Contents lists available at ScienceDirect

Current Research in Pharmacology and Drug Discovery

journal homepage: www.journals.elsevier.com/current-research-in-pharmacology-and-drug-discovery

Artificial intelligence-driven drug repurposing and structural biology for SARS-CoV-2

Kartikay Prasad, Vijay Kumar [*]

Amity Institute of Neuropsychology & Neurosciences, Amity University, Noida, UP, 201303, India

A B S T R A C T

It has been said that COVID-19 is a generational challenge in many ways. But, at the same time, it becomes a catalyst for collective action, innovation, and discovery. Realizing the full potential of artificial intelligence (AI) for structure determination of unknown proteins and drug discovery are some of these innovations. Potential applications of AI include predicting the structure of the infectious proteins, identifying drugs that may be effective in targeting these proteins, and proposing new chemical compounds for further testing as potential drugs. AI and machine learning (ML) allow for rapid drug development including repurposing existing drugs. Algorithms were used to search for novel or approved antiviral drugs capable of inhibiting SARS-CoV-2. This paper presents a survey of AI and ML methods being used in various biochemistry of SARS-CoV-2, from structure to drug development, in the fight against the deadly COVID-19 pandemic. It is envisioned that this study will provide AI/ML researchers and the wider community an overview of the current status of AI applications particularly in structural biology, drug repurposing, and development, and motivate researchers in harnessing AI potentials in the fight against COVID-19.

1. Introduction

The novel coronavirus disease (COVID-19) has become an unprecedented public health crisis affecting people's lives and causing a large number of deaths. Till June 2021, over 178 million confirmed cases was reported worldwide with more than 3.88 million deaths reported (https://covid19.who.int/). The numbers of infections and death are still increasing. With the continued growth of the COVID-19 pandemic, scientists and healthcare providers worldwide are working to better comprehend, alleviate, and suppress its spread. The usual symptoms of COVID-19 are pneumonia, shortness of breath, dry cough, tiredness, and fever (Huang et al., 2020a) along with several neurological complications (Khatoon et al., 2020). SARS-CoV-2 is a positive-sense single-stranded RNA virus consisting of ~30 kb genome encoding for four main structural proteins including spike (S) glycoprotein, small envelope (E) glycoprotein, membrane (M) glycoprotein, and nucleocapsid (N) protein, in addition to the sixteen non-structural proteins (NSPs) (Wu et al., 2020; Lu et al., 2020). These viral proteins have a specific role in the life cycle and pathogenicity of the virus and a complete understanding of their structure and function is essential for drug discovery.

The power of artificial intelligence (AI) approaches has been attributed to a wide range of applications across public health, disease prediction, and drug development. Fuchs in 2020 (Fuchs et al., 2020) elegantly summarized the role of AI in the current COVID-19 pandemic in six major areas, including early predictions and alerts, tracking, data dashboards, diagnosis and prognosis, treatments, and social control. Over the past decade, AI-based models have revolutionized drug discovery in general (Fleming, 2018; Lavecchia, 2019). Machine learning (ML), a subset of AI, has enabled the generation of models that can learn and study the patterns present in data and can make inferences from a large number of test data. With the advent of deep learning (DL), the automatic feature extraction from raw data leads to an increase in performance compared to other computer-aided models (Chen et al., 2018; Zhavoronkov et al., 2019). Different DL algorithms were utilized in fighting the COVID-19 pandemic including artificial neural network (ANN), convolutional neural network (CNN), and long short-term memory (LSTM).

The recent applications of AI in the case of COVID-19 include the virtual screening of both repurposed drugs as well as new chemical entities (Fig. 1) (Keshavarzi ArshadiWebb et al., 2020; Zhou et al., 2020a; Mottaqi et al., 2021; Piccialli et al., 2021). ML-based molecular docking has been utilized extensively for virtual screening and drug repurposing. This approach requires the following information: (a) dataset of
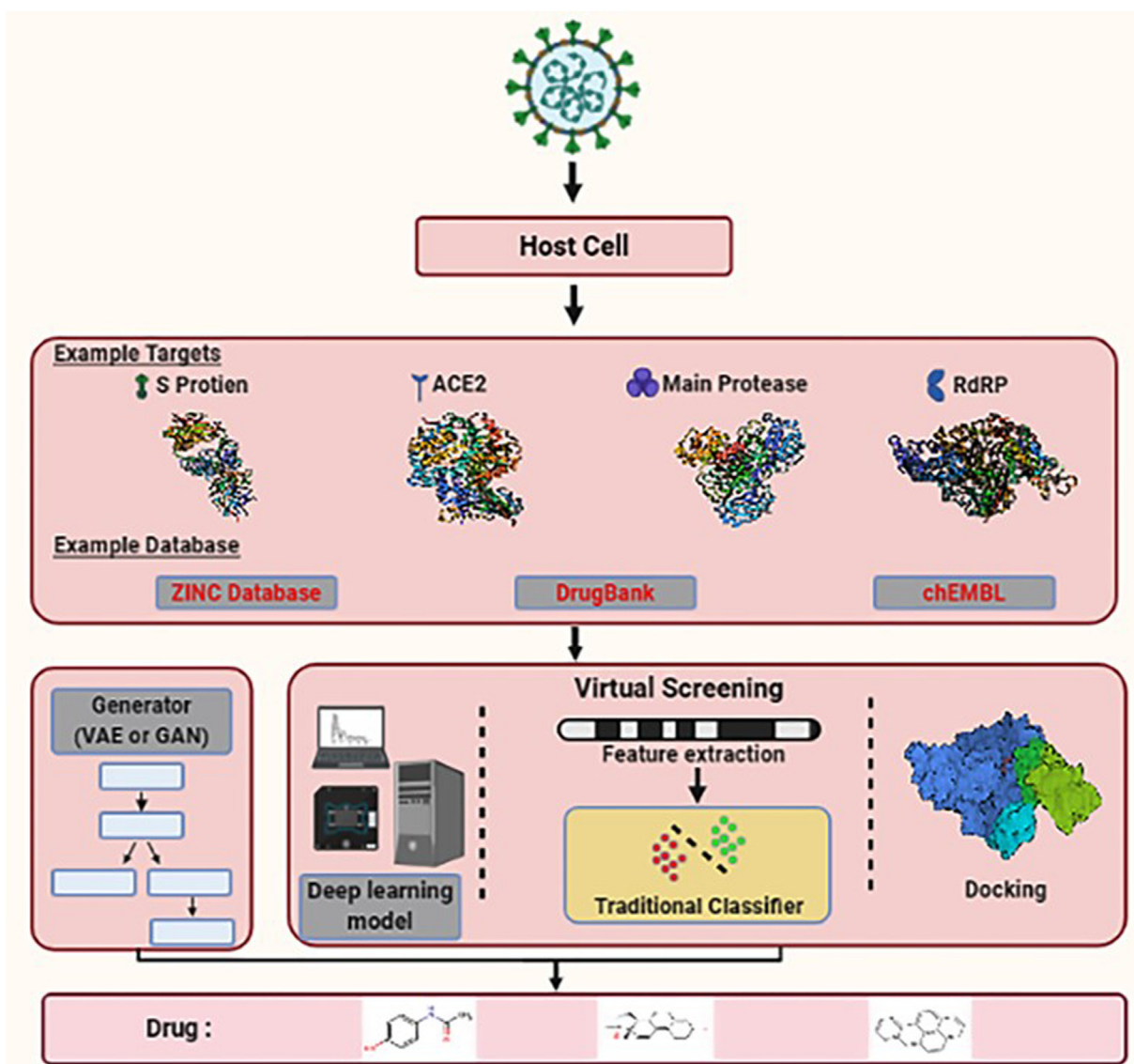

Fig. 1. The pipeline of AI and ML-based platforms for drug discovery and drug repurposing in COVID-19. The structural proteins of the SARS-CoV-2 were targeted through the drugs/small molecules present in different databases for drug repurposing and drug development. Drug repurposing can be feasible through virtual screening and structure-based molecular docking utilizing ML/DL approaches. De novo drug designing or drug discovery against the selected targets can be achieved through various generative approaches including Generative Adversarial Networks (GAN) and variational autoencoders (VAE).

approved drugs or drug-like molecules, (b) three-dimensional structure of the protein target, and (c) molecular docking software or program. The molecular docking-based studies allow the identification of several chemical molecules that binds to different SARS-CoV-2 proteins and thus can potentially inhibit viral replication and growth (Fig. 1).

In addition, DL-based applications have been the main focus for drug repurposing research, as their automatic feature extraction accelerates the process of drug discovery (Fig. 1). The application of DL has benefitted the *de novo* drug design approaches. The current design approach utilizes state-of-the-art DL models such as Generative Autoencoders (GAE) and Generative Adversarial Networks (GAN) to generate data-based molecules (Wrapp et al., 2020). More recent approaches use generative models such as variational autoencoders (VAE) to generate sequences of atoms (Fig. 1). This approach allows the creation of unique drug molecules with greater diversity (Griffiths and Hernandez-Lobato, 2020). These autoencoders instruct molecules into a vector that captures properties such as bond order, element properties, and functional group (Zhang and Lu, 2019; Baucum et al., 2020). To design a drug-generative network, the information about (a) collection of

drug-like molecules, (b) a feature-representation of these molecules in silico, (c) method to increase the diversity of the molecules, and (d) screening and modification of the altered molecules, are important (Fig. 1).

AI, in the field of computational biology and medicine, has been used to partially understand COVID-19 to discover drugs against the SARS-CoV-2 virus (Alimadadi et al., 2020; Mei et al., 2020; Ke et al., 2020; Nguyen, 2020). Equipped with a strong computational power to deal with large amounts of data, AI can help scientists to expand their knowledge about the coronavirus quickly. For example, by determining the protein structures of the virus, researchers would be able to find the machinery necessary for designing a drug or vaccine more accurately and effectively. The application of AI and its subsets are crucial for the current pandemic situation and rapid discovery of drugs against COVID-19 for several key reasons. The automatic feature extraction ability of DL can support models with better accuracy and reliable results. Also, the generative ability demonstrated by DL models can be largely exploited to create small molecule drugs and better epitope prediction, minimizing the chances of failure in trial experiments. Thus, the use of AI is essential

to find out potential drugs against COVID-19 in time and accurately.

In this review, we provide a survey of AI-based models for COVID-19 drug discovery and structure biology of SARS-CoV-2 proteins, that will provide rapid and cost-effective therapeutic interventions in COVID-19. We propose that these AI-based technological advances are in utmost need during the COVID-19 global pandemic.

2. Methodology

The systematic literature search and analysis was done through the Dimensions database (https, 2020) in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Liberati et al., 2009). Dimensions is the most comprehensive database developed by digital science with the help of over 100 leading research organizations, worldwide. This database provides access to various contents ranging from publications to grants, funding agencies, clinical trials, patents, datasets, and policy documents.

The literatures were retrieved from the dimensions database using search keywords with "OR/AND" operators. Keywords include "COVID-19", "SARS-CoV-2", "drug repurposing", "drug repositioning", "artificial intelligence", "machine learning", "deep learning", and "neural networks". The study period was restricted from January 1, 2020, to March 31, 2021. A staged literature search was performed for the first section of this systematic review and all the relevant studies were identified based on the set of inclusion/exclusion criteria and summarized narratively. Studies were included if they reported outcomes on any results on drug repurposing or drug discovery to COVID-19 utilizing AI/ML approaches. Whereas, the editorials, commentary, survey, and narrative reviews on AI/ML were excluded. The article selection for this review followed the three-stage analysis. The first stage analysis considered only the titles and abstracts of the articles to extract relevant articles. The second stage analysis enables the selection of articles based on the introduction and conclusion to further refine the selection in the first stage. At the third and final stage of analysis, articles were explored thoroughly and selected in terms of their relevance to the review aim. The article was selected if it reported an empirical application of AI/ML in drug repurposing or drug discovery to COVID-19.

3. Results

A total of 8841 articles were identified in the Dimensions database with the search term: "COVID-19" OR "SARS-CoV-2" AND (machine learning OR artificial intelligence OR deep learning OR neural network). After adding "drug repurposing" in the search term i.e., "COVID-19″ OR "SARS-CoV-2″ AND (machine learning OR artificial intelligence OR deep learning OR neural network) AND (drug repurposing), a total of 516 articles were found (Fig. 2). Out of these 516 articles, most of the articles were posted on an open-access online pre-print server and the publications cover mainly the field of medical and health sciences and services. In the initial analysis where the titles and abstracts were considered, 427 were excluded because they were not related to the drug repurposing work using the AI. The full text of the remaining 89 studies was then
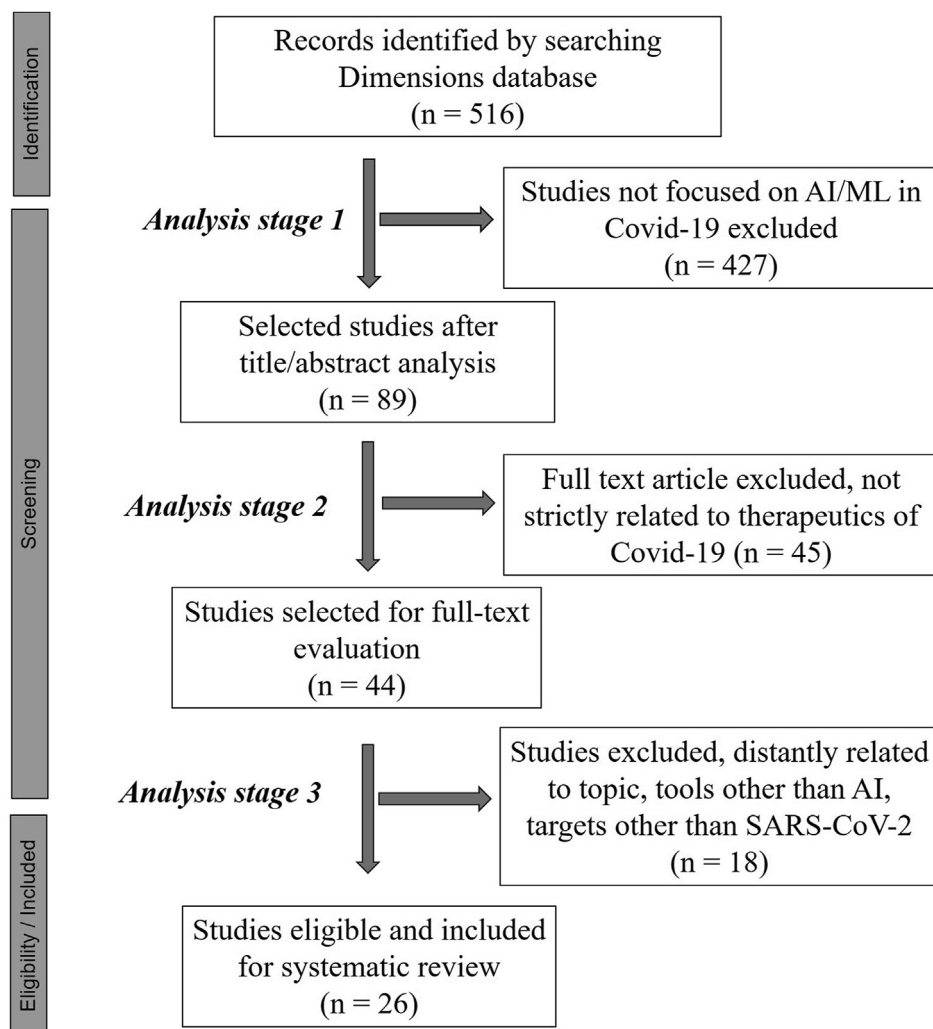

Fig. 2. PRISMA flow diagram for the systematic review of the role of AI/ML in drug repurposing and drug development for COVID-19.

analyzed; and finally, 26 studies met the final inclusion criteria, and were included in our study. Fig. 2 shows the total number of articles obtained from the Dimensions database and the final number of articles considered for the research after applying all the inclusion/exclusion criteria. Table 1 summarizes the findings of these studies.

3.1. AI/ML-based methods in COVID-19 drug repurposing and drug discovery

AI-based drug discovery and drug repurposing have been much publicized as an effective approach to accelerate the drug discovery process (Zhou et al., 2020b; Prasad et al., 2020; Abbasi, 2020; Richardson et al., 2020; Ge et al., 2020; Hong et al., 2019; Beck et al., 2020). Broadly, AI can be very useful for initial drug discovery in two main ways, one is through screening millions of chemical compounds available in different databases for potential drugs in simulation tests, and the other is to identify novel drugs that can latch onto the targets, to reduce their infectivity. The strategy of drug repurposing comes out as a powerful solution for COVID-19 (Zhou et al., 2020a, 2020b; Prasad et al., 2020; Abbasi, 2020). Within a month after the first COVID-19 report in China, two independent groups have used AI in different ways to find possible treatments for the SARS-CoV-2. Scientists from the AI drug discovery company BenevolentAI and Imperial College London utilized their

in-house developed algorithms to mine the data and find the enzyme, adaptor-associated protein kinase 1 (AAK1) as a possible target for the COVID-19. The program then reported Baricitinib as one of the best inhibitors out of 378 known AAK1 inhibitors (Richardson et al., 2020). Baricitinib is an approved drug against rheumatoid arthritis.

Ge et al. (2020) identified CVL218, a promising PARP1 (Poly (ADP-Ribose) Polymerase 1) inhibitor through a Natural Language Processing (NLP) model. Here, a Biomedical Entity Relation Extraction (BERE) approach (Hong et al., 2019) was applied to the PubMed database and filtered after searching the term - candidate drug compounds, coronaviruses, or the related proteins. Beck et al. (2020) have utilized their pre-trained deep learning-based drug-target interaction model known as Molecule Transformer-Drug Target Interaction (MT-DTI) to identify FDA-approved antivirals against SARS-CoV-2 proteins (the 3CL^{pro}, RdRP, helicase, 3′-to-5′ exonuclease, endoRNAse, and 2′-O-ribose methyltransferase). This model exploited simplified molecular-input line-entry system (SMILES) strings and amino acid sequences as 1D string inputs and thus can be easily applied to target proteins that do not have any 3D structures.

Another research group used a knowledge graph (KG) based deep learning method for drug repurposing in COVID-19 and termed as CoV-KGE (Zeng et al., 2020). The authors utilized a DL approach, RotatE, developed by Amazon supercomputing resource, AWS-AI (Wang et al.,

Table 1
Summary of important AI and ML-based studies for drug repurposing and drug development in COVID-19.

Author	Study keywords	Targets	No. of drugs	Reference
Ke, Y. Y	Deep neural network	FIP	80	Ke et al. (2020)
Ge, Y	Knowledge Graph, deep learning	SARS-CoV-2	64	Ge et al. (2020)
Beck, B. R	Hybrid CNN and RNN model called MT-DTI	3CLpro, RdRp, helicase, 3′-to-5′ exonuclease, endoRNAse, and 2′-O-ribose methyltransferase	3410	Beck et al. (2020)
Zeng, X	Deep learning-based knowledge graph (CoV-KGE),	SARS-CoV-2	41	Zeng et al. (2020)
Gao, K	2-D fingerprint, GBDT model, Recurrent Neural Network (RNN)	3CLpro	40	Gao et al. (2020)
Hofmarcher, M	Deep neural network, ChemAI	3CLpro, PLP	20	Hofmarcher et al. (2020)
Ton, A. T	deep learning platform – Deep Docking (DD)	Mpro	1000	Ton et al. (2020)
Hu, F	Deep learning-based multi-task models, Classification and Regression	3CLpro	10	Hu and Jiang (2020)
Gysi, D.M	Graph Neural Network	SARS-CoV-2	77	Gysi et al. (2020)
Huang, K	Deep Purpose, Python toolkit, CNN	3CLpro	13	Huang et al. (2021)
Batra, R	Random forest (RF) regression algorithm, ensemble docking	S, S-ACE2 complex	187	Batra et al. (2020)
Redka, D.S	Deep learning, Ligand Design, MatchMaker, PolypharmDB	3CLpro, Spike; ACE2, TMPRSS2, Cathepsin B	30	Redka et al. (2020)
Zhang, H	Dense fully CNN (DFCNN), DL, Virtual screening	3CLpro	+100 drugs & 20 peptides	Zhang et al. (2020a)
Nguyen, D. D	Mathematical Deep Learning (MathDL), CNN	3CLpro	15	Nguyen et al. (2020)
Artigas, L	Therapeutic Performance Mapping System (TPMS), ANN, GUILDify	SARS-CoV-2 interactome	12	Artigas et al. (2020)
Mahapatra, S	ML, Naïve Bayes classification algorithm, molecular docking	3CLpro	10	Mahapatra et al. (2020)
Blasiak, A	IDentif.AI, orthogonal array composite design (OACD, drug–dose relationships	SARS-CoV-2	12-drug set, over 530,000 drug combinations	Blasiak et al. (2021)
Chakravarty, K	BIOiSIM, in silico simulation, and modeling	ACE2, Spike protein	>2000	Chakravarty et al. (2021)
Zhavoronkov, A	Deep learning-based generative model, AAE, GENTRL	3CLpro, Mpro	1000	Zhavoronkov et al. (2020)
Bung, N	Deep neural network, New Chemical Entities (NCE)	3CLpro	31	Bung et al. (2020)
Tang, B	Advanced deep-Q learning network with the fragment-based drug design (ADQN-FDNN)	3CLpro	47	Tang et al. (2020)
Chenthamarakshan, V	Deep learning-based generative model, CogMol, VAE	Mpro, Spike, NSP9	3000	Chenthamarakshan et al. (2020)
Delijewski, M	MACCS fingerprints, deep learning, gradient-boosted tree learning	3CLpro	~290000	Delijewski and Haneczok (2021)
Laponogov, I	Corona-AI, DreamLab App platform,		52	Laponogov et al. (2021)
Kowalewski, A	Machine learning, SVM, REF	65 host SARS-CoV-2 targets	+10 million	Kowalewski and Ray (2020)
Cantürk, S	ANN, long-short term memory network (LSTM), CNN	SARS-CoV-2	12	Cantürk et al. (2020)

2019) to construct a KG from 24 million PubMed publications and DrugBank. This comprehensive KG includes 15 million edges across 39 types of relationships connecting drugs, diseases, genes/proteins, pathways, and expression profiles. Subsequently, a DL approach (RotatE in DGL-KE) was used to provide high-confidence drug candidates for drug repurposing. The authors then identified 41 candidate drugs through enrichment analysis of drug-gene signatures and SARS-CoV-2 induced transcriptome and proteomics data along with ongoing clinical data.

Gao et al. (2020) used structure-based drug repositioning (SBDR) ML models to evaluate the drug binding affinity to SARS-CoV-2 3CL^{pro}. In this work, the authors trained their 2-D fingerprint-based DL gradient-boosting decision tree (GBDT) model on 314 SARS-CoV-2/SARS-CoV-3CL^{pro} inhibitors to predict the binding affinities of potential protease inhibitors. A total of 8565 drugs (including 1553 FDA-approved drugs) from DrugBank were evaluated and the top 20 FDA-approved drugs along with the top 20 investigational off-market drugs were selected as potent inhibitors of SARS-CoV-2 3CL protease.

Hofmarcher et al. (2020) conducted a deep ligand-based virtual screening using DL network model "ChemAI," trained on more than 220 million data points across 3.6 million molecules from three public drug-discovery databases. They screened approximately 900 million compounds from the ZINC database and evaluated their inhibitory potentials to the SARS-CoV-2 3CL^{pro}, and the papain-like protease (PLP). Additionally, the authors also screened the DrugBank for drug repurposing. By ranking the compounds according to their predicted inhibitory potentials, toxicity, and closeness to known drugs, the authors created a list of 30,000 possible compounds for further screening. These top-ranked compounds were made available as a library at https://github.com/ml-jku/sars-cov-inhibitors-chemai.

Similar work has been reported by Ton et al. (2020) where they used a deep docking platform trained on a neural network to predict the results of docking simulations. From the ZINC database, the authors identified a set of 3 million candidate 3CL^{pro} inhibitors which were subsequently winnowed down to 1000 compounds for drug repurposing after docking simulation. Hu et al. (Hu and Jiang, 2020) used a multitask neural network model to predict protein-ligand binding affinities of viral proteins against a database of 4895 drugs. They suggested 10 potential drugs with strong binding affinity to their target proteins.

Gysi and colleagues (Gysi et al., 2020) utilized an AI-combined network medicine drug repurposing approach to rank 6340 drugs for their efficacy against SARS-CoV-2. The predictions were then validated on 918 experimentally validated drugs. They developed a multimodal approach that has combinations of different algorithms and further identified 77 potential repurposing drugs. Huang and colleagues (Huang et al., 2021) developed a Python-based DL toolkit, DeepPurpose that is based on an encoder-decoder framework and presented a case study on SARS-CoV-2 3CL^{pro} with 13 potential repurposing candidates identified. Batra et al. (2020) trained and validated a random forest algorithm on data from Smith et al. (Smith and Smith, 2020). The authors then executed their models to CureFFI and DrugCentral datasets containing 1495 and 3967 drugs, respectively. They also applied their model to screen compounds from the BindingDB dataset and identified 19,000 additional candidates that bind strongly to either the native S-protein or the human ACE2-S protein complex.

Redka et al. (2020) utilized DL platform, Ligand Design to identify the FDA-approved drugs and experimental medicines that have the potential to inhibit SARS-CoV-2 infection. They developed a resource, PoylpharmDB that contains over 10,224 drugs along with the computed list of ~8700 proteins predicted to interact with them. The interactions were generated with Cyclica's MatchMaker TM technology which is a DL model trained on the entire human proteome that combines structural and experimental data to predict the binding of drug molecules to protein pockets.

Zhang et al. (2020a) utilized the neural network algorithm trained on the PDBbind database and identified possible inhibitors of the SARS-CoV-2 3CL^{pro}. They then used the structural model of the 3CL^{pro},

explored the databases, ChemDiv, and TargetMol to find promising compounds targeting 3CL^{pro} protein. Similarly, Nguyen et al. (2020) applied the Mathematical Deep Learning (MathDL) approach to identify possible inhibitors for SARS-CoV-2 3CL^{pro}. Their model has been trained on two datasets, ChEMBL and PDBbind database using two different CNNs. Finally, they identified 15 promising drug candidates for SARS-CoV-2 3CL^{pro} using the trained CNN models on the DrugBank dataset. Artigas et al. (2020) utilized a systems biology and AI-based approach, the Therapeutic Performance Mapping System (TPMS) technology (Jorba et al., 2020) to repurpose drugs and drug combinations for COVID-19. TPMS method employs an ANN to measure the potential relationship between the nodes of a network (i.e., protein) grouped based on their association with a phenotype. This strategy has been then used to evaluate the effect of 6605 drugs present in the DrugBank and 122 human proteins retrieved through a literature search of the coronavirus-human interactome. A total of 12 approved drugs have been identified, out of which 4 are currently in COVID-19 clinical trials. Besides, they also identified the drug combinations using ANN of TPMS technology and suggested that a combination of drug pirfenidone with melatonin could be a good candidate against COVID-19 and their combined mechanism of action has been identified at the molecular level through the use of TPMS sampling-based models.

In a study, Scott D. Bembenek of Denovicon Therapeutics (San Diego, CA, 92130, USA) used the Denovicon computational platform to perform a molecular modeling-AI hybrid computational approach to find potential inhibitors of the SARS-CoV-2 main protease (Mpro, 3CL^{pro}) (Bembenek et al., 2020). Over 13,000 FDA-approved drugs and clinical candidates (approximately 30,000 protomers) were investigated and finally arrived at the five hits that may prove useful in the designing of future inhibitors of the main protease.

Recently, Mahapatra et al. (2020) reported the ML model based on the Naive Bayes algorithm, which predicts COVID-19 drugs with more than 70% accuracy. This approach suggested 10 FDA-approved drugs that can be repurposed to target COVID-19.

To optimize the use of drug combination therapy, an AI-based platform, Project Identif.AI (Identifying Infectious Disease Combination Therapy with Artificial Intelligence) were utilized for drug development and drug repurposing (Abdulla et al., 2020). It is a neural network approach built on a quadratic correlation between inputs defined by drugs and their doses and outputs defined by treatment efficacy and safety. The authors examined 12 drug/dose parameters and identified drug combinations that effectively inhibit vesicular stomatitis virus infection of A549 lung cells. Many of the studied drugs are currently used in COVID-19 clinical trials also. The authors also suggested the utilization of this platform for COVID-19 intervention. Blasiak et al. (2021) utilized IDentif.AI to evaluate over 530,000 drug combinations against the SARS-CoV-2 live virus collected from a patient sample. IDentif.AI identified the combination of remdesivir, ritonavir, and lopinavir as a potentially effective treatment against SARS-CoV-2 infection. Further experimental validation indicates that this drug combination exhibits a 6.5-fold enhanced efficacy over remdesivir alone. Also, the author showed that hydroxychloroquine and azithromycin were relatively ineffective against live SARS-CoV-2. Thus, Project IDentif.AI greatly cuts the number of in vitro assays required to evaluate the drug tolerability and efficacy and can be applied along with the in vitro investigations of drug validation.

Yi-Yu Ke et al. (2020) developed an AI system trained on two different learning databases. The first one is an antiviral database against SARS-CoV, SARS-CoV-2, HIV, influenza virus, and the second database contains 210 known 3CL^{pro} inhibitors. The authors identified a total of 80 potential antiviral drugs, among them, 8 drugs were shown to inhibit feline infectious peritonitis (FIP) virus in Fcwf-4 cells.

Bung et al. (2020) constructed a deep neural network-based generative and predictive model for SMILES input strings. The model was first trained on 1.6 million compounds from the ChEMBL database and then applied to a small dataset of protease inhibitors using transfer learning.

The authors used reinforcement learning to train the model and identified potential drug compounds. Based on the screening and docking results, 31 potential inhibitors of SARS-CoV-2 3CL^{pro} have been identified.

Very recently, Chakravarty et al. (2021) developed an AI-integrated Bio-simulation platform for drug development and repurposing of pulmonary hypertension therapies for COVID-19. The group conducted an in-silico modeling by using AI-integrated mechanistic modeling platform BIOiSIM with known preclinical in-vitro and in-vivo datasets for accurately simulating the systemic therapy disposition and site-of-action penetration of Angiotensin-Converting Enzyme (ACE) and calcium channel blockers (CCB) compounds to tissues playing role in COVID-19 pathogenesis. The group provides AI/ML-driven computational modeling for repurposing and accelerated the drug development process.

Zhavoronkov and colleagues (Zhavoronkov et al., 2020) have used an integrated AI-based drug discovery pipeline to generate novel drug compounds against SARS-CoV-2 3CL^{pro}. A total of 28 ML models including, GAE and GAN have generated molecular structures that were optimized with reinforcement learning (RL) approaches. Novel drug-like compounds made with these approaches were published at www.insilico.com/ncov-sprint/and have been continuously updated.

Tang et al. (2020) combined AI with the structure-based drug design (SBDD) to speed up the generation of potential candidate compounds against SARS-CoV-2. The authors generated a list of 284 molecules known to inhibit SARS-CoV-2 3CL^{pro}, break them into 316 fragments, and generated potential lead compounds via an advanced deep Q-learning network with fragment-based drug design (ADQN-FBDD). This framework rewards three aspects of discovered leads, the first is a drug-likeliness score, the second is the addition of pre-determined favorable fragments, and the last is the existence of known pharmacophores. This AI-based approach generated a library of 4922 covalent lead compounds with unique valid structures that are heuristically filtered, and finally, 47 lead compounds were evaluated with molecular docking and simulations. All these 47 top compounds and related derivatives based on SBOP were made available in the molecular library at https://github.com/tbwxmu/2019-nCov.

IBM research scientists from Singapore (Chenthamarakshan et al., 2020) applied deep learning generative modeling framework, Controlled Generation of Molecules (CogMol) as a drug discovery approach. This framework applies a variational autoencoder trained on SMILES strings to learn molecule embeddings. The authors on these embeddings train attribute regression models to predict drug properties and protein binding affinities. The authors then went for conditional sampling using Conditional Latent (attribute) Space Sampling (CLaSS) to generate samples with desired features. They used a multitask deep neural network (MT-DNN) to assess the toxicity of the generated molecules. The authors applied this framework to generate ~ 3000 novel drug candidates against the SARS-CoV-2 non-structural protein 9 (NSP9) replicase, the 3CL^{pro}, and the receptor-binding domain (RBD) of the S protein.

Delijewski and Haneczok from the Medical University of Selesia, Poland (Delijewski and Haneczok, 2021), identified Zafirlukast could be potent against COVID-19 infection due to its antiviral property and its ability to attenuate the cytokine storm. For the identification of Zafirlukast as a potential target against COVID-19, the researcher used the AI-based model for drug discovery. The AI model was based on MACCS fingerprints computed using the RDKit library and implementation of gradient-boosted tree learning method (XGBoost). The FDA-approved drug datasets were used in this study to identify the potential drug target against the COVID-19 infection.

Laponogov et al. (2021), has prepared a network machine learning method to target the SARS-CoV-2 host gene-gene interactome by identifying the potentially bioactive molecules in foods based on their anti-COVID-19 ability. The group performed the analysis using the ideal computational power of the unused thousands of smartphones by using a supercomputing DreamLab app platform. The machine learning model first identifies the anti-COVID-19 candidate drugs from the list of experimental and clinically approved drug lists, which can be used as a repurposed drug against COVID-19 interactome in a 5-fold cross-validated setting. Later the ML model screen the database of bioactive food-based molecules from varied chemical classes to target the SARS-CoV-2 interactome. This model then ultimately using the above information created an in-silico food map to play an important role in clinical studies of precision nutrition intervention against COVID-19.

Recently, Ray and Kowalewski (Kowalewski and Ray, 2020) have utilized their machine learning models to screen more than 10 million small molecules from the ZINC database that contains 200 million small molecules. With their AI-linked drug discovery pipeline, they identified the best-in-class hits for the 65 human proteins that interact with SARS-CoV-2. With their ML models, they prioritize the chemicals based on toxicity and volatility (vapor pressure). The chemical features of the drugs were computed and cross-validated with the recursive feature elimination (RFE) along with the random forest and support vector machine algorithms.

Many drug discovery companies are utilizing AI to accelerate drug development and drug repurposing against the SARS-CoV-2 to confronts the COVID-19 pandemic. At large, the success of AI platforms depends on the data that is being used to 'train' the algorithms. A limited data of SARS-CoV-2 can be a challenge. Some of the AI-based companies employing these intellectual approaches to find new drugs against COVID-19 are listed in supporting information, Table S1. These AI-drug discovery companies or AI-based start-ups are working to hasten a rational drug repurposing of the available drugs or discovery of novel drugs against the novel coronavirus.

3.2. AI/ML in vaccine development

Machine learning has also improved the field of vaccine design over the past two decades by creating the virtual frameworks of "Reverse Vaccinology" (RV) approaches. VaxiJen and Vaxign-ML are some of the examples of ML-based RV approaches (Doytchinova and Flower, 2007; Ong et al., 2020a). Various ML approaches like RF, SVM, RFE, and deep CNN (DCNN) have been used to identify the antigens from a given protein sequence (Bowick and Barrett, 2010; Rahman et al., 2019).

Since the outbreak of the COVID-19 pandemic, different ML-based approaches have been used to predict potential epitopes to design vaccines. Ong et al. (2020b) used Vaxign and Vaxign-ML-based RV to prioritize NSPs as vaccine candidates for SARS-CoV-2. They identified NSP3 as the most promising potential target for vaccine development after spike protein (Ong et al., 2020b). Malone et al. (2020) studied the SARS-CoV-2 proteome and offered a complete vaccine design blueprint for SARS-CoV-2 using the *NEC Immune Profiler* suite of tools to create an epitope map for different HLA alleles. Fast and Chen utilized neural network tools, MARIA and NetMHCPan4 to identify potential T-cell epitopes for SARS-CoV-2 spike receptor-binding domain (RBD) (Fast et al., 2020). Crossman (2020) utilized deep learning RNN and provided simulated sequences of S protein to identify possible targets for vaccine design. Rahman et al. (2020) applied immunoinformatic approaches to produce a SARS-CoV-2 anti-peptide vaccine of S, E (envelop), and M (membrane) protein. They used the ML-based Ellipro antibody epitope predictive method to predict B-specific epitopes in S-protein. Prachar et al. (2020) applied 19 joined epitope-HLA tools, including the Immune Epitope Database (IEDB), ANN (PyTorch), and position-specific weight matrices (PSSM) algorithm, to identify and validate 174 epitopes of SARS-CoV-2 binding strongly to 11 HLA alleles. In addition, Sarkar et al. (2020) applied an SVM technique to design the epitope-based vaccine of COVID-19 and predict the toxicity of designed epitopes.

We believe that an AI-based framework may quicken and improve the design and the development of the vaccine formulation which can enhance the immune response, and improve protection for prophylactic vaccines.

3.3. Applications of ML in SARS-CoV-2 protein structure determination

The SARS-CoV-2, like other coronaviruses, has four conserved structural proteins and 16 non-structural proteins, such as proteases (NSP3 and NSP5) and RdRp (NSP12) (Masters, 2006).

Computational models have been used to predict protein structures (Kuhlman and Bradley, 2019). There are primarily two modeling-based approaches available for the prediction of unknown protein structures (Haddad et al., 2020). The first one is template-based modeling, which predicts structure using similar proteins as a template, and the second one is template-free modeling, where no known related structures are available. Many of the SARS-CoV-2 proteins are close homologs to the proteins in related organisms with known structures. However, for some of the proteins, template-based modeling is not possible because of the lack of experimentally determined template structure. Recently, the prediction of structures for proteins where no template structures are available has been advanced significantly via novel ML methods. Senior et al. (2020) of DeepMind company, UK have developed a system called AlphaFold to predict a variety of protein structures related to COVID-19. AlphaFold has recently won the *Ab initio* category of CASP13 competition (Critical Assessment of Techniques for Protein Structure Prediction) and is based on a deep neural network, ResNet architecture (Senior et al., 2019). It is generally an unbiased model predictor and ignores similar structures when making predictions, which is indeed helpful for COVID-19, as very few similar protein structures are available. Central to AlphaFold lies three different layers of deep neural networks. The first layer is made up of a variational autoencoder weighted with an attention model which creates accurate fragments based on a single amino acid sequence. The second layer is divided into two sublayers, of which the first sublayer optimizes inter-residue distances utilizing a 1D convolutional neural network (CNN) on a contact map. The second sublayer help in optimizing the generated substructures against a protein using a 3D CNN. The third layer then scores the generated protein against the actual model (Senior et al., 2019).

The researchers at DeepMind cross-validated their AlphaFold's generated structure of SARS-CoV-2 spike protein with the experimentally determined spike structures by the Francis Crick Institute. Motivated with the positive results, DeepMind has applied AlphaFold to predict the other structures of SARS-CoV-2 proteins including the membrane protein, protein 3a, NSP2, NSP4, NSP6, and papain-like protease (Jumper et al., 2020). These protein structures can potentially contain druggable sites, and thus will help the drug development efforts to contain COVID-19.

Zhang et al. (2020b) have used C–I-TASSER (Zheng et al., 2019) to create structural models for the SARS-CoV-2 proteins, which are available at (Zhou et al., 2021). C–I-TASSER is an extended version of I-TASSER (Yang et al., 2015) and employs the deep convolutional neural network-based contact maps (Li et al., 2019) to guide the Monte Carlo fragment assembly simulations. C–I-TASSER, also known as "Zhang-Server", is the top-ranked computerized server for protein structure prediction in the CASP13 challenge.

Heo and Feig (2020) employed a deep-learning neural network approach built-in as part of the transform-restrained Rosetta (trRosetta) (Yang et al., 2020) pipeline, to predict the structure of the SARS-CoV-2 proteins. The dilated ResNet-enabled trRosetta network may allow for better performance as it has various output layers for the prediction of the distances and orientation between the residues of a protein. The accuracy of the predicted structure models was further improved by applying molecular dynamics simulation-based refinement. The refined trRosetta and AlphaFold's models in this study were further compared to the Zhang C–I-TASSER (Zheng et al., 2019) models. The authors showed significant variability among most of the predicted models, however, there is some similarity in the predicted structures of the M protein, nsp4, and papain-like protease available at (Jungnick et al., 2021).

To get deeper insights into the molecular structures of different human coronavirus spike (S) proteins, Chen et al. (Serena and Chen, 2020) have employed a combination of MD simulation and deep-learning methodology on S proteins of SARS-CoV-2, SARS-CoV-1, Middle East respiratory syndrome coronavirus (MERS-CoV), and human coronavirus HKU1. They have used unsupervised deep learning architecture based on a convolutional variational autoencoder to systematically compare S protein ensembles from MD simulations. The authors demonstrated large flexibility between the subunits of the S proteins and reveal important regions for S protein oligomerization which could be considered as potential targets for therapeutic interventions.

3.4. Deep neural network translate coronavirus protein structure into music

Based on a nanomechanical analysis of the structure and motions of atoms and molecules at different scales, MIT scientists used ML-based deep neural network models to create music to represent the structure of the SARS-CoV-2 spike protein (Buehler, 2020). The principal author, Markus Buehler specializes in developing ML models to design new proteins and has extensively used sonification to illuminate structural details that might otherwise remain elusive. Sonification is a method to translate protein structures into audible signals. In this study, they have used a unique approach of sonification termed "materiomusic", to use the actual vibrations and structures of molecules to create music. According to the study, the hierarchical organization structure of a protein is reminiscent of the music where the primary sequence of amino acids defines the notes and the secondary structure i.e., the coil of the helix or the flatness of a sheet defines the rhythm and pitch (Fig. 3A). Also, the overall vibrational motions of the molecules were defined by the Anisotropic Network Model (Eyal et al., 2015) approach and incorporated into an audio signal (Qin and Buehler, 2019). The signal is then imported into the Max device, and sounds are generated using Ableton Live (Ableton Live Digital Audio Workstation, 2020), which forms the basis for the secondary signal. The signals from the structure and vibrations of a protein were overlaid and played together, generating a multi-dimensional image of the protein's structure. Further sonification of the SARS-CoV-2 S protein in twelve-tone equal temperament tuning, a total number of 3,647,770 notes were generated in the raw musical coding (Fig. 3B). This work overall results in a nearly 2-h piece of classical music that was uploaded to the music sharing website- SoundCloud for the public to hear (SoundCloud, 2758).

The author also reports an ML-enabled nanomechanical vibrational spectrum of five different protein structures, which provides understanding into how genetic mutations and the binding of the SARS-CoV-2 S protein to the human ACE2 cell receptor directly influence the audio (Buehler, 2020). The authors further suggested that the musical representations of proteins could also be used as a tool to design effective drug therapies, development of *de novo* antibodies, identification of druggable sites within the coronavirus' structure, detecting mutations, and material design by manipulating sound.

These neural network-based interventions can convert protein structure to music rapidly and the Markus team has built up a database of over 10,000 protein songs. The team also developed a free app for the Android smartphone, called the Amino Acid Synthesizer, where users can create their own protein "compositions" from the sounds of amino acids (Kozakov et al., 2006).

4. CONCLUSION and FUTURE CHALLENGES

AI and ML are being applied in many COVID-19-related domains, two of which are accelerating the structural biology of SARS-CoV-2 proteins and structure-based drug discovery along with drug repurposing. This paper has presented a survey of literature review of AI applications in the field of computational biology and medicine. In particular, we have highlighted the role of AI in drug repurposing and the structure analysis of SARS-CoV-2 proteins.

AI/ML methods are generally relying on the application domain and the types of data. Text mining techniques and graph-based approaches
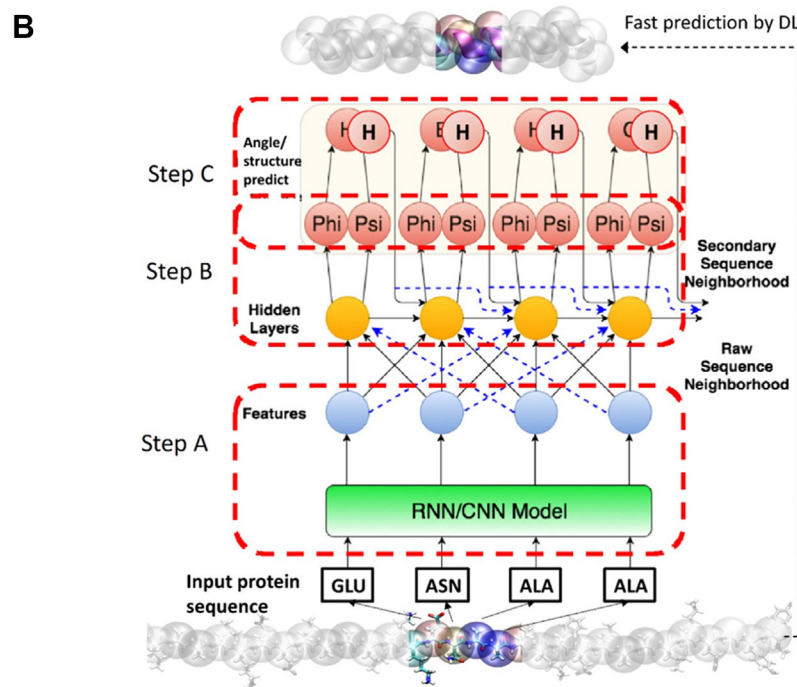
Fig. 3. The hierarchical structure of proteins and music. (A) The three-dimensional structure. of a protein can be translated into a musical score through a process known as sonification which involves a deep neural network model. (B) Neural Network platform (CNN/RNN) for translating protein structure to music. Adapted from. https://towardsdatascience.com/everyprotein-is-a-song-6d30ee9addd4.

are used in drug repurposing while, in predicting drug likeliness, drug target relationship, and generation of novel drug molecules, autoencoder approaches are largely helpful. DL like Graph Convolutional Network and MT-DTI approach proved to be successful to predict available antiviral drugs that could be effective against SARS-CoV-2. According to the so far studies, AI/ML subsets, homology modeling, virtual screenings, and molecular docking are the most used SARS-CoV-2 drug repurposing approaches to identify potentially effective drugs for the treatment of COVID-19 infection. AI/ML-based drug repurposing or drug discovery in the majority of the studies were not confirmed either by experimental methods or follow-up clinical studies. This illustrates the uncertainties regarding reproducibility and strong evidence of drug repurposing studies to tackle COVID-19. However, AI/ML technologies utilized within the drug development studies have greatly improved and could serve, shortly, as a decision support system for policymakers, healthcare

providers, and society at large. The development of effective and robust in vitro and in vivo models can decrease the failure rate of repurposed drugs in preclinical studies and clinical trials. However, challenges remain in developing these technologies, such as data and model harmonization, data heterogeneity and quality, data sharing and security, and biological interpretability of the models.

This is the chance to accomplish the visionary outlook of scientists to deliver an AI-based efficient and agile drug discovery process at an accelerated pace and at a price that every COVID-19 patient would appreciate.

Author contributions

CRediT authorship contribution statement

Kartikay Prasad: Data curation, Writing – original draft. **Vijay Kumar:** Conceptualization, Methodology, Supervision, Writing-Reviewing and Editing.

Declaration of competing interest

No potential conflict of interest was reported by the authors.

Acknowledgments

The authors sincerely thank Amity University, Noida for providing facilities.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crphar.2021.100042.

References

Abbasi, J., 2020. Drug repurposing study pinpoints potential COVID-19 antivirals. J. Am. Med. Assoc. 324, 928.

Ableton Live Digital Audio Workstation, https://www.ableton.com/en/live/, (2020).

Abdulla, Aynur, et al., 2020. Project IDentif. AI: harnessing artificial intelligence to rapidly optimize combination therapy development for infectious disease intervention. Advanced Therapeutics 3 (7), 2000034.

Alimadadi, A., Aryal, S., Manandhar, I., Munroe, P.B., Joe, B., Cheng, X., 2020. Artificial intelligence and machine learning to fight COVID-19. Physiol. Genom. 52, 200–202.

Artigas, L., Coma, M., Matos-Filipe, P., Aguirre-Plans, J., Farres, J., Valls, R., Fernandez-Fuentes, N., de la Haba-Rodriguez, J., Olvera, A., Barbera, J., Morales, R., Oliva, B., Mas, J.M., 2020. In-silico drug repurposing study predicts the combination of pirfenidone and melatonin as a promising candidate therapy to reduce SARS-CoV-2 infection progression and respiratory distress caused by cytokine storm. PloS One 15 e0240149.

Batra, R., Chan, H., Kamath, G., Ramprasad, R., Cherukara, M.J., Sankaranarayanan, S., 2020. Screening of therapeutic Agents for COVID-19 using machine learning and ensemble docking studies. J. Phys. Chem. Lett. 11, 7058–7065.

Baucum, M., Khojandi, A., Vasudevan, R., 2020. In: Improving deep reinforcement learning with Transitional variational autoencoders: a healthcare application. IEEE J, Biomed Health Inform, PP.

Beck, B.R., Shin, B., Choi, Y., Park, S., Kang, K., 2020. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. Comput. Struct. Biotechnol. J. 18, 784–790.

Bembenek, S.D., Repurposing, Drug, Strategies, New Therapeutic, 2020. For SARS-CoV-2 disease using a novel molecular modeling-AI hybrid Workflow. ChemRxiv Preprint.

Blasiak, A., Lim, J.J., Seah, S.G.K., Kee, T., Remus, A., Chye, H., Wong, P.S., Hooi, L., Truong, A.T.L., Le, N., Chan, C.E.Z., Desai, R., Ding, X., Hanson, B.J., Chow, E.K., Ho, D., 2021. IDentif.AI: rapidly optimizing combination therapy design against severe Acute Respiratory Syndrome Coronavirus 2 (SARS-Cov-2) with digital drug development. Bioeng Transl Med 6 e10196.

Bowick, G.C., Barrett, A.D., 2010. Comparative pathogenesis and systems biology for biodefense virus vaccine development. J. Biomed. Biotechnol. 2010, 236528.

Buehler, M.J., 2020. Nanomechanical sonification of the 2019-nCoV coronavirus spike protein through a materiomusical approach. arXiv, 2003.14258.

Bung, N., Krishnan, S.R., Bulusu, G.R., 2020. A, de novo design of new chemical entities (NCEs) for SARS-CoV-2 using artificial intelligence. ChemRxiv Preprint.

Cantürk, S., Singh, A., St-Amant, P., Behrmann, P., 2020. Machine-learning driven drug repurposing for covid-19. arXiv.

Chakravarty, K., Antontsev, V.G., Khotimchenko, M., Gupta, N., Jagarapu, A., Bundey, Y., Hou, H., Maharao, N., Varshney, J., Repurposing, Accelerated, Drug, 2021. Development of pulmonary hypertension therapies for COVID-19 treatment using an AI-integrated Biosimulation platform. Molecules 26.

Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., Blaschke, T., 2018. The rise of deep learning in drug discovery. Drug Discov. Today 23, 1241–1250.

Chenthamarakshan, V., Das, P., Padhi, I., Strobelt, H., Lim, K., Hoover, B., Hoffman, C.S., Mojsilovic, A., Target-Specific, Drug, Selective, 2020. Design for COVID-19 using deep generative models. arXiv preprint.

Crossman, L.C., 2020. Leverging deep learning to simulate coronavirus spike proteins has the potential to predict future Zoonotic sequences. bioRxiv.

Delijewski, M., Haneczok, J., 2021. AI drug discovery screening for COVID-19 reveals zafirlukast as a repurposing candidate. Med Drug Discov 9, 100077.

Doytchinova, I.A., Flower, D.R., 2007. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinf. 8, 4.

Eyal, E., Lum, G., Bahar, I., 2015. The anisotropic network model web server at 2015 (ANM 2.0). Bioinformatics 31, 1487–1489.

Fast, Ethan, Altman, Russ B., Chen, B., T-cell, Potential, Epitopes, B-cell, 2020. Of 2019-nCoV. bioRxiv.

Fleming, N., 2018. How artificial intelligence is changing drug discovery. Nature 557, S55–S57.

Fuchs, C., Life, Everyday, Communication, Everyday, 2020. In coronavirus Capitalism. tripleC: Communication, Capitalism & Critique. Journal for a Global Sustainable Information Society 18, 375–399.

Gao, K., Nguyen, D.D., Chen, J., Wang, R., Wei, G.W., 2020. Repositioning of 8565 existing drugs for COVID-19. J. Phys. Chem. Lett. 11, 5373–5382.

Ge, Y., Tian, T., Huang, S., Wan, F., Li, J., Li, S., 2020. Shen X Z, J, A data-driven drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. bioRxiv preprint.

Griffiths, R.R., Hernandez-Lobato, J.M., 2020. Constrained Bayesian optimization for automatic chemical design using variational autoencoders. Chem. Sci. 11, 577–586.

Gysi, D.M., Do Valle, I., Zitnik, M., Ameli, A., Gan, X., Varol, O., Sanchez, H., Baron, R.M., Ghiassian, D., Loscalzo, J., Barabasi, A.L., 2020. Network medicine framework for identifying drug repurposing opportunities for COVID-19. arXiv.

Haddad, Y., Adam, V., Heger, Z., 2020. Ten quick tips for homology modeling of high-resolution protein 3D structures. PLoS Comput. Biol. 16 e1007449.

Heo, L., Feig, M., 2020. Modeling of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) proteins by machine learning and physics-based Refinement. bioRxiv.

Hofmarcher, M., Mayr, A., Rumetshofer, E., Ruch, P., Renz, P., Schimunek, J., 2020. Hochreiter S, K, G, Large-scale ligand-based virtual screening for SARS-CoV-2 inhibitors using deep neural networks. arXiv.

Hong, L., Lin, J., Tao, J.Z., 2019. J., BERE: an accurate distantly supervised biomedical entity relation extraction network. arXiv preprint.

https:www.dimensions.ai/, (Accessed on 01 September 2020.).

Hu, F., Jiang, J.Y., 2020. P., prediction of potential commercially inhibitors against SARS-CoV-2 by multi-Task deep model. arXiv preprint.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Xiao, Y., Gao, H., Guo, L., Xie, J., Wang, G., Jiang, R., Gao, Z., Jin, Q., Wang, J., Cao, B., 2020a. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395, 497–506.

Huang, K., Fu, T., Glass, L.M., Zitnik, M., Xiao, C., Sun, J., 2021 Apr 1. DeepPurpose: a deep learning library for drug-target interaction prediction. Bioinformatics 36 (22-23), 5545–5547. https://doi.org/10.1093/bioinformatics/btaa1005. PMID: 33275143; PMCID: PMC8016467.

Jorba, G., Aguirre-Plans, J., Junet, V., Segu-Verges, C., Ruiz, J.L., Pujol, A., Fernandez-Fuentes, N., Mas, J.M., Oliva, B., 2020. In-silico simulated prototype-patients using TPMS technology to study a potential adverse effect of sacubitril and valsartan. PloS One 15 e0228926.

Jumper, J., Tunyasuvunakool, K., Kohli, P., Hassabis, D., 2020. a.t.A. team., Computational predictions of protein structures associated with COVID-19. DeepMind website. https://deepmind.com/research/open-source/computational-predictions-of-protein-structures-associated-with-COVID-19.

Jungnick, S., Hobmaier, B., Mautner, L., Hoyos, M., Haase, M., Baiker, A., Lahne, H., Eberle, U., Wimmer, C., Hepner, S., Sprenger, A., Berger, C., Dangel, A., Wildner, M., Liebl, B., Ackermann, N., Sing, A., Fingerle, V., 2021. Detection of the new SARS-CoV-2 variants of concern B.1.1.7 and B.1.351 in five SARS-CoV-2 rapid antigen tests (RATs), Germany, March 2021. Euro Surveill. 26.

Ke, Y.Y., Peng, T.T., Yeh, T.K., Huang, W.Z., Chang, S.E., Wu, S.H., Hung, H.C., Hsu, T.A., Lee, S.J., Song, J.S., Lin, W.H., Chiang, T.J., Lin, J.H., Sytwu, H.K., Chen, C.T., 2020. Artificial intelligence approach fighting COVID-19 with repurposing drugs. Biomed. J. 43 (4). https://doi.org/10.1016/j.bj.2020.05.001.

Keshavarzi Arshadi, A., Webb, J., Salem, M., Cruz, E., Calad-Thomson, S., Ghadirian, N., Collins, J., Diez-Cecilia, E., Kelly, B., Goodarzi, H, Y, J.S., 2020. Artificial intelligence for COVID-19 drug discovery and vaccine development. Front. Artif. Intell. 3.

Khatoon, F., Prasad, K., Kumar, V., 2020. Neurological manifestations of COVID-19: available evidences and a new paradigm. J. Neurovirol. 26, 619–630.

Kowalewski, J., Ray, A., 2020. Predicting novel drugs for SARS-CoV-2 using machine learning from a >10 million chemical space. Heliyon 6 e04639.

Kozakov, D., Brenke, R., Comeau, S.R., Vajda, S., 2006. PIPER: an FFT-based protein docking program with pairwise potentials. Proteins 65, 392–406.

Kuhlman, B., Bradley, P., 2019. Advances in protein structure prediction and design. Nat. Rev. Mol. Cell Biol. 20, 681–697.

Laponogov, I., Gonzalez, G., Shepherd, M., Qureshi, A., Veselkov, D., Charkoftaki, G., Vasiliou, V., Youssef, J., Mirnezami, R., Bronstein, M., Veselkov, K., 2021. Network machine learning maps phytochemically rich "Hyperfoods" to fight COVID-19. Hum. Genom. 15, 1.

Lavecchia, A., 2019. Deep learning in drug discovery: opportunities, challenges and future prospects. Drug Discov. Today 24, 2017–2032.

Li, Y., Zhang, C., Bell, E.W., Yu, D.J., Zhang, Y., 2019. Ensembling multiple raw coevolutionary features with deep residual neural networks for contact-map prediction in CASP13. Proteins 87, 1082–1091.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gotzsche, P.C., Ioannidis, J.P., Clarke, M., Devereaux, P.J., Kleijnen, J., Moher, D., 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration,. PLoS Med. 6 e1000100.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., Chen, J., Meng, Y., Wang, J., Lin, Y., Yuan, J., Xie, Z., Ma, J., Liu, W.J., Wang, D., Xu, W., Holmes, E.C., Gao, G.F., Wu, G., Chen, W., Shi, W., Tan, W., 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 395, 565–574.

Mahapatra, S., Nath, P., Chatterjee, M., Das, N., Kalita, D., Roy, P., Satapathi, S., 2020. Repurposing therapeutics for COVID-19: rapid prediction of commercially available drugs through machine learning and docking. medRxiv.

Malone, B., Simovski, B., Moline, C., Cheng, J., Gheorghe, M., Fontenelle, H., Vardaxis, I., Tennoe, S., Malmberg, J.A., Stratford, R., Clancy, T., 2020. Artificial intelligence predicts the immunogenic landscape of SARS-CoV-2 leading to universal blueprints for vaccine designs. Sci. Rep. 10, 22375.

Masters, P.S., 2006. The molecular biology of coronaviruses. Adv. Virus Res. 66, 193–292.

Mei, X., Lee, H.C., Diao, K.Y., Huang, M., Lin, B., Liu, C., Xie, Z., Ma, Y., Robson, P.M., Chung, M., Bernheim, A., Mani, V., Calcagno, C., Li, K., Li, S., Shan, H., Lv, J., Zhao, T., Xia, J., Long, Q., Steinberger, S., Jacobi, A., Deyer, T., Luksza, M., Liu, F., Little, B.P., Fayad, Z.A., Yang, Y., 2020. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. Nat. Med. https://doi.org/10.1101/2020.04.12.20062661.

Mottaqi, M.S., Mohammadipanah, F., Sajedi, H., 2021. Contribution of machine learning approaches in response to SARS-CoV-2 infection. Inform Med Unlocked 23, 100526.

Nguyen, T.T., 2020. Artificial intelligence in the battle against coronavirus (COVID-19): a survey and future research directions. Preprint.

Nguyen, D.D., Gao, K., Chen, J., Wang, R., Wei, G.W., 2020. Potentially highly potent drugs for 2019-nCoV. bioRxiv.

Ong, E., Wang, H., Wong, M.U., Seetharaman, M., Valdez, N., He, Y., 2020a. Vaxign-ML: supervised machine learning reverse vaccinology model for improved prediction of bacterial protective antigens. Bioinformatics 36, 3185–3191.

Ong, E., Wong, M.U., Huffman, A., He, Y., 2020b. COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. Front. Immunol. 11, 1581.

Piccialli, F., di Cola, V.S., Giampaolo, F., Cuomo, S., 2021. The role of artificial intelligence in fighting the COVID-19 pandemic. Inf. Syst. Front 1–31.

Prachar, M., Justesen, S., Steen-Jensen, D.B., Thorgrimsen, S., Jurgons, E., Winther, O., Bagger, F.O., 2020. Identification and validation of 174 COVID-19 vaccine candidate epitopes reveals low performance of common epitope prediction tools. Sci. Rep. 10, 20465.

Prasad, K., Khatoon, F., Rashid, S., Ali, N., AlAsmari, A.F., Ahmed, M.Z., Alqahtani, A.S., Alqahtani, M.S., Kumar, V., 2020. Targeting hub genes and pathways of innate immune response in COVID-19: a network biology perspective. Int. J. Biol. Macromol. 163, 1–8.

Qin, Z., Buehler, M.J., 2019. Analysis of the vibrational and sound spectrum of over 100,000 protein structures and application in sonification. Extrem. Mech. Lett. 100460.

Rahman, M.S., Rahman, M.K., Saha, S., Kaykobad, M., 2019. Antigenic: an improved prediction model of protective antigens. Artif. Intell. Med. 94, 28–41.

Rahman, M.S., Hoque, M.N., Islam, M.R., Akter, S., 2020. A.S.M. Rubayet Ul Alam, M.A. Siddique, O. Saha, M.M. Rahaman, M. Sultana, K.A. Crandall, M.A. Hossain, Epitope-based chimeric peptide vaccine design against S, M and E proteins of SARS-CoV-2, the etiologic agent of COVID-19 pandemic: an in silico approach. PeerJ 8 e9572.

Redka, D.S., MacKinnon, S.S., Landon, M., Windemuth, A., Kurji, N., Shahani, V., 2020. PolypharmDB, a deep learning-based resource, quickly identifies repurposed drug candidates for COVID-19. ChemRxiv Preprint.

Richardson, P., Griffin, I., Tucker, C., Smith, D., Oechsle, O., Phelan, A., Stebbing, J., 2020. Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. Lancet 395, e30–e31.

Sarkar, B., Ullah, M.A., Johora, F.T., Taniya, M.A., Araf, Y., 2020. The essential Facts of Wuhan novel coronavirus outbreak in China and epitope-based vaccine designing against 2019-nCoV. bioRxiv.

Senior, A.W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Zidek, A., Nelson, A.W.R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D.T., Silver, D., Kavukcuoglu, K., Hassabis, D., 2019. Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). Proteins 87, 1141–1148.

Senior, A.W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Zidek, A., Nelson, A.W.R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D.T., Silver, D., Kavukcuoglu, K., Hassabis, D., 2020. Improved protein structure prediction using potentials from deep learning. Nature 577, 706–710.

Serena, H., Chen, M., 2020. Todd Young, John Gounley, Christopher Stanley, a.D. Bhowmik, distinct structural flexibility within SARS-CoV-2 spike protein reveals potential therapeutic targets. bioRxiv.

Smith, M., Smith, J.C., 2020. Repurposing therapeutics for COVID-19: Supercomputer-based docking to the SARS-CoV-2 viral spike protein and viral spike protein-human ACE2 interface. ChemRxiv Preprint.

SoundCloud.com, https://soundcloud.com/user-275864738/viral-counterpoint-of-the-coronavirus-spike-protein-2019-ncov?in=user-275864738/sets/protein-counterpoint.

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.

Ton, A.T., Gentile, F., Hsing, M., Ban, F., Cherkasov, A., 2020. Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 Billion compounds. Mol Inform, 2000028.

Wang, M., Yu, L., Zheng, D., Gan, Q., Gai, Y., Ye, Z., Li, M., Zhou, J., Huang, Q., Ma, C., Huang, Z., Guo, Q., Zhang, H., Lin, H., Zhao, J., Li, J., Smola, A., Zhang, Z., 2019. Deep Graph Library: Towards Efficient and Scalable Deep Learning On graphs. ArXiv.

Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.L., Abiona, O., Graham, B.S., McLellan, J.S., 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 367, 1260–1263.

Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., Zhang, Y.L., Dai, F.H., Liu, Y., Wang, Q.M., Zheng, J.J., Xu, L., Holmes, E.C., Zhang, Y.Z., 2020. A new coronavirus associated with human respiratory disease in China. Nature 579, 265–269.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER Suite: protein structure and function prediction. Nat. Methods 12, 7–8.

Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., Baker, D., 2020. Improved protein structure prediction using predicted interresidue orientations. Proc. Natl. Acad. Sci. U. S. A. 117, 1496–1503.

Zeng, X., Song, X., Ma, T., Pan, X., Zhou, Y., Hou, Y., Zhang, Z., Li, K., Karypis, G., Cheng, F., 2020. Repurpose open data to discover therapeutics for COVID-19 using deep learning. J. Proteome Res. 4624–4636.

Zhang, Y., Lu, Z., 2019. Exploring semi-supervised variational autoencoders for biomedical relation extraction. Methods 166, 112–119.

Zhang, H., Saravanan, K.M., Yang, Y., Hossain, M.T., Li, J., Ren, X., Pan, Y., Wei, Y., 2020a. Deep learning based drug screening for novel coronavirus 2019-nCov. Interdiscip Sci.

Zhang, C., Zheng, W., Huang, X., Bell, E.W., Zhou, X., Zhang, Y., Structure, Protein, Reanalysis, Sequence, 2020b. Of 2019-nCoV genome Refutes Snakes as its intermediate host and the unique similarity between its spike protein insertions and HIV-1. J. Proteome Res. 19, 1351–1360.

Zhavoronkov, A., Ivanenkov, Y.A., Aliper, A., Veselov, M.S., Aladinskiy, V.A., Aladinskaya, A.V., Terentiev, V.A., Polykovskiy, D.A., Kuznetsov, M.D., Asadulaev, A., Volkov, Y., Zholus, A., Shayakhmetov, R.R., Zhebrak, A., Minaeva, L.I., Zagribelnyy, B.A., Lee, L.H., Soll, R., Madge, D., Xing, L., Guo, T., Aspuru-Guzik, A., 2019. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nat. Biotechnol. 37, 1038–1040. https://doi.org/10.1038/s41587-019-0224-x. Epub 2019 Sep 2. PMID: 31477924.

Zhavoronkov, A., Aladinskiy, V., Zhebrak, A., Zagribelnyy, B., Terentiev, V., Bezrukov, D.S., Yan, Potential, 2020. COVID-2019 3C-like protease inhibitors designed using generative deep learning approaches. ChemRxiv Preprint.

Zheng, W., Li, Y., Zhang, C., Pearce, R., Mortuza, S.M., Zhang, Y., 2019. Deep-learning contact-map guided protein structure prediction in CASP13. Proteins 87, 1149–1164.

Zhou, Y., Wang, F., Tang, J., Nussinov, R., Cheng, F., 2020a Dec. Artificial intelligence in COVID-19 drug repurposing. Lancet Digit Health 2 (12), e667–e676. https://doi.org/10.1016/S2589-7500(20)30192-8. Epub 2020 Sep 18. PMID: 32984792; PMCID: PMC7500917.

Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., Cheng, F., 2020b. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discov 6, 14.

Zhou, H., Dcosta, B.M., Samanovic, M.I., Mulligan, M.J., Landau, N.R., Tada, T., 2021. B.1.526 SARS-CoV-2 variants identified in New York City are neutralized by vaccine-elicited and therapeutic monoclonal antibodies. bioRxiv.