# Machine-Learning Models

## A classifier prediction model to predict the status of Coronavirus CoVID-19 patients in South Koread section

European Review for Medical and Pharmacological Sciences, 2020; 24: 3400-3403
https://www.europeanreview.org/wp/wp-content/uploads/3400-3403.pdf

This study uses official time series data from the Korea Centers for Disease Control and Prevention (KCDC) for 7869 Coronavirus patients in South Korea between 20/02/2020 and 09/03/2020. The dataset contains fifteen variables including patient ID, sex, birth, year, country, region, disease group infection reason, infection order, infected by, contact number, confirmation date, released date, deceased date and state. This study adopted seven variables as independent variables including sex, birth year, country, region, group, infection reason and confirmed date, where dependent variable is one of the following variables, namely death or recovered. The variables are chosen based on the most used variables in several researches. To avoid missing independent and dependent variables, only 659 and 649 patients are employed for recovered and death cases, respectively.

Classification Model: Our recommendation is to use this model to predict the status of the patients globally.

## One-shot screening of potential peptide ligands on HR1 domain in COVID-19 glycosylated spike (S) protein with deep siamese network

https://arxiv.org/pdf/2004.02136.pdf
(interesting paper)

## DeepPurpose: a Deep Learning Based Drug Repurposing Toolkit

https://arxiv.org/pdf/2004.08919
(interesting paper)

## Improving Coronavirus (COVID-19) Diagnosis using Deep Transfer Learning

https://www.medrxiv.org/content/10.1101/2020.04.11.20054643v1.full.pdf

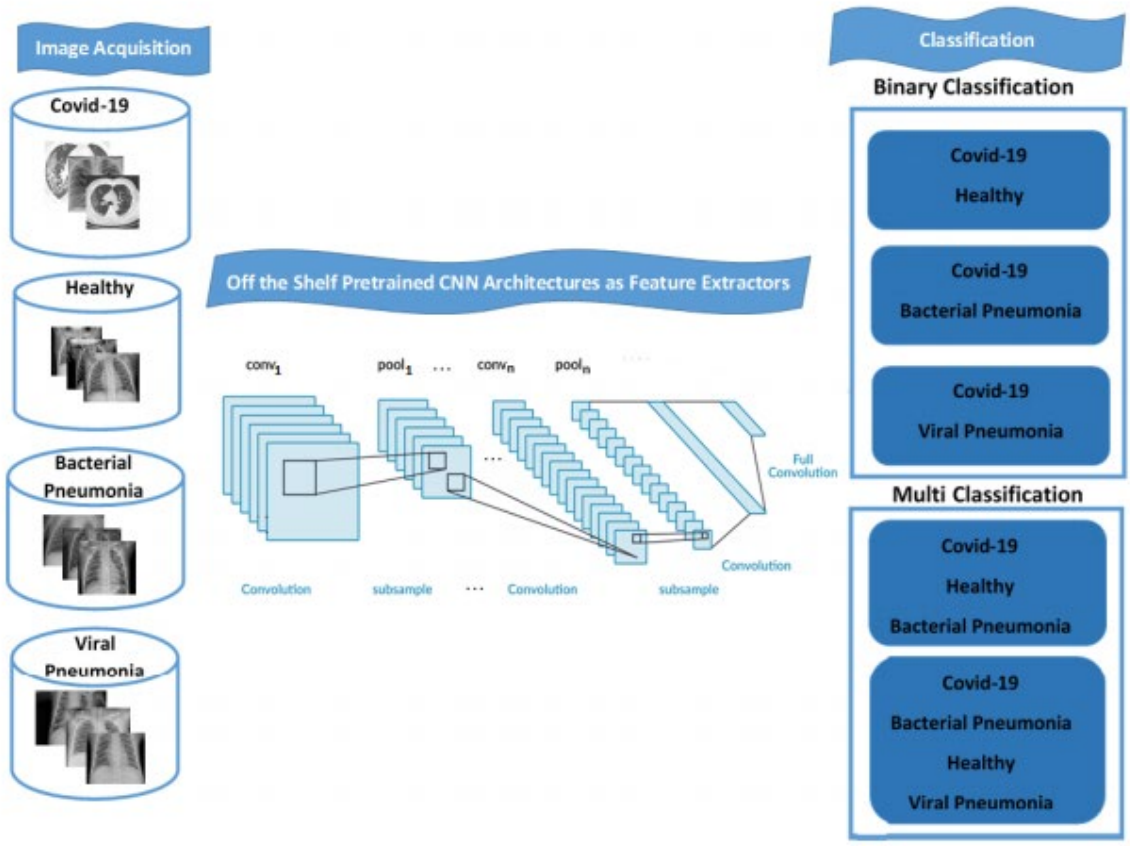


Fig. 1: Proposed system for detection and classification of COVID-19

## JCS: An Explainable COVID-19 Diagnosis System by Joint Classification and Segmentation

https://arxiv.org/pdf/2004.07054.pdf

## Deep Learning COVID-19 Features on CXR using Limited Training Data Sets

https://arxiv.org/pdf/2004.05758.pdf

Under the global pandemic of COVID-19, the use of artificial intelligence to analyze chest X-ray (CXR) image for COVID-19 diagnosis and patient triage is becoming important. Unfortunately, due to the emergent nature of the COVID-19 pandemic, a systematic collection of the CXR data set for deep neural network training is difficult. To address this problem, here we propose a patch-based convolutional neural network approach with a relatively small number of trainable parameters for COVID-19 diagnosis.

## Towards an Efficient Deep Learning Model for COVID-19 Patterns Detection in X-ray Images

https://arxiv.org/pdf/2004.05717.pdf

## Deep Learning System to Screen Coronavirus Disease 2019 Pneumonia

https://arxiv.org/ftp/arxiv/papers/2002/2002.09334.pdf

## Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CTon

https://doi.org/10.1148/radiol.2020200905

In this retrospective and multi-center study, a deep learning model, COVID-19 detection neural network (COVNet), was developed to extract visual features from volumetric chest CT exams for the detection of COVID-19. Community acquired pneumonia (CAP) and other non-pneumonia CT exams were included to test the robustness of the model. The datasets were collected from 6 hospitals between August 2016 and February 2020. Diagnostic performance was assessed by the area under the receiver operating characteristic curve (AUC), sensitivity and specificity

## A Machine Learning Application for Raising WASH Awareness in the Times of Covid-19 Pandemic

https://arxiv.org/pdf/2003.07074.pdf
(NLP based modeling)

## Predicting COVID-19 Incidence Through Analysis of Google Trends Data in Iran: Data Mining and Deep Learning Pilot Study section

https://publichealth.jmir.org/2020/2/e18828/
Comments: Data was collected from Google Trends and a NN model was built.

## An automatic COVID-19 CT segmentation based on U-Net with attention mechanism

https://arxiv.org/pdf/2004.06673.pdf

## Deep-Learning Models based on Chest Xrays

There are many papers. Growing every day. If we are interested, then we should do a fresh search to catch them all.

## COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning

https://www.biorxiv.org/content/10.1101/2020.03.20.000141v2.full#T3

## Rapid in silico design of antibodies targeting SARS-CoV-2 using machine learning and supercomputing

https://doi.org/10.1101/2020.04.03.024885

Rapidly responding to novel pathogens, such as SARS-CoV-2, represents an extremely challenging and complex endeavor. Numerous promising therapeutic and vaccine research efforts to mitigate the catastrophic effects of COVID-19 pandemic are underway, yet an efficacious countermeasure is still not available. To support these global research efforts, we have used a novel computational pipeline combining machine learning, bioinformatics, and supercomputing to predict antibody structures capable of targeting the SARS-CoV-2 receptor binding domain (RBD). In 22 days, using just the SARS-CoV-2 sequence and previously published neutralizing antibody structures for SARS-CoV-1, we generated 20 initial antibody sequences predicted to target the SARS-CoV-2 RBD. As a first step in this process, we predicted (and publicly released) structures of the SARS-CoV-2 spike protein using homology-based structural modeling. The predicted structures proved to be accurate within the targeted RBD region when compared to experimentally derived structures published weeks later. Next we used our *in silico* design platform to iteratively propose mutations to SARS-CoV-1 neutralizing antibodies (known not to bind SARS-Cov-2) to enable and optimize binding within the RBD of SARS-CoV-2. Starting from a calculated baseline free energy of −48.1 kcal/mol (± 8.3), our 20 selected first round antibody structures are predicted to have improved interaction with the SARS-CoV-2 RBD with free energies as low as −82.0 kcal/mole. The baseline SARS-CoV-1 antibody in complex with the SARS-CoV-1 RBD has a calculated interaction energy of −52.2 kcal/mole and neutralizes the virus by preventing it from binding and entering the human ACE2 receptor. These results suggest that our predicted antibody mutants may bind the SARS-CoV-2 RBD and potentially neutralize the virus. Additionally, our selected antibody mutants score well according to multiple antibody developability metrics. These antibody designs are being expressed and experimentally tested for binding to COVID-19 viral proteins, which will provide invaluable feedback to further improve the machine learning–driven designs. This technical report is a high-level description of that effort; the

Supplementary Materials includes the homology-based structural models we developed and 178,856 *in silico* free energy calculations for 89,263 mutant antibodies derived from known SARS-CoV-1 neutralizing antibodies.

## Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks

https://arxiv.org/pdf/2003.10849

The 2019 novel coronavirus (COVID-19), with a starting point in China, has spread rapidly among people living in other countries, and is approaching approximately 12,245,417 cases worldwide according to the statistics of European Centre for Disease Prevention and Control. There are a limited number of COVID-19 test kits available in hospitals due to the increasing cases daily. Therefore, it is necessary to implement an automatic detection system as a quick alternative diagnosis option to prevent COVID-19 spreading among people. In this study, three different convolutional neural network based models (ResNet50, InceptionV3 and InceptionResNetV2) have been proposed for the detection of coronavirus pneumonia infected patient using chest X-ray radiographs. ROC analyses and confusion matrices by these three models are given and analyzed using 5-fold cross validation. Considering the performance results obtained, it is seen that the pre-trained ResNet50 model provides the highest classification performance with 98% accuracy among other two proposed models (97% accuracy for InceptionV3 and 87% accuracy for Inception-ResNetV2).

## Repurpose Open Data to Discover Therapeutics for COVID-19 using Deep Learning

https://arxiv.org/ftp/arxiv/papers/2005/2005.10831.pdf

There have been more than 850,000 confirmed cases and over 48,000 deaths from the human coronavirus disease 2019 (COVID-19) pandemic, caused by novel severe acute respiratory syndrome coronavirus (SARS-CoV-2), in the United States alone. However, there are currently no proven effective medications against COVID-19. Drug repurposing offers a promising way for the development of prevention and treatment strategies for COVID-19. This study reports an integrative, network-based deep learning methodology to identify repurposable drugs for COVID-19 (termed CoV-KGE). Specifically, we built a comprehensive knowledge graph that includes 15 million edges across 39 types of relationships connecting drugs, diseases, genes, pathways, and expressions, from a large scientific corpus of 24 million PubMed publications. Using Amazon's AWS computing resources, we identified 41 repurposable drugs (including indomethacin, toremifene and niclosamide) whose therapeutic association with COVID-19 were validated by transcriptomic and proteomic data in SARS-CoV-2 infected human cells and data from ongoing clinical trials. While this study, by no means recommends specific drugs, it demonstrates a powerful deep learning methodology to prioritize existing drugs for further investigation, which holds the potential of accelerating therapeutic development for COVID-19.

## Accelerating drug repurposing for COVID-19 via modeling drug mechanism of action with large scale gene-expression profiles

https://arxiv.org/ftp/arxiv/papers/2005/2005.07567.pdf

The novel coronavirus disease, named COVID-19, emerged in China in December 2019, and has rapidly spread around the world. It is clearly urgent to fight COVID-19 at global scale. The development of methods for identifying drug uses based on phenotypic data can improve the efficiency of drug development. However, there are still many difficulties in identifying drug applications based on cell picture data. This work reported one state-of-the-art machine learning method to identify drug uses based on the cell image features of 1024 drugs generated in the LINCS program. Because the multi-dimensional features of the image are affected by non-experimental factors, the characteristics of similar drugs vary greatly, and the current sample number is not enough to use deep learning and other methods are used for learning optimization. As a consequence, this study is based on the supervised ITML algorithm to convert the characteristics of drugs. The results show that the characteristics of ITML conversion are more conducive to the recognition of drug functions. The analysis of feature conversion shows that different features play important roles in identifying different drug functions. For the current COVID-19, Chloroquine and Hydroxychloroquine achieve antiviral effects by inhibiting endocytosis, etc., and were classified to the same community. And Clomiphene in the same community inibited the entry of Ebola Virus, indicated a similar MoAs that could be reflected by cell image.

## Unveiling the molecular mechanism of SARS-CoV-2 main protease inhibition from 92 crystal structures

https://arxiv.org/pdf/2005.13653.pdf
(interesting paper)

Currently, there is no effective antiviral drugs nor vaccine for coronavirus disease 2019 (COVID-19) caused by acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Due to its high conservativeness and low similarity with human genes, SARS-CoV-2 main protease (Mpro) is one of the most favorable drug targets. However, the current understanding of the molecular mechanism of Mpro inhibition is limited by the lack of reliable binding affinity ranking and prediction of existing structures of Mpro-inhibitor complexes. This work integrates mathematics and deep learning (MathDL) to provide a reliable ranking of the binding affinities of 92 SARS-CoV-2 Mpro inhibitor structures. We reveal that Gly143 residue in Mpro is the most attractive site to form hydrogen bonds, followed by Cys145, Glu166, and His163. We also identify 45 targeted covalent bonding inhibitors. Validation on the PDBbind v2016 core set benchmark shows the MathDL has achieved the top performance with Pearson's correlation coefficient (Rp) being 0.858. Most importantly, MathDL is validated on a carefully curated SARS-CoV-2 inhibitor dataset with the averaged Rp as high as 0.751, which endows the reliability of the present binding affinity prediction. The present binding affinity ranking, interaction analysis, and fragment decomposition offer a foundation for future drug discovery efforts.

## MeSH descriptors indicate the knowledge growth in the SARS-CoV-2/COVID-19 pandemic.

https://arxiv.org/pdf/2005.06259.pdf

The scientific papers dealing with the novel betacoronavirus SARS-CoV-2 and the coronavirus disease 2019 (COVID-19) caused by this virus, published in 2020 and recorded in the database PUBMED, were retrieved on April 27, 2020. About 20% of the records contain Medical Subject Headings (MeSH),

keywords assigned to records in the course of the indexing process in order to summarise the articles' contents. The temporal sequence of the first occurrences of the keywords was determined, thus giving insight into the growth of the knowledge base of the pandemic.

## SAveRUNNER: a network-based algorithm for drug repurposing and its application to COVID-19

---

https://arxiv.org/ftp/arxiv/papers/2006/2006.03110.pdf

The novelty of new human coronavirus COVID-19/SARS-CoV-2 and the lack of effective drugs and vaccines gave rise to a wide variety of strategies employed to fight this worldwide pandemic. Many of these strategies rely on the repositioning of existing drugs that could shorten the time and reduce the cost compared to de novo drug discovery. In this study, we presented a new network-based algorithm for drug repositioning, called SAveRUNNER (Searching off-lAbel dRUg aNd NEtwoRk), which predicts drug–disease associations by quantifying the interplay between the drug targets and the disease-specific proteins in the human interactome via a novel network-based similarity measure that prioritizes associations between drugs and diseases locating in the same network neighborhoods. Specifically, we applied SAveRUNNER on a panel of 14 selected diseases with a consolidated knowledge about their disease-causing genes and that have been found to be related to COVID-19 for genetic similarity (i.e., SARS), comorbidity (e.g., cardiovascular diseases), or for their association to drugs tentatively repurposed to treat COVID-19 (e.g., malaria, HIV, rheumatoid arthritis). Focusing specifically on SARS subnetwork, we identified 282 repurposable drugs, including some the most rumored off-label drugs for COVID-19 treatments (e.g., chloroquine, hydroxychloroquine, tocilizumab, heparin), as well as a new combination therapy of 5 drugs (hydroxychloroquine, chloroquine, lopinavir, ritonavir, remdesivir), actually used in clinical practice. Furthermore, to maximize the efficiency of putative downstream validation experiments, we prioritized 24 potential anti-SARS-CoV repurposable drugs based on their network-based similarity values. These topranked drugs include ACE-inhibitors, monoclonal antibodies (e.g., anti-IFN$\gamma$, anti-TNF$\alpha$, anti-IL12, antiIL1$\beta$, anti-IL6), and thrombin inhibitors. Finally, our findings were in-silico validated by performing a gene set enrichment analysis, which confirmed that most of the network-predicted repurposable drugs may have a potential treatment effect against human coronavirus infections.

## Identifying Human Interactors of SARS-CoV-2 Proteins and Drug Targets for COVID-19 using Network-Based Label Propagation

---

https://arxiv.org/pdf/2006.01968.pdf

Motivated by the critical need to identify new treatments for COVID19, we present a genome-scale, systems-level computational approach to prioritize drug targets based on their potential to regulate hostvirus interactions or their downstream signaling targets. We adapt and specialize network label

propagation methods to this end. We demonstrate that these techniques can predict human-SARS-CoV2 protein interactors with high accuracy. The top-ranked proteins 1 arXiv:2006.01968v2 [q-bio.MN] 22 Jun 2020 that we identify are enriched in host biological processes that are potentially coopted by the virus. We present cases where our methodology generates promising insights such as the potential role of HSPA5 in viral entry. We highlight the connection between endoplasmic reticulum stress, HSPA5, and anti-clotting agents. We identify tubulin proteins involved in ciliary assembly that are targeted by anti-mitotic drugs. Drugs that we discuss are already undergoing clinical trials to test their efficacy against COVID-19. Our prioritized list of human proteins and drug targets is available as a general resource for biological and clinical researchers who are repositioning existing and approved drugs or developing novel therapeutics as anti-COVID-19 agents.

## Ontology-based systematic classification and analysis of coronaviruses, hosts, and host-coronavirus interactions towards deep understanding of COVID-19

---

https://arxiv.org/ftp/arxiv/papers/2006/2006.00639.pdf

Given the existing COVID-19 pandemic worldwide, it is critical to systematically study the interactions between hosts and coronaviruses including SARS-Cov, MERS-Cov, and SARS-CoV-2 (cause of COVID-19). We first created four host-pathogen interaction (HPI)-Outcome postulates, and generated a HPI-Outcome model as the basis for understanding host-coronavirus interactions (HCI) and their relations with the disease outcomes. We hypothesized that ontology can be used as an integrative platform to classify and analyze HCI and disease outcomes. Accordingly, we annotated and categorized different coronaviruses, hosts, and phenotypes using ontologies and identified their relations. Various COVID-19 phenotypes are hypothesized to be caused by the backend HCI mechanisms. To further identify the causal HCI-outcome relations, we collected 35 experimentally-verified HCI protein-protein interactions (PPIs), and applied literature mining to identify additional host PPIs in response to coronavirus infections. The results were formulated in a logical ontology representation for integrative HCI-outcome understanding. Using known PPIs as baits, we also developed

and applied a domain-inferred prediction method to predict new PPIs and identified their pathological targets on multiple organs. Overall, our proposed ontology-based integrative framework combined with computational predictions can be used to support fundamental understanding of the intricate interactions between human patients and coronaviruses (including SARS-CoV-2) and their association with various disease outcomes.

## Predicting potential drug targets and repurposable drugs for COVID-19 via a deep generative model for graphsection

---

https://arxiv.org/pdf/2007.02338.pdf

Coronavirus Disease 2019 (COVID-19) has been creating a worldwide pandemic situation. Repurposing drugs, already shown to be free of harmful side effects, for the treatment of COVID-19 patients is an important option in launching novel therapeutic strategies. Therefore, reliable molecule interaction data are a crucial basis, where drug-/protein-protein interaction networks establish invaluable, year-long carefully curated data resources. However, these resources have not yet been systematically exploited using high-performance artificial intelligence approaches. Here, we combine three networks, two of which are year-long curated, and one of which, on SARS-CoV-2-human host-virus protein interactions, was published only most recently (30th of April 2020), raising a novel network that puts drugs, human and virus proteins into mutual context. We apply Variational Graph AutoEncoders (VGAEs), representing most advanced deep learning based methodology for the analysis of data that are subject to network constraints. Reliable simulations confirm that we operate at utmost accuracy in terms of predicting missing links. We then predict hitherto unknown links between drugs and human proteins against which virus proteins preferably bind. The corresponding therapeutic agents present splendid starting points for exploring novel host-directed therapy (HDT) options.

## Machine-Learning Driven Drug Repurposing for COVID-19

---

https://arxiv.org/pdf/2006.14707.pdf

The integration of machine learning methods into bioinformatics provides particular benefits in identifying how therapeutics effective in one context might have utility in an unknown clinical context or against a novel pathology. We aim to discover the underlying associations between viral proteins and antiviral therapeutics that are effective against them by employing neural network models. Using the National Center for Biotechnology Information virus protein database and the DrugVirus database, which provides a comprehensive report of broad-spectrum antiviral agents (BSAAs) and viruses they inhibit, we trained ANN models with virus protein sequences as inputs and antiviral agents deemed safe-in-humans as outputs. Model training excluded SARS-CoV-2 proteins and included only Phases II, III, IV

and Approved level drugs. Using sequences for SARS-CoV-2 (the coronavirus that causes COVID-19) as inputs to the trained models produces outputs of tentative safe-in-human antiviral candidates for treating COVID-19. Our results suggest multiple drug candidates, some of which complement recent findings from noteworthy clinical studies. Our in-silico approach to drug repurposing has promise in identifying new drug candidates and treatments for other viruses.