Taylor & Francis
Taylor & Francis Group

Journal of Biomolecular Structure and Dynamics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tbsd20

Determining similarities of COVID-19 – lung cancer drugs and affinity binding mode analysis by graph neural network-based GEFA method

Cafer Budak, Vasfiye Mençik & Veysel Gider

To cite this article: Cafer Budak, Vasfiye Mençik & Veysel Gider (2021): Determining similarities of COVID-19 – lung cancer drugs and affinity binding mode analysis by graph neural network-based GEFA method, Journal of Biomolecular Structure and Dynamics, DOI: 10.1080/07391102.2021.2010601

To link to this article: https://doi.org/10.1080/07391102.2021.2010601

Published online: 08 Dec 2021.

JOURNAL OF BIOMOLECULAR STRUCTURE AND DYNAMICS
https://doi.org/10.1080/07391102.2021.2010601

Determining similarities of COVID-19 – lung cancer drugs and affinity binding mode analysis by graph neural network-based GEFA method

Cafer Budak[a], Vasfiye Mençik[b] and Veysel Gider[b]

[a]Department of Biomedical Engineering, Dicle University, Diyarbakır, Turkey; [b]Department of Electric-Electronic Engineering, Dicle University, Diyarbakır, Turkey

Communicated by Ramaswamy H. Sarma

ABSTRACT

COVID-19 is a worldwide health crisis seriously endangering the arsenal of antiviral and antibiotic drugs. It is urgent to find an effective antiviral drug against pandemic caused by the severe acute respiratory syndrome (Sars-Cov-2), which increases global health concerns. As it can be expensive and time-consuming to develop specific antiviral drugs, reuse of FDA-approved drugs that provide an opportunity to rapidly distribute effective therapeutics can allow to provide treatments with known preclinical, pharmacokinetic, pharmacodynamic and toxicity profiles that can quickly enter in clinical trials. In this study, using the structural information of molecules and proteins, a list of repurposed drug candidates was prepared again with the graph neural network-based GEFA model. The data set from the public databases DrugBank and PubChem were used for analysis. Using the Tanimoto/jaccard similarity analysis, a list of similar drugs was prepared by comparing the drugs used in the treatment of COVID-19 with the drugs used in the treatment of other diseases. The resultant drugs were compared with the drugs used in lung cancer and repurposed drugs were obtained again by calculating the binding strength between a drug and a target. The kinase inhibitors (erlotinib, lapatinib, vandetanib, pazopanib, cediranib, dasatinib, linifanib and tozasertib) obtained from the study can be used as an alternative for the treatment of COVID-19, as a combination of blocking agents (gefitinib, osimertinib, fedratinib, baricitinib, imatinib, sunitinib and ponatinib) such as ABL2, ABL1, EGFR, AAK1, FLT3 and JAK1, or antiviral therapies (ribavirin, ritonavir-lopinavir and remdesivir).

ARTICLE HISTORY
Received 1 June 2021
Accepted 21 November 2021

KEYWORDS
Drug similarity; drug repurposing; graph neural network; kinase inhibitors; drug affinity; COVID-19

1. Introduction

The new coronavirus infection, which was first reported in the Wuhan city of China in November 2019, has spread rapidly across the world within a few months and greatly affected many countries

. This virus, called as COVID-19, was denominated as Sars-Cov-2 by the Committee on Nomenclature of Viruses (Wu et al., 2020). This virus, which is believed to be transmitted from the bat through pangolin and belongs to beta coronaviruses, can spread among people in different environments. Sars-Cov (Simmons et al., 2004) and Mers-Cov (The WHO MERS-CoV Research Group, 2013) (a beta-coronavirus), which are similar to the Sars-Cov-2 virus, are thought to cause unprecedented health, economic and social repercussions. The World Health Organization (WHO) declared it as a pandemic on 11 March 2020 (Alhudhaif et al., 2021; Polat et al., 2021). The pandemic has developed dramatically, upon reporting of laboratory-confirmed cases of Sars-Cov-2 by more than 180 countries in the world. In coronavirus, which is divided into four sub-branches as alpha-, beta-, gamma- and delta-coronaviruses, the genome size ranges between 26 kb and 32 kb. Alpha and beta coronaviruses originate from mammals (especially bats), while gamma and delta viruses originate from pigs and birds. While beta-coronaviruses can cause severe illnesses and death, alpha-coronaviruses cause asymptomatic or mild symptomatic infections (Velavan & Meyer, 2020). In Sars-Cov-2, which consists of 16 nsp (nonstructural proteins) and has a single stand + RNA (Chan et al., 2020), proteins are responsible for different cellular functions ranging from self-replicating, infection, and host immune invasion. The rapid spread of the COVID-19 epidemic, the increasing number of cases and death rate, inadequate treatment methods and vaccine options have prompted many governments to take strict measures such as quarantine and travel to combat the pandemic (Karaman et al., 2021). All these results are crucial in determining effective treatment options to prevent the Sars-Cov-2 virus. Drug design and development is an important area of research for pharmaceutical companies and chemical scientists. However, low efficacy, off-target delivery, time consumption and high cost present obstacles and challenges affecting drug design and discovery. In addition, complex and large data obtained from genomics, proteomics, microarray data and clinical trials also pose an obstacle in the drug discovery line. Artificial intelligence and machine learning technology play a crucial role in drug discovery and development (Gupta et al., 2021). Recent advances in Experimental High Efficiency technologies have expanded

CONTACT Cafer Budak cafer.budak@dicle.edu.tr Department of Biomedical Engineering, Dicle University, Diyarbakır 21280, Turkey

the availability and quantity of molecular data in biology. Considering the importance of interactions in biological processes, such as interactions between proteins or bonds within a chemical compound, these data are often represented in the form of a biological network. The increase of this data created the need for new computational tools to analyze networks. One of the biggest trends in the field is to use deep learning for this purpose and more specifically to use methods that work with networks called as Graph Neural Networks (GNNs) (Muzio et al., 2021) COVID-19 causes symptoms such as acute respiratory disorder, fever, cough, sore throat, muscle pain and shortness of breath (Karaman, 2021). Previous studies repositioned many existing drugs to effectively treat infectious diseases caused by single-strand RNA viruses, such as Sars-Cov and Mers-Cov, which cause severe respiratory symptoms. Approaches on prediction of Drug-Target Interactions(DTI), a critical part of drug discovery in pharmaceutical researches for repositioning, predict drug-target interactions based on the similarity between ligands of target proteins (Keiser et al., 2009). While docking-based methods use 3D structure information of a target protein, Ligand and docking methods then run simulations to predict the likelihood of interacting with a particular drug based on binding affinity and strength (Cheng et al., 2007).

In recent years, several approaches have endeavored to exploit drug-drug and protein similarities with drug chemical structure and protein sequence and are based on the assumption of association in cases where similar drugs can share similar goals and vice versa. In this context, that predicts interactions for new drug or target candidates, NetLapRLS, which is a semi-supervised learning method (Xia et al., 2010), Gaussian interaction profile (GIP) kernel-based approach (van Laarhoven et al., 2011), and collaborative matrix factorization (MSCMF) (Zheng et al., 2013) drug-target interactions are methods suggested for predicting drug-target interactions. Some approaches present in the random walk with restart algorithm to predict drug-drug and protein-protein similarities and interactions (Chen et al., 2012). Drug-drug similarity studies aim to find drugs that exhibit similar pharmacological properties to the drug of interest and are guided by the hypothesis that similar drugs should be similar in terms of action mechanism. The drug-drug similarity, which has extensive application in various fields such as drug repositioning (Bibi et al., 2021), drug-drug interaction prediction (Ferdousi et al., 2017), drug target identification (Campillos et al., 2008), and drug side-effects prediction (Lounkine et al., 2012), can be calculated from different sources. Several calculations based on drug properties such as chemical structure characteristics (Zhang et al., 2014), gene expression profiles (Cha et al., 2014), side effect profiles (Tatonetti et al., 2012), and biological target (Sawada et al., 2015) have been applied to drug-drug similarity analytics. It has been assumed that similar drugs may have almost similar interactions and a neighbor recommendation method using molecular structure similarity analysis (Vilar et al., 2012), a computational framework for extracting drug interactions and related recommendations (Gottlieb et al., 2012), a heterogeneous Network-Assisted Inference (HNAI) framework (Cheng & Zhao, 2014) are other studies on drug-drug interactions. The search for the interaction between DNA-binding proteins (DBPs) that play a vital role in cell life activities such as DNA replication and RNA transcription and drugs has been an essential part of genomic drug discovery. Recently, there are studies on protein-drug interactions. In this context, network-based inference (NBI) (Cheng et al., 2012) and similarity indices (Lu et al., 2017) are suggested methods for predicting drug-target interactions. There are also studies suggesting integrates different ligand-based drug design strategies of some in-house chemicals (Amin et al., 2021).

Many approved kinase inhibitors with pharmacological effects that may be beneficial in recovering the life-threatening symptoms of COVID-19 have been suggested as important mediators of Sars-Cov and Mers-Cov in particular. Ideally, a kinase inhibitor with optimal pharmacokinetic properties can reduce infection directly through viral targeting. Kinase inhibitors can be reused as a bifunctional therapeutic that can provide clinical benefit by suppressing disease symptoms. Kinase inhibitors have properties such as anti-inflammatory and cytokine inhibitory activity that can reduce the likelihood of life-threatening conditions due to lung injury. For example, Osimertinib is a strong Epidermal Growth Factor Receptor (EGFR) inhibitor. Osimertinib is one of 24 the U.S. Food and Drug Administration (FDA) approved drugs showing in vitro activity against Sars-Cov-2.

The re-use of previously FDA-approved drugs as treatments for Sars-Cov-2 and related coronaviruses offers an opportunity for rapid distribution of effective therapeutics in the current pandemic environment where treatment options are largely limited. For this purpose, a combined data set from Mers-Cov, Sars-Cov and Sars-Cov-2 was used in this study. The drug data sets used in the study were obtained from PubChem (National Center for Biotechnology Information, 2021) and DrugBank (DrugBank Online, 2021). This combined data set was compared with the FDA-approved data set and the drug-drug similarity analysis was performed. This process was carried out using Tanimoto/Jaccard similarity analysis. The drugs obtained were compared with the drugs used in lung cancer, and the drugs were analyzed using the atom pair similarity method. These determined drugs are drugs which were previously used in the Sars-Cov virus and are still effective in kinase inhibitors and candidates to be effective. Drugs obtained by selecting various kinase inhibitors were investigated in terms of Drug-Protein Affinity Analysis using the Graph Neural Network-Based(GNN) Graph Early Fusion Affinity (GEFA) (Nguyen et al., 2021) model, consisting of four different experimental settings and using Davis dataset. The molecular characteristics that contribute to the most effective antibody response for COVID-19 caused by the Sars-Cov-2 virus are still unclear. The results obtained in this study can make it easier to help fight against the pandemic as quickly as possible during the pandemic crisis, in which the need and the urgency of time are essential for using repurposed drugs for the COVID-19 virus.

The second section of this study mentions about the methods used in the study, the third section mentions about

experimental results, and the fourth section mentions about results.

2. Material and method

The molecular (chemical) similarity has an important role in predicting the properties of chemical compounds, designing chemicals with predefined properties and especially conducting drug discovery studies. It is usually created by scanning extensive indexes that contain the structures of existing or potentially existing chemicals. At this stage, drug-drug similarity algorithms will then analyze drug-protein affinity states of drugs with high similarity.

2.1. Dataset

The drug in the data set and the drug structure used in the study were used as Simplified Molecular-Input Line-Entry System (SMILES). SMILES is a sequence representation of the 2D structure of the molecule. It matches any molecule (usually) to a unique particular string that matches back to the 2D structure. Sometimes, different molecules may be matched to the exact SMILES string, which can degrade the model's performance. In this study, the dataset obtained from the public databases DrugBank and PubChem was used to determine the similarities of selected FDA-approved drugs with COVID-19 and lung cancer drugs and analyze the affinity binding of these drugs. The DrugBank database combines detailed drug data with comprehensive drug target information. This database is a source for bioinformatics and chemical informatics. DrugBank Online contains 14.556 drug entries. Of these, 131 are nutraceuticals, 2.698 are approved small molecule drugs, 1.473 are approved biologics (proteins, peptides, vaccines and allergens), and more than 6.653 experimental drugs. PubChem is the knowledge base of chemical molecules. Structure and descriptive information of millions of compounds can be accessed here. The system is operated by the National Center for Biotechnology. This center is agency of the National Library of Medicine affiliated with the U.S. Department of Health. PubChem has 110.025.926 chemical structures and 96.561 protein targets. It also has 32.816.125 scientific publications linked to PubChem, 29.940.379 patents linked to PubChem and 803 organizations that contribute data to PubChem. DrugBank is a reliable database containing drug information such as drug targets, drug enzymes, drug interactions and drug carriers. PubChem is a database for drug structures. For similarity analysis, lung cancer drugs, FDA-approved drugs and drugs which were previously used in the treatment of Sars-Cov and Mers-Cov disease and currently used in clinical trials for Sars-Cov-2 were used. Table 1 shows information on the compound numbers of the drugs used.

Drug repositioning, a strategy of this study, can be considered as a valid alternative provided that the drug has been used frequently clinically. In particular, a remarkable number of drugs reconsidered for the treatment of COVID-19 are being used in cancer treatment. This is because the infected cells are forced to increase nucleic acid, protein and lipid synthesis and increase their energy metabolism in order to adapt to the 'viral program'. The same features are also seen in cancer cells. This makes it possible that drugs that interfere with specific cancer cell pathways may also be effective at defeating viral replication. COVID-19 can affect many organs such as the brain, kidney, liver, especially the lungs. The most affected organ involvement, which has an effect on mortality, is the lung. For this reason, since the treatment of COVID-19, which causes respiratory syndrome, is provided with drugs used to treat the symptoms of lung disease, drugs used in lung cancer were used in this study in order to prepare a list of the repurposed drugs.

2.2. Proposed method

The fact that the treatments for COVID-19 disease are drugs designed to treat the symptoms of lung disease justifies the re-use of FDA-approved drugs. This study aims to determine the similarities between FDA-approved drugs and COVID-19 and lung cancer drugs and help fight against the pandemic as quickly as possible by using these drugs that are repurposed for the rapidly spreading COVID-19 virus with molecular affinity binding mode analysis. For this purpose, the purposed method for determining repurposed drug candidates is shown in Figure 1.

The proposed method consists of two stages.

In the first stage, drug-drug similarity analysis was performed. For this process, drugs used in the treatment of COVID-19, FDA-approved drugs and drugs used in lung cancer were used. Bioinformatics analyses were performed to make a list of repurposed drug candidates. Since there is limited information about COVID-19, we focused our studies on similar pathogens and compared FDA-approved drugs with drugs used in the treatment of COVID-19, as shown in Figure 1, and made a list of similar reusable drugs. We then calculated the extended connectivity fingerprints for each drug compound using the Jaccard similarity coefficient. We calculated all binary chemical similarities using Tanimoto similarity, which has been proven to be a suitable choice for fingerprint-based similarity calculations. The drugs obtained were compared with the drugs used in lung cancer and analyzed by the Atom pair similarity method. Thus, a list of similar drugs used in both lung cancer and treatment of COVID-19 and approved by the FDA was obtained. This list of similar drugs was used in the next step.

In the second stage, Drug-protein Affinity analysis was performed to calculate the binding mode and affinity of drugs obtained from drug-drug similarity analysis. At this stage, the drug list obtained in the first stage and the Davis dataset specified in Table 1 were used. The Davis dataset includes the target-protein sequence of the drugs. Other methods, such as matrix factorization methods used for similar matrix completion problems, need to be retrained each time a new drug or target is added to a dataset. To overcome this problem, the GNN-based model was used. The type of GNN used is inductive in nature. This means that it can be used to make predictions about targets and compounds not seen during training, without repeating the training process. Specifically, a drug is modeled as a graph of

Table 1. Drug related public databases.

Databases	Drug properties	URL	The number of compounds
Lung Cancer Drugs	Molecular structure	https://go.drugbank.com/ https://pubchem.ncbi.nlm.nih.gov/	401
FDA Approved	Molecular structure	https://go.drugbank.com/ https://pubchem.ncbi.nlm.nih.gov/	294 662
Sars-Cov-2 M^{pro}	Molecular structure	https://go.drugbank.com/	101
Davis dataset	Molecular structure, Target	https://go.drugbank.com/	20 000

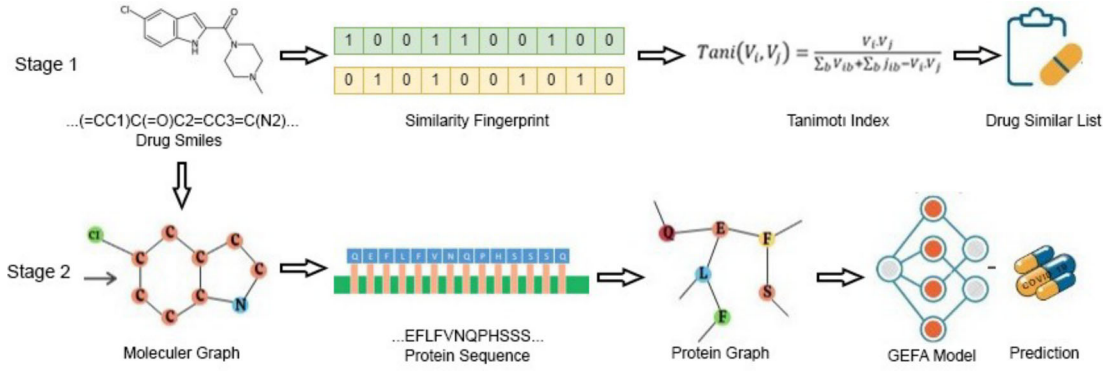

Figure 1. Block diagram of the proposed approach to identifying repurposed drug candidates.

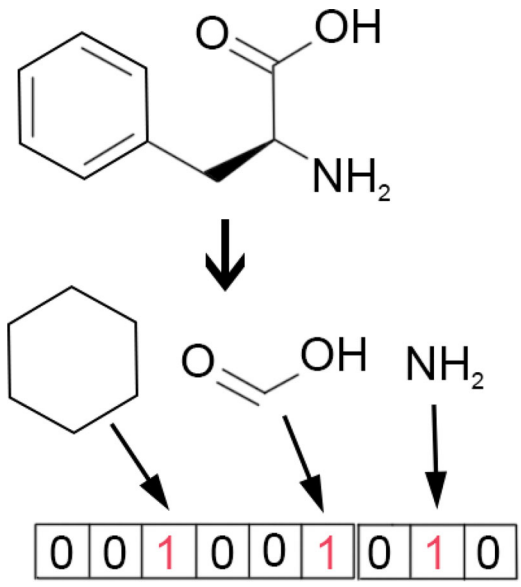

Figure 2. Molecular fingerprint structure.

atoms acting as a node in a larger graph of residues-drug complex. The resulting model is a meaningful deeply nested graph neural network. Trainings are conducted under different settings to evaluate scenarios such as new drugs or targets. The Dissociation constant (KD) parameter was used to calculate the Affinity value. As indicated in the flowchart of the proposed method, the combined method for drug-drug affinity (Atom pair, Tanimoto) and drug-target protein interaction (GNN) predicts the binding affinity of samples (ligands) with a particular molecular target.

2.3. Drug-drug similarity analysis

Clinical drug-drug similarity, which is associated with chemical similarity and drug similarity based on the literature, has many potential applications in evaluating drug therapy similarity and patient similarity. Chemical structures often represent features by inferring molecular fingerprints where structural features are converted into bits in a bit vector or numbers in a count vector. This representation enables to efficiently examine and compare of calculation of chemical structures. By using such fingerprints, the similarity between two molecules is extracted. The molecular fingerprint structure is shown in Figure 2.

A bit position is associated with a precisely predefined feature. '1' refers to the presence of a feature in the molecule and '0' refers to the absence of a feature in the molecule. If drug-drug interaction, one of the most critical issues in drug development and health, is set to zero, this suggests that there is no evidence of their interactions yet. Hence, they can interact with each other.

$$J(di, dj) = \frac{|di \cap dj|}{|di \cup dj|} \quad (1)$$

$$dj(di, dj) = 1 - J(di, dj) \quad (2)$$

In the study, the drug-drug distance was measured using the Jaccard similarity coefficient. This similarity coefficient performed in a significant drug-diagnosis relationship. To do this, the values were converted to binary bits and significant inputs were set to 1 and non-significant inputs to 0. The Jaccard similarity coefficient, which can be calculated as the valued bit rate of both drugs, can be calculated as the rate of bits with both drugs having a value of 1 in the same diagnosis code among drugs where at least one drug is 1, as specified in Equation 1. In this study, the drug-drug distance was calculated as specified in Equation 2.

2.3.1. Tanimoto similarity analysis

The concept of fingerprint is a way of representing the structure and properties of molecules with a binary (0 and 1) number system. This representation way was applied for data scanning operations. Many fingerprint algorithms and similarity analyses based on this algorithm are available. The mathematical formula of Tanimoto (Rogers & Tanimoto,

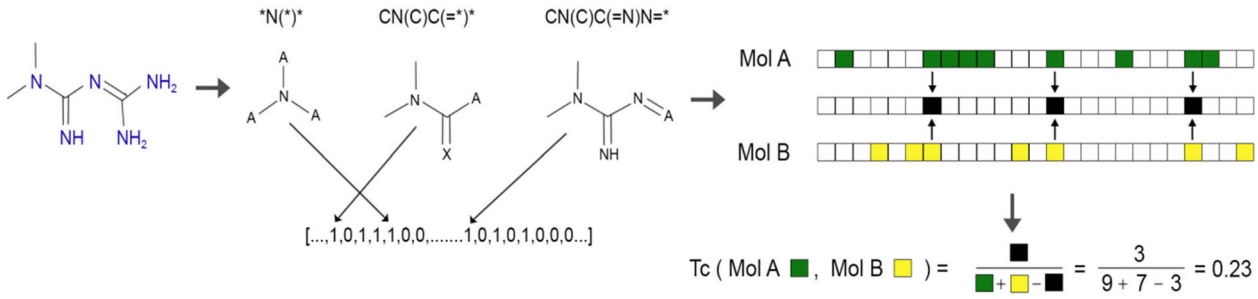

Figure 3. The structure of Tanimoto Algorithm.

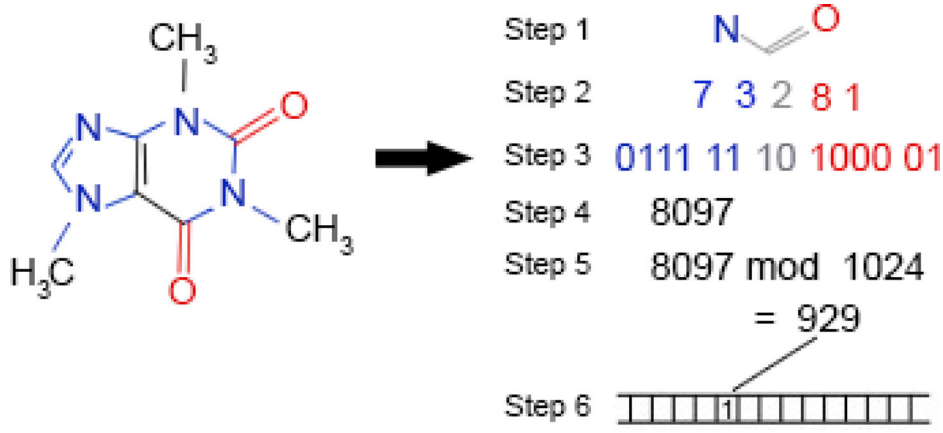

Figure 4. Fingerprint structure of Atom pair.

1960), which is the most widely used similarity algorithm, is as shown in Equation 3.

$$Tan\,i(V_i, V_j) = \frac{V_i.V_j}{\sum_b V_{ib} + \sum_b j_{ib} - V_i.V_j} \tag{3}$$

where V_i represents molecule A and V_j represents molecule B. The function here is a random number generator and applies to each property of a molecule, such as the existing bond and molecular structure types. Figure 3 shows the structure of the Tanimoto algorithm.

Figure 3 shows how the similarities of the two molecules can be assessed by creating molecular fingerprints. Each molecule has a hash function and then a fingerprint is generated according to the properties. The fingerprint generator in Figure 3 looks at a certain radius of bond distance and properties within that distance.

2.3.2. Atom pair similarity analysis

Representation of molecular (chemical) structures as sequences of bits (1's and 0's) is made with molecular fingerprints. The basic logic is to capture the structure information of a graph molecule and then encode it in a bit sequence to be used in assessment stage its similarity with a pair of compounds. The critical advantage of this process is that the storage of such a representation is very different. Comparing two molecular graphs proceeds rapidly compared to bit comparison, reducing time-consuming. We decided to use the best-performing Atomic Pair (AP) fingerprint (Riniker & Landrum, 2013) for coding of structural elements and it is relatively easier to implement. An AP configuration formula is as shown in Equation 4:

$$A = s\left[1 - \sum_{i=1}^{m}\left(\frac{\tau_i}{\sum \tau_i}.\frac{\Delta n_i}{n_i}\right)\right] \tag{4}$$

A is the activity relative to the standard drug, S is the similarity between the drug and its analog, τ_i is the pathway length of the atom pair, Δn_i is the difference of the standard drug from the analog of the i.heteroatom pair, and n_i is the number of the i.heteroatom pair in the standard drug. The general process is as outlined in Figure 4.

The process steps shown in Figure 4 are as follows. When creating an atom pair fingerprint, the following steps are performed for each pair of heavy atoms:

1. removing the given pair of atoms and the shortest pathway between them;
2. coding of identifiers (atomic type and the number of bonds for both atoms and their topological distance);
3. converting to bit strings;
4. combining the bit strings into a number;
5. hashing the number into the index field;
6. setting the corresponding position on the fingerprint as 1.

The primary rationale for this similarity is colored depending on how many of the bits set by the atom are present in the fingerprint. Figure 4 shows the 'weight' of an atom being normalized and the normalized weights then being used to color the atoms in a topography-like map; green indicates a positive difference (i.e. similarity or probability decreases when bits are removed), pink indicates a negative difference, and gray indicates no change. Visualization is shown for fingerprint types of atom pairs.

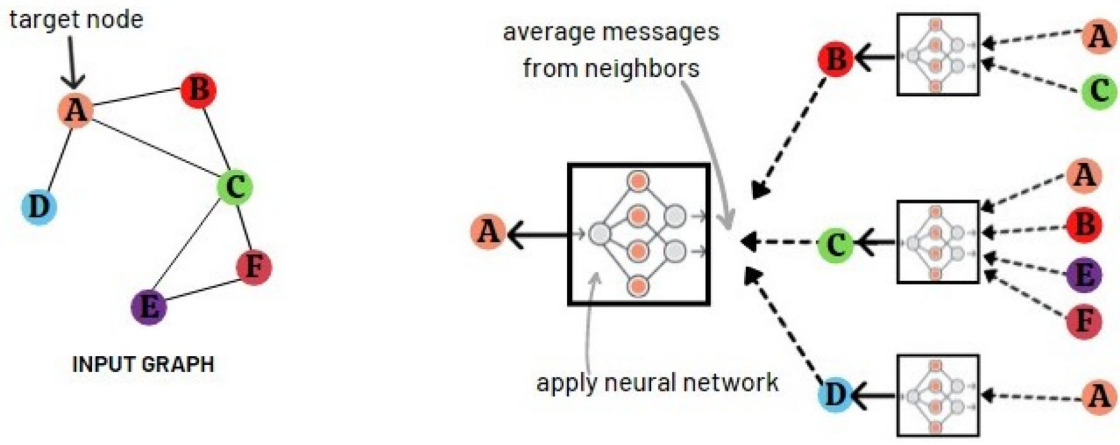

Figure 5. GNN structure.

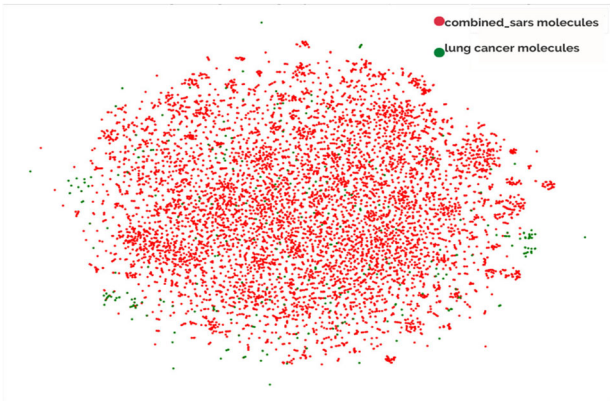

Figure 6. Similarity analysis of Sars drugs and lung cancer drugs combined using fingerprints with Tanimoto/Jaccard similarity method.

2.4. Drug-protein a finite analysis

The binding of drugs to proteins in blood, serum, or plasma has an important role in determining the activity states in the body, their distribution, the rate of excretion, and toxicity. With the development of GNN, recent studies on drug discovery have focused on using direct molecular graphic representation for both feature prediction and innovo design. The GEFA model was used in this study for Drug-protein Affinity analysis, which is very important in rapid drug reuse. Here, the drug is modeled as an atomic graph and then acts as a node in the residual-drug graph in later stages. The GEFA model used consists of four different training settings. Where both the protein and the drug are known by the model, only the proteins are known by the model, only the drug is known by the model, and finally both the protein and the drug are not known by the model. Drug-target affinity values were determined in the GEFA model according to the experimental setting in which the proteins are known, and the drug data set, which is obtained from the results of drug-drug similarity analysis and predicts its usability for Sars-Cov-2 treatment. Rational application of measurements that quantify the molecular properties required to obtain binding affinity accelerates the selection of fragments and hits. Since the outputs of networks are determined by the quantitative interaction of the many molecules and interactions that compose them, equilibrium constants are needed for the relationship between network components. For binding equilibrium, under conditions where one binding partner (here, protein, P) is too much than the other (RNA), the rate equation for approaching equilibrium is as follows:

$$k_{\text{equil}} = k_{\text{on}}[P] + k_{\text{off}}, \tag{5}$$

where k_{on} is the association rate constant, [P] is the protein concentration or expression of excessive binding, k_{off} is the dissociation rate constant. Since equilibration in Equation 5 is slowest at the lowest protein concentrations, equilibration times need to be established from the lower end of the concentration range. In practice, it is helpful to think of the limiting case where the protein concentration approaches zero ([P] $\sim$ 0) such that Equation 1 simplifies Equation 6.

$$k_{equil,\,limit} = k_{off} \tag{6}$$

The lower the Dissociation constant (KD), which expresses the binding strength of drug targets, the longer the incubation time required to reach equilibrium (Jarmoskaite et al., 2020).

$$K_D = \frac{k_{off}}{k_{on}} \tag{7}$$

The final measure of how well a compound binds a target is called as the dissociation constant KD is a real-numbered measurement used to represent the binding affinity between a drug and a target. The lower the KD value, the greater the binding. Therefore, this metric was used at this stage.

2.5. Graph Neural Network

In recent years, there has been an increase in the number of deep learning applications. Some deep learning applications are as follows: quantitative structure-activity relationships (QSAR), virtual screening, drug repositioning and in silico studies, prediction of pharmacokinetic properties (absorption, distribution, metabolism and excretion- and toxicity). Biological and chemical data have some features such as complexity, uncertainty, diversity and high dimensionality. The main advantage of deep learning is the complexity of

Table 2. The similarity values of combined Sars drugs and lung cancer drugs.

Lung cancer molecule	Combined Sars drugs	Similarity rate
Erlotinib	N-(3-Chlorophenyl)-6,7-dimethoxyquinazolin-4-amine hydrochloride	0.8013
Dasatinib	PubChem CID: 24816490	0.5520
Vandetanib	N-(3-Chlorophenyl)-6,7-dimethoxyquinazolin-4-amine hydrochloride	0.7342
Lapatinib	PubChem CID: 666049	0.5468
Gemcitabine	2′-Deoxycytidine hydrochloride	0.7420
Napabucasin	3-Acetyl-2-methylnaphtho[2,3-b]furan-4,9-dione	0.7822
Diflomotecan	Camptothecin	0.8784
Bortezomib	2-Acetylamino-3-phenyl-N-[1-(pyrazine-2-carbonyl)-piperidin-4-yl]-propionamide	0.6132
Acetaminophen	Phenacetin	0.7671
P-Toluenesulfonamide	Mafenide acetate	0.8193
Pomalidomide	Thalidomide	0.8180

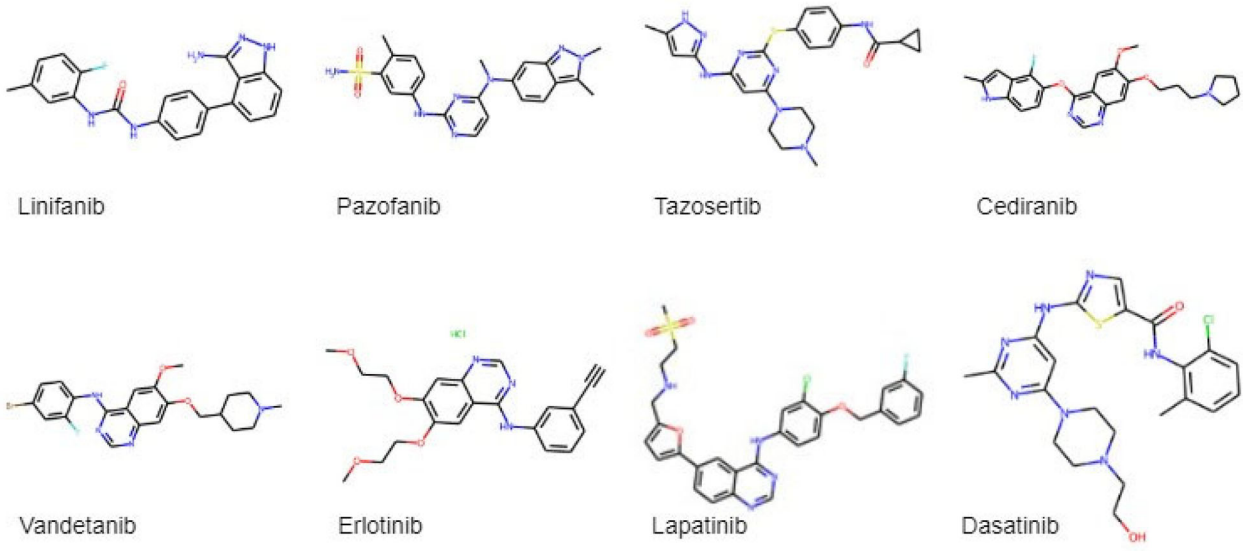

Figure 7. 2D molecular structures of drugs obtained by atom pair analysis.

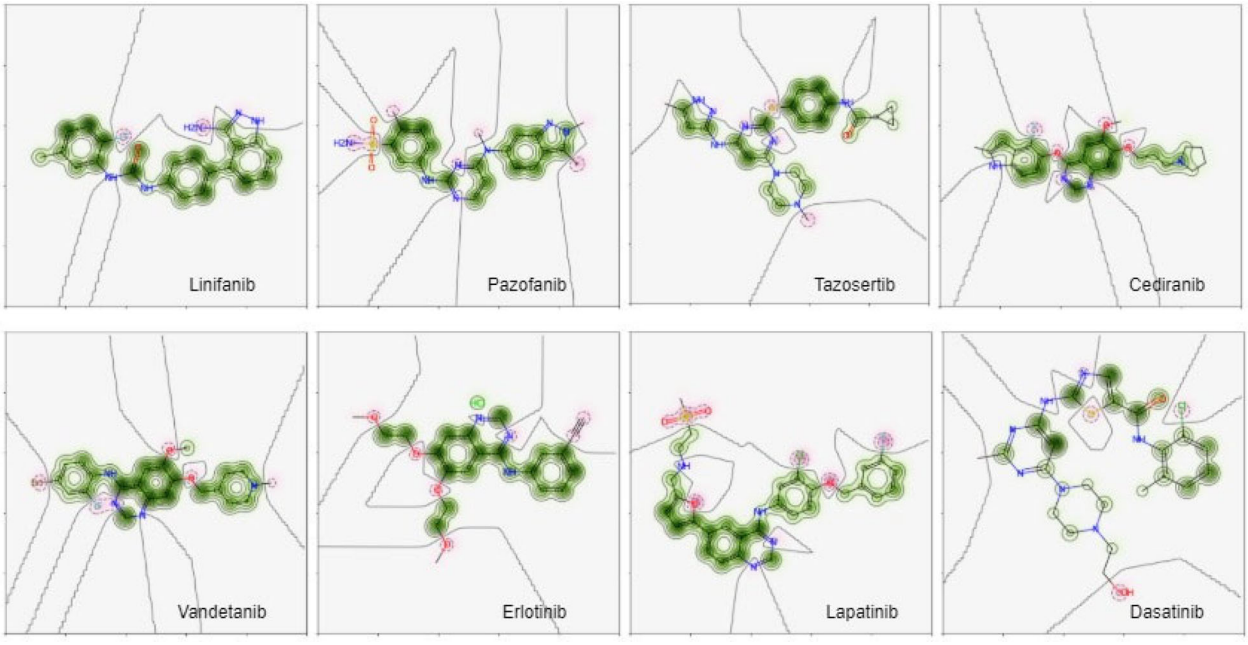

Figure 8. Atom pair similarity results.

the neural networks used and the flexible nature of their architecture that allows for adaptations to specific problems. Deep learning is effective in processing large chemical libraries to provide predictive computational models. Therefore, researchers use this methodology to find new chemotherapeutic agents in drug discovery. Deep learning performs well in predicting protein-ligand binding affinities. Deep learning scoring algorithms are GNNs that spatially and/or chemically

encode neighboring ligands and receptor atoms. Figure 5 shows the graph neural network structure (GNN) structure.

GNN aims to learn the representations of each atom. While doing this process, the atom combines the information from neighboring atoms encoded by the feature vector with the information of the bonds encoded by the bond feature vector. The state update of central atoms and the atom representations learned after reading can be used to predict molecular properties. Automatic learning of task-specific representations using graph convolutions without the need for fingerprints is an important feature of GNN. Computational methods created for the prediction of molecular properties, which is one of the main tasks in the field of drug discovery, can accelerate the process of finding better drug candidates quickly and cheaply. Let a graph be defined as G = (V, E). Where V denotes nodes and E denotes edges. The molecule can be thought of as a graph of nodes and edges. a € V is a node with feature vector x_a and ve Let b_{ua} € E be an edge point from u to a with the feature vector x_{ua}. The adjacency matrix, A, shows the connectivity of the nodes. Here this matrix is binary if the graph is not weighted. It is defined as a n × n matrix with Aua = 1 if euv∈E and Aua = 0 if $x_{ua} \notin$ E. The symmetrically normalized neighborhood matrix is defined as: $A_{sym} = D^{-1/2}AD^{-1/2}$. Where D is the degree matrix and is defined as: D∈$\mathbb{Z}$ ▫ V ▫ x ▫ V ▫ . Molecular graphs are generally undirected, weightless and often heterogeneous (Wieder et al., 2020).

3. Experimental results and discussion

In this study, two different methods were used: the Tanimoto algorithm and the Atom Pair algorithm. Affinity binding mode analysis was performed for re-use of existing useful drugs as COVID-19 therapeutics. In the first step of the proposed method, the Tanimoto similarity results for drug-drug similarity are shown in Figure 6.

Figure 6 shows the distribution of the drug molecules closest to each other according to the affinity distance according to the Tanimoto/Jaccard algorithm logic. Mers-Cov shown in red shows drugs used for Sars-Cov treatment and combined drugs from drugs approved in clinical trials and so far to treat Sars-Cov-2. Green-colored ones indicate the drugs used in lung cancer. Table 2 shows similarity values of these drugs.

The drug–drug distance calculated as indicated in Equation 2 shows how close the drug molecules are to each other, as seen in Figure 6. Very close to each other, almost similar drugs overlap. This similarity sheds light on the use of drugs used in lung cancer for treatment of infectious disease and COVID-19 disease that causes the acute respiratory syndrome. This is the first stage of the study and was used for the main result in the second stage. Figure 7 shows structures of Atom Pair 2D smiles of some drugs selected among sars drugs and drugs used in lung cancer combined with the Tanimoto/Jaccard similarity method.

As shown in Figure 7, the molecular structures of drugs obtained by Tanimoto/Jaccard similarity analysis are shown in 2D Smiles using AP fingerprint analysis, which indicates how the fragments should be encoded into strings.

The Atomic Pair fingerprints algorithm is also a demonstration of how fragments should be encoded into strings. The basic idea at this stage is to consider the structural features in molecular fingerprints (these are bond structure, atomic type, etc.) and take the values of each atom of a particular fragment. These are then encoded into a finite number of bits (for example, three bits are sufficient for the number of links) and combined to form the bit representation of the fragment index of the structure. The basic logic of the AP fingerprint structure is as follows:

1. Remove all atom pair fragments
2. Encode parts to integers (indexes)
3. Create a bit sequence of length *n*
4. Add hashes of indexes to a field of the bit string
5. Turn on the corresponding bit for each of the hash indexes,

In other words, the bits corresponding to the atom pairs in the molecule are turned on and the remaining bits are turned off. Figure 8 shows the atomic pair similarity results of the drugs obtained.

Coloring is done in these structures as specified in Section 2.3.2. In this context, in the Atom pair similarity results in Figure 8;

- Green indicates a positive difference (i.e. similarity or probability decreases when bits are removed)
- Pink indicates a negative difference,
- Gray indicates no change.

Green indicates how well the combined Sars drugs shown in Figure 8 overlap with the drugs used in lung cancer. In other words, the green color seen here indicates how similar the two drugs are. Similarity maps are an useful and easy-to-understand strategy for atomically visualizing fingerprint similarity between molecular structures. Atomic weights are generated by comparing the similarity resulting from removing bits belonging to the corresponding atom with the (unmodified) similarity of the previous fingerprint. Similarity maps can be created for each fingerprint, allowing bits to be traced back to a corresponding atom or substructure.

There are numerous protein kinases that can be blocked by FDA-approved drugs to target the viral life cycle in the COVID-19 outbreak and alleviate the symptoms of life-threatening lung-damaging infection. These inhibitors offer an attractive option for reuse, as they have been extensively studied for safety and are more easily available for treatment of patients and testing in clinical trials. Numerous kinases have been proposed as important mediators of Mers-Cov, Sars-Cov and various viral infections. It is estimated that these same proteins play a role in mediating the infection that causes Sars-Cov-2. Protein kinases, which make up 20%–30% of the drug discovery plans of major pharmaceutical companies, have become a very suitable target at this stage. There are many kinase inhibitors currently approved with pharmacological effects that may

2D Smiles	3D Conformer	Target Protein	Affinity
Erlotinib	Erlotinib	AAK1	5.0
Vandetanib	Vandetanib	AAK1	5.92081
Dasatinib	Dasatinib	ABL2	9.76955
Cediranib	Cediranib	JAK1	5.0
Lapatinib	Lapatinib	EGFR	8.92081
Linifanib	Linifanib	FLT3	7.95860
Pazopanib	Pazopanib	AAK1	5.53760
Tozasertib	Tozasertib	ABL1	8.10790

Figure 9. The binding values of kinase inhibitors to the target protein obtained by GEFA analysis.

be useful in improving the severe and potentially life-threatening symptoms of COVID-19, such as anti-inflammatory activity, cytokine suppression, and antifibrotic activity. Ideally, a kinase inhibitor with optimal pharmacokinetic properties could be reused as a dual-functional therapeutic that can reduce infection through direct viral targeting and also provide clinical

Table 3. Methods used in the literature for drug discovery.

Authors	Simulated tissue/purpose of the study	Used method	Application
Stebbing et al. (2020)	COVID-19: Combining antiviral and anti-inflammatory treatments	Artificial Intelligence Methods	Drug repurposing, protein affinity
Wang et al. (2020)	Drug repurposing report generation	Artificial Intelligence	Literature search, report
Domingo-Fernandez et al. (2020)	COVID-19 Knowledge Graph	BIKMI, OrientDB, Python Django	Drug repurposing, protein affinity, host-pathogen interactions
Hsieh et al. (2020)	Repurposable drugs by multiple SARS-CoV-2 and drug interactions	Graph Neural Network (GNN)	Drug repurposing, protein affinity
Zhou et al. (2020)	Artificial intelligence in COVID-19 drug repurposing	Artificial Intelligence	Drug repurposing

benefit by suppressing disease symptoms. Kinase inhibitors are types of proteins that are very commonly used as drug targets. Kinase inhibitors were used for drug-protein affinity analysis as they are drugs that bind and inhibit the activity of a kinase. Figure 9 shows the target protein binding values of the kinase inhibitors.

An ideal drug candidate would be a compound that has high binding affinity (low KD) with the desired target, but low binding affinity (high KD) with all other known biological targets. This minimizes the risk of the drug interacting with other targets and causing unwanted side effects. We may see some drugs emerge more than once, these would potentially be the most promising treatments as they can serve as multi-targeted inhibitors. What is promising is that many of the findings have already been suggested or confirmed in the literature. This is encouraging because it confirms our results.

In the proposed method, the listed drug smiles structures obtained by stage1 were trained with kinase inhibitors in the Davis dataset under various training settings. Here, target-protein analysis was performed with the kinase inhibitors in each smiles davis dataset and the KD value specified in Section 2.4, and the binding rate with all kinase inhibitors was obtained separately for each drug. The kinase inhibitor with a better binding rate than other inhibitors was selected. For example, for Cediranib, the JAK1 inhibitor yielded a higher binding rate than the others with a KD value of 5.0 among the kinase inhibitors. Similarly, Dasatinib achieved better binding to the ABL2 kinase inhibitor with a KD value of 9.7 compared to the other inhibitors (see Figure 9). Kinase inhibitors approved for treating various malignancies have properties such as anti-inflammatory and cytokine inhibitory activity that can reduce the likelihood of life-threatening conditions associated with lung damage caused by respiratory virus infections. For example, Osimertinib is a potent EGFR inhibitor. It has been reported that Osimertinib is one of 24 FDA-approved drugs showing in vitro activity against Sars-Cov-2. Lapatinib may also be recommended. According to this table, promising ones are used in treatment of patients suffering from Lung disease and Acute Myeloid Leukemia (AML). Among all results, the most promising EGFR (Lapatinib), AAK1 (Erlotinib, Vandetanib, Pazopanib), ABL2, and ABL1 (Dasatinib, Tozasertib), FLT3 (Linifanib), and JAK1 (Cediranib) can be used as an alternative for treatment of COVID-19.

3.1. Discussion

In the context of drug development, reuse of existing drugs for another disease may be faster than new drug discovery. When existing drugs are approved for other diseases, they can provide new treatments faster, but there are many factors to consider during reuse of approved drugs for a new indication. Identifying potentially inhibited key protein targets may be a good option for a new therapeutic application. However, the urgency of need and time during a pandemic crisis can make it challenging to conduct well-controlled studies with data that attribute efficacy to a drug. Various kinase inhibitors that target virus-associated proteins are under clinical investigation for COVID-19. Several vaccines are available to fight against the COVID-19 pandemic, but dealing with the pandemic is still challenging due to the emergence of mutant strains of the virus as well as difficulties in generating and distributing vaccines, and more. When the drug targets and protein structures associated with the disease of interest are known, it is possible to use structural bioinformatics to screen available drugs against these known targets using molecular docking. In this context, the toxicity profiles of known preclinical, pharmacokinetic, pharmacodynamic, and repurposed drugs can be an important solution to find rapid treatment against COVID-19 disease. The affinity of the Erlotinib drug, which we determined, was analyzed for AKT1 and can be used as an antiviral drug in the treatment of COVID-19. The most appropriate treatment for patients with newly diagnosed acute myeloid leukemia (AML) infected with severe acute respiratory syndrome Sars-Cov-2 is unknown. It has been concluded that single-agent Gilteritinib can be safely administered in patients applying due to de novo FLT3-ITD positive AML and may cause remission. Gilteritinib was considered to be a treatment option for patients with FLT3 mutated AML and severe COVID-19, where long-term chemotherapy-induced pancytopenia could adversely affect outcomes (Wilson et al., 2020). We believe that the drug Linifanib, whose affinity is obtained from drug-target similarities, such as Gilteritinib, an inhibitor to FLT3, may be effective in AML patients infected with COVID-19.

Long-term or permanent lung damage in the form of pulmonary fibrosis, a process mediated by Epidermal Growth Factor (EGFR), has been observed in patients recovering from Sars-Cov and Mers-Cov infections. EGFR-targeting inhibitors used by different viruses, including many respiratory viruses, have been observed by gefitinib and erlotinib (Lupberger et al., 2011). For example, Osimertinib is a potent EGFR inhibitor and has been reported to be one of 24 FDA-approved drugs that exhibit in vitro activity against Sars-Cov-2. Another inhibitor, lapatinib, which targets EGFR, can be used as an alternative for treatment of COVID-19. The kinase inhibitors (erlotinib, lapatinib, vandetanib, pazopanib,

cediranib, dasatinib, linifanib and tozasertib), obtained in the study can be used as an alternative for the treatment of COVID-19 as a combination of blocking agents such as ABL2, ABL1, EGFR, AAK1, FLT3 and JAK1 (gefitinib, osimertinib, fedratinib, baniticinib, imatinib, sunitinib and ponatinib) or antiviral treatments (ribavirin, ritonavir-lopinavir and remdesivir). Table 3 shows the methods recommended for the treatment of COVID-19 in the literature.

Stebbing et al. (2020) is a review article focusing on a proprietary AI algorithm. In this article, the authors were published as a commentary in the Lancet Infectious Diseases immediately before the onset of the pandemic. The authors previously described how BenevolentAI's proprietary Knowledge Graph (KG) questioned by a set of algorithms enabled the identification of a numb-associated kinase (NAK) inhibitor (baricitinib) to suppress clathrin-mediated endocytosis and thereby to prevent viral infection of cells. In this study, they examined again the affinity and selectivity of all approved drugs in their kg to identify those with both antiviral and anti-inflammatory properties, as the host inflammatory response has become a major cause of lung injury and then mortality for severe cases of COVID-19. The authors gave these three candidates: baricitinib, fedratinib and ruxolitinib.

The study of Wang et al. (2020) is the only article that mentions the creation of a report for drug reuse. The content of such an AI-generated drug report, mentioned in Wang et al. (2020) of their study, is useful for understanding why a drug reuse candidate was selected. The authors note that reports are reviewed by clinicians and medical students, however a more quantitative assessment can be made at a later stage. It is also unique in using shape images from publications to enrich article KG. The authors used the KGs to generate the drug reuse report. Such a report for a particular drug consisted of 11 typical questions they identified:

Domingo-Fernandez et al. (2020) created a KG that is a cause-and-effect knowledge model of the pathophysiology of COVID-19 and can then be applied for drug reuse. The authors noted that although KGs were originally developed to describe interactions between entities, new machine learning techniques can produce hidden, low-dimensional representations of KG that can then be used for downstream tasks such as clustering or classification. For the creation of the KG, scientific literature on COVID-19 was taken from open-access and freely available journals: additional COVID-19-specific reviews such as PubMed, Europe PMC and LitCovid. This corpus was then filtered based on available knowledge of potential drug targets for COVID-19, the biological pathways by which the virus interferes with replication in the human host, and information about various viral proteins along with their functions.

Hsieh et al. (2020) aimed to discover reusable drugs by integrating multiple Sars-Cov-2 and drug interactions, deep graph neural networks, and in vitro/population-based validations. They collected all available drugs ($n = 3635$) involved in COVID-19 treatment through the Comparative Toxicogenomics database. Candidate drugs can be divided into two broad categories: those that can directly target the virus replication cycle and those that rely on immunotherapy approaches aimed at enhancing innate antiviral immune responses or attenuating damage caused by dysregulated inflammatory responses. They created a Sars-Cov-2 KG based on interactions between virus baits, host genes, drugs, and phenotypes. The graph had four types of nodes and five types of edges based on interactions. They used a GNN approach to obtain a representation of the candidate drug based on biological interactions. To validate their approaches, they explained that in traditional network analysis, network proximity was defined by direct interactions, thus less attention was paid to a node's local role (e.g. neighbors, edge directions) and global location (e.g. overall topology or structure).

Zhou et al. (2020) conducted a review article for the Lancet Digital Health. In their review, they provided guidelines on how to use various forms of AI to accelerate drug reuse, with COVID-19 as an example. With regard to KGs in particular, they stated that KGs can be reduced to low-dimensional feature vectors and their similarity can be measured using feature vectors of drugs and diseases, thus effective drugs can be identified for a particular disease. One challenge they identified for the chart embedding method is scalability. The number of assets in a medical KG can be as many as several million. They mentioned that several systems were specifically designed to learn representations from large-scale graphs.

In this study, we performed bioinformatics analyses to make a list of repurposed drug candidates again by completing the limited information known about the COVID-19 virus, which has caused millions of deaths around the world, with data on associated viruses. Since there is limited information about COVID-19, we focused our studies on similar pathogens and prepared a list of similar reusable drugs by comparing drugs used in lung cancer with FDA-approved drugs used in the treatment of COVID-19. Graph neural network-based GEFA model was used to calculate the affinity of these drugs. Predicting the interaction between a compound and a target is crucial for rapid drug reuse. Deep learning has been successfully applied to the drug target affinity (DTA) problem. However, previous deep learning-based methods overlooked modeling direct interactions between drug and protein residues. This will lead to incorrect learning of target representation, which may change due to drug binding effects. Also, previous DTA methods neglect the use of proteins outside of DTA datasets. These methods learn protein representation based on a small number of protein sequences only in DTA datasets. GEFA was chosen in this study as it is a graph-in-graph neural network with attention mechanism to examine changes in target representation due to binding effects. Thus, the use of existing approved drugs can provide a faster treatment compared to the long time required for the discovery of a new drug in the fight against the pandemic. The use of repurposed drugs for the rapidly spreading COVID-19 virus can make it easier to help fight the pandemic as quickly as possible during the pandemic crisis, where need and urgency of time are important.

4. Conclusion

It takes an average of ten years and two billion dollars to develop a single FDA-approved drug. However, it is too slow to react to sudden global threats like the COVID-19 pandemic. COVID-19 vaccines are already distributed, but global vaccine distribution cannot be realized within a day. And in spite of all the challenges of vaccine delivery, hundreds of thousands of people worldwide have still suffered from the Sars-Cov-2 virus. An alternative to discover a new drug is to reuse the drug. It is to find drugs that have passed clinical trials for the treatment of other diseases and can be effective in the treatment of the new one. In this way, it remains only to test the drug's ability to treat a new disease, as the safety risk is already known. In this article, the GNN-based GEFA model was used to find new potential drug candidates that have already undergone clinical trials for the treatment of Sars-Cov-2. In this context, since there is limited information about COVID-19, we focused our studies on similar pathogens and prepared a list of similar reusable drugs by comparing drugs used in lung cancer and FDA-approved drugs used in the treatment of COVID-19 with similarity algorithms such as Tanimoto and Atom Pair. In order to calculate the binding rate of these drugs, the graph neural network-based GEFA model was applied and the results were evaluated. The drug datasets used in the study were taken from Pubchem and Drugbank. Clinical trials are ongoing for Sars-Cov-2 through various kinase inhibitors that target the main virus-associated proteins besides the proteins involved in the development of Sars-Cov-2- associated symptoms (including pneumonia, fibrosis and inflammation).

Kinase inhibitors (erlotinib, lapatinib, vandetanib, pazopanib, cediranib, dasatinib, linifanib and tozasertib), obtained in the study can be used as an alternative for the treatment of COVID-19 as a combination of blocking agents such as ABL2, ABL1, EGFR, AAK1, FLT3 and JAK1 (gefitinib, osimertinib, fedratinib, baniticinib, imatinib, sunitinib and ponatinib) or antiviral treatments (ribavirin, ritonavir-lopinavir and remdesivir). In conclusion, small molecules identified as and targeting major viral factors that could represent a potential target for future Sars-Cov-2 treatment may provide a basis and research for the design of new classes of treatments against Sars-Cov-2 infection. Through these results, we aimed to shed light on future treatments.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

References

Alhudhaif, A., Polat, K., & Karaman, O. (2021). Determination of COVID-19 pneumonia based on generalized convolutional neural network model from chest X-ray images. *Expert Systems with Applications*, *180*, 115141.

Amin, Sk. A., Ghosh, K., Gayen, S., & Jha, T. (2021). Chemical-informatics approach to COVID-19 drug discovery: Monte Carlo based QSAR, virtual screening and molecular docking study of some *in-house* molecules as papain-like protease (PLpro) inhibitors. *Journal of Biomolecular Structure & Dynamics*, *39*(13), 4764–4773. https://doi.org/10.1080/07391102.2020.1780946

Bibi, N., Farid, A., Gul, S., Ali, J., Amin, F., Kalathiya, U., & Hupp, T. (2021). Drug repositioning against COVID-19: A first line treatment. *Journal of Biomolecular Structure and Dynamics*, 1–15. https://doi.org/10.1080/07391102.2021.1977698

Campillos, M., Kuhn, M., Gavin, A. C., Jensen, L. J., & Bork, P. (2008). Drug target identification using side-effect similarity. *Science*, *321*(5886), 263–266. https://doi.org/10.1126/science.1158140

Cha, K., Kim, M. S., Oh, K., Shin, H., & Yi, G. S. (2014). Drug similarity search based on combined signatures in gene expression profiles. *Healthcare Informatics Research*, *20*(1), 52–60. https://doi.org/10.4258/hir.2014.20.1.52

Chan, J. F.-W., Kok, K.-H., Zhu, Z., Chu, H., To, K. K.-W., Yuan, S., & Yuen, K.-Y. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes & Infections*, *9* (1), 221–236.

Chen, X., Liu, M.-X., & Yan, G.-Y. (2012). Drug-target interaction prediction by random walk on the heterogeneous network. *Molecular Biosystems*, *8*(7), 1970–1978. https://doi.org/10.1039/c2mb00002d

Cheng, A. C., Coleman, R. G., Smyth, K. T., Cao, Q., Soulard, P., Caffrey, D. R., Salzberg, A. C., & Huang, E. S. (2007). Structure-based maximal affinity model predicts small-molecule druggability. *Nature Biotechnology*, *25*(1), 71–75. https://doi.org/10.1038/nbt1273

Cheng, F., & Zhao, Z. (2014). Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. *Journal of the American Medical Informatics Association*, *21*, 278–286.

Cheng, F., Liu, C., Jiang, J., Lu, W., Li, W., Liu, G., Zhou, W., Huang, J., & Tang, Y. (2012). Prediction of drug-target interactions and drug repositioning via network-based inference. *Plos Computational Biology*, *8*(5), e1002503. https://doi.org/10.1371/journal.pcbi.1002503

Domingo-Fernandez, D., Baksi, S., Schultz, B., Gadiya, Y., Karki, R., Raschka, T., Ebeling, C., & Hofmann-Apitius, M. (2020). COVID-19 knowledge graph: A computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology. *Bioinformatics*, *37*(9), 1332–1334.

DrugBank Online. (2021). Retrieved May 9, 2021, from https://go.drugbank.com/drugs

Ferdousi, R., Safdari, R., & Omidi, Y. (2017). Computational prediction of drug–drug interactions based on drugs functional similarities. *Journal of Biomedical Informatics*, *70*, 54–64. https://doi.org/10.1016/j.jbi.2017.04.021

Gottlieb, A., Stein, G. Y., Oron, Y., Ruppin, E., & Sharan, R. (2012). Ind1: A computational framework for inferring drug interactions and their associated recommendations. *Molecular Systems Biology*, *8*(1), 592. https://doi.org/10.1038/msb.2012.26

Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R. K., & Kumar, P. (2021). Artificial intelligence to deep learning: Machine intelligence approach for drug discovery. *Molecular Diversity*, *25*(3), 1–46.

Hsieh, K., Wang, Y., Chen, L., Zhao, Z., Savitz, S., Jiang, X., Tang, J., & Kim, Y. (2020). Drug repurposing for COVID-19 using graph neural network with genetic, mechanistic, and epidemiological validation. arXiv: 2009.10931

Jarmoskaite, I., AlSadhan, I., Vaidyanathan, P. P., & Herschlag, D. (2020). *How to measure and evaluate binding affinities* (Vol. 9, p. e57264). eLife Sciences Publications, Ltd.

Karaman, O. (2021). Boosting performance of transfer learning model for diagnosis of COVID-19 from computer tomography scans. *Süleyman Demirel Üniversitesi Fen Edebiyat Fakültesi Fen Dergisi*, *16*(1), 35–45.

Karaman, O., Alhudhaif, A., & Polat, K. (2021). Development of smart camera systems based on artificial intelligence network for social

distance detection to fight against COVID-19. *Applied Soft Computing*, *110*, 107610. https://doi.org/10.1016/j.asoc.2021.107610

Keiser, M. J., Setola, V., Irwin, J. J., Laggner, C., Abbas, A. I., Hufeisen, S. J., Jensen, N. H., Kuijer, M. B., Matos, R. C., Tran, T. B., Whaley, R., Glennon, R. A., Hert, J., Thomas, K. L. H., Edwards, D. D., Shoichet, B. K., & Roth, B. L. (2009). Predicting new molecular targets for known drugs. *Nature*, *462*(7270), 175–181. https://doi.org/10.1038/nature08506

Lounkine, E., Keiser, M. J., Whitebread, S., Mikhailov, D., Hamon, J., Jenkins, J. L., Lavan, P., Weber, E., Doak, A. K., Côté, S., Shoichet, B. K., & Urban, L. (2012). Large-scale prediction and testing of drug activity on side-effect targets. *Nature*, *486*(7403), 361–367. https://doi.org/10.1038/nature11159

Lu, Y., Guo, Y., & Korhonen, A. (2017). Link prediction in drug-target interactions network using similarity indices. *BMC Bioinformatics*, *18*(1), 39. https://doi.org/10.1186/s12859-017-1460-z

Lupberger, J., Zeisel, M. B., Xiao, F., Thumann, C., Fofana, I., Zona, L., Davis, C., Mee, J., Turek, M., Gorke, S., Royer, C., Fischer, B., Zahid, M. N., Lavillette, D., Fresquet, J., Cosset, F. L., Rothenberg, S. M., Pietschmann, T., Patel, A. H., … Baumert, T. F. (2011). EGFR and EphA2 are host factors for hepatitis C virus entry and possible targets for antiviral therapy. *Nature Medicine*, *17*(5), 589–595. https://doi.org/10.1038/nm.2341

Muzio, G., O'Bray, L., & Borgwardt, K. (2021). Biological network analysis with deep learning. *Briefings in Bioinformatics*, *22*(2), 1515–1530. https://doi.org/10.1093/bib/bbaa257

National Center for Biotechnology Information. (2021). Retrieved May 9, 2021, from https://pubchem.ncbi.nlm.nih.gov/

Nguyen, T. M., Nguyen, T., Le, T. M., & Tran, T. (2021). GEFA: Early fusion approach in drug-target affinity prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *5555*(01), 1–1. https://doi.org/10.1109/TCBB.2021.3094217

Polat, Ç., Karaman, O., Karaman, C., Korkmaz, G., Balcí, M. C., & Kelek, S. E. (2021). COVID-19 diagnosis from chest X-ray images using transfer learning: Enhanced performance by debiasing dataloader. *Journal of X-Ray Science and Technology*, *29*(1), 19–36.

Riniker, S., & Landrum, G. A. (2013). Similarity maps-a visualization strategy for molecular fingerprints and machine-learning methods. *Journal of Cheminformatics*, *5*(1), 1–7. https://doi.org/10.1186/1758-2946-5-43

Rogers, D. J., & Tanimoto, T. T. (1960). A computer program for classifying plants. *Science*, *132*(3434), 1115–1118. https://doi.org/10.1126/science.132.3434.1115

Sawada, R., Iwata, H., Mizutani, S., & Yamanishi, Y. (2015). Target-based drug repositioning using large-scale chemical-protein interactome data. *Journal of Chemical Information and Modeling*, *55*(12), 2717–2730.

Simmons, G., Reeves, J. D., Rennekamp, A. J., Amberg, S. M., Piefer, A. J., & Bates, P. (2004). Characterization of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) spike glycoprotein-mediated viral entry. *Proceedings of the National Academy of Sciences*, *101*(12), 4240–4245. https://doi.org/10.1073/pnas.0306446101

Stebbing, J., Phelan, A., Griffin, I., Tucker, C., Oechsle, O., Smith, D., & Richardson, P. (2020). COVID-19: Combining antiviral and anti-inflammatory treatments. *The Lancet. Infectious Diseases*, *20*(4), 400–402. https://doi.org/10.1016/S1473-3099(20)30132-8

Tatonetti, N. P., Ye, P. P., Daneshjou, R., & Altman, R. B. (2012). Data-driven prediction of drug effects and interactions. *Science Translational Medicine*, *4*(125), 125ra31. https://doi.org/10.1126/scitranslmed.3003377

The WHO MERS-CoV Research Group. (2013). State of knowledge and data gaps of Middle East respiratory syndrome coronavirus (MERS-CoV) in humans. *Plos Currents*, *5*.

van Laarhoven, T., Nabuurs, S. B., & Marchiori, E. (2011). Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics*, *27*(21), 3036–3043. https://doi.org/10.1093/bioinformatics/btr500

Velavan, T. P., & Meyer, C. G. (2020). The COVID-19 epidemic. *Tropical Medicine & International Health*, *25*(3), 278–280. https://doi.org/10.1111/tmi.13383

Vilar, S., Harpaz, R., Uriarte, E., Santana, L., Rabadan, R., & Friedman, C. (2012). Drug-drug interaction through molecular structure similarity analysis. *Journal of the American Medical Informatics Association*, *19*(6), 1066–1074. https://doi.org/10.1136/amiajnl-2012-000935

Wang, Q., Li, M., Wang, X., Parulian, N., Han, G., Ma, J., Tu, J., Lin, Y., Zhang, H., Liu, W., Chauhan, A., Guan, Y., Li, B., Li, R., Song, X., Fung, Y. R., Ji, H., Han, J., Chang, S.-F., … Onyshkevych, B. (2020). COVID-19 literature knowledge graph construction and drug repurposing report generation. arXiv, preprint. 2007.00576

Wieder, O., Kohlbacher, S., Kuenemann, M., Garon, A., Ducrot, P., Seidel, T., & Langer, T. (2020). A compact review of molecular property prediction with graph neural networks. *Drug Discovery Today: Technologies*, 1740–6749.

Wilson, A. J., Troy-Barnes, E., Subhan, M., Clark, F., Gupta, R., Fielding, A. K., Kottaridis, P., Mansour, M. R., O'Nions, J., Payne, E., Chavda, N., Baker, R., Thomson, K., & Khwaja, A. (2020). Successful remission induction therapy with gilteritinib in a patient with de novo FLT3-mutated acute myeloid leukaemia and severe COVID-19. *British Journal of Haematology*, 189–191.

Wu, Y., Ho, W., Huang, Y., Jin, D.-Y., Li, S., Liu, S.-L., Liu, X., Qiu, J., Sang, Y., Wang, Q., Yuen, K.-Y., & Zheng, Z.-M. (2020). SARS-CoV-2 is an appropriate name for the new coronavirus. *The Lancet*, *395*(10228), 949–950. https://doi.org/10.1016/S0140-6736(20)30557-2

Xia, Z., Wu, L.-Y., Zhou, X., & Wong, S. T. (2010). Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. *BMC Systems Biology*, *4*(S2), 6. https://doi.org/10.1186/1752-0509-4-S2-S6

Zhang, P., Wang, F., Hu, J., & Sorrentino, R. (2014). Toward personalized medicine: leveraging patient similarity and drug similarity analytics. *AMIA Joint Summits Translational Science Proceedings*, 132–136.

Zheng, X., Ding, H., Mamitsuka, H., & Zhu, S. (2013). *Collaborative matrix factorization with multiple similarities for predicting drug-target interactions*. Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1025–1033.

Zhou, Y., Wang, F., Tang, J., Nussinov, R., & Cheng, F. (2020). Artificial intelligence in COVID-19 drug repurposing. *The Lancet Digital Health*, *2*, 667–676.