



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

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



Comprehensive analysis of pathways in Coronavirus 2019 (COVID-19) using an unsupervised machine learning method



Golnaz Taheri^{a,b,*}, Mahnaz Habibi^{c,1}

^a Department of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden

^b Science for Life Laboratory, Stockholm, Sweden

^c Department of Mathematics, Qazvin Branch, Islamic Azad University, Qazvin, Iran

ARTICLE INFO

Article history:

Received 12 October 2021

Received in revised form 7 January 2022

Accepted 11 August 2022

Available online 17 August 2022

Keywords:

Coronavirus disease 2019

SARS-CoV-2

Machine learning

Unsupervised learning

ABSTRACT

The World Health Organization (WHO) introduced “Coronavirus disease 19” or “COVID-19” as a novel coronavirus in March 2020. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) requires the fast discovery of effective treatments to fight this worldwide crisis. Artificial intelligence and bioinformatics analysis pipelines can assist with finding biomarkers, explanations, and cures. Artificial intelligence and machine learning methods provide powerful infrastructures for interpreting and understanding the available data. On the other hand, pathway enrichment analysis, as a dominant tool, could help researchers discover potential key targets present in biological pathways of host cells that are targeted by SARS-CoV-2. In this work, we propose a two-stage machine learning approach for pathway analysis. During the first stage, four informative gene sets that can represent important COVID-19 related pathways are selected. These “representative genes” are associated with the COVID-19 pathology. Then, two distinctive networks were constructed for COVID-19 related signaling and disease pathways. In the second stage, the pathways of each network are ranked with respect to some unsupervised scoring method based on our defined informative features. Finally, we present a comprehensive analysis of the top important pathways in both networks. Materials and implementations are available at: <https://github.com/MahnazHabibi/Pathway>.

© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Coronavirus (CoV) is a large family of respiratory viruses that can cause a variety of illnesses ranging from mild to severe diseases including the common cold, to severe acute respiratory syndrome (SARS) [1]. CoV is found in different kinds of animals, including camels, pangolins, and bats [2]. Some of these viruses can evolve and infect humans on rare occasions. The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a newly evolved virus from this family. Pneumonia caused by the SARS-CoV-2 was introduced by the World Health Organization (WHO) as coronavirus disease (COVID-19).

Even though there are over thousands of clinical trials, there are no approved medications for the COVID-19 [3]. Although SARS-CoV-2 has a lower mutation rate than other coronaviruses, at the same time, genomic diversity is seen both among individual

patients and within the same virus class [4,5]. Genetic diversity allows viruses to adapt to a variety of hosts and circumstances within those hosts. This diversity is mostly associated with disease development, drug resistance, and treatment outcomes. As a result, even minor but continual virus alterations and mutations would reduce the efficiency of vaccines or commonly used drugs for COVID-19 treatments. Therefore, gathering information about the virus's development and pathology will be critical to controlling the pandemic.

Many researchers are sharing their findings to learn more about SARS-CoV-2's genome and evolution in the world. Several researchers are focusing on finding a treatment with the help of existing drugs with the drug repurposing method as a faster and less expensive technique [6]. Pathway analysis is a useful method for drug repurposing and understanding different patients' responses to the virus. Pathway analysis can improve the interpretation of SARS-CoV-2 data by identifying biological pathways of host cells affected by the virus. From a large amount of SARS-CoV-2 related data released, this kind of analysis can help to characterize possible drug targets and drug mechanisms of action. As a result, to find an effective treatment, obtaining knowledge from data that characterizes SARS-CoV-2 host infection is a beneficial approach.

* Corresponding author at: Department of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden.

E-mail addresses: golnazt@kth.se (G. Taheri), m_habibi@qiau.ac.ir (M. Habibi).

¹ The authors wish it to be known that, in their opinion, the first and second authors should be regarded as Joint First Authors.

Biological pathways are human representations of a cell's coordinated molecular operations, which might involve genes, proteins, molecules, tissues, and organs. Signaling pathways, disease pathways, metabolic pathways, and regulatory pathways are the four main groups of these pathways. Generic pathway databases, which include pathways from all groups, are more extensive and widely utilized. From all of the available pathway databases, each database has a distinct number of pathways. However, the pathways of each database only annotate a portion of the genome. When all major pathway databases are combined, most of the pathway databases focus on *Homo sapiens*-related pathways [7]. Pathway analysis methods can help researchers to identify the role of host genes in altering the biological response to SARS-CoV-2 and different outcomes of COVID-19.

Ranking important and more relevant pathways from all of the COVID-19 associated pathways proposed by recent studies will assist researchers to focus on select sets of genes for further inquiry. Furthermore, observing pathways as functional units of cellular systems may help researchers extract biologically relevant information from microarray data. Therefore, various research groups have been inspired to look at gene sets rather than single genes as a result of these considerations.

In this work, we developed an unsupervised machine learning approach to identify important disease pathways crosstalk between underlying diseases and COVID-19. We also tried to recognize important signaling pathways of the host protein which could help to identify effective COVID-19 treatment. For this purpose, we constructed two biological networks corresponding to the COVID-19 related signaling and disease pathways. Then, we defined six informative structural features for each pathway as a vertex in the network. Finally, we calculated the Laplacian Score with respect to our predefined features for each pathway and introduced the high score pathways with meaningful relationships to COVID-19 as candidate pathways for more investigation.

2. Preliminaries

Groups of proteins that interact in the body are referred to as "cellular processes". To understand the cell, we must understand how these spatially and temporally organized interactions lead to biological processes [8]. The conventional approach to studying cellular function has improved our understanding of disease, infection, drug development, and evolution [9]. In this work, we define data and networks as "molecular" if they are concerned with interactions between particular biological molecules. There are different approaches for analyzing biological processes using molecular interaction networks. Protein-protein interaction (PPI) data is often used to create networks in which proteins are shown interacting with other functionally associated proteins. This results in the formation of "functional modules", which are subnetworks that share a common function [10]. In PPI networks, the edges connect each protein to all of its other interacting proteins. Proteins may participate in various functions depending on the interactions they create in different cellular contexts and sub-cellular components [11]. In this work, we introduce a representation of cellular functions that uses pathways rather than genes. The pathway includes a set of proteins (and complexes) that interact with each other serially and create signaling, metabolic, and disease pathways [7]. This concept allows us to group sets of proteins known to interact under certain conditions. In this study, we did not use gene expression data and assumed that gene expression represents protein levels. The pathway model allows proteins to be characterized independently in multiple pathways.

Among many types of biological pathways, those that are involved in metabolism, regulation of genes, and the transmission of signals are well-known [12]. The chemical reactions that take

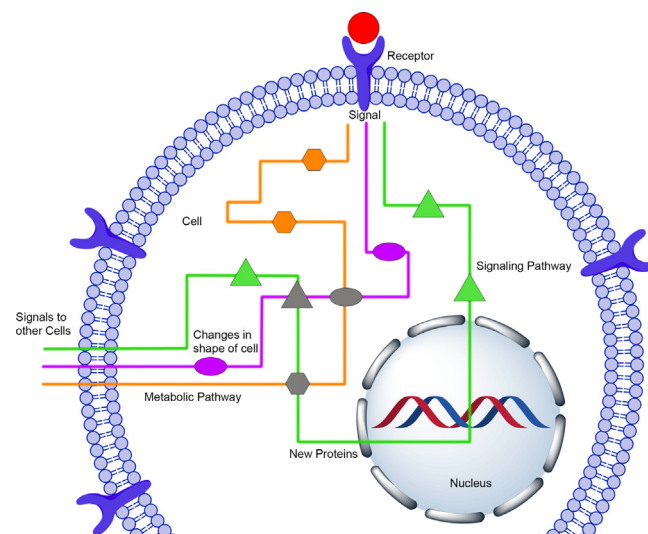


Fig. 1. A schematic view of a biological pathway.

place in our bodies are mediated by metabolic pathways. Some other metabolic pathways mostly help to produce molecules. Regulatory pathways turn genes on and off. This is such an important task because genes supply the formula by which cells produce proteins, and proteins are the key elements needed to carry out almost every task in the human body. Signaling pathways move a signal from the outside of the cells into the inside of them. Different cells are capable of receiving specific signals via structures on their surface called receptors. After interacting with these receptors, the signal travels into the cell. The message of a signal is transmitted by some kinds of spatial proteins that start a specific reaction in the cell. For example, a signal from outside the cell could lead the cell to create a particular protein inside the cell. After that, the produced protein may be a signal that causes the cell to move [12]. Fig. 1 shows a schematic view of a biological pathway as a series of interactions between different molecules in a cell that leads to a specified product or a change in a cell. As we mentioned earlier, the most common biological pathways are involved in metabolism, the regulation of gene expression, and the transmission of signals.

In the pathway analysis methods, the position of a given gene in pathways determines its function in some way [13]. Upstream genes can affect the downstream responses of pathways. In addition, genes with a high degree (hubs) in pathways participate in important biological processes that are associated with the diagnosis and therapy of different diseases [14]. For example, p53 is an important upstream gene with a high degree in the P53 signaling pathway, which is the most repeatedly mutated gene in different types of cancer. This gene plays an important role in tumor suppression through transcriptional regulation of its downstream target genes like MYC and EGR1 [15]. Another important upstream gene is RAS. This gene contains several important proteins (KRAS, NRAS, HRAS, RRAS, RRAS2, and MRAS). It has a high degree in the MAPK signaling pathway. The RAS mutation can cause a broad range of genetic disorders [16]. Recently, several methods consider the structure of pathways, such as centrality, betweenness, and weight-based pathway enrichment.

3. Related work

SARS-CoV-2 has become an existing health crisis and threat to the entire world. Thus, the world needs to be fast in recognizing suitable information and medication to restrict the spread

of this disease. The large-scale data of COVID-19 patients can be integrated and analyzed by state-of-the-art machine learning algorithms for a better understanding of the pattern of viral spread, and to improve diagnosis speed and accuracy. These algorithms also develop novel, effective therapeutic approaches and could identify the most sensitive people based on personalized genetics. After a short period of time since the COVID-19 outbreak, advanced machine learning techniques have been used in the classification of COVID-19 genomes [17], evaluation of CRISPR-based COVID-19 detection [18], survival prediction of severe COVID-19 patients, and finding potential drug candidates for COVID-19 treatments [19].

Current efforts have been made to develop effective and novel diagnostic approaches with the help of machine learning algorithms. As an example, machine learning-based screening of SARS-CoV-2 evaluation techniques using a CRISPR-based virus detection system was confirmed with high accuracy and speed [18]. Neural network-based classifiers were developed for large-scale screening of COVID-19 patients based on their specific respiratory patterns [20]. Another deep learning-based method for COVID-19 patients' CT image analysis was constructed for the automated diagnosis and monitoring over time [21]. The fast development of automated diagnostic systems based on artificial intelligence and machine learning methods can increase diagnostic accuracy and speed. These methods also save and protect healthcare workers by decreasing their contact with COVID-19 patients.

In [22] new drug-like compounds for COVID-19 treatment have been developed using a deep learning pipeline. Another deep learning system developed by Google DeepMind, AlphaFold, predicted protein structures associated with COVID-19 [23]. Traditional experimental methods can take months to complete this work. In addition, a new machine learning-based tool proposed COVID-19 vaccine candidates [24]. Data from worldwide COVID-19 treatment centers necessitates advanced machine learning techniques. These methods could be used to evaluate new patients by analyzing their personalized therapeutic effects.

4. Material and methods

As an input, we selected a set of candidate COVID-19 related genes. This candidate set contains most of the genes that have been suggested as drug targets for COVID-19 in the Drug Bank [25]. Then, we extracted the corresponding pathways from Kyoto Encyclopedia of Genes and Genomes dataset (KEGG) for these selected genes [26]. In the first step of this work, we constructed two biological networks corresponding to COVID-19 related signaling and disease pathways. We also defined six informative features for these two networks. In the second step, we used an unsupervised learning method and computed the Laplacian Score, then sorted pathways for each network with respect to the corresponding score. The workflow of the proposed methods to identify important pathways is illustrated in Fig. 2.

4.1. Protein–protein interaction network and important genes

The physical interaction between two or more protein molecules is called the PPI network. A set of proteins and the physical interactions between them are represented as a protein–protein interaction network $\mathcal{G} = \langle \mathcal{P}, \mathcal{I} \rangle$. Each vertex in the network represents a protein and each edge, uv , in the network shows the interaction between these two proteins, u and v . When two proteins interact physically, the two vertices, u and v , are referred to as adjacent. The number of interactions that a vertex, u , has with other vertices is called the degree. A set of high-degree vertices is called hub vertices. Hub vertices play an important role

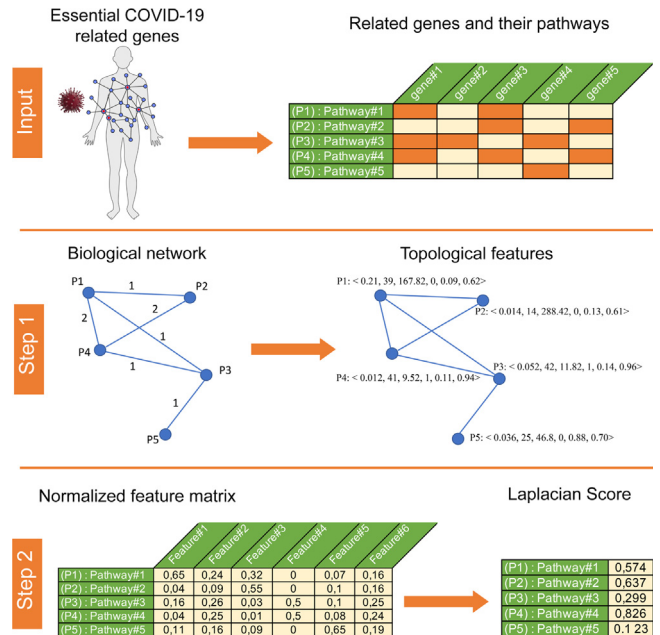


Fig. 2. The workflow of the proposed methods.

and contain important biological information for the cell. Generally, hub vertices and vertices with high centrality measures are selected as important genes in the PPI network. To identify and evaluate the important genes of each COVID-19 related pathway, we will identify the important vertices of the PPI network for all of the genes in that pathway.

4.2. Construction of biological network of COVID-19 related pathways

Suppose S represents a set of essential genes related to COVID-19, and $P = \{p_1, p_2, \dots, p_k\}$ defines a set of related pathways gathered from the KEGG dataset (See “Datasets” subsection). We construct a biological network to identify important pathways and related genes. Suppose $G^i = \{g_1^i, g_2^i, \dots, g_m^i\}$ represents a set of genes in set S associated with i th pathway p_i . A biological network associated with pathways is considered as an unidirectional weighted graph $G = (V, E, \omega)$. In graph G , each vertex represents a pathway. The two pathways p_i and p_j in graph G are connected by an edge $e_{ij} \in E$ whenever $G^i \cap G^j \neq \emptyset$. These two pathways are named neighbors, and the set of neighbors of each vertex is denoted by $N(p)$. The $\omega(e_{ij})$, as the weight of the edge e_{ij} , represents the number of genes in common between the two pathways p_i and p_j . For $H \subseteq V$ we have $N(H) = \bigcup_{s \in H} N(s)$. A path between two vertices u and v is a distinct sequence of graph vertices $u = v_0, v_1, \dots, v_k = v$ if each $v_i v_{i+1} (1 \leq i \leq k-1)$ is an edge of graph G . The length of a path is equal to the number of edges in this path. The distance between two vertices u and v is equal to the length of the shortest path between these two vertices, which is denoted by $d(u, v)$.

4.3. The topological features of biological network

In this part, for each pathway, we define six informative features that correspond to the pathway's position in the biological network:

1. The ratio of the number of genes in each pathway to the total number of COVID-19 related genes (S).

$$Pro(p_i) = \frac{|G^i|}{|S|}. \quad (1)$$

The larger value of this measure indicates that this pathway corresponds to the COVID-19 related genes.

- The number of neighbors of each pathway in a biological network is called the degree of that pathway. The degree of each pathway is another topological feature that we have considered. A large value of this measure indicates the importance of the COVID-19 related genes in this pathway in terms of the number of participants in the other pathways.
- Betweenness centrality value for each pathway, p_i , in a weighted graph G is defined as follows:

$$Bn(p_i) = \sum_{u,w \in V} \frac{\tau_{uw}(p_i)}{\tau_{uw}} \quad (2)$$

The value τ_{uw} indicates the total number of shortest paths from u to w and $\tau_{uw}(p_i)$ shows the number of shortest paths that pass through p_i . This measure shows the importance of the pathway p_i in terms of centrality.

- Let $L(G) = [l_{ij}]$ indicates the Laplacian matrix of weighted graph G , where:

$$l_{ij} = \begin{cases} \omega(e_{ij}) & \text{if } e_{ij} \in E \\ \sum_{e_{ij} \in N(p_i)} \omega(e_{ij}) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Assume that vector $u = (u_1, u_2, \dots, u_n)$ represents the normalized eigenvectors of matrix $L(G)$ and vector $(\lambda_1, \lambda_2, \dots, \lambda_n)$ represents the associated eigenvalues [27]. Based on the second smallest eigenvalue, we compute the eigenvectors of the Laplacian matrix $L(G)$ and store them in vector $X = (x_1, \dots, x_n)$. Then, we sort the elements of vector X and place half of the nodes in partition G_1 and the rest in partition G_2 . The subset C is selected from the vertices of two partitions G_1 and G_2 alternatively, such that removal of C separates the nodes of graph G into two disjoint parts of almost equal size. Now the essentiality of each pathway, p_i , is defined as follows:

$$Es(p_i) = \begin{cases} 1 & \text{if } p_i \in C \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

- Let $H = v \cup N(v)$, the Expansion of H is defined as follows:

$$Ex(H) = \frac{|N(H)|}{|H|} \quad (5)$$

where $|N(H)|$ and $|H|$ demonstrate the size of $N(H)$ and H respectively.

- The Closeness centrality measure for each vertex v of graph G is defined as follows:

$$Closeness(v) = \frac{|V| - 1}{\sum_{u \in V} d(u, v)} \quad (6)$$

If the $Closeness(v)$ value meets or exceeds a predefined threshold, we consider that the vertex v has the shortest distance with other vertices of the graph.

4.4. Proposed model

Suppose f_{ri} represents the r th feature of the i th sample ($1 \leq i \leq n$ and $1 \leq r \leq m$). We assign a feature vector $\vec{(p_i)} = (f_{1i}, \dots, f_{mi})$ to each sample p_i and define the column matrix $F_r = [f_{r1}, \dots, f_{rn}]^T$. To find the appropriate score for each feature we compute the following measures:

- The weight matrix $S = [s_{ij}]_{(n \times n)}$ is defined as follows:

$$s_{ij} = \begin{cases} e^{-\frac{|\vec{(p_i)} - \vec{(p_j)}|^2}{t}} & |\vec{(p_i)} - \vec{(p_j)}| < \epsilon \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

- Suppose that $J = [1, 1, \dots, 1]^T$ and let $L = D - S$ indicates the Laplacian matrix then we define $D = [d_{ij}]_{(n \times n)}$ as follows:

$$d_{ij} = \begin{cases} \sum_{k=1}^n s_{ik} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

- For the r th feature, we define F_r as follows:

$$\tilde{F}_r = F_r - \frac{F_r^T D J}{J^T D J} J \quad (9)$$

- For the r th feature, compute the Laplacian Score as follows:

$$L_r = \frac{\tilde{F}_r^T L \tilde{F}_r}{\tilde{F}_r^T D \tilde{F}_r} \quad (10)$$

4.5. Finding important COVID-19 related pathways

Since the problem of finding the appropriate set of COVID-19 related pathways is still an open question, it can be considered as a problem without a response variable or exact answer. Therefore, to find an efficient answer, we present a novel unsupervised learning method to find important COVID-19 related pathways. It is worth mentioning that, in supervised learning methods, feature selection has been studied extensively. Due to the lack of information about class labels to help the search for relevant knowledge in unsupervised learning methods, scoring features is a significantly more difficult challenge [28]. Our method comprises two main steps.

In the first step, we construct a biological network with respect to a set of COVID-19 related genes (S). Then we calculate six informative features for each pathway with respect to the topological features of the vertex in the biological network. It specifies a feature matrix $F = [f_{ir}]$ where f_{ir} represents the r th feature of the i th pathway.

In the second step, we use an unsupervised feature selection method on a feature matrix to calculate the Laplacian Score for each feature (L_r). Then, for the pathway p_i , compute the Laplacian Score as follows:

$$L_{p_i} = \sum_{r=1}^m \tilde{f}_{r_i} L_r \quad (11)$$

where \tilde{f}_{r_i} represents the normalized feature for the i th pathway.

Finally, the high score pathways are selected as important COVID-19 related pathways.

5. Results

5.1. Datasets

Identifying associated genes with disease pathology plays an important role in recognizing significantly related pathways and finding appropriate drugs. We collected four different sets of COVID-19 related genes. We utilized a set of experimental unapproved drugs used for COVID-19 treatment that are available on the Drug Bank [25]. This set is represented by Covid-Drug and includes 64 drugs. From these 64 drugs, 21 of them can target human proteins. These 21 drugs target 88 human proteins, which we defined as the first set of COVID-19 related genes and is denoted by the set Co . We also used a set of experimental unapproved drugs in clinical trials for COVID-19 treatment that are available on the Drug Bank [25]. This set includes 708 drugs, of which 347 drugs have been studied clinically in more than one clinic. From these 347 drugs, 213 drugs can target human

proteins. This group of drugs is represented by the Clinical-Drug. These 213 drugs target 930 human proteins, which we defined as another set of COVID-19 related genes, denoted by the set *Cl*. The third set, which we considered as a set of COVID-19 related genes, contains 332 human proteins identified by Gordon et al. as targets of the virus in human proteins and is denoted by the set *T* [29]. The last set that we have considered as a set of COVID-19 related genes includes 93 human proteins presented by Habibi et al. is represented by set *E* [30]. Finally, we placed all of the above-mentioned four sets into set *S*. This set contains 1300 genes, which we introduced as a final set of COVID-19 related genes. We also used the KEGG dataset, which contains complete information about pathways inside cells. From this dataset, we extracted information about signaling and disease pathways and related genes in each pathway. From the KEGG dataset, we selected two sets of signaling and disease pathways associated with COVID-19 that contain more than eight genes in the selected set *S*. In addition, to identify important genes in each pathway, we used the PPI network collected by Habibi et al. [30]. This network contains 304,730 interactions between 20,040 human proteins. They collected information on the physical interactions between proteins in 5 datasets, including Biological General Repository for Interaction Datasets (BioGRID) [31], Agile Protein Interactomes Data analyzer (APID) [32], Homologous interactions (Hint) [33], Human Integrated Protein-Protein Interaction reference (HIPPIE) [34], and Huri [35]. They mapped all of the proteins from all of the above-mentioned five datasets to their corresponding Universal Protein Resource (UniProt) ID [36] and removed proteins if they could not be mapped to UniProt IDs.

5.2. Pre-processing

Since the spread of COVID-19, different techniques have been used to identify COVID-19 related genes. In this study, we used four different groups of COVID-19 related genes. Fig. 3 shows the venn diagrams associated with these selected sets. Fig. 3 demonstrates that more than 50% of the selected genes in each set, *Cl*, *E*, and *T*, are not in the other sets. In other words, these sets almost contain unique genes. In total, these sets include 1300 genes associated with COVID-19, which we select as the COVID-19 related genes and denote by set *S*. Since many studies have been conducted to find out genes related to the pathology of COVID-19 and it is not possible to check all of them, we have reviewed a number of these studies. The results of our study show that the candidate genes in our selected set contain most of the important genes that are proposed through other studies. These studies used different methods and approaches for finding genes. In the following, we compare our candidate genes set with four different genes sets in other studies.

5.3. Evaluation of COVID-19 selected sets of proteins

As mentioned earlier, we compared our selected gene sets with four gene sets that have been presented in other studies. The authors of the first study presented a method that is based on informative statistical and topological features in a human-virus protein interaction network [37]. More than 96% of the genes presented in [37] are represented by our candidate genes. The second study identified potential targets for repurposing based on Mendelian randomization [38]. Our candidate genes contain more than 95% of the genes presented in [38]. The authors of the third study investigated the COVID-19 comorbidity-associated gene sets [39]. Our candidate genes contain more than 84% of the genes presented in [39]. The last study identified the hub gene in the pathways related to COVID-19 [40]. Our candidate genes include more than 94% of the genes listed in [40]. Therefore, it

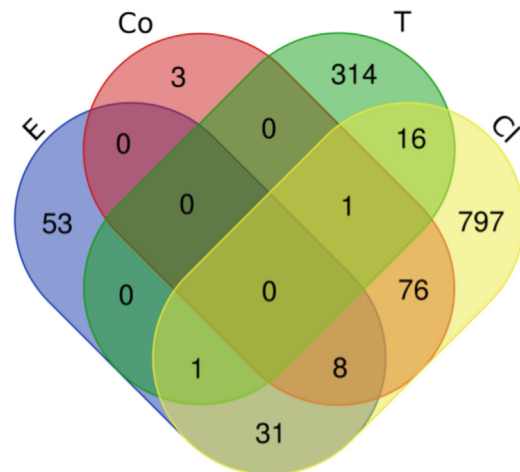


Fig. 3. Venn diagram of COVID-19 related genes in sets *E*, *Cl*, *Co*, *T* respectively.

Table 1

The number of related pathways in four COVID-19 related selected gene sets.

	Co	Cl	T	E	S
No. Signaling pathway	4	15	4	22	39
No. Disease pathway	12	23	2	31	56

can be concluded that our selected set *S* contains more than 92% of the genes reported in the above-mentioned studies. Thus, the selected set *S* could be a complete set of COVID-19 related genes.

We also studied six sets of disease-related genes for six underlying diseases: cardiovascular diseases, diabetes, hepatitis, lung diseases, kidney disease, and different types of cancers respectively. Then we compared the genes of these six sets with our selected set *S*. We found that from our 1300 candidate genes, 461 genes are related to these six underlying diseases. From these 461 genes, 12 genes (APOE, CCR5, CD14, IL1 A, IL1B, IL6, ITGB3, LDLR, MMP3, MTHFR, TNF) are commonly shared between all of the six disease-related genes. We also found 33 genes in more than 5 diseases out of 6 mentioned diseases. Fig. 4 shows the percentage of these shared genes between disease.

5.4. Evaluation of COVID-19 related pathways

One of the effective methods for finding therapeutic procedures in patients with severe COVID-19 is to understand the related molecular pathways. Identifying these pathways that are related to underlying COVID-19 could help to recognize potential drug targets.

Therefore, we studied the KEGG pathways associated with mentioned 1300 COVID-19 related genes. For these 1300 genes in *S*, there are 39 signaling pathways and 56 disease pathways. Table 1 presents the number of pathways related to our four candidate sets of COVID-19 related genes. Table 1 shows that the set of genes presented in set *S* covers the highest number of signaling pathways and the highest number of disease pathways as well. Most of these pathways have been studied as COVID-19 related pathways in other studies [41–43].

Then, we build two biological networks associated with each set of signaling and disease pathways and calculate the informative features for each network. The informative features of these two biological networks are modeled by the Laplacian matrix. The output of the model is the Laplacian Score for each feature. Table 2 shows the top ten signaling pathways with respect to the highest Laplacian Score and the number of genes in each pathway. Some other studies also introduced these pathways

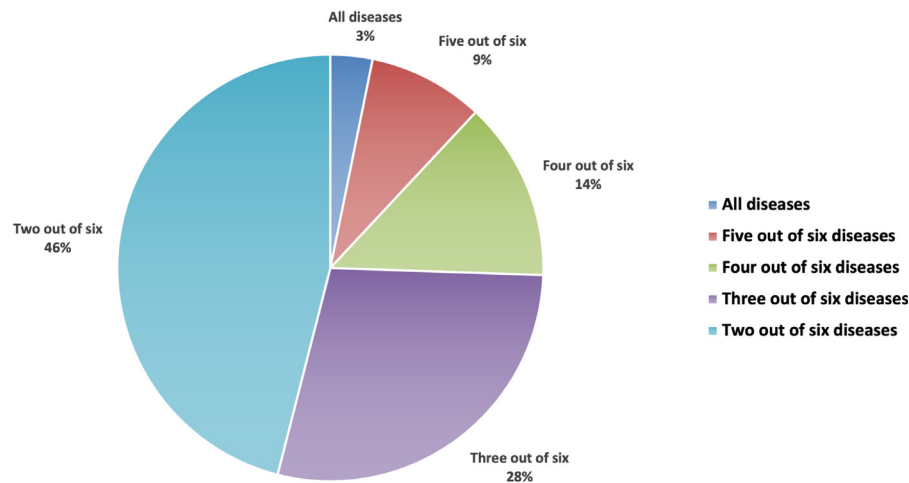


Fig. 4. Percentage of shared genes in more than 2, 3, 4, 5 and 6 out of 6 underlying diseases.

Table 2

List of top 10 signaling pathways with respect to the highest Laplacian Score.

Signaling pathway ID	No. genes	References
hsa04151:PI3K-Akt signaling pathway	80	[41]
hsa04010:MAPK signaling pathway	67	[42]
hsa04620:Toll-like receptor signaling pathway	31	[43]
hsa04621:NOD-like receptor signaling pathway	16	[44]
hsa04062:Chemokine signaling pathway	10	[45]
hsa04310:Wnt signaling pathway	18	[46]
hsa04152:AMPKs signaling pathway	28	[47]
hsa04630:Jak-STAT signaling pathway	30	[48]
hsa04668:TNF signaling pathway	32	[43]
hsa04064:NF- κ B signaling pathway	26	[49]

with other procedures and without measuring any priority as COVID-19 important related pathways. The last column of Table 2 shows the studies that presented these important pathways.

Tables 3 also shows the top ten disease pathways with respect to the highest Laplacian Score and some of the important genes in each pathway.

5.4.1. High score signaling pathways

In this subsection, we investigate three of the top high score signaling pathways that are selected through our method.

1. PI3K/AKT signaling pathway

The first pathway is the Phosphatidylinositol 3-Kinase PI3K/AKT signaling pathway as one of the possible therapeutic targets for COVID-19 treatment. This pathway is associated with different characteristics of virus entry into the cell and the evolution of immune responses. PI3K/AKT signaling pathway performs a major role in cell invasion, growth, migration, and proliferation. It can inhibit apoptosis and increase angiogenesis. Irregular PI3K/AKT signaling pathway can produce diseases, such as cancer, and inhibitors of this pathway perform a crucial role in cancer treatment. On the other hand, it has been shown that SARS-CoV-2 endocytosis happens through a clathrin-mediated pathway which is regulated by the PI3K/AKT signaling [41]. Therefore, suppression of this pathway can inhibit the entry of other viruses that use clathrin-mediated endocytosis. Other studies show that activation of the PI3K/AKT signaling pathway has been connected to the initiation of lung tissue fibrosis, which has been detected in COVID-19 patients. Based on this evidence, the use of anti-cancer and anti-viral drugs, to inhibit the PI3K/AKT signaling pathway and

eventually suppress inflammation in COVID-19 can be an effective treatment for COVID-19 patients.

2. MAPK signaling pathway

The second pathway is the p38 Mitogen-Activated Protein Kinase (MAPK). The p38 MAPK is the third major signaling cassette of the MAPK. This pathway performs a critical role in the release of pro-inflammatory cytokines like interleukin-6 (IL-6) and has been involved in myocardial dysfunction and acute lung injury. The strong inflammatory response in COVID-19 may be caused by abnormally upregulated p38 activity and can be explained by two mechanisms. First, during SARS-CoV-2 viral entry, angiotensin-converting enzyme 2 (ACE2) activity is lost. Second, as previously shown in SARS-CoV, p38 activity is upregulated via a viral protein, like other RNA respiratory viruses that may hijack p38 activity to promote replication [42]. Thus, SARS-CoV-2 may cause vast inflammation by directly activating p38 and down-regulating a key inhibitory pathway. The SARS-CoV-2 can simultaneously take advantage of p38 activity to replicate. Based on this evidence, p38 inhibitors should be considered for clinical trials in serious COVID-19 infections.

3. TLRs signaling pathway

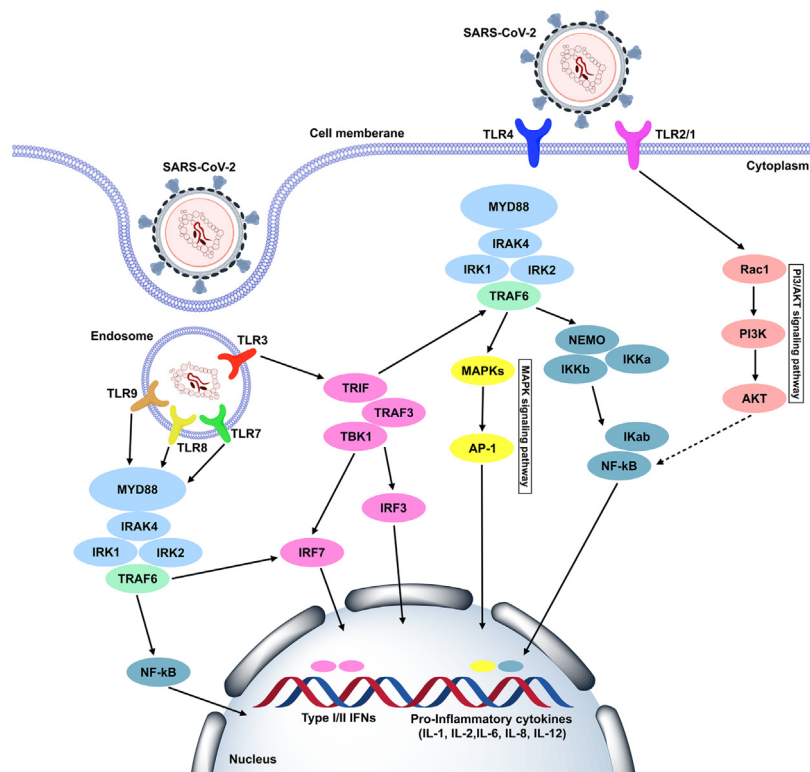
The next important pathway is the Toll-like Receptors (TLRs) as a family of ten members in humans. TLRs are expressed on different immune cells, like macrophages, dendritic cells, natural killer cells, B cells, and T cells. The innate immune system of this pathway is activated by many viruses, which is important for the removal of viruses. This procedure could be harmful to the host due to continuous inflammation and tissue damage. In patients with severe COVID-19, the inflammatory response begins with the release of cytokines, such as IL-6 and Tumoral Necrosis Factor-alpha (TNF- α). It is remarkable that these two main cytokines involved in severe COVID-19 (IL-6 and TNF- α) are downstream of TLR4. On the other hand, TLRs can activate the Janus kinase transducers (JAK/STAT), which could cause macrophage activation syndrome. Therefore, TLRs have a considerable dual role in viral infections [43]. Based on these evidences, TLRs properties make them possible candidates to treat inflammation in severe viral diseases like COVID-19.

Fig. 5 shows that TLR signaling pathways have cross-talk with both PI3K/AKT and MAPK signaling pathways. TLRs are divided into two groups: a MyD88-dependent pathway and a MyD88-independent pathway. The first one drives the production of

Table 3

List of top 10 disease pathways with respect to the highest Laplacian Score.

Disease pathway ID	No. genes	Important genes
hsa05200:Pathways in cancer	98	P53, STAT3
hsa05164:Influenza A	56	TNFRSF1A, IKKb, JAK1
hsa05166:HTLV-1 infection	49	P53, MYC, JAK1
hsa05152:Tuberculosis	44	TNFRSF1A, CASP3
hsa04931:Insulin resistance	34	MTOR, IKKb
hsa05169:Epstein-Barr virus infection	30	MYC, P53
hsa05321:Inflammatory bowel disease (IBD)	20	IL2, STAT3
hsa05323:Rheumatoid arthritis	28	CCL2, TNF, CCL5
hsa05144:Malaria	21	TGFB1, CXCL8
hsa05130:Pathogenic Escherichia coli infection	10	CD14, ACTB

**Fig. 5.** TLR signaling pathway cross-talks and immune deficiencies.

proinflammatory cytokines with the immediate activation of NF- κ B and MAPK. The second one associates with the induction of IFN- β and IFN-inducible and activation of NF- κ B and MAPK. In the MyD88-dependent pathway, the MyD88 protein recruits IRAK family proteins. The IRAK4 protein activates TRAF6 and this protein ultimately activates MAPKs and NF- κ B resulting in the production of proinflammatory cytokines. On the other hand, activated TRIF binds to TRAF3 and TBK1, ultimately activating IRF3 and IRF7 to initiate the production of interferons. TRIF also interacts with TRAF6 to stimulate MyD88-independent activation of MAPKs and NF- κ B. Our results in this subsection show the importance of TLRs in other mentioned signaling pathways. This relevance could confirm the potential link between the unknown COVID-19 pathways and the TLRs receptors.

5.4.2. High score disease pathways

In this subsection, we discuss three of the most important disease pathways and related important genes that have the highest score through our method. Identifying the shared pathways between underlying diseases and COVID-19 and also identifying the important genes related to each pathway gives us valuable information about possible drug targets.

1. Pathways in cancer

The first COVID-19 related disease pathway is the pathways in cancer (*hsa05200*). We find that 98 genes are shared between pathways in cancer and COVID-19. Fig. 6 shows the PPI network between these 98 genes. We evaluate two important genes TP53 and STAT3 with the valuable topological properties in this network. In Fig. 6 the topologically important genes are shown in red.

- TP53:** The first important gene that is shared between pathways in cancer and COVID-19 is the Tumor Protein P53 (TP53 or P53). The cellular tumor antigen P53 is a key member in INF type 1 [50]. Researchers believe that the key tumor-suppressor P53 protein will undergo degradation by SARS-CoV-2, leading to increased viral survival in host cells. Thus, P53 activators might effectively inhibit viral replication in the human respiratory tract and lung cells [51]. Therefore, TP53 could be a potential therapeutic target for COVID-19 treatment.
- STAT3:** The second important gene that is shared between pathways cancer and COVID-19 is the Signal Transducer and Activator of Transcription 3 (STAT3).

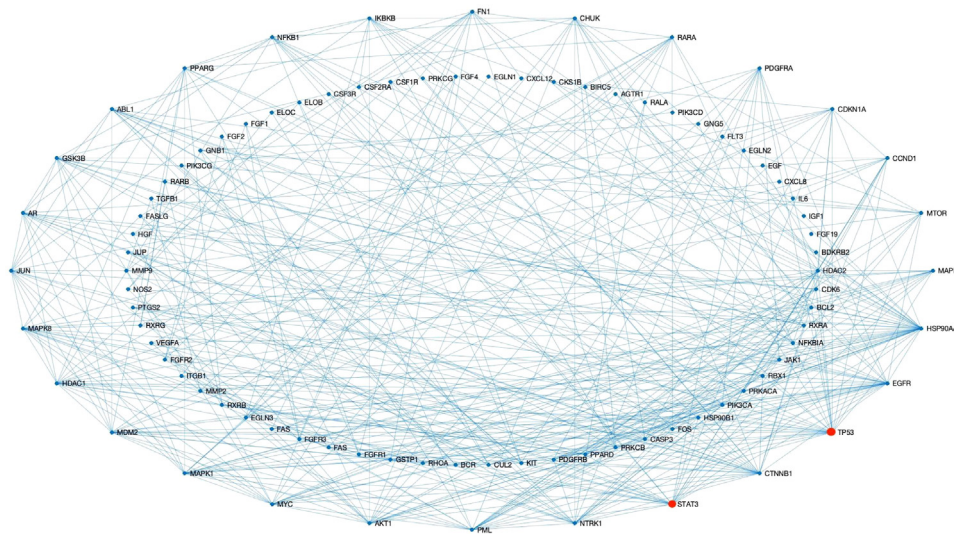


Fig. 6. The PPI network associated with 98 genes are shared between pathways in cancer and COVID-19.

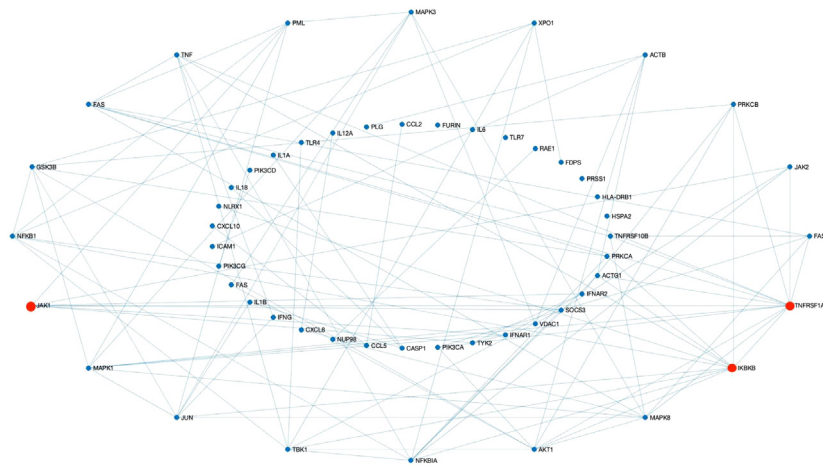


Fig. 7. The PPI network associated with 56 genes are shared between Influenza A and COVID-19.

Recent studies showed that the viral components of SARS-CoV-2 cause the abnormal hyperactivation of STAT3 through COVID-19 cytokines [52]. Abnormal activity of STAT3 elevates lymphopenia and reduces anti-virus immune responses. Therefore, STAT3 could be a potential therapeutic target for COVID-19 treatment.

2. Influenza A pathway

The second one is the Influenza A pathway (*hsa05164*), which is identified as an essential disease pathway related to COVID-19. We find that 57 genes are shared between Influenza A and COVID-19. We analyze the topologically important genes among these 57 reported genes. This analysis is accomplished with respect to the protein-protein interaction (PPI) network. As shown in Fig. 7, three genes, TNFRSF1A, JAK1, and IKK2, are known as important genes. In Fig. 7 the topologically important genes are shown in red.

(a) **TNFRSF1A**: The first important gene that is shared between Influenza A and COVID-19 is TNFRSF1A/TNFR1. The Tumor Necrosis Factor Receptor (TNFRs) is recognized as the receptor for TNFSF2/TNF- α [36]. TNF and its receptors (TNFR1 and TNFR2) are related

to proinflammatory responses and play a major role in the regulation of inflammation. Previous studies have shown the relation between the severity of COVID-19 and increasing TNFR1 [53]. Therefore, TNFRSF1A could be a potential therapeutic target for COVID-19 treatment.

(b) **JAK1**: The second important gene that is shared between COVID-19 and Influenza A is the Tyrosine-protein kinase JAK1 (JAK1). JAKs (JAK1 and JAK2) are transmembrane proteins that increase extracellular signals. Many of the cytokines, including interferons (IFNs), involved in the cytokine storm use protein kinases JAKs for signal transduction [48]. Therefore, blocking IFNs which are mediated by JAK could be an effective COVID-19 treatment.

(c) **IKK2**: The third important gene that is shared between Influenza A and COVID-19 is the Inhibitor of nuclear factor κ B kinase subunit beta (IKK2 or IKK β). IKK2 plays an essential role in the transcription factor nuclear factor- κ B (NF- κ B) pathway. This pathway changes multiple pathogenic activities such as inflammatory cytokines [49]. Several studies confirmed the benefit of IKKs as a downstream effector of this pathway to weaken COVID-19. Therefore,

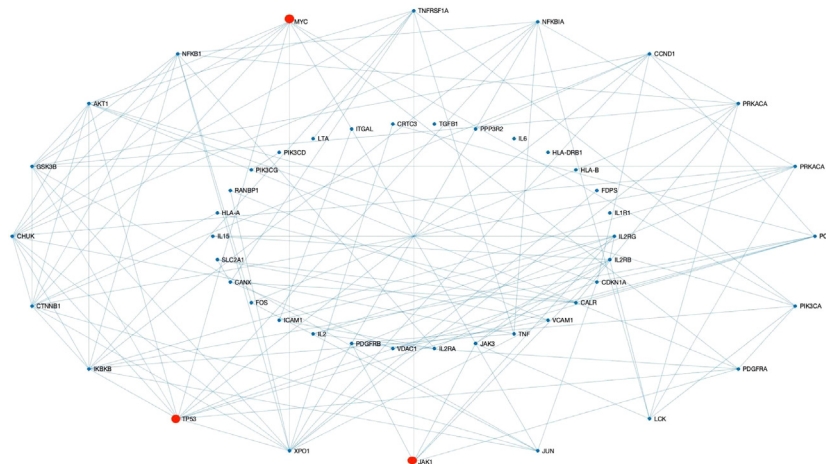


Fig. 8. The PPI network associated with 49 genes are shared between the HTLV-I infection pathway and COVID-19.

IKBKB could be a potential therapeutic target for COVID-19 treatment.

3. HTLV-I infection pathway

The third COVID-19 related disease pathway is the HTLV-I infection pathway (*hsa05166*). We find that 49 genes are shared between the HTLV-I infection pathway and COVID-19. Fig. 8 shows the protein interaction network between these genes. We evaluate three important genes, MYC, P53, and JAK1 with the valuable topological properties in this network. In Fig. 8 the topologically important genes are shown in red.

- (a) **MYC**: In the previous sections, we discussed P53 and JAK1. The next important gene that is shared between the HTLV-I infection pathway and COVID-19 is the Myc proto-oncogene protein (MYC). Recent studies have shown an increase level of MYC in COVID-19 patients. The regulation of MYC may contribute to cell proliferation, especially macrophages and neutrophils [54]. Therefore, MYC could be a potential therapeutic target for COVID-19 treatment.

We also compare and evaluate the COVID-19 related genes for three highly scored disease pathways. Fig. 9 shows the venn diagrams associated with these disease pathways. Fig. 9 demonstrates that 11 genes were identified through all three disease pathways. Table 4 shows the complete list of these 11 genes and detailed information about their corresponding functions.

6. Conclusion and discussion

There are knowledge gaps in the systems biology of COVID-19, and this makes the treatment of this disease more difficult. Continuous and comprehensive research with different points of view tries to identify these gaps and eliminate these shortcomings. Recently, many studies have been presented to identify common pathways between COVID-19 and other known signaling and disease pathways. In these studies, the potential relationship between COVID-19 and other diseases has been discussed. Therefore, a large number of signaling and disease pathways related to COVID-19 have been explored. Recognizing a smaller number of more related and important pathways can help researchers conduct more investigations. For this purpose, we constructed two biological networks corresponding to signaling and disease pathways and defined some informative features. Then, we proposed

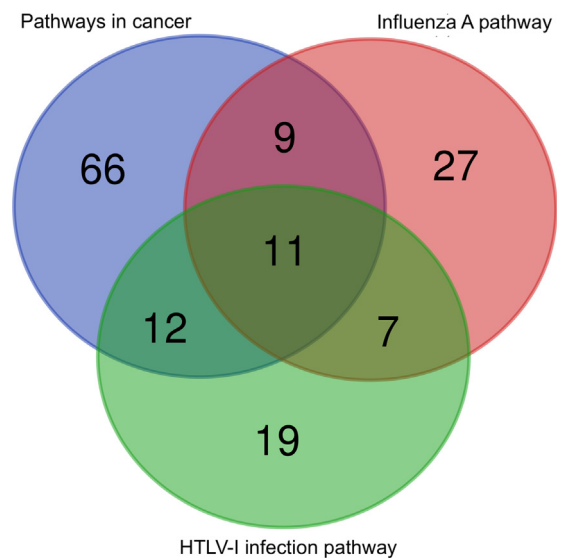


Fig. 9. Venn diagram of COVID-19 related genes in three highly scored disease pathways.

an unsupervised method for ranking these important COVID-19 related signaling and disease pathways. In the first step of this work, we selected a set of genes associated with COVID-19. This selected set contains most of the genes that have been presented as drug targets for COVID-19 in the Drug Bank or have been reported in multiple studies as COVID-19 related genes. Then, we extracted the corresponding pathways from KEGG for these selected genes. Each of the extracted pathways contains at least eight genes from the candidate gene set. Afterward, we constructed the two biological networks corresponding to these COVID-19 related signaling and disease pathways. We also defined six informative features for these networks. Then, we used an unsupervised learning method to score each feature and sort pathways for each network with respect to the corresponding score.

The results of our proposed model on a biological network corresponding to signaling pathways showed that three signaling pathways, PI3K/AKT, MAPK, and TLRs, were selected through our method as the top three signaling pathways. These pathways have the most cross-talk with COVID-19. Some researches on TLRs signaling pathways showed that MyD88 related pathways activate NF- κ B and MAPK signaling pathways rapidly, and

Table 4
The list of shared genes between three disease pathways and their functions.

Gene name	Function
NFKBIA	The NFKBIA helps keep NF- κ B bound in the IKK complex. NF- κ B regulates the activity of multiple genes, including genes that control the body's immune responses and inflammatory responses [55].
PIK3CD	The PIK3CD provides instructions for an enzyme called phosphatidylinositol 3-kinase (PI3K). PI3K is found in white blood cells, including immune system B cells and T cells. These cells identify and attack foreign invaders like viruses and bacteria and prevent infection [55].
JUN	The JUN encodes a protein that is highly similar to the viral protein. This gene interacts directly with precise target DNA sequences to control gene expression. JUN is mapped to a chromosomal region involved in both translocations and deletions in human malignancies [55].
JAK1	The JAK1 has a critical role in influencing the expression of genes that mediate inflammation. This gene is an important part of the (IL-6)/JAK1/STAT3 inflammation and immune response, and it could be an effective drug target for reducing cytokine storms [55].
AKT1	The AKT1 encodes one of the three members of the human AKT serine-threonine protein kinase family. AKT proteins control a large number of cellular functions like cell proliferation, metabolism, survival, and angiogenesis in both malignant and normal cells [55].
GSK3B	The GSK3B encodes a protein that belongs to the glycogen synthase kinase subfamily. It is involved in energy metabolism, inflammation, ER-stress, mitochondrial dysfunction, and apoptotic pathways [55].
IL6	The IL6 encodes a cytokine that operates in inflammation and in the maturation of B cells. The function of IL6 is implicated in a wide range of inflammation-associated diseases. High levels of the encoded protein have been found in virus infections like COVID-19 [55].
PIK3CG	The PIK3CG encodes a protein that is a class I catalytic subunit of PI3K. Since PI3K is found in the immune system, which identifies and attacks viruses and bacteria, it prevents infection.
NFKB1	The NFKB1 is a transcription regulator that is activated by various cellular stimuli like cytokines, bacterial and viral products. It has a major role in the regulation of the early response to viral infection. Inappropriate activation of NFKB1 has been associated with several inflammatory diseases [55].
IKKBK	The IKKBK encodes a protein that causes dissociation of the inhibitor and activation of NF- κ B which is activated by multiple stimuli such as inflammatory cytokines, bacterial or viral products [55].
PIK3CA	The PIK3CA provides instructions for PI3K. Since PI3K is found in the immune system, which identifies and attacks viruses and bacteria, it prevents infection [55].

these pathways produce proinflammatory cytokines. Other studies also showed the cross-talk between TLRs and PI3K/AKT signaling pathways. Fig. 5 shows the cross-talk between TLR signaling pathways and both PI3K/AKT and MAPK signaling pathways as well. Toll-like receptors (TLRs), as a family of ten members in humans, play an important role in identifying invasive pathogens and establishing an innate and adaptive connection. These receptors are expressed in immune cells and some other cells, but their regulation and expression in cells are associated with specific molecules derived from pathogens or damaged cells. The binding of the ligands to the TLRs activates the signaling cascade and initiates the body's innate immune response. This immune response of cells leads to the production of cytokine proteins that can limit the spread of the virus in the body. But some inflammatory cytokines such as TNF- α , IFNs, and IL6 are also stimulated and produced. Failure to regulate these inflammatory responses in the immune system and a sudden increase in cytokine production can lead to loss of lung function, shock, and even death in patients with COVID-19. Therefore, our results indicate the importance of TLRs in all other mentioned signaling pathways, which could demonstrate the potential link between the unknown COVID-19 pathways and the TLRs receptors. Several trials are being conducted regarding TLRs' pathways in COVID-19 that could lead to the finding of new drugs for COVID-19 treatment.

The results of our proposed model on a biological network corresponding to disease pathways showed that three disease pathways, pathways in cancer, Influenza A, and HTLV-I infection, were selected through our method as the top three disease pathways. Some of the important genes from these pathways are TNFR1, JAK1, IKK2, TP53, STAT3, and MYC, respectively. Recent studies have shown that these topologically important genes from our top disease pathways can be targeted by COVID-19 candidate drugs such as Baricitinib, Ruxolitinib, Tofacitinib, and Acetylcysteine, which are already approved or are undergoing clinical trials [25]. This could confirm that our method selected some important disease pathways and related genes. It can be

concluded that 11 common and important genes that are selected through our method in all three disease pathways (Table 4) could be good candidates as drug targets for further research in clinical trials for possible COVID-19 treatment.

CRedit authorship contribution statement

Golnaz Taheri: Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Mahnaz Habibi:** Data curation, Formal analysis, Methodology, Software, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] P. Habibzadeh, E.K. Stoneman, The novel coronavirus: a bird's eye view, *Int. J. Occup. Environ. Med.* 11 (2) (2020) 65.
- [2] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin, *Nature* 579 (7798) (2020) 270–273.
- [3] K. Thorlund, L. Dron, J. Park, G. Hsu, J.I. Forrest, E.J. Mills, A real-time dashboard of clinical trials for COVID-19, *Lancet Digit. Health* 2 (6) (2020) e286–e287.
- [4] S. Zijie, X. Yan, K. Lu, M. Wentai, S. Leisheng, Z. Li, Z. Zhuo, Y. Jing, Z. Jiaxin, Y. Donghong, et al., Genomic diversity of SARS-CoV-2 in coronavirus disease 2019 patients, *Clin. Infect. Dis.* (2020).
- [5] T. Phan, Genetic diversity and evolution of SARS-CoV-2, *Infect. Genet. Evol.* 81 (1) (2020) 104260.
- [6] R. Aghdam, M. Habibi, G. Taheri, Using informative features in machine learning based method for COVID-19 drug repurposing, *J. Cheminformatics* 13 (1) (2021) 1–14.
- [7] S. Rahmati, M. Abovsky, C. Pastrello, M. Kotlyar, R. Lu, C.A. Cumbaa, P. Rahman, V. Chandran, I. Jurisica, pathDIP 4: an extended pathway annotations and enrichment analysis resource for human, model organisms and domesticated species, *Nucleic Acids Res.* 48 (D1) (2020) D479–D488.

- [8] R. Stoney, D.L. Robertson, G. Nenadic, J.-M. Schwartz, Mapping biological process relationships and disease perturbations within a pathway network, *NPJ Syst. Biol. Appl.* 4 (1) (2018) 1–11.
- [9] X. Jiang, F. Feyertag, C.J. Meehan, G.P. McCormack, S.A. Travers, C. Craig, M. Westby, M. Lewis, D.L. Robertson, Characterizing the diverse mutational pathways associated with R5-tropic maraviroc resistance: HIV-1 that uses the drug-bound CCR5 coreceptor, *J. Virol.* 89 (22) (2015) 11457–11472.
- [10] M. Costanzo, A. Baryshnikova, J. Bellay, Y. Kim, E.D. Spear, C.S. Sevier, H. Ding, J.L. Koh, K. Toufighi, S. Mostafavi, et al., The genetic landscape of a cell, *Science* 327 (5964) (2010) 425–431.
- [11] G. Taheri, M. Ayati, L. Wong, C. Eslahchi, Two scenarios for overcoming drug resistance by co-targeting, *Int. J. Bioinform. Res. Appl.* 11 (1) (2015) 72–89.
- [12] M.A. García-Campos, J. Espinal-Enríquez, E. Hernández-Lemus, Pathway analysis: state of the art, *Front. Physiol.* 6 (1) (2015) 383.
- [13] Q. Zhou, F. Zhao, Z.-p. Lv, C.-g. Zheng, W.-d. Zheng, L. Sun, N.-n. Wang, S. Pang, F.M. de Andrade, M. Fu, et al., Association between APOC1 polymorphism and Alzheimer's disease: a case-control study and meta-analysis, *PLoS One* 9 (1) (2014) e87017.
- [14] B. Jin, W. Wang, G. Du, G. Huang, L. Han, Z. Tang, D. Fan, J. Li, S. Zhang, Identifying hub genes and dysregulated pathways in hepatocellular carcinoma, *Eur. Rev. Med. Pharmacol. Sci.* 19 (4) (2015) 592–601.
- [15] J. Liu, C. Zhang, W. Hu, Z. Feng, Tumor suppressor p53 and its mutants in cancer metabolism, *Cancer Lett.* 356 (2) (2015) 197–203.
- [16] Y. Aoki, T. Niihori, T. Banjo, N. Okamoto, S. Mizuno, K. Kurosawa, T. Ogata, F. Takada, M. Yano, T. Ando, et al., Gain-of-function mutations in RIT1 cause noonan syndrome, a RAS/MAPK pathway syndrome, *Am. J. Hum. Genet.* 93 (1) (2013) 173–180.
- [17] G.S. Randhawa, M.P. Soltysiak, H. El Roz, C.P. de Souza, K.A. Hill, L. Kari, Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study, *PLoS One* 15 (4) (2020) e0232391.
- [18] H.C. Metsky, C.A. Freije, T.-S.F. Kosoko-Thoroddsen, P.C. Sabeti, C. Myhrvold, CRISPR-based COVID-19 surveillance using a genomically-comprehensive machine learning approach, 2020, *BioRxiv*.
- [19] L. Yan, H.-T. Zhang, Y. Xiao, M. Wang, C. Sun, J. Liang, S. Li, M. Zhang, Y. Guo, Y. Xiao, et al., Prediction of survival for severe COVID-19 patients with three clinical features: development of a machine learning-based prognostic model with clinical data in Wuhan, 2020, *MedRxiv*.
- [20] Y. Wang, M. Hu, Q. Li, X.-P. Zhang, G. Zhai, N. Yao, Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 in an accurate and unobtrusive manner, 2020, *arXiv preprint arXiv:2002.05534*.
- [21] O. Gozes, M. Frid-Adar, H. Greenspan, P.D. Browning, H. Zhang, W. Ji, A. Bernheim, E. Siegel, Rapid ai development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis, 2020, *arXiv preprint arXiv:2003.05037*.
- [22] A. Zhavoronkov, B. Zagribelnyy, A. Zhebrak, V. Aladinskiy, V. Terentiev, Q. Vanhaelen, D.S. Bezrukov, D. Polykovskiy, R. Shayakhmetov, A. Filimonov, et al., Potential non-covalent SARS-CoV-2 3C-like protease inhibitors designed using generative deep learning approaches and reviewed by human medicinal chemist in virtual reality, 2020, *BioRxiv*.
- [23] A.W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Židek, A.W. Nelson, A. Bridgland, et al., Improved protein structure prediction using potentials from deep learning, *Nature* 577 (7792) (2020) 706–710.
- [24] E. Ong, M.U. Wong, A. Huffman, Y. He, COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning, *Front. Immunol.* 11 (2020) 1581.
- [25] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, et al., DrugBank 5.0: a major update to the DrugBank database for 2018, *Nucleic Acids Res.* 46 (D1) (2018) D1074–D1082.
- [26] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, KEGG as a reference resource for gene and protein annotation, *Nucleic Acids Res.* 44 (D1) (2016) D457–D462.
- [27] G. Taheri, M. Habibi, L. Wong, C. Eslahchi, Disruption of protein complexes, *J. Bioinform. Comput. Biol.* 11 (03) (2013) 1341008.
- [28] G.S. Tomar, N.S. Chaudhari, J.L.V. Barbosa, M.K. Aghwariya, International Conference on Intelligent Computing and Smart Communication 2019: Proceedings of ICSC 2019, Springer Nature, 2020.
- [29] D.E. Gordon, G.M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K.M. White, M.J. O'Meara, V.V. Rezelj, J.Z. Guo, D.L. Swaney, et al., A SARS-CoV-2 protein interaction map reveals targets for drug repurposing, *Nature* 583 (7816) (2020) 459–468.
- [30] M. Habibi, G. Taheri, R. Aghdam, A SARS-CoV-2 (COVID-19) biological network to find targets for drug repurposing, *Sci. Rep.* 11 (1) (2021) 1–15.
- [31] A. Chatr-Aryamontri, R. Oughtred, L. Boucher, J. Rust, C. Chang, N.K. Kolas, L. O'Donnell, S. Oster, C. Theesfeld, A. Sellam, et al., The BioGRID interaction database: 2017 update, *Nucleic Acids Res.* 45 (D1) (2017) D369–D379.
- [32] D. Alonso-López, F.J. Campos-Laborie, M.A. Gutiérrez, L. Lambourne, M.A. Calderwood, M. Vidal, J. De Las Rivas, APID database: redefining protein-protein interaction experimental evidences and binary interactomes, *Database* 2019 (D1) (2019) 1–10.
- [33] A. Patil, H. Nakamura, Hint: a database of annotated protein-protein interactions and their homologs, *Biophysics* 1 (1) (2005) 21–24.
- [34] G. Alanis-Lobato, M.A. Andrade-Navarro, M.H. Schaefer, HIPPIE v2. 0: enhancing meaningfulness and reliability of protein-protein interaction networks, *Nucleic Acids Res.* 45 (D1) (2016) D408–D414.
- [35] K. Luck, D.-K. Kim, L. Lambourne, K. Spirohn, B.E. Begg, W. Bian, R. Brignall, T. Cafarelli, F.J. Campos-Laborie, B. Charloteaux, et al., A reference map of the human binary protein interactome, *Nature* 580 (7803) (2020) 402–408.
- [36] U. Consortium, UniProt: a worldwide hub of protein knowledge, *Nucleic Acids Res.* 47 (D1) (2019) 506–515.
- [37] M. Habibi, G. Taheri, Topological network based drug repurposing for coronavirus 2019, *PLoS One* 16 (7) (2021) e0255270.
- [38] E. Pairo-Castineira, S. Clohisey, L. Klaric, A.D. Bretherick, K. Rawlik, D. Pasko, J. Walker, N. Parkinson, M.H. Fourman, C.D. Russell, et al., Genetic mechanisms of critical illness in COVID-19, *Nature* 591 (7848) (2021) 92–98.
- [39] M.E. Dolan, D.P. Hill, G. Mukherjee, M.S. McAndrews, E.J. Chesler, J.A. Blake, Investigation of COVID-19 comorbidities reveals genes and pathways coincident with the SARS-CoV-2 viral disease, *Sci. Rep.* 10 (1) (2020) 1–11.
- [40] D. Barh, A.A. Aljabali, M.M. Tambuwala, S. Tiwari, A. Serrano-Aroca, K.J. Alzahrani, B. Silva Andrade, V. Azevedo, N.K. Ganguly, K. Lundstrom, Predicting COVID-19-comorbidity pathway crosstalk-based targets and drugs: Towards personalized COVID-19 management, *Biomedicine* 9 (5) (2021) 556.
- [41] M.R. Khezri, PI3K/AKT signaling pathway: a possible target for adjuvant therapy in COVID-19, *Hum. Cell* 34 (2) (2021) 700–701.
- [42] Q. Tao, J. Du, X. Li, J. Zeng, B. Tan, J. Xu, W. Lin, X.-l. Chen, Network pharmacology and molecular docking analysis on molecular targets and mechanisms of huashi baidu formula in the treatment of COVID-19, *Drug Dev. Ind. Pharm.* 46 (8) (2020) 1345–1353.
- [43] S. Khanmohammadi, N. Rezaei, Role of Toll-like receptors in the pathogenesis of COVID-19, *J. Med. Virol.* 93 (5) (2021) 2735–2739.
- [44] E.A. Toraih, J.A. Sedhom, T.M. Dokunmu, M.H. Hussein, E.M. Ruiz, K. Muthusamy, M. Zerfaoui, E. Kandil, Hidden in plain sight: The effects of BCG vaccination in the COVID-19 pandemic, *J. Med. Virol.* 93 (4) (2021) 1950–1966.
- [45] Z. Zhou, L. Ren, L. Zhang, J. Zhong, Y. Xiao, Z. Jia, L. Guo, J. Yang, C. Wang, S. Jiang, et al., Heightened innate immune responses in the respiratory tract of COVID-19 patients, *Cell Host Microbe* 27 (6) (2020) 883–890.
- [46] A. Vallée, Y. Lecarpentier, J.-N. Vallée, Interplay of opposing effects of the WNT/ β -catenin pathway and PPAR γ and implications for SARS-CoV2 treatment, *Front. Immunol.* 12 (1) (2021) 666693.
- [47] S. Sharma, A. Ray, B. Sadasivam, Metformin in COVID-19: a possible role beyond diabetes, *Diabetes Res. Clin. Pract.* 164 (1) (2020) 108183.
- [48] K.D. Tuttle, R. Minter, K.A. Waugh, P. Araya, M. Ludwig, C. Sempeck, K. Smith, Z. Andrysiak, M.A. Burchill, B.A. Tamburini, et al., JAK1 inhibition blocks lethal sterile immune responses: implications for COVID-19 therapy, 2020, *BioRxiv*.
- [49] M. Kandasamy, NF- κ B signalling as a pharmacological target in COVID-19: potential roles for IKK β inhibitors, *Naunyn-Schmiedeberg's Arch. Pharmacol.* 394 (3) (2021) 561–567.
- [50] Z. Leng, R. Zhu, W. Hou, Y. Feng, Y. Yang, Q. Han, G. Shan, F. Meng, D. Du, S. Wang, et al., Transplantation of ACE2-mesenchymal stem cells improves the outcome of patients with COVID-19 pneumonia, *Aging Dis.* 11 (2) (2020) 216.
- [51] M.J. Ramaiah, mTOR inhibition and p53 activation, microRNAs: The possible therapy against pandemic COVID-19, *Gene Rep.* 20 (1) (2020) 100765.
- [52] T. Matsuyama, S.P. Kubli, S.K. Yoshinaga, K. Pfeffer, T.W. Mak, An aberrant STAT pathway is central to COVID-19, *Cell Death Differ.* 27 (12) (2020) 3209–3225.
- [53] Y. Palacios, A. Ruiz, L.A. Ramón-Luig, R. Ocaña-Guzman, O. Barreto-Rodríguez, A. Sánchez-Monciváis, B. Tecuatzí-Cadena, A.G. Regalado-García, R.D. Pineda-Gudiño, A. García-Martínez, et al., Severe COVID-19 patients show an increase in soluble TNFR1 and ADAM17, with a relationship to mortality, *Int. J. Mol. Sci.* 22 (16) (2021) 8423.
- [54] J. Park, H. Kim, S.Y. Kim, Y. Kim, J.-S. Lee, K. Dan, M.-W. Seong, D. Han, In-depth blood proteome profiling analysis revealed distinct functional characteristics of plasma proteins between severe and non-severe COVID-19 patients, *Sci. Rep.* 10 (1) (2020) 1–10.
- [55] G. Stelzer, N. Rosen, I. Plaschkes, S. Zimmerman, M. Twik, S. Fishilevich, T.I. Stein, R. Nudel, I. Lieder, Y. Mazor, et al., The GeneCards suite: from gene data mining to disease genome sequence analyses, *Curr. Protoc. Bioinform.* 54 (1) (2016) 1–30.