Scalable Deep Learning for RNA Secondary Structure Prediction

Anonymous Authors[1]

Abstract

The field of RNA secondary structure prediction has made significant progress with the adoption of deep learning techniques. In this work, we present the *RNAformer*, a deep learning architecture using a single model with axial attention and recycling in the latent space. We gain performance improvements by designing the architecture focusing on a high inductive bias for modeling the adjacency matrix in the latent space and by scaling the size of the model. Our approach achieves state-of-the-art performance on the popular TS0 benchmark dataset and even outperforms methods that use external information such as pre-trained sequence embeddings. Further, we show experimentally that the *RNAformer* can learn a biophysical model of the RNA folding process.

1. Introduction

RNA molecules play a central role in many cellular processes, including regulation of transcription, translation, epigenetics, or more general differentiation and development (Morris & Mattick, 2014). These functions strongly depend on the structure of the RNA, which is defined by the secondary structure that describes the intra-molecular base-pair interactions, determined by the sequence of nucleotides. Also, the secondary structure can provide important insights into RNA behavior and guide the design of RNA-based therapeutics and nanomachines (Kai et al., 2021). Therefore, the accurate prediction of the secondary structure is very desirable and a significant problem in computational biology (Bonnet et al., 2020).

Traditionally, the problem of secondary structure prediction is solved with dynamic programming approaches that minimize the free energy (MFE) of a structure, like the most widely used algorithm, RNAfold (Hofacker et al., 1994). The optimization is based on thermodynamic parameters derived from UV melting experiments (Szikszai et al., 2022). More recently, deep-learning-based approaches have conquered the field, showing superior performance on benchmark datasets, and can further incorporate additional information e.g. embeddings from large-scale RNA sequence models (Singh et al., 2019; Chen et al., 2022).

We present in this work a state-of-the-art deep learning architecture that outperforms other methods on a commonly used benchmark dataset, such as TS0 provided by Singh et al. (2019), without ensembling or making use of additional information. Our performance improvements are mainly based on an axial attention Transformer-like architecture which has a high inductive bias for the prediction of an adjacency matrix. In contrast to the conventional used CNNs, axial attention has a receptive field of the whole pair matrix at any time and does not need to build the receptive field by depth. Further, we gain improvement by recycling to simulate a larger depth and classical scaling in terms of more training data, model parameters, and longer training times.

However, some work in the field recently raised concerns about the performance improvements of deep learning methods, questioning if the learned predictions are a result of similarities between training and test data, and if the algorithms really learn a biophysical model of the folding process (Flamm et al., 2021). Since current datasets are typically curated with regard to sequence similarity only, the performance of models mainly assesses intra-family performance (Szikszai et al., 2022), while inter-family evaluations are rarely reported. Our suggestion is to show the capability to learn a biophysical model using sequences with predicted structures from the widely used, well-defined but simplified biophysical model RNAfold. To this end, we build a dataset based on RNA family information from the Rfam (Griffiths-Jones et al., 2003) database with structure predictions from RNAfold and demonstrate that our method is capable of learning the biophysical model of the folding process. Our main contributions are:

- We propose a novel architecture for RNA secondary structure prediction based on axial attention and recycling.
- We achieve state-of-the-art results on the commonly used benchmark dataset TS0 (Section 4.1).
- We show that our method is capable of learning the underlying folding dynamics of an MFE model in an inter-family prediction setting (Section 4.2).

[1] Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by *the 2023 ICML Workshop on Computational Biology*. Do not distribute.

2. Background & Related Work

Secondary structure prediction algorithms can be roughly divided into two classes: (1) *de novo* prediction methods that seek to predict the structures directly from the nucleotide sequence and (2) *homology modeling* methods that require a set of homologous RNA sequences for their predictions (Singh et al., 2021), called an RNA family. Predictions can then be applied either within given families (intra-family predictions) or across different families (inter-family prediction). *De novo* prediction methods are typically preferred since the search for homologous sequences is time-consuming and often, there is no family information available for novel RNAs. Until recently, the field of *de novo* RNA secondary structure prediction was dominated by Dynamic Programming (DP) approaches that either build on algorithms for predicting the MFE secondary structure (Zuker & Stiegler, 1981), or algorithms to find the most likely structure (maximum expected accuracy). One disadvantage of these algorithms is that they are typically limited to the prediction of nested RNA secondary structures, i.e. they cannot predict Pseudoknots (Staple & Butcher, 2005) out-of-the-box, which are present in around 40% of RNAs (Chen et al., 2020), overrepresented in functional important regions (Staple & Butcher, 2005) and known to assist folding into 3D structures (Fechter et al., 2001). Only recently, deep-learning-based approaches conquered the field, which benefit from making few assumptions on the underlying biophysical folding process, while not being restricted to only predict a subset of possible base pairs (Singh et al., 2019), and achieved state-of-the-art performance (Chen et al., 2022). We now briefly summarize some existing methods and refer the reader to more detailed related work in Appendix B.

RNAfold (Lorenz et al., 2011) uses a DP approach for the prediction of MFE secondary structures. The version we use here is based on the energy parameters provided by the Turner nearest-neighbor model (Turner & Mathews, 2010). *SPOT-RNA* (Singh et al., 2019) uses an ensemble of models with residual networks (ResNets) (He et al., 2016), bidirectional LSTM (Schuster & Paliwal, 1997), and dilated convolution (Yu & Koltun, 2015) architectures. *SPOT-RNA* was trained on a large set of intra-family RNA data for *de novo* predictions on a newly proposed test set, TS0. *ProbTransformer* (Franke et al., 2022) uses a probabilistic enhancement of the Transformer architecture for intra-family predictions. The model is trained on a large set of available secondary structure data and evaluated on TS0. *RNA-FM* (Chen et al., 2022) uses sequence embeddings of an RNA foundation model that is trained on 23 million RNA sequences to perform intra-family predictions of RNA secondary structures in a downstream task. The foundation model consists of a large Transformer architecture, while the downstream model uses a ResNet32 (He et al., 2016).

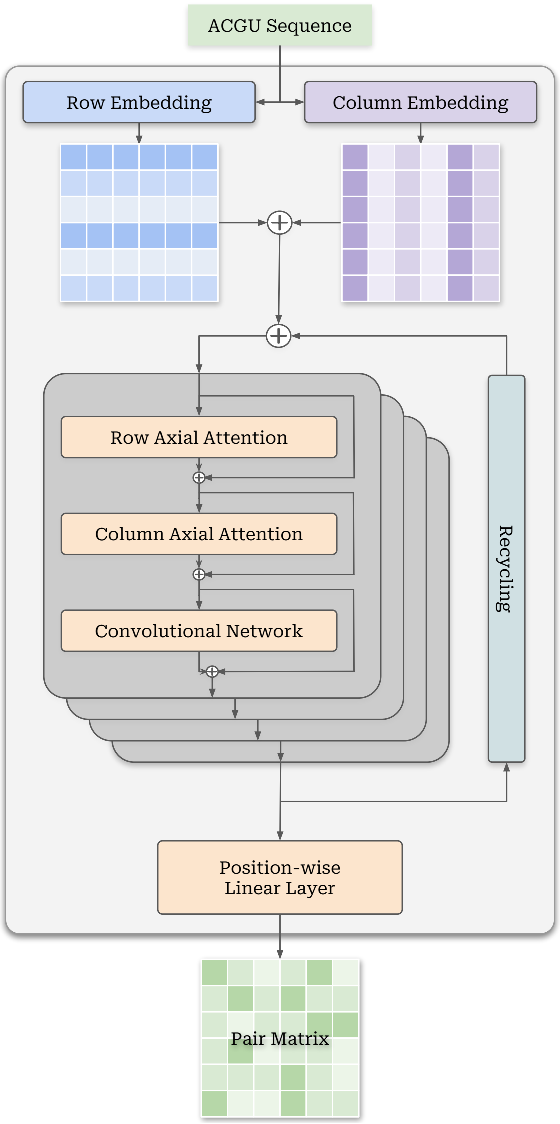

Figure 1. An overview of the *RNAformer* architecture.

3. RNAformer

Our model architecture is inspired by AlphaFold (Jumper et al., 2021), which models a multi-sequence alignment and a pair matrix in the latent space and processes it with the use of axial attention (Ho et al., 2019). In our approach, which we dub *RNAformer*[1], we simplify this architecture and only use axial attention for modeling a latent representation for the pairing between all nucleotides of the input RNA sequence. This construction leads to a potentially higher inductive bias since each layer adds some value to the latent representation of the adjacency matrix. To capture the dependency between the potential pairings we use two mechanisms: (1) *axial attention* and a (2) *convolutional*

[1] We will open source our code and model upon acceptance.

layer. Axial attention is a type of attention mechanism that captures dependencies between positions along a specific axis of the input data. In our case, we apply axial attention to the row and the column of the latent pairing matrix to create a dependency between all potential nucleotide pairings. To improve the modeling of local structures like stem-loops, we use a convolutional neural network with a kernel size of three instead of the position-wise feed-forward layer from the vanilla Transformer (Vaswani et al., 2017).

RNAformer embeds the RNA sequence with a linear layer twice and broadcasts them, one for a row-wise and one for a column-wise representation before we add them to the initial latent representation. Now we apply multiple Transformer-like blocks, each consisting of a row-wise axial attention, a column-wise axial attention, and a two-layer convolutional network. Lastly, we apply a linear layer and output the paring matrix of the secondary structure directly. Similar to AlphaFold, we apply *recycling* of the processed latent space to artificially increase the model depth and allow the model to reprocess and correct its own predictions. Therefore, we pass the latent representation multiple times through the block without gradient and calculate gradients only for the last recycle iteration. We apply dropout, pre-norm, and residual connections to all layers except the embedding and generator layers. For loss calculation, we masked 50% of the unpaired entries in the adjacency matrix before calculating the mean cross-entropy loss. This helps to increase the learning signal in the heavily imbalanced adjacency matrix. Refer to Figure 1 for an overview of our architecture.

4. Experiments

We evaluate the performance of our model in two settings. First, we evaluate the intra-family prediction capability based on the bpRNA dataset. Secondly, we assess the performance on inter-family predictions, as well as investigate the learning of a biophysical model by training the *RNAformer* on a dataset derived from Rfam database and the generated target secondary structures with RNAfold.

4.1. bpRNA Experiment

Data curation In order to generate a training dataset for intra-family predictions, we first collect a large data corpus from the following public sources: the bpRNA-1m (Danaee et al., 2018), the ArchiveII (Sloma & Mathews, 2016) and RNAStrAlign (Tan et al., 2017) dataset provided by Chen et al. (2020), all data from RNA-Strand (Andronescu et al., 2008), as well as all RNA containing data from PDB. Secondary structures for PDB samples were derived from the 3D structure information using DSSR (Lu et al., 2015). After removing duplicates we use the exact same protocol as Singh et al. (2019) to remove sequence similarities while we replace the training set TR0 with our own data. In particular, we apply a 80% similarity cutoff between the sequences using CD-HIT (Fu et al., 2012) and a homology search using BLASTN (Altschul et al., 1997) with a large e-value of 10, to further reject sequences from our training set that show homologies with the respective test sets. Most DL methods use the TS0 dataset for evaluations. However, similar to Franke et al. (2022), we did not cluster the training, validation, and test data internally to learn from the data diversity.

Table 1. The mean performance of three runs with different random seeds in comparison on the TS0 benchmark dataset. We evaluated all competitors based on their open-sourced models and will publish our evaluation script with the model and code.

Model	TS0	
	F1 Score	Solved
RNAformer $32M$+ $\circlearrowleft$	**0.728**	**17.2%**
RNAformer $32M$	0.717	16.6%
RNAformer $8M$	0.708	14.4%
RNAformer $2M$	0.677	11.4%
RNAformer $0.5M$	0.644	8.7%
RNA-FM	0.667	10.4%
ProbTransformer	0.625	11.8%
SPOT-RNA	0.597	0.05%
RNAfold	0.492	0.8%

Model & Training Setup We evaluate the *RNAformer* in a setup with 6 blocks and with different latent dimensions of 32, 64, 128, and 256, resulting in total parameter counts of roughly 0.5M, 2M, 8M, and 32M parameters, respectively. We applied recycling ($\circlearrowleft$) with 6 iterations to the largest model and sample the number of recycle iterations during the training uniformly from 2 to 6. We trained all models on 8 GPUs with a batch size of 500 tokens per GPU and a maximum sequence length of 500, for 50k steps. This limit is mainly due to the large memory footprint of the two-dimensional latent space, however, we note that the same cutoff was also applied in previous work (Singh et al., 2019). For optimization, we used AdamW (Loshchilov & Hutter, 2019) learning rate warm-up, a cosine learning rate decay, weight decay, and gradient clipping. Refer to Appendix A for all hyperparameter values.

Results We compared *RNAformer* to the models in the related work and present the results in Table 1. For a more comprehensive comparison refer to Appendix D. Our largest model with 32M parameters with the use of recycling achieves a new state-of-the-art result on the TS0 benchmark set. We solve 17.2% of the sequences completely without any mistakes. The recycling ($\circlearrowleft$) leads to a performance gain of $\sim 1\%$ and a steady increase of the parameter count from 0.5M to 32M also leads to a steady performance increase. This shows that we gain performance from overparameterization and enforces the fact that the inductive bias induced by the architecture is beneficial for this task.

4.2. Rfam Experiment

Data curation To evaluate the performance on inter-family predictions, as well as investigate the learning of a biophysical model, we derive a training dataset from families of the Rfam database version 14.9 (Kalvari et al., 2020). We first select all families with a covariance model with maximum *CLEN* of $\leq$ 500 and sample a large set of sequences for each family from the covariance models using Infernal (Nawrocki & Eddy, 2013). We then build a large set with two third sequences from families with *CLEN* $\leq$ 200 and one-third of sequences from the families with *CLEN* $>$ 200 to increase the number of families further. We randomly select 25 and 30 families from this set for validation and testing, respectively, and leaf all samples from other families for training. All sequences are folded using *RNAfold* (Lorenz et al., 2011). We apply a length cutoff at 200 nucleotides since we expect *RNAfold* predictions to be more reliable for sequences below this threshold, to save computational costs, and since all datasets of experimentally derived RNA structures from the literature show a maximum sequence length below 200 nucleotides. Singh et al. (2021) created a test set, TS-hard, in an inter-family manner similar to the data pipelines used by the Rfam database for RNA family assignments. We follow this pipeline to remove similar sequences between our training data and the validation- and test sets provided by Singh et al. (2021) using CD-HIT and BLASTN as described before. We then build an MSA of all sequences in TS-hard with BLASTN at an e-value of 0.1 using NCBI's nt database as a reference and build covariance models from the MSAs using Infernal. However, while Singh et al. (2021) used *SPOT-RNA* for predictions of the consensus structures of the MSA, which appears inappropriate since the method was built for *de-novo* predictions, we use *LocARNA-P* (Will et al., 2012), a commonly used tool to build MSAs based on sequence and structure-based alignments. The covariance models were then used to remove all sequences from the training data, using an e-value threshold of 0.1. We use this dataset to learn the underlying biophysical model of *RNAfold*, evaluated on the Rfam test data, and for evaluations on TS-hard. Again we avoid clustering the datasets internally to keep structural diversity. All datasets are described in more detail in Table 4 in Appendix C.

Model & Training Setup We used the same setup as in the first experiment with the difference of a maximum sequence length of 200 tokens, a batch size of 600 tokens per GPU, and a training time of 100k steps.

Table 2. We train different sizes of our model on the Rfam dataset on three different random seeds and report the mean performance.

Model	Rfam TS		TS-hard
	F1 Score	Solved	F1 Score
RNAformer $32M+\circlearrowleft$	**0.963**	**82.7%**	**0.651**
RNAformer $32M$	0.936	65.2%	0.642
RNAformer $8M$	0.925	60.0%	0.639
RNAformer $2M$	0.870	37.2%	0.625
RNAfold			0.636

Results As shown in Table 2, we can replicate the RNAfold algorithm increasingly better with growing model size. Our largest model achieves a mean F1 score of 94.8 on the test set and predicts 76.3% of the structures entirely correct. This result suggests that the *RNAformer* can learn the underlying biophysical model of the folding process. We observe similar results regarding scaling for the TS-hard dataset, where F1 scores increase with model size, resulting in a similar performance as RNAfold, which further supports our observation on the Rfam dataset. Interestingly, our larger models even slightly outperform RNAfold on TS-hard. However, these results require further investigations and a closer look at what the *RNAformer* layers models in detail, before we speculate about whether these results originate from inductive biases in the *RNAformer* architecture, or simply from slight deviations from the learned biophysical model.

5. Conclusion & Future Work

We introduced a new architecture for RNA secondary structure prediction and showed state-of-the-art performance on the TS0 benchmark set. The gain in performance is based on axial attention, a recycling of the latent space, and a larger dataset based on the same similarity criteria as used in related work. We also trained the *RNAformer* on a dataset derived from the Rfam database with RNAfold prediction to demonstrate that we can learn a biophysical model like RNAfold. The downside of our approach is a large memory footprint. Our approach could be further improved by the usage of additional information like MSA (Singh et al., 2021) or language embeddings with additional text information. We could also improve the architecture and enhance it with a probabilistic layer to capture ambiguities (Franke et al., 2022) or scale it even further. Another way to improve or adapt our model is finetuning, which is heavily used for large language models and could be applicable to fine-tuning high-quality data. However, besides methodological improvements, more effort in the generation and collection of high-quality data is required to achieve accurate predictions of RNA structures with deep learning.

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic acids research*, 25(17):3389–3402, 1997.

Andronescu, M., Bereg, V., Hoos, H. H., and Condon, A. Rna strand: the rna secondary structure and statistical analysis database. *BMC bioinformatics*, 9:1–10, 2008.

Bonnet, É., Rzażewski, P., and Sikora, F. Designing rna secondary structures is hard. *Journal of Computational Biology*, 27(3):302–316, 2020.

Chen, J., Hu, Z., Sun, S., Tan, Q., Wang, Y., Yu, Q., Zong, L., Hong, L., Xiao, J., King, I., et al. Interpretable rna foundation model from unannotated data for highly accurate rna structure and function predictions. *arXiv preprint arXiv:2204.00300*, 2022.

Chen, X., Li, Y., Umarov, R., Gao, X., and Song, L. Rna secondary structure prediction by learning unrolled algorithms. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=S1eALyrYDH.

Danaee, P., Rouches, M., Wiley, M., Deng, D., Huang, L., and Hendrix, D. bprna: large-scale automated annotation and analysis of rna secondary structure. *Nucleic acids research*, 46(11):5381–5394, 2018.

Fechter, P., Rudinger-Thirion, J., Florentz, C., and Giege, R. Novel features in the trna-like world of plant viral rnas. *Cellular and Molecular Life Sciences CMLS*, 58 (11):1547–1561, 2001.

Flamm, C., Wielach, J., Wolfinger, M. T., Badelt, S., Lorenz, R., and Hofacker, I. L. Caveats to deep learning approaches to rna secondary structure prediction. *Biorxiv*, pp. 2021–12, 2021.

Franke, J., Runge, F., and Hutter, F. Probabilistic transformer: Modelling ambiguities and distributions for rna folding and molecule design. *Advances in Neural Information Processing Systems*, 35:26856–26873, 2022.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. Cd-hit: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23):3150–3152, 2012.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S. R. Rfam: an RNA family database. *Nucleic Acids Research*, 31(1):439–441, 01 2003. ISSN 0305-1048.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016. doi: 10.1109/CVPR.2016.90.

Ho, J., Kalchbrenner, N., Weissenborn, D., and Salimans, T. Axial attention in multidimensional transformers. *arXiv preprint arXiv:1912.12180*, 2019.

Hofacker, I., Fontana, W., Stadler, P., Bonhoeffer, S., Tacker, M., and Schuster, P. Fast Folding and Comparison of RNA Secondary Structures. *Monatshefte fuer Chemie/Chemical Monthly*, 125:167–188, 02 1994.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

Kai, J., Yaya, H., Fei, W., Lihua, W., Chunhai, F., and Jiang, L. Structurally reconfigurable designer rna structures for nanomachines. *Biophysics Reports*, 7(1):21–34, 2021.

Kalvari, I., Nawrocki, E. P., Ontiveros-Palacios, N., Argasinska, J., Lamkiewicz, K., Marz, M., Griffiths-Jones, S., Toffano-Nioche, C., Gautheret, D., Weinberg, Z., Rivas, E., Eddy, S. R., Finn, R., Bateman, A., and Petrov, A. I. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Research*, 49(D1):D192–D200, 11 2020. ISSN 0305-1048. doi: 10.1093/nar/gkaa1047. URL https://doi.org/10.1093/nar/gkaa1047.

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. Viennarna package 2.0. *Algorithms for Molecular Biology*, 6 (1):26, Nov 2011. ISSN 1748-7188.

Loshchilov, I. and Hutter, F. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Lu, X.-J., Bussemaker, H. J., and Olson, W. K. Dssr: an integrated software tool for dissecting the spatial structure of rna. *Nucleic acids research*, 43(21):e142–e142, 2015.

Morris, K. V. and Mattick, J. S. The rise of regulatory rna. *Nature Reviews Genetics*, 15(6):423–437, 2014.

Nawrocki, E. P. and Eddy, S. R. Infernal 1.1: 100-fold faster rna homology searches. *Bioinformatics*, 29(22): 2933–2935, 2013.

Schuster, M. and Paliwal, K. K. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11):2673–2681, 1997.

Singh, J., Hanson, J., Paliwal, K., and Zhou, Y. Rna secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nature communications*, 10(1):1–13, 2019.

Singh, J., Paliwal, K., Zhang, T., Singh, J., Litfin, T., and Zhou, Y. Improved rna secondary structure and tertiary base-pairing prediction using evolutionary profile, mutational coupling and two-dimensional transfer learning. *Bioinformatics*, 37, 2021.

Sloma, M. F. and Mathews, D. H. Exact calculation of loop formation probability identifies folding motifs in rna secondary structures. *RNA*, 22(12):1808–1818, 2016.

Staple, D. W. and Butcher, S. E. Pseudoknots: Rna structures with diverse functions. *PLoS biology*, 3(6):e213, 2005.

Szikszai, M., Wise, M., Datta, A., Ward, M., and Mathews, D. H. Deep learning models for rna secondary structure prediction (probably) do not generalize across families. *Bioinformatics*, 38(16):3892–3899, 2022.

Tan, Z., Fu, Y., Sharma, G., and Mathews, D. H. Turbofold ii: Rna structural alignment and secondary structure prediction informed by multiple homologs. *Nucleic acids research*, 45(20):11570–11581, 2017.

Turner, D. H. and Mathews, D. H. Nndb: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic acids research*, 38 (suppl_1):D280–D282, 2010.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Will, S., Joshi, T., Hofacker, I. L., Stadler, P. F., and Backofen, R. Locarna-p: accurate boundary prediction and improved detection of structural rnas. *Rna*, 18(5):900–914, 2012.

Yu, F. and Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

Zuker, M. and Stiegler, P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9(1):133–148, 1981.

Appendix

A. Training

Group	Parameter	Value
Training	accelerator	GPU
	devices	8
	gradient_clip_val	1.0
	max_steps	50000 (100000)
	seed	1 / 2 / 3
Optimizer	optimizer	AdamW
	learning_rate	0.001
	weight_decay	0.1
	betas	[0.9, 0.98]
	eps	1.0e-09
	adam_w_mode	true
	num_warmup_steps	2000
	decay_factor	0.01
	schedule	cosine annealing
Model	vocab_size	5
	max_len	500 (200)
	model_dim	256 / 128 / 64 / 32
	n_layers	6
	num_head	4 / 2 / 1 / 1
	ff_kernel	3
	cycling	6
	resi_dropout	0.1
	embed_dropout	0.1
	relative position encoding	True
	ln_eps	1e-5
	softmax_scale	True
	key_dim_scaler	True
	flash_attn	True
	initializer_range	0.02
Data	dataset	bpRNA (Rfam)
	random_ignore_mat	0.5
	num_cpu_worker	32
	num_gpu_worker	8
	min_len	10
	max_len	500 (200)
	batch_token_size	500 (600)
	shuffle_pool_size	100

Table 3. The hyperparameters of the *RNAformer* training.

B. Related Work

As described in Section 2, RNA secondary structure prediction was previously dominated by dynamic programming approaches the either optimize for MFE or maximum expected accuracy (MEA) predictions. The runtime of these approaches in $\mathcal{O}\left(n^3\right)$. However, linear time approximations have been proposed Huang et al. (2019). Besides runtime, the major disadvantage of these algorithms is that they are typically limited to the prediction of nested RNA secondary structures, which strongly limits their accuracy (Szikszai et al., 2022). Some work, however, used heuristic approaches to overcome this issue, again at the price of runtime (Theis et al., 2010; Sato et al., 2011).

In this regard, deep learning approaches have strong advantages, especially when modeling the RNA secondary structure as an adjacency matrix, where all types of pairs and pseudoknots are represented identically. We now discuss existing deep learning approaches in more detail.

SPOT-RNA Singh et al. (2019) was the first algorithm using deep neural networks for end-to-end prediction of RNA secondary structures, using an ensemble of models with residual networks (ResNets) He et al. (2016), bidirectional LSTM- (Hochreiter & Schmidhuber, 1997) (BiLSTMs) (Schuster & Paliwal, 1997), and dilated convolution (Yu & Koltun, 2015) architectures. *SPOT-RNA* was trained on a large set of intra-family RNA data for *de novo* predictions on TS0, and further fine-tuned on a small set of experimentally-derived RNA structures, for predictions including tertiary interactions. However, the performance for these types of base pairs was rather poor and the currently available version of the algorithm excludes tertiary interactions from its outputs.

E2efold (Chen et al., 2020) uses a Transformer encoder architecture for *de novo* prediction of RNA secondary structures. The algorithm was trained on a dataset of homologous RNAs and showed strongly reduced performance across evaluation in multiple other publications (Sato et al., 2021; Fu et al., 2022), which indicates strong overfitting. We use the same data as the respective work for evaluations and thus exclude *E2efold* from our evaluations.

MXFold2 (Sato et al., 2021) seeks to learn the scoring function for a subsequent DP algorithm using a CNN/BiLSTM architecture. The network is trained to predict scores close to a set of thermodynamic parameters. In contrast to the previously described methods, *MXFold2* is restricted to predicting a limited set of base pairs due to the DP algorithm.

UFold (Fu et al., 2022) employs a UNet (Ronneberger et al., 2015) architecture for *de novo* secondary structure prediction, additionally reporting results for predictions on data that contains tertiary interactions after fine-tuning the model. In *UFold* an RNA sequence is an image of all possible base-pairing maps and an additional map for pair probabilities, represented as square matrices.

SPOT-RNA2 (Singh et al., 2021) is a *homology modeling* method that incorporates MSA features as well as sequence profiles (PSSM) and features derived from direct coupling analysis (DCA) for the prediction of RNA secondary structures. Similar to *SPOT-RNA*, predictions are based on an ensemble of models but using dilated convolutions only. Since *SPOT-RNA2*'s predictions are based on evolutionary features and homologous sequence information, the predictions can be considered intra-family wise independent of the curation of the dataset since homologies between the evolutionary information and the training or test sets were not explicitly excluded during evaluations. Nevertheless, we use the carefully designed test set, TS-hard, proposed by Singh et al. (2021) for our evaluations on inter-family predictions as described in Section C.

ProbTransformer (Franke et al., 2022) uses a probabilistic enhancement for either an encoder or decoder transformer architecture for intra-family predictions. The model is trained on a large set of available secondary structure data and evaluated on TS0. By learning a hierarchical joint distribution in the latent, the ProbTransformer is the first learning algorithm that is capable of sampling different structures of this latent distribution, which was shown by reconstructing structure ensembles of a distinct dataset with multiple structures for a given input sequence.

RNA-FM (Chen et al., 2022) uses sequence embeddings of an RNA foundation model that is trained on 23 million RNA sequences from 800000 species to perform intra-family predictions of RNA secondary structures in a downstream task. The foundation model consists of a 12-layer transformer architecture, while the downstream models use a ResNet32 architecture.

REDfold (Chen & Chan, 2023) uses a residual encoder-decoder architecture inspired by the UNet architecture of UFold. Interestingly, the model input is a $146 \times L \times L$ tensor, representing square matrices of all possible base pairs (10 combinations for dinucleotide pairs) and tetranucleotide combinations (136 combinations) without considering their order. The model is trained on highly homogeneous data, reporting strong performance on 4-fold cross-validation experiments, but also reporting strong results when considering sequence similarity. However, when we evaluated REDfold on TS0, we did not observe the same performance (see Table 5). Together with the results on unseen families provided by Chen & Chan (2023), this might indicate potential overfitting.

We note that there are other methods we do not consider here because they either showed inferior performance to methods we compare against (Zhang et al., 2019; Rezaur Rahman Chowdhury et al., 2019; Saman Booy et al., 2022; Wayment-Steele et al., 2022) or because their source code is not publicly available (Jung et al.).

C. Data

Dataset	# Samples	Min – Max Length	Mean Length	# Families
TS-hard	28	34 – 189	65.6	–
Rfam Test	3344	37 – 182	79.4	30
Rfam Valid	2727	34 – 160	80.2	25
Rfam Train	410408	22 – 200	95.2	3796
TS0	1305	22 – 499	136.1	–
VL0	1291	33 – 497	132.1	–
bpRNA Train	40836	13 – 500	123.0	–

Table 4. Dataset overview.

D. Experiments

Model	TS0	
	F1 Score	Solved
RNAformer $32M + \circlearrowleft$	**0.728**	**17.2%**
RNAformer $32M$	0.717	16.6%
RNAformer $8M$	0.708	14.4%
RNAformer $2M$	0.677	11.4%
RNAformer $0.5M$	0.644	8.7%
RNA-FM*	0.667	10.4%
ProbTransformer	0.625	11.8%
SPOT-RNA	0.597	0.05%
MXFold2	0.550	1.4%
UFold	0.588	3.8%
RNAfold	0.492	0.8%
LinearFold-C	0.509	1.2%
LinearFold-V	0.493	0.8%
RNAStructure	0.490	0.6%
pKiss	0.450	0.3%
CONTRAfold	0.522	0.8%
IpKnot	0.504	0.4%
REDfold	0.475	2.2%

Table 5. Performance comparison on the TS0 benchmark dataset. We report the mean performance out of the runs with different random seeds. *The number differs from their publication since we used their open-sourced model and our evaluation script which will be publicly available upon acceptance. We note, however, that the *RNAformer* also achieves a higher F1 score than reported in the publication of RNA-FM.

References

Chen, C.-C. and Chan, Y.-M. Redfold: accurate rna secondary structure prediction using residual encoder-decoder network. *BMC bioinformatics*, 24(1):1–13, 2023.

Chen, J., Hu, Z., Sun, S., Tan, Q., Wang, Y., Yu, Q., Zong, L., Hong, L., Xiao, J., King, I., et al. Interpretable rna foundation model from unannotated data for highly accurate rna structure and function predictions. *arXiv preprint arXiv:2204.00300*, 2022.

Chen, X., Li, Y., Umarov, R., Gao, X., and Song, L. Rna secondary structure prediction by learning unrolled algorithms. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=S1eALyrYDH.

Franke, J., Runge, F., and Hutter, F. Probabilistic transformer: Modelling ambiguities and distributions for rna folding and molecule design. *Advances in Neural Information Processing Systems*, 35:26856–26873, 2022.

Fu, L., Cao, Y., Wu, J., Peng, Q., Nie, Q., and Xie, X. Ufold: fast and accurate rna secondary structure prediction with deep learning. *Nucleic acids research*, 50(3):e14–e14, 2022.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016. doi: 10.1109/CVPR.2016.90.

Hochreiter, S. and Schmidhuber, J. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

Huang, L., Zhang, H., Deng, D., Zhao, K., Liu, K., Hendrix, D. A., and Mathews, D. H. Linearfold: linear-time approximate rna folding by 5’-to-3’dynamic programming and beam search. *Bioinformatics*, 35(14):i295–i304, 2019.

Jung, A. J., Lee, L. J., Gao, A. J., and Frey, B. J. Rtfold: Rna secondary structure prediction using deep learning with domain inductive bias.

Rezaur Rahman Chowdhury, F., Zhang, H., and Huang, L. Learning to fold rnas in linear time. *bioRxiv*, pp. 852871, 2019.

Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

Saman Booy, M., Ilin, A., and Orponen, P. Rna secondary structure prediction with convolutional neural networks. *BMC bioinformatics*, 23(1):58, 2022.

Sato, K., Kato, Y., Hamada, M., Akutsu, T., and Asai, K. Ipknot: fast and accurate prediction of rna secondary structures with pseudoknots using integer programming. *Bioinformatics*, 27(13):i85–i93, 2011.

Sato, K., Akiyama, M., and Sakakibara, Y. Rna secondary structure prediction using deep learning with thermodynamic integration. *Nature communications*, 12(1):1–9, 2021.

Schuster, M. and Paliwal, K. K. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11): 2673–2681, 1997.

Singh, J., Hanson, J., Paliwal, K., and Zhou, Y. Rna secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nature communications*, 10(1):1–13, 2019.

Singh, J., Paliwal, K., Zhang, T., Singh, J., Litfin, T., and Zhou, Y. Improved rna secondary structure and tertiary base-pairing prediction using evolutionary profile, mutational coupling and two-dimensional transfer learning. *Bioinformatics*, 37, 2021.

Szikszai, M., Wise, M., Datta, A., Ward, M., and Mathews, D. H. Deep learning models for rna secondary structure prediction (probably) do not generalize across families. *Bioinformatics*, 38(16):3892–3899, 2022.

Theis, C., Janssen, S., and Giegerich, R. Prediction of rna secondary structure including kissing hairpin motifs. In *Algorithms in Bioinformatics: 10th International Workshop, WABI 2010, Liverpool, UK, September 6-8, 2010. Proceedings 10*, pp. 52–64. Springer, 2010.

Wayment-Steele, H. K., Kladwang, W., Strom, A. I., Lee, J., Treuille, A., Becka, A., Participants, E., and Das, R. Rna secondary structure packages evaluated and improved by high-throughput experiments. *Nature Methods*, 19(10): 1234–1242, 2022.

Yu, F. and Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

Zhang, H., Zhang, C., Li, Z., Li, C., Wei, X., Zhang, B., and Liu, Y. A new method of rna secondary structure prediction based on convolutional neural network and dynamic programming. *Frontiers in genetics*, 10:467, 2019.