

“Predicting Repurposable Drugs with Machine Learning against the Novel Coronavirus”

Literature Review

Tahsin Saira

Introduction

The COVID-19 pandemic caused by the novel coronavirus (2019-nCoV/SARS-CoV-2) was accompanied by high mortality rates (2-4% for SARS-CoV-2), indicating the urgent need for effective treatment at the beginning of the outbreak to prevent the spread. However, with existing drug development and application systems, this is not possible, as it takes many years for newly created drugs to reach the market. Hence, drug repurposing or repositioning technique whereby already commercially existing drugs are used to treat emerging and challenging diseases, including COVID-19, has become a promising approach. **Further studies of these machine learning led bioinformatics research may offer a methodological pathway to identify repurposable drugs for future and neglected diseases underserved by the costs and extended timeline.**

Computer vision, speech recognition, natural language understanding, and digital pathology data analysis are just a few applications of ML uses. Similarly, by identifying hidden patterns and evidence from biomedical data, ML has revolutionized drug discovery. In this Review, we focus on 3 research journals on how ML applications in COVID-19 drug repurposing has provided quick and cost-effective therapeutic development solutions.

MT_DTI

Deep learning is the method of investigating data with layers of linear and non-linear transformations organized in a hierarchical manner, referred to as a subfield of machine learning. Bo Ram Beck and colleagues published a study where they used their pre-trained deep learning-based drug-target interaction model called Molecule Transformer-Drug Target Interaction (MT-DTI) to identify commercially available antiviral drugs that could act on viral proteins of SARS-CoV-2. The results showed that **atazanavir**, an antiretroviral drug used to treat and prevent HIV, is the most effective chemical compound, against the SARS-CoV-2 3C-like proteinase, followed by **remdesivir, efavirenz, ritonavir, and dolutegravir** [1].

MT_DTI can accurately predict binding affinities based on chemical sequences (SMILES) and amino acid sequences (FASTA) of a **target protein**, without their structural information [1]. **One possible drawback of this approach could be:** identifying drugs targeted for uncharacterized proteins with traditional three-dimensional (3D) structure-based docking approaches.

Nevertheless, the experiment goes into deep comparison analysis of the MT-DTI with deep learning-based (DeepDTA) approach and two traditional machine learning-based algorithms SimBoost and KronRLS, with the KIBA and DAVIS data sets and **concluded that MT-DTI showed the best performance**. Moreover, to confirm the performance of MT-DTI at least in silico

(experiment performed via computer simulation), 60 known FDA-approved antiviral drugs and 3410 FDA-approved drugs were evaluated by means of the MT-DTI deep learning-based affinity score, and AutoDock Vina (a widely used 3D structure-based docking algorithm) docking score [1]. **Since the results of both algorithms showed moderate similarities, it is not possible to determine which algorithm is more reliable without various experimental evaluations.**

DeepCE

Even though phenotypic screening has a slight advantage over target-based screening when it comes to identifying first-in-class drugs and cell-active compounds or where the disease target is unknown, target-based high-throughput screening dominates the conventional drug discovery process because Phenotypic screening is often met with low throughput and difficulty in target deconvolution [2].

To address these problems, Thai-Hoang Pham and colleagues designed a mechanism driven neural network-based model, DeepCE which utilizes a graph neural network, captures high-dimensional correlations between biological features, as well as non-linear relationships between biological features and outputs, in order to predict gene expression profiles when given a de novo dataset [2].

Gene expression profiling is the study of the pattern of genes expressed at the transcriptional level in a specific cell or under specific conditions to obtain a global picture of cellular activity. **Gene expression profiling can be used for drug repurposing, discovering drug structures, main compound identification and predicting side effects for preclinical compounds, therefore necessary** [2].

The performance of DeepCE25 model was explored by comparing it with TT_WOPT and it outperformed by a large margin not only in de novo chemical settings but also in the traditional setting. To further demonstrate the value of DeepCE, the researchers used a chemical-induced gene expression profile to discover potential drugs for COVID-19 treatment. Out of all the drugs identified, **nine drugs are antiviral drugs** and seven of them are used for treating hepatitis C, two drugs are **immunosuppressive agents**, two medicines have **anti-inflammatory** or **immuno-regulatory** properties and may be used to control the immune response in COVID-19 infection [2].

In the research paper, it is stated that to generate the patient profiles, eight SARS-CoV-2 patients and twelve healthy samples (population-based) as well as from one SARS-CoV-2 patient and two healthy samples (individual-based) we collected [2]. The size and variance of the dataset could be limiting here. Supervised learning algorithms like neural networks are still quite effective with small training data. However, more than the size of the data, what is important is the variations in the samples and it should have been reported in the experiment.

Nonetheless, a high-throughput, mechanism-driven phenotype compound screening method will provide new opportunities for discovering safe and effective therapeutics that modulate the biological system. In conclusion, DeepCE could be an effective instrument for phenotype-based compound screening.

Network-based Drug Repurposing

A journal based on Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2 by Yadi Zhou et al, presented an integrative, antiviral drug repurposing methodology implementing a systems pharmacology-based network medicine platform, quantifying the interplay between the HCoV–host interactome and drug targets in the human protein–protein interaction network [3].

Network proximity measures the distances between two modules, such as drug–target and disease–gene modules. Using network proximity analyses of drug targets and HCoV–host interactions in the human interactome, the researchers prioritized 16 potential anti-HCoV repurposable drugs (e.g., **melatonin, mercaptopurine, and sirolimus**) that are further validated by enrichment studies of drug–gene signatures and HCoV-induced transcriptomics data in human cell lines were used to confirm the findings. The authors further identify three potential **drug combinations** (e.g., sirolimus plus dactinomycin, mercaptopurine plus melatonin, and toremifene plus emodin) captured by the “Complementary Exposure” pattern: the targets of the drugs both hit the HCoV–host subnetwork but target separate neighbourhoods in the human interactome network [3]. The identified drugs were: **Mesalazine** (an approved drug for inflammatory bowel disease), **sirolimus** (an approved immunosuppressive drug), and **equilin** (an approved agonist of the estrogen receptor for menopausal symptoms) achieved the highest GSEA (Gene Set Enrichment Analysis) scores of 3, followed by **paroxetine** and **melatonin** with GSEA scores of 2 [3].

Drawbacks of the study can be summarized as predictions being non-SARS-CoV-2 specific due to lack of the known host proteins on SARS-CoV-2 (predictions are rather for a pan-HCoV dataset), also that there is no experimental validation for the predicted drugs.

DeepCE and Network-based model were published earlier in 2020 when science had less understanding of SARS-CoV-2. Even so, we can draw some potential conclusions. MT_DTI was able to identify only antiviral drugs, whereas DeepCE was able to identify not only antiviral drugs, but also drugs with **immunosuppressive, anti-inflammatory, or immuno-regulatory agents** that may prove effective against SARS-CoV-2. The network proximity analysis allows us to even identify drug combinations. The second paper identified anti-inflammatory agents, and the third paper identified melatonin (also an anti-inflammatory agent). Viral infections associated with immune-inflammatory injury can be treated by so. Although melatonin cannot eradicate or even limit the viral replication, it may prolong patients’ survival time, which may provide a chance for patients’ immune systems to recover and eventually eradicate the virus.

References

- [1] Beck, B., Shin, B., Choi, Y., Park, S., & Kang, K. (2020). Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. *Computational And Structural Biotechnology Journal*, 18, 784-790. doi: 10.1016/j.csbj.2020.03.025
- [2] Pham, T., Qiu, Y., Zeng, J., Xie, L., & Zhang, P. (2021). A deep learning framework for high-throughput mechanism-driven phenotype compound screening and its application to COVID-19 drug repurposing. *Nature Machine Intelligence*, 3(3), 247-257. doi: 10.1038/s42256-020-00285-9
- [3] Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., & Cheng, F. (2020). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discovery*, 6(1). doi: 10.1038/s41421-020-0153-3