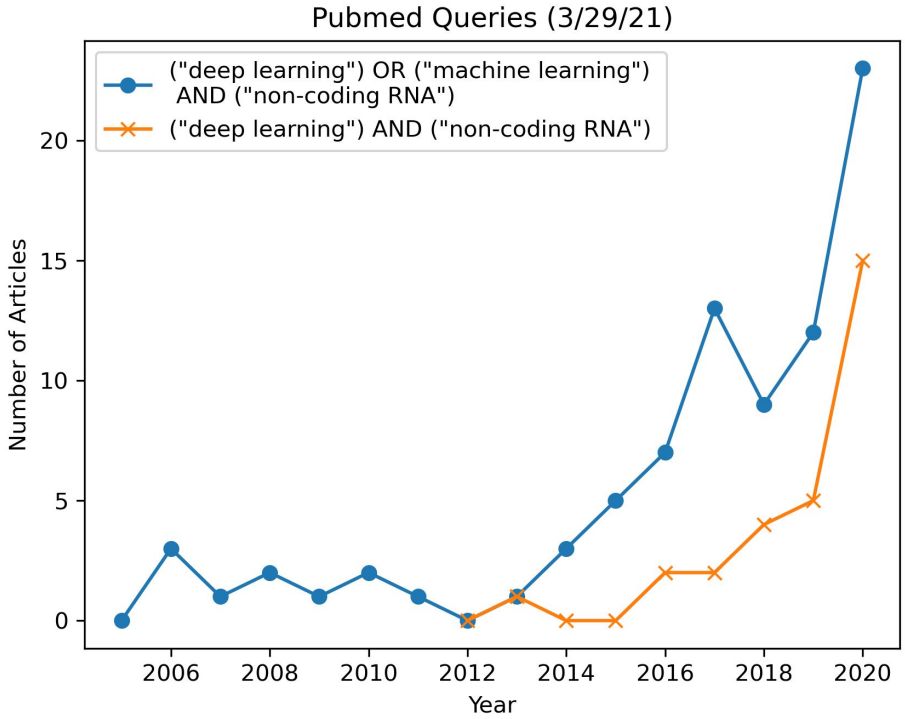
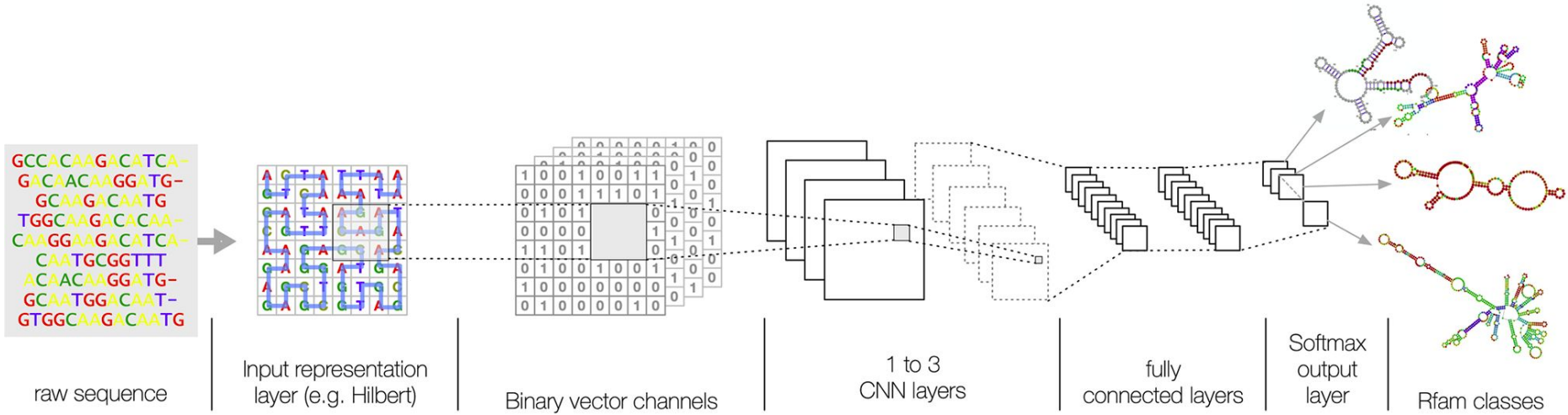
ML and DL for ncRNA prediction and classification problems:



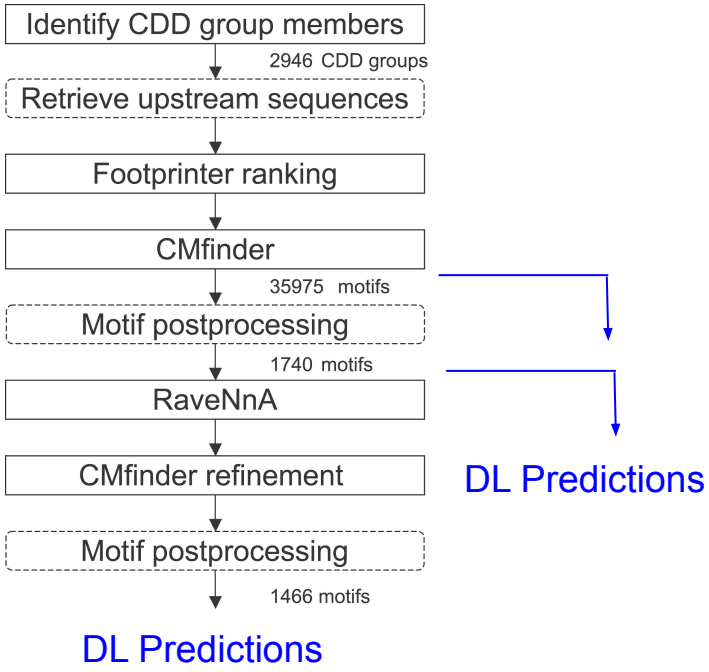
Deep learning (DL) models are growing in popularity for ncRNA prediction and classification problems



Noviello TMR, Ceccarelli F, Ceccarelli M, Cerulo L (2020) Deep learning predicts short non-coding RNA functions from only raw sequence data. PLOS Computational Biology 16(11): e1008415. https://doi.org/10.1371/journal.pcbi.1008415

DL combined with ncRNA motif finding:

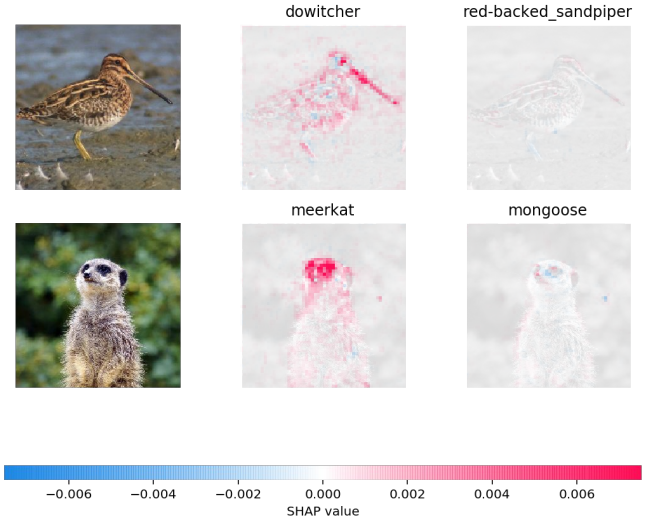
- Use pre-trained classification and function prediction models to complement the ncRNA discovery pipeline from (Yao *et al.*, 2007)
- Some suggested steps:
  - Review state-of-the-art DL (or ML) models for short ncRNA function prediction/classification models
  - For discovered motifs, create test sets of your motif example sequences and predict them



Yao Z, Barrick J, Weinberg Z, Neph S, Breaker R, et al. (2007) A Computational Pipeline for High-Throughput Discovery of *cis*-Regulatory Noncoding RNA in Prokaryotes. PLOS Computational Biology 3(7): e126. https://doi.org/10.1371/journal.pcbi.0030126

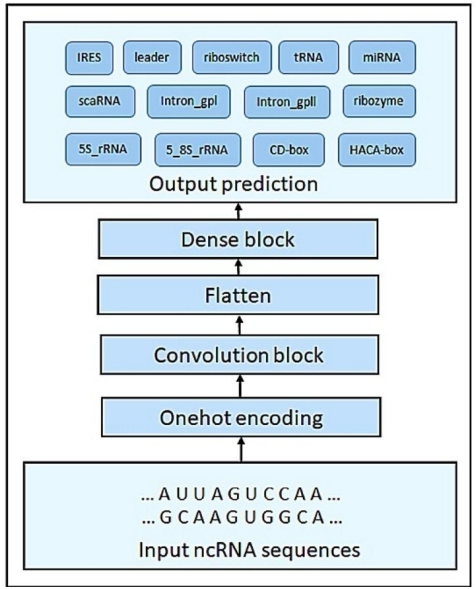
Potential interesting questions:

- *Feature attribution*: figuring out what inputs were the most important for a model's prediction
- For sequence based ncRNA DL predictors, how do their predictions relate to RNA secondary structures? Can we apply current feature attribution methods to figure this out?



https://github.com/slundberg/shap

- Some suggested steps:
  - Review state of the art DL models for those which use only sequence features for predictions & select some to try
  - Review state of the art feature attribution methods for DL networks (deepSHAP, deepLIFT, Integrated Gradients, etc.) & select some to try
  - Gather a test set of structures and sequences
  - Predict the function/families and run the attribution methods for these predictions
  - Visualize importance
  - Look for what the networks might learn implicitly about secondary structure



a.

Chantsalnyam T, Lim DY, Tayara H, Chong KT (2020) ncRDeep: Non-coding RNA classification with convolutional neural network. Computational Biology and Chemistry 88: 107364. https://doi.org/10.1016/j.compbiolchem.2020.107364.