Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

Deep learning and attention mechanisms in RNA secondary structure prediction: A critical survey

Musaab Nabil Ali Askar [1], Azian Azamimi Abdullah [1, 2, *], Mohd Yusoff Mashor [1], Zeti-Azura Mohamed-Hussein [3], Zeehaida Mohamed [4], Wei Chern Ang [5], Shigehiko Kanaya [6]

[1]*Faculty of Electronic Engineering and Technology, Universiti Malaysia Perlis, Arau, Malaysia*
[2]*Advanced Sensor Technology, Centre of Excellence (CEASTech), Universiti Malaysia Perlis (UniMAP), Arau, Malaysia*
[3]*Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia*
[4]*Department of Medical Microbiology and Parasitology, School of Medical Sciences, Universiti Sains Malaysia, George Town, Malaysia*
[5]*Clinical Research Centre, Hospital Tuanku Fauziah, Ministry of Health Malaysia, Perlis, Kangar, Malaysia*
[6]*Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma, Nara, Japan*

ARTICLE INFO

Article history:
Received 1 March 2025
Received in revised form
25 July 2025
Accepted 7 August 2025

Keywords:
RNA structure
Deep learning
Transformer models
Thermodynamic methods
Pseudoknot prediction

ABSTRACT

The secondary structure of ribonucleic acid (RNA) plays a key role in understanding gene regulation, cellular processes, and the development of new treatments. Traditional thermodynamic methods, especially those using Minimum Free Energy (MFE) algorithms, have provided a reliable physics-based approach for predicting RNA structures. Although these methods remain important, there is increasing interest in using deep learning models to detect new structural patterns, such as pseudoknots and long-range interactions, in large RNA datasets. Building on thermodynamic principles, these models aim to extend current knowledge and offer new ways to study RNA structure and function. In particular, attention-based transformer models are effective at capturing both short- and long-distance relationships, making them well-suited for modeling complex RNA sequences. This review highlights recent advances in RNA secondary structure prediction using transformer-based approaches, focusing on key models such as E2EFold, ATTFold, RNAformer, and DEBFold. It also discusses current challenges, future research directions, and the impact of attention-based deep learning on the field of RNA structural bioinformatics.

1. Introduction

The growing awareness of ribonucleic acid (RNA) significance in gene regulation has inspired a rush of interest in studying its secondary structure and more complex tertiary conformation (Mailler et al., 2019; Morris and Mattick, 2014). The rapidly expanding field of RNA research and structural analysis has opportunities for further growth. Recent research highlights transformers' effectiveness in modeling complex genetic relationships, motivating their use in RNA structural analysis (Choi and Lee, 2023). This indicates the ability to comprehend data by dynamically analyzing different genetic regions and their relationships and entails improving the creation of next-generation treatments and comprehending the process of RNA degradation and its connection to molecular structure, among other things (He et al., 2023).

RNA plays an important role in gene regulation, making its secondary and tertiary structures key research topics. Recent advances show that transformer models can capture complex genetic relationships, improving RNA structural prediction. This review introduces RNA structure, examines transformer-based approaches, and discusses their potential, challenges, and future directions in enhancing RNA bioinformatics and therapeutic development.

We review key attention-based models, assessing their ability to improve RNA structure prediction in both accuracy and scalability. Finally, we address ongoing challenges, explore future opportunities, and emphasize how progress in attention-based deep learning is expected to significantly advance RNA structural bioinformatics research and applications.

* Corresponding Author.
Email Address: azamimi@unimap.edu.my (A. A. Abdullah)
https://doi.org/10.21833/ijaas.2025.09.006
Corresponding author's ORCID profile:
https://orcid.org/0000-0001-5851-7705

2. RNA structure

A single-stranded RNA molecule undergoes RNA folding when it creates base-pair interactions and connections to form specific three-dimensional configurations (Schärfen and Neugebauer, 2021). The biological activity of RNA depends on its folded structure because this feature determines RNA interactions with proteins or small ligands, hence making this process vital for RNA's practicality. The combination of RNA nucleotide sequence and its environment, with ion presence and other elements, affects how RNA molecules dynamically self-fold. The ability of RNA molecules to perform biological activities, including gene expression and virus replication, alongside RNA-protein interactions, depends completely on the RNA molecules' folding process (Graf and Kretz, 2020; Hou and Jaffrey, 2023).

The chemical bonds between nucleotides cause RNA to develop hairpin loops, together with bulges and internal loops for structural maintenance. Studying RNA structure alterations under changing environmental conditions and chemical situations provides essential information about RNA structure functions. Loop size and composition, together with the neighboring elements, determine loop stability (Balcerowicz et al., 2021). Hairpin loops play a role in gene regulation by enabling ribozymes to catalyze reactions. On the other hand, interior loops and bulges provide RNA with the flexibility required for chemical interactions (Kuznetsov et al., 2008). Watson-Crick base pairing within RNA helices maintains the entire structure by preserving RNA integrity. The collection of structural motifs performs essential roles in RNA biological functions, making them crucial to study in more detail. RNA molecules use their three-dimensional arrangement to form tertiary structures in addition to their hairpins and loop configurations. RNA performs its molecular functions through these patterns that create spatial arrangements and establish contact points with other biomolecules. A loop functions by linking nucleotides with sequences that exist beyond the loop structure to generate pseudoknots (Bravo et al., 2021; Peselis and Serganov, 2014).

Similarly, RNA's extraordinary ability to catalyze specific chemical reactions on its own is demonstrated by ribozymes, which are catalytically active RNA molecules. This is particularly evident in RNA's capacity to self-cleave without substantial protein interference (Weinberg et al., 2019). Moreover, the intersection of many strands of RNA forms complicated junctions that are crucial for the determination of the tertiary shape of the molecule, thereby determining how stable it is, its activity, and its interaction with other molecules. The merging of structural characteristics underscores the elaborate architecture of RNA and its essential involvement in the cell environment. In Fig. 1, tertiary structure folding and interactions, as well as secondary structure motifs, including hairpin loops, bulges, and internal loops, are shown very vividly.

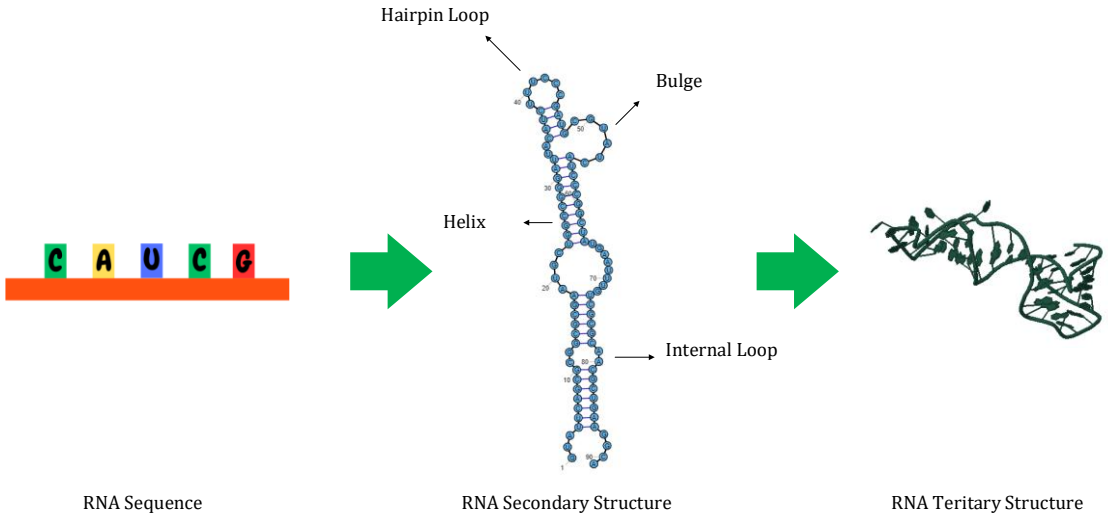

Fig. 1: RNA structures visualization

In Fig. 1, the left panel shows the basic RNA sequence characterized by its single-stranded nature and its nucleotides. The middle panel showcases RNA secondary structural motifs, including hairpin loops, bulges, internal loops, and helices, as predicted by RiboSketch (Lu et al., 2018). The right panel provides a conceptual representation of RNA's tertiary structure, highlighting the complex folding and interactions as rendered by RNAComposer (Sarzynska et al., 2023).

2.1. Thermodynamic models: Minimum free energy (MFE)

The beginnings of RNA secondary structure prediction tools in computational biology were intended to understand the process of RNA folding into its thermodynamically most stable conformation. The field initially focused on the identification of minimum free energy (MFE), like conformations of RNA molecules, because RNA takes

the conformation that results in the minimum free energy under physiological conditions. This was an initial approach, and fundamental, foundational tools such as Mfold, RNAfold, tRNAscan-SE, and RNA structure were developed to implement these rules using dynamic programming techniques. These methods, while pioneering in their own respect, defined a base from which improvements to RNA computational biology could be made (Hofacker, 2003; Reuter and Mathews, 2010; Zuker and Stiegler, 1981). Zuker and Stiegler (1981) set the basis for this work in 1981, initiating the tradition of an effective computational method for the corresponding future works. They assumed that RNA molecules fold to conform to the structure that corresponds to the minimum free energy; this is a shift from experimentation to the computational method (Zuker and Stiegler, 1981). To construct their model, they used several assumptions or components. Initially, they applied the available thermodynamic parameters to analyze the free energy of specificity stems, loops, and junctions in RNA molecules. This data supported the guiding assumption that RNA would adopt the lowest-energy conformation, which is a principle that originates from the molecular stability principle commonly found in physical chemistry. This thermodynamic framework brought further biological roots to MFE approaches and enabled predictions to meet with RNA stability inclination. Aside from the thermodynamic predictions, they used the dynamic programming algorithm to make the prediction process manageable (Nussinov and Jacobson, 1980). This algorithm enabled them to search for the potential base-pairing configurations without assessing each of them separately, a complication imposed by the fact that longer RNA sequences can potentially fold into millions or billions of conformations. In this manner, by dividing the problem into recursive subproblems, they were able to calculate the free energy for each segment of the RNA sequence and guarantee that the values would cover the configuration, which minimized the total free energy. They also adhered to conventional RNA base-pairing guidelines; they used Watson-Crick pairing and occasional GU wobble pairs. These rules also helped to eliminate more pairings that the program did not have to consider because they improved the algorithm's performance. This was the algorithm, one of the first and still one of the most widely used RNA structure prediction methods.

2.2. Machine learning in RNA secondary structure prediction

Following the discussion of thermodynamic principles governing RNA folding, attention is now turned to machine learning techniques that utilize data-driven strategies in RNA secondary structure prediction. MFE-based models have traditionally provided a sound thermodynamic basis for a priori prediction of RNA secondary structure, reporting details about probable stable folds. Nevertheless, Real RNA molecules can adopt other conformations, including pseudoknots, that could be biologically relevant but are not captured by single-conformation methods. In parallel, access to RNA sequencing data facilitated machine learning strategies, initially mooted in the early 2000s, trained to discover patterns within large datasets directly. Data-driven methods extend predictive capabilities and complement the thermodynamic view, particularly for structural features and variations beyond the realm of classical MFE predictions.

Among the first RNA structure prediction methods using machine learning is CONTRAfold (Do et al., 2006). It uses a conditional log-linear model, a type of generalized probabilistic model, to predict RNA secondary structures by computing probabilities for potential structures based on factors such as base-pairing propensity and sequence-specific features that are derived from the training data. Contrary to the conventional models, which depend on a predetermined set of thermodynamic parameters, it is based on the feature representation inferred from the current set of RNA structures. Moreover, it adopts a data-driven approach, assigning appropriate weights to the features using structural data. Given an RNA sequence, CONTRAfold computes the partition function to get the sum probability of all possible structures that are useful for probable pairing and pointing out the structures with the sum expected accuracy in their structures. It made use of probabilistic graphical models in connection with supervised learning, which was beneficial as it provided better structures in the structures of diverse RNA sequences than other thermodynamic models. This innovation marked a significant shift: Instead of simply seeking the minimum energy configuration, it is possible to train machine learning models that evaluate the probability of distinct structural configurations through characteristics obtained from realistic data.

2.3. Deep learning in RNA secondary structure prediction

Extending the foundation of machine learning, deep learning architectures have significantly enhanced the capability to model intricate RNA folding designs. It is a branch of machine learning sciences that has recently come into focus in data-oriented approaches owing to the program's capacity to learn intricate nonlinear patterns (Pan et al., 2022). It has been particularly beneficial in the medical field in the analysis of medical pictures with improvements in the component element characterization and medical application (Ursuleanu et al., 2021). It encompasses various architectures, including Convolutional Neural Networks (CNN) (O'shea and Nash, 2015) and Recurrent Neural Networks (RNN) (Sherstinsky, 2020). CNNs have turned out to be triumphant in Computer Vision tasks, including Object Recognition, Image Classification, and Semantic Segmentation (Chai et

al., 2021). Coupled deep neural networks have been applied efficiently and effectively in deep learning techniques, becoming very helpful in predicting RNA secondary structure, followed by convolutional neural networks and then deep neural networks (Bliss et al., 2020; Chen et al., 2020b; Mao et al., 2022; 2020; Zhang et al., 2019; Zhao et al., 2023).

It has also been used for thermodynamics and deep learning RNA secondary structure profile prediction, including Bidirectional Long Short-Term Memory and Residual Neural Networks (Wang et al., 2021). Most important in this respect is the Atomic Rotationally Equivariant Scorer, ARES, described by Townshend et al. (2021), which illustrated how RNA structure prediction can very effectively be improved by machine learning when limited or no training data are available. Moreover, a reliable method for RNA secondary structure prediction has been made possible by the combination of deep learning and thermodynamic integration (Sato et al., 2021). Similarly, the studies by Ou et al. (2022), Solayman et al. (2022), Xu et al. (2022), and Yu et al. (2020) showed how deep learning can be used to help gain more accurate and efficient RNA structure prediction, further insight into structures and functions of RNA, and mix artificial intelligence techniques with traditional bioinformatics tools. Having outlined the underlying foundations of attention mechanisms, we go on to introduce and critically review the prevailing models for RNA secondary structure prediction that incorporate these techniques.

2.4. The integration of attention mechanisms and transformer architectures

The attention mechanism has been applied to an extensive range of topics in the domains of machine learning, natural language processing, and computer vision. It was applied to various architectures of neural networks, such as convolutional neural networks, recurrent neural networks, and graph neural networks (Fu et al., 2023; Guo et al., 2022; Li et al., 2020a; 2020b; Liao and Deng, 2020; Qian et al., 2022; Zhang et al., 2020; Zhong et al., 2020). In contrast to using a single context vector, the attention mechanism changes this focus dynamically on each step of output production on different segments of the input sequence. First, scores are computed between each encoder state and the current decoder state. This is then normalized to create a weight distribution. These weights are utilized to compute a context vector, the weighted sum of the states of an encoder, directing the creation of every output word (Hernández and Amigó, 2021). Moreover, the application of the attention mechanism to RNA secondary structure prediction algorithms has shown promising results in terms of enhancing the overall performance of the algorithm, handling long RNA sequence and pseudoknot issues, and predictive accuracy. For example, one of the algorithms called "ATTfold" overcomes the long sequence prediction problem by considering the global information of RNA sequences through the attention mechanism, which focuses on the base-pair correlation (Wang et al., 2020).

Transformers-based models are a breakthrough in machine learning, especially in the area of natural language processing. The Transformer model introduced by Vaswani et al. (2017) revolutionized sequence modeling by enabling parallel processing through self-attention mechanisms, making it highly suitable for analyzing RNA sequences. It is based on the idea that the attention processes can be applied to efficiently process sequential data like language. It allows parallel computing and solves other issues with previous RNN models, such as long-term dependency. Word embedding is the first step through which the transformer model takes words and changes them to their numeric forms. Each phrase or symbol in the predefined lexicon has its own numeric vector. This is done by some kind of basic neuron network, where each input (word or symbol) obtains a certain weight from some layer, and it will carry out a specific numeric representation for each word. The embeddings cater to the process of preparing the model to work with numerical data in the neural networks.

The order of words carries significant meaning. The positioning of the words within the sentence is very important. Positional encoding is used to address this in numerical forms. This is done by adding function values that denote the position of the words in each sentence within the word embeddings. Specifically, the functions for this position-specific value setting are in the form of sine and cosine with certain restrictions. Therefore, the Transformer model retains the sequencing in which the input is presented with the inclusion of positional encoding. The secret lies in the self-attention mechanism of the Transformer architecture, so that the model is able to evaluate the relative weights of the various words in a sentence while keeping the overall context of the sentence in perspective. This method calculates three sets of values for each word: query, key, and value. Interaction among these sets determines the self-attention scores, which guide the model to attend only to relevant parts of an input sequence. Therefore, it can handle longer sequences well, noting that the process is highly effective and very parallelizable. An exemplary implementation of a Transformer is way ahead of RNNs in managing very long sequences. The core idea of multi-head attention is to parallelize the computation of multiple self-attentive LSTM models. In multi-head attention, the mechanism of attending to oneself is executed several times. For all input elements, it calculates different query, key, and value vectors in each head position. Overall, various linear transformations can be executed. Finally, attention-weighted outputs from every head are concatenated. The encoder and decoder stack is an encoder-decoder structure where an input sequence is translated into an output sequence. The encoder processes the input while the decoder generates the

output. The critical stage in this process is attention, which aligns the input and output sequences. This guarantees that relevant features in the input will be duly reflected in the output, as the decoder is capable of attending to important parts of the input at each step of generating a word in the output. Multiple components beyond the basic transformer architecture can be added to increase performance, especially on complex tasks. Normalization layers are usually added after each stage to preserve the learning process within a stable range. The scaling of the attention mechanism can be taken further, and more neural network layers can be densely added to the encoder and decoder networks to increase the capacity of the model further. Fig. 2 illustrates the transformer architecture.

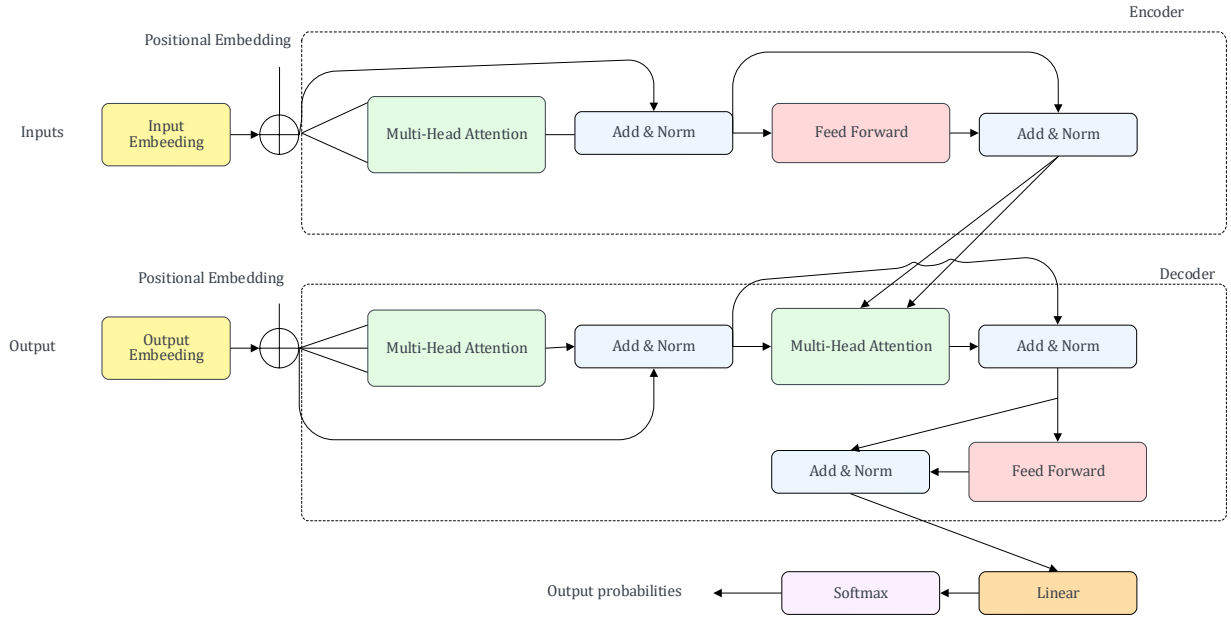

Fig. 2: Transformer architecture

2.5. Overview of approaches based on attention mechanisms and transformer architectures

Having introduced transformer architecture, we now examine how these models have been specifically adapted for RNA secondary structure prediction. The current review aims to extend the previous systematic literature review (SLR) performed by Budnik et al. (2024), which systematically assessed RNA structure prediction approaches. The original study was based on a search in Web of Science and Google Scholar, which identified 295 papers. From these, 33 were selected using strict inclusion and exclusion criteria. This review focuses on methods that apply attention mechanisms, a rapidly advancing area in RNA structure prediction. In addition to re-examining the methods discussed in the original study, this review also considers more recent approaches, offering a broader picture of current developments in the field. By combining earlier studies with the latest research, the review highlights the main trends, strengths, and limitations of attention-based methods for predicting RNA secondary and tertiary structures.

2.6. RNA structure prediction models

The advancement through incorporating attention mechanisms into RNA secondary structure prediction has dramatically improved the field of solving problems such as long-range dependencies, pseudoknot prediction, and computational scalability. This section considers their advancement based on performance metrics, literature, architecture, and the problems that they solve. Datasets are the foundation of RNA secondary structure prediction as they make up the training sets and the benchmarks that models are compared to. The reviewed models employ diverse datasets, and the differences in the structure of datasets, including families under consideration and their coverage, define the performance and, therefore, the extent of the model's applicability. The data sets used by the reviewed models are summarized in Table 1.

All of the reviewed models rely on at least one of these datasets for both training and evaluation. The datasets used during training influence how well a model performs on specific RNA families and structural types. Some models are trained only on canonical (non-pseudoknotted) structures from RNAStralign and ArchiveII, while others incorporate the pseudoknot-rich data from bpRNA-1m to enhance their capabilities. In contrast, the pretraining of UNI-RNA uses very large collections, combining multiple sequence databases from RNAcentral and NCBI, in order to extend the model's applicability.

Fig. 3 demonstrates that Rfam (multiple versions) together with bpRNA/bpRNA-1m is used most frequently by models, while new models primarily train using enormous unannotated sequence databases for pretraining.

Table 1: Datasets of the RNA secondary structure

Dataset	Description	Usage in models
Rfam (Griffiths-Jones et al., 2003; Kalvari et al., 2021)	A repository of non-coding RNA families with annotated consensus secondary structures. Versions (e.g., 14.4, 14.5, 14.9) provide incremental improvements in sequence diversity and annotation.	Training for canonical structures
bpRNA (Danaee et al., 2018)	Contains over 102,318 sequences derived from experimental RNA structures, covering diverse RNA families.	Training for general RNA families and benchmarking
bpRNA-1m (Danaee et al., 2018)	An extended variant of bpRNA with over one million sequences, offering enhanced structural diversity and coverage.	Training for pseudoknot prediction
PDB (Lu et al., 2015)	High-resolution, experimentally determined 3D RNA structures; crucial for refining secondary structure predictions.	Refining secondary structure predictions
RNAStralign (Tan et al., 2017)	A curated collection of 37,149 RNA structures with sequence alignments and secondary structure annotations. Sequence lengths range from 30 to 1851 nucleotides.	Benchmarking models
ArchiveII (Sloma and Mathews, 2016)	Includes 3,975 sequences from 10 conserved RNA families. Frequently used for benchmarking canonical structures.	Validation against experimental benchmarks
RNA STRAND (Andronescu et al., 2008)	A small-scale, manually curated database with high-quality RNA structure annotations, valuable for validation tasks.	Large-scale pretraining

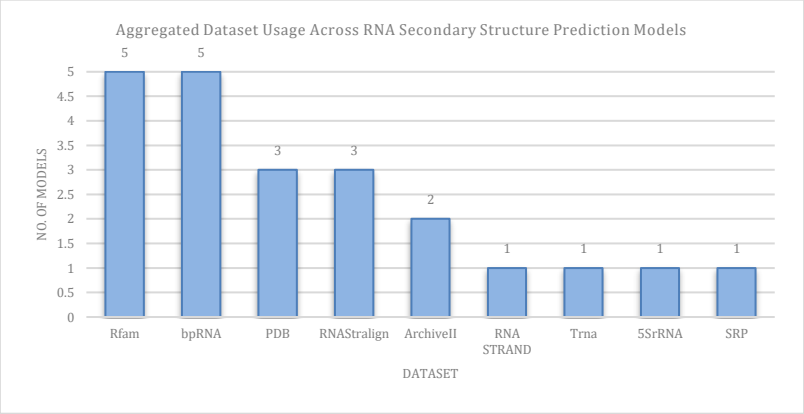

Fig. 3: Aggregated dataset usage across RNA secondary structure prediction modes

The performance evaluation of models primarily relies on the Precision and Recall measures together with their harmonic mean F1 score. The F1 score proves valuable because it combines precision with recall statistics, where precision represents the correct predictions among all predicted base pairs, while recall demonstrates the percentage of true base pairs that the model correctly identifies. The F1 score acts as a standard metric for comparing models since it utilizes per-based-pair calculations to provide an overall prediction accuracy summary. Fig. 4 presents a bar chart graphical representation of the F1 scores obtained by each of the models. All bars are graphically differentiated to be easily distinguished from one another. This chart was obtained by plotting the actual extracted performance metrics without making approximations in the process, and hence, the results are as accurate as possible during comparisons.

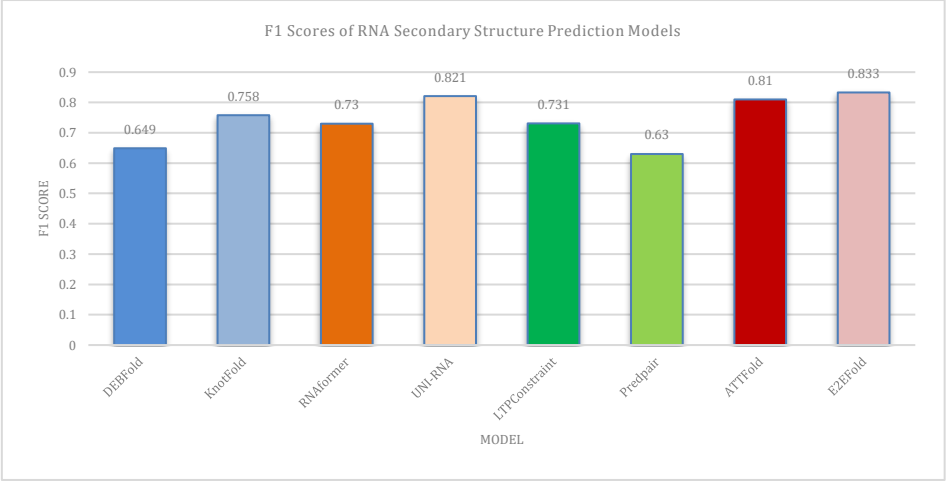

Fig. 4: F1 scores of RNA secondary structure prediction models

Table 2 shows the results of F1 between the models. Early models such as ATTfold reached a comparatively average F1 score of 0.810 on benchmark datasets, which suggested that there is immense potential for attention mechanisms, due to which there were restrictions on datasets as well as

complex architecture. Using only canonical structures, PredPair, taking advantage of CNNs, biLSTMs, and attention mechanisms, proved to provide high accuracy while being reasonably successful in pseudoknot recognition as well. Other models like E2EFold followed the trend of reaching the given benchmark of F1 equal to 0.833 while using multiple-head attention and constrained optimization for the prediction of both canonical and pseudoknotted structures. Likewise, UNI-RNA reported an F1 score of 0.821, which is ensured by large pretraining datasets where the large scale of the data led to more accurate prediction. The most recent DEBFold model proposed accurately evaluated correlations in F1 scores and achieved near canonical structure assessment with the balance of F1 scores and accurate attention integrated with thermodynamics.

Table 2: Comparison of RNA secondary structure prediction models based on attention mechanism results

Reference	Dataset	Model	Results
Yang (2024)	Rfam 14.4; bpRNA; Protein Data Bank (PDB)	DEBFold	F1 score: 0.649 F1 score: 0.758
Gong et al. (2024)	Rfam 14.5; bpRNA-1m	KnotFold	Precision: 0.784 Recall: 0.734
Franke et al. (2024)	Rfam 14.9; bpRNA; ArchiveII; RNA STRAND; PDB	RNAformer	F1 score: 0.730 F1 score: 0.821
Wang et al. (2023)	bpRNA; RNAStralign	UNI-RNA	Precision: 0.894 Recall: 0.801 F1 score: 0.731
Fei et al. (2022)	Rfam 14.5; bpRNA; tRNA; 5SrRNA; SRP; PDB	LTPConstraint	Precision: 0.747 Recall: 0.715
Grigorashvili et al. (2022)	Rfam	PredPair	Accuracy: 0.630
Wang et al. (2020)	RNAStralign	ATTfold	F1 score: 0.810 F1 score: 0.833
Chen et al. (2020a)	RNAStralign; ArchiveII	E2EFold	Precision: 0.880 Recall: 0.798

The subsequent parts of this section detail each significant attention-based model. We present an analysis of model architecture featuring attention mechanisms, which includes testing contexts with specified datasets along with performance metrics as F1 score assessments, and special strengths or weaknesses regarding the handling of pseudoknots and computational resource utilization.

2.6.1. E2EFold

E2EFold is a significant advancement that deals with limitations in accuracy, pseudoknot structures, and computational complexity of RNA secondary structure. This end-to-end deep learning model, which incorporates the transformer-based Deep Score Network and the constrained optimization Post-Processing Network, enables the prediction of RNA base-pair matrices directly while constraining severe complications such as non-overlapping pairs, min loop lengths, and nucleotide compatibility. The Deep Score Network utilizes transformer encoders to obtain sequence dependencies as well as global context through positional embeddings and utilizes 2D convolutions to compute all base-pair scores. The Post-Processing Network utilizes an unrolling of several primal-dual optimization steps, which are entirely differentiable; therefore, so is the end-to-end training and guaranteeing structural validity for the secondary structures of the RNA molecules, including the pseudoknots. The model is tested on the RNAStralign data set of 30451 (Tan et al., 2017), sequence length up to 1851, and evaluated on the ArchiveII data set without retraining of model of 3975, length up to 2968 (Sloma and Mathews, 2016). The training follows a two-step process: The Deep Score Network is trained with weighted binary cross-entropy for handling the data imbalance issue and, after that, fine-tuned in collaboration with the Post-Processing Network on a differentiable F1 loss function, which considers the specific error rates for optimization. The evaluation shows outstanding performance in terms of predictive accuracy with F1 scores of 0.821 and 0.686 on RNAStralign (Tan et al., 2017) and ArchiveII (Sloma and Mathews, 2016), respectively, outcompeting all state-of-the-art algorithms like CONTRAfold (Do et al., 2006) and LinearFold (Huang et al., 2019). This is mostly notable in the performance of the pseudoknot, where it scores an F1 of 0.71 compared to RNAstructure (Bellaousov et al., 2013), scoring an F1 of 0.472. Scalability to long sequences with an almost equivalent runtime with LinearFold (Huang et al., 2019) and the incorporation of hard constraints to prevent over-fitting and produce biologically realistic folds. The applicability of the model has been demonstrated in synthetic biology, drug discovery, and functional genomics; however, the current method has drawbacks that keep it from generalizing for other RNA families, particularly those underrepresented in the database, and is constrained by scalability issues in terms of sequence length beyond 3,000 bases. This coherent design and training framework emphasizes the capacity of E2Efold to establish a new gold standard for RNA structural bioinformatics.

2.6.2. ATTFold

ATTFold successfully overcomes some of the long-standing limitations of RNA secondary structure modeling, such as dealing with long sequences and predicting pseudoknots. Constructing upon an Encoder-Decoder-Constrain structure, ATTFold adopts a transformer-learning-encoder combined with an attention mechanism to generate

relatively high-dimensional feature vectors of RNA sequences, where the positional embeddings are also applied to preserve RNA sequence order. This is followed by a convolutional neural network (CNN) decoder, which produces a symmetric base-pairing scoring matrix to predict the likelihood of each base pair being paired. This unconstrained matrix is then refined in the constraint module, where certain biological constraints such as canonical base pairing (A-U, G-C, G-U pairings), minimum loops of a hairpin, and single pair restrictions are implemented, applying the hard constraint algorithm with Lagrange Multiplier optimization. Training on the RNAStralign (Tan et al., 2017) dataset of 25,425 exact RNA sequences obtained by removing stem-loops and filtering out sequences with lengths exceeding 512 bases, 20,340 of these sequences were used for training, 2,543 for validation, and 2,542 for testing, with redundant sequences culled to prevent leakage of information. The model optimization followed a fitting of the loss function involved computational complexity through gradient descent and backpropagation to optimize a composite loss function that combines fitting accuracy and structural constraints. On short sequences like tRNA and 5S_rRNA, ATTFold achieved an F1 score more than 0.9, which was 20% better than traditional methods, including RNAfold (Zuker and Stiegler, 1981) and Probknot (Bellaousov and Mathews, 2010). However, on longer, more complicated folds containing pseudoknots like telomerase RNA, ATTFold still demonstrated considerable gains, and F1 scores over 0.8. These results affirm the high scalability and versatility of ATTFold for different RNA families. However, there are issues that have not been adequately solved in predicting highly complex long sequences, mainly due to the scarcity of datasets and the complexity of RNA secondary structures. Future work seeking to augment long-sequence datasets, as well as incorporate experimental constraints, could also strengthen the model, positioning ATTFold as a foundational technology for RNA structural bioinformatics for use in biological research, synthetic biology, and therapeutic RNA design.

2.6.3. PredPair

The PredPair model contains a deep learning architecture that utilizes CNNs, biLSTMs, and attention mechanisms to predict base-pairing probabilities that originate from sequence data in the context of RNA secondary structure prediction. In contrast to the thermodynamic or homology-based approaches, PredPair is free from such pre-learned information as stacking energies or spatial layouts and factors, receiving as the sole input one-hot encoded RNA sequence with a unique mark on a particular nucleotide. The architecture then applies a convolutional layer to extract the local features, an attention layer that guides the model to the nucleotide complementarity, and biLSTMs to capture long-distance features, leading to a certainty matrix of pairing propensity. Training used 2,147 Rfam (Kalvari et al., 2018) families without sequences below 20/ 20 paired nucleotides and split into training, validation, and test sets in the ratio 60/20/20, respectively. Each RNA sequence was encoded into input-output pairs where the target represented specific base-pair interactions and the source text the sequence, producing 857,307 training samples. For optimization, categorical cross-entropy was used as a loss function, and the Adam optimizer had a learning rate of 0.1 and L2 regularization to prevent overfitting the model. Using evaluations of performance, it was indicated that there was a top-1 accuracy of 0.58 and a top-2 accuracy of 0.7. A rise was observed to 0.63 when making sure that the certainty matrix was symmetrical. Compared with benchmarks, PredPair obtained preferable F1 scores; however, slightly higher false positive ratios than RNAplfold (Lorenz et al., 2011) and SPOT-RNA (Singh et al., 2019). Notably, it was capable of predicting pseudoknots with reasonable accuracy at 0.78, and this is a problematic structure type to predict as it was not heavily represented in the training sets. Additional gradient-based saliency map analysis for three representative recognition cases confirmed that the model learned semi-quantitative representations of biologically relevant biochemical features, Watson-Crick base pairing and wobble pairings, and stacking energy hierarchies. Applying t-SNE to adjust clustering underlined the ability of the constructed network to arrange sequences according to their Rfam families, even though their similarity, rather than Rfam structural motifs, could prevail in those clusters. The usefulness of the model was confirmed again through an analysis of the DMS-seq data from E. coli (Burkhardt et al., 2017); predicted co-occupied nucleotides corresponded to experimental inaccessible sites.

2.6.4. LTPConstraint

The LTPConstraint model is used here as it is a novel state-of-the-art RNA secondary structure prediction model that utilizes deep learning methods like Bi-LSTM, Transformer, Transfer Learning, and biologically motivated constraints. By training the model on datasets drawn from Rfam v14.5 (Kalvari et al., 2021), bpRNA (Danaee et al., 2018), and RNAStralign (Tan et al., 2017), diverse data preprocessing methods, including redundancy deletion and length division of the input sequences, are incorporated to have high-quality data fed into the model. The architecture consists of three interconnected modules: A global semantic extraction module based on Bi-LSTM and Transformer layers, which capture sequential and positional information; a local feature extraction module that calculates base-pair scores using a U-net-inspired generator; and a hard constraint layer that translates RNA pairing rules. The Bi-LSTM used in this study is a type of bidirectional RNN that is able to process sequences forwards and backwards,

making sure that essential long-range dependencies and contextual relationships in RNA sequences are not ignored. This mechanism is crucial in unravelling the intricacies of folding patterns that are typical of RNA secondary structures.

The training TSA is 80-10-10 for training, validation, and testing with optimizers such as Adam and customized losses such as weighted logistic loss and F1 loss for better predictions, especially where the datasets are highly imbalanced. Transfer learning improves the model even more by training the model on big, general data sets (for example, Rfam) and then updating the model based on a small data set of chosen RNA families. It helps to avoid the need for large volumes of data that are usually specific to one family and, at the same time, increases the rate of convergence and accuracy. The model performs nicely in predicting pseudoknot structures because it achieves higher F1 scores than benchmark models such as ProbKnot (Bellaousov and Mathews, 2010) and Knotty (Jabbari et al., 2018); specifically, 0.9034 for 128-length sequences and 0.7310 for 512-length sequences. Quantitatively, the use of the adjacency matrix may add computational steps and the consideration of multiple neural components. At the same time, qualitatively, for both benchmark and chimeric RNA structures, LTPConstraint offers sound and precise predictions.

2.6.5. UNI-RNA

The Uni-RNA model is a model for RNA secondary structure prediction that uses a pre-trained architecture based on transformers and optimised for different RNA datasets. Uni-RNA is based on one of the most used NLP models in the last years, the BERT model (Bidirectional Encoder Representations from Transformers), which pretrains on masked tokens and learn contextual relations within the sequences. In RNA tasks, Uni-RNA generalizes this concept to nucleic acid sequences and identifies previously concealed structural and functional features with masked nucleic acid modeling. Integration of ~1 billion RNA sequences from RNAcentral (Sweeney et al., 2021), the National Center for Biotechnology Information (NCBI) (Sayers et al., 2021), and other databases of Uni-RNA, the training updates included MMseqs2 clustering for sequence clustering and readaptation of RNA sequences mapped to a DNA character set. The architecture varies from 8 to 24 transformer layers and applies rotary embeddings for positional encoding, using fast attention for memory-efficient computation and fused layer normalization for better training performance.

The best models of comparison used in benchmarking were RNA-FM (Chen et al., 2022) and Uni-RNA, using totally different approaches like UFold (Fu et al., 2022), CONTRAfold (Do et al., 2006), and SpliceBERT (Chen et al., 2024). Since RNA-FM aims to perform RNA tasks using language understanding, we utilize a pre-trained language model for RNA-FM to extract sequence embeddings for RNA tasks. UFold utilizes a deep learning framework to predict the secondary structures of RNA using an unrolled algorithm to enhance precise determination. CONTRAfold, an earlier work that involves the use of machine learning, forecasts secondary structures without adopting thermodynamic models, bearing a difference to physics-based models. To enhance the prediction of RNA secondary structures, SpliceBERT is another transformer-based model that applies BERT for DNA sequences to recognize splice sites, indicating its flexibility across different species.

Uni-RNA did better than these models, with 0.894 of Precision, 0.801 recall, and an F1 score of 0.821, and made RNA-FM's F1 score benchmark on bpRNA-1m (Danaee et al., 2018) and RNAStralign (Tan et al., 2017) to 0.694. In the long-range contact map, Uni-RNA attained a maximum L/5 precision of 0.709, which is significantly higher than one-hot encoding, with 0.647 superior in modelling intricate structural dependencies. In splice site prediction, Uni-RNA performs the best with an F1 score of 0.9635, while SpliceBERT shows 0.9568, which proves good cross-sectional compatibility. The training was done on 128 A100 GPUs, engaging the Adam optimization method coupled with the dropout of 0.1, gradient clipping, and the learning rate of 0.004 to maintain convergence stability. Uni-RNA at present, utilizing architectural novelty, vast data sets, and high performance, emerges as a front-rank tool in computational RNA biology; this innovative ally equips techniques in RNA structure prediction, functional annotation, and therapeutics use with unparalleled Precision and capacity.

2.6.6. RNAformer

RNAformer is a leap forward in the area of RNA secondary structure prediction based on a new architecture that combines axial self-attention and latent space recycling, thus allowing for the modelling of long-range interdependencies and multiple passes over the sequence to refine base-pair predictions. Unlike most of the current approaches, RNAformer does not require side information such as multiple sequence alignment (MSA); instead, it uses a two-dimensional embedding of RNA sequences, and this rationality makes it more interpretable and scalable. Training was based on representative datasets from Rfam (v14.9) (Kalvari et al., 2021), together with RNAStralign (Tan et al., 2017), bpRNA (Danaee et al., 2018), ArchiveII (Sloma and Mathews, 2016), and RNA STRAND (Andronescu et al., 2008), as well as on secondary structures of PDB entries supported by DSSR (Lu et al., 2015). Outlier removal processes like BLAST for homology filtering and covariance models helped in optimizing the mentioned dataset for minimizing the bias in model assessment. The model also proved rather portable in terms of training on finite-defined sets of canonical and complex shapes, pseudoknots in

particular, TS0 for intraspecific/subsets and TS-hard for interaspecific comparisons.

The Binary Cross Entropy loss with masking is utilized to address the problem of both paired and unpaired base entries, and the model is further trained, as a result of this RNAformer, with Dropout layers of 0.4 and the AdamW optimizer and Cosine Annealing, unique to learning rate scheduling of 0.001. This is achieved through the axial-attention-based transformer blocks of which the present model has six, as well as latent space recycling, which is borrowed from AlphaFold (Jumper et al., 2021). This mechanism of refinement acted during fine-tuning and was computationally lighter than recycling during the pretraining phase. Similar to previous experiments, evaluation based on F1 and F1-shifted scores also showed that RNAformer surpassed other de novo RNA structure prediction methods like RNA-FM and UFold (Fu et al., 2022) in terms of both scores, as it was able to achieve an F1 score of 0.73 on the TS0 dataset and 0.743 on the TS1, while the arrangement also indicates the comparative performance with homology-based methods like SPOT-RNA2 (Singh et al., 2021). In addition, inference speed is relatively fast: Less than one second for one sequence, and memory-efficient FlashAttention makes it fit for use in a large-scale context. Nevertheless, there remain challenges to attaining scalability with the use of RNAformer, especially when handling longer sequences, as well as making more accurate predictive estimations with respect to highly intricate characteristics.

2.6.7. KnotFold

KnotFold is a state-of-the-art RNA secondary structure prediction model capable of flagging canonical base pairs and recognizing complex pseudoknots using a transformer-based architecture and self-attention. The model predicts an L x L base-pair probability matrix, leveraging eight transformer encoder layers (embedding dimension: 256, eight attention heads per layer, and 6.5 million parameters for each model to learn, capturing long-range interaction dependencies, and estimating the corresponding probabilities P(b(i, j) | x). As trained and validated on the bpRNA-1m dataset (102,318 identified RNA structures, reduced to 23,819 for training and 1,131 for validation) (Danaee et al., 2018), KnotFold incorporates these probabilities into a potential function that employs the minimum-cost flow algorithm in optimizing the secondary structures predicated on the iterative addition and removal of base pairs. Its workflow was given on RNA CP000097.1_937913-937973 (Kalvari et al., 2021), which predicted that there would be 18 base pairs. However, there was one false positive, and after 36 optimization steps, the cost was finally tuned to -351.6. KnotFold's performance was benchmarked against four state-of-the-art RNA prediction tools. The four RNA folding tools that have been compared include MXfold2 (Sato et al., 2021), SPOT-RNA (Singh et al., 2019), UFold (Fu et al., 2022), and Jabbari et al. (2018). MXfold2 uses thermodynamic principles along with deep learning to predict structures with proclivity toward energy minimization. Several analyses were performed on the constructed dataset, PKTest, which comprises 1009 pseudoknotted RNAs from BpRNA (Danaee et al., 2018) and Rfam (Kalvari et al., 2021). In PKTest, KnotFold obtained the F1 score of 0.758, which was higher than MXfold2 (0.654), SPOT-RNA (0.579), and UFold (0.602). KnotFold employs AdamW optimization with (learning rate: 0.001, warm-up distance: 30000 steps, weight decay: 0.01, batch size: 4, policy for overfitting: Dropout and ensemble. Transformer-based attention mechanisms, as well as iterative flow optimization, make KnotFold suited for dealing with complex pseudoknots and for generalization to new RNA families, outperforming other existing methods and confirming its practical applications in synthetic biology, RNA function analysis, as well as in drug discovery.

2.6.8. DEBFold

DEBFold also marks an improvement over traditional and deep-learning relatives by effectively making use of a two-step approach to RNA secondary structure prediction. The first stage also utilizes a deep convolutional network with a self-attention mechanism to predict base-pairing probabilities from RNA sequences represented as 6×l×10 tensors in which l is the sequence length, capturing the results of six conventional thermodynamic predictors, such as RNAfold and IPknot. Taking advantage of a large-scale family-wise strategy on the (bpRNA-1m), including 43,269 sequences belonging to 2,125 structural families, DEBFold alleviates overfitting arising from deep-learning models trained on sequence information. The dataset was carefully split based on 25-fold cross-validation; all family members were independent and diverse; further testing sets (TestSetβ and TestSetγ) were generated from bpRNA-new and PDB, respectively, for assessment. The second stage converts these pairing probabilities into SHAPE-like scores to guide an optimization process that incorporates thermodynamic principles. In the proposed model, training was conducted for over 250 epochs using the Adam optimizer with an enhanced cosine-decay learning rate scheduler. Dropout regularization and batch normalization were also employed to improve generalization.

The performance benchmarks also put DEBFold higher than classical and deep-learning-based methods every time. Compared with other classical thermodynamic tools like RNAfold, IPknot, and TurboFold, DEBFold got higher median F1 scores (e.g., 0.649 for TestSetα) and depicted better overall cross-family generalization. DEBFold also outperforms these deep-learning approaches and achieves an F1 score of 0.557 on fully curated datasets like TestSetβ and 0.779 on the TestSetγ datasets. On a more primary note, DEBFold exceeded

a comparable performance to SPOT-RNA2 in the context of experimentally validated PDB sequences. At the same time, the time to make computations was significantly shorter and equal to 67 seconds, on average, not rising to 142 hours, 16% of the time for specific sequences explored in the subset of SPOT-RNA2. The use of variable sources thermal modelling integrated with deep learning forms the core of DEBFold, providing it with best-in-class accuracy and scalability. DEBFold proves to be computationally efficient but shows the best performance in tasks involving only canonical structures of RNA and sequences up to 512 nucleotides long. This unique combination of deep learning and thermodynamics, which can obtain state-of-the-art performance, also shows the potential for further application in RNA structural biology and functional annotation, even if there are still several difficulties, including problems with pseudoknots and long noncoding RNA prediction. This study shows that by improving the issues cited in previous approaches systematically and achieving better performance than benchmarks in different data sets, DEBFold not only offers new ways and thoughts for RNA secondary structure prediction but also has a substantial application value.

In addition to the explanations provided above, Fig. 5 presents a collective graphical summary of the eight attention-based models for RNA secondary structure prediction. Fig. 5 delineates the fundamental computational components, i.e., input encoders, attention mechanisms, intermediate neural layers, constraint modules, and output stages, that each respective model employs.

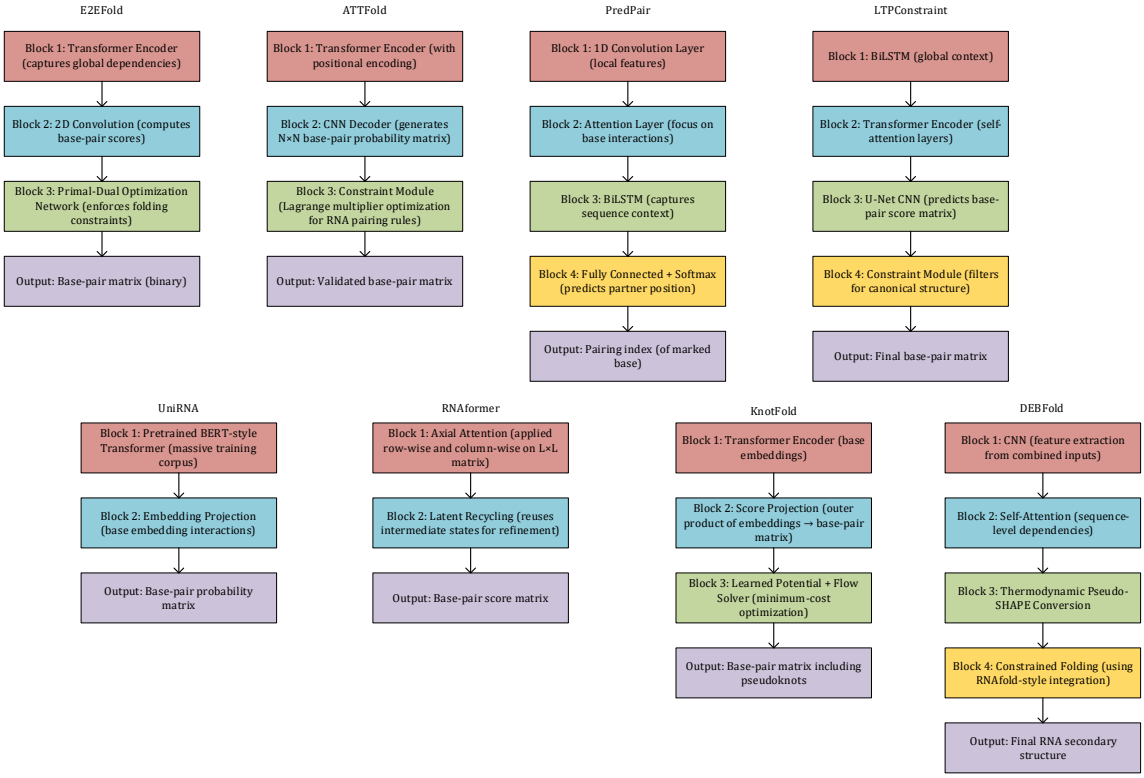

Fig. 5: Architecture diagrams of the attention-based RNA secondary structure prediction models

2.7. Comprehensive technical summary of the RNA secondary structure prediction models

Table 3 compares the models across several categories: pseudoknot support, maximum sequence length, computational cost, and main innovation. Each model involves trade-offs. Some are designed for faster performance on CPUs, while others can process very long sequences or more complex structures, though with higher computational demands. Table 3 shows which model is most suitable for different tasks, such as rapid folding or managing pseudoknots.

3. Discussion

The comparative flowchart in Fig. 6 illustrates the workflows of RNA secondary structure prediction approaches, including MFE-based algorithms, machine learning, deep learning, and attention mechanisms, highlighting their unique processes, iterative decision-making steps, and output visualizations.

The innovation of recent GPU hardware and large data technologies has enabled transformer-based models to be trained over millions of parameters, enabling complicated biological difficulties like huge pseudoknots and long-distance base pairings that were out of reach for conventional algorithms. The initial efforts, such as the ViennaRNA Package, served as a blueprint for the combination of thermodynamics and covariation through multidisciplinary cooperation that established the groundwork for integrating domain expertise and computational advancement.

Table 3: Comprehensive technical summary of the RNA secondary structure prediction models

Model (year)	Pseudoknot support	Long RNA support	Computational cost	Key strength/Innovation
DEBFold (2024)	No	Partial. Validated up to 512 nt; longer sequences out of scope.	Low. Extremely efficient ensemble model; 7500× faster than deep competitors.	Two-stage ensemble with family-wise generalization; fast, robust canonical structure prediction.
KnotFold (2024)	Yes	Yes. Efficient for ~1000 nt; cubic time scaling may limit >1kb inputs.	Moderate–High. Combines $O(n^2)$ neural scoring with $O(n^3)$ flow optimization.	Neural scoring with exact flow-based pseudoknot prediction: breakthrough in efficient complex folding.
RNAformer (2023)	No	Yes. Supports kilobase sequences using axial attention; evaluated on the TS0 benchmark.	Moderate–High. Transformer layers (8–24); axial attention improves efficiency.	Axial-attention (scans rows/columns separately) Transformer with latent recycling; no reliance on alignment or external data.
UNI-RNA (2023)	Yes	Yes. Trained on billion-scale data; supports long sequences with FlashAttention.	High. Pretraining on billions of sequences; 400M parameters; GPU-intensive.	Massive foundation model for RNA; learned from 1B+ sequences; transfer-learns RNA structure and function.
LTPConstraint (2022)	No	Partial. Optimized for moderate-length RNAs (e.g., <1000 nt); performance not evaluated on very long RNAs.	Low. Lightweight design using BiLSTM, Transformer, and CNN modules; fast inference.	Combines constraint learning with a lightweight architecture for efficient and accurate canonical structure prediction.
PredPair (2022)	Yes	Partial. Tested up to ~800 nt; performance declines with length; not optimized for >1kb.	Low–Moderate. Lightweight CNN+LSTM+attention; inference feasible on CPUs.	Learns base-pairing rules de novo; interpretable architecture that discovers thermodynamic features.
ATTfold (2020)	Yes	Yes. Targets long sequences up to ~rRNA length; accuracy declines >1000 nt.	Moderate. Transformer-like network; fast GPU inference; polynomial complexity.	Early attention-based model integrating structural constraints; handles long-range pairing and pseudoknots.

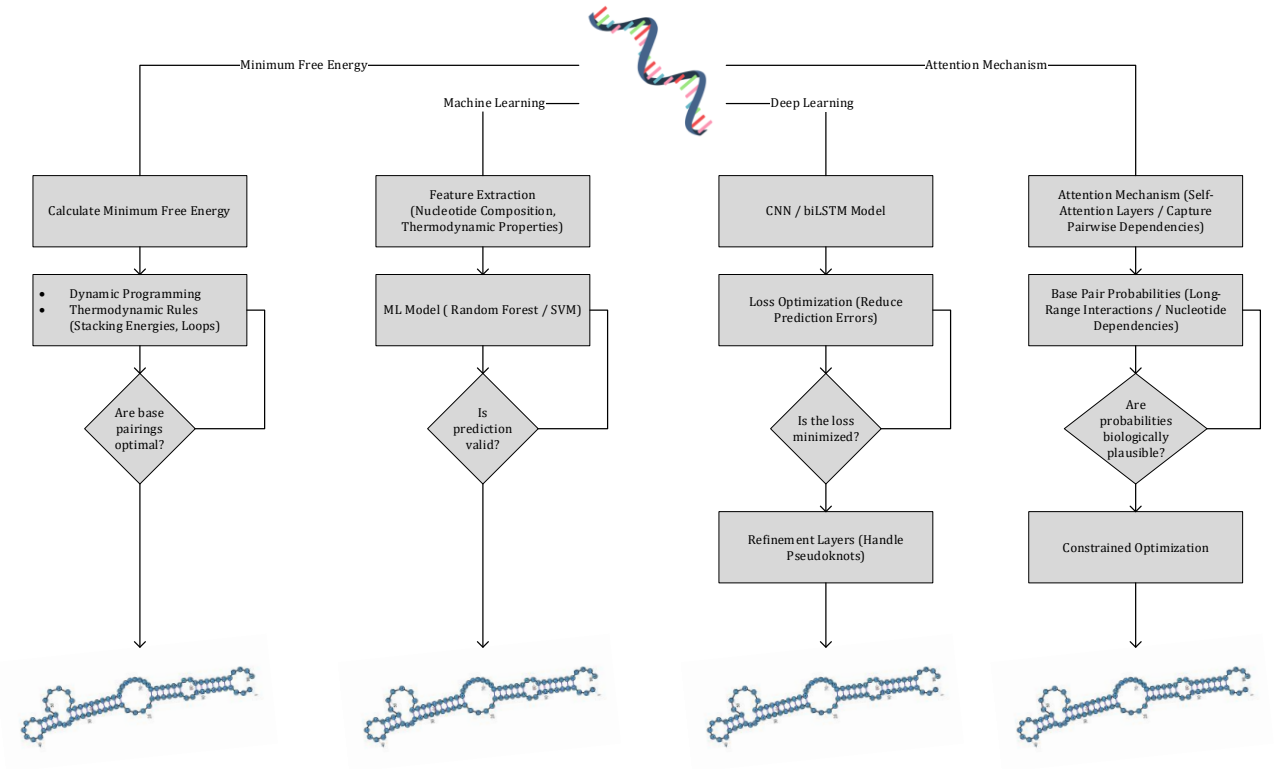

Fig. 6: The workflows of RNA secondary structure prediction approaches

Attention-based deep learning approaches have outperformed previous approaches by using data-driven methods in detecting local as well as distal dependencies within RNA sequences. By using weight sharing across the sequence and global self-attention mechanisms, transformers are able to detect complex patterns of base pairing, such as distal interactions and pseudoknots, that were challenging for simpler machine learning and thermodynamic models to represent. Additionally, parallel calculation of entire sequences provides computational benefit over sequential RNN-based methods by enabling quicker training and inference on big RNA datasets. Pretrained transformer models like UNI-RNA can generalize RNA knowledge into specialized tasks with fine-tuning, which further increases their versatility.

Despite these strengths, attention-based models also come with significant challenges and limitations that must be critically analyzed. The advanced efficacy of E2EFold and RNAformer depends on significant computational expenses, together with multiple scalability needs. Using GPUs and lots of training time, together with tens of thousands of sequences and many attention layers, makes this process difficult for small research groups that lack

extensive resources. Long RNAs exceeding thousands of nucleotides continue to present challenges for predicting their structures when deployed for use. The extended axial attention adaptations in RNAformer create difficulties for processing RNA sequences that surpass multiple thousands of lengths. The attention schemes implemented by E2EFold and RNAformer (multi-head and axial, respectively) require extensive resources to operate effectively. These efficiency limitations demonstrate that accuracy-boundary breakthroughs in these models need complementary technology to create either new minimalistic models or adaptation approaches that enhance memory efficiency for practical applications. Another limitation regarding biases and generalizability. Deep learning algorithms obtain their performance level from the quality of available data. Most attention-based predictors receive training from the same sets of data, including RNAStralign, bpRNA, and Rfam families. The models experience difficulties with RNA families and structural motifs that occur infrequently during training. A model that primarily encounters short pseudoknots during training might struggle to find long-range pseudoknots even if the input data contains them. The same holds true for E2EFold, which lacks the ability to predict previously unseen motifs because they are outside its scope. The elimination of sequence stretches longer than 1000 nucleotides during preprocessing leads to models receiving limited exposure to distant interaction patterns during training. The knowledge base of reference structures suffers from two major issues: first, they derive from predictive algorithms that could potentially embed system errors, and second, the reference experiments generate structures biased toward MFE assumptions. Model generalizability receives negative consequences from this procedure. The ability of UNI-RNA to process extensive unannotated data sequences suggests a strategy that involves training on diverse sequence sets to minimize bias effect, yet fails to ensure a complete solution. Extra care is needed when analyzing high test set performance stemming from data similarity to training sets because real generality needs to be demonstrated through new RNA performance assessment (multiple DEBFold tests employed this approach).

Different models showcase unique trade-offs between complexity, scalability, and accuracy:

- E2EFold achieves state-of-the-art performance on pseudoknots but incurs massive computational overhead and struggles beyond 3000 nucleotides.
- RNAformer incorporates axial attention to improve scalability, yet still faces challenges with extremely long RNAs and highly complex pseudoknotted structures.
- DEBFold balances attention mechanisms with thermodynamic priors, offering reliable and efficient performance on canonical structures without maximizing peak accuracy.
- KnotFold specializes in pseudoknot prediction by combining attention mechanisms with iterative optimization, excelling in this specific task.
- PredPair and LTPConstraint offer lightweight alternatives suitable for standard secondary structure prediction, though their flexibility and accuracy are lower for highly complex RNAs.
- UNI-RNA, via extensive pretraining, achieves broad generalization but requires substantial computational investment and still demands careful task-specific fine-tuning.

These comparisons highlight that no single model is universally optimal; model choice depends on specific research needs, available computational resources, and target RNA complexity.

Additionally, the interpretability and validation capabilities of these models remain limited because attention weights offer restricted diagnostic features, although they occasionally reveal which nucleotides the model pairs based on inspection of attention matrices. Researchers must have trust in predictive models but require an understanding of their operations, particularly when designing future experiments. The PredPair saliency analysis demonstrates progress by showing that it identified known pairing rules. Making sense of the underlying reasoning behind model-predicted structures remains a complex task for scientific interpretation. The validation process depends on experimental testing through SHAPE and DMS chemical probing and comparative sequence analysis. The prediction models currently do not supply uncertainty measurements other than basic pairing probability statistics. Researchers actively pursue methods to interpret complex models while establishing confidence measures since model complexity continues to increase.

RNA secondary structure modeling still poses one of the most difficult tasks, pseudoknot prediction. Although domain-specific strategies are in models such as KnotFold and E2EFold, other models, such as UNI RNA and PredPair, focus more on canonical structures for computational simplicity. Since accuracy in this domain is sensitive to well-behaved specializations, flexibility in other domains is traded off, but specialized architectures for pseudoknots yield better accuracy. It therefore offers validation to pseudoknot prediction with the important reminder that task-specific approaches will continue to be needed, and the range of model scope will be well assessed.

Attention-based RNA models utilize a tremendous amount of advanced training techniques to stabilize learning and prevent overfitting. Adam and variant AdamW are the optimizers the most used, sometimes along with a learning rate schedule, cosine annealing (as in RNAformer too), to keep training momentum. To also make the model more robust against noisy data, regularization strategies like dropout and batch normalization are used. Choosing these optimizations fits well with the

complexity of the architecture and diversity of the dataset encountered during RNA structure modeling.

The circumstances observed demonstrate why performance benchmarks must include additional information. The evaluation of F1 scores by models requires a specific mention regarding which structural combinations provided those results. The F1 score of 0.8 seems to indicate successful structure prediction for canonical structures, although the same model demonstrates reduced capability when dealing with pseudoknots or longer sequences. The review demonstrates that models E2EFold and KnotFold excel at identifying pseudoknots while DEBFold and LTPConstraint focus mainly on canonical forms because of their emphasis on consistency and efficiency. The results showed that ATTFold achieved exceptional accuracy on short RNAs containing simple motifs to the extent of outperforming physical models by a 20 percent margin. For valid comparisons, it is necessary to examine how many pseudoknots exist in the benchmark set and what length of sequences were included.

4. Conclusions

Recent developments have enhanced the field of RNA structure prediction comprehensively because of computational technology, which includes attention mechanisms and transformation systems. Such approaches have raised the bar for precision, allowing for the representation of structures like pseudoknots and long-range contacts. However, some of the major problems still need to be solved today: the requirements for computational resources of current methods and the need for sufficiently diverse and high-quality datasets. Overcoming these limitations is of paramount importance for the enhancement of the availability and application of RNA prediction tools and the achievement of groundbreaking advancements in synthetic biology, drug discovery, and RNA-based therapeutic technologies.

On the horizon for future work, the solution requires multiple focused approaches to resolve these difficulties. Using sparse attention along with axial attention as a computational architecture reduces both hardware memory and running time requirements. These attention mechanisms minimize attention scope to nucleotide neighbors in addition to splitting base-pair matrix analysis between rows and columns. Model performance stability remains intact when adding low-rank approximation and pruning methods to optimize large model structures. Transfer learning enables researchers to overcome data scarcity by training models on massive unlabeled RNA datasets prior to fine-tuning them specifically for structure prediction. The training process enhances its ability to discover important biological patterns through the incorporation of experimental probing data, such as SHAPE or DMS reactivity profiles pertaining to underrepresented RNAs. A comprehensive pipeline of interdisciplinary work combines machine learning models to produce experimental testable predictions that generate empirical data, which cycles back into training data. Proper strategy implementation will generate RNA folding tools with high accuracy while providing large-scale capabilities to more scientists in the field.

List of abbreviations

AI	Artificial intelligence
ARES	Atomic rotationally equivariant scorer
BERT	Bidirectional encoder representations from transformers
BiLSTM	Bidirectional long short-term memory
bpRNA	A database containing RNA sequences and structures
CEASTech	Centre of excellence (advanced sensor technology)
CNN	Convolutional neural network
CONTRAfold	An RNA secondary structure prediction method
CPU	Central processing unit
DMS	Dimethyl sulfate
DMS-seq	Dimethyl sulfate sequencing
DNA	Deoxyribonucleic acid
DSSR	An algorithm for dissecting the spatial structure of RNA
E2EFold	An end-to-end deep learning model for RNA secondary structure prediction
FRGS	Fundamental research grant scheme
GPU	Graphics processing unit
IPknot	An RNA secondary structure prediction method including pseudoknots
LSTM	Long short-term memory
LTPConstraint	A transfer learning-based method for RNA secondary structure prediction
MFE	Minimum free energy
MMseqs2	A software suite for fast and deep clustering and searching of large protein and nucleotide sequence sets
MSA	Multiple sequence alignment
NCBI	National center for biotechnology information
NLP	Natural language processing
nt	Nucleotide(s)
PDB	Protein data bank
PredPair	A neural network model for predicting RNA secondary structure
ProbKnot	A method for fast prediction of RNA secondary structure including pseudoknots
Rfam	A database of RNA families
RNA	Ribonucleic acid
RNA-FM	An RNA foundation model
RNAcentral	A database of non-coding RNA sequences
RNAfold	A program for predicting minimum free energy secondary structures of RNA
RNAformer	A deep learning model for RNA secondary structure prediction
RNAplfold	A program for computing local base pairing probabilities in RNA
RNN	Recurrent neural network
RNAstructure	A software package for RNA secondary structure prediction and analysis
RNA STRAND	The RNA secondary structure and statistical analysis database
SHAPE	Selective 2'-hydroxyl acylation analyzed

	by primer extension
SLR	Systematic literature review
SPOT-RNA	A method for RNA secondary structure prediction using an ensemble of neural networks
tRNA	Transfer RNA
tRNAscan-SE	A program for detecting transfer RNA genes in genomic sequences
TSNE	T-distributed stochastic neighbor embedding
UFold	A fast and accurate RNA secondary structure prediction method with deep learning
UNI-RNA	Universal pre-trained models for RNA research

Acknowledgment

The authors would like to thank the Ministry of Higher Education for providing financial support under the Fundamental Research Grant Scheme (FRGS) (FRGS/1/2021/TKO/UNIMAP/02/65).

Compliance with ethical standards

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Andronescu M, Bereg V, Hoos HH, and Condon A (2008). RNA STRAND: The RNA secondary structure and statistical analysis database. BMC Bioinformatics, 9: 340. https://doi.org/10.1186/1471-2105-9-340 **PMid:18700982 PMCid:PMC2536673**

Balcerowicz M, Di Antonio M, and Chung BY (2021). Monitoring real-time temperature dynamics of a short RNA hairpin using Förster resonance energy transfer and circular dichroism. Bio-Protocol, 11(6): e3950. https://doi.org/10.21769/BioProtoc.3950 **PMid:33855112 PMCid:PMC8032495**

Bellaousov S and Mathews DH (2010). ProbKnot: Fast prediction of RNA secondary structure including pseudoknots. RNA, 16(10): 1870-1880. https://doi.org/10.1261/rna.2125310 **PMid:20699301 PMCid:PMC2941096**

Bellaousov S, Reuter JS, Seetin MG, and Mathews DH (2013). RNAstructure: Web servers for RNA secondary structure prediction and analysis. Nucleic Acids Research, 41(W1): W471-W474. https://doi.org/10.1093/nar/gkt290 **PMid:23620284 PMCid:PMC3692136**

Bliss N, Bindewald E, and Shapiro BA (2020). Predicting RNA shape scores with deep learning. RNA Biology, 17(9): 1324-1330. https://doi.org/10.1080/15476286.2020.1760534 **PMid:32476596 PMCid:PMC7549691**

Bravo JP, Bartnik K, Venditti L, Acker J, Gail EH, Colyer A, Davidovich C, Lamb DC, Tuma R, Calabrese AN, and Borodavka A (2021). Structural basis of rotavirus RNA chaperone displacement and RNA annealing. Proceedings of the National Academy of Sciences, 118(41): e2100198118. https://doi.org/10.1073/pnas.2100198118 **PMid:34615715 PMCid:PMC8521686**

Budnik M, Wawrzyniak J, Grala Ł, Kadziński M, and Szóstak N (2024). Deep dive into RNA: A systematic literature review on RNA structure prediction using machine learning methods. Artificial Intelligence Review, 57: 254. https://doi.org/10.1007/s10462-024-10910-3

Burkhardt DH, Rouskin S, Zhang Y, Li GW, Weissman JS, and Gross CA (2017). Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. eLife, 6: e22037. https://doi.org/10.7554/eLife.22037 **PMid:28139975 PMCid:PMC5318159**

Chai J, Zeng H, Li A, and Ngai EW (2021). Deep learning in computer vision: A critical review of emerging techniques and application scenarios. Machine Learning with Applications, 6: 100134. https://doi.org/10.1016/j.mlwa.2021.100134

Chen J, Hu Z, Sun S et al. (2022). Interpretable RNA foundation model from unannotated data for highly accurate RNA structure and function predictions. Arxiv Preprint Arxiv:2204.00300. https://doi.org/10.48550/arXiv.2204.00300

Chen K, Zhou Y, Ding M, Wang Y, Ren Z, and Yang Y (2024). Self-supervised learning on millions of primary RNA sequences from 72 vertebrates improves sequence-based RNA splicing prediction. Briefings in Bioinformatics, 25(3): bbae163. https://doi.org/10.1093/bib/bbae163 **PMCid:PMC11009468**

Chen X, Li Y, Umarov R, Gao X, and Song L (2020a). RNA secondary structure prediction by learning unrolled algorithms. Arxiv Preprint Arxiv:2002.05810. https://doi.org/10.48550/arXiv.2002.05810

Chen Z, Zhao P, Li F, Wang Y, Smith AI, Webb GI, Akutsu T, Baggag A, Bensmail H, and Song J (2020b). Comprehensive review and assessment of computational methods for predicting RNA post-transcriptional modification sites from RNA sequences. Briefings in Bioinformatics, 21(5): 1676-1696. https://doi.org/10.1093/bib/bbz112

Choi SR and Lee M (2023). Transformer architecture and attention mechanisms in genome data analysis: A comprehensive review. Biology, 12(7): 1033. https://doi.org/10.3390/biology12071033 **PMid:37508462**

Danaee P, Rouches M, Wiley M, Deng D, Huang L, and Hendrix D (2018). bpRNA: Large-scale automated annotation and analysis of RNA secondary structure. Nucleic Acids Research, 46(11): 5381-5394. https://doi.org/10.1093/nar/gky285 **PMid:29746666 PMCid:PMC6009582**

Do CB, Woods DA, and Batzoglou S (2006). CONTRAfold: RNA secondary structure prediction without physics-based models. Bioinformatics, 22(14): e90-e98. https://doi.org/10.1093/bioinformatics/btl246

Fei Y, Zhang H, Wang Y, Liu Z, and Liu Y (2022). LTPConstraint: A transfer learning based end-to-end method for RNA secondary structure prediction. BMC Bioinformatics, 23: 354. https://doi.org/10.1186/s12859-022-04847-z **PMid:35999499 PMCid:PMC9396797**

Franke JK, Runge F, Köksal R, Backofen R, and Hutter F (2024). RNAformer: A simple yet effective deep learning model for RNA secondary structure prediction. BioRxiv. https://doi.org/10.1101/2024.02.12.579881

Fu L, Cao Y, Wu J, Peng Q, Nie Q, and Xie X (2022). UFold: Fast and accurate RNA secondary structure prediction with deep learning. Nucleic Acids Research, 50(3): e14. https://doi.org/10.1093/nar/gkab1074 **PMid:34792173 PMCid:PMC8860580**

Fu X, Wang G, Wang C, Xu H, and Li H (2023). Multi-scale hybrid three-dimensional-two-dimensional-attention boosted convolutional neural network for hyperspectral image classification. Journal of Applied Remote Sensing, 17(2): 026513. https://doi.org/10.1117/1.JRS.17.026513

Gong T, Ju F, and Bu D (2024). Accurate prediction of RNA secondary structure including pseudoknots through solving minimum-cost flow with learned potentials. Communications Biology, 7: 297. https://doi.org/10.1038/s42003-024-05952-w **PMid:38461362 PMCid:PMC10924946**

Graf J and Kretz M (2020). From structure to function: Route to understanding lncRNA mechanism. BioEssays, 42(12): 2000027. https://doi.org/10.1002/bies.202000027 **PMid:33164244**

Griffiths-Jones S, Bateman A, Marshall M, Khanna A, and Eddy SR (2003). Rfam: An RNA family database. Nucleic Acids Research, 31(1): 439-441. https://doi.org/10.1093/nar/gkg006 **PMid:12520045 PMCid:PMC165453**

Grigorashvili EI, Chervontseva ZS, and Gelfand MS (2022). Predicting RNA secondary structure by a neural network: What features may be learned? PeerJ, 10: e14335. https://doi.org/10.7717/peerj.14335 **PMid:36530406 PMCid:PMC9756865**

Guo MH, Xu TX, Liu JJ, Liu ZN, Jiang PT, Mu TJ, Zhang SH, Martin RR, Cheng MM, and Hu SM (2022). Attention mechanisms in computer vision: A survey. Computational Visual Media, 8(3): 331-368. https://doi.org/10.1007/s41095-022-0271-y

He S, Gao B, Sabnis R, and Sun Q (2023). RNAdegformer: Accurate prediction of mRNA degradation at nucleotide resolution with deep learning. Briefings in Bioinformatics, 24(1): bbac581. https://doi.org/10.1093/bib/bbac581 **PMCid:PMC9851316**

Hernández A and Amigó JM (2021). Attention mechanisms and their applications to complex systems. Entropy, 23(3): 283. https://doi.org/10.3390/e23030283 **PMid:33652728 PMCid:PMC7996841**

Hofacker IL (2003). Vienna RNA secondary structure server. Nucleic Acids Research, 31(13): 3429-3431. https://doi.org/10.1093/nar/gkg599 **PMid:12824340 PMCid:PMC169005**

Hou Q and Jaffrey SR (2023). Synthetic biology tools to promote the folding and function of RNA aptamers in mammalian cells. RNA Biology, 20(1): 198-206. https://doi.org/10.1080/15476286.2023.2206248 **PMid:37129556 PMCid:PMC10155629**

Huang L, Zhang H, Deng D, Zhao K, Liu K, Hendrix DA, and Mathews DH (2019). LinearFold: Linear-time approximate RNA folding by 5'-to-3' dynamic programming and beam search. Bioinformatics, 35(14): i295-i304. https://doi.org/10.1093/bioinformatics/btz375 **PMid:31510672 PMCid:PMC6681470**

Jabbari H, Wark I, Montemagno C, and Will S (2018). Knotty: Efficient and accurate prediction of complex RNA pseudoknot structures. Bioinformatics, 34(22): 3849-3856. https://doi.org/10.1093/bioinformatics/bty420

Jumper J, Evans R, Pritzel A et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596: 583-589. https://doi.org/10.1038/s41586-021-03819-2 **PMid:34265844 PMCid:PMC8371605**

Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, Bateman A, Finn RD, and Petrov AI (2018). Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Research, 46(D1): D335-D342. https://doi.org/10.1093/nar/gkx1038 **PMCid:PMC5753348**

Kalvari I, Nawrocki EP, Ontiveros-Palacios N et al. (2021). Rfam 14: Expanded coverage of metagenomic, viral and microRNA families. Nucleic Acids Research, 49(D1): D192-D200. https://doi.org/10.1093/nar/gkaa1047 **PMCid:PMC7779021**

Kuznetsov SV, Ren CC, Woodson SA, and Ansari A (2008). Loop dependence of the stability and dynamics of nucleic acid hairpins. Nucleic Acids Research, 36(4): 1098-1112. https://doi.org/10.1093/nar/gkm1083 **PMid:18096625 PMCid:PMC2275088**

Li Y, Chen S, Liu N, Ma L, Wang T, Veedu RN, Li T, Zhang F, Zhou H, Cheng X, and Jing X (2020a). A systematic investigation of key factors of nucleic acid precipitation toward optimized DNA/RNA isolation. Biotechniques, 68(4): 191-199. https://doi.org/10.2144/btn-2019-0109 **PMid:32066262**

Li Z, Gao T, Guo C, and Li HA (2020b). A gated recurrent unit network model for predicting open channel flow in coal mines based on attention mechanisms. IEEE Access, 8: 119819-119828. https://doi.org/10.1109/ACCESS.2020.3004624

Liao G and Deng X (2020). Leveraging social relationship-based graph attention model for group event recommendation. Wireless Communications and Mobile Computing, 2020: 8834450. https://doi.org/10.1155/2020/8834450

Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, and Hofacker IL (2011). ViennaRNA Package 2.0. Algorithms for Molecular Biology, 6: 26. https://doi.org/10.1186/1748-7188-6-26 **PMCid:PMC3319429**

Lu JS, Bindewald E, Kasprzak WK, and Shapiro BA (2018). RiboSketch: Versatile visualization of multi-stranded RNA and DNA secondary structure. Bioinformatics, 34(24): 4297-4299. https://doi.org/10.1093/bioinformatics/bty468 **PMid:29912310 PMCid:PMC6289134**

Lu XJ, Bussemaker HJ, and Olson WK (2015). DSSR: An integrated software tool for dissecting the spatial structure of RNA. Nucleic Acids Research, 43(21): e142. https://doi.org/10.1093/nar/gkv716 **PMid:26184874 PMCid:PMC4666379**

Mailler E, Paillart JC, Marquet R, Smyth RP, and Vivet-Boudou V (2019). The evolution of RNA structural probing methods: From gels to next-generation sequencing. WIREs RNA, 10: e1518. https://doi.org/10.1002/wrna.1518 **PMid:30485688**

Mao K, Wang J, and Xiao Y (2020). Prediction of RNA secondary structure with pseudoknots using coupled deep neural networks. Biophysics Reports, 6: 146-154. https://doi.org/10.1007/s41048-020-00114-x

Mao K, Wang J, and Xiao Y (2022). Length-dependent deep learning model for RNA secondary structure prediction. Molecules, 27(3): 1030. https://doi.org/10.3390/molecules27031030 **PMid:35164295**

Morris KV and Mattick JS (2014). The rise of regulatory RNA. Nature Reviews Genetics, 15(6): 423-437. https://doi.org/10.1038/nrg3722 **PMid:24776770 PMCid:PMC4314111**

Nussinov R and Jacobson AB (1980). Fast algorithm for predicting the secondary structure of single-stranded RNA. Proceedings of the National Academy of Sciences, 77(11): 6309-6313. https://doi.org/10.1073/pnas.77.11.6309 **PMid:6161375 PMCid:PMC350273**

O'shea K and Nash R (2015). An introduction to convolutional neural networks. Arxiv Preprint Arxiv:1511.08458. https://doi.org/10.48550/arXiv.1511.08458

Ou X, Zhang Y, Xiong Y, and Xiao Y (2022). Advances in RNA 3D structure prediction. Journal of Chemical Information and Modeling, 62(23): 5862-5874. https://doi.org/10.1021/acs.jcim.2c00939 **PMid:36451090**

Pan Z, Wang Y, Wang K, Ran G, Chen H, and Gui W (2022). Layer-wise contribution-filtered propagation for deep learning-based fault isolation. International Journal of Robust and Nonlinear Control, 32(17): 9120-9138. https://doi.org/10.1002/rnc.6328

Peselis A and Serganov A (2014). Structure and function of pseudoknots involved in gene expression control. WIREs RNA, 5: 803-822. https://doi.org/10.1002/wrna.1247 **PMid:25044223 PMCid:PMC4664075**

Qian X, Zhang C, Chen L, and Li K (2022). Deep learning-based identification of maize leaf diseases is improved by an attention mechanism: Self-attention. Frontiers in Plant Science, 13: 864486. https://doi.org/10.3389/fpls.2022.864486 **PMid:35574079 PMCid:PMC9096888**

Reuter JS and Mathews DH (2010). RNAstructure: Software for RNA secondary structure prediction and analysis. BMC Bioinformatics, 11: 129. https://doi.org/10.1186/1471-2105-11-129 **PMid:20230624 PMCid:PMC2984261**

Sarzynska J, Popenda M, Antczak M, and Szachniuk M (2023). RNA tertiary structure prediction using RNAComposer in CASP15. Proteins, 91(12): 1790-1799. https://doi.org/10.1002/prot.26578 **PMid:37615316**

Sato K, Akiyama M, and Sakakibara Y (2021). RNA secondary structure prediction using deep learning with thermodynamic integration. Nature Communications, 12: 941. https://doi.org/10.1038/s41467-021-21194-4 **PMid:33574226 PMCid:PMC7878809**

Sayers EW, Beck J, Bolton EE et al. (2021). Database resources of the National Center for Biotechnology Information. Nucleic Acids Research, 49(D1): D10-D17. https://doi.org/10.1093/nar/gkaa892 **PMid:33095870 PMCid:PMC7778943**

Schärfen L and Neugebauer KM (2021). Transcription regulation through nascent RNA folding. Journal of Molecular Biology, 433(14): 166975. https://doi.org/10.1016/j.jmb.2021.166975 **PMid:33811916 PMCid:PMC8184640**

Sherstinsky A (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 404: 132306. https://doi.org/10.1016/j.physd.2019.132306

Singh J, Hanson J, Paliwal K, and Zhou Y (2019). RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. Nature Communications, 10: 5407. https://doi.org/10.1038/s41467-019-13395-9 **PMid:31776342**

Singh J, Paliwal K, Zhang T, Singh J, Litfin T, and Zhou Y (2021). Improved RNA secondary structure and tertiary base-pairing prediction using evolutionary profile, mutational coupling and two-dimensional transfer learning. Bioinformatics, 37(17): 2589-2600. https://doi.org/10.1093/bioinformatics/btab165 **PMid:33704363**

Sloma MF and Mathews DH (2016). Exact calculation of loop formation probability identifies folding motifs in RNA secondary structures. RNA, 22: 1808-1818. https://doi.org/10.1261/rna.053694.115 **PMid:27852924 PMCid:PMC5113201**

Solayman M, Litfin T, Singh J, Paliwal K, Zhou Y, and Zhan J (2022). Probing RNA structures and functions by solvent accessibility: An overview from experimental and computational perspectives. Briefings in Bioinformatics, 23(3): bbac112. https://doi.org/10.1093/bib/bbac112 **PMid:35348613 PMCid:PMC9116373**

Sweeney BA, Petrov AI, Ribas CE et al. (2021). RNAcentral 2021: Secondary structure integration, improved sequence search and new member databases. Nucleic Acids Research, 49(D1): D212-D220. https://doi.org/10.1093/nar/gkaa921 **PMid:33106848 PMCid:PMC7779037**

Tan Z, Fu Y, Sharma G, and Mathews DH (2017). TurboFold II: RNA structural alignment and secondary structure prediction informed by multiple homologs. Nucleic Acids Research, 45(20): 11570-11581. https://doi.org/10.1093/nar/gkx815 **PMid:29036420 PMCid:PMC5714223**

Townshend RJ, Eismann S, Watkins AM, Rangan R, Karelina M, Das R, and Dror RO (2021). Geometric deep learning of RNA structure. Science, 373: 1047-1051. https://doi.org/10.1126/science.abe5650 **PMid:34446608 PMCid:PMC9829186**

Ursuleanu TF, Luca AR, Gheorghe L, Grigorovici R, Iancu S, Hlusneac M, Preda C, and Grigorovici A (2021). Unified analysis specific to the medical field in the interpretation of medical images through the use of deep learning. E-Health Telecommunication Systems and Networks, 10(2): 41-74. https://doi.org/10.4236/etsn.2021.102003

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, and Polosukhin I (2017). Attention is all you need. In the 31st Conference on Neural Information Processing Systems, Long Beach, USA: 1-15.

Wang L, Zhong X, Wang S, Zhang H, and Liu Y (2021). A novel end-to-end method to predict RNA secondary structure profile based on bidirectional LSTM and residual neural network. BMC Bioinformatics, 22: 169. https://doi.org/10.1186/s12859-021-04102-x **PMid:33789581 PMCid:PMC8011163**

Wang X, Gu R, Chen Z, Li Y, Ji X, Ke G, and Wen H (2023). UNI-RNA: Universal pre-trained models revolutionize RNA research. bioRxiv. https://doi.org/10.1101/2023.07.11.548588

Wang Y, Liu Y, Wang S, Liu Z, Gao Y, Zhang H, and Dong L (2020). ATTfold: RNA secondary structure prediction with pseudoknots based on attention mechanism. Frontiers in Genetics, 11: 612086. https://doi.org/10.3389/fgene.2020.612086 **PMid:33384721 PMCid:PMC7770172**

Weinberg CE, Weinberg Z, and Hammann C (2019). Novel ribozymes: Discovery, catalytic mechanisms, and the quest to understand biological function. Nucleic Acids Research, 47(18): 9480-9494. https://doi.org/10.1093/nar/gkz737 **PMCid:PMC6765202**

Xu B, Zhu Y, Cao C et al. (2022). Recent advances in RNA structurome. Science China Life Sciences, 65(7): 1285-1324. https://doi.org/10.1007/s11427-021-2116-2 **PMid:35717434 PMCid:PMC9206424**

Yang TH (2024). DEBFold: Computational identification of RNA secondary structures for sequences across structural families using deep learning. Journal of Chemical Information and Modeling, 64(9): 3756-3766. https://doi.org/10.1021/acs.jcim.4c00458 **PMid:38648189 PMCid:PMC11094721**

Yu K, Wei T, Li Z, Li J, Wang Z, and Dai Z (2020). Construction of molecular sensing and logic systems based on site-occupying effect-modulated MOF–DNA interaction. Journal of the American Chemical Society, 142(51): 21267-21271. https://doi.org/10.1021/jacs.0c10442

Zhang H, Zhang C, Li Z, Li C, Wei X, Zhang B, and Liu Y (2019). A new method of RNA secondary structure prediction based on convolutional neural network and dynamic programming. Frontiers in Genetics, 10: 467. https://doi.org/10.3389/fgene.2019.00467 **PMid:31191603 PMCid:PMC6540740**

Zhang H, Zhang Q, Shao S, Niu T, and Yang X (2020). Attention-based LSTM network for rotatory machine remaining useful life prediction. IEEE Access, 8: 132188-132199. https://doi.org/10.1109/ACCESS.2020.3010066

Zhao Q, Mao Q, Zhao Z, Yuan W, He Q, Sun Q, Yao Y, and Fan X (2023). RNA independent fragment partition method based on deep learning for RNA secondary structure prediction. Scientific Reports, 13: 2861. https://doi.org/10.1038/s41598-023-30124-x **PMid:36801945 PMCid:PMC9938198**

Zhong G, Lin X, Chen K, Li Q, and Huang K (2020). Long short-term attention. In: Ren J, Hussain A, Zhao H, Huang K, Zheng J, Cai J, Chen R, and Xiao Y (Eds.), Proceedings of the 10th International Conference on Brain Inspired Cognitive Systems

(BICS), Springer, Guangzhou, China: 45-54. https://doi.org/10.1007/978-3-030-39431-8_5

Zuker M and Stiegler P (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Research, 9(1): 133-148. https://doi.org/10.1093/nar/9.1.133 **PMid:6163133 PMCid:PMC326673**