

Recent omics-based computational methods for COVID-19 drug discovery and repurposing

Hilal Tayara[†], Ibrahim Abdelbaky[†] and Kil To Chong

Corresponding author: Kil To Chong, Department of Electronics and Information Engineering, Jeonbuk National University, Jeonju, Jeollabukdo 54896, Republic of Korea; Advances Electronics and Information Research Center, Jeonbuk National University, Jeonju 54896, Republic of Korea; E-mail: kitchong@jbnu.ac.kr

[†]These authors contributed equally to this work.

Abstract

The coronavirus disease 2019 (COVID-19) pandemic, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is the main reason for the increasing number of deaths worldwide. Although strict quarantine measures were followed in many countries, the disease situation is still intractable. Thus, it is needed to utilize all possible means to confront this pandemic. Therefore, researchers are in a race against the time to produce potential treatments to cure or reduce the increasing infections of COVID-19. Computational methods are widely proving rapid successes in biological related problems, including diagnosis and treatment of diseases. Many efforts in recent months utilized Artificial Intelligence (AI) techniques in the context of fighting the spread of COVID-19. Providing periodic reviews and discussions of recent efforts saves the time of researchers and helps to link their endeavors for a faster and efficient confrontation of the pandemic. In this review, we discuss the recent promising studies that used Omics-based data and utilized AI algorithms and other computational tools to achieve this goal. We review the established datasets and the developed methods that were basically directed to new or repurposed drugs, vaccinations and diagnosis. The tools and methods varied depending on the level of details in the available information such as structures, sequences or metabolic data.

Key words: SARS-CoV-2; COVID-19; drug discovery; omics data analysis; deep learning

Introduction

The viral family Coronaviridae hit humanity by three deadly and highly pathogenic viruses namely, SARS-CoV, MERS-CoV and SARS-CoV-2. Coronaviruses (Covs) belong to a family of enveloped, single-stranded and positive-sense RNA viruses. There are four available genres of Covs: Alpha-CoV, Beta-CoV, Gamma-Cov and delta-Cov [1]. Alpha-CoV and Beta-CoV can cross animal-human barriers and emerge to be human pathogens [2, 3], so they are under close investigation.

In December 2019, SARS-COV-2 was declared as the causative agent of Coronavirus and was responsible for coronavirus disease 2019 (COVID-19). It develops flu-like symptoms such as sore throat, cough, headache and fever. These symptoms can be developed into a severe respiratory failure [4]. Since then, the reported number of the confirmed cases by the World Health Organization (WHO) increased exponentially. Thus, it resulted in global and dreadful threats in terms of economy and health. Therefore, the researchers from different disciplines united

Hilal Tayara received his Ph.D. in Electronics and Information Engineering from Jeonbuk National University, South Korea. He is currently an assistant professor at the School of International Engineering and Science at Jeonbuk National University, South Korea. His research fields are Bioinformatics, Computational Biology, Deep Learning and Image Processing.

Ibrahim Abdelbaky received his Ph.D. degree from the Department of Computer Science in Cairo University, Egypt, in 2019. He is currently a lecturer at Artificial Intelligence Department, Faculty of Computers and Artificial Intelligence, Benha University, Banha. His main research includes Machine learning techniques, and their application in Computational Drug Discovery and Bioinformatics.

Kil To Chong received Ph.D. from Texas A&M University in Mechanical Engineering. Currently, he is a Professor of Electronics Engineering Division, president of Electronics and IT New Technologies Research Center in Jeonbuk National University, South Korea. His research fields are Bioinformatics, Computational Biology, Deep Learning and Medical Image Processing.

Submitted: 22 May 2021; **Received (in revised form):** 9 July 2021

together to find therapeutic drugs and vaccines to combat the virus. In this regard, artificial intelligence (AI) with conventional computational tools have proven to be promising methods for accelerating drug discovery, and repurposing [5–8].

AI-based drug discovery and repurposing are powerful tools for the identification of hit molecules for COVID-19 for rapid and cost-effective identification. Therefore, many researchers have already started using AI for developing different novel methods for new drug discovery, drug repurposing and vaccine development. In general, AI has been used in almost all steps of the drug discovery pipeline. It has been utilized for designing a drug with a set of predefined properties including toxicity, bioactivity and bioavailability. These properties have been used for supervising the generation of new drugs using generative models [9–11]. Furthermore, various AI-based models have been used for quantitative structure–activity relationship (QSAR) studies. Here, the matched molecular pair (MMP) analysis [12] is used to evaluate the impact of altering a single localized modification in a candidate drug on its bioactivity and molecular properties [13, 14]. Another contribution of AI in drug discovery is the prediction of the toxicity of a compound because it is considered the most time-consuming and expensive task. For instance, the deep learning-based model called DeepTox [15] achieved outstanding results in Tox21 Data Challenge [15]. Also, AI has a significant contribution in drug design such as the prediction of target protein 3D structure and drug–target interaction (DTI). The 3D structure prediction of a target protein is the main step in structure-based drug discovery (SBDD) because it makes the process of designing new drug molecules relatively attainable if the ligand-binding site is provided [16, 17]. Although traditional methods such as *de novo* protein design and homology modeling have been widely used [18–20], the obtained predictions were not always satisfying. Conversely, deep learning-based methods showed a substantial leap in prediction accuracy. The recent Critical Assessment of Protein Structure Prediction (CASP 14) competition showed a scientific breakthrough made by AlphaFold2. This model achieved the best prediction for 88 out of 97 structures with accuracy comparable to the X-ray crystallography experimental technique. The quantum mechanics and the hybrid quantum mechanics with molecular mechanics DTI methods employed AI in different ways as training AI models to reproduce quantum mechanics energies from atomic coordinates so the calculation time is close to molecular mechanics models with the accuracy of quantum mechanics models [21]. AI in the pharma industry is becoming an essential part of the drug discovery pipeline, therefore, different AI companies are using AI-based tools for finding a potential treatment for COVID-19. These companies concentrate on repurposing existing drugs or designing new drugs. For instance, Benevolent AI company, located in the UK, got the Food and Drug Administration (FDA) approval for the use of their proposed Baricitinib drug in combination with Remdesivir where recovery rate increased for the hospitalized COVID-19 patients [22].

On the other hand, computer-aided drug discovery (CADD) methods have been used as efficient tools to support drug discovery. These tools rely mainly on structural biological information and act in different phases of the drug discovery process. CADD methods are used to reduce the cost and time of traditional high-throughput screening (HTS) where a large number of compounds are tested for a specific activity in the wet lab. CADD methods are categorized into two main subcategories: structure-based and ligand-based. In SBDD, the 3D structure of the target should be available by either experimental or modeling methods. Different techniques are used within the scope of SBDD

such as molecular docking, *de novo* ligand design, molecular dynamics simulations and virtual high throughput screening (VHTS). Ligand-based drug discovery (LBDD) is used when the target structure is not attainable, but information about ligands that bind to the target is provided. LBDD includes different methods such as similarity searchers, QSAR and pharmacophore modeling [23].

Drug Repurposing is performed with the same techniques used for drug discovery but considering fewer drugs based on the application and desired results. Drug repurposing efforts focus only on approved or previously tested compound libraries supported by previous knowledge about their properties including safety. Drug repurposing has many advantages over drug discovery as around 90% of drugs that get to clinical tests fail to satisfy the FDA approval measures. Repurposing of approved drugs can help to save the most portion of costs and time spent for *de novo* drug design [24]. Thus, in a situation such as COVID-19, the need is urgent to apply faster solutions with the least available resources to rapidly confront the pandemic.

In this review, we explored the recent research efforts that utilized AI or other computational methods for finding COVID-19 treatments. We started by stating some of the recent reviews in the field to give as a collective presentation of related studies. After that, our review states the recent efforts in three directions: omics studies, drug discovery and repurposing, and vaccine development. In addition, we list the databases and resources that contain COVID-19-related information and could be used for constructing various AI- and computational-based tools.

Related reviews

Coronavirus has a vital impact on all areas of life: social, economic and health. It prompted researchers from all fields to address this pandemic according to their specializations. The previous works can be roughly categorized into three main categories: pandemic management, image-based diagnosis and drug discovery and repurposing. The general theme of pandemic management research is the utilization of AI in tracking, screening and predicting future patients, and also the roles of new technologies such as drones, the Internet of things (IoT), Blockchain and 5G in managing the impacts of the pandemic. Furthermore, the researchers employed the advances achieved in medical image processing using deep learning to analyze chest x-ray (CXR), computed tomography (CT) and positron emission tomography (PET) images for COVID-19 diagnosis. The most promising direction in the fight against COVID-19 is the utilization of AI in drug discovery and repurposing. Many studies have been carried out to find a new drug or repurpose already existed one for COVID-19 treatment. Table 1 summarizes the previous reviews in these three categories.

SARS-CoV-2

COVID-19 disease is caused by the SARS-CoV-2 virus which is a member of the Covs family. It is characterized as pathogenic enveloped RNA genome viruses. SARS-CoV-2 was found to be more pathogenic than its predecessors in the same family: SARS-CoV, 2002 and MERS-CoV, 2013. The virus spreads among humans by direct or close contact. This spread is taking an exponential manner as it has a person-to-person spread factor (RO) of 2.6. The understanding of the virus mechanism of progress and pathogenesis is substantial for devising possible treatments [51].

Table 1. The previous reviews of computational methods for COVID-19

Category	Focus
Managing the pandemic	<ul style="list-style-type: none"> • The role of AI-based application in managing the pandemic [25–30]. • The role of drones, IoT, AI, Blockchain, robotics and 5G in managing the impacts of pandemic [31–33]. • AI and mathematical modeling for tracking, screening and forecasting [34–38]
Image-based diagnosis	<ul style="list-style-type: none"> • Deep learning for different image modality assessment (CXR, CT and PET) [39–42] • Data acquisition and segmentation [43]
Drug discovery, repurposing and vaccine development	<ul style="list-style-type: none"> • Therapeutic candidate using deep learning [44–49] • Drug repurposing using deep learning [49, 50]

The length of the SARS-CoV-2 genome size is 30 kb [52]. The Sequencing of the SARS-CoV-2 genome showed about 80% identity with SARS-CoV, and 50% identity with MERS-CoV [53]. Most portion of its genome encodes 16 nonstructural proteins (NSP), while the remaining parts encode four structural proteins (SP) in addition to six or seven accessory proteins [52]. The four structural proteins are spike, envelope, membrane and nucleocapsid proteins, denoted as S, E, M and N, respectively [53]. The viral spike (S) protein has characteristics specific to the virus and distinctive importance because it is responsible for enabling the attachment and entry of the virus into the host cells. This role makes it a major factor in the high pathogenic level of SARS-CoV-2 [49, 54]. The main subunit in the spike protein (S1) contains a receptor-binding domain (RBD). The RBD domain works as the mediator for attaching the spike protein to the host receptor angiotensin-converting enzyme 2 (ACE2). The crystal structure of the RBD-ACE2 complex was determined by [55]. In addition to the ACE2, the transmembrane protease serine 2 (TMPRSS2) is another host protein that helps the entry of the virus into the cell at the membrane surface. After the viral genome is released into the cell cytosol, it is translated into various viral proteins that work with some other host elements to facilitate the particle formation and replication of the virus [53]. The roles of ACE2 and TMPRSS2 in viral attachment and entry to the cell made them potential COVID-19 therapeutic targets [56]. However, ACE2 was found in other organs such as the heart and kidney [57]. Thus, it is not preferred as a target because it could lead to undesirable side effects if inhibited [54]. The 3CL protease (M^{Pro}), NSP, was identified as the main protease with an important role in the virus replication process. M^{Pro} was considered as a potential drug target for antiviral drugs in several studies [58].

Omics data analysis

Omics studies such as genomics, epigenomics, transcriptomics, proteomics and metabolomics are key resources for understanding COVID-19. They help in understanding the origin of the virus. In addition, they help in predicting the 3D structure of the proteins of the virus, identifying the sequence of the virus and its mutational variants. Omics data could be processed individually or integrated using different computational- and AI-based tools for providing several biological insights such as origin, genetics variants, protein structures and identification of SARS-CoV-2 sequence. These biological insights are essential for drug discovery, repurposing and vaccine development. A broad overview of omics data analysis workflow is shown in Figure 1.

It is essential to understand the evolutionary origin of SARS-CoV-2 to identify its single nucleotide polymorphisms (SNP) [67, 68]. This process is carried out using phylogeny and mutant variation analysis. Phylogenetic trees mainly

depend on sequence alignment and many tools have been utilized for the alignment of the SARS-CoV-2 genome [69–71]. Conversely, alignment-free tools compare the sequences using features derived from these sequences [72]. Randhawa et al. [73] combined digital signal processing method with supervised machine learning for taxonomic classification of genomic sequences by linking each genomic sequence to discrete values representing its genomic signals. They tested different machine learning methods to detect SARS-CoV-2 and identify its origin.

Understanding the mutant variants in SARS-CoV-2 is essential for vaccine development. Islam et al. reported significant variants [74] based on Genome-wide analysis of SARS-CoV-2. Recently, Hie et al. developed a natural language processing (NLP)-based model to identify the mutations that could affect the immune system of already infected or previously vaccinated people [75].

Furthermore, understanding the functions of all parts of the SARS-CoV-2 genome is an essential step in our battle against the virus. Therefore, various computational tools worked in this direction. Lopez-Rincon et al. [61] used convolutional neural networks (CNN) for identifying representative genomic sequences in SARS-CoV-2. They trained the model on 553 sequences extracted from the National Genomics Data Center repository to separate the genome of the coronavirus family from other different virus strains with an accuracy of 98.73%. Then, they analyzed the trained model to find the sequences that the model used to identify SARS-CoV-2. Whata et al. [62] developed a hybrid CNN-BiLSTM model for classifying SARS CoV-2 among Cows and then used the trained model for discovering regulatory motifs in the SARS CoV-2 genome. Arslan et al. [63] proposed a K-nearest neighbor (KNN) integrated with CpG features for identifying the SARS-CoV-2 genome. Naeem et al. [76] proposed an automated diagnostic system to distinguish between the SARS-CoV, MERS-CoV and SARS-CoV-2 using their genomic sequences. They extracted the features using discrete cosine transform (DCT), discrete Fourier transform (DFT) and seven-moment invariants. These features were passed to two classifiers namely Cascade-forward backpropagation network and KNN.

Another important topic in understanding COVID-19 is protein structure prediction taking into consideration that non-synonymous mutations can alter the function and the structure of the resulting protein [77]. Protein structure identification using wet lab experiments is expensive and time-consuming. Therefore, computational tools are alternative methods for predicting the 3D structures of SARS-CoV-2 proteins. Deep learning-based tools such as AlphaFold [64] and trRosetta [65] have been used for predicting 3D structures of SARS-CoV-2 proteins. In addition, existing computational structure and homology modeling tools have been used for the same purpose such as PyMOL [78], SWISS-MODEL [79], COMPOSER [80] and ITasser [81].

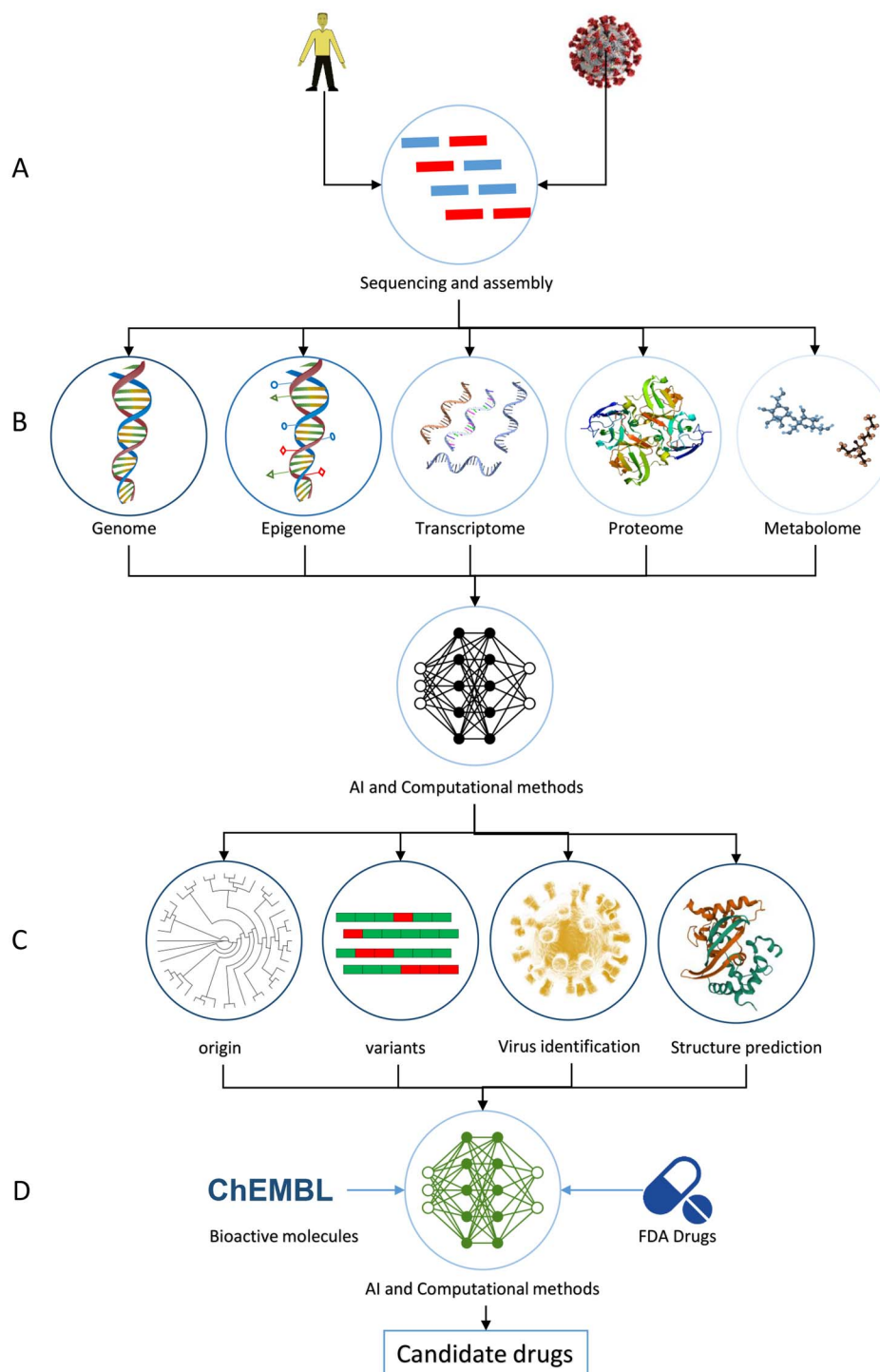


Figure 1. Drug discovery and repurposing using Omics data analysis: (A) the host genome and the virus genome are sequenced and assembled. (B) Various Omics data analysis is performed and AI- and computational-based tools are utilized for extracting different biological insights such as the origin of the virus, the functional variants, virus detection and 3D structure prediction (C). These information are fed into another AI- and computational-based tools in combination with FDA-approved drugs and bioactive molecules datasets for new drug discovery, repurposing and vaccine development (D).

Metabolomics and transcriptomics data analysis have been adopted to provide additional therapeutic strategies for COVID-19 treatment. Transcriptomics data analysis studies the role of a set of genes in different functional pathways and organs on COVID-19 disease. Various studies have been carried out based on the available transcriptomics data such as [82–84]. Loganathan et al. [85] carried out differential expressed gene

analysis of the SARS-CoV-2 and other respiratory infection viruses and produced dysregulated genes in disease conditions. This study identified 31 upregulated host factors including eight pro-viral factors in SARS-CoV-2. Using Connectivity Map-based, they identified repurposed drugs for SARS-CoV-2 infection treatment. More specifically, they suggested that the inhibition of PTGS2 can be considered for treating viral infection and

Table 2. Summary of the AI-based methods in omics data analysis

Task	Method	Description
Origin and mutant variation analysis	Machine learning integrated with digital signal processing method [59] [60] Natural Language processing (NLP) models	- Taxonomic classification of genomic sequences. - The origin of the SARS CoV-2. - Escape mutation in SARS-CoV-2
Identification	Convolution neural network [61]	- Identifying the genome of the coronavirus family with an accuracy of 98.73%. - Designing a specific primer for the identification of SARS-CoV-2
	hybrid CNN-BiLSTM [62]	- Classifying SARS CoV-2 among Coronaviruses - Discovering regulatory motifs in the SARS CoV-2 genome
	K-nearest neighbor (KNN) [63]	- Using CpG feature with KNN for identifying SARS CoV-2
	Cascade-forward backpropagation network and KNN.	- Distinguish Coronaviruses using genomics sequence only
Protein structure prediction	Convolution neural network, AlphaFold [64]	- Predicting the distances between pairs of residues instead of contact information
	Convolution neural network, trRosetta [65]	- Predicting inter-residue orientations and distances
Metabolomics and transcriptomics	Random Forest [66]	- Identified 7 metabolites and 22 proteins related to SARS-CoV-2 infection.

therefore they proposed six approved PTGS2 inhibitors that could be repositioned for treatment of SARS-CoV-2 infection. Jia et al. [86] performed transcriptomics differential expression analysis for healthy and COVID-19 groups. They found that lysosome and endocytosis pathways participate in the disease and they are parts of the disease and disruption of the gene regulation involved in neutrophil degranulation. Then, they used co-expression drug repositioning analysis and reported Saquinavir and Ribavirin antiviral drugs and other candidate drugs. The study of Gordon et al. [87] investigated the map of PPIs for the 26 of the SARS-CoV 2 viral proteins with the human host cell proteins using mass spectrometry. They identified 332 PPIs of which 66 human proteins are targeted by 69 compounds divided as FDA-approved or undergoing clinical tests. The viral assays screening showed a subset of these compounds that can be studied in more detail as potential therapies for COVID-19. They used a variety of tools and servers during the study to perform the required tasks. This included annotation and codon optimization of SARS-Cov 2 genome, tools to predict transmembrane or hydrophobic regions and signal peptides (TMHMM Server v.2.0 [88], SignalP v.5.0 [89]), PPI scoring tools (SAINTexpress (v.3.6.3) [90] and MiST [91]), secondary structure prediction (JPRED) [92], sequence alignment (Clustal Omega) [93], cheminformatics analyses and molecular docking (DOCK3.7) [94].

Various metabolic studies have been carried out for a better understanding of SARS-CoV-2 infection [84, 95–99]. The study by Shen et al. [66] reported seven metabolites and 22 proteins by analyzing metabolomic and proteomic data from 13 server and 18 non-server patients using random forest. Table 2 summarize AI-based tools in omics data analysis.

Drug discovery

The typical process of a new drug discovery takes 13 years and costs 1.3 billion USD on average [112]. However, the COVID-19 pandemic requires rapid steps toward providing new therapeutics to alleviate its consequences on the economy, health

and society. Therefore, using advanced experimental technologies with AI is projected to provide cheaper and quicker new therapeutics for COVID-19 and other complex diseases. Drug discovery pipeline consists of broadly four main steps: target identification, potential compound screening followed by lead optimization, animal trials and clinical trials. In the first step, the disease of interest is studied comprehensively and the target protein is identified and validated. In the second step, methods such as virtual screening, HTS and combinatorial chemistry screen molecular libraries for hit identification. Then, the selected hit molecule goes into an interactive process for functional properties improvement. In the third step, animal models are used for performing *in vivo* studies including pharmacokinetics and toxicity. If the drug candidate passed the first three steps successfully, clinical tests start on patients in the last step. Here, the candidate drug should successfully pass three phases to get the final approval by agencies such as FDA. In the first phase, a small number of people test the drug safety, while in the second phase, the drug efficacy is carried out on a small number of patients. The last phase tests the drug on a large number of patients. The long and complex drug discovery process should be shortened in general and especially for the COVID-19 epidemic. Therefore, AI-based methods in combination with conventional methods can be a boon for reducing the needed time and cost for COVID-19 treatment and they can be used in almost all steps of the drug discovery pipeline. In addition, drug repurposing is a fast alternative in which already approved drugs can be reused for COVID-19 treatment. The workflow of drug repurposing is shown in Figure 2. In this section, we introduce the recent computational tools for COVID-19 drug discovery and repurposing.

Nguyen et al. [100] proposed a mathematical deep learning model for generating a low-dimensional representation of high-dimensional chemical/physical interactions. They integrated this representation into different deep learning models such as CNN and generative adversarial networks (GAN) for predicting the pose and energy of the interaction. They applied this model for finding inhibitors for 3CLpro of SARS-CoV-2 [113].

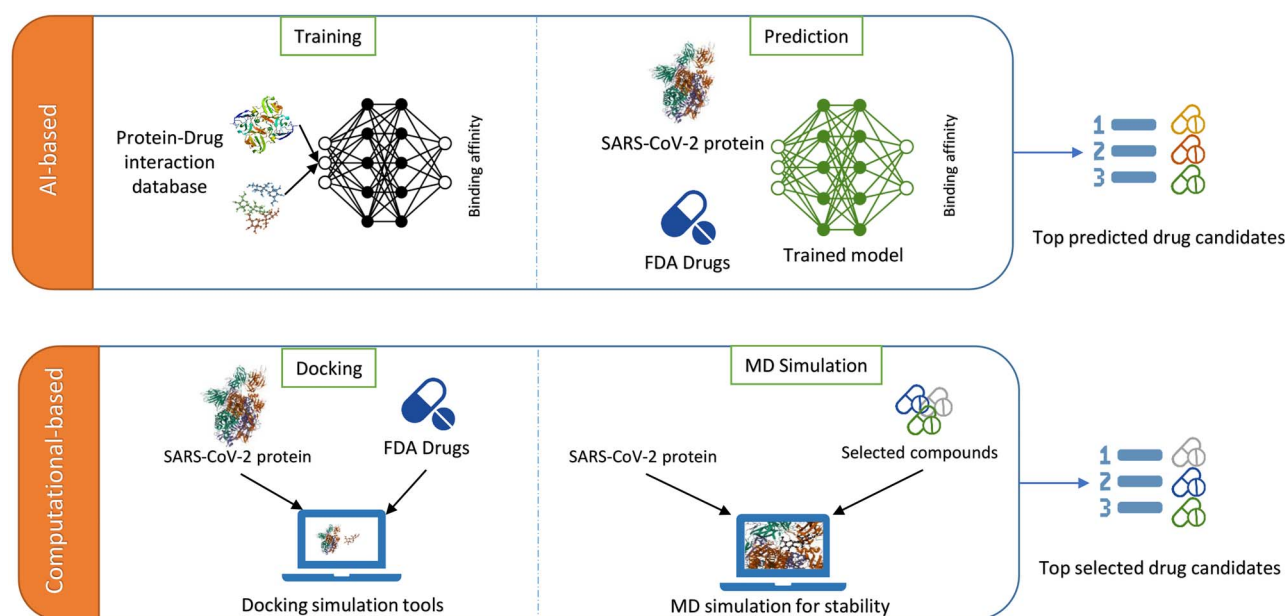


Figure 2. Drug repurposing workflow using AI- and computational-based tools. In AI-based tools, a deep learning model is trained on experimentally available protein-drug interaction dataset. Then, the trained model is used for finding the potential drugs that could bind with the target SARS-CoV-2 protein or host protein. The standard computational tools have two steps: Docking and MD simulation. Docking is used for performing virtual screening and selecting the compounds that could bind a target protein. MD simulation is used to study the complex stability. The final selected drugs are ranked based on the binding affinity scores or docking scores.

Beck et al. [101] searched for commercially available antiviral drugs that interact with SARS-CoV-2 proteins. They targeted different proteins of SARS-CoV-2. They used the pre-trained Molecule transformer–drug target interaction (MT-DTI) [114] for predicting binding affinity. MT-DTI relies only on the sequence information of the target protein and the simplified molecular-input line-entry system (SMILE) of the drugs. The reported results were verified using AutoDock Vina. For instance, they reported that Atazanavir, a human immunodeficiency virus (HIV) treatment, is the best inhibitor for the 3C-like proteinase of SARS-CoV-2.

Pham et al. [102] designed DeepCE, a publically available neural network model to detect high-dimensional associations and nonlinear relationships between biological features to predict gene expression profiles for new chemical compounds. The features of chemical substructures were extracted using a graph convolutional neural network (GCNN), while the associations between genes and genes to chemical substructures were detected by attention mechanism. The gene expression values were predicted using a multilayer feed-forward neural network. The model was applied for COVID-19 clinical phenotypes. They screened the compounds in DrugBank [115] to assign higher priority for promising compounds. The results obtained confirm the available clinical information. After training the DeepCE model on gene expression profiles, they used it to predict gene expression profiles for the drugs in the Drugbank (11 179 drugs). Also, they used SARS-COV-2 gene expression datasets from National Genomics Data Center (NGDC) [116] and the National Center for Biotechnology Information (NCBI) to calculate differential gene expressions for patients. The model predicted 10 and 15 repurposed drugs for population and individual analyses, respectively, where most of the predictions have previous antiviral activity.

The work of Zhang et al. [103] proposed a novel method for COVID-19 drug repurposing based on literature knowledge.

They used the knowledge from SemMedDB [117], which contains semantic predications for the PubMed entries. They used the extracted biomedical knowledge and the COVID-19 literature to construct a knowledge graph. Then, the knowledge graph completion method supported by different neural network-based algorithms was applied to obtain repurposed drugs for COVID-19. This approach predicted drugs that have been already tested for COVID-19 in addition to new suggested drugs.

The study by Auwul et al. [104] identified the major genes and targets involved in the COVID-19 activity. A bioinformatics and machine learning workflow was suggested to achieve their goal. The flow included RNA-sequencing datasets for analyzing gene expression weights and constructing co-expression networks. The key gene modules were selected and analyzed with dedicated tools including DAVID tools [118]. A total of 10 hub genes were determined according to their membership in the modules by applying statistical methods and Protein-Protein Interactions (PPI) network analysis. The validation of hub gene signatures was performed using machine learning methods including SVM and RF. The evaluation was done using common performance measures such as ROC-AUC and accuracy. Potential regulators of the hub genes were identified from analyzing dedicated databases such as JASPAR [119], Tarbase [120] and miRTarbase [121]. Five top repurposed drugs were determined for the important genes by studying gene–drug relations and searching the LINCS-L1000 data [122].

Delijewski and Haneczok [105] performed drug repurposing for antiviral FDA-approved drugs from the DrugBank. A dataset of 290 000 inactive and 405 active compounds was used to train a machine learning (ML) model to predict active inhibitors for the SARS-CoV-2 3CLpro protein. The active compounds used for training the model included active inhibitors against SARS-CoV 3CLpro that have 96% sequence identity to SARS-CoV-2 3CLpro. They calculated MACCS fingerprints for the drugs in the dataset using RDKit library [123]. The used XGBoost for prediction. The

drug zafirlukast was selected as the best potential repurposed drug in terms of prediction score and median lethal dose.

Kim et al. [106] applied two different bioinformatics techniques to suggest potential COVID-19 therapies from the FDA-approved drugs. The first techniques aimed to find inhibitors that can act as blockers for the virus entry to the cells by targeting ACE2 and TMPRSS2. The prediction of the inhibitors was done by a deep learning-based QSAR model, Fluency. The model was trained on data from ChEMBL [124] and then used to predict the binding score between the input proteins and compounds. The second technique aimed to find drugs that can reduce the expression of virus-induced genes by utilizing the Disease Cancelling Technology (DCT) platform. The results of the first technique identified a set of inhibitors, including antiviral agents, for further experimental assessment as beta-lactam antibiotic, Fosamprenavir, glutathione and others. The second technique suggested using Vitamin E, ruxolitinib and glutamine that emphasized the selection of glutathione by the first technique.

Mall et al. [107] proposed a machine learning model that uses the induced vector embeddings of deep learning to represent features of compounds and viral proteins. The model was used to predict the activity of compounds against viral proteins. The selected compounds were ranked by consensus framework. The proposed framework could predict the activity of compounds against viral proteins with high accuracy (Pearson Correlation 0.917 and mean R^2 of 0,84).

Ton et al. [108] used a deep learning-based platform, Deep Docking (DD), to quickly predict the docking score of 1.3 billion compounds from ZINC 15 library [125] against the SARS-CoV-2 M^{Pro} protein. The DD platform rapidly predicts the docking score estimated by any docking program by using docking scores in different databases to train QSAR models. This method enables a fast screening of a large number of compounds. The top 1000 compounds were identified and made available for further scientific inspections.

A broad computational docking study was performed by Berber and Doluca [109] as they obtained 7900 drugs that are FDA-approved or under clinical investigations. The drugs were docked against Dihydroorotate dehydrogenase (DHODH) which is a suggested target for COVID-19 because it is involved in virus replication. A total of 20 DHODH structures were obtained from Protein Databank (PDB) [126] for the docking which was performed with AutoDock Vina [127]. The results selected 28 FDA-approved drugs in addition to 79 of the clinically investigated drugs for further analysis. The interactions of the drugs with the targets were explored using AutoDock4 [128] and DS visualizer. The 28 FDA-approved drugs were suggested for a more detailed experimental examination. They included nine serotonin, dopamine receptor antagonists that are used for treating depression and schizophrenia.

In a molecular docking study, Elfiky et al. [110] suggested a list of compounds as inhibitors for the SARS-CoV-2 RNA dependent RNA polymerase (RdRp). A homology model was built for the SARS-CoV-2 RdRp using the Swiss Model web server [129]. The docking was performed using AutoDock vina [127] considering the built model for SARS-CoV-2 RdRp as the target. The set of tested compounds included 24 approved or clinically investigated drugs in addition to physiological nucleotides. Most of the compounds used in the study showed activity against RdRp from different viruses. The docking results were examined using the Protein-Ligand Interaction Profiler (PLIP) webserver [130]. The study suggested a list of compounds including Ribavirin,

Remdesivir, Setrobuvir, IDX-18 and others as potential inhibitors for the SARS-CoV-2 RdRp.

Wang et al. [111] performed a molecular docking study on the SARS-CoV-2 main protease after its 3D structure was discovered in complex with ligand N3 (PDB ID: 6LU7) [131]. Approved and clinical trials drugs were screened against the target using Glide [132]. The top docked candidates were then studied using molecular dynamics simulations. The results showed promising outcomes with a list of several potential inhibitors. The best results were obtained with carfilzomib and eravacycline. All methods reviewed in this paper are summarized in Table 3.

The methods in the reviewed studies can be broadly divided into learning-based and structure-based. The main actor in selecting the technique to use is the type of available data for the target problem. Learning-based methods mostly ignore the important information contained in the macromolecular structures. Additionally, the number of samples used for training the learning-based models should be of proper size and represent all possible varieties, which is not always available. Another important requirement for such methods is the verification of data quality and integrity. These mentioned factors affect the robustness of the results obtained by different methods. In addition, the difference in data formats needed by different methods hinders the use of several methods on the same data. Though there are promising results for AI-based methods applied in COVID-19 treatments, the impact of AI-based techniques is not yet highly observed because there is a shortage of available data. Suitable strategies are required to consider privacy and public health issues when providing data for AI-based methods to find COVID-19 treatments [133, 134]. The data problem also tackles structural-based methods such as docking and virtual screening. A robust result requires a verified high-quality 3D structure of the target protein(s). To obtain these structures, crystallization and other methods should be applied which is expected to take some time to be available for more COVID-19 targets. Other options such as homology modeling or using similar proteins could give initial indications as a starting point. An important issue that should be observed when studying a host target(s) is the possible side effects that could occur [135–137]. Generally, robust outcomes of AI-based and computational-based strategies depend mainly on the verified data. It should also use efficient and accessible tools to produce predictions. Validations of the predictions should be possible with the knowledge of experts to explain the results and judge their robustness and applicability [138].

AI and computational tools were used extensively in the race for finding COVID-19 treatments. In the course of drug design or repurposing, there are few FDA approved and a variety of suggested lists for further clinical trials. Among these suggestions is Remdesivir which was approved by FDA for treating COVID-19 hospitalized cases. Remdesivir was suggested by different computational-based methods as it got the approval of combined use with Baricitinib that was proposed by Benevolent AI company [22] and it was also suggested by Elfiky et al. [110]. Additional efforts by companies that use AI in developing COVID-19 treatments resulted in a set of drugs that are being validated or clinically tested. Innoplexus: 3 drug combinations, Deargen: atazanavir and Gero: Nine drugs including niclosamide and nitazoxanide are examples of such efforts [133].

Vaccine development

The immune system in humans fights the virus when it enters the body. One type of cells in the immune system is the white

Table 3. The summary of the reviewed drug discovery methods

Study and Method	Summary	Results
Mathematical deep learning model [100]	Search for Inhibitors for Viral Proteins 3CLpro of SARS-CoV-2	Identified fifteen potential inhibitors ranked based on the predicted binding energy. The top predicted inhibitors were Bortezomib, Flurazepam, Ponatinib and Sorafenib.
Molecule transformer–drug target interaction (MT-DTI) [101] DeepCE, Graph neural network and multihead attention mechanism [102]	Repurposing of commercially available antiviral drugs SARS-CoV-2 proteins Predict associations between gene expressions and chemical compounds. It was applied to repurpose drugs for SARS-CoV-2 proteins	Reported Atazanavir as the best inhibitor for the 3C-like proteinase of SARS-CoV-2. The model predicted 10 and 15 repurposed drugs for population and individual analyses, respectively. Most of them have previous antiviral activity.
Neural knowledge graph completion [103]	Building a biomedical knowledge graph of COVID-19 and PubMed articles for COVID-19 drug repurposing	A list of predicted drugs that have been already tested or newly suggested as treatments for COVID-19 based on literature.
Bioinformatics tools and machine learning (RF, SVM) [104]	Identify genes and targets involved in the activity of COVID-19 for drug repurposing	Identified the major genes and targets involved in the COVID-19 activity. Then, five top repurposed drugs were determined for the important genes
Machine Learning XGBoost model [105]	Predict active compounds against SARS-CoV-2 proteins 3CLpro protein from FDA-approved antiviral drugs	The drug zafirlukast was selected as the best potential repurposed drug in terms of prediction score and median lethal dose.
Deep learning-based QSAR model [106]	Predict host ACE2 and TMPRSS2 inhibitors and identify compounds that reduce virus induced genes.	Antiviral agents were suggested for further experimental assessment such as beta-lactam antibiotic, Fosamprenavir, glutathione, and others. Additionally, Vitamin E, ruxolitinib, and glutamine were suggested.
Deep learning model for protein and compound embedding and machine and deep learning for binding prediction [107] Deep Docking (DD) and Docking tools such as Glide [108]	Identified a list of 47 COVID-19 inhibitors such as Rifabutin Rapid screening of compound libraries for SARS-CoV-2 M ^{Pro} protein	The proposed framework could predict the activity of compounds against viral proteins with high accuracy (Pearson Correlation 0.917 and mean R2 of 0,84). The top 1000 compounds were identified and made available for further scientific inspections.
Docking using AutoDock Vina [109]	Docking of FDA-approved, or clinically tested drugs against host protein DHODH as COVID-19 target.	The results selected 28 FDA-approved drugs in addition to 79 of the clinically investigated drugs for further analysis. The 28 FDA-approved drugs included nine serotonin, dopamine receptor antagonists that are used for treating depression and schizophrenia.
Docking using AutoDock Vina [110]	Docking of previous antivirals against homology model of SARS-CoV-2 protein RdRp	The study suggested a list of compounds including Ribavirin, Remdesivir, Setrobutvir, IDX-18, and others as potential inhibitors for the SARS-CoV-2 RdRp.
Docking using Glide [111]	Docking and MD simulations for approved and clinically tested drugs to inhibit SARSCoV-2 main protease	The results showed promising outcomes with a list of several potential inhibitors. The best results were obtained with carfilzomib and eravacycline.

blood cells which work to fight virus infection in different ways by macrophages, B-cells or T-cells. The immune system takes time after infection, days or weeks, to learn how to fight the virus. Then, the immune system memorizes how to fight the virus infection quickly if it happened again. A COVID-19 vaccine does the teaching task for the immune system to identify the SARS-CoV-2 virus if the infection happened, and hence it responds immediately by fighting the virus. To save millions

of people's lives during the pandemic, there is an urgent need to design safe vaccines for COVID-19. As of 29 June 2021, there are 127 vaccine candidates, 367 trials and 19 approved vaccines. There are 35 vaccines in Phase 1, 50 vaccines in Phase 2 and 37 vaccines in Phase 3. Out of 19 approved vaccines, there are eight inactivated vaccines, two protein subunit vaccines, six non-replicating viral vector vaccines and three RNA vaccines (<https://covid19.trackvaccines.org/>).

Here, we review the latest computational methods applied for vaccine development. Magar et al. [139] utilized the proteomics sequences of the SARS-CoV-2, to find antibody sequences that can neutralize the virus. A variety of antibody sequences for different viruses were collected in a dataset, (VirusNet) with their patient and IC50 data. Various machine and deep learning models were tested on the collected data and the best performing model was selected. The selected model was then used to find potential antibodies from a set created based on the SARS 2006 antibody scaffold [140]. Candidates selected by the model were then tested for structural stability using MD simulations, which showed a list of nine suggested SARS-CoV-2 neutralizing antibodies.

In the study of Ong et al. [141], they surveyed the results of vaccination clinical trials that were performed against SARS and MERS viruses. The efforts for vaccination focused on targeting the whole viruses, or any of the spike, nucleocapsid or membrane proteins. The results showed concerns about safety because of the lack of full protection. They applied a new technique that was developed based on machine learning for predicting potential vaccines for COVID-19, Vaxign-ML. The analysis results showed that some viral proteins have a level of conservation among SARS-CoV-2, SARS-CoV and MERS-CoV. The level of protection of the studied proteins was estimated by the Vaxign-ML. The study suggested a combination vaccine for structural proteins (sp), nonstructural proteins (nsp) and spike protein (S) considering nsp3, S, nsp8 as promising vaccine targets.

The study of Yang et al. [142] designed an approach based on deep learning for multi-epitope vaccine design, named Deep-VacPred. The vaccine prediction is based on the sequence of the SARS-CoV-2 spike protein that resulted in suggesting 26 vaccine subunits. Additional *in silico* methods were used to examine the suggested vaccines, where 11 of them were selected for designing a multi-epitope vaccine. The designed vaccine was then tested for different quality aspects using bioinformatics methods. The coverage, toxicity, secondary structures and other properties were assessed showing good qualities. The 3D structure was predicted by using computational tools. Finally, the ability of the designed vaccine was tested against the recent SARS-CoV-2 mutations. Data and several tools used in the study were from The Immune Epitope Database (IEDB) [143].

Data resources

Computational methods rely on the availability of high-quality data to analyze and extract hidden patterns that provide the biologist with clear insights into the problem of interest. Since the pandemic, researchers produced big data sources and used them in different ways to understand and fight COVID-19. In general, they relied on the already established data sources for developing various computational models such as UniProt, Protein Data Bank (PDB), GenBank, LINCS L1000 database and Genotype-Tissue Expression (GTEx). Here, we focus on the recent studies and websites that provide specific data resources for COVID-19.

Chen et al. [144] curated an up-to-date database of the COVID-19-related information published in PubMed. It is named Lit-Covid and accessed at <https://www.ncbi.nlm.nih.gov/research/coronavirus/>. This dataset is updated daily and categorized into general information, mechanism of COVID-19 disease, diagnosis, treatment, prevention, case report and forecasting. Raybould et al. [145] constructed the CoV-AbDab dataset for the known coronavirus-binding antibody. This dataset collects data from patented/published nanobodies and antibodies that bind to betacoronavirus. Korn et al. [146] constructed the COVID-KOP

dataset for integrating biomedical knowledge graphs of Reasoning Over Biomedical Objects linked in Knowledge Oriented Pathways (ROBOKOP) with biomedical literature in the COVID-19 collection (<https://allenai.org/data/cord-19>). Gordon et al. [87] reported 332 protein-protein interactions between human proteins and SARS-CoV-2. Further, they reported that 69 compounds (28 preclinical, 12 in clinical trials and 28 approved by the United States FDA) target 66 host factors or proteins.

Messina et al. [147] proposed a network-based model for understanding viral-host interactome. They reported that host interactome highlights innate immunity pathway components such as chemokines, cytokines, Toll-Like receptors. Ostaszewski et al. [148] provided a COVID-19 disease map by constructing a repository of SARS-CoV-2 virus-host interaction mechanisms. Martin et al. [149] constructed a dataset of potential drugs for COVID-19. They provide up-to-date information on *in vivo*, *in vitro*, clinical trials and computational predictions <https://cordite.mathematik.uni-marburg.de/>.

In addition, different websites provide COVID-19-related data. National Institute of Health (NIH) (<https://datascience.nih.gov/COVID-19-open-access-resources>) provides a big source of COVID-19 data categorized in 13 groups: bioactivity, case studies, chemical structure data, clinical studies, dashboards and visualization tools, digital images, epidemiology, genomics, healthcare resources, literature, participant-level clinical data, RNA-seq and expression counts and social sciences. The COVID-19 browser provides analysis of published drug researches related to COVID-19 (<https://covidtib.c19hcc.org>). COVID-19 data portal (<https://www.covid19dataportal.org/>) provides updated information of the COVID-19 such as viral sequences, host sequences, proteins, expression, images, biochemistry and literature. OverCOVID website (<http://bis.zju.edu.cn/overcovid/>) provides accumulated information about covid-19 for data scientists and bioinformaticians such as biological data, epidemiological data and databases. Nextstrain [150] website (<https://nextstrain.org/>) provides an open-source real-time tracker of pathogen evolution for different viruses including SARS-CoV-2.

The data resource in this review are summarized in Table 4.

Discussion

The irreversible effect of the COVID-19 pandemic shed light on the possible role of AI- and computational-based tools to accelerate therapeutic solutions. Thus, pharmaceutical companies equipped their arsenal with AI algorithms for accelerating drug discovery and repurposing. Successful stories started to emerge as Baricitinib that was proposed by Benevolent AI company. This success demonstrated that the failure rate of drug repurposing could be reduced significantly by robust *in vivo* and *in vitro* model development.

Different methods could be applied depending on the availability of data and knowledge that becomes clear day by day. The scarcity of information at the beginning of the pandemic caused the use of less accurate or multi-level prediction methods. When more knowledge becomes available, such as genomic and proteomic sequences, structural data and experimental or clinical results, the use of more accurate methods turns to be possible. The availability of protein structures allowed using computational docking and virtual screening for drug repurposing against COVID-19 or host cell proteins. While AI methods could also support structure-based methods, they can do the task by utilizing clinical and experimental results when they are available. AI and machine learning could also infer hidden

Table 4. Summary of the reviewed COVID-19 data sources

Data source	Summary	Link
LitCovid	collects and classifies PubMed articles of COVID-19	https://www.ncbi.nlm.nih.gov/research/coronavirus/
CoV-AbDab	Coronavirus-binding antibody	http://opig.stats.ox.ac.uk/webapps/coronavirus
COVID-KOP	Integrates Reasoning Over Biomedical Objects (ROBOKOP) with biomedical literature in the COVID-19 collection	https://covidkop.renci.org/
National Institute of Health (NIH)	Computational approaches and open access data resources for COVID-19	https://datascience.nih.gov/COVID-19-open-access-resources
COVID-19 browser	Analysis of published drug researches related to COVID-19	https://covidtib.c19hcc.org
COVID-19 data portal	Collects data generated from SARS-CoV-2 experiments https://www.covid19dataportal.org/	
OverCOVID	Bioinformatics resources for COVID-19 related researches	http://bis.zju.edu.cn/overcovid/
Nextstrain	Tracking of pathogen evolution of SARS-CoV-2	https://nextstrain.org/

patterns for building accurate prediction and therapeutic models from the raw genomic and proteomic sequences. This diversity of the applicability of computational and AI-based methods made them usable at many stages in the virus tackling efforts according to the types and sizes of available data. Our review shows such diversity and broad applicability by describing several research efforts that utilized different techniques. The overall idea gained from this review aims to help the integration of efforts for faster production of a safe, wide spectrum, and efficient COVID-19 treatment.

Bioinformatics tools have played important role in multi-omics data analysis [151–154]. Multi-omics analysis using AI has several challenges that need to be handled for accelerating the understanding of COVID-19. For instance, multi-omics data are heterogeneous [155] because of using different normalization and scaling methods such as the cases in the transcriptomics and proteomics data. Also, sparse data could be generated from some omics such as metabolomics [156]. Furthermore, outliers should be detected and null values should be imputed [157] before the integration of multi-omics data. Another challenge is a class imbalance in the multi-omics dataset [158] because training machine/deep learning model on imbalance dataset may overfit. Therefore, various techniques could be used for dealing with this issue such as collecting more data if possible, using normalized metrics to measure the machine/deep learning performance such as F1-Score [159], oversampling of underrepresented class, undersampling or overrepresented class, or using methods such as SMOTE [160] or ADASYN [161] for generating synthetic samples of the underrepresented class. The curse of dimensionality in most multi-omics datasets is another challenge that should be handled by applying appropriate feature extraction and selection methods [162]. The storage and computation cost in the case of applying machine/deep learning algorithms on multi-omics data is an additional challenge that should be taken into consideration [163]. The most challenging part in applying AI algorithms on multi-omics data is selecting the appropriate machine/deep learning algorithm. For that, many reviews in literature analyzed the weaknesses and strengths of different ML/DL algorithms using single- and multi-omics data [164, 165]. To obtain a more robust testable hypothesis from multi-omics analysis, a larger and broader COVID-19 patient

population should be considered. Multi-omics analyses help in building COVID-19 knowledge base [166], understanding cellular hallmarks of severe COVID-19, making multi-omics COVID-19 data more accessible and readier for data-driven biological research [167–169].

Some challenges need to be tackled while using AI for drug discovery and repurposing. Therefore, accurate measurements should be implemented in order to accelerate the utilization of AI-based models for COVID-19 or other pandemics. The main challenge is biological interpretation. Biological systems are composed of multiple levels ranging from DNA sequences to organisms. Similarly, the drug discovery process involves multiple levels of interactions between chemical compounds and biological systems. Therefore, the developed AI models should utilize information on the interactions between different entities at different levels. Although the majority of reviewed researches focused on drug repurposing, the development of AI models for drug repurposing for COVID-19 is a challenging task. In general, the repurposed drugs are originally optimized for a certain target with a certain dose. Also, it is possible that cellular or animal tests do not accurately reflect the virus's host environment in people. *In vitro* tests, for example, demonstrate that hydroxychloroquine has anti-SARS-CoV-2 activity [170]. In preclinical and clinical trials, however, hydroxychloroquine has demonstrated little or no efficacy [171]. Therefore, the developed AI tools should take into consideration these challenges. For instance, the existence of diverse populations with varying genetic origins may potentially influence clinical outcomes. Thus, clinical trial success rates might be improved even further using genotype drug repurposing [172]. The construction of effective and reliable *in vitro* and *in vivo* AI-based models for COVID-19 might minimize the failure rate of drug repurposing between preclinical and clinical trials [173, 174]. Another challenge in using AI for COVID-19 is data integration, sharing and security. Data come from different sources and it is important to construct a unified database. This guarantees that the developed models can work in different settings. Furthermore, data security and privacy should be addressed. Questions such as what sort of data will be gathered, if the data are essential, who will collect the data, how the data will be kept, used and transferred, and what would be the rights of the

person whose data are being collected should all be thoroughly addressed.

COVID-19 pandemic has proven that the collaboration between scientists from different disciplines and the sharing of high-quality data are essentials for the fast response toward fighting this pandemic. Data and tools availability are the main essential step for any computational tool. Therefore, the fast generation of data in standard forms is a crucial step for fighting the pandemic. Also, the integration of the experimental data from different laboratories is also important. Therefore, there should be a unified pipeline for data collection and integration. Further, the utilization of the available AI-based software and computational tools in clinical trials is limited during the pandemic. This requires developing more robust tools that provide more accurate results. For example, the need for developing tools that can predict binding affinity for new drugs with different scaffolds from the training set is a significant contribution. Further studies should concentrate on designing more accurate AI-based models for predicting the physical properties of a new drug molecule. Building these models requires collecting more data that cover large chemical space and taking into consideration subcellular compartments [175] or the particular tissue of interest. AI-based models should consider the identification of synergistic drug combinations as this is more effective than concentrating on monotherapies [176, 177]. Also, the limited side effect data and annotations for drug action targets should be explored using AI-based techniques.

Key Points

- We introduced a comprehensive review of the recent computational tools proposed for COVID-19.
- We provided a checkpoint to link the efforts of the researchers working on COVID-19 treatments by AI and related methods.
- We analyzed the utilized approaches in Omics data such as genomics, transcriptomics and proteomics.
- We discussed the employed approaches for drug discovery, repurposing and vaccine development using various AI- and computational-based tools.
- We summarized the latest data resources that help data scientists in developing computational tools for COVID-19.

Funding

This work was supported in part by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2020R1A2C2005612) and in part by the Brain Research Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. NRF-2017M3C7A1044816).

Data Availability

No new data were generated or analysed in support of this research.

References

1. Weiss SR. Forty years with coronaviruses. *Journal of Experimental Medicine* 2020; **217**(5):e20200537.
2. Coleman CM, Frieman MB. Coronaviruses: important emerging human pathogens. *J Virol* 2014; **88**(10): 5209–12.
3. Gorbalenya AE, Baker SC, Baric RS, et al. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-ncov and naming it sars-cov-2. *Nat Microbiol* 2020; **5**(4): 536.
4. Kong W-H, Li Y, Peng M-W, et al. Sars-cov-2 detection in patients with influenza-like illness. *Nat Microbiol* 2020; **5**(5): 675–8.
5. Rifaioğlu AS, Atas H, Martin MJ, et al. Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases. *Brief Bioinform* 2019; **20**(5): 1878–912.
6. Zhong F, Xing J, Li X, et al. Artificial intelligence in drug design. *Sci China Life Sci* 2018; **61**(10): 1191–204.
7. Lavecchia A. Deep learning in drug discovery: opportunities, challenges and future prospects. *Drug Discov Today* 2019; **24**(10): 2017–32.
8. Walters WP, Murcko M. Assessing the impact of generative ai on medicinal chemistry. *Nat Biotechnol* 2020; **38**(2): 143–5.
9. Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. *Sci Adv* 2018; **4**(7):eaap7885.
10. Sanchez-Lengeling B, Aspuru-Guzik A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* 2018; **361**(6400): 360–5.
11. Joulin A, Mikolov T. Inferring algorithmic patterns with stack-augmented recurrent nets arXiv preprint arXiv:1503.01007. 2015.
12. Tyrchan C, Evertsson E. Matched molecular pair analysis in short: algorithms, applications and limitations. *Comput Struct Biotechnol J* 2017; **15**:86–90.
13. Pereira JC, Caffarena ER, Santos CND. Boosting docking-based virtual screening with deep learning. *J Chem Inf Model* 2016; **56**(12): 2495–506.
14. Sheridan RP, Wang WM, Liaw A, et al. Extreme gradient boosting as a method for quantitative structure–activity relationships. *J Chem Inf Model* 2016; **56**(12): 2353–60.
15. Mayr A, Klambauer G, Unterthiner T, et al. Deeptox: toxicity prediction using deep learning. *Front Environ Sci* 2016; **3**:80.
16. Chan HCS, Li Y, Dahoun T, et al. New binding sites, new opportunities for gpcr drug discovery. *Trends Biochem Sci* 2019; **44**(4): 312–30.
17. Chan HCS, Wang J, Palczewski K, et al. Exploring a new ligand binding site of g protein-coupled receptors. *Chem Sci* 2018; **9**(31): 6480–9.
18. Kufareva I, Katritch V, Stevens RC, et al. Advances in gpcr modeling evaluated by the gpcr dock 2013 assessment: meeting new challenges. *Structure* 2014; **22**(8): 1120–39.
19. Zheng Y, Lasker K, Schneidman-Duhovny D, et al. Ucsf chimera, modeller, and imp: an integrated modeling system. *J Struct Biol* 2012; **179**(3): 269–78.
20. Cavasotto CN, Phatak SS. Homology modeling in drug discovery: current trends and applications. *Drug Discov Today* 2009; **14**(13–14): 676–83.
21. Zhang Y-J, Khorshidi A, Kastlunger G, et al. The potential for machine learning in hybrid qm/mm calculations, 241740. *J Chem Phys* 2018; **148**(24).
22. Richardson P, Griffin I, Tucker C, et al. Baricitinib as potential treatment for 2019-ncov acute respiratory disease. *Lancet (London, England)* 2020; **395**(10223):e30.
23. Leelananda SP, Lindert S. Computational methods in drug discovery. *Beilstein journal of organic chemistry* 2016; **12**(1): 2694–718.
24. Cha Y, Erez T, Reynolds IJ, et al. Drug repurposing from the perspective of pharmaceutical companies. *Br J Pharmacol* 2018; **175**(2): 168–80.

25. Kumar A, Gupta PK, Srivastava A. A review of modern technologies for tackling covid-19 pandemic. *Diabetes Metab Syndr Clin Res Rev* 2020; **14**(4): 569–73.
26. Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artificial intelligence for covid-19 (sars-cov-2) pandemic: A review. *Chaos, Solitons & Fractals*, page 2020; **110059**.
27. Bragazzi NL, Dai H, Damiani G, et al. How big data and artificial intelligence can help better manage the covid-19 pandemic. *Int J Environ Res Public Health* 2020; **17**(9): 3176.
28. Naudé W. Artificial intelligence vs covid-19: limitations, constraints and pitfalls. *AI Soc* 2020; **35**(3): 761–5.
29. Rahmatizadeh S, Valizadeh-Haghi S, Dabbagh A. The role of artificial intelligence in management of critical covid-19 patients. *Journal of Cellular & Molecular Anesthesia* 2020; **5**(1): 16–22.
30. Albahri AS, Hamid RA, Alwan JK, et al. Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (covid-19): a systematic review. *J Med Syst* 2020; **44**:1–11.
31. Chamola V, Hassija V, Gupta V, et al. A comprehensive review of the covid-19 pandemic and the role of iot, drones, ai, blockchain, and 5g in managing its impact. *Ieee access* 2020; **8**:90225–65.
32. Sufian A, Ghosh A, Sadiq AS, et al. A survey on deep transfer learning to edge computing for mitigating the covid-19 pandemic. *Journal of Systems Architecture* 2020; **108**:101830.
33. Elavarasan RM, Pugazhendhi R. Restructured society and environment: A review on potential technological strategies to control the covid-19 pandemic. *Sci Total Environ* 2020; **725**:138858.
34. Mohamadou Y, Halidou A, Kapen PT. A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of covid-19. *Applied Intelligence* 2020; **50**(11): 3913–25.
35. Vaishya R, Javaid M, Khan IH, et al. Artificial intelligence (ai) applications for covid-19 pandemic. *Diabetes Metab Syndr Clin Res Rev* 2020; **14**(4): 337–9.
36. Shinde GR, Kalamkar AB, Mahalle PN, et al. Forecasting models for coronavirus disease (covid-19): a survey of the state-of-the-art. *SN Computer Science* 2020; **1**(4): 1–15.
37. Ahmad A, Garhwal S, Ray SK, et al. The number of confirmed cases of covid-19 by using machine learning: Methods and challenges. *Archives of Computational Methods in Engineering* 2020;1–9.
38. Devaraj J, Elavarasan RM, Pugazhendhi R, et al. Forecasting of covid-19 cases using deep learning models: Is it reliable and practically significant? *Results in Physics* 2021; **21**:103817.
39. Shaikh F, Anderson M, Sohail MR, et al. Current landscape of imaging and the potential role for artificial intelligence in the management of covid-19. *Curr Probl Diagn Radiol* 2020.
40. Dong D, Tang Z, Wang S, et al. The role of imaging in the detection and management of covid-19: a review. In: *IEEE reviews in biomedical engineering*, 2020.
41. Salehi AW, Baglat P, Gupta G. Review on machine and deep learning models for the detection and prediction of coronavirus. *Materials Today: Proceedings* 2020; **33**:3896–901.
42. Benameur N, Mahmoudi R, Zaid S, et al. Sars-cov-2 diagnosis using medical imaging techniques and artificial intelligence: A review. *Clin Imaging* 2021.
43. Shi F, Wang J, Shi J, et al. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. In: *IEEE reviews in biomedical engineering*, 2020.
44. Chukwudozie OS, Duru VC, Ndiribe CC, et al. The relevance of bioinformatics applications in the discovery of vaccine candidates and potential drugs for covid-19 treatment. *Bioinformatics and Biology Insights* 2021; **15**:11779322211002168.
45. Shorten C, Khoshgoftaar TM, Furht B. Deep learning applications for covid-19. *Journal of Big Data* 2021; **8**(1): 1–54.
46. Das JK, Tradigo G, Veltri P, et al. Data science in unveiling covid-19 pathogenesis and diagnosis: evolutionary origin to drug repurposing. *Brief Bioinform* 2021; **22**(2): 855–72.
47. Tayarani-N M-H. Applications of artificial intelligence in battling against covid-19: a literature review. *Chaos, Solitons & Fractals*, page 2020; **110338**.
48. Kaushik AC, Raj U. Ai-driven drug discovery: A boon against covid-19? *AI Open* 2020; **1**:1–4.
49. Arshadi AK, Webb J, Salem M, et al. Artificial intelligence for covid-19 drug discovery and vaccine development. *Frontiers in Artificial Intelligence* 2020; **3**.
50. Mohanty S, Rashid MHA, Mridul M, et al. Application of artificial intelligence in covid-19 drug repurposing. *Diabetes Metab Syndr Clin Res Rev* 2020.
51. Naqvi AAT, Fatima K, Mohammad T, et al. Insights into sars-cov-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 2020;165878.
52. Arora N, Banerjee AK, Narasu ML. The role of artificial intelligence in tackling covid-19. *Future Medicine* 2020.
53. Harrison AG, Lin T, Wang P. Mechanisms of sars-cov-2 transmission and pathogenesis. *Trends Immunol* 2020.
54. Md Asif Ahsan, Yongjing Liu, Cong Feng, et al. Bioinformatics resources facilitate understanding and harnessing clinical research of sars-cov-2. *Brief Bioinform*, 2021.
55. Shang J, Ye G, Shi K, et al. Structural basis of receptor recognition by sars-cov-2. *Nature* 2020; **581**(7807): 221–4.
56. Hoffmann M, Kleine-Weber H, Schroeder S, et al. Sars-cov-2 cell entry depends on ace2 and tmprss2 and is blocked by a clinically proven protease inhibitor. *Cell* 2020; **181**(2): 271–80.
57. Hamming I, Wim T, Bulthuis MLC, et al. Tissue distribution of ace2 protein, the functional receptor for sars coronavirus. a first step in understanding sars pathogenesis. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland* 2004; **203**(2): 631–7.
58. Ma C, Sacco MD, Hurst B, et al. Boceprevir, gc-376, and calpain inhibitors ii, xii inhibit sars-cov-2 viral replication by targeting the viral main protease. *Cell Res* 2020; **30**(8): 678–92.
59. Randhawa GS, Hill KA, Kari L. Ml-dsp: Machine learning with digital signal processing for ultrafast, accurate, and scalable genome classification at all taxonomic levels. *BMC Genomics* 2019; **20**(1): 267.
60. Randhawa GS, Hill KA, Kari L. Mldsp-gui: an alignment-free standalone tool with an interactive graphical user interface for dna sequence comparison and analysis. *Bioinformatics* 2020; **36**(7): 2258–9.
61. Lopez-Rincon A, Tonda A, Mendoza-Maldonado L, et al. Classification and specific primer design for accurate detection of sars-cov-2 using deep learning. *Sci Rep* 2021; **11**(1): 1–11.
62. Whata A, Chimedza C. Deep learning for sars cov-2 genome sequences. *IEEE Access* 2021.

63. Arslan H, Arslan H. A new covid-19 detection method from human genome sequences using cpG island features and knn classifier. *Engineering Science and Technology, an International Journal* 2021.
64. Senior AW, Evans R, Jumper J, et al. Improved protein structure prediction using potentials from deep learning. *Nature* 2020; **577**:706–10.
65. Yang J, Anishchenko I, Park H, et al. Improved protein structure prediction using predicted interresidue orientations. *Proc Natl Acad Sci* 2020; **117**(3): 1496–503.
66. Shen B, Yi X, Sun Y, et al. Proteomic and metabolomic characterization of covid-19 patient sera. *Cell* 2020; **182**(1): 59–72.
67. Ceraolo C, Giorgi FM. Genomic variance of the 2019-ncov coronavirus. *J Med Virol* 2020; **92**(5): 522–8.
68. Gonzalez JM, Gomez-Puertas P, Cavanagh D, et al. A comparative sequence analysis to revise the current taxonomy of the family coronaviridae. *Arch Virol* 2003; **148**(11): 2207–35.
69. Wu C, Liu Y, Yang Y, et al. Analysis of therapeutic targets for sars-cov-2 and discovery of potential drugs by computational methods. *Acta Pharmaceutica Sinica B* 2020; **10**(5): 766–88.
70. Zhang T, Wu Q, Zhang Z. Probable pangolin origin of sars-cov-2 associated with the covid-19 outbreak. *Curr Biol* 2020; **30**(7): 1346–51.
71. Waterhouse AM, Procter JB, Martin DMA, et al. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009; **25**(9): 1189–91.
72. Zhou S, Wang Y, Zhu T, et al. Ct features of coronavirus disease 2019 (covid-19) pneumonia in 62 patients in wuhan, china. *Am J Roentgenol* 2020; **214**(6): 1287–94.
73. Randhawa GS, Soltysiak MPM, Roz, et al. Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: Covid-19 case study. *Plos one* 2020; **15**(4): e0232391.
74. Islam MR, Hoque MN, Rahman MS, et al. Genome-wide analysis of sars-cov-2 virus strains circulating worldwide implicates heterogeneity. *Sci Rep* 2020; **10**(1): 1–9.
75. Hie B, Zhong ED, Berger B, et al. Learning the language of viral evolution and escape. *Science* 2021; **371**(6526): 284–8.
76. Naeem SM, Mabrouk MS, Marzouk SY, et al. A diagnostic genomic signal processing (gsp)-based system for automatic feature analysis and detection of covid-19. *Brief Bioinform* 2020.
77. Tiwari M, Mishra D. Investigating the genomic landscape of novel coronavirus (2019-ncov) to identify non-synonymous mutations for use in diagnosis and drug design. *J Clin Virol* 2020; **128**:104441.
78. Janson G, Zhang C, Maria Giulia Prado, et al. Pymol 2.0: improvements in protein sequence-structure analysis and homology modeling within pymol. *Bioinformatics* 2017; **33**(3): 444–6.
79. Waterhouse A, Bertoni M, Bienert S, et al. Swiss-model: homology modelling of protein structures and complexes. *Nucleic Acids Res* 2018; **46**(W1): W296–303.
80. Sutcliffe MJ, Haneef I, Carney D, et al. Knowledge based modelling of homologous proteins, part i: Three-dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Engineering, Design and Selection* 1987; **1**(5): 377–84.
81. Yang J, Yan R, Roy A, et al. The i-tasser suite: protein structure and function prediction. *Nat Methods* 2015; **12**(1): 7–8.
82. Cavalli E, Petralia MC, Basile MS, et al. Transcriptomic analysis of covid-19 lungs and bronchoalveolar lavage fluid samples reveals predominant b cell activation responses to infection. *Int J Mol Med* 2020; **46**(4): 1266–73.
83. Gardinassi LG, Souza COS, Sales-Campos H, et al. Immune and metabolic signatures of covid-19 revealed by transcriptomics data reuse. *Front Immunol* 2020; **11**:1636.
84. Wu D, Shu T, Yang X, et al. Plasma metabolomic and lipidomic alterations associated with covid-19. *Natl Sci Rev* 2020; **7**(7): 1157–68.
85. Loganathan T, Ramachandran S, Shankaran P, et al. Host transcriptome-guided drug repurposing for covid-19 treatment: a meta-analysis based approach. *PeerJ* 2020; **8**:e9357.
86. Jia Z, Song X, Shi J, et al. Transcriptome-based drug repositioning for coronavirus disease 2019 (covid-19). *Pathogens and disease* 2020; **78**(4):ftaa036.
87. Gordon DE, Jang GM, Bouhaddou M, et al. A sars-cov-2 protein interaction map reveals targets for drug repurposing. *Nature* 2020; **583**(7816): 459–68.
88. Krogh A, Larsson B, Von Heijne G, et al. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol* 2001; **305**(3): 567–80.
89. Armenteros JJA, Tsirigos KD, S CK, et al. Signalp 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 2019; **37**(4): 420–3.
90. Teo G, Liu G, Zhang J, et al. Saintexpress: improvements and additional features in significance analysis of interactome software. *J Proteomics* 2014; **100**:37–43.
91. Verschuere E, Von Dollen J, Cimermancic P, et al. Scoring large-scale affinity purification mass spectrometry datasets with mist. *Curr Protoc Bioinformatics* 2015; **49**(1): 8–19.
92. Drozdetskiy A, Cole C, Procter J, et al. Jpred4: a protein secondary structure prediction server. *Nucleic Acids Res* 2015; **43**(W1): W389–94.
93. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol Syst Biol* 2011; **7**(1): 539.
94. Mysinger MM, Shoichet BK. Rapid context-dependent ligand desolvation in molecular docking. *J Chem Inf Model* 2010; **50**(9): 1561–73.
95. Kimhofer T, Lodge S, Whitley L, et al. Integrative modeling of quantitative plasma lipoprotein, metabolic, and amino acid data reveals a multiorgan pathological signature of sars-cov-2 infection. *J Proteome Res* 2020; **19**(11): 4442–54.
96. Shu T, Ning W, Wu D, et al. Plasma proteomics identify biomarkers and pathogenesis of covid-19. *Immunity* 2020; **53**(5): 1108–22.
97. Doğan HO, Şenol O, Bolat S, et al. Understanding the pathophysiological changes via untargeted metabolomics in covid-19 patients. *J Med Virol* 2021; **93**(4): 2340–9.
98. Asim M, Sathian B, Banerjee I, et al. A contemporary insight of metabolomics approach for covid-19: Potential for novel therapeutic and diagnostic targets. *Nepal Journal of Epidemiology* 2020; **10**(4): 923.
99. Zhao Y, Shang Y, Ren Y, et al. Omics study reveals abnormal alterations of breastmilk proteins and metabolites in puerperant women with covid-19. *Signal Transduct Target Ther* 2020; **5**(1): 1–3.
100. Nguyen DD, Gao K, Wang M, et al. Mathdl: mathematical deep learning for d3r grand challenge 4. *J Comput Aided Mol Des* 2020; **34**(2): 131–47.

101. Beck BR, Shin B, Choi Y, et al. 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* 2020; **18**:784–90.
102. Pham T-H, Qiu Y, Zeng J, et al. A deep learning framework for high-throughput mechanism-driven phenotype compound screening and its application to covid-19 drug repurposing. *Nature Machine Intelligence* 2021; **3**(3): 247–57.
103. Zhang R, Hristovski D, Schutte D, et al. Drug repurposing for covid-19 via knowledge graph completion. *J Biomed Inform* 2021; **115**:103696.
104. Auwul MR, Rahman MR, Gov E, et al. Bioinformatics and machine learning approach identifies potential drug targets and pathways in covid-19. *Brief Bioinform* 2021.
105. Delijewski M, Haneczok J. Ai drug discovery screening for covid-19 reveals zafirlukast as a repurposing candidate. *Medicine in Drug Discovery* 2021; **9**:100077.
106. Kim J, Zhang J, Cha Y, et al. Advanced bioinformatics rapidly identifies existing therapeutics for patients with coronavirus disease-2019 (covid-19). *J Transl Med* 2020; **18**(1): 1–9.
107. Mall R, Elbasir A, Almeer H, et al. A modelling framework for embedding-based predictions for compound-viral protein activity. *Bioinformatics* 2021.
108. Ton A-T, Gentile F, Hsing M, et al. Rapid identification of potential inhibitors of sars-cov-2 main protease by deep docking of 1.3 billion compounds. *Molecular informatics* 2020; **39**(8): 2000028.
109. Berber B, Doluca O. A comprehensive drug repurposing study for covid19 treatment: novel putative dihydroorotate dehydrogenase inhibitors show association to serotonin-dopamine receptors. *Brief Bioinform* 2021.
110. Elfiky AA. Ribavirin, remdesivir, sofosbuvir, galidesivir, and tenofovir against sars-cov-2 rna dependent rna polymerase (rdrp): A molecular docking study. *Life Sci* 2020; **253**:117592.
111. Wang J. Fast identification of possible drug treatment of coronavirus disease-19 (covid-19) through computational drug repurposing study. *J Chem Inf Model* 2020; **60**(6): 3277–86.
112. Hopkins AL. Predicting promiscuity. *Nature* 2009; **462**(7270): 167–8.
113. Nguyen D, Gao K, Chen J, et al. Potentially highly potent drugs for 2019-ncov BioRxiv, 2020.
114. Shin B, Park S, Kang K, et al. Self-attention based molecule representation for predicting drug-target interaction. In: *Machine Learning for Healthcare Conference*. PMLR, 2019, 230–48.
115. Wishart DS, Knox C, Guo AC, et al. Drugbank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 2006; **34**(suppl_1): D668–72.
116. Zhou Z, Ren L, Zhang L, et al. Heightened innate immune responses in the respiratory tract of covid-19 patients. *Cell Host Microbe* 2020; **27**(6): 883–90.
117. Kilicoglu H, Shin D, Fiszman M, et al. Semmeddb: a pubmed-scale repository of biomedical semantic predications. *Bioinformatics* 2012; **28**(23): 3158–60.
118. Sherman BT, Lempicki RA, et al. Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nat Protoc* 2009; **4**(1): 44.
119. Khan A, Fornes O, Stigliani A, et al. Jaspas 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res* 2018; **46**(D1): D260–6.
120. Sethupathy P, Corda B, Hatzigeorgiou AG. Tarbase: A comprehensive database of experimentally supported animal microRNA targets. *RNA* 2006; **12**(2): 192–7.
121. Hsu S-D, Lin F-M, Wu W-Y, et al. mirtarbase: a database curates experimentally validated microRNA–target interactions. *Nucleic Acids Res* 2011; **39**(suppl_1): D163–9.
122. Wang Z, Lachmann A, Keenan AB, et al. L1000fwd: fireworks visualization of drug-induced transcriptomic signatures. *Bioinformatics* 2018; **34**(12): 2150–2.
123. Landrum G, et al. Rdkit: Open-source cheminformatics, 2006.
124. Mendez D, Gaulton A, Bento AP, et al. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res* 2019; **47**(D1): D930–40.
125. Sterling T, Irwin JJ. Zinc 15–ligand discovery for everyone. *J Chem Inf Model* 2015; **55**(11): 2324–37.
126. Berman H, Henrick K, Nakamura H. Announcing the worldwide protein data bank. *Nat Struct Mol Biol* 2003; **10**(12): 980–0.
127. Trott O, Olson AJ. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 2010; **31**(2): 455–61.
128. Morris GM, Huey R, Lindstrom W, et al. Autodock4 and autodocktools4: Automated docking with selective receptor flexibility. *J Comput Chem* 2009; **30**(16): 2785–91.
129. Biasini M, Bienert S, Waterhouse A, et al. Swiss-model: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 2014; **42**(W1): W252–8.
130. Salentin S, Schreiber S, Haupt VJ, et al. Plip: fully automated protein–ligand interaction profiler. *Nucleic Acids Res* 2015; **43**(W1): W443–7.
131. Liu X, Zhang B, Jin Z, et al. The crystal structure of covid-19 main protease in complex with an inhibitor n3. *Protein DataBank* 2020.
132. Friesner RA, Banks JL, Murphy RB, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *J Med Chem* 2004; **47**(7): 1739–49.
133. Kaushik AC, Raj U. Ai-driven drug discovery: A boon against covid-19? *AI Open* 2020; **1**:1–4.
134. Naudé W. Artificial intelligence vs covid-19: limitations, constraints and pitfalls. *AI Soc* 2020; **35**(3): 761–5.
135. Ahsan MA, Liu Y, Feng C, et al. Bioinformatics resources facilitate understanding and harnessing clinical research of sars-cov-2. *Brief Bioinform* 2021; **22**(2): 714–25.
136. Hamming I, Wim T, Bulthuis MLC, et al. Tissue distribution of ace2 protein, the functional receptor for sars coronavirus. a first step in understanding sars pathogenesis. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland* 2004; **203**(2): 631–7.
137. Hoffmann M, Kleine-Weber H, Schroeder S, et al. Sars-cov-2 cell entry depends on ace2 and tmprss2 and is blocked by a clinically proven protease inhibitor. *Cell* 2020; **181**(2): 271–80.
138. Galindez G, Matschinske J, Rose TD, et al. Lessons from the covid-19 pandemic for advancing computational drug repurposing strategies. *Nature Computational Science* 2021; **1**(1): 33–41.
139. Magar R, Yadav P, Farimani AB. Potential neutralizing antibodies discovered for novel corona virus using machine learning. *Sci Rep* 2021; **11**(1): 1–11.
140. Hwang WC, Lin Y, Santelli E, et al. Structural basis of neutralization by a human anti-severe acute respiratory

- syndrome spike protein antibody, 80r. *J Biol Chem* 2006; **281**(45): 34610–6.
141. Ong E, Wong MU, Huffman A, et al. Covid-19 coronavirus vaccine design using reverse vaccinology and machine learning. *Front Immunol* 2020; **11**:1581.
 142. Yang Z, Bogdan P, Nazarian S. An in silico deep learning approach to multi-epitope vaccine design: a sars-cov-2 case study. *Sci Rep* 2021; **11**(1): 1–21.
 143. Vita R, Mahajan S, James A, Overton, et al. The immune epitope database (iedb): 2018 update. *Nucleic Acids Res* 2019; **47**(D1): D339–43.
 144. Chen Q, Allot A, Lu Z. Litcovid: an open database of covid-19 literature. *Nucleic Acids Res* 2021; **49**(D1): D1534–40.
 145. Raybould MIJ, Kovaltsuk A, Marks C, et al. Cov-abdab: the coronavirus antibody database. *Bioinformatics (Oxford, England)* btaa739.
 146. Korn D, Bobrowski T, Li M, et al. Covid-kop: integrating emerging covid-19 data with the robokop database. *Bioinformatics* 2020.
 147. Messina F, Giombini E, Agrati C, et al. Covid-19: viral–host interactome analyzed by network based–approach model to study pathogenesis of sars-cov-2 infection. *J Transl Med* 2020; **18**(1): 1–10.
 148. Ostaszewski M, Mazein A, Gillespie ME, et al. Covid-19 disease map, building a computational repository of sars-cov-2 virus-host interaction mechanisms. *Scientific data* 2020; **7**(1): 1–4.
 149. Martin R, Löchel HF, Welzel M, et al. Cordite: the curated corona drug interactions database for sars-cov-2. *Iscience* 2020; **23**(7): 101297.
 150. Ahsan MA, Liu Y, Feng C, et al. Overcovid: an integrative web portal for sars-cov-2 bioinformatics resources. *J Integr Bioinform* 2021; **18**(1): 9–17.
 151. Stephenson E, Reynolds G, Rachel A Botting, et al. Single-cell multi-omics analysis of the immune response in covid-19. *Nat Med* 2021; **27**(5): 904–16.
 152. Yapeng S, Chen D, Yuan D, et al. Multi-omics resolves a sharp disease-state shift between mild and moderate covid-19. *Cell* 2020; **183**(6): 1479–95.
 153. Overmyer KA, Shishkova E, Miller IJ, et al. Large-scale multi-omic analysis of covid-19 severity. *Cell systems* 2021; **12**(1): 23–40.
 154. Ahmed Z. Practicing precision medicine with intelligently integrative clinical and multi-omics data analysis. *Hum Genomics* 2020; **14**(1): 1–5.
 155. Bersanelli M, Mosca E, Remondini D, et al. Methods for the integration of multi-omics data: mathematical aspects. *BMC bioinformatics* 2016; **17**(2): 167–77.
 156. Antonelli J, Claggett BL, Henglin M, et al. Statistical workflow for feature selection in human metabolomics data. *Metabolites* 2019; **9**(7): 143.
 157. Liew AW-C, Law N-F, Yan H. Missing value imputation for gene expression data: computational techniques to recover missing data from available information. *Brief Bioinform* 2011; **12**(5): 498–513.
 158. Haas R, Zelezniak A, Iacovacci J, et al. Designing and interpreting ‘multi-omic’ experiments that may change our understanding of biology. *Current Opinion in Systems Biology* 2017; **6**:37–45.
 159. Jeni LA, Cohn JF, De La Torre F. Facing imbalanced data–recommendations for the use of performance metrics. In: 2013 Humaine association conference on affective computing and intelligent interaction, Vol. 2013. IEEE, 245–51.
 160. Chawla NV, Bowyer KW, Hall LO, et al. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 2002; **16**:321–57.
 161. He H, Yang B, Garcia EA, et al. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence). IEEE, 2008, 1322–8.
 162. Misra BB, Langefeld C, Olivier M, et al. Integrated omics: tools, advances and future approaches. *J Mol Endocrinol* 2019; **62**(1): R21–45.
 163. Herrmann M, Probst P, Hornung R, et al. Large-scale benchmark study of survival prediction methods using multi-omics data. *Brief Bioinform* 2021; **22**(3): bbaa167.
 164. Uddin S, Khan A, Hossain ME, et al. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak* 2019; **19**(1): 1–16.
 165. Sathyanarayanan A, Gupta R, Thompson EW, et al. A comparative study of multi-omics integration tools for cancer driver gene identification and tumour subtyping. *Brief Bioinform* 2020; **21**(6): 1920–36.
 166. Overmyer KA, Shishkova E, Miller IJ, et al. Large-scale multi-omic analysis of covid-19 severity. *Cell systems* 2021; **12**(1): 23–40.
 167. Deshpande C. Thromboembolic findings in covid-19 autopsies: pulmonary thrombosis or embolism? *Ann Intern Med* 2020; **173**(5): 394–5.
 168. Leppkes M, Knopf J, Naschberger E, et al. Vascular occlusion by neutrophil extracellular traps in covid-19. *EBioMedicine* 2020; **58**:102925.
 169. Meyerholz DK, McCray PB, Jr. Illuminating covid-19 lung disease through autopsy studies. *EBioMedicine* 2020; **57**.
 170. Jia Liu, Ruiyuan Cao, Mingyue Xu, et al. Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting sars-cov-2 infection in vitro. *Cell discovery*, **6**(1):1–4, 2020.
 171. Rosenberg ES, Dufort EM, Udo T, et al. Association of treatment with hydroxychloroquine or azithromycin with in-hospital mortality in patients with covid-19 in new york state. *JAMA* 2020; **323**(24): 2493–502.
 172. Hou Y, Zhao J, Martin W, et al. New insights into genetic susceptibility of covid-19: an ace2 and tmprss2 polymorphism analysis. *BMC Med* 2020; **18**(1): 1–8.
 173. Maisonnasse P, Guedj J, Contreras V, et al. Hydroxychloroquine use against sars-cov-2 infection in non-human primates. *Nature* 2020; **585**(7826): 584–7.
 174. Hoffmann M, Mösbauer K, Hofmann-Winkler H, et al. Chloroquine does not inhibit infection of human lung cells with sars-cov-2. *Nature* 2020; **585**(7826): 588–90.
 175. Jarada TN, Rokne JG, Alhaji R. A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions. *J Chem* 2020; **12**(1): 1–23.
 176. Wei Sun, Philip E Sanderson, and Wei Zheng. Drug combination therapy increases successful drug repositioning. *Drug Discov Today*, **21**(7): 1189–95, 2016.
 177. Liu H, Zhang W, Nie L, et al. Predicting effective drug combinations using gradient tree boosting based on features extracted from drug-protein heterogeneous network. *BMC bioinformatics* 2019; **20**(1): 1–12.