Part 2: Critical Thinking

Due 10/16/25 @ Noon
Email the Completed Document to jemohl@utep.edu

Name:

Article Review (50 points)

a) Title of Article:

ModiDeC: A multi-RNA modification classifier for direct nanopore sequencing.

b) First Author's Name:

Nicolò Alagna

c) What was the main motivation of the paper?

The main motivation behind this paper is to overcome several challenges in the field of epitranscriptomics. Being the three main challenges: The ***lack of distinctions*** of certain RNA regions, the ***inaccurate on predictions*** for certain regions or certain RNA types in Direct RNA sequencing for improvement in RNA diagnosis and the ***evolving nature of nanopore sequencing chemistry***.

The goal was to create a reliable, easy-to-use, and highly **customizable deep-learning tool (ModiDeC)** to address these issues and advance the analysis of the epitranscriptome for both research and clinical diagnostics.

d) What was the major conclusion from the paper and why is it useful?

The major conclusion for this project was:

ModiDeC, can **accurately recognize and distinguish multiple RNA modifications** (m^6A, inosine, pseudouridine, 2'-O-methylguanosine, and m^1A) simultaneously and at single-nucleotide resolution. The tool achieved high accuracy (>91% of modification position/type correctly identified in training) and demonstrated a very low false-positive prediction rate (ranging from 0% to 5.5%) on synthetic, HEK293T cell, and human blood samples.

"(...) A result of particular note is that ModiDeC seems to have versatility in identifying motifs that differ from those used for training but contain variations in the sequence (...). This indicates that ModiDeC has potential for identifying features that have not been explicitly learned from the training pool and would thus also be suitable for use in human data, in which a large number of genetic variants, such as single nucleotide polymorphisms, must be assumed."

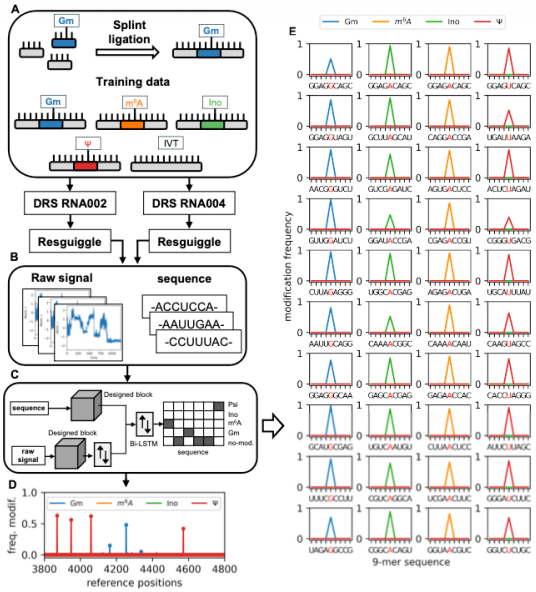
This conclusion is useful because quantitatively describes the accuracy of the model and improves the "prediction accuracy" and the "distinctions" which is the original aim and purpose, but now also extending the obtained conclusion from certain sequences to another modified sequences with different features, which probes the utility of the model.

e) What was the key methods described and why were they used?

1. **Isolation of nuclear and cytoplasmic rRNA:** In this method rRNA was refined, homogenized and prepared for being processed.
2. **DRS library preparation:** Important because these libraries will be the basis for training the model.
3. **Neural Network training and processing:** In this step the NN was trained, then is the "*main key process*" method because a variation in training can lead to a very good or very poor classifier.

f) Which of the figures/tables was most informative and why was it?

Figure 1:



This is the most informative image because it summarizes the whole process from beginning to end of this paper, also giving a clear idea of what is really happening during all of the steps.

g) What is one thing you wish they would have done and why do you think it would have been beneficial?

One thing that I'd like to see is a definition of a ***"healthy volunteer"*** because in the methods section they mention ***"Peripheral blood was obtained from a healthy volunteer"*** and then the selection criteria in my very humble opinion turns obscure, like a Blackbox, also maybe using more than one volunteer (I think at least 10), because this person can have some mutation or

extremely rare condition that can bias the whole experiment. With these changes I think the paper will become more solid in terms of methodology, also more trustable.