# Hybrid BiLSTM machine learning with RNAfold-based thermodynamic modeling for RNA secondary structure prediction

Jophy Lin
Science & Engineering
Manalapan High School
Englishtown, NJ
126jlin@frhsd.com

**Abstract**

RNA secondary structure plays an important role in various biological fields, such as gene regulation, catalysis, and RNA–protein interactions, yet accurate prediction is challenging due to long-range base-pairing and structural complexity. This work presents a hybrid framework that integrates RNAfold's thermodynamic modeling with machine learning to improve RNA secondary structure prediction and interpretability. Using a balanced dataset of approximately 9,700 RNA sequences spanning multiple RNA families ($\approx$ 4.09 million nucleotides), traditional thermodynamic prediction with RNAfold was evaluated alongside a bidirectional long short-term memory (BiLSTM) neural network trained for base-wise pairing prediction. While RNAfold achieved a baseline test accuracy of 0.77, the BiLSTM reached 0.91 accuracy. Building on these results, several hybrid approaches were developed that selectively combine RNAfold and neural predictions, including a base-wise selector, a sequence-level meta-learner, and a Monte Carlo (MC) dropout uncertainty method. The best-performing hybrid model, the MC dropout uncertainty method, achieved a test accuracy of 0.917, outperforming both standalone approaches. This framework has been deployed through an interactive web interface, enabling users to input RNA sequences and compare prediction methods in real-time. To enhance interpretability, predicted structures are converted from dot-bracket notation into annotated visual diagrams depicting the corresponding secondary structure motifs. This study demonstrates that hybrid modeling with uncertainty-aware selection can improve RNA secondary structure prediction while maintaining accessibility and interpretability for later biological analysis.

**Index Terms**

deep learning, RNA, thermodynamic modeling, ribonucleic acid, secondary structure, BiLSTM neural network, Monte Carlo, machine learning, hybrid, RNAfold