, we applied K-means clustering to the UMAP-reduced features and computed the corresponding Silhouette scores (**Supplementary Table 1**). Comparatively higher Silhouette values were observed for the liver, brain, and kidney datasets, indicating better feature separability in these tissues, whereas lower Silhouette values were found for cell line datasets, suggesting reduced separability in these cases. Subsequently, a hard-voting ensemble of the five base learners' predictions was employed to produce the final outputs on the independent test sets, followed by a comprehensive performance analysis. The experimental results

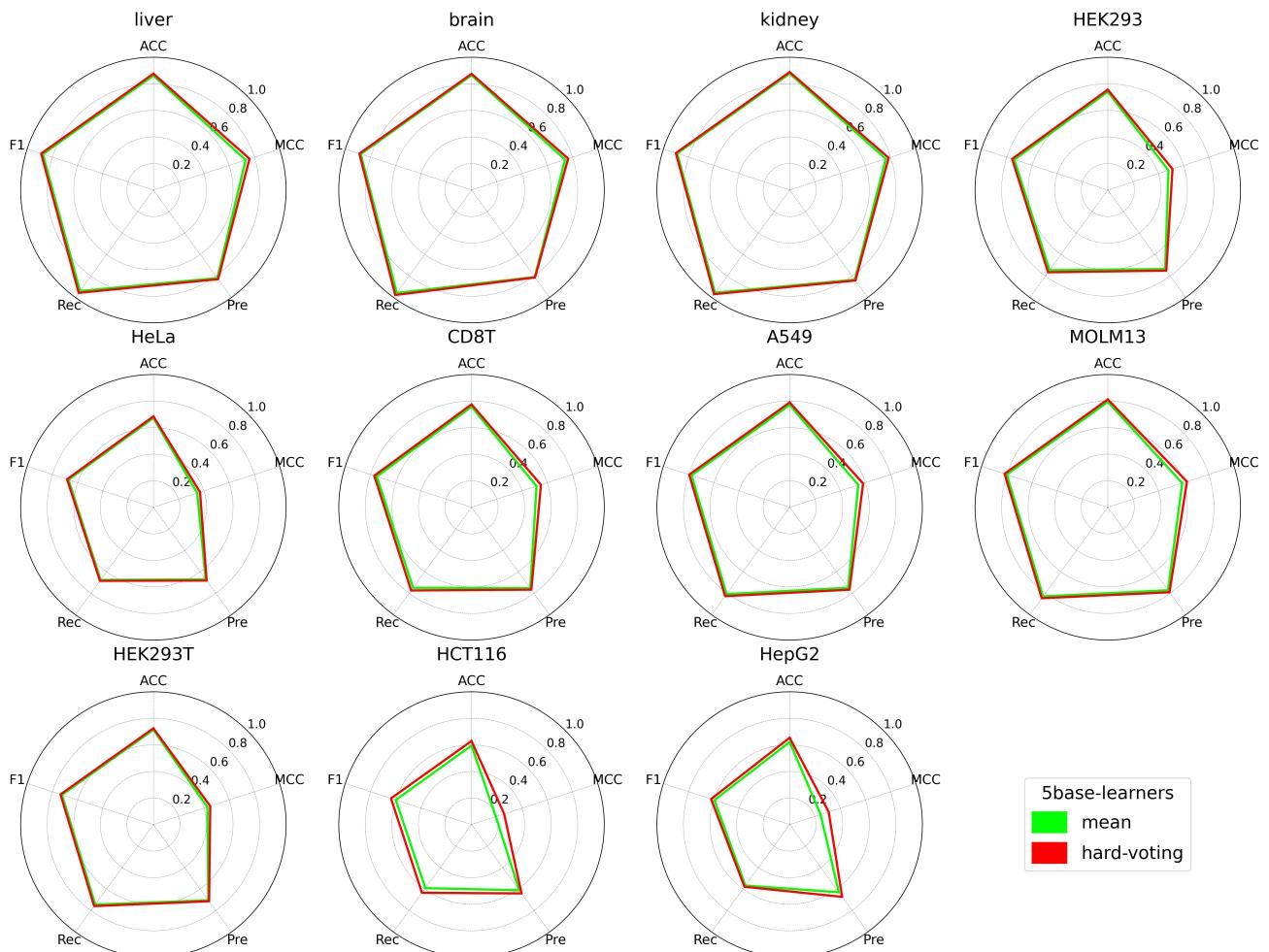


Fig. 12. Performance comparison between the average performance of the five base learners and the ensemble hard-voting results across diverse cell lines and tissues. Each panel presents the performance metrics of the DT-m6A model for a specific cell line or tissue, including Liver, Brain, Kidney, HEK293, HeLa, CD8T, A549, MOLM13, HEK293T, HCT116, and HepG2. Generated using Python.

demonstrate that the model's performance on the independent test set is highly consistent with that observed during cross-validation. In particular, models achieving higher AUC values on validation sets also maintained strong overall performance on independent datasets across multiple metrics, including ACC, MCC, Precision, Recall, and F1-score. Based on classification accuracy (ACC) on the independent test sets, the performance of DT-m6A can be categorized into three tiers: high (ACC >85%, encompassing Liver, Brain, and Kidney tissues), moderate (70% < ACC < 85%, encompassing HEK293, CD8T, A549, MOLM13, and HEK293T cell lines), and low (ACC <70%, encompassing HeLa, HCT116, and HepG2 cell lines). DT-m6A achieves higher performance on tissue-specific datasets (Liver, Brain, and Kidney), moderate performance on several cell-line datasets (HEK293, CD8T, A549, MOLM13, and HEK293T), and relatively lower accuracy on HeLa, HCT116, and HepG2. This performance gap reflects dataset-dependent variability and highlights that prediction difficulty differs across cell lines and tissues.

Interestingly, DT-m6A also demonstrated strong performance on tissue-specific datasets, showing particularly high accuracy on the kidney dataset, where it achieved ACC, MCC, and F1-score values of 0.8884, 0.7845, and 0.9013, respectively. Among all evaluated cell lines, DT-m6A again achieved the best performance on the MOLM13 dataset, with ACC, MCC, and F1 scores of 0.8126, 0.6270, and 0.8176, respectively. Consistent with the validation results observed during training, DT-m6A exhibited distinct performance between HEK293 and HEK293T, despite both being derived from the same parental cell line, achieving ACC values of 0.756 and 0.724, respectively. This strong alignment between validation and independent test performance further underscores the stability and reliability of the DT-m6A framework.

Although the liver, brain, and kidney datasets are substantially smaller than the cell-line datasets, they nevertheless exhibit higher predictive accuracy. To ensure that this observation was not attributable to data leakage or overfitting, we performed additional quality checks. Sequence-

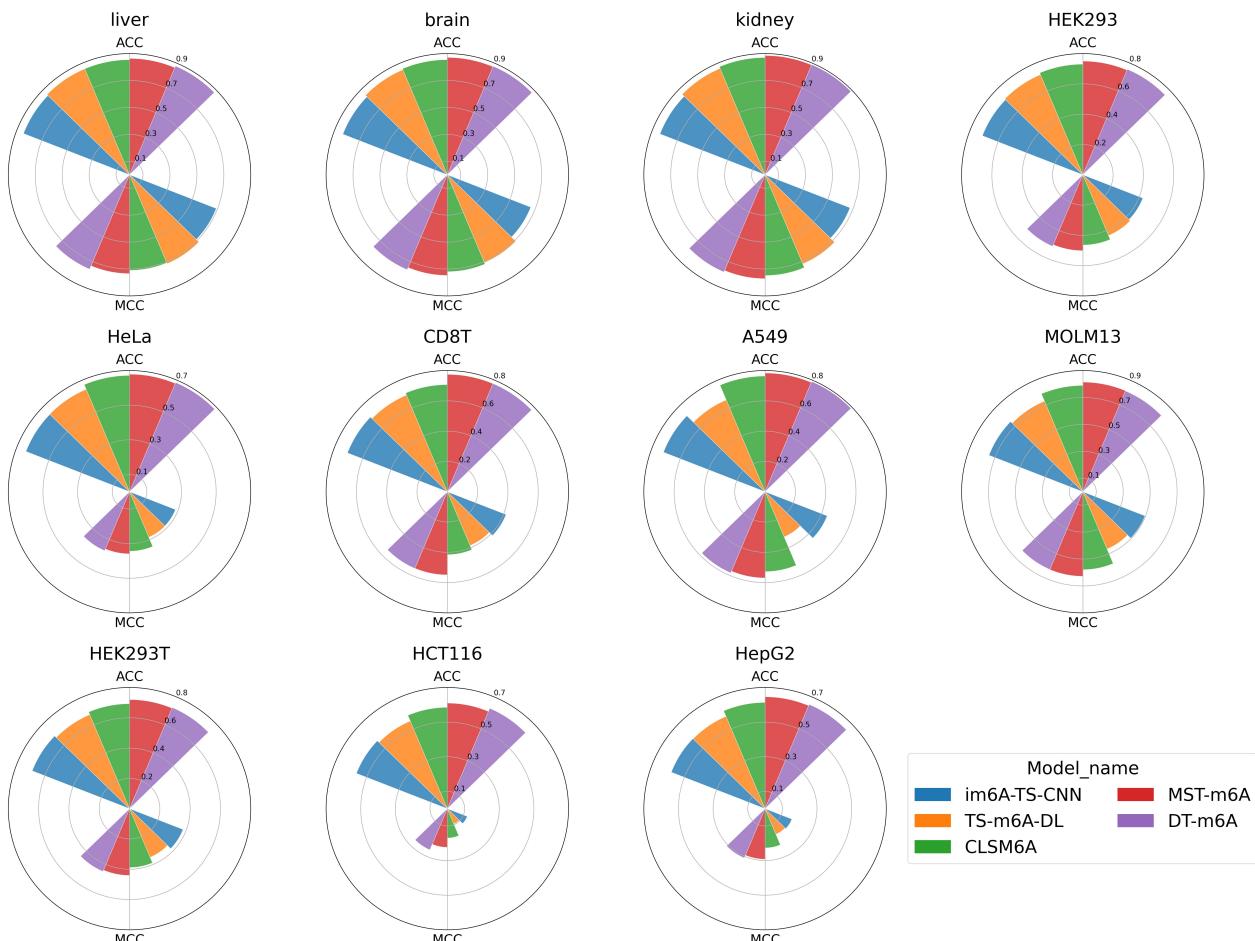


Fig. 13. Performance of DT-m6A on independent test datasets compared with state-of-the-art methods in terms of accuracy (ACC) and Matthews correlation coefficient (MCC). Each panel presents the comparative performance of DT-m6A and state-of-the-art approaches on a specific independent test dataset, including Liver, Brain, Kidney, HEK293, HeLa, CD8T, A549, MOLM13, HEK293T, HCT116, and HepG2. Generated using Python.

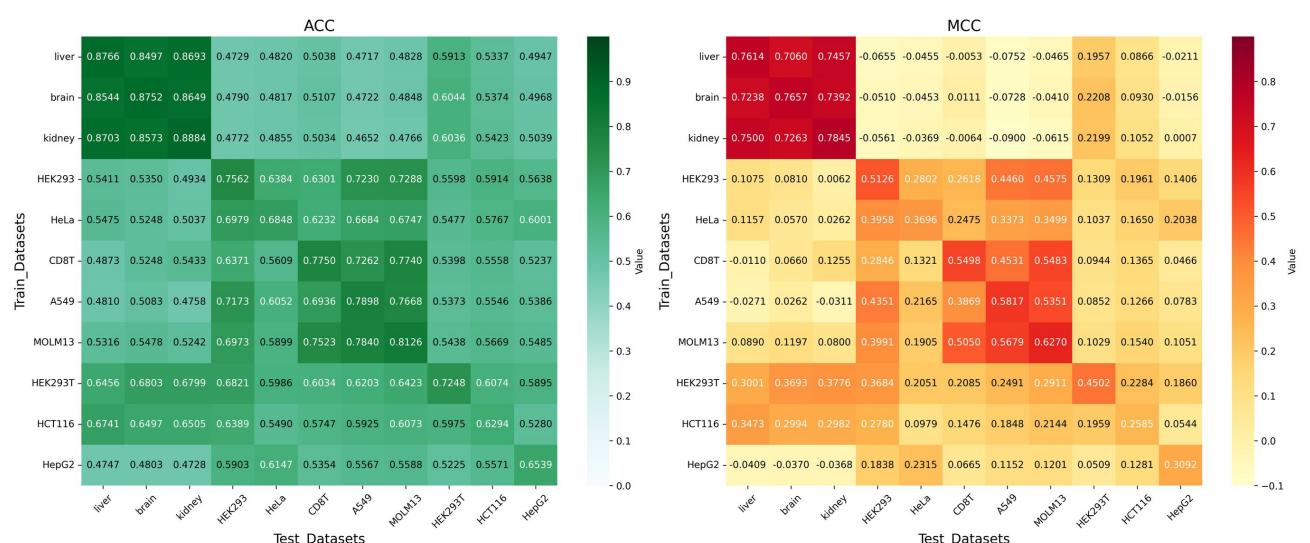


Fig. 14. Tissue and Cell-specific models, which are trained on their respective datasets (columns) and tested on both their own and other independent cell/tissue datasets (rows), have their performance assessed using Matthews correlation coefficient (MCC) and accuracy (ACC). Generated using Python.

Table 4. Performance comparison between DT-m6A and state-of-the-art methods.

Dataset	Method	ACC	MCC	PRE	REC	F1
Liver	CLSM6A	0.8544	0.7099	0.8343	0.8957	0.8639
	MST-m6A	0.8639	0.7342	0.8225	0.9386	0.8767
	DT-m6A	0.8766	0.7614	0.8298	0.9571	0.8889
Brain	CLSM6A	0.8548	0.7197	0.8047	0.9422	0.8681
	MST-m6A	0.8713	0.7476	0.8337	0.9321	0.8801
	DT-m6A	0.8752	0.7657	0.8138	0.9774	0.8881
Kidney	CLSM6A	0.8708	0.7484	0.8277	0.9525	0.8857
	MST-m6A	0.8855	0.7717	0.8645	0.9273	0.8948
	DT-m6A	0.8884	0.7845	0.8422	0.9693	0.9013
HEK293	CLSM6A	0.7301	0.4639	0.7040	0.7900	0.7445
	MST-m6A	0.7502	0.5012	0.7361	0.7765	0.7557
	DT-m6A	0.7562	0.5126	0.7500	0.7656	0.7577
HeLa	CLSM6A	0.6716	0.3435	0.6804	0.6397	0.6594
	MST-m6A	0.6789	0.3588	0.6941	0.6327	0.6620
	DT-m6A	0.6848	0.3696	0.6823	0.6847	0.6835
CD8T	CLSM6A	0.7633	0.5279	0.7422	0.7884	0.7646
	MST-m6A	0.7746	0.5490	0.7664	0.7734	0.7699
	DT-m6A	0.7750	0.5498	0.7666	0.7743	0.7704
A549	CLSM6A	0.7636	0.5274	0.7701	0.7451	0.7574
	MST-m6A	0.7834	0.5692	0.7592	0.8239	0.7902
	DT-m6A	0.7898	0.5817	0.7668	0.8272	0.7958
MOLM13	CLSM6A	0.7894	0.5790	0.7813	0.7993	0.7902
	MST-m6A	0.8140	0.6280	0.8080	0.8198	0.8138
	DT-m6A	0.8126	0.6270	0.7909	0.8462	0.8176
HEK293T	CLSM6A	0.6929	0.3885	0.7183	0.6438	0.6790
	MST-m6A	0.7201	0.4405	0.7121	0.7473	0.7292
	DT-m6A	0.7248	0.4502	0.7138	0.7589	0.7357
HCT116	CLSM6A	0.5853	0.1695	0.5948	0.6062	0.6005
	MST-m6A	0.6098	0.2231	0.6361	0.5632	0.5974
	DT-m6A	0.6294	0.2585	0.6410	0.6348	0.6379
HepG2	CLSM6A	0.6136	0.2267	0.6116	0.5905	0.6009
	MST-m6A	0.6461	0.2919	0.6454	0.6250	0.6350
	DT-m6A	0.6539	0.3092	0.6728	0.5790	0.6224
Average	CLSM6A	0.7445	0.4913	0.7335	0.7630	0.7467
	MST-m6A	0.7634	0.5287	0.7525	0.7782	0.7641
	DT-m6A	0.7697	0.5427	0.7518	0.7976	0.7726
<i>p</i> -value		0.0023	0.0012	0.8778	0.0780	0.0099

Performance comparison between DT-m6A and state-of-the-art methods on independent test datasets, as measured by ACC, precision (PRE), recall (REC), F1 score (F1), and MCC. Bold values indicate the best performance for each metric across diverse cell lines and tissues. The *p*-values represent the statistical significance of performance differences between DT-m6A and MST-m6A across all evaluation metrics.

level deduplication confirmed that there are no overlapping sequences between the training and testing sets. Moreover, the model demonstrated consistently strong and comparable performance on both sets, suggesting that the high accuracy is not caused by overfitting. Motif analysis further revealed that the tissue datasets share highly similar RRACH-like

motif structures, resulting in lower sequence diversity compared with that of the more heterogeneous cell-line datasets. This motif redundancy likely simplifies the prediction task, which plausibly explains the relatively higher accuracy observed on these smaller datasets.

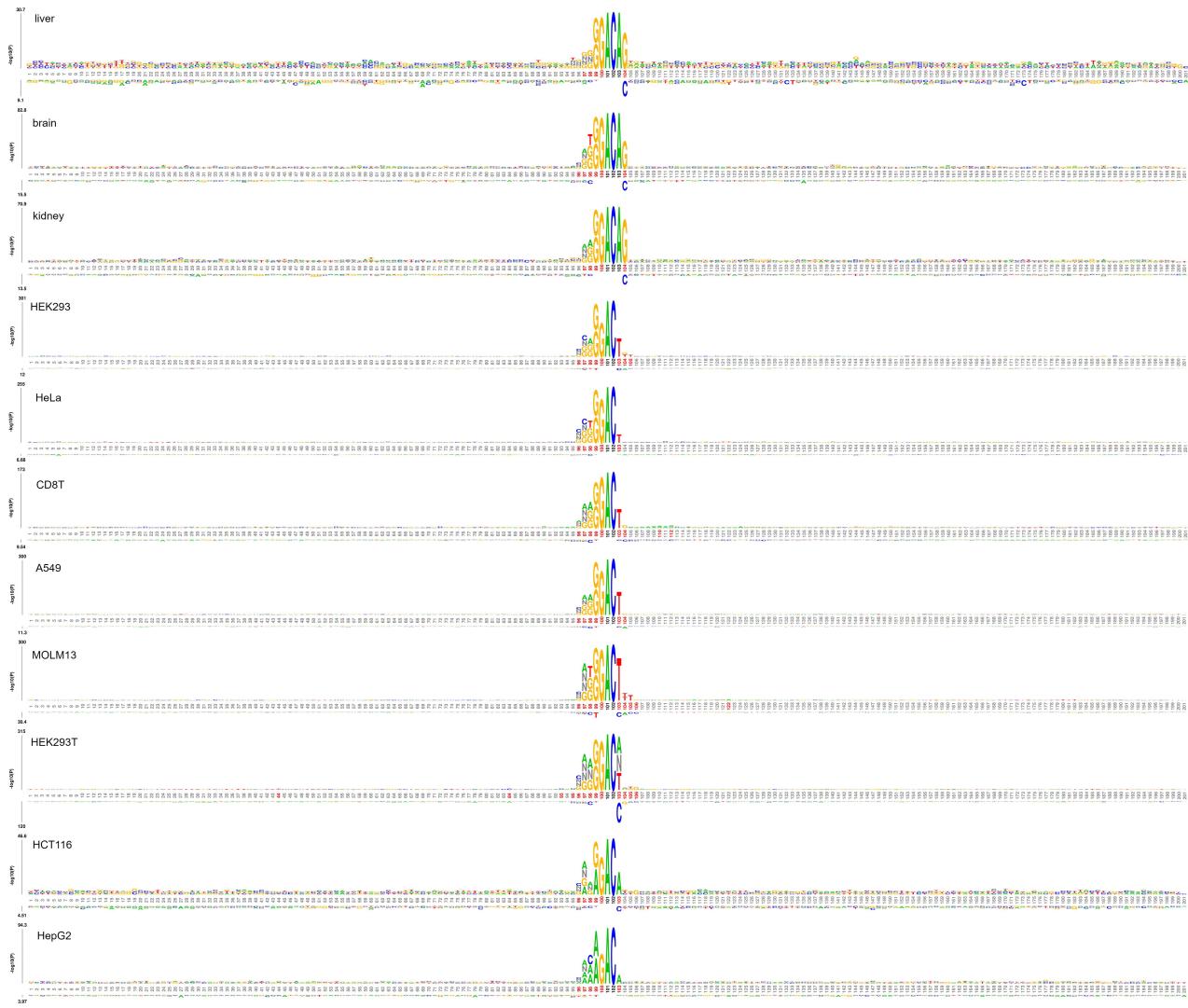


Fig. 15. Visualization of k-mer sequence logos for positive samples across three tissues and eight cell lines (Liver, Brain, Kidney, HEK293, HeLa, CD8T, A549, MOLM13, HEK293T, HCT116, and HepG2) generated by kpLogo. Created using the web tool available at <http://kplogo.wi.mit.edu/manual.html>.

To further examine the effectiveness of our ensemble strategy, we compared the performance obtained by averaging the prediction performance of the five base learners with that produced by the hard-voting ensemble (Fig. 12). The hard-voting approach consistently delivered superior results, indicating that integrating discrete model decisions can better exploit the complementary characteristics learned by different models and enhance overall predictive robustness.

3.4 Performance Comparison Between DT-m6A and the State-of-the-Art Predictors on the Same Independent Datasets

We systematically compared DT-m6A with four state-of-the-art m6A modification site predictors, namely MST-m6A, CLSM6A, im6A-TS-CNN, and TS-m6A-DL. The performance results of these predictors were derived from previous studies [34]. As shown in Table 4, we focused on

comparing the prediction performance of three predictors (DT-m6A, MST-m6A, and CLSM6A) based on five core metrics: accuracy (ACC), Matthews correlation coefficient (MCC), precision (PRE), recall (REC), and F1 score. Additionally, Fig. 13 further presents the comprehensive comparison results between DT-m6A and all four comparative methods regarding the two key metrics of ACC and MCC.

DT-m6A exhibited excellent predictive performance across multiple independent datasets. In tissue-specific datasets (including those of liver, brain, and kidney), the accuracy (ACC) and Matthews correlation coefficient (MCC) of DT-m6A were significantly superior to those of the CLSM6A and MST-m6A methods. Specifically, compared with MST-m6A, DT-m6A improved ACC by 0.29%–1.27% and MCC by 1.28%–2.72%. In comparison with CLSM6A, it enhanced ACC by 1.76%–2.22%, while the improvement in MCC was more significant, reaching 3.61%–5.15%.

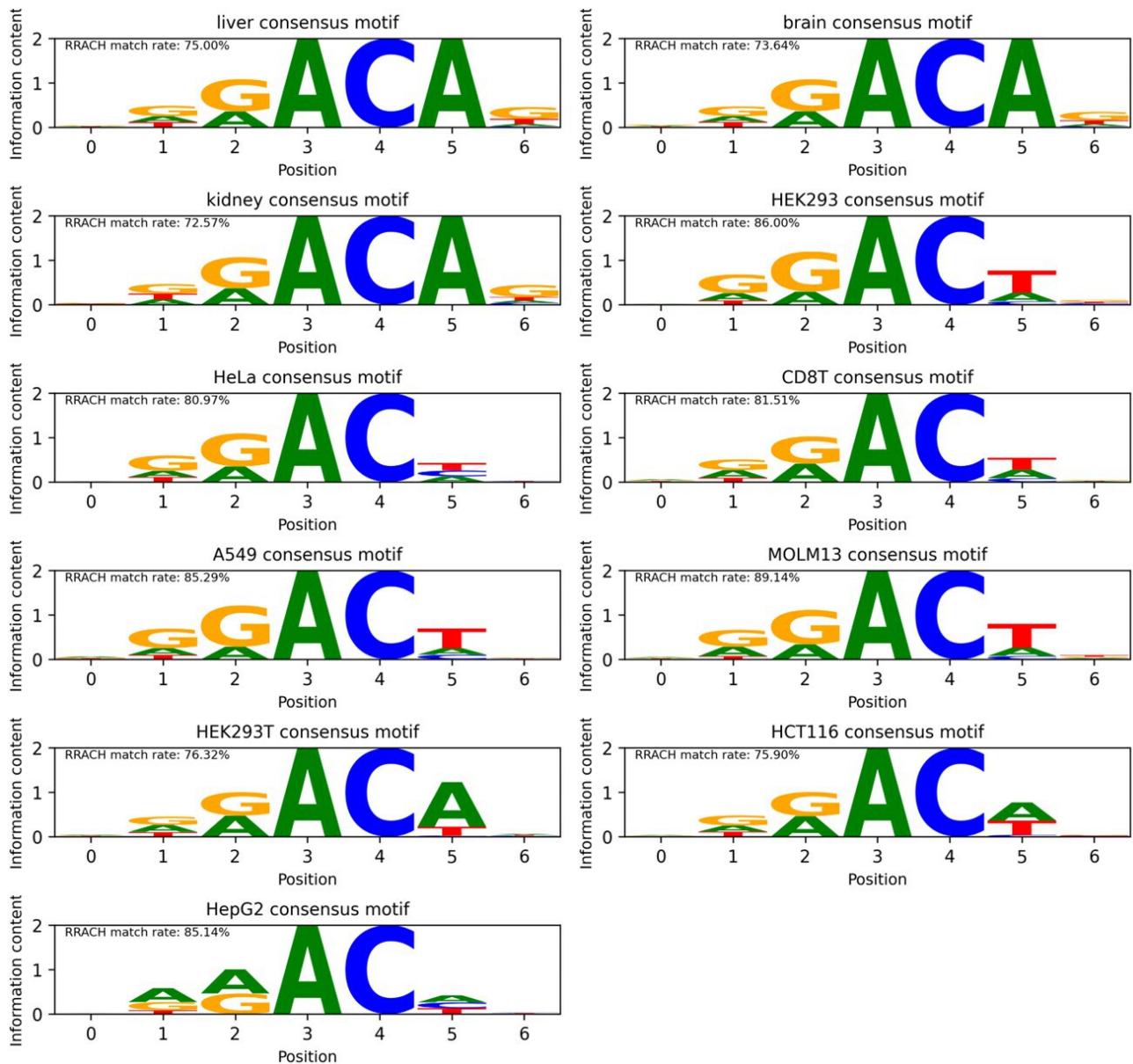


Fig. 16. Distinct m6A sequence motifs identified from the positive samples predicted by the model on tissue and cell line test sets (Liver, Brain, Kidney, HEK293, HeLa, CD8T, A549, MOLM13, HEK293T, HCT116, and HepG2). The motif window was set to 7 bp, and the prediction threshold for positive samples was 0.5. Generated using Python.

DT-m6A also exhibited strong performance in cell line dataset analyses. In the six cell lines (HEK293, HeLa, A549, HEK293T, HCT116, and HepG2), DT-m6A outperformed the other two methods: compared with MST-m6A, its ACC improved by 0.47%–1.96% and MCC by 0.97%–3.54%; relative to CLSM6A, ACC increased by 1.32%–4.41% and MCC by 2.61%–8.9%.

Although DT-m6A achieves consistent improvements on most independent datasets, its performance on the CD8T and MOLM13 datasets is comparable to that of MST-m6A. The comparable performance of DT-m6A and MST-m6A on CD8T and MOLM13 reflects the dataset-dependent effectiveness of sequence-based m6A prediction. DT-m6A

relies exclusively on nucleotide chemical property (NCP) encoding, and sequence-only features may be less discriminative in cell types with higher regulatory complexity. Moreover, the model does not incorporate RNA secondary structure or epigenomic context, which may play a more prominent role in m6A regulation in certain datasets. Potential dataset biases, including differences in sample size, class balance, and experimental noise, may further influence performance comparisons. Finally, although Batch Normalization improves training stability, its reliance on batch-level statistics may limit effectiveness under heterogeneous sequence distributions, partially explaining the comparable performance observed for these datasets.

DT-m6A Sequence Prediction

Select Cell Type:

liver

Enter Sequences (one per line):

Submit Manual Sequences

Or Upload a FASTA File:

Choose File

No file selected

Submit FASTA File

Fig. 17. A web interface for m6A site prediction is available at <https://precious-robust-lynx.ngrok-free.app>.

To ensure a more comprehensive and fair comparison, we re-trained MST-m6A and implemented three additional baseline models, including a CNN-based model, a DenseNet-LSTM hybrid model, and a CNN-Transformer model. All baseline methods were evaluated using the same five-fold cross-validation and ensemble strategy as DT-m6A. As shown in **Supplementary Table 2**, DT-m6A consistently achieves superior performance in terms of ACC and MCC across all datasets compared with the baseline models.

3.5 Cross-Model Evaluation

In this study, we employed an independent training and testing strategy. For each tissue and cell line, a dedicated model was trained on its corresponding dataset to ensure it captured the patterns present in that dataset. To rigorously evaluate the generalization capability of these

models, we performed comprehensive independent tests. Specifically, each trained model was assessed not only on the test set of its source tissue or cell line but also on the test sets of all other tissues and cell lines. This design established a systematic framework for evaluating cross-tissue and cross-cell-line prediction performance.

During model evaluation, we focused on two key performance metrics: accuracy (ACC) and Matthews correlation coefficient (MCC). All ACC and MCC values were compiled into a confusion matrix, as shown in Fig. 14, providing an intuitive visualization of model performance across datasets. Consistently, the highest performance for each test set was achieved by the model trained on the corresponding tissue or cell line. Combined with the motif analysis in Fig. 15, these results indicate that the models exhibit pronounced cross-datasets transferability primarily due to shared sequence motif patterns. In particular, liver,

brain, and kidney sequences contain highly similar motifs, and models trained on these tissues achieve strong cross-tissue predictive performance, with ACC values ranging from 84.97% to 88.84%. Positive sequences from different tissues exhibit highly consistent motif patterns, enabling the model to learn robust and transferable motif representations. Similarly, models trained on five cell lines (HEK293, HeLa, CD8T, A549, and MOLM13) demonstrate reasonably good cross-cell-line transferability, although the ACC of the HeLa test set drops below 60% when evaluated using the CD8T and MOLM13 models. In contrast, the remaining three cell line models show limited transferability to other datasets. Overall, these findings mainly reflect the similarity and differences between datasets: the observed transferability of certain tissue- or cell line-trained models is largely driven by shared sequence motifs rather than underlying biological mechanisms.

3.6 Sequence Analysis of m6A Sites

To further investigate the sequence characteristics underlying m6A site prediction, we employed the kplogo tool [52] to systematically analyze m6A-positive sequence data derived from test sets across multiple tissues and cell lines. The corresponding visualization results are presented in Fig. 15. In the experimental setup, adenine (A) was positioned at the center of each RNA sequence fragment, with 100 nucleotides upstream and downstream.

At the tissue level, the liver, brain, and kidney exhibited highly consistent motif characteristics. Significant enrichment of specific k-mers was observed at positions 99–104, with the enriched k-mer types being largely identical—mainly GG, G, C, and A—while C was markedly depleted at position 104. These findings indicate that the three tissues share a conserved motif composition at m6A sites. Such conservation likely reflects similar methylation recognition mechanisms and may explain the strong cross-tissue transferability of their respective models, which achieved accuracies ranging from 84.97% to 88.84%.

In contrast, at the cell line level, the k-mer enrichment patterns displayed both complexity and regularity across different cell lines. Specifically, HEK293, HeLa, CD8T, A549, and MOLM13 all showed pronounced enrichment of specific k-mers at positions 98–103. At position 98, HEK293, CD8T, and MOLM13 were enriched for AGG, whereas HeLa and MOLM13 showed enrichment for TGG, revealing distinct cell line–specific variations at this site. In contrast, positions 99–103 demonstrated clear commonalities across all five cell lines, with consistent enrichment of k-mers such as GG, G, C, and T, indicating a high degree of motif pattern similarity within this region. This conserved enrichment pattern may underlie the moderate cross-cell-line transferability observed among the corresponding models.

Additionally, the HEK293T, HCT116, and HepG2 cell lines exhibited notable k-mer enrichment signals.

HEK293T displayed pronounced enrichment of GG, G, and C at positions 99–102. Both HCT116 and HepG2 showed enrichment at positions 99–103, though with distinct k-mer preferences: HCT116 was primarily enriched for GA, G, C, and A, whereas HepG2 was enriched for AA, G, C, and A. Moreover, all three cell lines exhibited varying degrees of C depletion at position 103. These variations in motif composition likely contribute to the lower cross-cell-line transferability observed for their respective models.

Building upon these observations, we further examined the sequence motifs learned by the model itself to determine whether they align with experimentally derived patterns. Motif analysis was performed on high-confidence positive samples from the test set (predicted as positive by at least three out of five models). For each predicted site, a short window centered on the candidate position was extracted to characterize its local sequence context. The nucleotide frequencies within this window were summarized into a position frequency matrix and visualized as sequence logos using Logomaker (Fig. 16).

Interestingly, the identified motifs exhibited distinct tissue- and cell line–specific patterns. In liver, brain, and kidney tissues, the identified motifs shared a common sequence containing the ACA motif. This observation is consistent with the results shown in Fig. 14, further demonstrating the strong transferability of the model across these three tissues. Similarly, in the other eight cell lines, consistent common sequences were observed, all of which contained the AC motif—a finding that also aligns with Fig. 14. The main difference between tissue and cell line motifs was observed at position 5, where tissue motifs predominantly featured an A, whereas cell line motifs exhibited greater variability with A, T, or C. Furthermore, over 70% of the predicted positive sequences were consistent with the canonical RRACH motif (R = A/G, H = A/C/T), which is widely recognized as a hallmark of m6A modification. The strong concordance between the motifs enriched in DT-m6A–predicted positive sites and established RRACH patterns supports the biological relevance of the learned sequence features. Notably, a substantial proportion of predicted positives do not strictly conform to the canonical motif, suggesting that DT-m6A does not rely solely on deterministic motif matching but captures broader contextual sequence patterns associated with m6A modification.

3.7 An Available Web Server for DT-m6A

We have developed an online tool (Fig. 17) dedicated to m6A modification site prediction, providing researchers with a convenient and accurate platform for sequence analysis. The website supports prediction for three tissues (liver, brain, kidney) and eight cell lines (HEK293, HeLa, CD8T, A549, MOLM13, HEK293T, HCT116, and HepG2), comprehensively covering a wide range of common biological samples and accommodating diverse experimental scenarios.

In terms of user workflow, the website is designed with both convenience and flexibility in mind. Users first select the relevant tissue type or cell line corresponding to their research material, ensuring accurate background parameter settings for the prediction algorithm. Next, they can flexibly choose an input method based on the data scale: for single or small numbers of RNA sequences, users may directly input sequences of any length via a text box, offering an intuitive and unrestricted experience; for large-scale analyses (such as those involving high-throughput sequencing data), the platform supports FASTA file uploads, enabling efficient batch processing and greatly enhancing analytical throughput.

Once the sequence information is submitted, the system rapidly initiates the analysis and returns results within a short time. Specifically, it automatically calculates and displays the sequence length, providing users with basic sequence feature information, while simultaneously identifying potential m6A modification sites and delivering prediction outcomes with high precision.

4. Conclusion

M6A modification is a form of post-transcriptional modification that plays a vital role in gene expression regulation and metabolism processes. Given its biological importance, we proposed a deep learning model that combines the feature extraction capabilities of DenseNet and the multi-head attention mechanism of Transformer to develop a high-precision m6A prediction tool, aiming to more accurately identify m6A sites and promote epigenetic research. DT-m6A has the following advantages: (1) DT-m6A uses a simple and efficient nucleotide chemical property-based coding scheme (NCP), which consists of discrete 0-1, is intuitive and easy to use, and does not require complex feature engineering. In addition, NCP has more advantages in terms of computational overhead, which helps to improve the training efficiency and prediction performance of the model. (2) DT-m6A utilizes a lightweight DenseNet-Transformer architecture, where DenseNet achieves feature reuse and the Transformer module models long-range dependencies. Innovatively, BatchNorm is used instead of the traditional LayerNorm in the normalization layer of the Transformer, which significantly improves the convergence speed and performance of the model. This hybrid design significantly reduces computational overhead and memory consumption, leading to improved efficiency and scalability in large-scale m6A site prediction tasks. We have also developed an online m6A prediction website. This platform enables real-time access to and application of the DT-m6A model and automatically predicts the corresponding RNA sequence uploaded by the users. Users only need to upload the sequence and click submit to obtain the m6A site prediction results. Our analyses show that the observed cross-datasets generalization is largely driven by the similarity of m6A-associated sequence motifs across tissues and

cell-lines. These findings also imply that future work may explicitly leverage motif-consistent datasets or incorporate motif-aware learning strategies to further enhance the robustness and generalization capacity of m6A site prediction models.

Although DT-m6A exhibits excellent prediction performance, it also has some potential limitations. Specifically, a limitation of this study is that all datasets were derived from the CLSM6A resource. Although CLSM6A includes diverse tissues and cell lines, at present there is no publicly available high-quality m6A dataset with the same sequence-resolution and label format as CLSM6A that would allow a fully consistent evaluation. Therefore, the generalization observed in this study should be viewed as empirical within-dataset portability rather than verified real-world generalization. Future work will incorporate external datasets comparable, once high-quality m6A annotations become available. Its performance on the CD8T and MOLM13 cell lines is comparable to that of the state-of-the-art method MST-m6A, indicating that the model generalization ability across specific cellular contexts still requires improvement. Subsequent research can be expanded in the following directions: First, cutting-edge methods such as self-supervised learning can be used to improve the generalization performance and anti-interference ability of the model by incorporating more comprehensive and diverse datasets. Secondly, we can focus on studying the fusion strategy of RNA secondary structure characteristics and other multimodal data, and use deep learning architectures such as graph neural networks or two-dimensional convolution to systematically explore the intrinsic relationship between nucleic acid sequences, spatial conformations and biological functions, so as to achieve accurate analysis of the spatial localization characteristics of m6A modification sites and their dynamic regulatory networks.

Availability of Data and Materials

Data and source code are publicly available at https://github.com/242004-t/tqy_first.git. The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Author Contributions

Conceptualization, JJ and QT; methodology, JJ and QT; formal analysis, QT; validation, JJ; resources, JJ; data curation, QT; writing—original draft preparation, QT and JJ; writing—review and editing, JJ and QT; visualization, QT; supervision, JJ. Both authors have read and approved the final manuscript. Both authors have contributed sufficiently to the work and agree to be accountable for all aspects of its integrity and accuracy.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

The authors sincerely appreciate the constructive comments and valuable suggestions provided by the reviewers, which have greatly improved the quality of this work.

Funding

This work was partially supported by the Scientific Research Plan of the Department of Education of Jiangxi Province, China (GJJ2400909, GJJ2402711). These funders had no role in the study design, data collection and analysis, decision to publish or preparation of manuscript.

Conflict of Interest

The authors declare no conflict of interest.

Declaration of AI and AI-Assisted Technologies in the Writing Process

During the preparation of this manuscript, the authors used ChatGPT-4 to assist with spelling and grammar checking. The authors subsequently reviewed and revised all content as necessary and take full responsibility for the final version of the manuscript.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at [https://doi.org/10.31083/FBL48029](https://doi.org/10.31083/fbl48029).

References

- [1] Wiener D, Schwartz S. The epitranscriptome beyond m⁶A. *Nature Reviews. Genetics*. 2021; 22: 119–131. <https://doi.org/10.1038/s41576-020-00295-8>.
- [2] He PC, He C. m⁶ A RNA methylation: from mechanisms to therapeutic potential. *The EMBO Journal*. 2021; 40: e105977. <https://doi.org/10.15252/embj.2020105977>.
- [3] Hamar R, Varga M. The role of post-transcriptional modifications during development. *Biologia Futura*. 2023; 74: 45–59. <https://doi.org/10.1007/s42977-022-00142-3>.
- [4] Jiang X, Liu B, Nie Z, Duan L, Xiong Q, Jin Z, *et al.* The role of m6A modification in the biological functions and diseases. *Signal Transduction and Targeted Therapy*. 2021; 6: 74. <https://doi.org/10.1038/s41392-020-00450-x>.
- [5] Murakami S, Jaffrey SR. Hidden codes in mRNA: Control of gene expression by m⁶A. *Molecular Cell*. 2022; 82: 2236–2251. <https://doi.org/10.1016/j.molcel.2022.05.029>.
- [6] Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*. 2012; 149: 1635–1646. <https://doi.org/10.1016/j.cell.2012.05.003>.
- [7] Engel M, Eggert C, Kaplick PM, Eder M, Röh S, Tietze L, *et al.* The Role of m⁶A/m-RNA Methylation in Stress Response Regulation. *Neuron*. 2018; 99: 389–403.e9. <https://doi.org/10.1016/j.neuron.2018.07.009>.
- [8] Yang Y, Hsu PJ, Chen YS, Yang YG. Dynamic transcriptomic m⁶A decoration: writers, erasers, readers and functions in RNA metabolism. *Cell Research*. 2018; 28: 616–624. <https://doi.org/10.1038/s41422-018-0040-8>.
- [9] Li C, Zhu M, Gao C, Lu F, Chen H, Liu J, *et al.* N6-Methyladenosine Regulator-Mediated Methylation Modifica-
- tion Patterns with Distinct Prognosis, Oxidative Stress, and Tumor Microenvironment in Renal Cell Carcinoma. *Frontiers in Bioscience (Landmark edition)*. 2024; 29, 33. <https://doi.org/10.31083/j.fbl2901033>.
- [10] Lee Y, Choe J, Park OH, Kim YK. Molecular Mechanisms Driving mRNA Degradation by m⁶A Modification. *Trends in Genetics: TIG*. 2020; 36: 177–188. <https://doi.org/10.1016/j.tig.2019.12.007>.
- [11] Yang C, Hu Y, Zhou B, Bao Y, Li Z, Gong C, *et al.* The role of m⁶A modification in physiology and disease. *Cell Death & Disease*. 2020; 11: 960. <https://doi.org/10.1038/s41419-020-03143-z>.
- [12] Hong J, Xu K, Lee JH. Biological roles of the RNA m⁶A modification and its implications in cancer. *Experimental & Molecular Medicine*. 2022; 54: 1822–1832. <https://doi.org/10.1038/s12276-022-00897-8>.
- [13] Patil DP, Chen CK, Pickering BF, Chow A, Jackson C, Guttman M, *et al.* m(6)A RNA methylation promotes XIST-mediated transcriptional repression. *Nature*. 2016; 537: 369–373. <https://doi.org/10.1038/nature19342>.
- [14] Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, *et al.* N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature Chemical Biology*. 2011; 7: 885–887. <https://doi.org/10.1038/nchembio.687>.
- [15] Lee JH, Hong J, Zhang Z, de la Peña Avalos B, Proietti CJ, Deamicis AR, *et al.* Regulation of telomere homeostasis and genomic stability in cancer by N⁶-adenosine methylation (m⁶A). *Science Advances*. 2021; 7: eabg7073. <https://doi.org/10.1126/sciadv.abg7073>.
- [16] Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*. 2012; 485: 201–206. <https://doi.org/10.1038/nature11112>.
- [17] Linder B, Grozhik AV, Olarerin-George AO, Meydan C, Mason CE, Jaffrey SR. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nature Methods*. 2015; 12: 767–772. <https://doi.org/10.1038/nmeth.3453>.
- [18] Ke S, Alemu EA, Mertens C, Gantman EC, Fak JJ, Mele A, *et al.* A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes & Development*. 2015; 29: 2037–2053. <https://doi.org/10.1101/gad.269415.115>.
- [19] Zhang Z, Chen LQ, Zhao YL, Yang CG, Roundtree IA, Zhang Z, *et al.* Single-base mapping of m⁶A by an antibody-independent method. *Science Advances*. 2019; 5: eaax0250. <https://doi.org/10.1126/sciadv.aax0250>.
- [20] Meyer KD. DART-seq: an antibody-free method for global m⁶A detection. *Nature Methods*. 2019; 16: 1275–1280. <https://doi.org/10.1038/s41592-019-0570-0>.
- [21] Chen W, Feng P, Ding H, Lin H, Chou KC. iRNA-Methyl: Identifying N(6)-methyladenosine sites using pseudo nucleotide composition. *Analytical Biochemistry*. 2015; 490: 26–33. <https://doi.org/10.1016/j.ab.2015.08.021>.
- [22] Feng P, Ding H, Yang H, Chen W, Lin H, Chou KC. iRNA-PseColl: Identifying the Occurrence Sites of Different RNA Modifications by Incorporating Collective Effects of Nucleotides into PseKNC. *Molecular Therapy. Nucleic Acids*. 2017; 7: 155–163. <https://doi.org/10.1016/j.omtn.2017.03.006>.
- [23] Chen W, Ding H, Zhou X, Lin H, Chou KC. iRNA(m6A)-PseDNC: Identifying N⁶-methyladenosine sites using pseudo dinucleotide composition. *Analytical Biochemistry*. 2018; 561–562: 59–65. <https://doi.org/10.1016/j.ab.2018.09.002>.
- [24] Qiang X, Chen H, Ye X, Su R, Wei L. M6AMRFS: Robust Prediction of N6-Methyladenosine Sites With Sequence-Based Features in Multiple Species. *Frontiers in Genetics*. 2018; 9: 495. <https://doi.org/10.3389/fgene.2018.00495>.
- [25] Chen K, Wei Z, Zhang Q, Wu X, Rong R, Lu Z, *et al.* WHISTLE:

a high-accuracy map of the human N6-methyladenosine (m6A) epitranscriptome predicted using a machine learning approach. *Nucleic Acids Research*. 2019; 47: e41. <https://doi.org/10.1093/nar/gkz074>.

[26] Rehman MU, Tayara H, Chong KT. DL-m6A: Identification of N6-Methyladenosine Sites in Mammals Using Deep Learning Based on Different Encoding Schemes. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 2023; 20: 904–911. <https://doi.org/10.1109/TCBB.2022.3192572>.

[27] Zhang Y, Hamada M. DeepM6ASeq: prediction and characterization of m6A-containing sequences using deep learning. *BMC Bioinformatics*. 2018; 19: 524. <https://doi.org/10.1186/s12859-018-2516-4>.

[28] Xiong Y, He X, Zhao D, Tian T, Hong L, Jiang T, *et al.* Modeling multi-species RNA modification through multi-task curriculum learning. *Nucleic Acids Research*. 2021; 49: 3719–3734. <https://doi.org/10.1093/nar/gkab124>.

[29] Abbas Z, Tayara H, Zou Q, Chong KT. TS-m6A-DL: Tissue-specific identification of N6-methyladenosine sites using a universal deep learning model. *Computational and Structural Biotechnology Journal*. 2021; 19: 4619–4625. <https://doi.org/10.1016/j.csbj.2021.08.014>.

[30] Song Z, Huang D, Song B, Chen K, Song Y, Liu G, *et al.* Attention-based multi-label neural networks for integrated prediction and interpretation of twelve widely occurring RNA modifications. *Nature Communications*. 2021; 12: 4011. <https://doi.org/10.1038/s41467-021-24313-3>.

[31] Mahmoudi O, Wahab A, Chong KT. iMethyl-Deep: N6 Methyladenosine Identification of Yeast Genome with Automatic Feature Extraction Technique by Using Deep Learning Algorithm. *Genes*. 2020; 11: 529. <https://doi.org/10.3390/genes11050529>.

[32] Wang H, Zhao S, Cheng Y, Bi S, Zhu X. MTDeepM6A-2S: A two-stage multi-task deep learning method for predicting RNA N6-methyladenosine sites of *Saccharomyces cerevisiae*. *Frontiers in Microbiology*. 2022; 13: 999506. <https://doi.org/10.3389/fmicb.2022.999506>.

[33] Zhou Y, Zeng P, Li YH, Zhang Z, Cui Q. SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. *Nucleic Acids Research*. 2016; 44: e91. <https://doi.org/10.1093/nar/gkw104>.

[34] Wei L, Chen H, Su R. M6APred-EL: A Sequence-Based Predictor for Identifying N6-methyladenosine Sites Using Ensemble Learning. *Molecular Therapy. Nucleic Acids*. 2018; 12: 635–644. <https://doi.org/10.1016/j.omtn.2018.07.004>.

[35] Wang H, Liu H, Huang T, Li G, Zhang L, Sun Y. EMDLP: Ensemble multiscale deep learning model for RNA methylation site prediction. *BMC Bioinformatics*. 2022; 23: 221. <https://doi.org/10.1186/s12859-022-04756-1>.

[36] Chen J, Zou Q, Li J. DeepM6ASeq-EL: prediction of human N6-methyladenosine (m6A) sites with LSTM and ensemble learning. *Frontiers of Computer Science*. 2022; 16: 162302. <https://doi.org/10.1007/s11704-020-0180-0>.

[37] Su Q, Phan LT, Pham NT, Wei L, Manavalan B. MST-m6A: A Novel Multi-Scale Transformer-based Framework for Accurate Prediction of m6A Modification Sites Across Diverse Cellular Contexts. *Journal of Molecular Biology*. 2025; 437: 168856. <https://doi.org/10.1016/j.jmb.2024.168856>.

[38] Ji Y, Zhou Z, Liu H, Davuluri RV. DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome. *Bioinformatics (Oxford, England)*. 2021; 37: 2112–2120. <https://doi.org/10.1093/bioinformatics/btab083>.

[39] Zhang Y, Wang Z, Zhang Y, Li S, Guo Y, Song J, *et al.* Interpretable prediction models for widespread m6A RNA modification across cell lines and tissues. *Bioinformatics (Oxford, England)*. 2023; 39: btad709. <https://doi.org/10.1093/bioinformatics/cs/btad709>.

[40] Tang Y, Chen K, Song B, Ma J, Wu X, Xu Q, *et al.* m6A-Atlas: a comprehensive knowledgebase for unraveling the N6-methyladenosine (m6A) epitranscriptome. *Nucleic Acids Research*. 2021; 49: D134–D143. <https://doi.org/10.1093/nar/gkaa692>.

[41] Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics (Oxford, England)*. 2012; 28: 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>.

[42] Bari AG, Reaz MR, Choi H-J, Jeong B-S. DNA encoding for splice site prediction in large DNA sequence. In *International Conference on Database Systems for Advanced Applications*. Springer: Berlin, Heidelberg. 2013.

[43] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA. 2017.

[44] Wang H, Zhao H, Yan Z, Zhao J, Han J. MDCAN-Lys: A Model for Predicting Succinylation Sites Based on Multilane Dense Convolutional Attention Network. *Biomolecules*. 2021; 11: 872. <https://doi.org/10.3390/biom11060872>.

[45] Jia J, Lei R, Qin L, Wei X. i5mC-DCGA: an improved hybrid network framework based on the CBAM attention mechanism for identifying promoter 5mC sites. *BMC Genomics*. 2024; 25: 242. <https://doi.org/10.1186/s12864-024-10154-z>.

[46] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA. 2016.

[47] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, *et al.* Attention is all you need. *Advances in neural information processing systems*. 2017; 30. <https://doi.org/10.48550/arXiv.1706.03762>. (preprint)

[48] Fu H, Ding Z, Wang W. Trans-m5C: A transformer-based model for predicting 5-methylcytosine (m5C) sites. *Methods (San Diego, Calif.)*. 2025; 234: 178–186. <https://doi.org/10.1016/j.ymeth.2024.12.010>.

[49] Balasamy K, Suganyadevi S. Multi-dimensional fuzzy based diabetic retinopathy detection in retinal images through deep CNN method. *Multimedia Tools and Applications*. 2025; 84: 19625–19645. <https://doi.org/10.1007/s11042-024-19798-1>.

[50] Le NQK, Nguyen TTD, Ou YY. Identifying the molecular functions of electron transport proteins using radial basis function networks and biochemical properties. *Journal of Molecular Graphics & Modelling*. 2017; 73: 166–178. <https://doi.org/10.1016/j.jmgm.2017.01.003>.

[51] Le NQK, Ou YY. Incorporating efficient radial basis function networks and significant amino acid pairs for predicting GTP binding sites in transport proteins. *BMC Bioinformatics*. 2016; 17: 501. <https://doi.org/10.1186/s12859-016-1369-y>.

[52] Wu X, Bartel DP. kpLogo: positional k-mer analysis reveals hidden specificity in biological sequences. *Nucleic Acids Research*. 2017; 45: W534–W538. <https://doi.org/10.1093/nar/gkx323>.