



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

## CHAPTER 8

# Role of artificial intelligence in fast-track drug discovery and vaccine development for COVID-19

Alka Bali<sup>a</sup> and Nishu Bali<sup>b</sup>

<sup>a</sup>University Institute of Pharmaceutical Sciences, UGC Center of Advanced Study, Panjab University, Chandigarh, India

<sup>b</sup>Department of Computer Sciences, Chitkara University, Patiala, India

## 1 Introduction

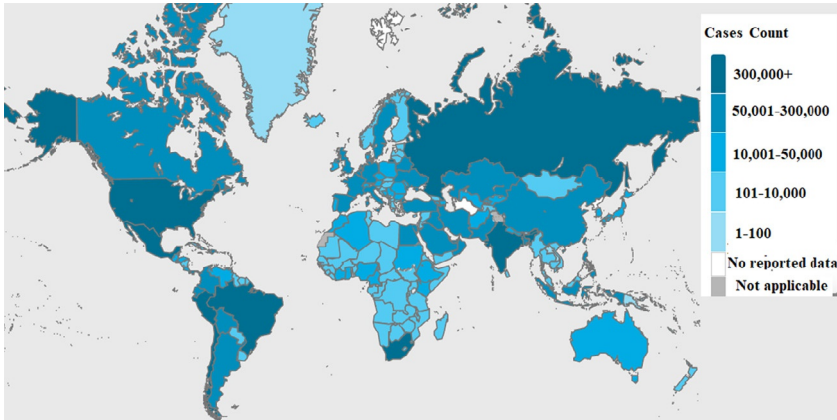
Recent times have seen a surge of reports employing computational methods and artificial intelligence (AI) for drug design. This stems from the fact that big data or data sets from modern biological techniques have become huge, complex and multidimensional and the classical computational methods do not have the capability to process them with acceptable levels of accuracy. Machine learning can accurately predict several compound properties like target binding affinity, membrane transport properties, toxicity aspects, pharmacokinetic or pharmacodynamic parameters without the need for an actual synthesis of the compounds or their evaluation.

Coronavirus Disease 2019, also known as COVID-19, was reported first from China in December 2019 and today, it has become a major pandemic currently afflicting 213 countries. The World Health Organization had pronounced this disease as a global public health emergency on January 30, 2020, presenting a huge risk for countries with vulnerable healthcare arrangements (Fig. 1) (Sohrabi, Alsafi, & O'Neill, 2020).

This pandemic has more than 22 million cases worldwide (6.48 million active cases), claiming more than 7.78 lakh lives in a short span of eight months, as per the data available on August 18, 2020 (Worldometer, 2020).

## 2 Artificial intelligence in COVID-19

Artificial intelligence is a powerful technique that has practically invaded all major spheres of our lives today. It makes a computer study, learn from the

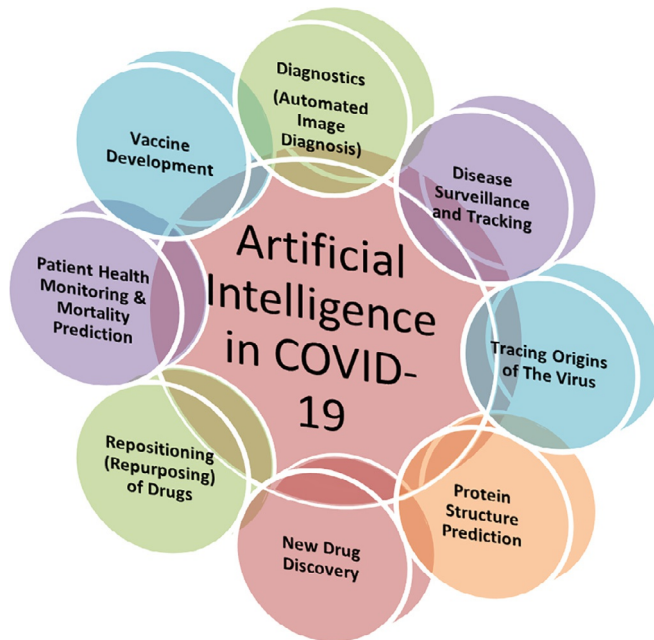


**Fig. 1** Global data on COVID-19 cases available from the World Health Organization. As of August 18, 2020, 21,732,472 cases of COVID-19 have been reported worldwide, including 770,866 fatalities.

existing dataset and their patterns, “think” and apply the acquired knowledge to make certain predictions. Today AI has profoundly advanced the field of healthcare through the introduction of a greater degree of automation, convenience and higher accuracy and precision in diagnostic and therapeutic care (Davenport & Kalakota, 2019). With the introduction of robotic technologies, medical personnel are being relieved of usual routine tasks, thus allowing them to concentrate more on more crucial aspects of patient care, but the robots are actually assisting surgeons during surgical procedures. In some cases, computer algorithms are actually outmatching the capabilities of radiologists in the detection of malignancy in tumors. Automated disinfection of patient rooms, collection and transport of blood samples, patient sample storage, and analysis are a few areas where AI makes its presence felt by reducing the risks posed to patients and healthcare personnel via human intervention. Industry 4.0, often referred to as the 4th industrial revolution, represents the heralding of the age of digitization led by various types of disruptive technologies, encompassing greater automation through cyber-physical systems (enabling autonomous information exchange), Industrial Internet of Things (IoT), Internet of Medical Things (IoMT), intelligent networking of machines, information and communication technologies (ICT), Drone technology and smart machine-driven decision-making processes (Alcácerac & Cruz-Machado, 2019; Oztemel & Gursev, 2020).

As the world grapples with the enormous spread of this disease coupled with virtually nonexistent remedial measures insight, artificial intelligence-based tools are being employed for tackling the problem at various levels (Fig. 2) and such AI-based applications in the COVID-19 pandemic have been widely reviewed (Kumar, Gupta, & Srivastava, 2020; Lin & Hou, 2020; Naudé, 2020; Nguyen, Waurn, & Campus, 2020; Ting, Carin, Dzau, & Wong, 2020; Vaishya, Javaid, Khan, & Haleem, 2002; Vashisht et al., 2020).

Deep learning tools are being employed to aid in accurate diagnosis by helping analyze computational tomography (CT) scans, X-ray images predicting the advancement of the infection (Jin et al., 2020; Li, Zhang, Goncalves, et al., 2020; Narin, Kaya, & Pamuk, 2021; Wang, Kang, & Ma, 2020; Wang, Zha, Li, et al., 2020) and (or) mortality based on the patient's clinical data. For example, Dhiman et al. have reported a multiobjective optimization- and deep learning-based technique for detecting COVID-19 infection in patients employing chest X-ray images (radiographs) (Dhiman, Chang, Singh, & Shankar, 2021). COVID-19 is known to cause pulmonary parenchymal ground-glass opacification in the lungs



**Fig. 2** Summary of potential applications of artificial intelligence in the management of COVID-19.

with subsequent consolidation on the advancement of the disease. In this study, eleven different types of CNN (convolutional neural network) based models have been developed, cross-validated and tuned using MOSHO (multiobjective spotted hyena optimizer). The reported models have generated good performance metrics, thus showing their usefulness in the real-time classification of COVID-19 disease. In another study, Abdel-Basset et al. have proposed a hybrid approach based on the thresholding technique in order to address the image segmentation problem for chest X-ray images (Abdel-Basset, Chang, & Mohamed, 2020). In this work, the integration of the slime mold algorithm (SMA) with the whale optimization algorithm has yielded results superior to several other algorithms.

Wang et al. have based their development of a fully automatic deep learning system for diagnosis and prognosis of COVID-19 on computed tomography images from a selected group of COVID-19 patients (Wang, Zha, et al., 2020). A semi-supervised few-shot segmentation (FSS) approach has been applied to carry out accurate and efficient segmentation of SARS-CoV2 infection (FSS-2019-nCov) with potential applications in limited data situations on a small dataset of annotated lung CT scans (Abdel-Basset, Chang, Hawash, Chakraborty, & Ryan, 2021). Further, the deep learning system could not only distinguish COVID-19 from other types of pneumonias but was also able to segregate the patients into high- and low-risk propensity. AI-based methods are being employed for tracking COVID-19 spread over time and place (Chen, Lu, Chang, & Liu, 2020; *Metabiota Epidemic Tracker*, 2020; Rao & Vazquez, 2020; Tuli, Tuli, Tuli, & Gill, 2020). AI is further being used for disease surveillance, to scan public spaces for people with potential COVID-19 infection, and to enforce various social distancing measures or imposition of lockdown (Gurucharan, 2020; Maslan, 2020). Researchers have employed a myriad of disruptive technologies such as AI, IoT, IoMT, virtual reality, big data, blockchain, 5G, robots, etc. to devise an intelligent framework for detection and prevention of the spread of COVID-19 disease (Abdel-Basset, Chang, & Nabeeh, 2021). The artificial intelligence-enabled mobile application called Aarogya Setu launched by the Government of India, is a good example of which helps the registered users to check their safety status based on whether they have crossed paths with Covid-19 positive patients. Thermal cameras possessing AI-based multisensory technology are being deployed in hospitals, airports, and public places to overcome the disadvantages of manually operated cameras. Artificial intelligence is also playing a very important role in tracing the origin of the novel coronavirus. Nguyen et al. have used

artificial intelligence (AI) to carry out an investigation of the origin of the COVID-19 virus (Nguyen et al., 2020). The researchers have used raw genomic sequences of the virus, exceeding three hundred in number, collected from infected cases across different countries and analyzed by unsupervised clustering methods. Their results support the scientific hypotheses postulating bats and pangolins as the probable hosts for the SARS-CoV2 virus. Another similar study to track the virus origin has utilized 334 complete genomic sequences of the virus taken across the world, which have been lodged in the GenBank database. This analysis has been carried out based on unsupervised clustering methods, employing a hierarchic clustering algorithm twinned with DBSCAN (density-based spatial clustering of applications with noise), which has progressively narrowed down the search parameters, progressing from higher to lower levels of taxonomic classification. This study has concluded that SARS-CoV-2 belongs to the Riboviria realm or cluster (earlier represented by the coronavirus responsible for the Middle East respiratory syndrome or MERS) and progressive narrowing of cut-off parameters showed SARS-CoV-2 clustering with only two viruses. These included the bat CoV RaTG13 virus and the Guangdong pangolin CoV. Further pinning down of search suggested that SARS-CoV-2 virus groups exclusively with the bat. CoV-RaTG13 suggested that bats could be the most probable reservoir host of the novel coronavirus. The important role of AI in COVID-19 has been acknowledged by the WHO, creating a new technology access pool for the pandemic (The Pharma Letter, 2020).

### **3 Artificial intelligence in drug discovery**

The most important application of AI in combating the SARS-CoV-2 virus is the development of potential treatment approaches in the arena of drug discovery and the development of vaccines. Considering that COVID-19 is a major health emergency, it is imperative that effective treatment measures be discovered in the shortest possible time frame. Recent times have seen a spurt in drug discovery institutes/ pharmaceutical industry collaborations with artificial intelligence platforms. For example, Scripps Research, with its drug repurposing collection ReFRAME has recently announced a collaboration with Repurpose.AI ActivPred artificial intelligence (AI) drug discovery platform to discover drug candidates effective against the novel coronavirus (The Science Advisory Board, 2020). The company, French AI startup Iktos, which is specifically engaged in applying artificial

intelligence approaches for identifying novel drug candidates, has recently announced its collaboration agreement with a research institute in California, SRI International, to accelerate the development of novel antivirals against various types of viruses, including the influenza virus and the novel coronavirus (COVID-19).

#### 4 Chemical structure input for data processing

Several repositories are available from which information relating to chemical structures of various viral proteins as well as drug-target interaction may be retrieved, e.g., the Protein Data Bank (PDB), NCBI PubMed, and NCBI PubMed Central (PMC), PharmGKB database, Therapeutic Target Database (TTD), etc. and these can be used for docking analyses or repurposing strategies. Several small-molecule libraries are there, from which structures of active molecules or approved/investigational drugs may be obtained, e.g., ChEMBL, DrugBank, drugCentral, GHDDI (virus-specific dataset), etc. The chemical structures of drugs with SMILES (Simplified Molecular Input Line Entry System) format can be extracted from DrugBank (Weininger, Weininger, & Weininger, 1989). Multistage text processing and data analysis of abstracts and full papers included in public repositories have been used to identify common host-protein targets between HIV-1 and COVID-19 infections by feeding queries built by employing Python scripts. Full-text articles or abstracts can be retrieved in machine-readable XML format from several of these repositories. Tarasova, Ivanov, Filimonov, and Poroikov (2002) have employed the Lingpipe-4.1.2 algorithm fine-tuned on GENIA corpus (Kim, Ohta, Tateisi, & Tsujii, 2003) to extract out protein sets and to identify overlapping proteins. They have further saved species-specific proteins related to the viruses under study and the host proteins out of this complete dataset by prior checking on UniProt API. Hu, Jiang, & Yin, 2020 have trained their multitask deep model consisting of shared layers and task-specific layers by a virus-specific dataset to carry out drug prediction against the COVID-19 virus. The binding sites have been predicted in the selected viral proteins by exploring critical input sequences by applying a nonparametric method occlusion. In another study, standard protein interaction networks have been utilized for identifying repurposing opportunities based on interactions of SARS-CoV-2 proteins with other proteins and the interaction networks between molecules in the cell (Kumar, 2020).

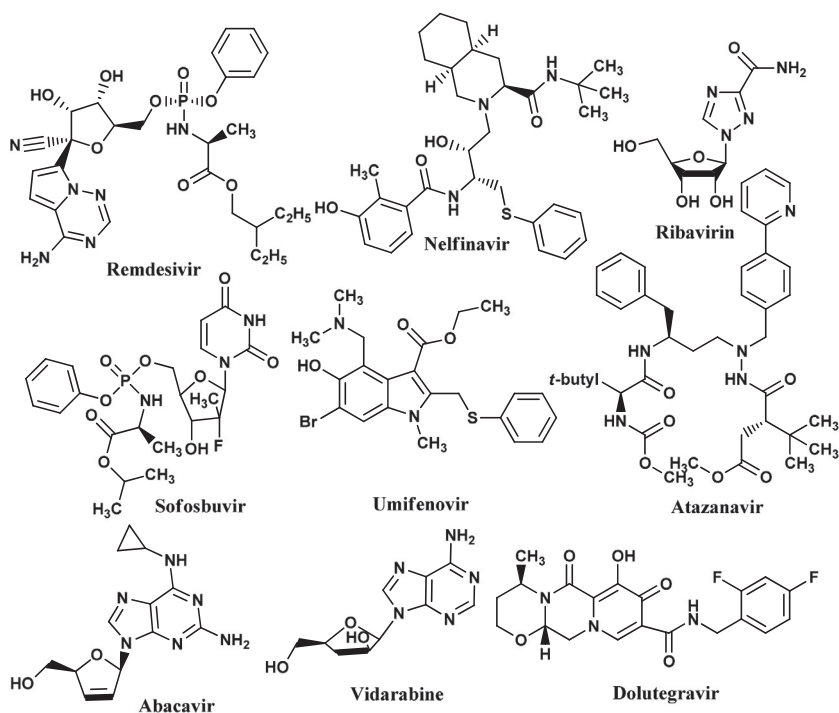
Generative deep learning approaches have been utilized for arriving at novel drug-like candidates (Zhavoronkov, Aladinskiy, Zhebrak, et al., 2020).

Active ligand complexed within the crystal structure of SARS-CoV main protease was extracted for carrying out ligand-based generation. A pocket algorithm has been used to annotate the binding site. Employing an X-ray structure of 4MDS from SARS-CoV M<sup>pro</sup> (co-crystallized with a noncovalently bound ligand) as a template, a homology model has been constructed for SARS-CoV-2 3C-like protease-ligand complex using SWISS-MODEL. Further, a dataset of protease-binding molecules has been constructed by extraction from repositories and filtering out highly nondrug-like molecules (such as radicals, metals, hydrazines, etc.) via the application of Mild medicinal chemistry filters (MCFs). Several ML models, including generative autoencoders, generative adversarial networks, genetic algorithms, etc. were used to create novel molecular structures followed by optimization through reinforcement learning (RL) employing the reward functions based on scores for medicinal chemistry & drug-likeness, active chemistry (utilizing self-organizing maps previously fine-tuned on protease peptidomimetics dataset), structure-fitting to receptor pockets, novelty (e.g., the penalty for similarity to existing molecules), etc. The low reward was given to molecules with structural alerts (Guex, Peitsch, & Schwede, 2009). In another study, generative and predictive models for de novo design of novel molecules against the SARS-CoV-2 3CL protease have been reported (Bung, Krishnan, Bulusu, & De Roy, 2020), based on a deep neural network. For preprocessing of data, Python (RDKit library) was used, and molecules were input in SMILES notation, followed by introducing chemical filters and removing SMILES strings longer than 100 symbols. The molecules in SMILE format were then used to train the generative model, which was then retrained with another dataset of protease inhibitors (ChEMBL) by transfer learning. Drug-target interaction information in various studies is generally used to build the drug-target network and this is generally based on binding affinity criteria like  $K_i$ ,  $K_d$ ,  $IC_{50}$ , or  $EC_{50}$  (should be low; in range of nM or at submicromolar concentrations). For example, Zhou, Hou, and Shen (2020) have employed the input criterion that these affinity values should not exceed 10  $\mu$ M. In another deep learning approach for virtual drug screening (Zhang, Saravanan, & Yang, 2020a), the workers retrieved the viral RNA sequences from Global Initiative on Sharing All Influenza Data (GISAID) database (Shu & McCauley, 2017), which were then translated to their respective amino acid sequences employing a web tool (Translate).



## 5 Artificial intelligence in repositioning approaches

Considering the large-scale and rapid spread of COVID-19, it is increasingly recognized that there is a need for immediate discovery of drug treatment before any vaccine is developed. Hence, the major focus targeted towards mitigating the virus has been on the repositioning approaches, i.e., identifying new clinical uses for existing drugs approved by the US Food and Drug Administration (US-FDA) (Fig. 3). Drug repurposing is recognized as an efficient strategy for drug discovery which can significantly reduce the time and costs in comparison to the de novo drug discovery approaches and randomized clinical trials (Cheng, Hong, Yang, & Wei, 2017; Cheng, Murray, & Rubin, 2016). Simple and computational drug repurposing has been previously utilized to identify plausible drug candidates with the potential to be effective against other viral infections such as dengue, Ebola, ZIKA and influenza (Johansen, Brannan, Delos, et al., 2013; Madrid, Chopra, Manger, et al., 2013; Mani, Wadhvani, & Krishnamurthy, 2019).



**Fig. 3** Antiviral drugs repurposed against SARS-CoV-2 infection through artificial intelligence approaches.

Experimental repurposing approaches are relatively time-demanding and costlier compared to the computational approaches (Santos et al., 2020) and hence, it may be followed after zeroing on viable drug candidates through the application of computational approaches. Computational repurposing has earlier shown promising results against infections by the SARS-CoV and MERS-CoV viruses (Dyall, Coleman, & Hart, 2014; Nukoolkarn, Lee, & Malaisree, 2008) and following the SARS-CoV2 outbreak, computational repurposing is being applied in search of novel treatment strategies. One of the more widely used drugs in COVID-19 treatment has been hydroxychloroquine, an antimalarial drug, though its efficacy and safety aspects have been debated. This drug was identified through AI-based and big data platform (Medicircle, 2020) through analysis of drug pathways and interactions to identify targets for human/virus proteins. Further, this AI-based platform has carried out the evaluation of the clinical potential of drugs previously approved for the treatment of other diseases for postulating potential drug combinations for COVID-19. Pertinent suggestions include a combination of Chloroquine with an antiarthritic drug Tocilizumab to counter cytokine release syndrome, a well-known complication of COVID-19 (Zhang, Wu, Li, Zhao, & Wang, 2020). Other potential drug combinations that have been suggested include, a combination of chloroquine with the antiviral drug remdesivir, and a combination of hydroxychloroquine with a semisynthetic macrolide antibiotic and antiretroviral drug clarithromycin (Medicircle, 2020).

The phylogenetic analysis of fifteen human coronaviruses (HCoV) genome sequences has revealed that the novel coronavirus (SARS-CoV2) collected from patients at the beginning of the outbreak has a maximum (79.7%) sequence identity with respect to the SARS-CoV (Jaimes, André, Chappie, Millet, & Whittake, 2020; Zhou, Yang, & Wang, 2020). Further, two evolutionarily conserved regions of SARS-CoV-2, i.e., the envelope and nucleocapsid proteins, possess much higher sequence identities with respect to SARS-CoV (96% and 89.6%, respectively) (Jaimes et al., 2020). Thus, drugs that have previously shown promise against SARS-CoV should have the potential to be effective against this novel coronavirus as well.

Recently, network-based repurposing of drugs employing network proximity analyses of the drug targets and interactions between Human CoV and the host has been employed to prioritize sixteen potential anti-HCoV drugs that could be repurposed COVID-19 (e.g., melatonin, mercaptopurine, and sirolimus) (Zhou, Hou, & Shen, 2020). The prioritized

drug molecules have been validated through enrichment analyses carried out with drug-gene signatures and the transcriptomics data in human cell lines induced by the human CoV. Other useful information has been obtained from the “Complementary Exposure” pattern, i.e., when targets of two drugs hit the sub-network between Human CoV and the host. However, they target separate neighborhoods in the network of the human interactome. This information has been utilized to identify three potential drug combinations for effectiveness in SARS-CoV2 infection, i.e., sirolimus and dactinomycin, mercaptopurine and melatonin, toremifene and emodin.

Advanced molecular simulations are being combined with AI-guided search for potentially active drug molecules to facilitate the identification of molecules that can be repositioned as prospective antiviral drugs (Koppal, 2020). Research by an AI drug discovery company BenevolentAI has utilized AI software to discover an approved rheumatoid arthritis drug, Baricitinib, having the capability to diminish the infectivity of the virus (Richardson, Corbellino, & Stebbing, 2020; Richardson, Griffin, & Tucker, 2020).

Recently, four small molecular drugs have been identified on the basis of their high binding affinity for the SARS-CoV main protease via high-throughput screening of eight thousand approved/experimental drugs and small molecules acquired from Drugbank. These include the fluoroquinolone antibiotic Prulifloxacin, anti-HIV-1 drugs Bictegravir and Nelfinavir, and the antiviral drug Tegobuvi (Li, Zhang, Wang, et al., 2020). Molecular similarity search was performed on the basis of the similarity in sequences of the structure-divulged molecules. Out of the resulting 690 compounds, only 50 were subjected to docking studies and the remaining compounds were discarded based on their toxic profiles, strong side effects, antitumor properties, etc. All these drugs have been widely employed in clinical practice and have been proposed as promising candidates for the treatment of the infection from 2019-novel coronavirus. Nelfinavir has also been predicted to be a prospective inhibitor of SARS-CoV 3C-linked main protease (3CL<sup>Pro</sup> or M<sup>Pro</sup>) by another integrative computational-based approach based on a combination of homology modeling, docking and binding free energy calculations (Xu, Peng, Shi, et al., 2020). In this report, the SARS-CoV main protease structures have been modeled, taking as a template the SARS homolog (PDB ID: 2GTB). Fifteen drugs were initially selected from docking score and three-dimensional (3D) similarity analyses and these were again docked with ten additional models of the protease enzyme. Six drugs have been chosen for further studies. These include Praziquantel,

Pitavastatin, Zopiclone, Nelfinavir, Eszopiclone and Perampanel. Binding free energy calculations on four of these drugs employing MM/GBSA and SIE approaches retrieved nelfinavir as the most promising candidate (Xu et al., 2020).

As mentioned previously, two inhibitors of HIV-1 protease enzyme, i.e., lopinavir and ritonavir had been identified as inhibitors of the SARS-CoV M<sup>Pro</sup> and these have also been found to be clinically effective against SARS-CoV2 (having 96% similarity to SARS-CoV) (Nukoolkarn et al., 2008). Further, the three-dimensional structure of the binding pocket where lopinavir/ritonavir interact with the enzyme is also conserved (the conserved amino acids being Threonine (Metabiota Epidemic Tracker, 2020)-Asparagine (Abdel-Basset, Chang, & Nabeeh, 2021) and Asparagine; numbering as per SARS-CoV protease) between the main protease of both SARS-CoV and SARS-CoV2. Amino acids Threonine (Metabiota Epidemic Tracker, 2020), Threonine (Gurucharan, 2020), and Asparagine<sup>119</sup> have been predicted to bind to the two drugs through hydrogen bonding. Based on these results, ten clinical drugs have been identified from the Drug-Bank database that can forge hydrogen bonds with conserved amino acid residues in the binding pocket of the enzyme SARS-CoV2 M<sup>Pro</sup> along with the possibility of higher tolerance to resistance mutations (Liu & Wang, 2020). There is a significant amount of data on molecular mechanisms of HIV-1 infections and several anti-HIV drugs have been found to inhibit the SARS-CoV-2 virus. Considering this relationship, text and data mining workflow have been used to identify as many as 46 targets. The study has shown that the two viruses share some common molecular pathways for the generation of their inflammatory, immune and cell cycle responses (Tarasova et al., 2002). Studies have evoked that the virus can enter the human cells by binding with ACE2 (angiotensin-converting enzyme 2), a cell-surface molecule (Lu & Sun, 2020; Zhao et al., 2020) and this is crucial for the human cell entry of the virus. The study by the BenevolentAI group had adapted its search to the available data on ACE2. In this regard, the enzyme AAK1 (adaptor-associated protein kinase 1), which is responsible for the regulation of endocytosis (entry of material into the cells) was identified as a possible disease target. Baricitinib emerged as the final drug out of 378 known AAK1 inhibitors included in the study, and the researchers have proposed a clinical investigation of the drug against the virus (Stebbing, Phelan, & Grifn, 2020). Accelerated clinical trials are proposed to be conducted on baricitinib to test its effectiveness in COVID-19 by the National Institute of Health, UK. This drug is presently being explored by several other

countries including Canada, Italy, and the US (Smartcity, 2020) and clinical studies also demonstrate its usefulness along with other repurposed drugs for COVID-19 (Sodani, Mucci, & Girolimetti, 2020).

Recently, another computational AI model has been reported for drug repurposing for mitigation of SARS-CoV-2 infection. For screening a database of more than a hundred approved antiviral drugs, the AI model has followed the approach of drug-target interaction prediction, i.e., it retrieved drugs effective against viruses having a genomic similarity with respect to the novel coronavirus. The model initially computes structural complementarity of the chemical structures of the drugs and the genomic structures of different viruses along with SARS-CoV-2 and then considers the diachronic data relating to the efficacy of the drugs against various types of viruses. All the drugs selected by the developed model (based on structural complementarity to viruses having genomic similarity to SARS-CoV-2) as prospective cures for COVID-19, i.e., sofosbuvir, remdesivir, umifenovir and ribavirin are presently undergoing clinical investigation against COVID-19 infection. Furthermore, AI could also offer a solution to the problem of mutation of the novel coronavirus as drugs effective against the original strain that surfaced in December 2019. These are considerably different from those effective against the more recent strains encountered from June 2020 onwards (The Indian Practitioner, 2020).

Deep learning-based models have been employed to carry out prediction of binding affinities based on SMILES chemical sequences of the target protein along with the sequences of amino acids (FASTA). Drugs like Atazanavir, Remdesivir, Kaletra (Lopinavir/ritonavir combination), Rapamycin and tiotropium bromide have been identified as potential inhibitors of the SARS-CoV2 virus. Amongst these drugs, remdesivir has recently been approved by US FDA for use in COVID-19 patients (Narayana et al., 2020). Beck, Shin, Choi, Park, and Kang (2020) have also predicted the antiretroviral drug atazanavir as the best candidate for inhibition of the SARS-CoV2 3C-like protease enzyme followed by efavirenz, ritonavir and dolutegravir.

Nguyen, Gao, and Chen (2020) have applied a Deep Learning approach based on mathematical techniques (MathDL) for the identification of drugs with inhibition potential against 3CL like protease enzyme. For this purpose, they have employed a homology model of 3CL like protease, prepared to take the crystal structure of SARS-CoV 3CL protease as the template. Training sets were prepared from eighty four inhibitors of SARS coronavirus protease enzyme taken from the ChEMBL database. Another general dataset containing more than fifteen thousand ligand-protein

complexes was taken from the PDBbind database. Based on these, deep learning models have been developed by combining algebraic topology with (i) deep convolutional neural networks (CNNs) and (ii) deep multitask CNNs. Evaluation of a set of 1465 FDA-approved drugs for their binding affinities has identified a list of fifteen prospective drug candidates included in the DrugBank dataset, the top-ranking candidate being Bortezomib followed by flurazepam and ponatinib in this order (Nguyen, Gao, & Chen, 2020).

In another similar study (Zhang, Saravanan, & Yang, 2020c), the RNA sequences of 2019-nCoV from 18 patients lodged in a public domain database (GISAID) were first retrieved, followed by a translation into corresponding protein sequences and multiple sequence alignment. This study indicated 3C-like protease as a key therapeutic target. Further, an indigenously built homology model of the 3CL protease was taken and a deep learning model was employed for performing large-scale virtual screening to identify protein-ligand interacting pairs. Drug screening has been carried out from four chemical compound databases using a deep learning model founded on the Dense Fully Convolutional Neural Network (DFCNN). The potential ligands identified from this study include the small molecule drugs like Meglumine (amino sugar), D-Sorbitol (sugar alcohol), Adenosine nucleoside, D-Mannitol (sugar), Sodium gluconate (chelating agent), Ganciclovir (antiviral drug) & Chlorobutanol (sedative) and peptide drugs, (comprised of lysine, isoleucine, and proline).

PolypharmDB is a recently developed repository of clinically evaluated molecules along with their multiple pharmacological profiles for identifying their repurposing opportunities. Deep learning engine, MatchMaker™, which has been trained on millions of reported drug-target interactions in humans, has been used along with PolypharmDB MatchMaker to perform a rapid investigation of human target proteins and various viral proteins for potential therapeutic significance in COVID-19 (Kurji, 2020). This study has resulted in a set of molecules with potential of interacting with COVID-19 at various targets.

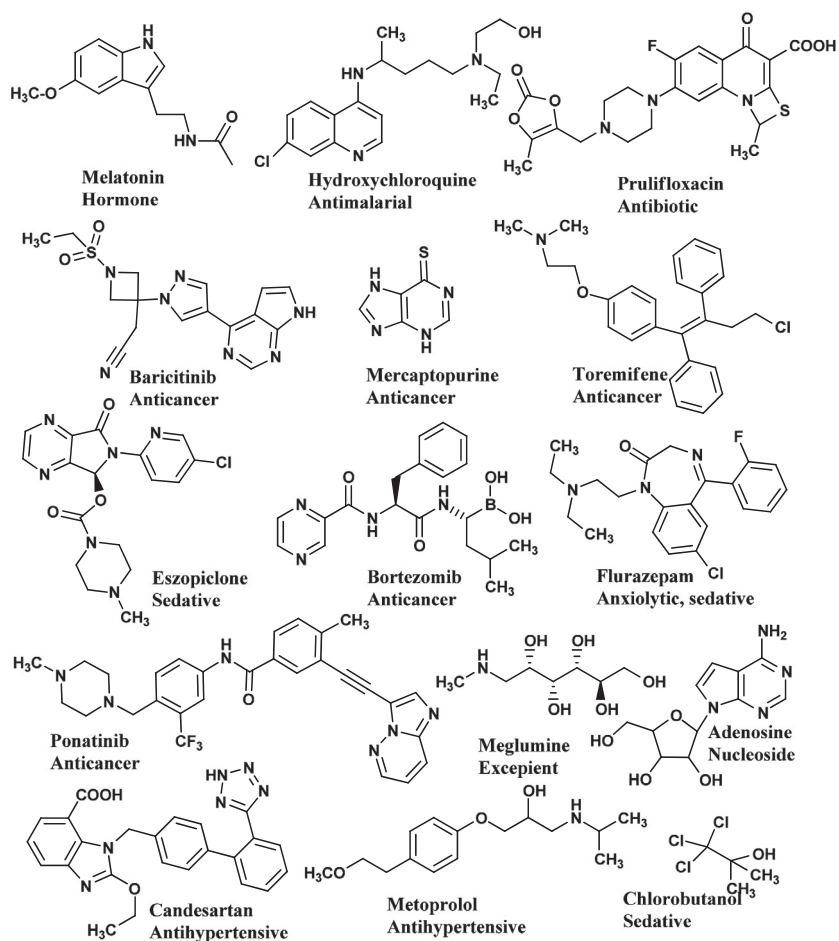
Hu, Jiang, & Yin, 2020 have also reported a prediction study. They first fine-tuned their pre-trained multitask deep model using a dataset of marketed drugs that have demonstrated in vitro inhibition of the novel coronavirus through in vitro studies. This re-trained model was then used to predict the affinities of a dataset of 4895 commercially available drugs against eight prospective protein targets present in SARS-CoV-2. Ten promising drug candidates have emerged from this study. Amongst these, abacavir, a

powerful reverse transcriptase inhibitor (anti-HIV), was predicted to possess a high affinity of binding towards several proteins of SARS-CoV-2, including RdRp, 3CL main protease, papain-like protease and helicase. Another anti-HIV drug, darunavir, a protease inhibitor, has also been predicted to target 3CL protease, RdRp and papain-like protease enzymes with nanomolar affinity. Other drugs include the antiasthmatic drug Fibroflapon sodium (predicted to have a potential affinity to Papain-like protease), daclatasvir used against Hepatitis-C virus predicted to inhibit RdRp, almitrine (respiratory stimulant) with a predicted affinity towards 3C-like proteinase and RdRp, Itraconazole with a predicted affinity towards 3C-like proteinase, roflumilast (antiinflammatory) with a predicted affinity towards Papain-like protease.

RdRp and metoprolol (antihypertensive) (Fig. 4) with a predicted affinity towards Papain-like protease (Hu, Jiang, & Yin, 2020).

A combination of network algorithms and human curation has been used to search integrated knowledge graphs and the information has been used for identifying drug repurposing opportunities against SARS-CoV-2. Eight potential repurposing opportunities have been identified in the form of drug classes (e.g., cathepsin L inhibitors; calmodulin inhibitors) and individual drugs like fostamatinib. In this study, a semantic knowledge graph was constructed in Neo4j, including the data of already existing, approved drugs (Skelton, Alsobhe, & Anastasi, 2020). Various data sources included for this purpose include UniProt (protein database), DrugBank (drugs), Monarch Disease Ontology (MONDO) (diseases/disorders), Drug Central (drugs), OMIM (human genes and genetic phenotypes), DisGeNET (database related to gene-disease associations). A literature search from the selected databases was used to identify existing know-how on genetic and proteomic mechanisms of pathogenesis in *Homo sapiens* and SARS-CoV-2 and this knowledge were used to devise semantic queries.

A high-throughput docking approach has been applied (Cavasotto & Filippa, 2020) employing a novel quantum mechanical grading for *silico* repositioning of drugs for COVID-19. In this study, they have screened a chemical library of more than eleven thousand compounds comprising of drugs approved by the US-FDA and molecules under clinical evaluation against three target proteins of the SARS-CoV-2. These proteins include the spike protein called the S-protein, the 3C-like proteinase, and the Papain-like protease. The docking library of 11,552 compounds has been constructed by coalescing molecular subsets obtained from 4 chemical compound libraries, i.e., (i) ChEMBL, which included molecules reaching at



**Fig. 4** Drugs not previously documented as antivirals repurposed against SARS-CoV-2 infection through artificial intelligence approaches.

least one of the phases of clinical trials; (ii) DrugBank, which included all approved as well as investigational drugs; (iii) DrugCentral database comprising of approved drugs and (iv) the library of FDA approved drugs from Selleck Chem. Potential inhibitors identified by this study include the antiviral drugs sofosbuvir, samatasvir, elbasvir and saquinavir and several other drugs with diverse indications like the  $AT_1$  receptor antagonist candesartan, antidiuretic drug desmopressin, thrombin inhibitor flovagatran and the anti-neoplastic drug brilacidin (Cavasotto & Filippo, 2020).



Another AI platform has been established to repurpose old drugs against COVID-19 by using two different learning databases; first, comprising compounds reported to have activity against SARS coronaviruses (SARS-CoV and SARS-CoV-2), HIV, influenza virus, and second comprising the known inhibitors of 3C-like main protease (Ke, Peng, Yeh, et al., 2020). activity against a feline strain of coronavirus and the results were used for relearning of the AI model. Eighty marketed drugs were predicted to have potential against the virus and amongst these, eight drugs, i.e., bedaquiline, celecoxib, gemcitabine, clofazimine, tolcapone, brequinar, conivaptan, and vismodegib, actually demonstrated in vitro activities against the virus responsible for feline infectious peritonitis (FIP). Further, boceprevir, chloroquine, homoharringtonine, tilorone, and salinomycin have also been predicted to be active through the developed AI model (Ke et al., 2020).

## 6 Artificial intelligence for accelerating computer modeling

Besides the repositioning approaches, which appear to be the most promising, efforts are also afoot to screen a large volume of small molecules for their potential to interact with the virus, particularly the viral proteins, to disrupt its activity. Computer-aided molecular design has been the most common approach in recent times for designing/screening molecules against macromolecular targets involved in diverse types of diseases. However, for an unknown virus-like SARS-CoV-2, where no particular lead molecule is known yet, one needs to screen billions of small molecules against more than twenty known proteins in the virus to assess their potential effectiveness in COVID-19 infection. The utilization of high-performance computational tools like supercomputers coupled with AI and machine-learning can considerably speed up this process by enabling the prediction of key viral proteins and quick screening of billions of molecules to arrive at pertinent drug candidates (Zhavoronkov, 2018). Four categories of structural proteins present in the SARS-CoV-2 virus are important, i.e., the spike proteins (S), nucleocapsid proteins (N), envelope proteins (E) and membrane proteins (M) (Lu, Zhao, Li, et al., 2020; Zhavoronkov et al., 2020). Further, some nonstructural proteins (NSPs) are there which are crucially involved in the viral pathogenesis. These include the C30 endopeptidase, more commonly referred to as 3-chymotrypsin-like protease (called 3C-like protease or 3CLPro or MPro (the main protease), or nsp5) & the papain-like protease (PLpro, component of nsp3). At the human level, research is being directed

towards the angiotensin-converting enzyme 2 (ACE2), a receptor known to facilitate the entry of the virus into the cells of the host. Several studies suggest that the SARS-CoV-2 utilizes this ACE2 receptor for gaining entry into the host cells and the Transmembrane protease serine 2 gene (TMPRSS2) is used for S-protein priming (Hoffmann, Kleine-Weber, Schroeder, et al., 2020).

Recently, a collaborative work has been conducted by integrating artificial intelligence and machine-learning tools, i.e., Deep-Learning-Guided Adaptive Molecular Simulations for Protein Folding (DeepDriveMD) with drug docking based on physics concepts and molecular dynamics simulation tools to fast-track the discovery of promising molecules against the virus (Dubrow, 2020). Millions simulations have been run as a part of this hybrid AI- and physics-based modeling-based project undertaken with some of the most powerful supercomputers like Frontera, Longhorn, Summit, Theta and Comet at multiple laboratories, which can simulate lakhs of molecules per hour. The results from these simulations have been used to train the machine learning system to identify crucial molecular parameters for activity. This can be utilized to screen other molecules for their potential activity. This work has yielded 30 molecules (out of a billion molecules taken in the initial stage of screening) with the potential to block a crucial enzyme involved in viral replication, adenosine diphosphate ribose 1" phosphatase (ADRP) (Dubrow, 2020).

Besides the repurposing approaches or screening of existing chemical databases for potential hits against salient viral proteins, generative chemistry approaches or de novo approaches are also being explored to encompass a virtually limitless number of chemical structures. Such deep generative models employ huge datasets for fine-tuning to carry out in silico designing of de novo chemical structures with required pharmacological or biochemical characteristics (Zhavoronkov et al., 2020). Recently, an HPC wire report has suggested the potential of quantum machine learning, a combination of machine learning and quantum information processing approaches for quicker and economical generation of novel complex compounds for COVID-19 (HPC Wire Staff Report, 2020). Huge investments are being made in this area by research institutes and researchers are employing machine learning methods to gain additional knowledge on "molecular aspects" of the virus and the mechanism behind the harmful effects of the virus on lungs (Rosenblatt, 2020). The study on the viral effect on lungs has been directed towards identifying a protein network interacting with the COVID-19 virus and enabling it to spread further to determine which

proteins could possibly be targeted to arrest the viral activity through drug treatments.

Currently, work is in progress to build an all-encompassing, all-atom model of the envelope of SARS-CoV-2 coronavirus, approximated to be containing more than two hundred million atoms using the Frontera super-computer of Texas Advanced Computing Centre ([The Science Advisory Board staff, 2020](#)). This work is on the lines of a previous work in which an all-atom simulation of the influenza virus envelope had been reported ([Durrant et al., 2020](#)). This information could be useful for the scientists working on the design of new drugs against SARS-CoV-2 and can also shed some light on the mechanism of interaction of current drugs and new potential drug combinations with the virus.

## 7 Artificial intelligence: De novo design of novel small molecules

Recent advancements in artificial intelligence (AI) & machine learning enable efficient mining of existing knowledge to explore an unlimited volume of chemical space. This can be used to develop novel small molecular agents or new chemical entities (NCEs) having required pharmacological and physicochemical properties ([Popova, Isayev, & Tropsha, 2018](#); [Stokes et al., 2020](#)). This approach has been applied recently to design novel small molecules to target SARS-CoV-2. One research group has employed generative and predictive models based on a deep neural network for carrying out de novo designing of novel chemical molecules for inhibiting the 3CL protease, an enzyme (biocatalyst) required for the growth of the virus ([Bung et al., 2020](#)). This generative model trained on a dataset consisting of around sixteen lakh molecules possessing drug-like properties (ChEMBL) was re-trained using another dataset consisting of 2515 molecules with inhibitory properties against protease enzymes by transfer learning ([Gaulton, Bellis, Bento, et al., 2012](#)). Further, the generative model was modulated by reinforcement learning to generate molecules having the desired properties. These small molecules generated through artificial intelligence were filtered on the basis of various physicochemical parameters and docked within the 3CL protease model (PDB ID: [6LU7](#)). This study has generated thirty-one potentially active candidate compounds to be synthesized and evaluated against SARS-CoV-2. Two of these small molecules have also shown high similarity with respect to Aurantiamide, an antiviral compound of herbal origin, obtained from *Baphicacanthus cusia*, an extensively used herb

in the Chinese system of medicine for the management of colds, flu and hyperthermia (Zhou, Yang, Feng, et al., 2017).

Another approach based on an advanced deep Q-learning network coupled with the fragment-based drug designing (ADQN-FBDD) has been employed in amalgamation with variational autoencoder (KL annealing & circular annealing) with the aim to generate lead molecules targeting the chymotrypsin-like protease (3CL main protease of SARS-CoV-2) (Verma & Bansal, 2020). This method employed SBOP (a structure-based optimization policy) in reinforcement learning. This work has generated three molecular structures which can potentially inhibit viral growth. Amongst these, one of the molecules bears functional similarity to the drugs remdesivir, umifenovir and elbasvir, which the US FDA approves for use in COVID-19 patients during emergency situations.

In silico Medicine has also reported utilizing its previously validated generative chemistry approaches to design novel drug-like inhibitors of COVID-19 with SARS-CoV-2 main protease (Mpro) as the target protein. These include crystal-derived pocked-based generators, generation based on homology modeling, and ligand-based approaches (Zhavoronkov et al., 2020).

Generation of potential lead compounds against SARS-CoV-2 3CL (PDB ID: 6LU7) has also been carried out by utilizing the ADQN-FBDD approach (Tang et al., 2020). This was followed by a structure-based optimization policy (SBOP) to obtain forty-seven top lead compounds, which have been assessed through molecular docking simulations.

Two novel molecule candidates have been generated by deep generative models by Innoplexus that can stop the growth of the virus by targeting a specific virus protein. Preliminary results from in silico predictions of ADME profile and docking studies have demonstrated an increase in efficacy over existing drugs and further testing is being carried out (Medicircle, 2020).

## 8 Artificial intelligence in protein structure prediction

Information about viral protein structures is crucial to the understanding of protein functions and mechanisms of their inhibition by drugs or vaccines, but the experimental determination of protein structures employing techniques like X-ray crystallography, solid-state NMR, etc. is a time-consuming process. Moreover, these techniques will not be applicable if the pure isolated form of protein is unavailable for analysis. Hence,

computational approaches are employed to predict protein structures either from sequences of component amino acids or through “template modeling” based on already existing structures of similar proteins.

Deep learning approaches of artificial intelligence have been utilized to carry out structure-based predictions of several proteins associated with SARS-CoV-2. AlphaFold system by Deep Mind (a division of Google Inc.) (Senior, Evans, & Jumper, 2020), is based on ResNet architecture (Yu & Koltun, 2016) and focuses on a template-free protein modeling approach, i.e., it can predict structure for those proteins which do not have any known related structure. It uses the sequences of component amino acids and MSA (alignment of multiple sequences) based on the features extracted from similar amino acid sequences to predict the spatial distance and angular alignment amongst the amino acid residues. This system has released structure predictions of six under-studied proteins associated with SARS-CoV-2, including the SARS-CoV-2 membrane protein, Nsp2, Nsp4, Nsp6, protein 3a, and the C-terminal domain of Papain-like protease (Deep Mind, 2020). Their system had also provided accurate structure prediction of the spike protein of the SARS-CoV-2, the experimentally determined structure of which is lodged in PDB (Protein Data Bank) (RCSB Id: 7BZ5).

The transform-restrained Rosetta (trRosetta) pipeline system also uses a dilated ResNet architecture for the prediction of the structures of the proteins mentioned above along with the ORF6, ORF7b, ORF8, and ORF10 proteins.

## 9 Artificial intelligence in vaccine development

Besides the very important role of artificial intelligence in designing drugs against COVID-19, attempts are now being directed towards the utilization of AI programs to facilitate the development of the COVID-19 vaccine. Virus-neutralizing antibodies are produced by the body via B-cells (humoral immunity) and T-cells (cellular immunity). A subtype of T-cells, termed memory cells are responsible for the recognition of an antigen of the eliminated pathogen, and any type of re-exposure to the pathogen quickly activates additional effector T-cells. All these processes are utilized for designing vaccines. The major histocompatibility complex proteins (MHC I and MHC II proteins) are the helper proteins that present the binding regions of the antigens, termed as epitopes, to the antibodies, B- or T-cells for

binding and their consequent neutralization. AI programs allow the identification of epitopes, i.e., the viral antigenic portions that have a greater probability of exposure on the infected cell surface and can attach to the antibodies. One example is that of an interactive AI platform, “Epitopes world” intended to make predictions for possible vaccine targets and thus reduce the time taken and the expense involved in creating vaccine candidates ([Epitopes World, 2020](#)). This platform has been designed on an algorithm called “CAMAP” that had been originally applied for cancer immunotherapy.

Magar et al. have devised a machine learning model to explore antibodies to neutralize the antigens ([Magar, Yadav, & Farimani, 2021](#)). A training dataset (VirusNet) was prepared consisting of more than nineteen hundred previously reported sequences of antigen-antibodies from several allied diseases, including HIV, influenza, SARS and Ebola. Graph featurization has been used with various machine learning methods for the screening of a large number of hypothetical antibody sequences ranging in thousands. Eight antibodies have been proposed with potential against COVID-19. The stability of the candidate antibodies has been finally assessed through a combination of bioinformatics, structural biology, and Molecular Dynamics (MD) simulations.

Fast et al. have utilized previously-developed neural networks for the prediction of major histocompatibility complex protein presentation, i.e., NetMHCpan4 and MARIA for identification of more than four hundred potential epitopes on the T-cells that can be presented by the major histocompatibility complex proteins (MHC I & MHC II), and two epitopes on the S-protein, targeting B-cells ([Fast, Altman, & Chen, 2020](#)). Further, 68 genomic variants of the SARS-CoV-2 have been examined to investigate the mutational characteristics of the virus in order to recognize which viral parts are less susceptible to phylogenesis. It has been observed that mutations in the case of the virus are more probable in those regions which have a strong antigenic presentation and the S-protein epitopes should make good vaccine targets as no nearby mutations have been found near the specific domain of the spike protein involved in receptor binding ([Fast et al., 2020](#)).

Reverse vaccinology has revolutionized the vaccine design approach in recent times. It focuses on identifying promising candidates for vaccine development through bioinformatics-based analysis of the genome of the pathogen. A recent study has employed reverse vaccinology and machine learning (ML) approaches for designing of vaccines against COVID-19 ([Ong, Wong, & Huffman, 2020](#)). In this study, Vaxin RV (a web-based

program) has been employed along with Vaxign-ML (Ong, Wang, & Wong, 2020), a machine learning tool developed by the researchers to enhance prediction accuracy, for predicting the viral proteins which can serve as effective targets for COVID-19 vaccine. In this study also, the spike protein (or S-protein) has been found to demonstrate the highest protective antigenicity score, thus, emerging as the best vaccine candidate. Additionally, five nonstructural proteins or NSPs, i.e., nsp3, 3CL-pro, and nsp8–10, have been identified as vaccine targets. The spike protein and the nonstructural proteins-3 and 8 (nsp3 & nsp8) have been predicted by Vaxign-ML to be the antigens with high protective antigenicity. The nsp3 protein has been more conserved amongst SARS-CoV-2, SARS-CoV, and MERS-CoV than amongst fifteen coronaviruses that can infect humans/animals; therefore, nsp3 protein has been selected as the second-best or alternative vaccine candidate to the S-protein.

Malone et al. have reported AI-enabled prediction of blueprints for designing universal vaccines for SARS-CoV-2 (Malone, Simovski, Moliné, et al., 2020). These are designed to include a reasonably extensive repertory of T-cell epitopes to ensure that the global population is broadly covered and protected. In this study, the presentation of the antigen to the host cell surface infected by the virus and predictions regarding immune response generation from *NEC Immune Profiler* has been used to profile the complete proteome of SARS-CoV-2 followed by the use of robustness Monte Carlo as well as digital twin simulations to identify a subset of epitope hotspots that have the maximum probability of being immunogenic amongst a broad spectrum of HLA allele types. These could be used for the formulation of the SARS-CoV-2 vaccine that can provide wide coverage of the population around the globe.

## 10 Summary

The novel coronavirus disease (COVID-19) has spread to more than 200 countries worldwide, thus assuming the proportions of an enormous global pandemic. Virtually, all spheres of life have been affected by the disease and hence, it is attracting the attention of researchers not only in pharmaceutical academia/industry/health profession but also in AI technology. AI has been utilized extensively for numerous extra-clinical applications in modeling viral transmission/spread of the disease, monitoring quarantine measures, and general data analysis. The health/pharmaceutical professionals are the frontline runners in this race as they are trying to contain this disease through

novel drug discovery, repositioning of existing drugs and vaccine development. Artificial intelligence has a major role to play not only in evolving all these clinical strategies for combating the disease but also for medical diagnostics based on radiologic imaging data. The most common application has been in the area of repurposing of existing drugs approved by the US-FDA for SARS-CoV-19 through various AI tools. Machine-learning models based on binding patterns of existing antiviral drugs with protein sequences from diverse virus species have been found to give fairly accurate predictions regarding the potential effectiveness of molecules against the proteomic sequence of the SARS-CoV-2 virus. Besides drug-treatment approaches, another powerful weapon against the virus is vaccines which normally take years to develop. Machine learning tools have been used to look for the immunogenic components of the virus, which could be exploited as good vaccine candidates. Thus, AI-driven modeling platforms are being used to accelerate drug design as well as vaccine design research for COVID-19.

## 11 Conclusions and future directions

Artificial intelligence has established itself as the foundation for research in all fields today and this includes its phenomenal contribution towards mitigating the challenges of the COVID-19 pandemic. This chapter has presented a review of AI applications in drug/vaccine development against the novel SARS-CoV-2 virus. AI has extensively contributed towards the understanding of the profile of the SARS-CoV-2-host protein interactions, which has been modeled on the lines of interactions of human cells with other viruses such as the Zika virus, Ebola virus, HIV virus, etc. Text- and data mining techniques have helped identify crucial proteins and molecular mechanisms that the SARS-CoV-2 virus has in common with previously known viruses. A huge list of molecular targets has built up from such studies in just a few months and is projected to uprise at an exponential rate as more studies are conducted. Prediction accuracy is critically dependent upon existing data and it should be understood that antiviral research is a particularly complex and ambiguous area of medicine. Even the mechanisms of action of existing drugs against various types of viruses are not unequivocal because the constantly evolving field of protein research will keep introducing newer protein targets with plausible newer mechanisms. The majority of the computational and biological studies have identified the M<sup>PRO</sup> as a crucial enzyme needed for the replication of coronavirus and having a huge potential for drug/vaccine targeting through de novo designing and repurposing.



By considering the timeline of novel drug development for establishing its efficacy and safety profile, a new molecule would take at least another 15–20 years before it reaches the patients. Hence, results from de novo design strategies don't seem attractive, at least for now, or till a suitable vaccine candidate arrives. Currently, repurposing appears to be a more attractive and time-effective approach due to the emergent requirement for an immediate cure for the disease. Several drugs predicted to be efficacious through AI or pharmacological screenings have been found to be useful in the clinical suppression of the disease. Repurposing strategies are generally based on homologies between molecular mechanisms or protein targets shared by the novel coronavirus and other viruses. Most of these studies may not be definitive in nature and could change in a future course. For example, the SARS-CoV-2 belongs to a different virus class from the HIV-1 virus and the homology between the two viruses is also disputed (Xiao et al., 2020). Inclusion of newer data will help design improvised AI models with predictions more accurately synced with clinical studies.

## References

- Abdel-Basset, M., Chang, V., Hawash, H., Chakraborty, R. K., & Ryan, M. (2021). FSS-2019-nCov: A deep learning architecture for semi-supervised few-shot segmentation of COVID-19 infection. *Knowledge-Based Systems*, 212, 106647.
- Abdel-Basset, M., Chang, V., & Mohamed, R. (2020). HSMA\_WOA: A hybrid novel slime mould algorithm with whale optimization algorithm for tackling the image segmentation problem of chest X-ray images. *Applied Soft Computing*, 95, 106642.
- Abdel-Basset, M., Chang, V., & Nabeeh, N. A. (2021). An intelligent framework using disruptive technologies for COVID-19 analysis. *Technological Forecasting and Social Change*, 163, 120431.
- Alcácerac, V., & Cruz-Machado, V. (2019). Scanning the industry 4.0: A literature review on technologies for manufacturing systems. *Engineering Science and Technology, An International Journal*, 22, 899–919.
- Beck, B. R., Shin, B., Choi, Y., Park, S., & Kang, K. (2020). Predicting commercially available antiviral drugs that may act on the novel coronavirus (2019-nCov), Wuhan, China through a drug-target interaction deep learning model. *bioRxiv*. <https://doi.org/10.1101/2020.01.31.929547>. 2020.01.31.929547.
- Bung, N., Krishnan, S. R., Bulusu, G., & De Roy, A. (2020). Novo design of new chemical entities (NCEs) for SARS-CoV-2 using artificial intelligence. *ChemRxiv*. <https://doi.org/10.26434/chemrxiv.11998347.v2>. Preprint.
- Cavasotto, C., & Filippo, J. D. (2020). *In silico* drug repurposing for COVID-19: Targeting SARS-CoV-2 proteins through docking and quantum mechanical scoring. *ChemRxiv*. <https://doi.org/10.26434/chemrxiv.12110199.v2>. Preprint.
- Chen, Y. C., Lu, P. E., Chang, C. S., & Liu, T.-H. (2020). A time-dependent SIR model for COVID-19 with undetectable infected persons. <http://arxiv.org/abs/2003.00122>.
- Cheng, F., Hong, H., Yang, S., & Wei, Y. (2017). Individualized network-based drug repositioning infrastructure for precision oncology in the panomics era. *Briefings in Bioinformatics*, 18, 682–697.

- Cheng, F., Murray, J. L., & Rubin, D. H. (2016). Drug repurposing: New treatments for zika virus infection? *Trends in Molecular Medicine*, 22, 919–921.
- Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, 6(2), 94–98.
- Deep Mind. (2020). *Computational predictions of protein structures associated with COVID-19*. <https://deepmind.com/research/open-source/computational-predictions-of-protein-structures-associated-with-COVID-19>. Accessed August 04 2020.
- Dhiman, G., Chang, V., Singh, K. K., & Shankar, A. (2021). ADOPT: Automatic deep learning and optimization-based approach for detection of novel coronavirus COVID-19 disease using X-ray images. *Journal of Biomolecular Structure & Dynamics*. <https://doi.org/10.1080/07391102.2021.1875049>.
- Dubrov, A. (2020). *AI fast-tracks drug discovery to fight COVID-19*. <https://www.tacc.utexas.edu/-/ai-fast-tracks-drug-discovery-to-fight-covid-19>.
- Durrant, J. D., Kochanek, S. E., Casalino, L., Jeong, P. U., Dommer, A. C., & Amaro, R. E. (2020). All-atom influenza virus simulations suggest new substrate binding mechanism. *ACS Central Science*, 6, 189–196.
- Dyall, J., Coleman, C. M., & Hart, B. J. (2014). Repurposing of clinically developed drugs for treatment of middle east respiratory syndrome coronavirus infection. *Antimicrobial Agents and Chemotherapy*, 58, 4885–4893.
- Epitopes World. (2020). *COVID-19 vaccine targets prediction*. <https://epitopes.world/en>.
- Fast, E., Altman, R. B., & Chen, B. (2020). Potential T-cell and B-cell epitopes of 2019-nCoV. *bioRxiv*. <https://doi.org/10.1101/2020.02.19.955484>. 2020.02.19.955484.
- Gaulton, A., Bellis, L. J., Bento, A. P., et al. (2012). ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Research*, 40, D1100–D1107.
- Guex, N., Peitsch, M. C., & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis*, 30, S162–S173.
- Gurucharan, M. K. (2020). *COVID-19: AI-enabled social distancing detector using OpenCV*. <https://towardsdatascience.com/covid-19-ai-enabled-social-distancing-detector-using-opencv-ca2abd827d34>.
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*, 181, 271–280.
- HPC Wire Staff Report. (2020). *Quantum machine learning: A new weapon against COVID-19?*. <https://www.hpcwire.com/2020/05/06/quantum-machine-learning-a-new-weapon-against-covid-19/>.
- Hu, F., Jiang, J., & Yin, P. (2020). *Prediction of potential commercially inhibitors against SARS-CoV-2 by multi-task deep model*. arXiv:2003.00728v2 [q-bio.MN]. Accessed August 26, 2021.
- Jaimes, J. A., André, N. M., Chappie, J. S., Millet, J. K., & Whittake, G. R. (2020). Phylogenetic analysis and structural modeling of SARS-CoV-2 spike protein reveals an evolutionary distinct and proteolytically sensitive activation loop. *Journal of Molecular Biology*, 432, 3309–3325.
- Jin, C., Chen, W., Cao, Y., Xu, Z., Zhang, X., & Deng, L. (2020). *Development and evaluation of an AI system for COVID-19 diagnosis* (p. 1e23). <https://doi.org/10.1101/2020.03.20.20039834>.
- Johansen, L. M., Brannan, J. M., Delos, S. E., et al. (2013). FDA-approved selective estrogen receptor modulators inhibit ebola virus infection. *Science Translational Medicine*, 5, 190ra179.
- Ke, Y., Peng, T., Yeh, T., et al. (2020). Artificial intelligence approach fighting COVID-19 with repurposing drugs. *Biomedical Journal*, 43, 355–362.

- Kim, J. D., Ohta, T., Tateisi, Y., & Tsujii, J. (2003). GENIA corpus—semantically annotated corpus for bio-textmining. *Bioinformatics*, *19*, 180–182.
- Koppal, T. (2020). *Artificial intelligence: Applications in the pandemic and beyond*. <https://www.biocompare.com/Editorial-Articles/565713-Artificial-Intelligence-Applications-in-The-Pandemic-and-Beyond/>.
- Kumar, S. (2020). *COVID-19: A drug repurposing and biomarker identification by using comprehensive gene-disease associations through protein-protein interaction network analysis*. Preprints 2020030440. <https://doi.org/10.20944/preprints202003.0440.v1>.
- Kumar, A., Gupta, P. K., & Srivastava, A. (2020). A review of modern technologies for tackling COVID-19 pandemic. *Diabetes and Metabolic Syndrome: Clinical Research & Reviews*, *14*, 569e573.
- Kurji, N. (2020). *China's institute of materia medica partners with cyclica on innovative drug repurposing for COVID-19*. <https://www.cio.com/article/3292616/how-singapore-is-using-artificial-intelligence.html>.
- Li, Y., Zhang, J. G., Goncalves, J., et al. (2020). An interpretable mortality prediction model for COVID-19 patients. *Nature Machine Intelligence*, *2*, 283–288.
- Li, Y., Zhang, J., Wang, N., et al. (2020). Therapeutic drugs targeting 2019-ncov main protease by high-throughput screening. *bioRxiv*. <https://doi.org/10.1101/2020.01.28.922922>. 2020.01.28.922922.
- Lin, L., & Hou, Z. (2020). Combat COVID-19 with artificial intelligence and big data. *Journal of Travel Medicine*, 1–4. <https://doi.org/10.1093/jtm/taaa080>.
- Liu, X., & Wang, X. (2020). Potential inhibitors against 2019-ncov coronavirus m protease from clinically approved medicines. *Journal of Genetics and Genomics*, *47*(2), 119–121.
- Lu, J., & Sun, P. D. (2020). High affinity binding of SARS-CoV-2 spike protein enhances ACE2 carboxypeptidase activity. *The Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.RA120.015303>.
- Lu, R., Zhao, X., Li, J., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *Lancet*, *395*, 565–574.
- Madrid, P. B., Chopra, S., Manger, I. D., et al. (2013). A systematic screen of FDA-approved drugs for inhibitors of biological threat agents. *PLoS One*, *8*, e60579.
- Magar, R., Yadav, P., & Farimani, A. B. (2021). Potential neutralizing antibodies discovered for novel corona virus using machine learning. *Scientific Reports*, *11*(1), 5261.
- Malone, B., Simovski, B., Moliné, C., et al. (2020). Artificial intelligence predicts the immunogenic landscape of SARS-Cov-2 leading to universal blueprints for vaccine designs. *Scientific Reports*, *10*, 22375.
- Mani, D., Wadhvani, A., & Krishnamurthy, P. T. (2019). Drug repurposing in antiviral research: A current scenario. *Journal of Young Pharmacists*, *11*, 117–121.
- Maslan, C. (2020). *Social distancing detection for COVID-19*. <https://blog.camio.com/social-distancing-detection-for-covid-19-a5c9e8d6c760>.
- Medicircle. (2020). *Innoplexus announces major findings in potential combination therapies for COVID-19*. <https://www.innoplexus.com/news/innoplexus-announces-major-findings-in-potential-combination-therapies-for-covid-19/>. Accessed August 04, 2020.
- Metabiota Epidemic Tracker. (2020). <https://www.epidemictracker.com>. Accessed August 04, 2020.
- Narayana, D., Arpna, D., Peeyushi, S., Samiksha, S., Sanjay, V., & Sneha, P. (2020). *Artificial intelligence as COVID-19 drug discovery booster—Express healthcare*. <https://mc.ai/artificial-intelligence-as-covid-19-drug-discovery-booster-express-healthcare/>.
- Narin, A., Kaya, C., & Pamuk, Z. (2021). Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. *Pattern Analysis and Applications*, *24*, 1207–1220.

- Naudé, W. (2020). *Artificial Intelligence vs COVID-19: Limitations, constraints and pitfalls*. <https://doi.org/10.1007/s00146-020-00978-0>.
- Nguyen, T. T., Abdelrazek, M., Nguyen, D. T., Aryal, S., Nguyen, D. T., & Khatami, A. (2020). Origin of novel coronavirus (COVID-19): A computational biology study using artificial intelligence. *bioRxiv*. <https://doi.org/10.1101/2020.05.12.091397>. 2020.05.12.091397.
- Nguyen, D. D., Gao, K., & Chen, J. (2020). Potentially highly potent drugs for 2019-nCoV. Preprint *bioRxiv*. <https://doi.org/10.1101/2020.02.05.936013>. 2020.02.05.936013.
- Nguyen, T. T., Waurn, G., & Campus, P. (2020). *Artificial intelligence in the battle against coronavirus (COVID-19): A survey and future research directions*. <https://doi.org/10.13140/R.G.2.2.36491.23846>.
- Nukoolkarn, V., Lee, V. S., & Malaisree, M. (2008). Molecular dynamic simulations analysis of ritonavir and lopinavir as SARS-CoV 3CL(pro) inhibitors. *Journal of Theoretical Biology*, 254, 861–867.
- Ong, E., Wang, H., & Wong, M. U. (2020). Vaxign-ML: Supervised machine learning reverse vaccinology model for improved prediction of bacterial protective antigens. *Bioinformatics*, 36(10), 3185–3191.
- Ong, E., Wong, M. U., & Huffman, A. (2020). COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. Preprint *bioRxiv*. <https://doi.org/10.1101/2020.03.20.000141>. 2020.03.20.000141. 2020.
- Oztemel, E., & Gursev, S. (2020). Literature review of industry 4.0 and related technologies. *Journal of Intelligent Manufacturing*, 31(4), 127–182.
- Popova, M., Isayev, O., & Tropsha, A. (2018). Deep reinforcement learning for de novo drug design. *Science Advances*, 4(7), eaap7885.
- Rao, A. S. R. S., & Vazquez, J. A. (2020). Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey in the populations when cities/towns are under quarantine. *Infection Control and Hospital Epidemiology*, 2020, 1–5. <https://doi.org/10.1017/ice.2020.61>.
- Richardson, P., Corbellino, M., & Stebbing, J. (2020). Baricitinib for COVID-19: A suitable treatment? authors' reply. *The Lancet Infectious Diseases*, 2(393), 52e65.
- Richardson, P., Griffin, I., & Tucker, C. (2020). Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *Lancet*, 395, E30–E31.
- Rosenblatt, L. (2020). *New research institute puts \$5 million toward AI to battle COVID-19 spread and its social impact*. <https://www.post-gazette.com/business/tech-news/2020/06/23/artificial-intelligence-COVID19>.
- Santos, J., Brierley, S., Gandhi, M. J., Cohen, M. A., Moschella, P. C., & Declan, A. B. L. (2020). Repurposing therapeutics for potential treatment of SARS-CoV-2: A review. *Viruses*, 12(7), 705.
- Senior, A. W., Evans, R., & Jumper, J. (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, 577, 706–710.
- Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data—From vision to reality. *Euro Surveillance*, 22(13), 30494.
- Skelton, D. J., Alsobhe, A., & Anastasi, E. (2020). *Drug repurposing prediction for COVID-19 using probabilistic networks and crowdsourced curation*. arXiv:2005.11088 [q-bio.MN].
- Smartcity. (2020). *Can AI help the world find the medicine to treat COVID-19?*. <https://www.smartcity.press/ai-accelerating-covid-vaccine/>.
- Sodani, P., Mucci, L., & Girolimetti, R. (2020). Successful recovery from COVID-19 pneumonia after receiving baricitinib, tocilizumab, and remdesivir. A case report: Review of treatments and clinical role of computed tomography analysis. Version 2. *Respiratory Medicine Case Reports*, 31, 101115.

- Sohrabi, C., Alsafi, Z., & O'Neill, N. (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery*, *76*, 71–76.
- Stebbing, J., Phelan, A., & Grifn, I. (2020). COVID-19: Combining antiviral and anti-inflammatory treatments. *Lancet*, *20*, 400–401.
- Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., & Donghia, N. M. (2020). A deep learning approach to antibiotic discovery. *Cell*, *180*, 688–702.
- Tang, B., He, F., Liu, D., Fang, M., Wu, Z., & Xu, D. (2020). AI-aided design of novel targeted covalent inhibitors against SARS-CoV-2. *BioRxiv*. <https://doi.org/10.1101/2020.03.03.972133>. Preprint.
- Tarasova, O., Ivanov, S., Filimonov, D. A., & Poroikov, V. (2002). Data and text mining help identify key proteins involved in the molecular mechanisms shared by SARS-CoV-2 and HIV-1. *Molecules*, *25*, 2944–2958.
- The Indian Practitioner. (2020). <https://theindianpractitioner.com/2020/07/08/iiit-delhi-develops-artificial-intelligence-model-to-repurpose-existing-drugs-for-treating-covid-19/>. Accessed August 02, 2020.
- The Pharma Letter. (2020). <https://www.thepharmaletter.com/article/who-launches-new-technology-access-initiative-for-covid-19>. Accessed August 04, 2020.
- The Science Advisory Board. (2020). *Scripps research expands drug repurposing efforts for COVID-19*. <https://www.scienceboard.net/index.aspx?sec=sup&sub=cell&pag=dis&ItemID=665>.
- The Science Advisory Board staff. (2020). <https://www.scienceboard.net/index.aspx?sec=sup&sub=BioChem&pag=dis&ItemID=587>. Accessed September 09, 2020.
- Ting, D. Z. W., Carin, L., Dzau, V., & Wong, T. Y. (2020). Digital technology and COVID-19. *Nature Medicine*, *26*, 458–464.
- Tuli, S., Tuli, S., Tuli, R., & Gill, S. S. (2020). Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. *Internet of Things*, *11*, 100222.
- Vaishya, R., Javaid, M., Khan, I. H., & Haleem, A. (2002). Artificial intelligence (AI) applications for COVID-19 pandemic. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, *14*, 337e339.
- Vashisht, A., Bagaria, A., Kumari, S., Singh, O. K., Gupta, A., & Bhagat, N. (2020). Artificial intelligence & Covid- 19—A review. *International Journal of Research in Health and Allied Sciences*, *6*(3), 14–17.
- Verma, M., & Bansal, D. (2020). Novel potential inhibitors against SARS-Cov-2 using artificial intelligence. *ChemRxiv*. <https://doi.org/10.26434/chemrxiv.12228362.v4>. Preprint.
- Wang, S., Kang, B., Ma, J., et al. (2020). *A deep learning algorithm using CT images to screen for corona virus disease (COVID-19)*. 1e28, medRxiv 2020.02.14.20023028.
- Wang, S., Zha, Y., Li, W., et al. (2020). A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *The European Respiratory Journal*. <https://doi.org/10.1183/13993003.00775-2020>. In press.
- Weininger, D., Weininger, A., & Weininger, J. L. (1989). Smiles. 2. Algorithm for generation of unique smiles notation. *Journal of Chemical Information and Computer Sciences*, *29*, 97–101.
- Worldometer. (2020). [https://www.worldometers.info/coronavirus/?utm\\_campaign=homeAdvegas1](https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1). Accessed August 18, 2020.
- Xiao, C., Li, X., Liu, S., Sang, Y., Gao, S., & Gao, F. (2020). HIV-1 did not contribute to the 2019-nCoV genome. *Emerging Microbes & Infections*, *9*(1), 378–381.
- Xu, Z., Peng, C., Shi, Y., et al. (2020). Nelfinavir was predicted to be a potential inhibitor of 2019-ncov main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation. *bioRxiv*. <https://doi.org/10.1101/2020.01.27.921627>.

- Zhang, H., Saravanan, K. M., & Yang, Y. (2020a). Deep learning based drug screening for novel coronavirus 2019-nCov. *Interdisciplinary Sciences, Computational Life Sciences*. <https://doi.org/10.1007/s12539-020-00376-6>.
- Zhang, H., Saravanan, K. M., & Yang, Y. (2020c). Deep learning based drug screening for novel coronavirus 2019-nCov. *Interdisciplinary Sciences, Computational Life Sciences*. <https://doi.org/10.1007/s12539-020-00376-6>.
- Yu, F., & Koltun, V. (2016). Multi-scale context aggregation by dilated convolutions. *International Conference on Learning Representations*. arXiv:1511.07122v3 [cs.CV].
- Zhang, C., Wu, Z., Li, J., Zhao, H., & Wang, G. (2020). Cytokine release syndrome in severe COVID-19: Interleukin-6 receptor antagonist tocilizumab may be the key to reduce mortality. *International Journal of Antimicrobial Agents*, 55(5), 105954.
- Zhao, Y., Zhao, Z., Wang, Y., Zhou, Y., Ma, Y., & Zuo, W. (2020). Single-cell RNA expression profiling of ACE2, the receptor of SARS-Cov-2. *bioRxiv*. <https://doi.org/10.1101/2020.01.26.919985>. 2020.01.26.919985.
- Zhavoronkov, A. (2018). Artificial intelligence for drug discovery, biomarker development, and generation of novel chemistry. *Molecular Pharmaceutics*, 15(10), 4311–4313.
- Zhavoronkov, A., Aladinskiy, V., Zhebrak, A., et al. (2020). Potential COVID-2019 3C-like protease inhibitors designed using generative deep learning approaches. *ChemRxiv*. <https://doi.org/10.26434/chemrxiv.11829102.v2>. Preprint.
- Zhou, Y., Hou, Y., & Shen, J. (2020). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discovery*, 6(14), 2020.
- Zhou, B., Yang, Z., Feng, Q., et al. (2017). Aurantiamide acetate from baphicacanthus cusia root exhibits anti-inflammatory and antiviral effects via inhibition of the NF- $\kappa$ B signaling pathway in influenza A virus-infected cells. *Journal of Ethnopharmacology*, 199, 60–67.
- Zhou, P., Yang, X., & Wang, X. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579, 270–273.