Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 20(11): 19320–19340.
DOI: 10.3934/mbe.2023854
Received: 24 August 2023
Revised: 08 October 2023
Accepted: 12 October 2023
Published: 17 October 2023

Research article

TransUFold: Unlocking the structural complexity of short and long RNA with pseudoknots

Yunxiang Wang[1], Hong Zhang[1,*], Zhenchao Xu[1], Shouhua Zhang[2] and Rui Guo[3]

[1] School of Cyber Security and Computer, Hebei University, Baoding, Hebei, China
[2] Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland
[3] College of Life Sciences, Institute of Life Science and Green Development, Hebei University, Baoding, China

* **Correspondence:** Email: hzhang@hbu.edu.cn.

Abstract: The RNA secondary structure is like a blueprint that holds the key to unlocking the mysteries of RNA function and 3D structure. It serves as a crucial foundation for investigating the complex world of RNA, making it an indispensable component of research in this exciting field. However, pseudoknots cannot be accurately predicted by conventional prediction methods based on free energy minimization, which results in a performance bottleneck. To this end, we propose a deep learning-based method called TransUFold to train directly on RNA data annotated with structure information. It employs an encoder-decoder network architecture, named Vision Transformer, to extract long-range interactions in RNA sequences and utilizes convolutions with lateral connections to supplement short-range interactions. Then, a post-processing program is designed to constrain the model's output to produce realistic and effective RNA secondary structures, including pseudoknots. After training TransUFold on benchmark datasets, we outperform other methods in test data on the same family. Additionally, we achieve better results on longer sequences up to 1600 nt, demonstrating the outstanding performance of Vision Transformer in extracting long-range interactions in RNA sequences. Finally, our analysis indicates that TransUFold produces effective pseudoknot structures in long sequences. As more high-quality RNA structures become available, deep learning-based prediction methods like Vision Transformer can exhibit better performance.

Keywords: RNA secondary structure prediction; pseudoknot; Vision Transformer; Long-range interactions; deep learning

1. Introduction

RNA plays a critical role in transferring genetic information from DNA to proteins [1], but it also has other functions such as enzyme activity [2] and cellular regulation [3]. To understand the function of RNA, it is crucial to obtain its structure. RNA structure can be divided into three levels: Primary, secondary and tertiary. Predicting the tertiary structure is challenging due to the involvement of multiple factors [4]. Experimental methods like X-ray crystallography [5] and NMR [6] are time-consuming and expensive. Therefore, it is crucial to predict RNA secondary structure accurately. In order to address the demand for high-throughput data [7], computational methods for RNA secondary structure prediction have been created.

It is expected to predict higher-order RNA secondary structures through the primary structure of RNA. The most common methods are based on thermodynamic models. These models assume that RNA secondary structures only contain nested base pairs and employ dynamic programming [8] to minimize free energy. The Nearest Neighbor Thermodynamic Model (NNTM) [9] is a state-of-the-art technique that uses experimental parameters to describe the free energy of nearest-neighbor loops (Figure 1), which are then added together to represent the entire free energy of the RNA secondary structure. Other more efficient tools like Mfold [10], UNAfold [11], RNAfold [12], RNAstructure [13] and LinearFold [14] are also based on this approach. However, they cannot predict pseudoknots in RNA secondary structures, which are non-nested structures (Figure 2) that make predictions based on energy minimization an NP-complete problem [15]. The Nearest Neighbor Thermodynamic Model has recently been redesigned by introducing additional parameters such as PKNOTS [16], NUPACK [17] and VFold [18]. However, these algorithms still exhibit a time complexity of $O(n^4)$ to $O(n^6)$ when computing the secondary structure of an RNA molecule containing n bases [19,20]. There are also algorithms that mitigate the computational challenges by employing heuristic strategies, such as HotKnots [21] and IPKnot [22]. Although these algorithms are remarkably fast, they do not guarantee the quality of the predicted secondary structures [20]. As the number of known RNA secondary structures gradually increases, another class of machine learning-based methods has been proposed. ContraFold [23] and ContextFold [24] improve RNA secondary structure prediction accuracy by training energy parameter scores based on known structures. There is also a type of hybrid method that integrates thermodynamics with learning-based techniques, such as MXfold [25] and MXfold2 [26]. These methods can evaluate substructures not seen during training. However, these methods still rely on dynamic programming algorithms to minimize free energy and struggle to predict pseudoknot structures.

Due to the rapid advancements in deep learning techniques, many previously explored topics have been revisited, resulting in significant breakthroughs. The accumulation of large amounts of RNA secondary structure data has provided favorable conditions for applying deep learning to predict RNA secondary structure. CDPFold [27] utilized a convolutional neural network but represented the resulting structure in dot-bracket notation which failed to express pseudoknot structures. SPOT-RNA [28] utilized ResNet [29] and bidirectional LSTM [30] while E2Efold [31] combined Transformer [32] with convolutional networks to design an end-to-end model that effectively considered the inherent constraints through unrolled algorithms. Both regarded RNA secondary structure prediction as binary classification and could predict pseudoknot structures. UFold [33] introduced a U-net architecture and represented input sequence data in an "image-like" format, significantly improving prediction performance.

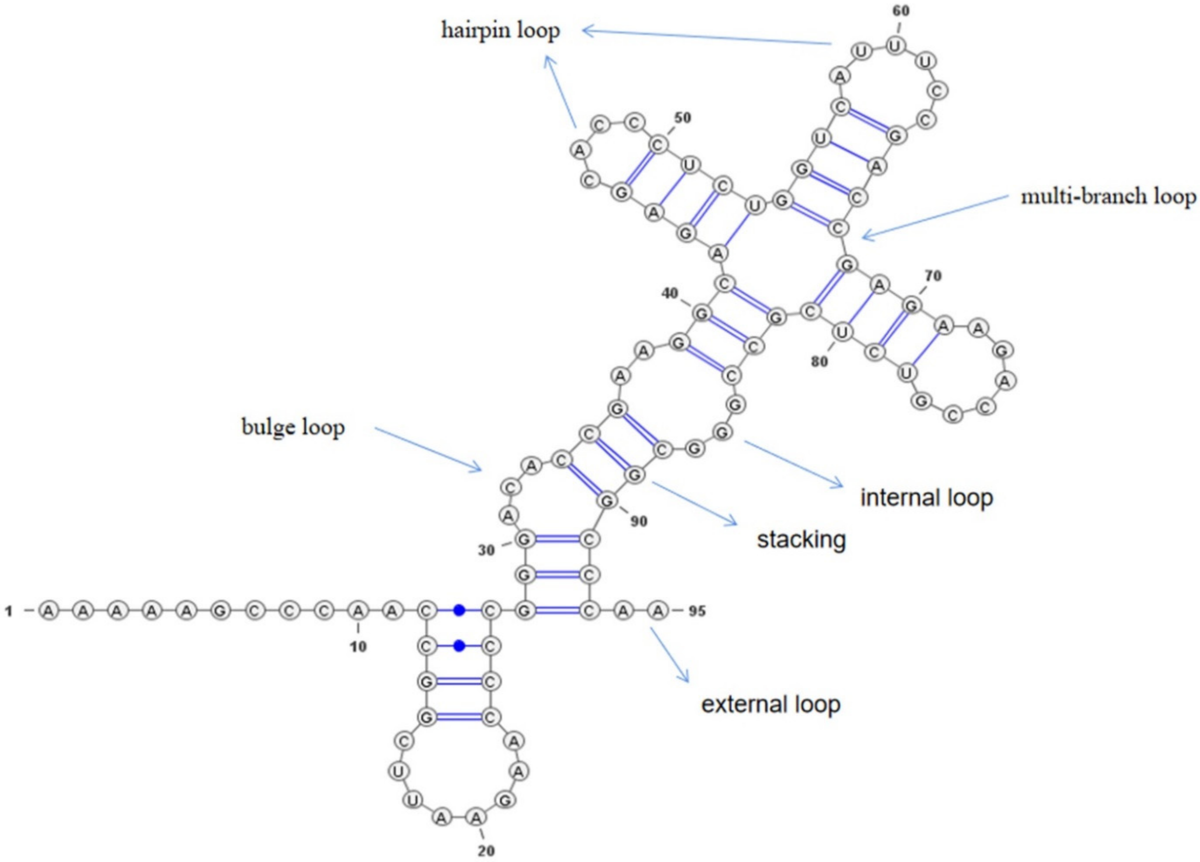

Figure 1. Decomposition of an RNA secondary structure into nearest structural motifs. An RNA secondary structure can be decomposed into several types of nearest-neighbor loops, including bulge loops (e.g., bases 30–34 and 90–91), hairpin loops (e.g., bases 45–50 and 57−64), multi-branch loops (e.g., bases 41–41, 53–54, 64−68 and 81–82), internal loops (e.g., bases 36–39 and 84–88), base-pair stackings (e.g., bases 34–36 and 88–90) and external loops (e.g., bases 94–95). This diagram was drawn using VARNA [34].

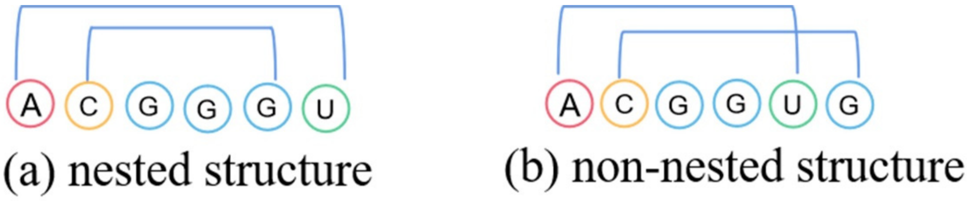

Figure 2. Examples of nested and non-nested secondary structures. The blue lines represent base pairing, the red circles with A represent Adenosine monophosphate, the yellow circles with C represent Cytidine monophosphate, the blue circles with G represent Guanosine monophosphate and the green circles with U represent Uridine monophosphate.

In this work, we propose a novel method named TransUFold, which combines Vision Transformer with lateral connections for predicting RNA secondary structures. This method can transform onedimensional RNA sequences into ‘image-like’ data with 16 channels as input. For methods using pure convolution, it is difficult to capture the long-range interactions and folding information in RNA sequences due to the limitation of the receptive field, which leads to defects in predicting long sequences and long-distance base pairings. We employ a Vision Transformer [35] framework based on the self-attention mechanism [32] to extract more comprehensive features from images. The Self-Attention Mechanism is a crucial technique in deep learning used for processing sequence data and capturing internal data relationships. The core idea of this mechanism is to compute weights for each element in a sequence, reflecting its degree of association with other elements in the sequence. Specifically, it calculates these weights by comparing each element with the others and then normalizing the results to obtain the final weight values. This enables the model to dynamically assign different attention weights to each element based on the relationships between different positions, without the need for manually specifying weights or positions. Therefore, the advantage of the Self-Attention Mechanism lies in its ability to capture relationships between any two positions in the input sequence, making it highly effective in handling long-range dependencies. The Vision Transformer (ViT) is a recent deep learning architecture by introducing the self-attention mechanism that has gained prominence in the field of computer vision. ViT aims to better understand images by allowing the model to capture global relationships between pixels. To encode a base of the input RNA sequence, the self-attention mechanism allows the model to focus directly on other bases in RNA sequence. In our approach, we also introduce lateral connections from a series of full convolutions to discover short-range interactions in RNA. Convolutions excel at capturing short-range interactions. These lateral connections enable the decoder to compensate for potential deficiencies in short-range interactions that may exist in the primary encoder, which is advantageous for predicting complex pseudoknot structures. Due to this, our model can predict RNA secondary structures containing pseudoknots. We conducted a series of experiments to compare TransUFold with other state-of-the-art methods. The results showed that TransUFold achieved superior performance in predicting RNA secondary structures, indicating potential influence on advancing RNA research.

2. Materials and methods

2.1. Datasets

To examine the accuracy of our approach, we conduct the experiments based on reliable RNA sequences and associated structural information of various families from two benchmark datasets: RNAStralign [36] and ArchiveII [37]. After removing redundant sequence structures, we summarize the datasets in Table 1. In addition to evaluating the performance of our method in cross-family prediction, we also employ dataset bpRNA-new from Rfam 14.2 [26,38], which includes sequences from 1500 new RNA families that are not present in any other datasets. The redundancy among the datasets RNAStralign, ArchiveII and bpRNA-new has been eliminated.

We randomly split the dataset RNAStralign into a training set and a test set at 4:1 to evaluate the accuracy of our method for RNA secondary structural prediction. Furthermore, another two datasets ArchiveII and bpRNA-new are introduced as test sets to examine the accuracy of prediction for families with different distributions and other unused families during training. Then, we select 10,879 RNAs with sequence length of 512−1600 nt in dataset RNAStralign and divide them into a training set and a test set in a ratio of 4:1 at random to test the performance in long sequence prediction.

Table 1. Dataset statistics.

TYPE	RNAStralign	ArchiveII
5SrRNA	9385	1283
16SrRNA	11,620	110
tRNA	6443	557
grp1	1502	98
SRP	468	928
tmRNA	572	462
RNaseP	434	454
telomerase	37	37
23SrRNA	-	35
grp2	-	11
ALL	30,461	3975

2.2. Input and output representation

RNA secondary structure prediction is the task of predicting the base pairing pattern for a given RNA sequence. Most methods, such as E2Efold, ATTfold [39] and MXfold2, treat the RNA sequence $S = (\mathrm{s}_1, \mathrm{s}_2, \ldots \mathrm{s}_l),\ \mathrm{s}_l \in \{A, U, C, G\}$ as a simple sequence for input. However, UFold introduces a novel method to convert an RNA sequence into an "image". Like UFold, our approach represents S using one-hot encoding as an $L \times 4$ binary matrix $X \in \{0,\ 1\}^{L \times 4}$ and then performs a Kronecker product on X with itself to transform S into a $16 \times L \times L$ tensor (as shown in Figure 3).

$$K = X \otimes X \tag{1}$$

A noteworthy feature of the UFold-structure is its ability to simulate various long-distance interactions among nucleotides within the RNA sequence as local patterns in the image. Additionally, it takes every base pairing (including both canonical and non-canonical) into account by representing S as a 16-channel image, where each channel represents a base pairing. In Figure 3, $K \in \{0,\ 1\}^{16 \times L \times L}$ denotes a 16-channel UFold -structure image, where $K(i, j, k)$ represents whether s_j and s_k form a base pair according to the $i-th$ base pairing rule. The final output of our model is the secondary structure matrix $U \in [0,\ 1]^{L \times L}$, where U_{ij} represents whether s_i and s_j in S exist a base pair.

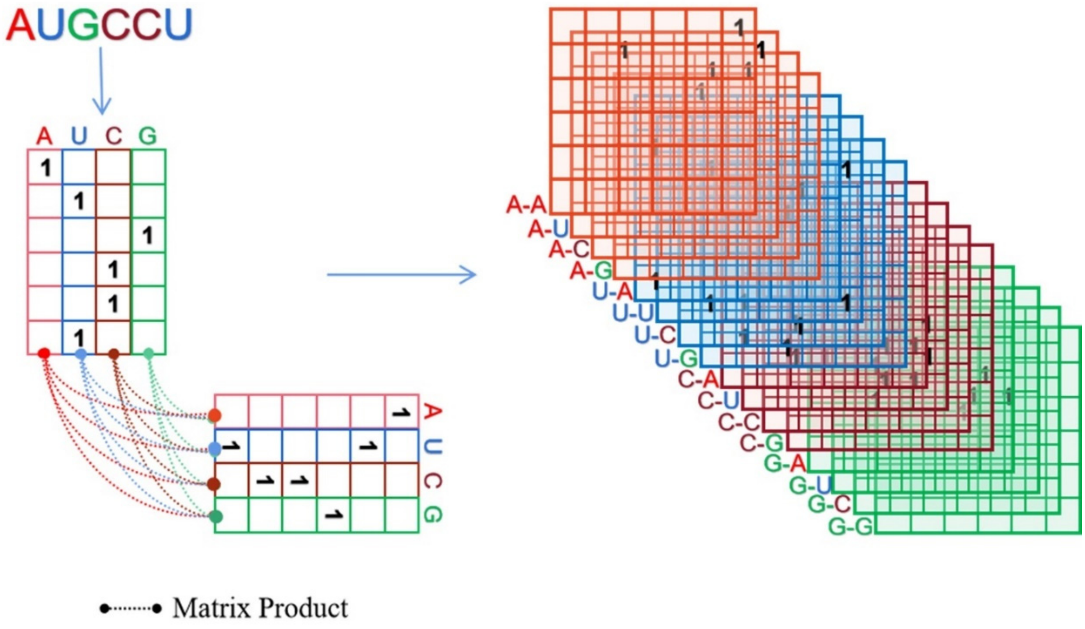

Figure 3. The TransUFold input with a 16-channel image-like representation. Illustration of base sequence transformation into a one-hot matrix (blue arrows) and subsequent conversion into a “16-channel image” (rightward blue arrows). The red, blue, brown and green dashed lines express the vectors corresponding to the positions of the bases in the sequence, undergoing Matrix Product with other vectors (Adenine, Uracil, Cytosine and Guanine). The red, blue, brown and green matrices represent the four pairing rules associated with the respective bases (Adenine, Uracil, Cytosine and Guanine).

2.3. Network architecture and post-processing

We design a brand-new encoder-decoder architecture with two encoders: a primary encoder and an auxiliary encoder. The primary encoder applies a vision transformer to focus on global features like long-range interactions from the input of RNA sequence. Specifically, our vision transformer consists of six Transformer Encoders, each of which undergoes a series of key steps to effectively extract features. First, the input features pass through a layer normalization module, which helps balance the distribution of features and enhances training stability. Next, a multi-head self-attention computation module is introduced, where 16 self-attention heads are set to capture the global contextual information of the input sequence. Such a design allows the network to simultaneously focus on features at different positions, thereby better capturing extensive semantic relationships. After the self-attention computation, a residual connection is established between the original input and the self-attention output features. This type of connection facilitates information flow and helps maintain feature stability. Subsequently, another layer normalization module is applied to maintain feature consistency. Following this, a Multi-Layer Perceptron (MLP) module is introduced for non-linear transformations to capture more comprehensive feature information. This MLP module comprises two linear transformation layers, with the first layer mapping input features to a higher-dimensional feature space. Subsequently, a GELU activation function is applied for non-linear transformation, further enriching feature representation. To prevent overfitting, a Dropout layer is introduced after the MLP. Finally, the feature is mapped to the final output dimension through the second linear layer. This design not only allows the network to adaptively capture crucial features from the input data but also connects the original input of the MLP with its output through residual connections, further enhancing feature expressiveness. The combination of these steps enables each Transformer Encoder to efficiently extract information from input features and utilize it for the task at hand, thereby enhancing the network's performance. To further enrich local features, we also design an auxiliary encoder consisting of a series of fully convolutional downsampling layers as lateral connections for the input of the decoder. The convolutional neural network encoder block is a crucial module designed to effectively extract features through a sequence of hierarchical operations. First, it comprises two stacked convolutional layers that utilize 3 × 3 convolutional kernels to perform convolutions on the feature maps. This aids in capturing local features within the image while preserving the spatial dimensions of the feature maps. After each convolutional layer, batch normalization is applied to normalize the distribution of features, enhancing model stability and training speed. Following batch normalization, the ReLU activation function is introduced to bring about non-linear transformations. This facilitates the model in learning richer and more complex feature representations to adapt to different image patterns and structures. The overall design of the encoder block follows the sequential arrangement of convolutional layers, batch normalization and activation functions, resulting in a compact and effective feature extraction process. The output of the convolutional neural network encoder block is supplemented by the decoder through lateral connections. A 2 × 2 max-pooling is applied after each encoder block to perform downsampling operations. By stacking four such encoder blocks, the neural network progressively extracts more semantically meaningful features and supplements them to the decoder. In the decoder section, comprised of four decoder convolution blocks, each block follows a set of key steps in its design: To begin with, an upsampling operation is applied to double the dimensions of the input feature map. This operation aids in restoring the spatial resolution of the feature map and contributes additional information for subsequent stages. Following this, a 3 × 3 convolutional kernel is utilized to perform convolutional computations on the upsampled feature map. After the convolutional operation, batch normalization is employed to normalize the distribution of features, thereby enhancing the model's stability. Subsequently, a ReLU activation function is introduced, introducing non-linear transformations that enable the model to acquire more intricate and enriched feature representations, suited for adapting to various image patterns and structures. In summary, the decoder convolution block progressively transforms low-dimensional feature mappings into high-resolution outputs through a sequence of operations. We use the output of the primary encoder as the main input and use the output of the auxiliary encoder to supplement the decoder layer by layer. The network outputs an $L \times L$ matrix, which is then multiplied by its transpose to make a symmetric matrix as the contact score matrix U shown in Figure 4. The loss function we applied is Binary CrossEntropyLoss to minimize the loss between the contact score matrix U and the true pairing matrix A through training with stochastic gradient descent. Before computing the loss, the output of the last layer must activate with a sigmoid function to ensure that the contact score of Matrix U is strictly positive.

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{2}$$

$$Loss(U, A) = -\sum_{ij}[A_{ij}\log(U_{ij}) + (1 - A_{ij})\log(1 - U_{ij})] \tag{3}$$

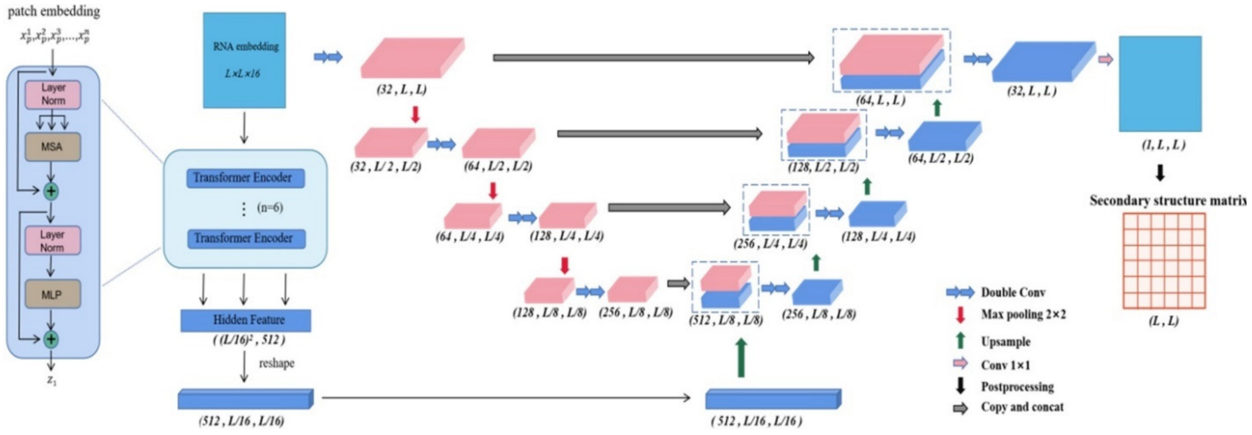

Figure 4. The overall architecture of TransUFold. The input is a 16 × L × L tensor obtained by transforming the original sequence, and the output is a L × L symmetric score matrix, which is post-processed to obtain the final secondary structure matrix.

To ensure that the output satisfies the RNA pairing constraints, we add a post-processing network to filter out non-standard base pairings, resulting in the final RNA secondary structure matrix. The post-processing considers three hard constraints of RNA secondary structure:

(i) Only three base pairing modes are allowed to exist: A-U, G-C [40] and U-G [41];

(ii) no sharp loops are allowed， $\forall |i-j| < 4, U_{ij} = 0$;

(iii)no overlapping pairs are allowed, $\forall i, \sum_{j=1}^{L} A_{ij} \le 1$.

We obtain a symmetric contact score matrix U in our network architecture. For constraint (i), we implement base pairing modes on each output sequence through a predefined $L \times L$ matrix, where we fill 1 at locations satisfying A-U, G-C and U-G pairings, and 0 elsewhere. By element-wise multiplying this matrix by the contact score matrix U, we preserve predicted values that satisfy the base pairing constraints. For constraint (ii), we employ a similar approach as constraint (i). We design an $L \times L$ matrix with its diagonal to 0 to indicate that a base does not pair with itself, and assign 0 to all entries within 3 positions away from the diagonal and 1 to all other entries. To ensure constraint (ii) is satisfied, we multiply (element-wise) this matrix by the contact score matrix U. For considering both constraints (i) and (ii), we define a nonlinear transformation T used in E2Efold in Eq (4):

$$\mathrm{T}(\hat{U}) := \frac{1}{2}\left(\hat{U} \circ \hat{U} + (\hat{U} \circ \hat{U})^T\right) \circ M(x) \tag{4}$$

Where $\circ$ represents element-wise multiplication. The matrix M is defined to satisfy the constraints (i) and (ii), that is, $M(x)_{ij} := 1$ if $x_i x_j$ {AU,UA} ∪ {GC,CG} ∪ {GU,UG} and $|i-j| > 4$, and $M(x)_{ij} := 0$ otherwise. The final constraint is depicted in Eq (5) to ensure each base has at most one pairing.

$$\mathrm{relu}(\mathrm{Ul} - 1) = 0, \mathrm{l} = 1, \ldots, \mathrm{L} \tag{5}$$

where Ul represents the number of base pairings, satisfying Ul ≤ 1. We transform it into a linear programming problem to find the optimal scoring matrix that maximizes its similarity with $\mathrm{T}(\hat{U})$.

$$\max_{\hat{U}\in R^{L\times L}} \frac{1}{2}\left\langle U-s, \mathrm{T}(\hat{U})\right\rangle - \rho\left\|\hat{U}\right\|, subject\ to, \boldsymbol{Ul} \leq 1 \qquad (6)$$

where s represents a threshold such that $U_{ij} > s$ indicates a pairing existing between base i and base j, otherwise not. The similarity between the scoring matrix and $\mathrm{T}(\hat{U})$ is represented by their inner product, with $L1$ regularization penalty term to control the sparsity of the output matrix through hyperparameters. For this single-constraint optimization problem, it can be solved using the method of Lagrange multipliers. Finally, the optimal matrix U^* with the highest similarity to $\mathrm{T}(\hat{U})$ is our final RNA secondary structure prediction matrix.

$$\min_{\lambda\geq 0} \max_{\hat{U}\in R^{L\times L}} \frac{1}{2}\left\langle U-s, \mathrm{T}(\hat{U})\right\rangle - \left\langle \lambda, relu(Ul-1)\right\rangle - \rho\left\|\hat{U}\right\| \qquad (7)$$

2.4. Evaluation

Our experiments are executed on a computer with 64-bit AMD EPYC 7551P processor, Nvidia RTX A4000-16G graphics card, 36 GB RAM and Ubuntu Operating System. Our model is trained for 100 epochs on dataset RNAStralign and the best model is selected through the validation set as our final model. To better evaluate the predicted RNA secondary structure by TransUFold, we apply three evaluation metrics: Precision, Recall and F1 score, shown in Eqs (8)−(10). The definitions of TP, FN, TN and FP are demonstrated in Table 2. TP represents the correctly predicted base pairs. FN represents incorrectly predicted actual base pairs. TN represents correctly predicted non-base pair positions. FP represents incorrectly predicted base pairs that actually do not exist.

Table 2. Specific representations of each parameter in performance metrics.

Predict	**True**	
	P	N
P	TP	FP
N	FN	TN

$$Precision = \frac{TP}{TP+FP} \qquad (8)$$

$$Recall = \frac{TP}{TP+FN} \qquad (9)$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{10}$$

3. Results

We evaluate the performance of our model on two benchmark datasets (RNAStralign and ArchiveII) and a cross-family dataset (bpRNA-new). Due to differences in length distributions (as shown in Figure 5), padding the input according to maximum sequence lengths would greatly increase the sparsity of the input matrix and training complexity. Therefore, we select non-redundant RNA sequences with lengths less than 512 nt from the dataset RNAStralign for one model and others for another model. We train the model using the Adam optimizer for 100 epochs with a learning rate of 0.001. The same settings are applied to other learning-based methods.

Specifically, to simulate real-world scenarios for short RNA sequences, two different scenarios are designed to predict new RNA sequence structures for known and unknown families, respectively. For the former scenario, we evaluate the accuracy of the trained model directly based on dataset ArchiveII with different data distributions from dataset RNAStralign. To assess our model in the latter scenario, we introduce dataset bpRNA-new, which contains families that are not included in other datasets. Additionally, we evaluate TransUFold's performance on long RNA sequences from dataset RNAStralign that range in length from 512 to 1600 nt. Finally, to verify whether TransUFold truly generates pseudoknots, we analyze the results of the RNA sequence structures with pseudoknots in the test set.

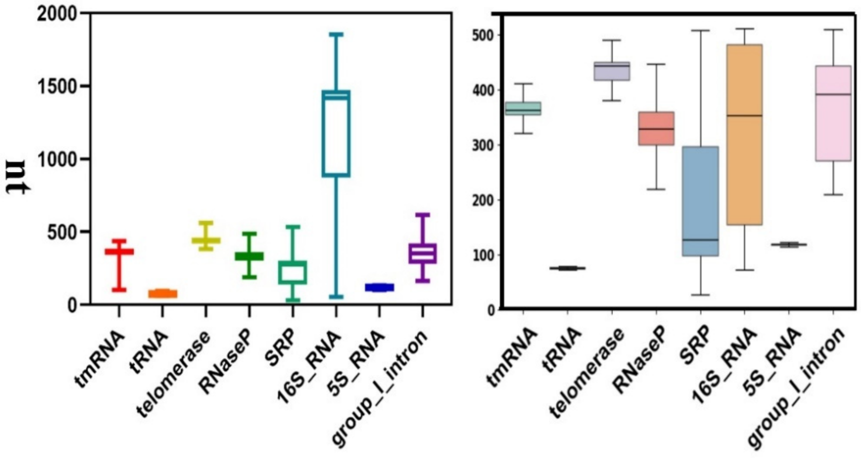

Figure 5. The length distribution of sequences in dataset RNAStralign (left) and Archive II (right) for each family.

3.1. The performance for predicting secondary structure of short sequences on RNAStralign

In this section of the experiment, we present the results of TransUFold on dataset RNAStralign and compare them with other state-of-the-art methods, including thermodynamic-based methods such as RNAfold, RNAstructure and LinearFold, machine learning-based methods such as CONTRAfold and ContextFold, hybrid machine learning and thermodynamic-based methods MXfold and MXfold2, and recently developed deep learning methods E2Efold, ATTfold and UFold.

The experimental results are presented in Table 3 and Figure 6. We find that the conventional thermodynamic-based methods yield F1 scores ranging from 0.671 to 0.719. In contrast, machine learning-based methods CONTRAfold and ContextFold achieve better performance, with F1 scores of 0.726 and 0.904, respectively. Hybrid machine learning and thermodynamic-based methods MXfold and MXfold2 also performs better than pure thermodynamic-based methods, with F1 score of 0.764 and 0.835, respectively. All mentioned deep learning-based methods obtain F1 scores exceeding 0.8. The F1 scores of E2Efold, ATTfold and UFold are 0.840, 0.813 and 0.945, respectively. Our TransUFold reaches the highest F1 score 0.951, which also outperforms other methods on the Recall and Precision indicators.

Table 3. Results on RNAStralign test set.

Method	Precision	Recall	F1
TransUFold	0.959	0.944	0.951
UFold	0.956	0.936	0.945
MXfold2	0.851	0.827	0.835
ATTfold	0.808	0.824	0.813
E2Efold	0.827	0.863	0.840
MXfold	0.809	0.734	0.764
ContextFold	0.927	0.887	0.904
CONTRAfold	0.745	0.718	0.726
LinearFold	0.781	0.685	0.719
RNAstructure	0.699	0.645	0.665
RNAfold	0.697	0.656	0.671

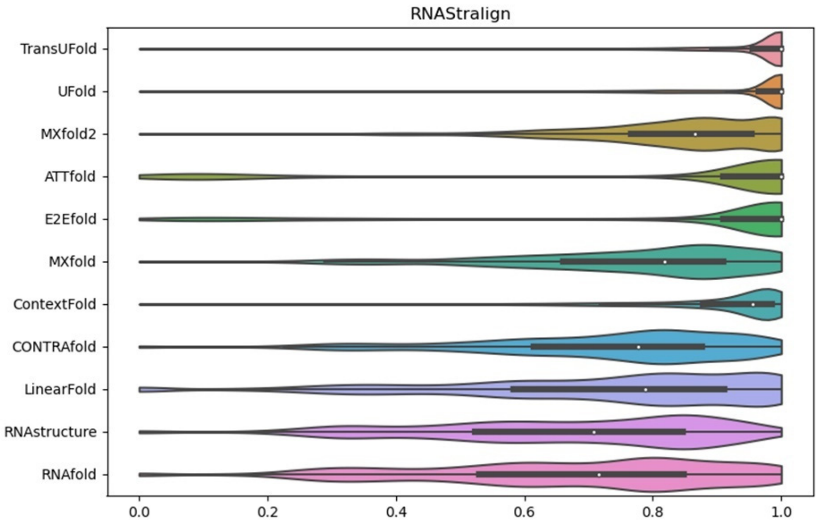

Figure 6. Violin plot visualization of F1 value of RNA secondary structure predictions methods on dataset RNAStralign.

3.2. The performance of RNA secondary structures prediction for known families

In this section, we compare our TransUFold to the methods mentioned above on dataset ArchiveII with different distributions of the same families and illustrate the results in Table 4 and Figure 7. The experiments show similar results that the F1 scores of traditional thermodynamic-based methods ranged from 0.623 to 0.646. Although the majority of machine learning-based and deep learning-based methods obtain better performance, E2Efold and ATTfold achieving F1 scores of just 0.552 and 0.524, respectively, lower than the performance of traditional thermodynamic-based methods. The probability distributions of F1 scores for E2Efold and ATTfold display significant polarization in Fig. 7, which can explain the fact that they are unable to forecast RNA secondary structures accurately when the RNA secondary structure homology is low. Our proposed TransUFold still achieves the best results, with F1 score of 0.866.

Table 2. The performance of methods on ArchiveII.

Method	Precision	Recall	F1
TransUFold	0.886	0.854	0.866
UFold	0.879	0.833	0.853
MXfold2	0.807	0.767	0.781
ATTfold	0.515	0.548	0.524
E2Efold	0.547	0.583	0.552
MXfold	0.670	0.764	0.707
ContextFold	0.874	0.821	0.842
CONTRAfold	0.698	0.654	0.668
LinearFold	0.730	0.604	0.646
RNAstructure	0.662	0.602	0.623
RNAfold	0.663	0.614	0.631

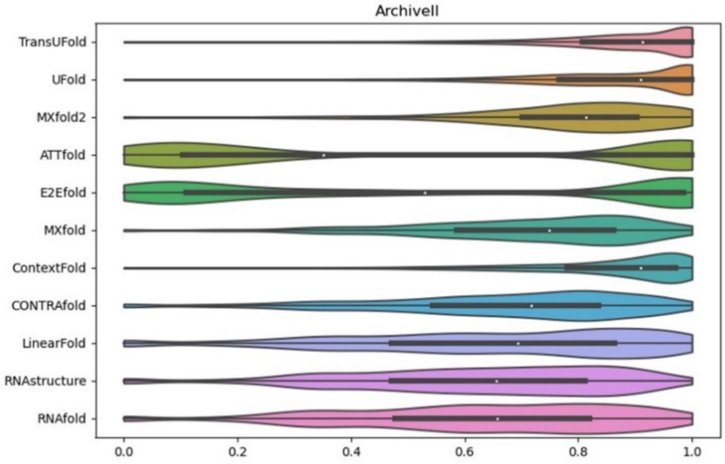

Figure 7. Violin plot visualization of F1 value of 11 RNA secondary structure predictions methods on dataset ArchiveII.

3.3. The predicted accuracy of RNA secondary structures for unknown families

Since dataset bpRNA-new consists of 1500 families of sequence structures that are not present in any other datasets, we utilize it to demonstrate the performance of TransUFold for unknown families, shown in Table 5 and Figure 8. Traditional thermodynamic methods achieve the similar accuracy, but the performance of machine learning-based methods and deep learning methods all decrease. The F1 score of the hybrid method MXfold reaches 0.663, showing the best performance. Unfortunately, the performance of all deep learning methods has significantly declined. This indicates that solely relying on deep learning methods cannot predict RNA secondary structures accurately when no prior knowledge is provided in the training process. E2Efold and ATTfold only have F1 scores of 0.051 and 0.059, respectively, which are too low to predict RNA secondary structures in unknown families. In comparison to other deep learning techniques, our TransUFold achieves an F1 score of 0.421, demonstrating relatively better performance.

Table 5. Results on bpRNA-new test set.

Method	Precision	Recall	F1
TransUFold	0.437	0.428	0.421
UFold	0.453	0.413	0.423
MXfold2	0.621	0.718	0.654
ATTfold	0.054	0.069	0.059
E2Efold	0.074	0.061	0.051
MXfold	0.638	0.717	0.663
ContextFold	0.596	0.636	0.603
CONTRAfold	0.620	0.736	0.661
LinearFold	0.686	0.646	0.633
RNAstructure	0.579	0.718	0.630
RNAfold	0.592	0.720	0.640

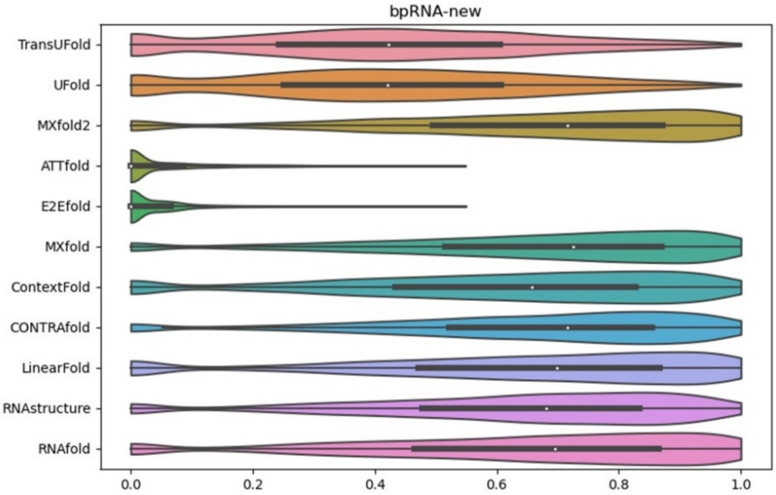

Figure 8. Violin plot visualization of F1 value of RNA secondary structure predictions methods on dataset bpRNA-new.

3.4. The performance for predicting secondary structures of long RNA sequences

In RNA secondary structure prediction, the lengths of RNA sequences can vary greatly. To accommodate deep learning methods, sequences must be padded to a uniform length, which not only increases training complexity but also adds too much meaningless information to short sequences, potentially reducing prediction accuracy. Therefore, most deep learning methods do not support long sequence prediction. For example, ATTfold models only RNA sequences with less than 512 nt. However, RNA sequences with longer lengths are equally important, so we evaluate the secondary structures of RNA sequences up to 1600 nt. In the experimental results, F1, Precision and Recall score 0.954, 0.932 and 0.978, respectively. Notably, compared with sequences shorter than 512 nt in the dataset RNAStralign, we even achieve better performance on long sequences. This is because the self-attention mechanism in Transformer provides long-range interactions for long RNA sequences.

3.5. Pseudoknot prediction analysis

In this section, we evaluate whether our model produces real and effective pseudoknots on the dataset RNAStralign. There exists a total of 3894 sequences less than 512 nt in the test set. After filtering of pseudoknots, 478 sequences with pseudoknots exist. We then access the performance of various methods (ProbKnot, E2Efold, Attfold, UFold and TransUFold) for generating pseudoknots on these sequences. We preserve all the bases that form pseudoknots in these sequences and only analyze the performance indicators of these bases' prediction, shown in Table 6. Our proposed TransUFold outperforms other methods in terms of F1 score, Precision and Recall.

Finally, we analyze TransUFold's performance on long sequences with pseudoknots between 512 and 1600 nt in length. We discover that 2131 of 2178 sequences in the test set contain pseudoknots, which means almost all long sequences contain pseudoknots. Therefore, the accuracy of pseudoknot prediction is crucial for predicting the secondary structures of long sequences. Our analysis of the 2131 sequences' base pairs directly involved in constructing pseudoknots yields F1, Precision and Recall of 0.960, 0.942 and 0.996, respectively. The training set containing a sufficient number of sequences with pseudoknots accounts for the greater accuracy compared to short sequences.

Table 6. Evaluation of base pairs involved in pseudoknot formation.

Method	**Precision**	**Recall**	**F1**
TransUFold	0.455	0.952	0.560
UFold	0.434	0.938	0.551
E2Efold	0.200	0.657	0.267
Attfold	0.067	0.387	0.103
ProbKnot	0.140	0.814	0.172

3.6. Ablation study

To validate the effectiveness of the Vision Transformer-based encoder we employed, we conducted an ablation experiment in which we removed the Vision Transformer module (TransUFold-WVT) and compared it with TransUFold. This experiment helps us assess the contribution of the Vision Transformer to the system's performance. The experimental results are presented in Table 7 and

Figure 9. TransUFold achieved superior results with an F1 score of 0.952. As the sequence length increased, the advantage of TransUFold became more noticeable, indicating that the attention-based Vision Transformer is more effective in handling long-range interactions between bases.

Table 7. Ablation study on VIT (RNAStralign testing set).

Method	Precision	Recall	F1
TransUFold	0.949	0.956	0.952
TransUFold-WoVT	0.935	0.941	0.937

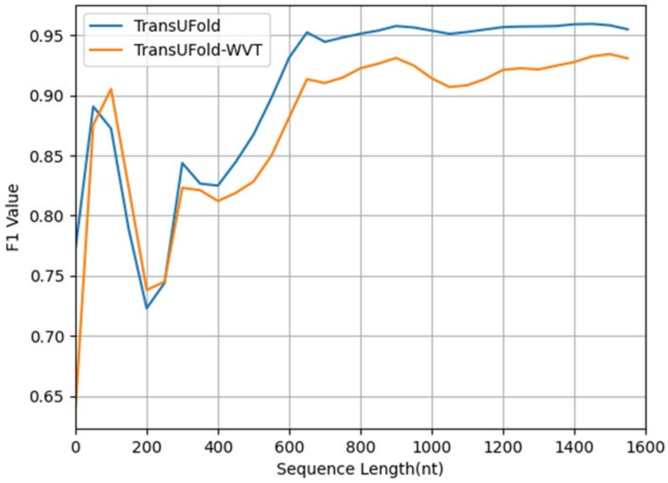

Figure 9. Analyzing F1 scores based on sequence length.

3.7. Visualization

In this section, we compare the visualized output RNAs of TransUFold to MXfold2, E2Efold, ATTfold, UFold, Contextfold and RNA structure methods. We convert the predicted RNA secondary structure matrix into a bpseq format and visualize RNA sequence P00855 (Figure 10) from dataset RNAStralign and 5s_Methanosarcina-acetivorans-2 (Figure 11) from dataset ArchiveII using VARNA, a visualization tool of RNA sequences. In both examples, TransUFold produces results that are most similar to the actual RNA secondary structure. Finally, we also visualize an 872-length RNA sequence in Figure 12, which is recorded in dataset RNAStralign as AY807427. TransUFold also yields high similarity to the actual structure on longer sequences.

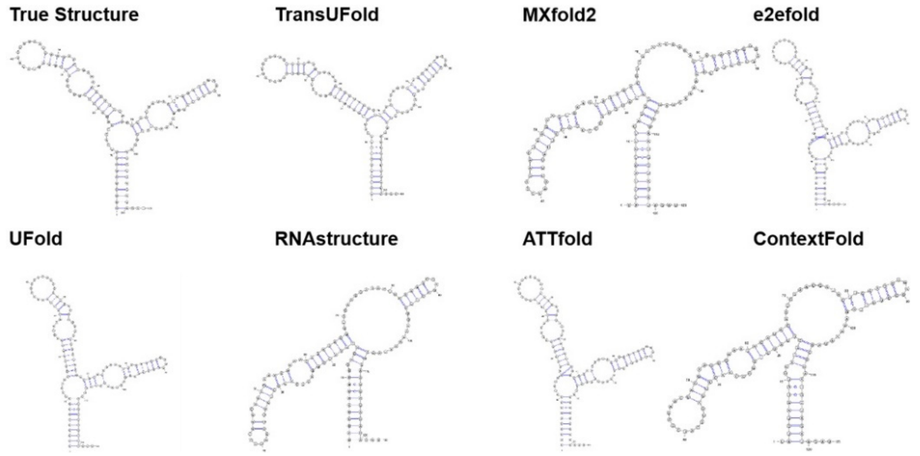

Figure 10. Visualization of RNA secondary structure prediction for P00855 in dataset RNAStralign.

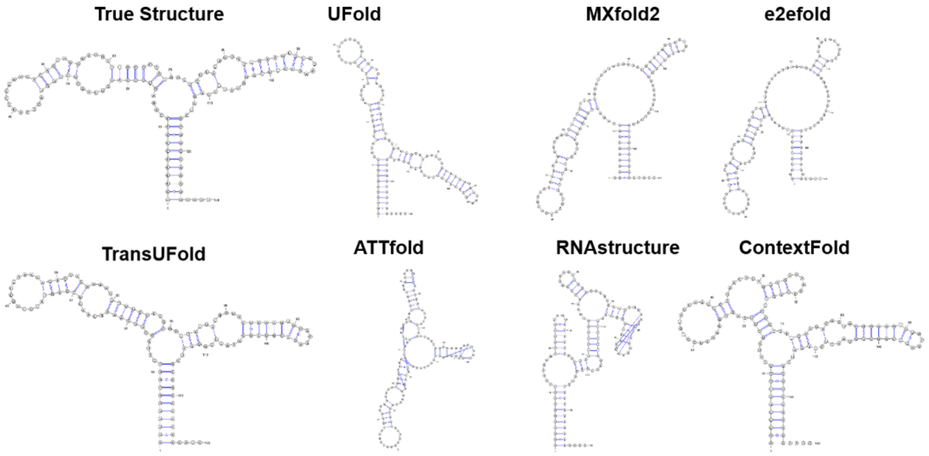

Figure 11. Visualization of RNA secondary structure prediction for 5s_Methanosarcina - acetivorans-2 in dataset ArchiveII.

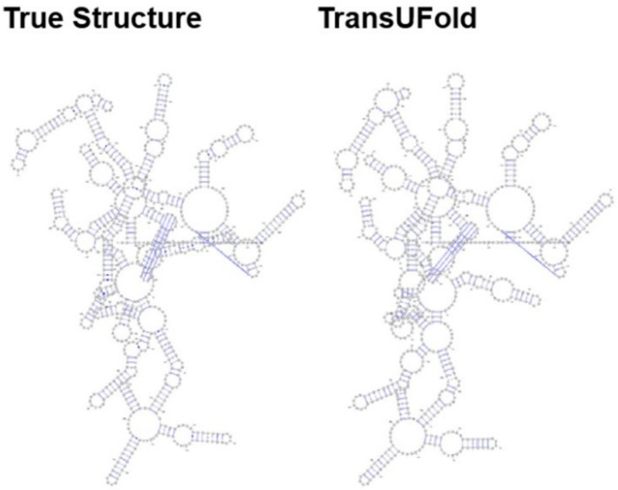

Figure 12. Visualization of RNA secondary structure prediction for AY807427 in dataset RNAStralign, with a sequence length of 872.

4. Discussion

In our experiment, dataset RNAStralign is used to train the proposed model TransUFold, and F1 score is applied as the main performance metric during testing. TransUFold performs 20−30% more accurately than thermodynamic methods. TransUFold achieves 5% to 20% improvement when comparing with other machine learning methods. Our TransUFold performs up to 10% better than other cutting-edge deep learning methods. We also design experiments for different scenarios. For instance, when predicting the structure of newly discovered RNA sequences belonging to known families in dataset ArchiveII, TransUFold outperforms the other methods. When predicting the structures of RNA sequences from previously unknown families in dataset bpRNA-new, TransUFold's performance decreases by over 20% compared to traditional thermodynamic methods. However, TransUFold performs relatively better in terms of accuracy (F1 = 0.4), compared with other deep learning methods like E2Efold and ATTfold with F1 scores less than 0.1. When predicting longer RNA sequences up to 1600 nt, TransUFold performs well, since the primary encoder can capture long-range interactions in RNA sequences. The F1 score is 0.954 for sequences longer than 512 nt, which is higher than the performance for short sequences. Finally, we verify the practicability of pseudoknots prediction. In short sequences, our method achieves the best performance, but it is still challenging to predict pseudoknots precisely. In long sequences up to 1600 nt, the prediction of TransUFold for pseudoknots achieves the same performance as predicting other base pairs and can generate effective pseudoknots.

Our TransUFold achieves superior performance compared to previous methods due to its distinctive network architecture. We employ an image-like input transformed from the original RNA sequence instead of the raw sequence used by the majority of methods. The advantage of this approach is that all base pairing patterns are explicitly represented in the image-like input, allowing the proposed model to select all potential base pairing rules that can contribute to the construction of RNA secondary structure. For the output of RNA secondary structure prediction, traditional methods and other deep learning methods utilize dot-bracket notation to represent RNA secondary structures without pseudoknots, which cannot accurately reflect the genuine activity of RNA. In contrast, our TransUFold method outputs a 2D base pairing matrix after complying with realistic rules through post-processing, which easily conveys the structure of pseudoknots. Furthermore, in our proposed model, Vision Transformer is applied as the primary encoder to capture long-range interactions in RNA sequences, and convolutions with lateral connections are introduced as the auxiliary encoder for extracting the extra short-range interactions to the decoder. This network architecture that combines local and global features is particularly effective for predicting RNA secondary structures. However, TransUFold does have some limitations like other deep learning-based methods. The prediction performance decreases significantly in the absence of enough available known structures, as demonstrated in our experiment's second scenario. Fortunately, high-throughput methods for determining RNA structures have begun to emerge [42,43]. With a large number of available structures, TransUFold's performance can be further improved.

In the future, we consider combining deep learning with meta-learning to achieve adaptive learning and generalization capabilities. Environmental conditions, such as temperature, ion concentration and pH value, also have an impact on RNA secondary structure in actual studies. It is meaningful to take this information into account to make deep learning tools more useful in practical applications. It is important to note that the constraint "no sharp loops are allowed" in our model is indeed a simplified description and does not consider structures like bulges and internal loops. We plan to design another classifier to loosen this constraint later. Overall, we demonstrate the outstanding potential of deep learning methods to address the challenge of RNA secondary structure prediction. Our proposed network architecture that utilizes a Vision Transformer fused with an auxiliary encoder with lateral convolutions provides a solution for capturing both long-range and short-range interactions within RNA sequences and outperforms other state-of-the-art methods in RNA structure prediction. Without prior knowledge of RNA sequences within the same family, it is difficult for deep learning methods to make accurate predictions. However, along with the development of technologies for RNA structure detection, the performance of deep learning methods like TransUFold can be further enhanced. The implemented code and experimental dataset are available online at https://github.com/traveltheroad/TransUFold.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Author contributions statement

Y. W., H. Z. and R. G. conceived the idea and supervised the work. Y. W. and Z. X. performed the research and analyzed the data. Y. W. and H. Z. wrote the first draft of the manuscript. H. Z., R. G and S. Z. reviewed and revised the manuscript. All authors approved the manuscript.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (#32171181, RG), the Hundred Talents Funding Program of Hebei Province (#E2019050010, RG) and the Interdisciplinary Research Program of Natural Science of Hebei University (#DXK202105, R).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. J. A. Shapiro, Revisiting the central dogma in the 21st century, *Ann. N. Y. Acad. Sci.*, **1178** (2009), 6−28. https://doi.org/10.1111/j.1749-6632.2009.04990.x
2. T. A. Lincoln, G. F. Joyce, Self-sustained replication of an RNA enzyme, *Science*, **323** (2009), 1229−1232. https://doi.org/10.1126/science.1167856
3. P. V. Ryder, D. A. Lerit, RNA localization regulates diverse and dynamic cellular processes, *Traffic*, **19** (2018), 496−502. https://doi.org/10.1111/tra.12571
4. E. Westhof, P. Auffinger, RNA tertiary structure, in *Encyclopedia of Analytical Chemistry*, (2000), 5222−5232. https://doi.org/10.1002/9780470027318.a1428
5. F. E. Reyes, C. R. Schwartz, J. A. Tainer, R. P. Rambo, Methods for using new conceptual tools and parameters to assess RNA structure by small-angle X-ray scattering, *Methods Enzymol.*, **549** (2014), 235−263. https://doi.org/10.1016/B978-0-12-801122-5.00011-8

6. C. Helmling, S. Keyhani, F. Sochor, B. Fürtig, M. Hengesbach, H. Schwalbe, Rapid NMR screening of RNA secondary structure and binding, *J. Biomol. NMR*, **63** (2015), 67−76. https://doi.org/10.1007/s10858-015-9967-y
7. R. Stark, M. Grzelak, J. Hadfield, RNA sequencing: the teenage years, *Nat. Rev. Genet.*, **20** (2019), 631−656. https://doi.org/10.1038/s41576-019-0150-2
8. M. Zuker, P. Stiegler, Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information, *Nucleic Acids Res.*, **9** (1981), 133−148. https://doi.org/10.1093/nar/9.1.133
9. D. H. Turner, D. H. Mathews, NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure, *Nucleic Acids Res.*, **38** (2010), D280−D282. https://doi.org/10.1093/nar/gkp892
10. M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Res.*, **31** (2003), 3406−3415. https://doi.org/10.1093/nar/gkg595
11. N. R. Markham, M. Zuker, UNAFold: software for nucleic acid folding and hybridization, in *Bioinformatics*, **453** (2008), 3−31. https://doi.org/10.1007/978-1-60327-429-6_1
12. I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, P. Schuster, Fast folding and comparison of RNA secondary structures, *Monatsh. Chem. Mon.*, **125** (1994), 167−188. https://doi.org/10.1007/BF00818163
13. S. Bellaousov, J. S. Reuter, M. G. Seetin, D. H. Mathews, RNAstructure: web servers for RNA secondary structure prediction and analysis, *Nucleic Acids Res.*, **41** (2013), W471−W474. https://doi.org/10.1093/nar/gkt290
14. L. Huang, H. Zhang, D. Deng, K. Zhao, K. Liu, D. A. Hendrix, et al., LinearFold: linear-time approximate RNA folding by 5'-to-3'dynamic programming and beam search, *Bioinformatics*, **35** (2019), i295−i304. https://doi.org/10.1093/bioinformatics/btz375
15. X. Wang, J. Tian, Dynamic programming for NP-hard problems, *Procedia Eng.*, **15** (2011), 3396−3400. https://doi.org/10.1016/j.proeng.2011.08.636
16. E. Rivas, S. R. Eddy, A dynamic programming algorithm for RNA structure prediction including pseudoknots, *J. Mol. Biol.*, **285** (1999), 2053−2068. https://doi.org/10.1006/jmbi.1998.2436
17. R. M. Dirks, N. A. Pierce, A partition function algorithm for nucleic acid secondary structure including pseudoknots, *J. Comput. Chem.*, **24** (2003), 1664−1677. https://doi.org/10.1002/jcc.10296
18. X. Xu, P. Zhao, S. J. Chen, Vfold: a web server for RNA structure and folding thermodynamics prediction, *PloS One*, **9** (2014), e107504. https://doi.org/10.1371/journal.pone.0107504
19. K. Sato, M. Hamada, Recent trends in RNA informatics: a review of machine learning and deep learning for RNA secondary structure prediction and RNA drug discovery, *Briefings Bioinf.*, **24** (2023). https://doi.org/10.1093/bib/bbad186
20. T. Gong, F. Ju, D. Bu, Accurate prediction of RNA secondary structure including pseudoknots through solving minimum-cost flow with learned potentials, *bioRxiv*, (2022). https://doi.org/10.1101/2022.09.19.508461
21. J. Ren, B. Rastegari, A. Condon, H. H. Hoos, HotKnots: heuristic prediction of RNA secondary structures including pseudoknots, *RNA*, **11** (2005), 1494−1504. https://doi.org/10.1261/rna.7284905

22. K. Sato, Y. Kato, M. Hamada, T. Akutsu, K. Asai, IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming, *Bioinformatics*, **27** (2011), i85−i93. https://doi.org/10.1093/bioinformatics/btr215
23. C. B. Do, D. A. Woods, S. Batzoglou, CONTRAfold: RNA secondary structure prediction without physics-based models, *Bioinformatics*, **22** (2006), e90−e98. https://doi.org/10.1093/bioinformatics/btl246
24. S. Zakov, Y. Goldberg, M. Elhadad, M. Ziv-Ukelson, Rich parameterization improves RNA structure prediction, in *Research in Computational Molecular Biology*, **18** (2011), 1525−1542. https://doi.org/10.1007/978-3-642-20036-6_48
25. M. Akiyama, K. Sato, Y. Sakakibara, A max-margin training of RNA secondary structure prediction integrated with the thermodynamic model, *J. Bioinf. Comput. Biol.*, **16** (2018), 1840025. https://doi.org/10.1142/S0219720018400255
26. K. Sato, M. Akiyama, Y. Sakakibara, RNA secondary structure prediction using deep learning with thermodynamic integration, *Nat. Commun.*, **12** (2021), 941. https://doi.org/10.1038/s41467-021-21194-4
27. H. Zhang, C. Zhang, Z. Li, C. Li, X. Wei, B. Zhang, et al., A new method of RNA secondary structure prediction based on convolutional neural network and dynamic programming, *Front. Genet.*, **10** (2019), 467. https://doi.org/10.3389/fgene.2019.00467
28. J. Singh, J. Hanson, K. Paliwal, Y. Zhou, RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning, *Nat. Commun.*, **10** (2019), 5407. https://doi.org/10.1038/s41467-019-13395-9
29. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), 770−778. https://doi.org/10.1109/CVPR.2016.90
30. Z. Huang, W. Xu, K. Yu, Bidirectional LSTM-CRF models for sequence tagging, preprint, arXiv:1508.01991.
31. X. Chen, Y. Li, R. Umarov, X. Gao, L. Song, RNA secondary structure prediction by learning unrolled algorithms, preprint, arXiv:2002.05810.
32. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., Attention is all you need, preprint, arXiv:1706.03762.
33. L. Fu, Y. Cao, J. Wu, Q. Peng, Q. Nie, X. Xie, et al., UFold: fast and accurate RNA secondary structure prediction with deep learning, *Nucleic Acids Res.*, **50** (2022), e14. https://doi.org/10.1093/nar/gkab1074
34. K. Darty, A. Denise, Y. Ponty, VARNA: interactive drawing and editing of the RNA secondary structure, *Bioinformatics*, **25** (2009), 1974. https://doi.org/10.1093/bioinformatics/btp250
35. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al., An image is worth 16x16 words: transformers for image recognition at scale, preprint, arXiv:2010.11929.
36. Z. Tan, Y. Fu, G. Sharma, D. H. Mathews, TurboFold II: RNA structural alignment and secondary structure prediction informed by multiple homologs, *Nucleic Acids Res.*, **45** (2017), 11570−11581. https://doi.org/10.1093/nar/gkx815
37. M. F. Sloma, D. H. Mathews, Exact calculation of loop formation probability identifies folding motifs in RNA secondary structures, *RNA*, **22** (2016), 1808−1818. https://doi.org/10.1261/rna.053694.115

38. I. Kalvari, E. P. Nawrocki, N. Ontiveros-Palacios, J. Argasinska, K. Lamkiewicz, M. Marz, et al., Rfam 14: expanded coverage of metagenomic, viral and microRNA families, *Nucleic Acids Res.*, **49** (2021), D192−D200. https://doi.org/10.1093/nar/gkaa1047
39. Y. Wang, Y. Liu, S. Wang, Z. Liu, Y. Gao, H. Zhang, et al., ATTfold: RNA secondary structure prediction with pseudoknots based on attention mechanism, *Front. Genet.*, **11** (2020), 612086. https://doi.org/10.3389/fgene.2020.612086
40. J. D. Watson, F. H. C. Crick, Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid, *Nature*, **171** (1953), 737−738. https://doi.org/10.1038/171737a0
41. G. Varani, W. H. McClain, The G·U wobble base pair, *EMBO Rep.*, **1** (2000), 18−23. https://doi.org/10.1093/embo-reports/kvd001
42. E. J. Strobel, A. M. Yu, J. B. Lucks, High-throughput determination of RNA structures, *Nat. Rev. Genet.*, **19** (2018), 615−634. https://doi.org/10.1038/s41576-018-0034-x
43. S. Lusvarghi, J. Sztuba-Solinska, K. J. Purzycka, J. W. Rausch, S. F. J. Le Grice, RNA secondary structure prediction using high-throughput SHAPE, *Biology*, **2013** (2013), e50243. https://doi.org/10.3791/50243-v