BMC Bioinformatics

RESEARCH

Open Access

Mke-resnet: a lightweight and interpretable deep learning framework for efficient RNA m6A site identification

Xiao Gao[1], Jianhua Jia[1*], Cong Hui[1] and Yang Lin[1]

*Correspondence:
Jianhua Jia
jjh163yx@163.com
[1]School of Information Engineering, Jingdezhen Ceramic University, Jingdezhen 333403, China

Abstract

Background N6-methyladenosine (m^6A) plays a crucial role in enriching RNA functional and genetic information. While deep learning has advanced m^6A site identification, current state-of-the-art methods, particularly those based on large-scale Pre-trained Language Models (PLMs), often suffer from high computational complexity and over-parameterization, limiting their scalability for genome-wide analysis.

Results In this work, we propose MKE-ResNet, an ultra-lightweight end-to-end framework designed to balance predictive performance with extreme computational efficiency. MKE-ResNet integrates a deep residual network with an innovative Multi-Kernel Efficient Channel Attention (MKE-ECA) module to adaptively capture multi-scale sequence patterns. Extensive experiments on 22 datasets demonstrate that MKE-ResNet achieves competitive performance against complex SOTA models (including MST-m6A). Notably, our model exhibits exceptional stability against sequence noise perturbation and demonstrates superior generalization capability on independent test sets (leading in 8 out of 11 cases in terms of AUROC). Furthermore, interpretability analysis reveals that MKE-ResNet moves beyond the static central motif to capture dynamic sequence patterns resembling RBP binding motifs (e.g., SRSF1 and PUM2). The source code and datasets are available at https://github.com/Gxttk/MKE-Resnet.

Conclusions MKE-ResNet provides a biologically interpretable and computationally efficient solution for m^6A site identification, offering a pragmatic tool for large-scale epitranscriptome screening.

Keywords m^6A modification sites, Deep learning, Lightweight model, Computational efficiency, Model interpretability

Introduction

Among the diverse landscape of RNA modifications in eukaryotic transcriptomes, N^6-methyladenosine (m^6A) stands out as a predominant internal modification [1, 2]. It orchestrates a wide array of RNA metabolic processes, including splicing regulation, nuclear export, translation efficiency, and stability maintenance [3, 4]. The precise mapping of m^6A sites is fundamental to decoding its regulatory mechanisms [5] and

uncovering its implications in various pathologies, including carcinogenesis [6, 7]. While high-throughput sequencing technologies like MeRIP-seq [8] and miCLIP-seq [9] have revolutionized m^6A detection, they remain resource-intensive and dependent on antibody specificity. Similarly, nanopore-based direct RNA sequencing (DRS) has emerged as a powerful alternative, enabling single-molecule detection without PCR amplification [10, 11]. While tools like Tombo and Epinano [12, 13] represent significant progress, robustly distinguishing modified bases from raw ionic current signals remains computationally challenging. Consequently, developing robust and efficient computational predictors has become an imperative strategy to complement experimental efforts for genome-wide m^6A identification.

Deep learning has superseded traditional machine learning approaches [14, 15] by automatically learning high-level abstractions from sequence data. *Pioneering frameworks, such as Deepm6ASeq* [16], *successfully employed Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to capture local and long-range sequence dependencies.* Subsequently, models such as iN6-Methyl [17] and m6A-word2vec [18] applied Natural Language Processing (NLP) techniques, treating RNA sequences as sentences to extract local contextual features. More recently, pre-trained language models (PLMs) have been introduced; for example, MST-m6A [19] utilizes DNABERT embeddings to capture deep semantic information. Another category of tools focuses on tissue-specific prediction, such as m6A-TSHub [20], or fuses diverse encoding schemes like k-mer frequencies (e.g., TS-m6A-DL [21]). Furthermore, prediction accuracy has been enhanced by integrating physicochemical properties—such as Nucleotide Chemical Property (NCP) and Electron-Ion Interaction Potential (EIIP)—into the input layer, a strategy employed by pm6A-CNN [22] and DL-m6A [23]. While some studies rely on streamlined CNN architectures like CLSm6A [24] and AdaptRM [25], more advanced architectures have also emerged; for example, GR-m6A [26] leverages residual networks to capture deeper structural information from spatial molecular features.

However, these methods, while effective, often entail design limitations. For instance, treating RNA simply as text or relying on k-mer encoding may prioritize local sequence information while neglecting broader context. Moreover, most methods that fuse multiple features employ simple concatenation, failing to adaptively assess the relative importance of different biochemical attributes. Critically, the majority of these models use fixed-size convolutional kernels, which introduces an inherent scale bias in learning sequence patterns and makes it difficult to capture biological signals of varying lengths simultaneously. To address this, recent studies have introduced increasingly complex architectures, such as large-scale pre-trained language models (e.g., MST-m6A [19]). While these approaches effectively capture long-range dependencies, this accuracy gain often comes at the cost of massive parameterization and high computational latency, creating a barrier for rapid, large-scale genome-wide screening. Consequently, the field is facing a challenge where the marginal improvement in accuracy cannot justify the exponential growth in model complexity.

Therefore, the core scientific challenge today is designing efficient frameworks that balance high performance with low computational cost and biological interpretability. To address this challenge, we propose a novel end-to-end deep learning framework named MKE-ResNet. Unlike previous parameter-heavy models, MKE-ResNet utilizes

a lightweight design driven by our innovative Multi-Kernel Efficient Channel Attention (MKE-ECA) module. This architecture provides a unified solution to two critical needs: (1) Efficiency: it drastically reduces parameter redundancy while adaptively fusing multi-source features; (2) Interpretability: it decodes black-box predictions by identifying biologically meaningful motifs (validated against RBP binding profiles) rather than merely fitting statistical correlations.

Our extensive experiments on 11 benchmark datasets demonstrate that MKE-ResNet achieves performance comparable to or even better than current state-of-the-art complex models (including PLM-based methods) while remaining significantly more efficient. The main contributions of this paper are summarized as follows:

1. We propose a lightweight deep learning framework, MKE-ResNet, which strikes an excellent balance between performance and efficiency, addressing the computational bottlenecks of existing complex models.
2. We design an innovative Multi-Kernel Efficient Channel Attention module, which simultaneously solves the key problems of adaptive fusion of multi-source features and multi-scale pattern capture within a unified architecture.
3. Extensive experimental results show that our model achieves performance comparable to state-of-the-art methods (e.g., MST-m6A) on multiple benchmark datasets while offering superior interpretability.

Materials and methods

Problem and datasets description

The identification of $\mathrm{m^6A}$ sites is formulated as a binary classification task, where the objective is to predict the methylation status of the central adenosine (A) within a given RNA sequence. To facilitate a rigorous comparison with state-of-the-art methods, we utilized the standard benchmark datasets originally established by Dao et al. [27]. These datasets have been widely adopted as a reliable baseline in recent studies. Crucially, to prevent potential data leakage and ensure rigorous evaluation, high-similarity sequences between the training and testing sets were removed using CD-HIT clustering with a strict threshold of 80%, following the original construction protocol.

The benchmark collection encompasses class-balanced data from three distinct species: *Homo sapiens* (human), *Mus musculus* (mouse), and *Rattus norvegicus* (rat), covering multiple tissue types. Consistent with standard protocols, each sample consists of a 41-nucleotide (nt) sequence window centered on the adenosine of interest. Detailed statistics for each sub-dataset are summarized in Table 1. Furthermore, to assess the generalization capability of our model on unseen data, we employed independent test sets constructed following the protocol described in DL-m6A [23]. These independent sets strictly adhere to the same structural criteria (41-nt length, balanced classes) but originate from different sources than the training benchmarks.

Sequence feature encodings

To comprehensively describe the RNA sequences from various biochemical aspects, we adopted four widely-validated feature encoding schemes as the model's input. This multi-view strategy aims to provide richer information to enhance the model's discriminative power. For a standard input sequence of length $L = 41$, the specific dimensions of each encoding are detailed below:

Table 1 The detailed information of benchmark datasets

Species	Tissues	Name	Positive	Negative
Human	Brain	H-b	4605	4605
	Kidney	H-k	4574	4574
	Liver	H-l	2634	2634
Mouse	Brain	M-b	8025	8025
	Heart	M-h	2201	2201
	Kidney	M-k	3953	3953
	Liver	M-l	4133	4133
	Testis	M-t	4707	4707
Rat	Brain	R-b	2352	2352
	Kidney	R-k	3433	3433
	Liver	R-l	1762	1762

- *1. One-hot Encoding* ($\mathrm{x1\_one\_hot}$): This is the most fundamental sequence representation. The four nucleotides A, C, G, and U are encoded into orthogonal 4-dimensional vectors: [1, 0, 0, 0], [0, 1, 0, 0], [0, 0, 1, 0], and [0, 0, 0, 1], respectively. This encoding preserves the original sequence composition, resulting in a feature matrix of dimension 4×41 (4 channels).
- *2. Nucleotide Chemical Properties* ($\mathrm{x2\_chem}$): This encoding classifies nucleotides based on their chemical structures. According to ring structure (purine/pyrimidine), functional group (amino/keto), and hydrogen bond strength (strong/weak), each nucleotide is mapped into a 3-dimensional chemical property space, thereby capturing the physicochemical properties of the sequence. This produces a feature matrix of dimension 3×41 (3 channels).
- *3. Electron-Ion Interaction Pseudopotential* ($\mathrm{x3\_eiip}$): This is a numerical representation that reflects the energy distribution of free electrons in nucleotides. It assigns a scalar value to each nucleotide, representing its contribution to electron-ion interaction, which is related to the secondary structure and function of the RNA molecule. The resulting dimension is 1×41 (1 channel).
- *4. Enhanced Nucleic Acid Composition* ($\mathrm{x4\_enac}$): This method considers the local composition of the sequence. For each position in the sequence, the frequencies of the four nucleotides within an upstream window are calculated, thus combining positional information with local statistical information. This encoding yields a feature matrix of dimension 4×41 (4 channels).

For each 41-nt RNA sequence, the four encoding methods generate four feature matrices of distinct dimensions. These four matrices serve as the inputs to the four parallel feature extraction channels of MKE-ResNet.

Model architecture

The proposed MKE-ResNet utilizes a multi-stream deep learning framework designed to extract and fuse high-level representations from heterogeneous feature encodings. As illustrated in Fig. 1, the workflow comprises three key stages: (1) Parallel Multi-stream Feature Extraction, (2) Multi-scale Feature Fusion and Refinement via the MKE-ECA module, and (3) Classification.

Let the four input feature matrices be denoted as $\mathbf{X}_i \in \mathbb{R}^{C_i \times L}$, where $i \in \{1, 2, 3, 4\}$ corresponds to One-hot, Chemical Properties, EIIP, and ENAC encodings, respectively, and $L = 41$ is the sequence length.

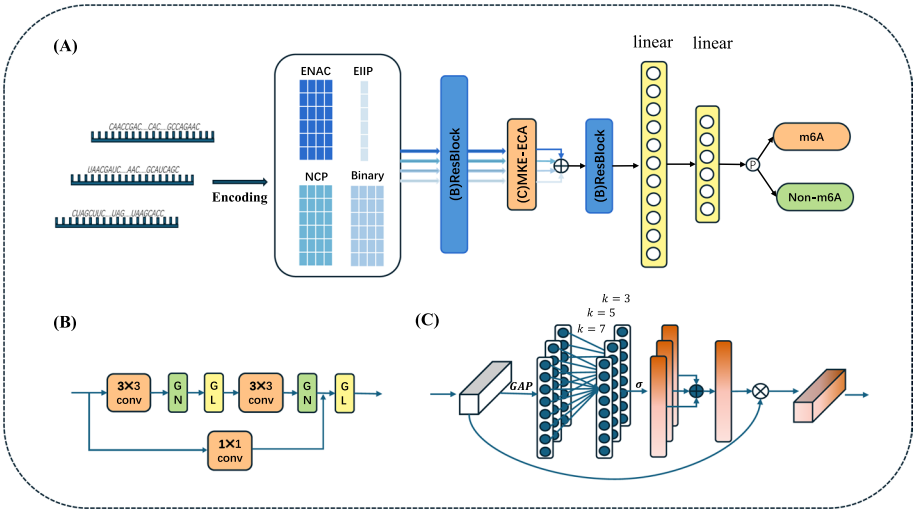

Fig. 1 The overall architecture of the proposed MKE-ResNet framework

Parallel multi-stream feature extraction

Unlike simplistic concatenation strategies, we employ separate convolutional streams to process specific encoding types, preventing immediate semantic interference. Each stream Φ_i consists of a stack of Residual Blocks (ResBlocks) and Max-Pooling layers to downsample the temporal dimension and expand the channel dimension:

$$\mathbf{F}_i = \Phi_i(\mathbf{X}_i), \quad \mathbf{F}_i \in \mathbb{R}^{C' \times L'} \tag{1}$$

where C' and L' denote the transformed channel and sequence dimensions after the parallel extraction stage. Specifically, separate ResBlocks are applied to extract local patterns, followed by pooling operations to abstract the features.

Residual block formulation

To mitigate the vanishing gradient problem and facilitate the propagation of features, we employ Residual Blocks (ResBlock) as the fundamental building unit. For an input tensor $\mathbf{Z}_{in}$, the ResBlock output $\mathbf{Z}_{out}$ is defined as:

$$\mathbf{Z}_{out} = \mathrm{GELU}(\mathcal{F}(\mathbf{Z}_{in}, \{W_j\}) + \mathcal{W}_s(\mathbf{Z}_{in})) \tag{2}$$

where $\mathcal{F}(\cdot)$ represents the residual mapping function involving two stacked 1D convolutional layers, employing Group Normalization (GN) and GELU activation. The operation allows the network to learn identity mappings effectively:

$$\mathcal{F}(\mathbf{Z}) = \mathrm{GN}(\mathrm{Conv}_2(\mathrm{GELU}(\mathrm{GN}(\mathrm{Conv}_1(\mathbf{Z}))))) \tag{3}$$

The term $\mathcal{W}_s(\mathbf{Z}_{in})$ represents the shortcut connection. If the dimensions of the input and output match, $\mathcal{W}_s$ is an identity mapping; otherwise, a 1×1 convolution is applied to align the spatial and channel dimensions for element-wise addition.

Feature fusion and MKE-ECA module

The extracted features from four streams are concatenated along the channel dimension to form a unified semantic tensor $\mathbf{F}_{merged}$:

$$\mathbf{F}_{merged} = \mathrm{Concat}([\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3, \mathbf{F}_4]) \tag{4}$$

This merged tensor is further processed by additional ResBlocks and pooling layers. To adaptively emphasize informative features and suppress noise, we introduce the *Multi-Kernel Efficient Channel Attention (MKE-ECA)* module, which sequentially integrates Channel Attention and Multi-Scale Spatial Attention.

1. *Channel Attention (SE-Block)* We first apply a Squeeze-and-Excitation mechanism to model inter-channel dependencies. A global average pooling (GAP) generates a channel descriptor $\mathbf{z} \in \mathbb{R}^{C_{merge}}$, which is then processed by a multi-layer perceptron (MLP) with a reduction ratio $r = 16$:

   $$\mathbf{w}_c = \sigma(\mathbf{W}_2 \cdot \mathrm{ReLU}(\mathbf{W}_1 \cdot \mathrm{GAP}(\mathbf{F}))) \tag{5}$$

   where σ denotes the Sigmoid function, and $\mathbf{W}_1, \mathbf{W}_2$ are learned weights. The channel-refined feature is obtained via element-wise multiplication: $\mathbf{F}' = \mathbf{F} \odot \mathbf{w}_c$.
2. *Multi-Scale Spatial Attention* To capture context-aware spatial dependencies, we propose a multi-branch architecture. We first aggregate channel information via average pooling and max pooling across the channel dimension, generating spatial descriptors $\mathbf{S}_{avg}, \mathbf{S}_{max} \in \mathbb{R}^{1\times L}$. These are concatenated and fed into three parallel convolutional branches with varying kernel sizes $k \in \{3, 5, 7\}$:

   $$\mathbf{M}_k = \mathrm{Conv}_{1d}^{k}([\mathbf{S}_{avg}; \mathbf{S}_{max}]), \quad k \in \{3, 5, 7\} \tag{6}$$

   The multi-scale spatial maps are fused via a 1×1 convolution and normalized by a Sigmoid function to generate the final spatial attention map $\mathbf{M}_{spatial}$:

   $$\mathbf{M}_{spatial} = \sigma(\mathrm{Conv}_{1\times1}([\mathbf{M}_3; \mathbf{M}_5; \mathbf{M}_7])) \tag{7}$$

   The final attention-refined output is given by $\mathbf{F}'' = \mathbf{F}' \odot \mathbf{M}_{spatial}$.

Classification module

The recalibrated feature map $\mathbf{F}''$ is flattened into a vector $\mathbf{v} \in \mathbb{R}^d$. The final prediction is generated by a classifier consisting of fully connected (FC) layers with Batch Normalization (BN) and Dropout regularization:

$$\hat{y} = \mathrm{Softmax}(\mathbf{W}_{out} \cdot \mathrm{ReLU}(\mathrm{BN}(\mathbf{W}_{hidden} \cdot \mathbf{v}))) \tag{8}$$

This architecture enables MKE-ResNet to learn hierarchical representations from local motifs to global semantic patterns effectively.

Experimental setup

To objectively evaluate the model's performance, we employed a 5-fold cross-validation scheme for each dataset. For the benchmark datasets, we report the average performance of the 5-fold cross-validation. For the independent test sets, we adopted an ensemble learning strategy: the five models trained during the 5-fold cross-validation process were combined into a voting ensemble, and the performance was evaluated based on the final predictions of this ensemble model.Crucially, to ensure a strictly fair comparison, all comparative baseline models (including MST-m6A, m6A-TSHub, etc.)

were re-trained and evaluated on the identical 5-fold cross-validation splits and independent test sets generated for MKE-ResNet. This rigorous protocol eliminates performance variations arising from different data partitioning schemes.All models were implemented using the PyTorch 2.0.1 framework with CUDA 11.7 and were trained and tested on a single NVIDIA GeForce RTX 4060 Ti GPU. To determine the optimal hyperparameters, we conducted a grid search based on the performance of the validation set. The search space for the batch size included $\{64, 128, 256, 512\}$, and the learning rate was selected from $\{1e{-}3, 1e{-}4, 1e{-}5\}$. Consequently, the batch size was set to 64, and the initial learning rate was set to 0.001 ($1e{-}3$). We utilized the Adam optimizer with a weight decay of $1e{-}5$, employing the binary cross-entropy (BCE) loss function. The training process was capped at 100 epochs. To ensure training stability and efficiency, we employed a ReduceLROnPlateau scheduler (patience=10) and an Early Stopping mechanism (patience=20) to prevent overfitting. A Model Checkpointing strategy was strictly applied to save and evaluate only the model weights that achieved the best performance on the validation set.

Evaluation metrics

To comprehensively measure the predictive capability of our model, we adopted six widely-used binary classification metrics:

- Accuracy (Acc): Measures the proportion of correctly classified samples.

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

- Matthews Correlation Coefficient (MCC): Takes into account all four classification outcomes and is considered a more robust and balanced metric, especially for imbalanced datasets. Its value ranges from -1 to $+1$.

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

- Sensitivity (Sn) / Recall: Measures the proportion of actual positive samples that are correctly identified.

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

- Specificity (Sp): Measures the proportion of actual negative samples that are correctly identified.

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

- Area Under the ROC Curve (AUROC): Quantifies the model's discriminative ability across all classification thresholds. It plots the True Positive Rate against the False Positive Rate.

- Area Under the Precision-Recall Curve (AUPR): Reflects the trade-off between precision and recall, providing a reliable performance measure particularly for imbalanced datasets.

Where TP (True Positives), TN (True Negatives), FP (False Positives), and FN (False Negatives) represent the number of true positives, true negatives, false positives, and false negatives, respectively.

Result

Performance comparison on benchmark datasets

To systematically evaluate the performance of our proposed MKE-ResNet model, we conducted a comprehensive comparison with five state-of-the-art (SOTA) and classic baseline methods across 22 datasets (11 benchmark and 11 independent test sets). The compared methods include the recently proposed PLM-based *MST-m6A* [19] (representing the current SOTA), the Transformer-based *m6A-TSHUB* [20], the efficient residual network *GR-m6A* [26], as well as classic deep learning models *Deepm6ASeq* [16] and *pm6A-CNN* [22]. All methods were evaluated using a 5-fold cross-validation scheme, with Accuracy (Acc) and Matthews Correlation Coefficient (MCC) as the primary metrics.

The detailed comparative results are presented in Table 2. In the cross-validation tests on the 11 benchmark datasets, MKE-ResNet demonstrated strong competitiveness against the heavyweight MST-m6A, achieving the highest Acc and MCC scores on 7 of them ($\mathrm{H-b}$, $\mathrm{H-l}$, $\mathrm{M-b}$, $\mathrm{M-h}$, $\mathrm{M-l}$, $\mathrm{M-t}$, and $\mathrm{R-b}$).

To further assess the model's generalization ability, which better reflects its performance on unseen data, we conducted validations on the independent test sets. The results reveal a more pronounced advantage for MKE-ResNet. It outperformed the SOTA model MST-m6A on 8 of the 11 independent test sets. Notably, MKE-ResNet consistently surpasses the baseline even in challenging scenarios. For instance, on the '$\mathrm{iH-b}$' dataset, the MCC of MKE-ResNet reached 0.510, outperforming MST-m6A (0.498); on the '$\mathrm{iM-h}$' dataset, the MCC achieved 0.555, showing a clear margin over MST-m6A (0.525). This series of results strongly demonstrates the excellent generalization performance and robustness of MKE-ResNet. We attribute this performance advantage to its unique architectural design: the deep residual network effectively extracts high-level representations of sequence features, and our innovative multi-kernel attention mechanism enables the model to more intelligently integrate information from different encoding schemes by modeling multi-scale channel dependencies, thereby amplifying key discriminative features.

Furthermore, to evaluate the model using threshold-independent metrics, we report the Area Under the ROC Curve (AUC) and Precision-Recall (PR) values for the independent test sets (Table 3). Consistent with the threshold-dependent metrics, MKE-ResNet achieved the highest AUC scores on 8 out of the 11 independent test sets, further confirming its superior robustness and ranking capability compared to other baseline methods.

To intuitively visualize this performance advantage, we plotted the Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves for the representative independent Human-Brain (iH-b) dataset, as shown in Fig. 2. It can be clearly observed that the curve of MKE-ResNet (represented in red) consistently encompasses the largest area

Table 2 Performance comparison of MKE-ResNet and five baseline models on benchmark and independent datasets

			MKE-ResNet		MST-m6A		m6A-TSHUB		GR-m6A		Deepm6ASeq		pm6A-CNN	
Datasets	**Species**	**Name**	**Acc**	**MCC**	**Acc**	**MCC**	**Acc**	**MCC**	**Acc**	**MCC**	**Acc**	**MCC**	**Acc**	**MCC**
Benchmark	Human	H-b	**0.754**	**0.514**	*0.749*	*0.505*	0.729	0.459	0.743	0.488	0.738	0.483	0.747	0.493
		H-k	*0.835*	*0.674*	**0.843**	**0.686**	0.800	0.601	0.804	0.613	0.800	0.604	0.808	0.617
		H-l	**0.840**	**0.680**	*0.833*	*0.667*	0.805	0.613	0.802	0.612	0.806	0.609	0.809	0.623
	Mouse	M-b	**0.828**	**0.661**	*0.818*	*0.639*	0.798	0.591	0.798	0.601	0.796	0.593	0.800	0.607
		M-h	**0.782**	**0.568**	*0.767*	*0.530*	0.747	0.500	0.750	0.507	0.751	0.507	0.750	0.513
		M-k	*0.841*	*0.690*	**0.851**	**0.703**	0.812	0.626	0.818	0.635	0.816	0.631	0.823	0.645
		M-l	**0.742**	**0.493**	*0.735*	*0.469*	0.724	0.445	0.733	0.468	0.722	0.450	0.734	0.464
		M-t	**0.784**	**0.581**	0.762	0.528	0.768	0.551	*0.776*	0.554	0.774	0.552	0.774	*0.561*
	Rat	R-b	**0.809**	**0.621**	*0.805*	*0.614*	0.776	0.560	0.775	0.559	0.775	0.555	0.783	0.563
		R-k	*0.852*	*0.704*	**0.867**	**0.732**	0.832	0.666	0.841	0.680	0.831	0.667	0.837	0.678
		R-l	*0.850*	*0.701*	**0.856**	**0.715**	0.813	0.625	0.816	0.634	0.816	0.635	0.834	0.667
Independent	Human	iH-b	**0.751**	**0.510**	*0.745*	*0.498*	0.723	0.447	0.737	0.479	0.735	0.472	0.744	0.496
		iH-k	*0.834*	*0.670*	**0.841**	**0.682**	0.793	0.586	0.800	0.601	0.795	0.590	0.802	0.606
		iH-l	**0.831**	**0.663**	*0.823*	*0.646*	0.800	0.601	0.799	0.599	0.800	0.600	0.803	0.608
	Mouse	iM-b	**0.829**	**0.661**	*0.814*	*0.631*	0.789	0.577	0.794	0.590	0.790	0.582	0.795	0.592
		iM-h	**0.776**	**0.555**	*0.761*	*0.525*	0.742	0.487	0.746	0.491	0.747	0.494	0.749	0.499
		iM-k	*0.833*	*0.667*	**0.844**	**0.692**	0.807	0.615	0.813	0.625	0.808	0.617	0.817	0.634
		iM-l	**0.733**	**0.473**	0.724	0.455	0.718	0.437	0.726	*0.458*	0.718	0.439	*0.728*	0.456
		iM-t	**0.782**	**0.563**	0.759	0.529	0.766	0.540	*0.770*	0.546	0.768	0.538	0.769	*0.548*
	Rat	iR-b	**0.801**	**0.601**	*0.793*	*0.600*	0.773	0.547	0.772	0.547	0.770	0.543	0.775	0.552
		iR-k	*0.851*	*0.703*	**0.854**	**0.711**	0.828	0.657	0.834	0.668	0.828	0.657	0.833	0.667
		iR-l	**0.848**	**0.696**	*0.844*	*0.692*	0.805	0.613	0.811	0.623	0.809	0.620	0.827	0.654

The best result is remarked by bold font, and the second-best result is italic

under the curve, achieving an AUROC of **0.840** and an AUPR of **0.823**. This visually confirms that our model maintains a lower false positive rate while ensuring high sensitivity compared to other competing methods, including the PLM-based MST-m6A. The steep ascent of the ROC curve in the low false-positive range further indicates the model's reliability in practical screening scenarios where specificity is paramount.

Ablation study

To investigate the contribution of each core component within the MKE-ResNet model, we designed a series of ablation studies. To comprehensively assess the robustness of each component, we conducted tests on three different Mouse tissue datasets ($\mathrm{M-b}$, $\mathrm{M-h}$, and $\mathrm{M-k}$). As illustrated in Fig. 3, we constructed four variant models by removing or replacing key modules and compared their performance on four core metrics (Acc, MCC, Sn, and Sp) against the full model.

The results clearly and consistently show that across all tested datasets and all evaluation metrics, the full MKE-ResNet model (red bars) outperforms all four variant models. This finding strongly demonstrates the integrity of the model design and the necessity of each component.

Specifically, the variant labeled 'w/o Multi-feature', which relies solely on the fundamental One-hot encoding by discarding the auxiliary physicochemical features (Chemical, EIIP, and ENAC), exhibits the most significant performance drop. This substantial decline highlights the fundamental role of fusing multi-source biochemical information for accurate $\mathrm{m^6A}$ site identification.

Similarly, removing the residual connections ('w/o Res') and the $\mathrm{MKE-ECA}$ module ('w/o ECA') also results in a general performance decay, which validates the critical contributions of the deep residual structure for learning high-quality feature representations and the channel attention mechanism for enhancing the model's discriminative power, respectively. Furthermore, it is noteworthy that the 'Single-Kernel ECA' variant consistently performs worse than our full model with the multi-kernel design, indicating that capturing channel interactions at multiple scales is more effective than at a single scale. In summary, this consistent set of experimental results demonstrates that the multi-feature input strategy, the residual backbone, and the multi-kernel attention mechanism of MKE-ResNet each make an indispensable contribution to its overall performance.

Robustness analysis against sequence noise

In real-world high-throughput sequencing scenarios, data is often prone to random errors or genetic variations such as Single Nucleotide Polymorphisms (SNPs). Therefore, assessing the model's robustness against input noise is crucial for evaluating its reliability.

To this end, we conducted a fine-grained random substitution noise experiment on the representative independent Human-Brain ($\mathrm{iH-b}$) dataset. Considering the short length of the input sequences (41 nt), we adopted a probabilistic mutation strategy where random mutations were introduced to the test sequences at noise rates ranging from 0% to 5% (with a 1% interval). Crucially, after perturbing the raw sequences, we re-extracted all four feature groups (One-hot, Chemical, EIIP, and ENAC) for each perturbed sample to ensure that the noise was consistently propagated through the entire feature engineering pipeline.

Table 3 Detailed performance comparison of MKE-ResNet and baseline models on independent test sets

			MKE-ResNet		MST-m6A		m6A-TSHUB		GR-m6A		Deepm6ASeq		pm6A-CNN	
Datasets	**Species**	**Name**	**AUC**	**PR**	**AUC**	**PR**	**AUC**	**PR**	**AUC**	**PR**	**AUC**	**PR**	**AUC**	**PR**
Independent	Human	iH-b	**0.840**	**0.823**	*0.833*	0.807	0.814	0.787	0.828	*0.814*	0.825	0.811	0.828	*0.814*
		iH-k	*0.914*	*0.906*	**0.926**	**0.921**	0.862	0.840	0.884	0.869	0.877	0.855	0.880	0.863
		iH-l	**0.909**	**0.900**	*0.904*	*0.896*	0.876	0.850	0.876	0.863	0.879	0.858	0.881	0.865
	Mouse	iM-b	**0.904**	**0.896**	*0.893*	*0.885*	0.875	0.848	0.873	0.854	0.873	0.845	0.873	0.858
		iM-h	**0.853**	*0.838*	*0.846*	**0.838**	0.833	0.795	0.824	0.804	0.832	0.804	0.827	0.802
		iM-k	*0.913*	*0.904*	**0.930**	**0.926**	0.876	0.849	0.892	0.873	0.895	0.880	0.900	0.883
		iM-l	*0.810*	**0.789**	0.805	*0.789*	0.790	0.750	0.804	0.784	0.803	0.764	**0.810**	0.777
		iM-t	**0.857**	**0.837**	0.838	0.827	0.834	0.797	0.849	*0.830*	*0.851*	0.819	0.851	0.829
	Rat	iR-b	**0.878**	**0.867**	*0.874*	*0.866*	0.848	0.818	0.851	0.833	0.854	0.830	0.855	0.831
		iR-k	*0.931*	*0.926*	**0.935**	**0.930**	0.909	0.895	0.912	0.903	0.922	0.903	0.909	0.895
		iR-l	**0.930**	**0.925**	*0.925*	*0.920*	0.884	0.862	0.892	0.878	0.900	0.876	0.902	0.887

The best results are highlighted in bold, and the second-best results are italic

The results, as visualized in Fig. 4, demonstrate the exceptional stability of MKE-ResNet. Both the AUC and Accuracy metrics exhibit a gradual, linear decline rather than a sharp drop as the noise rate increases. Quantitatively, the AUC decreased only marginally from the baseline of 0.840 to approximately 0.800 even at a 5% noise rate. This represents a retention of over 95% of the original predictive performance despite significant sequence perturbation. These results confirm that MKE-ResNet relies on robust, distributed feature patterns captured by the multi-stream architecture rather than being overly sensitive to specific single-point mutations, ensuring its applicability in practical biological research.

Generalization and interpretability analysis

The efficacy of MKE-ResNet on source datasets has been demonstrated. To further elucidate whether the sequence features captured by the model possess species- and tissue-sharing properties, we devised cross-species (CS) and cross-tissue (CT) identification tasks. For instance, in CS tasks, we used a model trained on one species' dataset (e.g., $\mathrm{M}-\mathrm{k}$) to predict on another's (e.g., $\mathrm{H}-\mathrm{k}$). Table 4 show the results of these cross-validation tasks. The data reveal that, compared to scenarios where training and validation sets are from the same source, the model's performance in cross tasks decreases slightly but remains at a high level. This finding strongly implies that the discriminative features captured by MKE-ResNet—namely, the multi-scale contextual information intelligently fused by the multi-kernel attention mechanism—possess a degree of universality and are shared across different species and tissues.

To substantiate this and investigate the inner-working mechanism of MKE-ResNet, we conducted a series of interpretability analyses. First, we analyzed the nucleotide enrichment in the $\mathrm{H}-\mathrm{b}$ dataset using TwoSampleLogo [28] (Fig. 5). Interestingly, the results show that the 'ACA' motif is significantly enriched in *both* positive and negative samples near the central site. This phenomenon indicates that while 'ACA' is part of the "RRACH" conserved motif, its presence in both classes limits its discriminative power as a standalone feature. This makes the model's task more challenging and implies that it must rely on broader contextual information for differentiation. To visualize the model's ability to learn from this context, we used UMAP to visualize the feature representations of the $\mathrm{M}-\mathrm{b}$ test set (Figs. 6 and 7). The results clearly show that the deep feature representations learned by the model (before the final classification layer) can effectively separate the highly overlapped positive and negative sample points from the raw data into two distinct, well-separated clusters.

How, then, does the model use contextual information to achieve this separation? We utilized the Grad-CAM technique [29] to visualize the model's attention weights. As a qualitative demonstration, Fig. 8 illustrates the attention heatmaps for representative samples from the $\mathrm{H}-\mathrm{b}$ dataset. It is clearly observable that MKE-ResNet does not focus solely on the central 'ACA' motif but also assigns significant importance to specific flanking regions.

Building on this visual evidence, we proceeded to a quantitative global analysis to systematically identify the biological relevance of these high-attention regions. We implemented an adaptive boundary detection strategy across the entire test set. Instead of arbitrarily extracting fixed-length segments, we defined a dynamic extraction window based on a strict saliency threshold (average contribution rate > 0.85). These

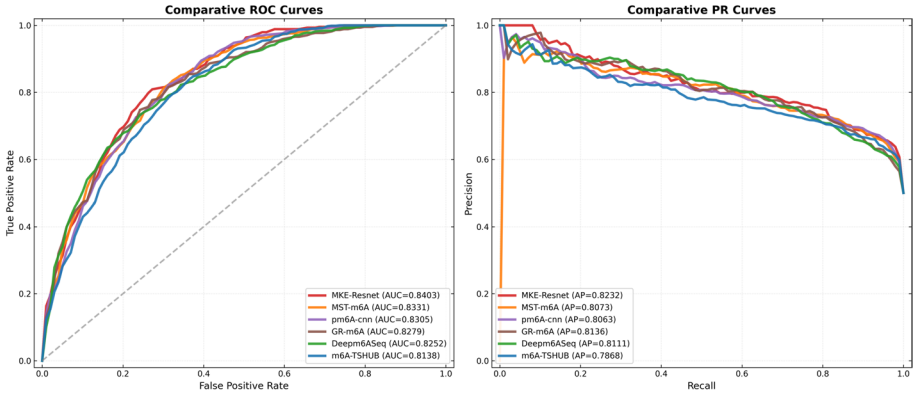

Fig. 2 ROC and Precision-Recall (PR) curves evaluating the performance of MKE-ResNet and baseline models on the representative independent Human-Brain (iH-b) test set. The AUROC and AUPR values for each model are displayed in the legend, demonstrating the robustness and discriminative capability of MKE-ResNet on unseen data. Notably, MKE-ResNet (red curve) encloses the largest area, outperforming state-of-the-art baselines

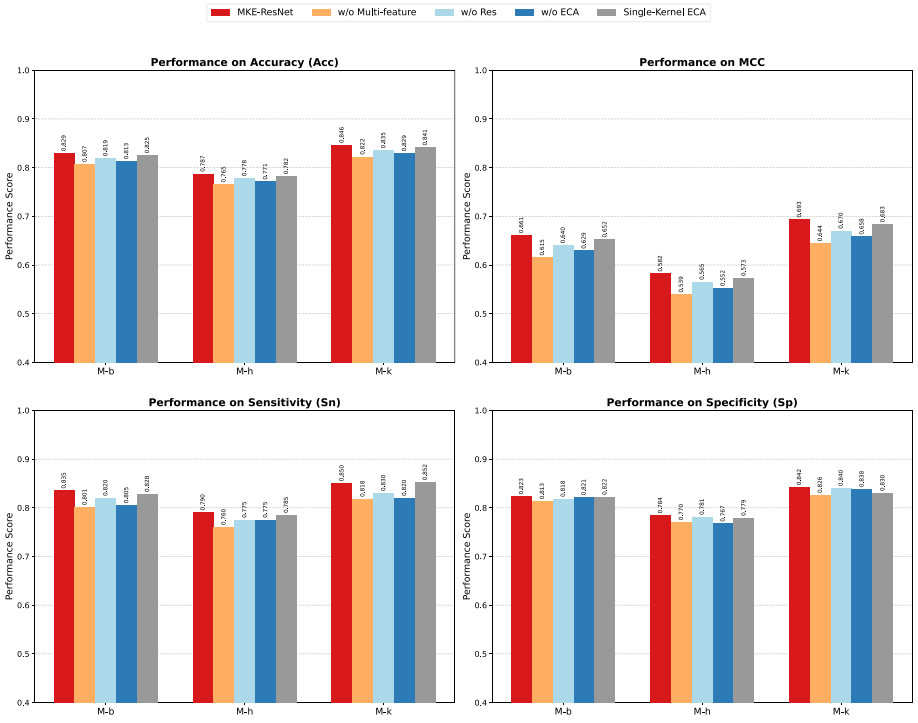

Fig. 3 Results of the ablation study comparing the full MKE-ResNet model against four variants across three Mouse tissue datasets

high-contribution fragments (ranging from 4-nt to 8-nt) were then aligned against the Ray2013 database [30] using the Tomtom tool [31] (Fig. 9 and Table 5).

The comparison revealed multiple matches with high statistical significance, suggesting potential crosstalk between m^6A sites and specific RBPs. For instance, in human data, we identified a motif (GGAGGAA) matching the binding site of SRSF1 ($p \approx 2.48 \times 10^{-5}$), a splicing factor often modulated by m^6A [30]. Another motif (UGUAAUU) matched PUM2 ($p \approx 4.12 \times 10^{-3}$), a key regulator of RNA stability [32]. Similarly, in mouse and rat datasets, we observed patterns resembling the binding motifs of HNRNPK and TARDBP [33].

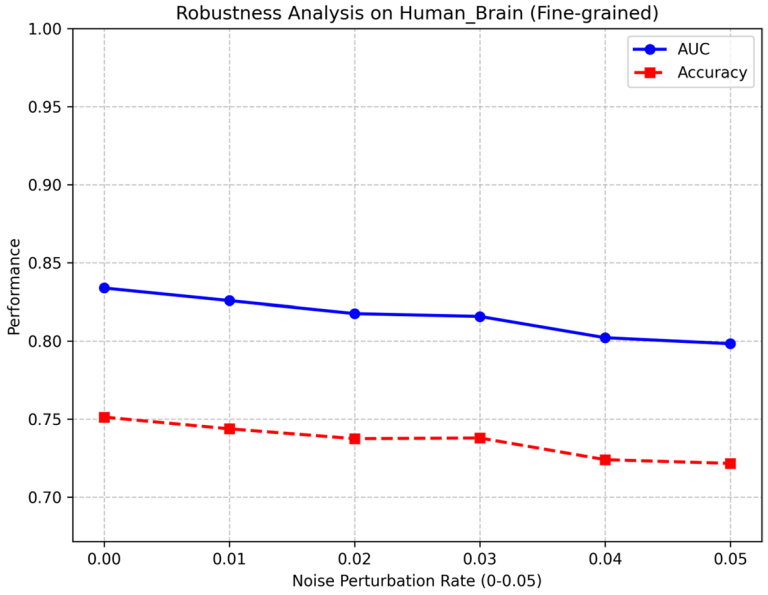

Fig. 4 Performance trend of MKE-ResNet under varying noise perturbation rates (0% - 5%) on the representative Human_Brain dataset. The metrics (AUC and Accuracy) exhibit a gradual, linear decline, demonstrating the model's stability against sequence variations

Table 4 The results of Cross-Species (CS) Tasks and Cross-Tissue (CT) Tasks

Part 1: Cross-Species (CS) Tasks										
Tasks		**Human-Brain**			**Mouse-Brain**			**Rat-Brain**		
Name	**Metrics**	**H-H**	**M-H**	**R-H**	**M-M**	**H-M**	**R-M**	**R-R**	**H-R**	**M-R**
CS Tasks	Acc	0.752	0.799	0.784	0.736	0.828	0.784	0.729	0.801	0.809
	MCC	0.508	0.604	0.569	0.473	0.664	0.561	0.461	0.601	0.620
	Sn	0.828	0.869	0.832	0.761	0.903	0.774	0.795	0.809	0.840
	Sp	0.675	0.729	0.772	0.711	0.753	0.787	0.665	0.792	0.778
Part 2: Cross-Tissue (CT) Tasks										
Tasks		**Rat-brain**			**Rat-kidney**			**Rat-liver**		
Name	**Metrics**	**b-b**	**k-b**	**l-b**	**k-k**	**b-k**	**l-k**	**l-l**	**b-l**	**k-l**
CT Tasks	Acc	0.829	0.819	0.731	0.796	0.830	0.727	0.782	0.792	0.737
	MCC	0.661	0.638	0.463	0.592	0.661	0.463	0.565	0.611	0.481
	Sn	0.835	0.827	0.760	0.782	0.850	0.825	0.965	0.939	0.822
	Sp	0.823	0.812	0.703	0.811	0.811	0.629	0.598	0.649	0.655

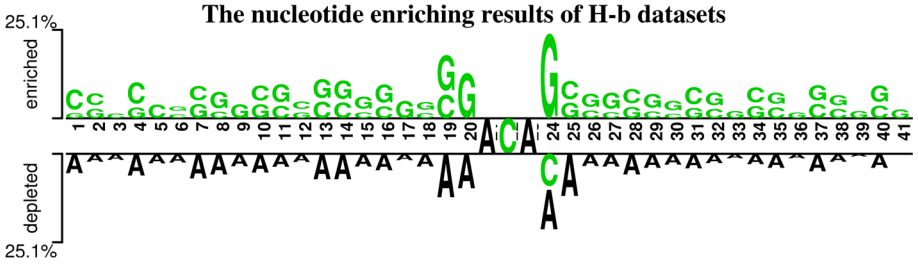

Fig. 5 Nucleotide enrichment analysis using Two Sample Logo for the $\mathrm{H}-\mathrm{b}$ dataset, indicating motif enrichment near the central site

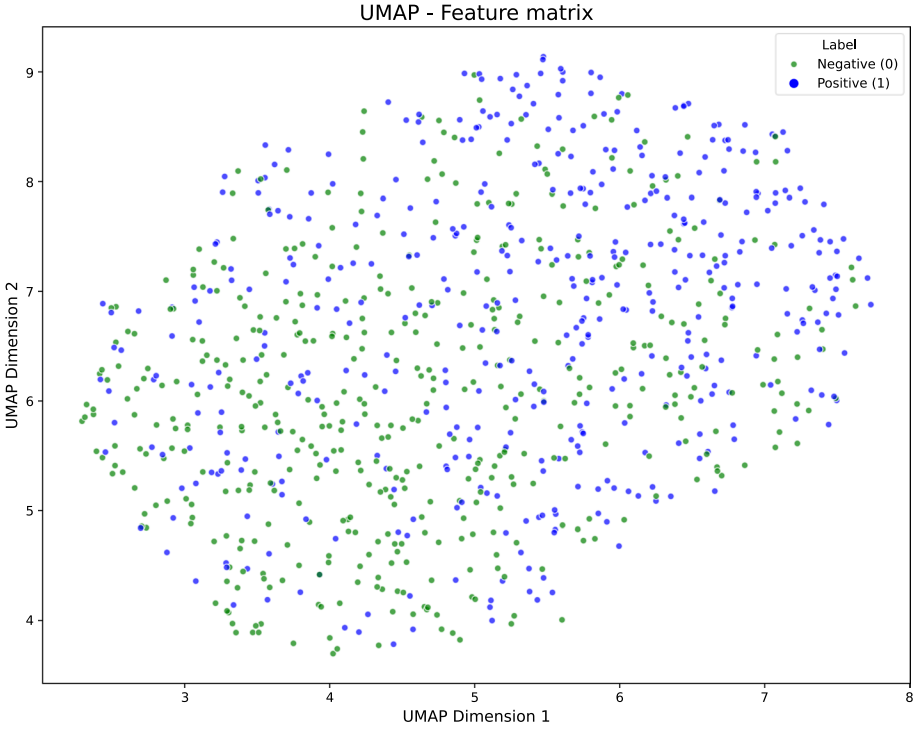

Fig. 6 Interpretability visualization results: UMAP visualization of raw sequence feature representations (encoded features) on the M-b test set

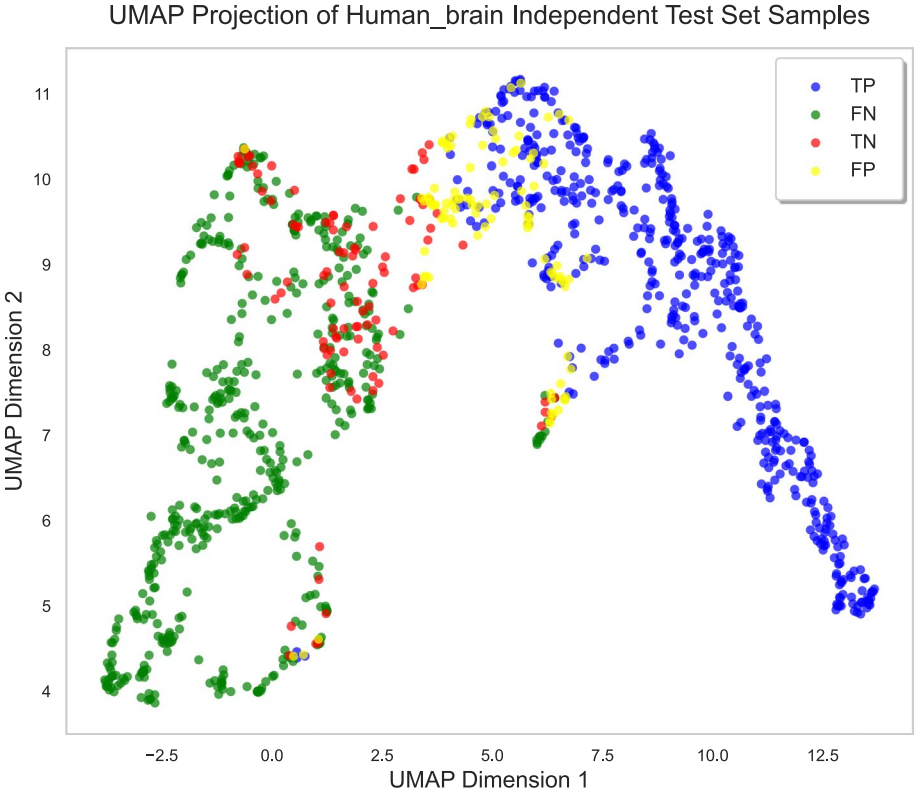

Fig. 7 Interpretability visualization results: UMAP visualization of deep feature representations (trained features) learned by the MKE-ResNet model on the M-b test set

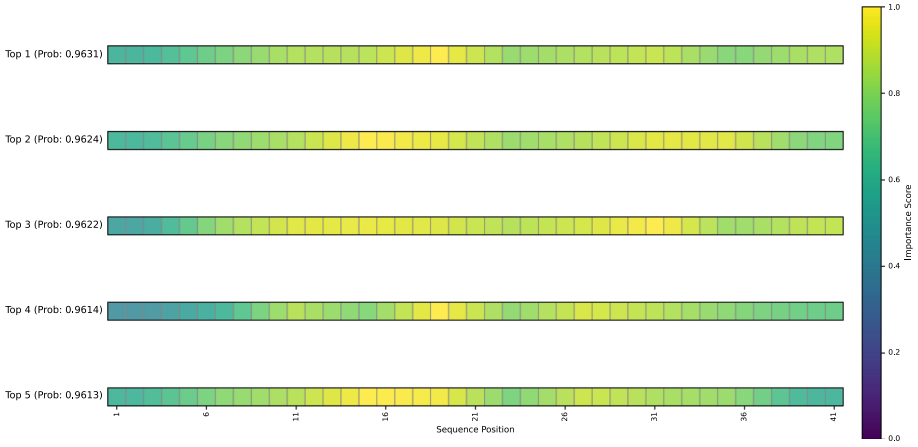

Fig. 8 Grad-CAM visualization of attention weights on samples from the H-b dataset, highlighting specific contextual regions outside the central motif

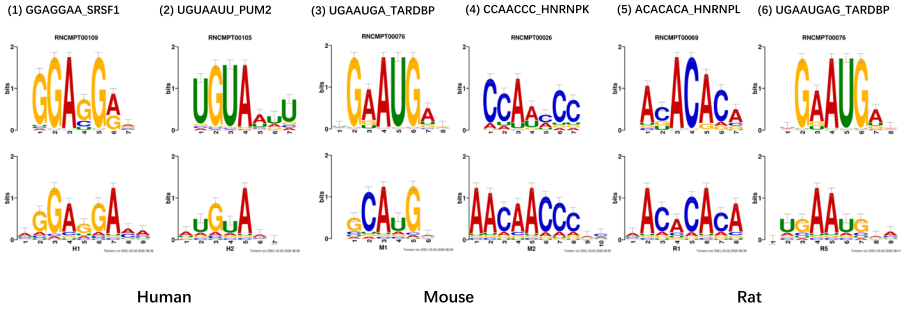

Fig. 9 Comparison of adaptive motif fragments identified by MKE-ResNet against known RBP motifs in the Ray2013 database. The matches were identified using the Tomtom tool with high statistical significance

Table 5 The detailed information of identified motifs and corresponding effector proteins from TomTom (Ray2013 database)

Species	Motif	Effector proteins	Positions	*P*-value
Human	GGAGGAA	SRSF1 [30]	23–29	2.48×10^{-5}
	UGUAAUU	PUM2 [32]	32–38	4.12×10^{-3}
	GACGACGG	SRSF7 [30]	12–19	2.04×10^{-3}
	GAGUAGA	RBM28 [30]	30–36	2.89×10^{-2}
	GACGACA	RBM45 [30]	5–11	2.61×10^{-2}
Mouse	UGAAUGA	TARDBP [33]	19–24	7.95×10^{-3}
	CCAACCC	HNRNPK [30]	10–16	8.05×10^{-4}
	ACUAACA	Tv_0259 [33]	33–49	6.28×10^{-3}
	UGUAAUU	PUM2 [32]	35–41	4.12×10^{-3}
	AUGCACA	MEC-8 [30]	2–8	2.28×10^{-2}
Rat	ACACACA	HNRNPL [33]	1–7	9.04×10^{-6}
	AAUCAAU	RBM46 [30]	12–18	1.32×10^{-2}
	GUGUGUG	RBM38 [30]	28–34	5.52×10^{-3}
	AGUGUGA	RBM24 [30]	15–21	1.15×10^{-4}
	UGAAUGAG	TARDBP [33]	19–26	1.15×10^{-4}

While these findings are computationally derived and require further wet-lab validation, they strongly suggest that MKE-ResNet learns to capture biologically meaningful signals related to the m^6A regulatory network. The model's ability to generalize successfully likely stems from its transcendence of simple reliance on the ambiguous central

motif. Instead, it learns to identify the putative binding signals of various effector proteins encoded in the contextual sequences. This highlights our model's potential in discovering interpretable regulatory patterns.

Conclusion

In current research on $\mathrm{m^6A}$ site identification, most computational models struggle to balance predictive performance with computational efficiency. To address this core challenge, this paper proposes a lightweight deep learning framework named MKE-ResNet. Through a concise serial architecture, this model integrates multi-feature inputs, a deep residual network, and our innovative Multi-Kernel Efficient Channel Attention ($\mathrm{MKE-ECA}$) mechanism. Extensive experiments on multiple benchmark and independent datasets show that MKE-ResNet maintains predictive accuracy comparable to state-of-the-art methods while offering significant advantages in efficiency. Notably, compared to Pre-trained Language Model (PLM) strategies, MKE-ResNet achieves a superior trade-off: it matches the performance of heavy, fine-tuned models (e.g., MST-m6A) without their massive computational costs, while significantly outperforming frozen embedding strategies (e.g., RiNALMo). Furthermore, our interpretability analysis reveals the intrinsic mechanism behind its strong generalization ability: the model moves beyond simple reliance on the ambiguous central motif and instead learns to capture sequence patterns resembling the binding motifs of various effector proteins (such as SRSF1 and PUM2) related to the $\mathrm{m^6A}$ regulatory network.

Despite these promising results, our study has certain limitations that outline directions for future research. First, regarding the scope of comparison, our work focused strictly on sequence-based classification. Consequently, we did not perform a direct comparison with gene-expression-based regression models like Exp2RM, although we acknowledge their distinct contribution to quantifying methylation levels via transcript abundance.

Second, a methodological limitation lies in the data source. Our model was trained on benchmark datasets derived from antibody-dependent m6A-REF-seq, which has been reported to suffer from inherent specificity issues [34]. We fully recognize that direct RNA sequencing (DRS) using Nanopore technology is emerging as the new gold standard for detecting RNA modifications without PCR bias [10, 11]. Therefore, a critical future direction is to extend MKE-ResNet to incorporate high-fidelity Nanopore sequencing data, enabling the capture of a more dynamic and accurate epitranscriptomic landscape.

In conclusion, MKE-ResNet not only provides an efficient and accurate computational tool for $\mathrm{m^6A}$ site identification but also, by revealing contextual regulatory patterns, offers a new computational perspective and valuable biological insights for future studies.

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

Author contributions
X.G. conceived the study, implemented the MKE-ResNet model, performed the experiments, and drafted the manuscript. J.J. supervised the project, provided funding support, and revised the manuscript. C.H. participated in the data analysis and validation. Y.L. assisted in project administration. All authors read and approved the final manuscript.

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.

Data availability
The benchmark datasets analyzed during the current study were originally developed by Dao et al. (2020) and are publicly available in the "Computational and Structural Biotechnology Journal" (DOI: 10.1016/j.csbj.2020.04.015). The specific processed datasets used for training and testing in this study, along with the model code, are available in the GitHub repository: https://github.com/Gxttk/MKE-Resnet.git.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing of interest.

Received: 2 December 2025 / Accepted: 27 February 2026
Published online: 19 March 2026

References

1. 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(7397):201–6.
2. Boccaletto P, Stefaniak F, Ray A, Cappannini A, Mukherjee S, Purta E, Kurkowska M, Shirvanizadeh N, Destefanis E, Groza P, Avşar G, Romitelli A, Pir P, Dassi E, Conticello SG, Aguilo F, Bujnicki JM. MODOMICS: A database of RNA modification pathways. 2021 update. Nucleic Acids Research 2021;50(D1):231–235 https://academic.oup.com/nar/article-pdf/50/D1/D231/42058545/gkab1083.pdf
3. Meyer KD, Jaffrey SR. The dynamic epitranscriptome: N6-methyladenosine and gene expression control. Nat Rev Mol Cell Biol. 2014;15(5):313–26.
4. Geula S, Moshitch-Moshkovitz S, Dominissini D, Mansour AA, Kol N, Salmon-Divon M, et al. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. Science. 2015;347(6225):1002–6.
5. 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. Nat Chem Biol. 2011;7(12):885–7.
6. Pan T, Wu F, Li L, Wu S, Zhou F, Zhang P, et al. The role m6A RNA methylation is CNS development and glioma pathogenesis. Mol Brain. 2021;14(1):119.
7. Kostyusheva A, Brezgin S, Glebe D, Kostyushev D, Chulanov V. Host-cell interactions in HBV infection and pathogenesis: the emerging role of m6A modification. Emerg Microbes Infections. 2021;10(1):2264–75.
8. Meng J, Lu Z, Liu H, Zhang L, Zhang S, Chen Y, Rao MK, Huang Y. A protocol for RNA methylation differential analysis with MeRIP-Seq data and exomePeak R/Bioconductor package. Methods (San Diego, Calif.) 2014;69(3):274–281
9. Linder B, Grozhik AV, Olarerin-George AO, Meydan C, Mason CE, Jaffrey SR. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. Nat Methods. 2015;12(8):767–72.
10. Zhang Y, Jiang J, Ma J, Wei Z, Wang Y, Song B, et al. etal Directrmdb: a database of post-transcriptional rna modifications unveiled from direct rna sequencing technology. Nucleic Acids Res. 2023;51(D1):106–16.
11. Zhang Y, Wu Y, Ma J, Wu Y, Li L, Wang H, et al. etal Directrm: integrated detection of landscape and crosstalk between multiple rna modifications using direct rna sequencing. Nat Commun. 2025;16(1):9450.
12. Parker MT, Knop K, Sherwood AV, Schurch NJ, Mackinnon K, Gould PD, Hall AJ, Barton GJ, Simpson GG. Nanopore direct RNA sequencing maps the complexity of Arabidopsis mRNA processing and m^6A modification. eLife 2020;9:49658
13. Senol Cali D, Kim JS, Ghose S, Alkan C, Mutlu O. Nanopore sequencing technology and tools for genome assembly: computational analysis of the current state, bottlenecks and future directions. Briefings Bioinf. 2019;20(4):1542–59.
14. Zhou Y, Zeng P, Li Y-H, Zhang Z, Cui Q. SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. Nucleic Acids Res. 2016;44(10):91–91.
15. Wei L, Chen H, Su R. M6APred-EL: a sequence-based predictor for identifying N6-methyladenosine sites using ensemble learning. Mol Therapy Nucleic Acids. 2018;12:635–44.
16. Zhang Y, Hamada M. Deepm6aseq: prediction and characterization of m6a-containing sequences using deep learning. BMC Bioinf. 2018;19(Suppl 19):524.
17. Nazari I, Tahir M, Tayara H, Chong KT. iN6-Methyl (5-step): identifying RNA N6-methyladenosine sites using deep learning mode via Chou's 5-step rules and Chou's general PseKNC. Chemom Intell Lab Syst. 2019;193:103811.
18. Tahir M, Hayat M, Chong KT. Prediction of N6-methyladenosine sites using convolution neural network model based on distributed feature representations. Neural Netw. 2020;129:385–91.
19. Su Q, Pham NT, Wei L, Manavalan B. etal Mst-m6a: a novel multi-scale transformer-based framework for accurate prediction of m6a modification sites across diverse cellular contexts. J Mol Biol. 2025;437(6):168856.
20. Song B, Huang D, Zhang Y, Wei Z, Su J, Magalhães J, et al. m6a-tshub: unveiling the context-specific m 6 a methylation and m 6 a-affecting mutations in 23 human tissues. Genomics Proteomics Bioinf. 2023;21(4):678–94.
21. Abbas Z, Tayara H, Zou Q, Chong KT. TS-m6A-DL: tissue-specific identification of N6-methyladenosine sites using a universal deep learning model. Comput Struct Biotechnol J. 2021;19:4619–25.

22. Alam W, Ali SD, Tayara H, Chong K. A CNN-based RNA N6-methyladenosine site predictor for multiple species using heterogeneous features representation. IEEE access : practical innovations, open solutions 2020;8:138203–138209 https://doi.org/10.1109/ACCESS.2020.3002995
23. 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 Trans Comput Biol Bioinf. 2023;20(2):904–11. https://doi.org/10.1109/TCBB.2022.3192572.
24. 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. 2023;39(12):709.
25. Song Y, Wang Y, Wang X, Huang D, Nguyen A, Meng J. Multi-task adaptive pooling enabled synergetic learning of RNA modification across tissue, type and species from low-resolution epitranscriptomes. Brief Bioinform. 2023;24(3):105.
26. Qiu S, Liu R, Liang Y. GR-m6A: prediction of N6-methyladenosine sites in mammals with molecular graph and residual network. Comput Biol Med. 2023;163:107202.
27. Dao F-Y, Lv H, Yang Y-H, Zulfiqar H, Gao H, Lin H. Computational identification of N6-methyladenosine sites in multiple tissues of mammals. Comput Struct Biotechnol J. 2020;18:1084–91.
28. Vacic V, Iakoucheva LM, Radivojac P. Two sample logo: a graphical representation of the differences between two sets of sequence alignments. Bioinformatics. 2006;22(12):1536–7.
29. Sattarzadeh S, Sudhakar M, Plataniotis KN, Jang J, Jeong Y, Kim H. Integrated grad-cam: Sensitivity-aware visual explanation of deep convolutional networks via integrated gradient-based scoring. In: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021;1775–1779 . https://doi.org/10.1109/ICASSP39728.2021.9415064
30. Ray D, Kazan H, Cook KB, Weirauch MT, Najafabadi HS, Li X, et al. A compendium of RNA-binding motifs for decoding gene regulation. Nature. 2013;499(7457):172–7.
31. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between motifs. Genome Biol. 2007;8(2):24.
32. Cook KB, Kazan H, Zuberi K, Morris Q, Hughes TR. RBPDB: a database of RNA-binding specificities. Nucleic Acids Res. 2011;39(suppl1):301–8.
33. Sasse A, Ray D, Laverty KU, Tam CL, Albu M, Zheng H, Lyudovyk O, Dalal T, Nie K, Magis C, Notredame C, Weirauch MT, Hughes TR, Morris Q. Reconstructing the sequence specificities of RNA-binding proteins across eukaryotes. bioRxiv 2024. arxiv: 10.1101/2024.10.15.618476.full.pdf
34. Liu C, Sun H, Yi Y, Shen W, Li K, Xiao Y, et al. etal Absolute quantification of single-base m6a methylation in the mammalian transcriptome using glori. Nat Biotechnol. 2023;41(3):355–66.

Publisher's Note