



# OpenVaccine: COVID-19

## mRNA Vaccine Degradation Prediction

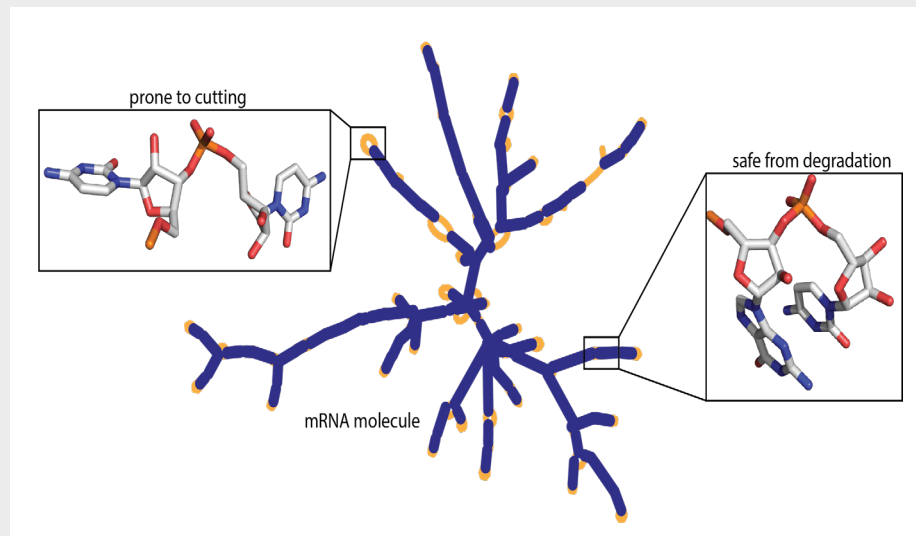
Stuart Hopkins

### Introduction

Winning the fight against the COVID-19 pandemic will require an effective vaccine that can be equitably and widely distributed. Building upon decades of research has allowed scientists to accelerate the search for a vaccine against COVID-19, but every day that goes by without a vaccine has enormous costs for the world nonetheless. We need new, fresh ideas from all corners of the world.

mRNA vaccines have taken the lead as the fastest vaccine candidates for COVID-19, but currently, they face key potential limitations. One of the biggest challenges right now is how to design super stable messenger RNA molecules (mRNA). Conventional vaccines (like your seasonal flu shots) are packaged in disposable syringes and shipped under refrigeration around the world, but that is not currently possible for mRNA vaccines.

The object of this model is to predict the which mRNA connections are prone to cutting. With this information, vaccine researchers will be able to make a more stable and transportable vaccine. Below is a diagram showing which connections are prone to cutting and which are safe from degradation.

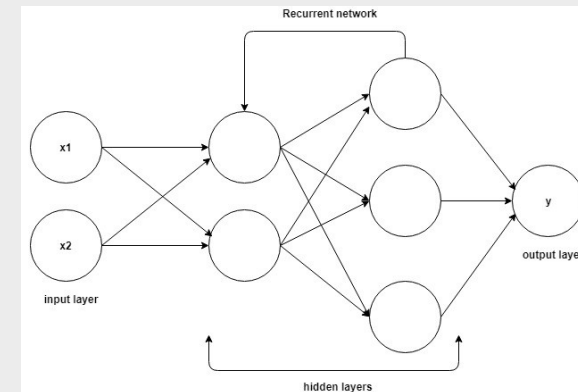


(<https://www.kaggle.com/c/stanford-covid-vaccine>)

### Approach

To achieve a good result, I used an LSTM network. This allowed the model to have control over many aspects of the training process.

(<https://towardsdatascience.com/machine-learning-recurrent-neural-networks-and-long-short-term-memory-lstm-python-keras-example-86001ceaaebc>)



### Results

Name	Submitted	Wait time	Execution time	Score
submission.csv	a few seconds ago	0 seconds	4 seconds	0.27658
Complete				
<a href="#">Jump to your position on the leaderboard</a>				

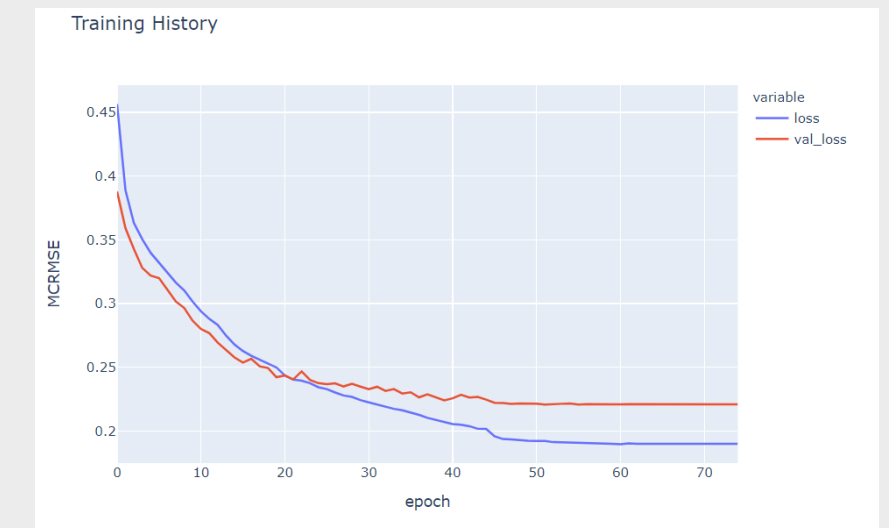
The results were posted to a kaggle competition (titled the same as this poster). Based on their hidden tests this LSTM based model achieved a score of .27658. This score is based off of mean squared error (MSE) of normalized data. This means that in practice, this model is around very accurate at predicting mRNA connection survival.

This score of .27658 places this LSTM model in 1,363th place. Based on my lack of experience in datascience this is a very good result.

1329	Jose Carmona	< openVaccine v0.3...	0.27591	13	2mo
1330	Horia Velicu		0.27598	16	2mo
1331	Lucius lu		0.27611	5	2mo
1332	Abhinay		0.27623	9	2mo
1333	wilsonsujames		0.27657	1	3mo
1334	raccoon		0.27660	1	2mo
1335	greyhound		0.27666	1	2mo
1336	ChongAiH		0.27678	5	2mo
1337	RM		0.27682	37	2mo
1338	Belinda Trotta		0.27695	14	2mo

To improve upon this models accuracy, I would use a more suffisticated neural network (NN). Because of the complexity of data given, a neural network with a larger hidden layer would most likely produce better results. Some of the top models for this data set are using multiple NNs such as Gated NN and Convolutional NN combined to achieve a higher accuracy.

### Results



As seen above, the chart models accuracy as time goes on. Each epoch is an iteration of the model. With traditional models, extended processing of data can result in overfitting. However, the observer can see dropout's stablizing effect around epoch 45.

We are measuring the performance based on the mean squared error (MSE). The observer can see that the val\_loss is close to .223 MSE which is very good. This model achieves over good accuracy when approximating mRNA connection survival.

### Conclusions

Based on the results, this Long Short Term Memory (LSTM) NN model is good but could be improved upon. The score of 0.27658 it received is nothing special; however, it is around average for this competition. The parameters used maximized its efficacy and dropout stablized its runtime. In order to achieve a better accuracy, it is recommended to use a combination of networks. This will allow the mRNA data to be processed more effectively.