

COVID-19 MRNA VACCINE DEGRADATION PREDICTION

Sankalp Biswal(17103314),
Shobhit Agarwal(17103359)

JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY,
NOIDA

Introduction

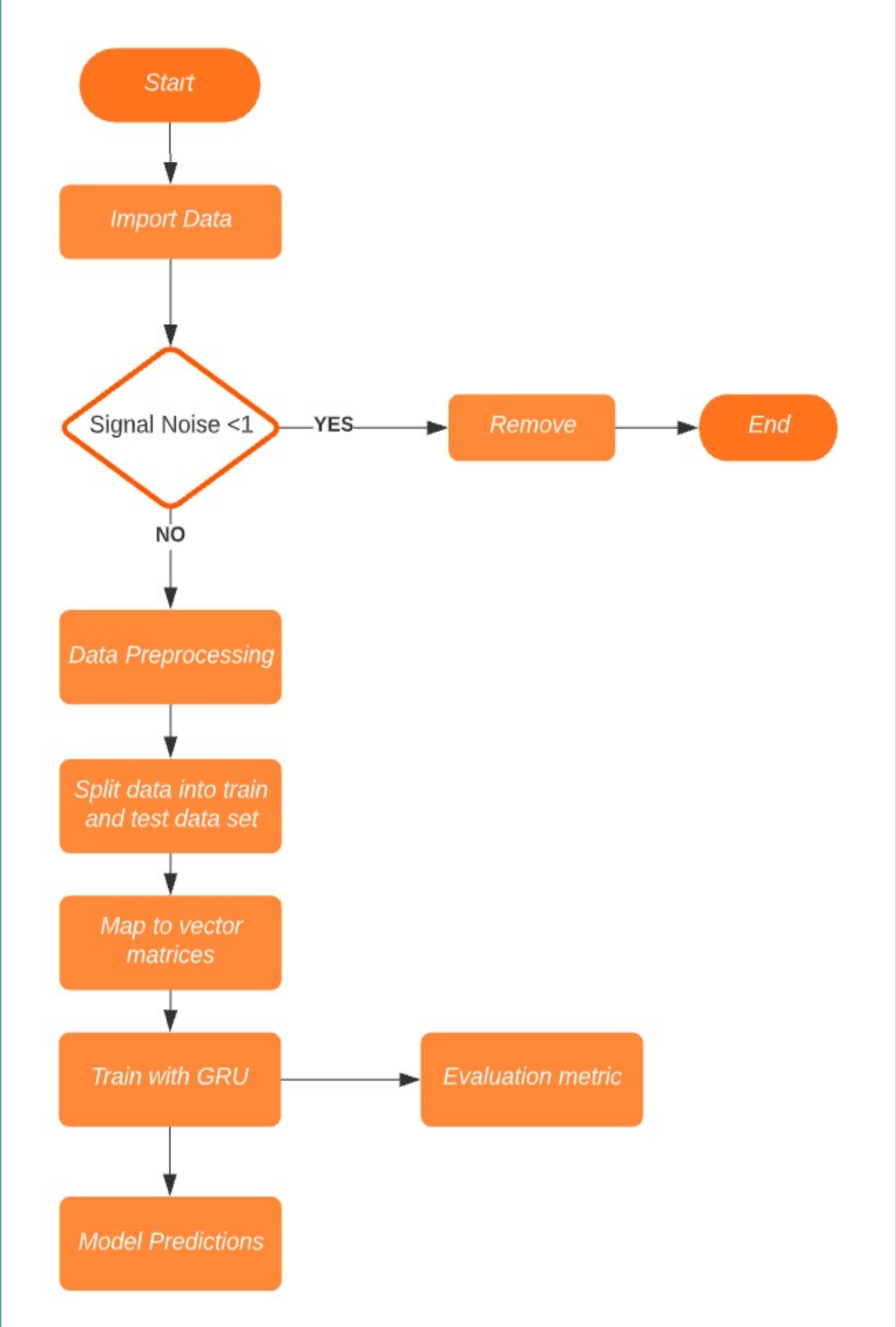
- mRNA vaccines are receiving increased interest as potential alternatives to conventional methods for the prevention of several diseases, including Covid-19.
- This paper proposes and evaluates Gated Recurrent Unit networks as a method to predict the stability/reactivity and risk of degradation of sequences of RNA.

Problem Statement

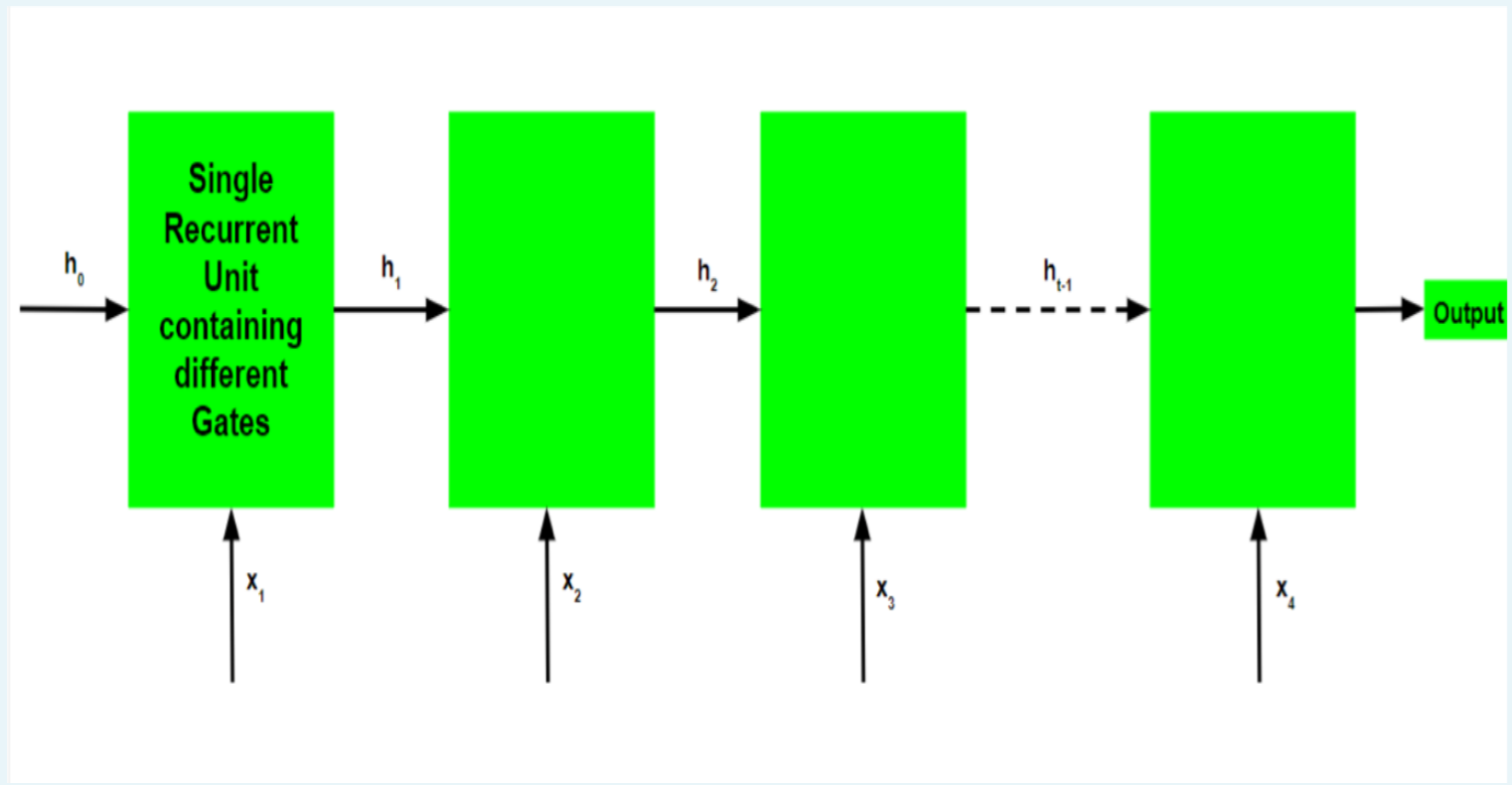
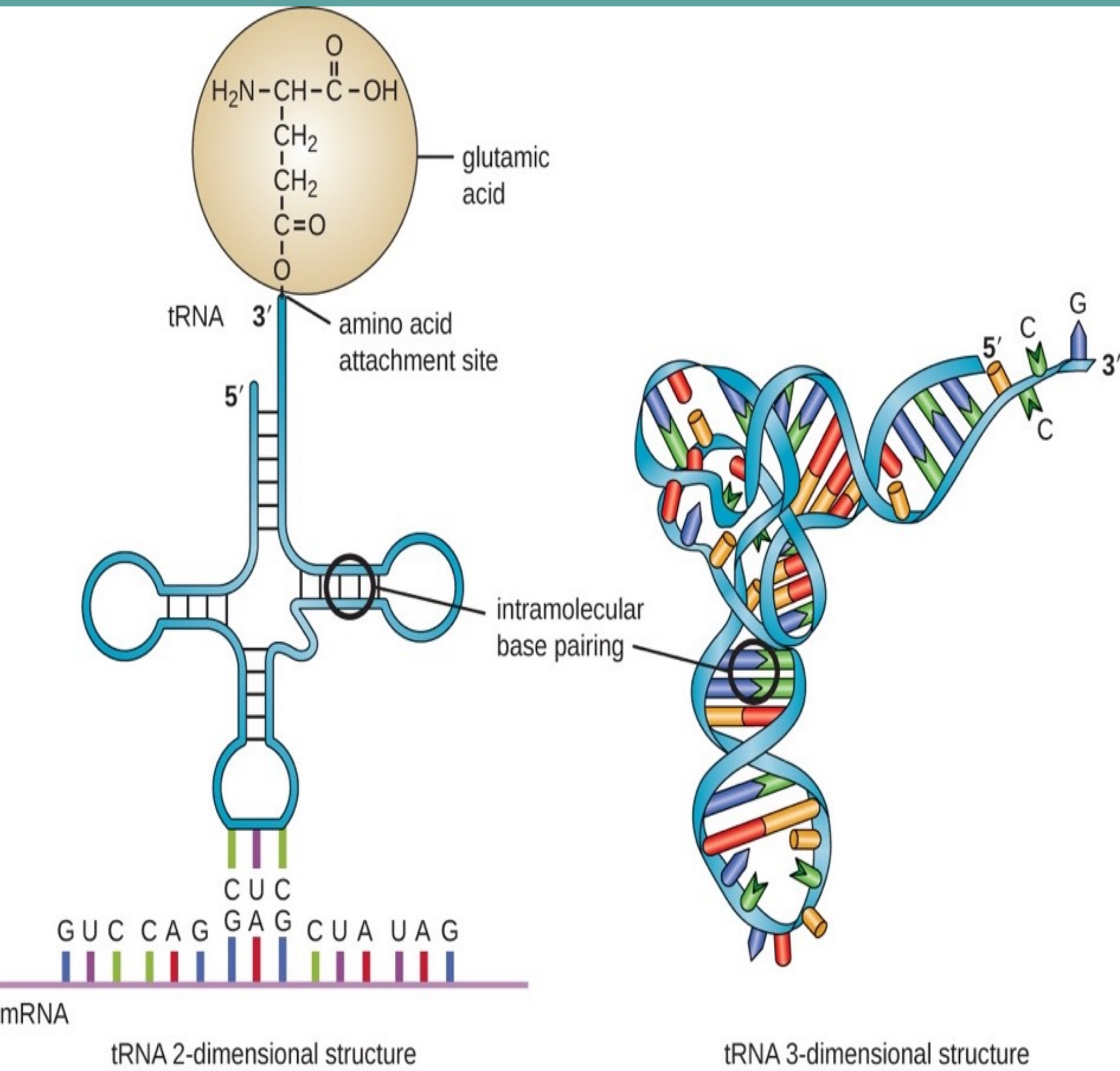
- This project proposes method to predict the stability and risk of degradation of sequences of RNA.
- The main task of the project is to take the sequence and other structural features of an RNA molecule and predict its stability.

Brief Description of the Solution Approach

- We are proposing a system that will predict likely degradation rates at each base of an RNA molecule, trained on a subset of an Eterna dataset comprising over 3000 RNA molecules.
- The aim of this project is to present a Deep Learning approach to this problem through the usage of the dataset.
- The model applied is GRU to assess whether Machine Learning methods can provide helpful results in predicting the reactivity and degradation of mRNA molecules.



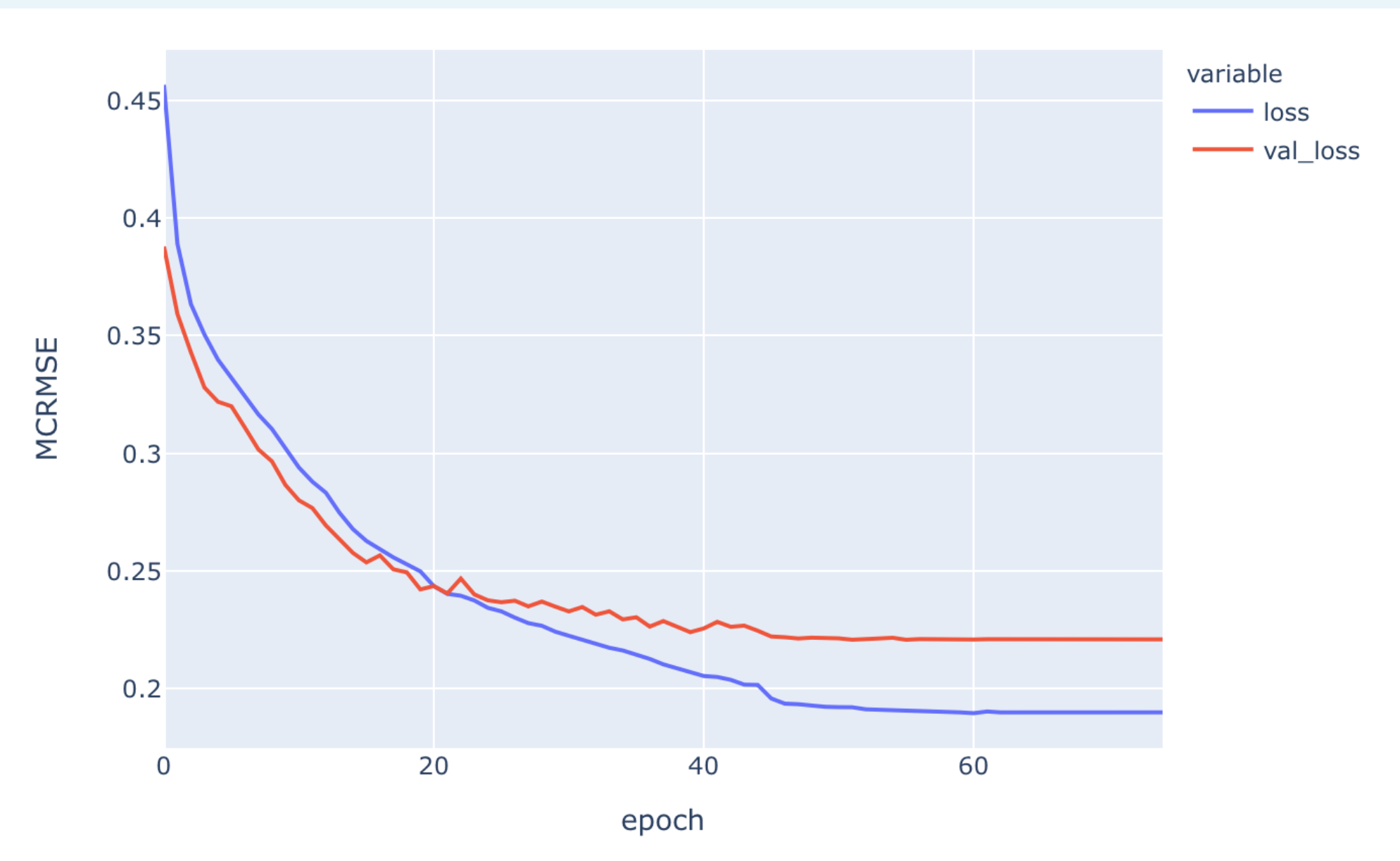
High level diagram of the overall model architecture



Flow of GRU

Method

- We have used Bi-directional GRU for prediction of values of various predictor columns.
- For bidirectional GRU networks, the recurrent networks were duplicated, made to work in the opposing direction and combined with the original forward networks by concatenating their outputs. It has 3 layers and has a dropout.
- The implementation of Bi-directional GRU has been completed.
- The different hyperparameters were tested and the minimal characteristics were selected after their performances ceased to improve.
- Epoch value of 50 has been selected along with adam optimizer and sigmoid activation function and loss metric as MCRMSE.



Conclusion

- Through this project we've predicted the values of degradation rates at various locations along the RNA sequence and calculated the training loss=0.1716 and validation loss=0.2162.
- The difference in the gap is the generalization gap. Through the predictions, newer sequences have been validated and can be used for building stable RNA sequences.