**DE NO DRUG MOLECULE DISCOVERY FOR COVID-19**

Adamu Abubakar Abba, February 2020

**ABSTRACT**

Artificial intelligence is currently the go to technology for solving a lot of tasks that previously seemed unachievable in the past. Deep Learning in particular has achieved success in a range of fields across different industries and sectors. In Drug discovery today sequence generation models can been used to synthesize chemical molecules with important and unique characteristics that would qualify them as ideal drug candidates.

In this report, I have used an LSTM-based RNN model for SMILES sequence generation, particularly training and validating on the GDBChEMBL SMILES datasets [1], after generation, Open Babel GUI software was used to convert the SMILES to pdbqt (AutoDock) formats and later PyRx accepts the pdbqt formats and plots the 3D structures of the generated molecules. Finally the COVID-19 FASTA genome [2] is also loaded and the binding affinity of the two is computed, and after a few iterations, I was able to identify some candidate molecules can could serve as potential Inhibitors for the COVID-19.

1. **INTRODUCTION**

Artificial Intelligence is no doubt the talk of the town right now, the hottest tech topic showing tremendous progress with a variety of applications and use cases. With the recent breakthroughs particularly in deep learning, in the areas of image recognition and object localization/segmentation using Convolutional Neural Networks (CNNs), Language Translation, Sentiment Analysis, Image Generation using Generative Adversarial Networks (GANs) and Variational Auto-Encoders (VAEs), Sequence Generation using both CNNs and Recurrent Neural Networks (RNNs) particularly Long-Term Short-Term Memory (LSTM) based RNNs and in Reinforcement Learning (RN) as in the case of an RN model defeating the world champion at AlphaGO. Deep Learning has definitely proved itself to be the most successful Machine Learning subset, and is no doubt attracting a lot more attention from the community.

In the field of Animal Science particularly in medicine, image recognition models have shown a wide range of applications in object identification, for instance in tumor image analysis for the identification of malignant or benign tumors, thereby helping speed up the time taken to thoroughly analyze cells image and identify distinguishing features.

Recently they have also been shown to be very good at molecular generation, harnessing their sequence generation characteristics used in the case of language models for text sequence generation. In molecular sequence generation, both CNN and LSTM models have been used to generate SMILES (Simplified Molecular Input Line Entry System) molecular representations for prospective ligands.

**­­­­­­COVID-19­­­­­­­­­**

The COVID-19 or nCoV-19 is a novel virus belonging to the large class of coronaviruses of which there are seven (7) currently identified, including the Severe Acute Respiratory Syndrome (SARS-CoV) and Middle East Respiratory Syndrome (MERS-CoV) which all have a flu-like / common cold symptoms and are also known for their respiratory limiting disorders, fever and shortness of breath amongst their patients. The viruses have been identified to be mostly common amongst animals like bats, cats but have recently evolved to be contactable by humans as in the cases of SARS and COVID-19 etc.

The COVID-19 was first reported in late December of 2019 at a local fish market in Wuhan, Hubei Province of China, and was officially identified and named by the World Health Organization (WHO) in January 2020. As of 29 February 2020, more than 86,000 cases have been confirmed in 62 territories, of which 8,000 were classified as serious. At least 2,900 deaths have been attributed to the disease, most in mainland China with more than 100 deaths in other countries. More than 39,000 people have recovered so far. [1]

**LIGANDS**

Ligands are chemical molecules that have unique structures and show some important chemical activities in the interactions between their composing atoms, these chemical activities can be harnessed to develop molecular vaccines that could be used to fight some specific viruses that have affinity for such ligands.

Ligands bind with the proteomes (virus, bacteria etc.) in an appropriate chemical environment and binding site to form a complex, during this chemical association energy is released and sometimes these energy could be sufficient enough to alter the normal function or operation of such a virus.

When designing a possible ligand to target a particular virus both the binding affinity and the quantity in nano-molecules (nM) required to successfully bind is very crucial. A ligand with a very high binding affinity for the target virus and having a very low required quantity to bind or occupy 50% or more of the binding site is usually more preferable.

**­­­­**

1. **METHODS**

**THE DATASETS**

ChemBL datasets [1] were used to train the model

**THE MODEL**

An LSTM-based RNN model was used as the generator model.

**DICUSSIONS**

**RESULTS AND FINDINGS**

**CONCLUSION AND RECOMMENDATION**

**REFERENCES**

1. <https://www.ebi.ac.uk/pdbe/node/1>
2. <https://www.ncbi.nlm.nih.gov/genbank/>
3. <https://www.wikipedia.org>