Original Paper

Toward Preparing a Knowledge Base to Explore Potential Drugs and Biomedical Entities Related to COVID-19: Automated Computational Approach

Junaed Younus Khan[1], BSc; Md Tawkat Islam Khondaker[1], BSc; Iram Tazim Hoque[1], BSc; Hamada R H Al-Absi[2], MSc; Mohammad Saifur Rahman[1], PhD; Reto Guler[3,4,5], PhD; Tanvir Alam[2], PhD; M Sohel Rahman[1], PhD

[1]Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

[2]College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

[3]International Centre for Genetic Engineering and Biotechnology, Cape Town Component, Cape Town, South Africa

[4]Division of Immunology and South African Medical Research Council Immunology of Infectious Diseases, Department of Pathology, Institute of Infectious Diseases and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

[5]Wellcome Centre for Infectious Diseases Research in Africa, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

Corresponding Author:
Tanvir Alam, PhD
College of Science and Engineering
Hamad Bin Khalifa University
PO Box 34110
Education City
Doha
Qatar
Phone: 974 44542277
Email: talam@hbku.edu.qa

Abstract

Background: Novel coronavirus disease 2019 (COVID-19) is taking a huge toll on public health. Along with the non-therapeutic preventive measurements, scientific efforts are currently focused, mainly, on the development of vaccines and pharmacological treatment with existing drugs. Summarizing evidences from scientific literatures on the discovery of treatment plan of COVID-19 under a platform would help the scientific community to explore the opportunities in a systematic fashion.

Objective: The aim of this study is to explore the potential drugs and biomedical entities related to coronavirus related diseases, including COVID-19, that are mentioned on scientific literature through an automated computational approach.

Methods: We mined the information from publicly available scientific literature and related public resources. Six topic-specific dictionaries, including human genes, human miRNAs, diseases, Protein Databank, drugs, and drug side effects, were integrated to mine all scientific evidence related to COVID-19. We employed an automated literature mining and labeling system through a novel approach to measure the effectiveness of drugs against diseases based on natural language processing, sentiment analysis, and deep learning. We also applied the concept of cosine similarity to confidently infer the associations between diseases and genes.

Results: Based on the literature mining, we identified 1805 diseases, 2454 drugs, 1910 genes that are related to coronavirus related diseases including COVID-19. Integrating the extracted information, we developed the first knowledgebase platform dedicated to COVID-19, which highlights potential list of drugs and related biomedical entities. For COVID-19, we highlighted multiple case studies on existing drugs along with a confidence score for their applicability in the treatment plan. Based on our computational method, we found Remdesivir, Statins, Dexamethasone, and Ivermectin could be considered as potential effective drugs to improve clinical status and lower mortality in patients hospitalized with COVID-19. We also found that Hydroxychloroquine could not be considered as an effective drug for COVID-19. The resulting knowledgebase is made available as an open source tool, named COVID-19Base.

Conclusions: Proper investigation of the mined biomedical entities along with the identified interactions among those would help the research community to discover possible ways for the therapeutic treatment of COVID-19.

(JMIR Med Inform 2020;8(11):e21648) doi: 10.2196/21648

KEYWORDS

COVID-19; 2019-nCoV; coronavirus; SARS-CoV-2; SARS; remdesivir; statin; statins; dexamethasone; ivermectin; hydroxychloroquine

Introduction

SARS-CoV-2 initially spread widely in China, then in Italy, and has since been reported worldwide [1,2]. SARS-CoV-2 is a novel coronavirus that causes COVID-19 [3]. Although SARS-CoV-2 has gained attention as a consequence of the global COVID-19 pandemic, other known human coronaviruses, including betacoronaviruses (SARS-CoV, MERS, OC43, HKU1) and alphacoronaviruses (229E, NL63), have resulted in severe respiratory syndrome in patients and been of public health concern [4]. To combat COVID-19, an urgent solution is needed for the detection and therapeutic treatment of this disease, which requires a comprehensive experimental investigation of relevant biomedical entities (eg, genes, noncoding ribonucleic acids [ncRNA], viruses, drugs) [5]. However, this is a relatively slow process due to the inherent nature of experimental validation. As an alternative, faster in silico methods can be applied [6,7], which can act as a filter prior to wet lab validation. Virtual screening, molecular docking, and other in silico methods have already been investigated to discover drugs that may work against COVID-19 [8]. Still, this is a daunting task due to the large number of possible combinations of biomedical entities (eg, drug-gene pairs) that need to be examined [9]. To enable comprehensive exploration of potential therapeutic treatments, knowledge base solutions are proposed; these would allow the scientific community to focus on a relatively smaller number of potential biomedical entities that may lead to the discovery of a novel treatment for COVID-19.

Databases that focus on virus-related diseases for multiple hosts already exist. For example, in ViRBase [10], the authors highlighted the association between ncRNAs and viruses in 20 hosts. The VISDB database, based on literature mining, integrated the virus interaction site in humans for five DNA oncoviruses and four RNA retroviruses [11]. Virus Pathogen Resources (VIPR) developed a portal that collected a comprehensive set of information related to coronavirus and hepatitis C virus (HCV), as well as other viruses [12,13]. However, none of the abovementioned databases are particularly useful for COVID-19/SARS-CoV-2, as those databases were not specific to the novel coronavirus, or they provided very limited information about the associated genes, or they did not include other factors involved in coronavirus-related diseases, drugs, and drug side effects. Moreover, there is no one knowledge base that has integrated all biomedical entities specific to COVID-19/SARS-CoV-2. To address this gap, we explored the potential of machine intelligence to automatically mine the scientific literature, with the goal of developing the first comprehensive knowledge base that integrates several biomedical entities associated with COVID-19/SARS-CoV-2. To achieve this, we leveraged state-of-the-art natural language processing algorithms, sentiment analysis, and deep learning–based techniques and applied them to a large corpus of coronavirus-related scientific literature.

Methods

Data Sets

For this study, we used the COVID-19 Open Research Dataset (CORD-19) [14], generated by the Allen Institute for AI. The data set contains over 138,000 scholarly articles related to COVID-19 and the coronavirus family of viruses. The data set was collected using the following query to search PubMed, PubMed Central (PMC), bioRxiv, and medRxiv: “COVID-19” OR “Coronavirus” OR “Corona virus” OR “2019-nCoV” OR “SARS-CoV” OR “MERS-CoV” OR “Severe Acute Respiratory Syndrome” OR “Middle East Respiratory Syndrome.” This query covers most research articles related to COVID-19 and other coronaviruses (eg, MERS, SARS) and we searched up until June 9, 2020. Unless otherwise specified, we considered both the abstract and full body of the manuscripts (when available) for downstream analysis.

Source of Dictionaries

We collected gene names from HUGO Gene Nomenclature Committee (HGNC) [15], Protein Data Bank (PDB) entries from PDB [16], micro ribonucleic acids (miRNAs) from miRBase [17], disease names from Disease Ontology (DO) [18], drug names from DrugBank [19], and drug side effects from Side Effect Resource (SIDER) [20].

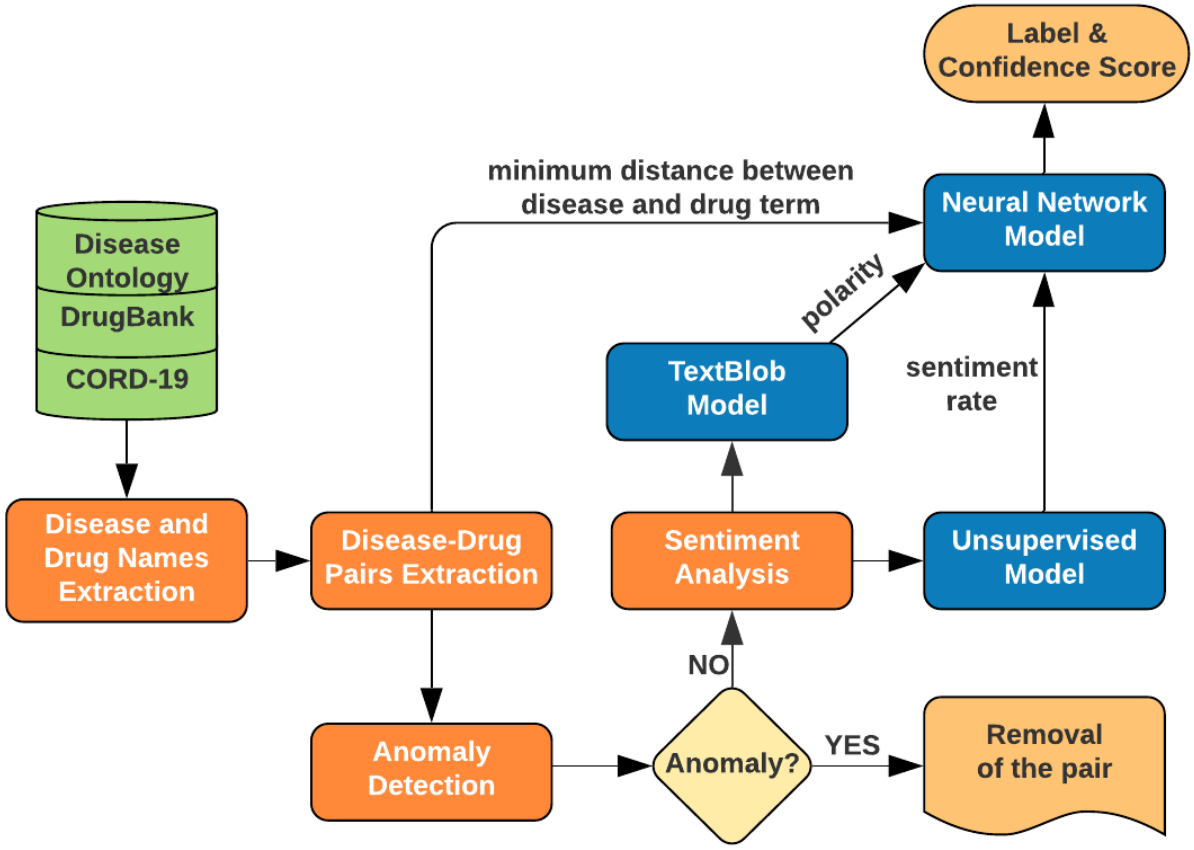
Overview of Methodology

We extracted disease-drug, disease-gene, drug-PDB pairs, and their corresponding sentences from the CORD-19 literature in a co-occurrence–based approach. To evaluate the effectiveness of the disease-drug pairs, we used both a pretrained model (TextBlob) and an unsupervised model (developed by the authors using the Word2Vec model and K-means clustering) to determine the sentiment scores of the sentences extracted for each pair. We further used these sentiment scores along with the minimum distance between the disease and drug term in the corresponding sentences as input features of our neural network model, which we used for the final classification of the disease-drug pairs (as positive or negative). To determine the confidence level of the extracted disease-gene associations, we transformed each disease and gene of a pair into two separate vectors using the Word2Vec model and calculated their cosine similarity. We used the known disease-gene associations from the DisGeNET database as the gold standard to determine the confidence level of the new associations on the basis of cosine similarity measures. Finally, we extracted the side effects of the drugs that were found by our mining from SIDER. Additionally, a feedback mechanism was incorporated into COVID-19Base to collect feedback from users for future use.

Extracting Disease-Drug Interactions

We extracted disease-drug interactions from the CORD-19 literature and classified them into one of two categories (labels): positive and negative. The positive label means the drug is potentially effective against COVID-19, and the negative label means the opposite. We also determined a confidence score, which indicates our level of confidence in that automatic label. Figure 1 shows the workflow of extracting disease-drug interactions and predicting the effectiveness of drugs against diseases with confidence scores.

Figure 1. Flowchart of extracting disease-drug interactions and predicting the effectiveness of drugs against diseases with confidence scores. CORD-19: COVID-19 Open Research Dataset.



Disease and Drug Name Extraction

To extract relevant disease-drug pairs from the CORD-19 literature, we employed a dictionary-based approach to detect mentions of diseases and drugs in the literature. We used Disease Ontology [18] and DrugBank [19] to prepare the disease and drug dictionaries. We leveraged the Aho-Corasick algorithm [21] to search the drug and disease names, considering the large size of the drug and disease dictionaries and the corpus itself. The Aho-Corasick algorithm is a string-searching algorithm that efficiently locates multiple patterns in a large amount of text. The time complexity of the algorithm is $O(n + m + z)$, where n is the length of the text, m is the total length of all the patterns to be searched, and z is the total number of occurrences of the patterns in the text.

Disease-Drug Pairs Extraction

After extracting the disease and drug names separately, we wanted to mine the literature and identify the sentences that contain the disease and drug pairs to semantically evaluate their interactions. For this purpose, we searched for every disease-drug pair from our disease and drug list in the CORD-19 literature and collected every sentence where a co-occurrence was found. We then created a document for every disease-drug pair, combining all extracted sentences. Thus, we built a disease-drug pair to document mapping. We did not use a pattern-based approach here (as was done previously in [22]) as this could result in missing some sentences containing disease-drug pairs.

Anomaly Removal

As we automatically extracted the sentences containing the disease-drug pairs, there was a possibility of errors in our extracted data; therefore, we decided to check and remove any abnormalities from our collected data before moving on to the next stage of the pipeline. We used unsupervised anomaly detection [23] for this task. Unsupervised anomaly detection detects anomalies in an unlabeled data set by looking for instances that seem to fit the remainder of the data set the least, under the assumption that the majority of the instances in the data set are “normal.” We used the K-means clustering algorithm

[24], as it has been used for anomaly detection in several studies [25-29]. We proceeded as follows. First, we used Doc2Vec [30] to create a numeric representation of each document associated with each disease-drug pair. We then fitted these representations into our K-means model and observed two clear clusters of easily discriminable sizes, where the smaller one consisted of only 189 instances. As we know that anomalies differ from the normal instances significantly and occur very rarely in the data, we could assume that the instances of the smaller cluster were indeed anomalies. We also checked a number of instances manually to verify our assumption. We discarded these 189 instances from any further consideration.

Sentiment Analysis

Overview

We applied sentiment analysis to automatically assess the effectiveness of a drug to treat a particular disease in the context of each extracted drug-disease pair. First, we applied the concept of transfer learning. We used TextBlob [31], which is a pretrained sentiment analysis tool provided as a Python library. However, it showed some inconsistency in some cases as expected from a pretrained model and we felt it necessary to perform unsupervised sentiment analysis, which is the second model in our pipeline. We obtained a polarity score from the TextBlob model and a sentiment rate from our unsupervised model for each disease-drug pair, which were subsequently fed to our neural network model to predict the final label.

TextBlob Model

TextBlob is a Python library that is widely used in natural language processing tasks such as part-of-speech (POS) tagging, noun phrase extraction, sentiment analysis, classification, and translation. Given the sentences that we mined for each disease-drug pair as input, TextBlob gives a polarity score between –1 and 1. We recorded the polarity scores for each disease-drug pair to use it as a feature for our neural network model.

Unsupervised Model

We used the concept of K-means clustering again for unsupervised sentiment analysis. First, we trained the Word2Vec [32] model with our mined literature and got a vector representation of every word. We then ran K-means clustering on the estimated word vectors and found two clusters (positive and negative). The positive cluster was decided on the basis of the presence of several positive words (in the context of a disease-drug pair), including “cure,” “preclude,” “inhibit,” “prescribe,” “reduce,” and “modest.” On the other hand, the negative cluster contained words like “risky,” “kill,” and “danger.” We then assigned each word a sentiment value, either +1 or –1, based on the cluster (positive or negative) they belong to. We weighed this value by dividing it by the distance between the word and the centroid of its cluster to describe the extent of its potential positiveness or negativeness. We then calculated the term frequency–inverse document frequency (tf–idf) score [33,34] of each word in the sentence collection to consider the significance of the unique words. Next, we built a tf–idf representation, T, for each disease-drug pair by replacing each word of the corresponding sentences with its tf–idf score and a sentiment value representation, S, by replacing each word with its sentiment value. Finally, we took their dot product ($T \cdot S$) as the final sentiment rate of our unsupervised model.

Neural Network Model for Automatic Labeling and Confidence Score

Overview

We used a deep neural network (DNN) model to automatically predict the label and confidence score for our disease-drug pairs. We used a relatively simpler neural network with two hidden layers as such models commonly perform better for smaller data sets compared to neural networks with many layers and parameters [35,36].

Training Data

We manually labeled 200 disease-drug pairs to train our neural network model. Among them, there were 110 positive instances and the rest were negative.

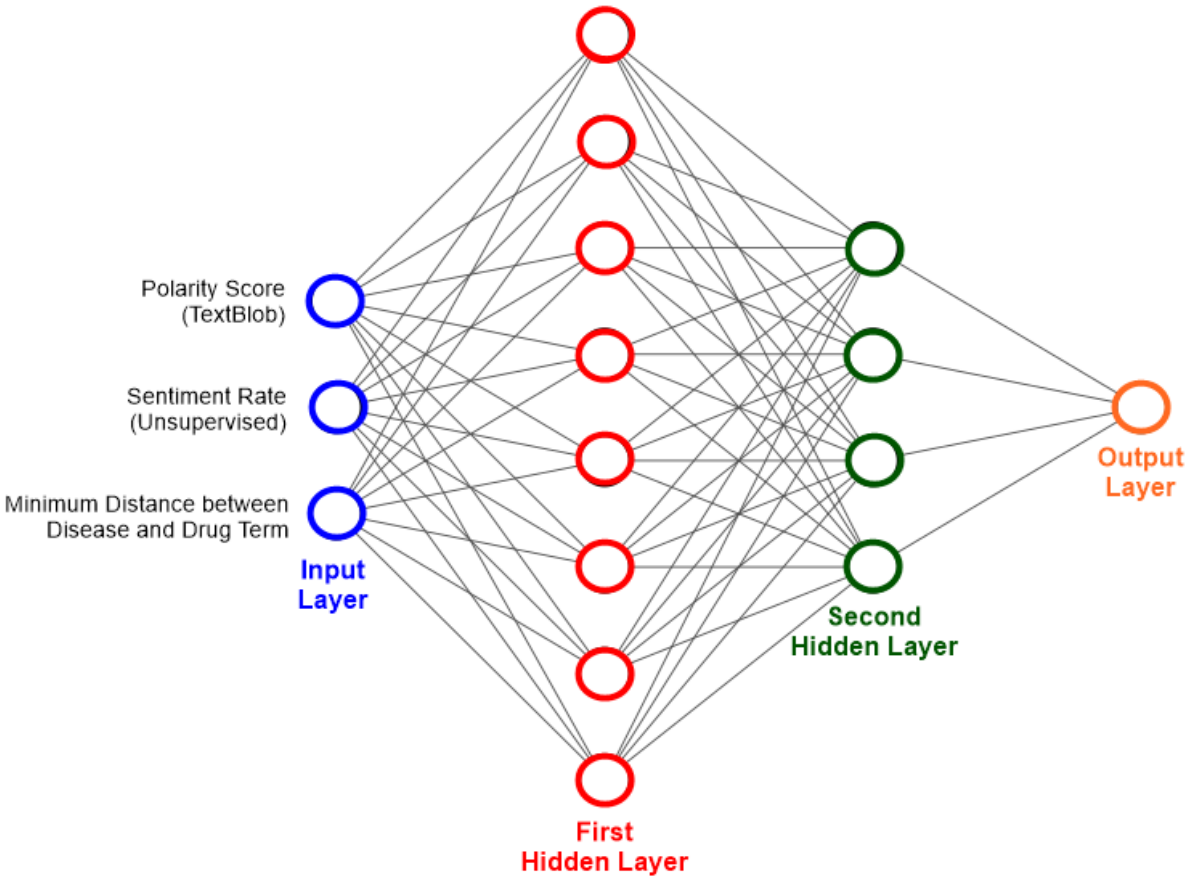
Input Features

We used the polarity or sentiment score given by the TextBlob and unsupervised models as the input features for our neural network model, along with the minimum distance between the disease and drug term in the corresponding document.

Model Setup and Output

The DNN structure used in this study is similar to that shown in Figure 2. It consists of one input layer with three neurons (each neuron corresponds to one input feature), two hidden layers with eight and four neurons respectively, and one output layer containing one neuron for binary classification (positive or negative). The transfer functions of the first and second hidden layers were the rectified linear unit (ReLU) [37] and hyperbolic tangent function (tanh) [38], respectively. The transfer function of the output layer was a sigmoid function [39]. We trained the DNN model using Xavier initialization [40], which tries to make the variance of the outputs of a layer equal to the variance of its inputs. We used Adam optimizer [41] and the maximum training epoch was set to 500. We split our labeled data into training and test sets on an 80:20 ratio. We trained our model on the training data and achieved 75% accuracy on the test set.

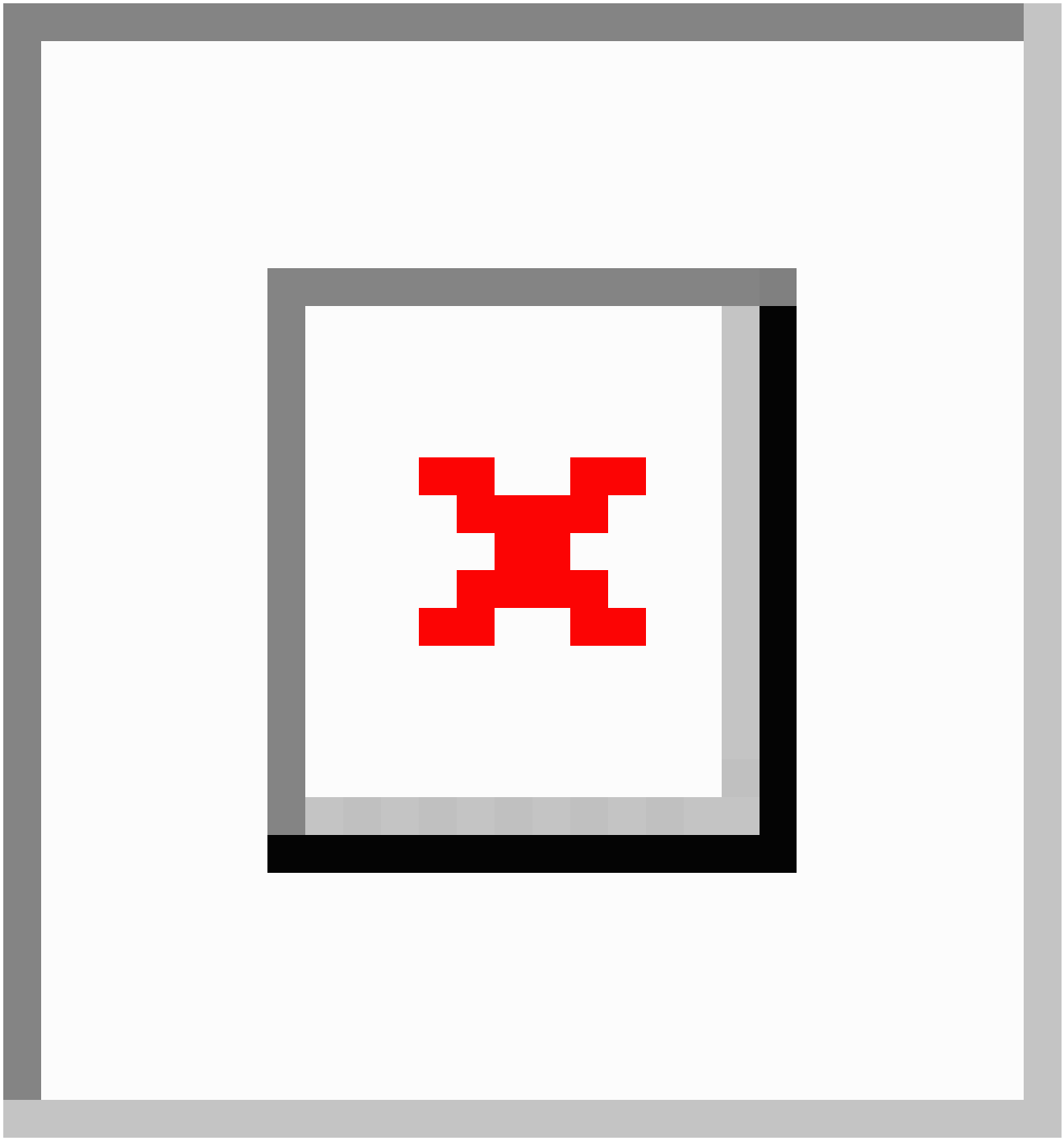
Figure 2. Schematic diagram of the deep neural network used to predict the effectiveness of drugs against diseases.



Extracting Disease-Gene Associations

Figure 3 shows the workflow of extracting disease-gene associations. We extracted gene names along with miRNAs from the CORD-19 literature in a dictionary-based approach using HGNC [15] and miRBase [17]. We then extracted their associations with diseases in a similar process to the one we had used to extract the disease-drug pairs and collected all the abstracts where a co-occurrence was found. Next, we applied the concept of cosine similarity [42] to confidently infer the associations. We transformed each disease into vector V_1, each gene (and miRNA) into vector V_2, and then calculated the cosine similarity of V_1 and V_2 for each pair. To create the vector representations, we trained a Word2Vec model with all the collected abstracts. We used the DisGeNET [43] database as the gold standard to evaluate the performance of cosine similarity in predicting the gene-disease linkage. First, we calculated the maximum, average, and minimum cosine similarity of the pairs that were common both in our findings and in the DisGeNET database. We found that 99.7% of the newly discovered pairs lie within this range (determined from DisGeNET) in terms of cosine similarity. We further classified the associations into three classes (high, medium, and low) in terms of confidence as follows: pairs having cosine similarity closest to the maximum (minimum) of the known ones were considered as high (low) confidence associations, and the remaining ones (those closest to the average) as medium confidence associations. Moreover, pairs that were also found in the DisGeNET database were labeled as verified associations.

Figure 3. Flowchart of extracting disease-gene and disease-miRNA associations and determining their confidence levels. CORD-19: COVID-19 Open Research Dataset; HGNC: HUGO Gene Nomenclature Committee; miRNA: micro ribonucleic acid.



Extracting Drug-Protein Associations

We also extracted drug-protein associations from the CORD-19 literature, applying the same co-occurrence–based approach as mentioned above. We used PDB IDs from the Protein Data Bank [16] for extracting protein names. Unlike the disease-gene associations, we did not apply the concept of cosine similarity here as we did not find any suitable data set that could be used as the gold standard in this case.

Extracting Side Effects of Drugs

The drugs we are suggesting through this literature mining may come with different side effects. Therefore, we also explored the possible side effects of the drugs. We collected the drugs with their corresponding side effects from SIDER [20] and mapped them with the drugs mentioned in the CORD-19 literature to extract the possible side effects.

Feedback Mechanism

We implemented a feedback mechanism in COVID-19Base for future improvement. This mechanism enables expert users from the scientific community to share their valuable feedback on the label (positive or negative) for a particular interaction determined by the automatic natural language processing–based approach. The users can voluntarily label each sentence that is mined from the literature as a source of an interaction. This feedback will be recorded and further processed to enrich the labeled data set, which can be leveraged in the next version of

COVID-19Base to further improve the prediction quality for determining effective disease-drug interactions. The accompanying tutorial (user manual) on COVID-19Base highlighted an example of how a user can use the feedback mechanism.

Results

Terms and Interactions Highlighted in CORD-19 Data Set

Based on our computational workflow, we identified 1805 diseases, 2454 drugs, 1910 genes, 11 miRNAs, and 70 PDB entries from the CORD-19 literature (Table 1). Among the disease-drug pairs, 21,581 were positive and 1318 were negative. Among the disease-gene associations, 2088 were verified, and 82 associations were found with high-confidence, 12,231 with medium-confidence, and 1488 with low-confidence. More results are shown in Table 1. Notably, a small proportion (1.5%) of the findings were manually labeled. Interestingly, we found 194 drug-PDB pairs for coronavirus-related diseases, which indicates the rapid growth of experimental work to understand the interaction mechanisms of drugs and target proteins.

Table 1. Pairs of terms as identified in the analyzed set of documents[a].

Interaction or association	Number of extracted pairs of terms
Disease-drug	22,899 (21,581 positive, 1318 negative)
Disease-gene	15,889 (2088 verified, 82 high, 12,231 medium, 1488 low)
Disease-miRNA	56 (48 medium, 8 low)
Drug-Protein Data Bank	194

[a]Positive (negative) indicates an (in)effective association. High, medium, and low refer to confidence associations.

COVID-19–Related Terms and Interactions

Our computational workflow identified 514 drugs and 417 genes that are directly associated with COVID-19 (Table 2). Among the 514 drugs, 492 were found to have a positive association and 22 had a negative association. Among the 417 genes, 347 were medium-confidence associations and 70 were low-confidence associations.

Table 2. Biomedical terms that are related to COVID-19[a].

Interaction or association	Number of extracted pairs of terms
COVID-19–drug	514 (492 positive, 22 negative)
COVID-19–gene	417 (347 medium, 70 low)
COVID-19–miRNA	3 (2 medium, 1 low)

[a]Positive (negative) indicates an (in)effective association. High, medium, and low refer to confidence associations.

Genes Related to COVID-19

Our automated workflow identified C-reactive protein (CRP) as one of the COVID-19–associated genes with "medium" confidence. CRP is a known clinical biomarker for SARS [44] and the level of CRP increases significantly in patients with SARS. The level of CRP was also higher for patients with COVID-19 in some clinical cases [45,46]. More than 25 papers (from the CORD-19 data set) related to the association between CRP and COVID-19 were identified through our computational workflow. Furthermore, the genes *ELANE*, *AZU1*, *MPO*, *PRTN3*, *CTSG*, and *TCN1* were shown to be significantly altered in patients with COVID-19 [47], and our automatically prepared knowledge base highlights all of them as associated with COVID-19 with "medium" or "low" confidence. The *ACE2* and *TMPRSS2* genes are known to be involved in SARS-CoV-2 infection [48]; in fact, SARS-CoV-2 uses angiotensin-converting enzyme 2 (ACE2) as a receptor for entry into host cells [49,50]. The spike protein of SARS-CoV-2 binds with the ACE2 receptor and the protease TMPRSS2 mediates the infection process [51]. It is important to note that *ACE2* and *TMPRSS2* were not directly listed in DisGeNET as genes associated with COVID-19. In spite of that, our data-driven approach based on a gold-standard data set from DisGeNET was able to infer the association of *ACE2* and *TMPRSS2* with COVID-19 with "medium" confidence, which suggests that our approach is efficacious. Analyzing the complete *ACE2* interaction network, Wicik et al [52] listed several element genes (*ACE2*, *ANPEP*, *DPP4*, *CCL2*, *MEPIA*, *TFRC*, *ADAM17*, *NPC1*, *FABP2*, *TMPRSS2*, *CLEC4M*) and all of these genes were identified as COVID-19–associated in our automatically prepared knowledge base. In addition, we mined three miRNAs (hsa-miR-4661-3p, hsa-miR-429, and hsa-miR-183) that were mentioned in the abstracts of COVID-19–related literature.

Case Studies

In this section, we discuss interesting and useful findings from our automatically prepared knowledge base in the context of potential drugs that can be investigated for the potential therapeutic treatment of COVID-19.

Case Study 1: Dexamethasone Can be Considered an Effective Drug for COVID-19

Dexamethasone, an inexpensive and commonly used steroid, is a major breakthrough in the fight against COVID-19. We found dexamethasone to be a positive (ie, effective) drug for COVID-19, automatically labeled as such through our computational workflow with a confidence score of 77.61%. Our computational workflow also discovered the effectiveness of this drug against pneumonia, respiratory failure, and diarrhea, which are strongly correlated to COVID-19 [53,54]. Thus, further exploration of this drug to fight COVID-19 is likely to be fruitful. Recent studies suggest that dexamethasone reduces the risk of death from COVID-19 from 40% to 28% for patients on ventilators and from 25% to 20% for patients needing oxygen [55].

Case Study 2: Ivermectin Might be Considered an Effective Drug for COVID-19

Ivermectin is an effective drug against pneumonia and diarrhea, and has recently been claimed to successfully treat patients with COVID-19 as well [56]. It is a US Food and Drug Administration (FDA)–approved drug used for parasitic infections, which has the potential to be repurposed. Ivermectin inhibits the replication of SARS-CoV-2 in vitro [57]. Recently, a team of medical doctors in Bangladesh reported quick recoveries of patients with COVID-19 using this drug [58]. We found ivermectin to be a positive (ie, effective) drug for COVID-19, automatically labeled with a confidence score of 77.91%. It was also labeled a positive drug for pneumonia and diarrhea in our knowledge base.

Case Study 3: Remdesivir Seems Effective Against COVID-19

Remdesivir has been identified as a positive (ie, effective) drug for COVID-19, automatically labeled as such through our pipeline, with a confidence score of 68.18%. Thus, it seems to be a promising drug for further investigation for treating COVID-19. Interestingly, it was recently being considered as an effective drug for treating COVID-19 [59]. Notably, remdesivir is an antiviral drug originally developed for Ebola treatment [60,61]. A recent clinical trial conducted by the National Institute of Allergy and Infectious Diseases (NIAID) showed that remdesivir helped patients with COVID-19 recover faster and improved their survival rates. Adult patients treated with remdesivir were found to recover 4 days faster, an improvement of 31% compared to other patients; in addition, the overall death rate dropped from 11.6% to 8% [62]. Remdesivir is now under consideration for use against COVID-19 in more than ten clinical trials [63]. We found 6LU7 was one of the PDB entries for remdesivir. After exploring the corresponding literature [64], we found that remdesivir was shown to be an effective inhibitor of the main SARS-CoV-2 protease using molecular docking [65,66].

Case Study 4: Hydroxychloroquine Is Not an Effective Treatment for COVID-19

Antimalaria drug hydroxychloroquine, which is one of the most talked-about drugs for treating COVID-19, was also found in our mining, albeit with a negative interaction. Our model found it is a negative (ie, ineffective) drug with 64.67% confidence. Additionally, it also revealed that this drug has 111 side effects including anemia, hemorrhage, liver disorder, hepatitis fulminant, cardiomyopathy, and cardiac failure, which makes it a risky option, especially for patients with heart and liver complications. Although the FDA had previously granted authorization to use this drug for COVID-19, it has recently cautioned against its use outside of a hospital setting or a clinical trial due to its side effects and risk factors [67].

Case Study 5: Statins Drugs Could be Effective Against COVID-19

Statins are effective as lipid-lowering drugs and mainly used for the treatment of cardiovascular diseases [68]. Statins are also well known for their anti - inflammatory effects [69] and some studies have supported the use of these drugs as part of a COVID - 19 treatment protocol [70]. Multiple clinical trials (eg, NCT04343001, NCT04380402) have been launched to determine the efficacy of statins against COVID-19 [71,72]. In our knowledge base, the majority of statin classes were shown to be effective against COVID-19. For example, ulinastatin, rosuvastatin, fluvastatin, and lovastatin were labeled as positive (ie, effective) drugs against COVID-19 with 94.04%, 79.38%, 78.88%, and 70.75% confidence scores, respectively. Through our automated computational workflow, we found only one mention of atorvastatin in the literature [73]. In that single article, Deliwala et al [73] mentioned atorvastatin as part of a prevention plan against cortical stroke for a 31-year-old female patient with COVID-19, without referring to the effectiveness of atorvastatin against COVID-19. Consequently, our knowledge base labeled atorvastatin with a negative sentiment and a rather low confidence score (61.22%) for COVID-19. We anticipate that as the number of articles related to atorvastatin use in COVID-19 treatment protocols increases, our model will be able to effectively infer the sentiment (effective versus ineffective) of this drug. Based on our finding, it is safe to state that statins, as low-cost and well-tolerated drugs, should be investigated in more detail in clinical trials; such drugs may help low- and middle-income countries in particular, where expensive drugs might not be affordable.

Discussion

Principal Findings

In our knowledge base, through a computational workflow, we not only extracted the drugs and other biomedical terms that are mentioned in the literature, but also identified “term pairs” based on their co-occurrence, which will allow the scientific community to investigate in depth the associations between term pairs like disease-gene and disease-drug. Many drugs were associated with COVID-19, representing the cumulative effort of the scientific community to repurpose existing drugs rather than pursue novel drug discoveries, which is a rational approach in a pandemic situation [74]. We leveraged an automated approach to highlight the effectiveness of drugs against the disease based on sentiment analysis of the text in the literature. Through this literature mining, we found dexamethasone, ivermectin, remdesivir, and others in the list of potential drugs for COVID-19 treatment. We highlighted hydroxychloroquine

as an ineffective drug against COVID-19. We extracted disease-gene associations from the literature and, based on cosine similarity against the gold-standard DisGeNET data set, provided a confidence level for the associations between diseases and genes. We found 194 drug-PDB associations, which highlighted the large amount of work performed by the scientific community to understand the mechanism behind drug-target interactions and virus-host protein interaction mechanisms for coronavirus-related diseases. Surprisingly, we found few miRNAs related to COVID-19, indicating the primary focus of the scientific community is toward protein-based drugs rather than RNA-based drugs, though there have been successful RNA-based antiviral drugs. One such drug is Miravirsen, which binds miR-122 to prevent it from hybridizing with the RNA genome of HCV, depriving HCV of its essential cellular cofactor and blocking HCV replication [75]. We expect more research along these lines in the coming months.

Research Implications

Currently, we are facing the largest public health emergency since the 1918 influenza outbreak [76]. From the beginning of this outbreak, the scientific community has invested large amounts of effort to create vaccines and identify therapeutic solutions. Vaccines for SARS-CoV-2 might come too late to have any effect on the first wave of the COVID-19 pandemic [77]. However, vaccines might be useful in subsequent waves of COVID-19 or in a postpandemic scenario in which COVID-19 becomes a seasonal virus [77]. In this scenario, the identification of drugs with good efficacy and minimal side effects is a rational goal that can be achieved in the near future to combat SARS-CoV-2 [48]. Although promising pharmacological results with repurposed drugs are emerging every day, unfortunately, no drug has been approved thus far for the treatment of COVID-19. Repurposed drugs are under investigation worldwide, many in preclinical and clinical stages [78]. With increasing information about SARS-CoV-2, along with publications about similar respiratory diseases (eg, pneumonia, SARS), it will be essential to investigate existing drugs that are already known to be effective against other respiratory diseases. As a prime example, dexamethasone, an FDA-approved drug, was known to be effective against pneumonia [79], respiratory failure [80], and other diseases. However, there was no evidence of its effectiveness against COVID-19 until its recent breakthrough in a clinical trial [55]. Although final approval of the drug is still pending, had it been investigated earlier, more lives could have been saved.

The research in this study is expected to support the scientific community and decision makers in identifying candidate drugs with proper evidence from the scientific literature. This will also help stakeholders explore existing drugs that are already known to be effective against other respiratory diseases. Although careful manual curation of the identified associations of biomedical entities is the ultimate goal, our novel approach estimates the effectiveness of drugs for coronavirus-related diseases based on natural language processing, sentiment analysis, and deep learning to help the scientific community shorten the potential list of drugs, ultimately saving time and resources.

Tool and Availability

We made our computational workflow and the resulting database an open source tool named COVID-19Base for use by the scientific community [81,82]. It not only identifies the terms and associations, but also highlights the relevant literature through its digital object identifier (DOI) so that any researcher using this tool can easily check the original source for more detailed information. As the number of scientific publications related to COVID-19 is constantly increasing, we will update the knowledge base on a monthly basis and integrate all recent updates in the knowledge base. COVID-19Base has already gone through its first transformation (from COVID-19Base 1.0 to COVID-19Base 2.0), as the CORD-19 data set was updated during the manuscript preparation phase. The earlier version of the CORD-19 data set contained about 44,000 papers, whereas the current version includes more than 138,000. The knowledge base materials and the source code of our computational approach are available on GitHub [83].

Limitations

Understandably, our findings as presented in the knowledge base may have some errors due to the inherent limitations of the methods and approaches adopted. This is why the identified inferences and associations are made available to users for review and a feedback mechanism is included in COVID-19Base.

Conclusions

We proposed a dictionary-based automated computational workflow to find the associations of six different thematic areas related to COVID-19/SARS-CoV-2 and other coronavirus-related diseases in humans. We prepared a knowledge base and made it available as a tool for the scientific community. We believe this knowledge base will help the research community explore the existing drugs and biomedical entities for coronavirus-related diseases, and the lessons learned before this outbreak will allow us to find an effective treatment for COVID-19.

Acknowledgments

The open access publication of this article was funded by the College of Science and Engineering, Hamad Bin Khalifa University.

Conflicts of Interest

None declared.

References

1. Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos Solitons Fractals 2020 May;134:109761 [FREE Full text] [doi: 10.1016/j.chaos.2020.109761] [Medline: 32308258]
2. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: 10.1126/science.aba9757] [Medline: 32144116]
3. Heymann DL, Shindo N. COVID-19: what is next for public health? The Lancet 2020 Feb;395(10224):542-545. [doi: 10.1016/s0140-6736(20)30374-3]
4. de Wit E, Rasmussen AL, Falzarano D, Bushmaker T, Feldmann F, Brining DL, et al. Middle East respiratory syndrome coronavirus (MERS-CoV) causes transient lower respiratory tract infection in rhesus macaques. Proc Natl Acad Sci U S A 2013 Oct 08;110(41):16598-16603. [doi: 10.1073/pnas.1310744110] [Medline: 24062443]
5. Alam I, Kamau AA, Kulmanov M, Jaremko Ł, Arold ST, Pain A, et al. Functional Pangenome Analysis Shows Key Features of E Protein Are Preserved in SARS and SARS-CoV-2. Front Cell Infect Microbiol 2020;10:405 [FREE Full text] [doi: 10.3389/fcimb.2020.00405] [Medline: 32850499]
6. Muralidharan N, Sakthivel R, Velmurugan D, Gromiha MM. Computational studies of drug repurposing and synergism of lopinavir, oseltamivir and ritonavir binding with SARS-CoV-2 protease against COVID-19. J Biomol Struct Dyn 2020 Apr 16:1-6. [doi: 10.1080/07391102.2020.1752802] [Medline: 32248766]
7. Stebbing J, Phelan A, Griffin I, Tucker C, Oechsle O, Smith D, et al. COVID-19: combining antiviral and anti-inflammatory treatments. The Lancet Infectious Diseases 2020 Apr;20(4):400-402. [doi: 10.1016/s1473-3099(20)30132-8]
8. Kandeel M, Al-Nazawi M. Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease. Life Sci 2020 Jun 15;251:117627 [FREE Full text] [doi: 10.1016/j.lfs.2020.117627] [Medline: 32251634]
9. Lavecchia A, Di Giovanni C. Virtual screening strategies in drug discovery: a critical review. Curr Med Chem 2013 Jun 01;20(23):2839-2860. [doi: 10.2174/09298673113209990001] [Medline: 23651302]
10. Li Y, Wang C, Miao Z, Bi X, Wu D, Jin N, et al. ViRBase: a resource for virus-host ncRNA-associated interactions. Nucleic Acids Res 2015 Jan;43(Database issue):D578-D582 [FREE Full text] [doi: 10.1093/nar/gku903] [Medline: 25274736]
11. Tang D, Li B, Xu T, Hu R, Tan D, Song X, et al. VISDB: a manually curated database of viral integration sites in the human genome. Nucleic Acids Res 2020 Jan 08;48(D1):D633-D641 [FREE Full text] [doi: 10.1093/nar/gkz867] [Medline: 31598702]
12. Zhang Y, Zmasek C, Sun G, Larsen CN, Scheuermann RH. Hepatitis C Virus Database and Bioinformatics Analysis Tools in the Virus Pathogen Resource (ViPR). Methods Mol Biol 2019;1911:47-69. [doi: 10.1007/978-1-4939-8976-8_3] [Medline: 30593617]
13. Pickett B, Greer D, Zhang Y, Stewart L, Zhou L, Sun G, et al. Virus pathogen database and analysis resource (ViPR): a comprehensive bioinformatics database and analysis resource for the coronavirus research community. Viruses 2012 Nov 19;4(11):3209-3226 [FREE Full text] [doi: 10.3390/v4113209] [Medline: 23202522]
14. Lu Wang L, Lo K, Chandrasekhar Y, Reas R, Yang J, Eide D, et al. CORD-19: The Covid-19 Open Research Dataset. ArXiv Preprint posted online on April 22, 2020. [Medline: 32510522]
15. Yates B, Braschi B, Gray KA, Seal RL, Tweedie S, Bruford EA. Genenames.org: the HGNC and VGNC resources in 2017. Nucleic Acids Res 2017 Jan 04;45(D1):D619-D625 [FREE Full text] [doi: 10.1093/nar/gkw1033] [Medline: 27799471]
16. Goodsell DS, Zardecki C, Di Costanzo L, Duarte JM, Hudson BP, Persikova I, et al. RCSB Protein Data Bank: Enabling biomedical research and drug discovery. Protein Sci 2020 Jan 29;29(1):52-65. [doi: 10.1002/pro.3730] [Medline: 31531901]
17. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic Acids Res 2008 Jan 23;36(Database issue):D154-D158 [FREE Full text] [doi: 10.1093/nar/gkm952] [Medline: 17991681]
18. Schriml L, Mitraka E, Munro J, Tauber B, Schor M, Nickle L, et al. Human Disease Ontology 2018 update: classification, content and workflow expansion. Nucleic Acids Res 2019 Jan 08;47(D1):D955-D962 [FREE Full text] [doi: 10.1093/nar/gky1032] [Medline: 30407550]
19. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 2006 Jan 01;34(Database issue):D668-D672 [FREE Full text] [doi: 10.1093/nar/gkj067] [Medline: 16381955]
20. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. Nucleic Acids Res 2016 Jan 04;44(D1):D1075-D1079 [FREE Full text] [doi: 10.1093/nar/gkv1075] [Medline: 26481350]
21. Aho AV, Corasick MJ. Efficient string matching. Commun ACM 1975 Jun;18(6):333-340. [doi: 10.1145/360825.360855]
22. Wang P, Hao T, Yan J, Jin L. Large-scale extraction of drug-disease pairs from the medical literature. Journal of the Association for Information Science and Technology 2017 Jun 06;68(11):2649-2661. [doi: 10.1002/asi.23876]
23. Zimek A, Schubert E, Kriegel H. A survey on unsupervised outlier detection in high-dimensional numerical data. Statistical Analy Data Mining 2012 Aug 27;5(5):363-387. [doi: 10.1002/sam.11161]
24. Lloyd S. Least squares quantization in PCM. IEEE Trans Inform Theory 1982 Mar;28(2):129-137. [doi: 10.1109/tit.1982.1056489]
25. Han L. Using a dynamic K-means algorithm to detect anomaly activities. : IEEE; 2012 Presented at: Seventh International Conference on Computational Intelligence and Security; December 3-4, 2011; Hainan, China. [doi: 10.1109/cis.2011.233]

26. Lima M, Zarpelão BB, Sampaio LDH, Rodrigues JJPC, Abrão T, Proença ML. Anomaly detection using baseline and k-means clustering. : IEEE; 2010 Presented at: SoftCOM 2010, 18th International Conference on Software, Telecommunications and Computer Networks; 2010; Split, Dubrovnik, Croatia p. 305-309.
27. Lu W, Traore I. Unsupervised anomaly detection using an evolutionary extension of k-means algorithm. IJICS 2008;2(2):107. [doi: 10.1504/ijics.2008.018513]
28. Syarif I, Prugel-Bennett A, Wills G. Unsupervised Clustering Approach for Network Anomaly Detection. In: Benlamri R, editor. Networked Digital Technologies. Berlin, Heidelberg: Springer; 2012.
29. Yasami Y, Mozaffari SP. A novel unsupervised classification approach for network anomaly detection by k-Means clustering and ID3 decision tree learning methods. J Supercomput 2009 Oct 9;53(1):231-245. [doi: 10.1007/s11227-009-0338-x]
30. Le QV, Mikolov T. Distributed Representations of Sentences and Documents. ArXiv Preprint posted online on May 16, 2014. [FREE Full text]
31. Textblob: simplified text processing. URL: https://textblob.readthedocs.io/en/dev/ [accessed 2020-11-03]
32. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. ArXiv Preprint posted online on January 16, 2013. [FREE Full text]
33. Sparck Jones K. A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation 1972 Jan;28(1):11-21. [doi: 10.1108/eb026526]
34. Luhn HP. A Statistical Approach to Mechanized Encoding and Searching of Literary Information. 1957 Oct;1(4):309-317. [doi: 10.1147/rd.14.0309]
35. Pasupa K, Sunhem W. A comparison between shallow and deep architecture classifiers on small dataset. 2016 Presented at: 8th International Conference on Information Technology and Electrical Engineering (ICITEE); 2016; Yogyakarta, Indonesia. [doi: 10.1109/iciteed.2016.7863293]
36. Feng S, Zhou H, Dong H. Using deep neural network with small dataset to predict material defects. 2019 Jan;162:300-310. [doi: 10.1016/j.matdes.2018.11.060]
37. Hahnloser RHR, Sarpeshkar R, Mahowald MA, Douglas RJ, Seung HS. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. Nature 2000 Jun 22;405(6789):947-951. [doi: 10.1038/35016072] [Medline: 10879535]
38. Karlik B, Olgac AV. Performance analysis of various activation functions in generalized MLP architectures of neural networks. International Journal of Artificial Intelligence and Expert Systems 2011;1(4):111-122 [FREE Full text]
39. Menon A, Mehrotra K, Mohan CK, Ranka S. Characterization of a Class of Sigmoid Functions with Applications to Neural Networks. Neural Networks 1996 Jul;9(5):819-835. [doi: 10.1016/0893-6080(95)00107-7]
40. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. 2010 Presented at: Proceedings of the 13th International Conference on Artificial Intelligence and Statistics; 2010; Sardinia, Italy.
41. Kingma DP, Ba J. Adam: A method for stochastic optimization. ArXiv Preprint posted online on December 22, 2014. [FREE Full text]
42. Singhal A. Modern information retrieval: A brief overview. IEEE Data Eng Bull 2001;24(4):1-43.
43. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res 2017 Jan 04;45(D1):D833-D839 [FREE Full text] [doi: 10.1093/nar/gkw943] [Medline: 27924018]
44. Wang J, Sheng W, Fang C, Chen Y, Wang J, Yu C, et al. Clinical manifestations, laboratory findings, and treatment outcomes of SARS patients. Emerg Infect Dis 2004 May;10(5):818-824 [FREE Full text] [doi: 10.3201/eid1005.030640] [Medline: 15200814]
45. Chen C, Chen C, Yan JT, Zhou N, Zhao JP, Wang DW. [Analysis of myocardial injury in patients with COVID-19 and association between concomitant cardiovascular diseases and severity of COVID-19]. Zhonghua Xin Xue Guan Bing Za Zhi 2020 Jul 24;48(7):567-571. [doi: 10.3760/cma.j.cn112148-20200225-00123] [Medline: 32141280]
46. Wang G, Wu C, Zhang Q, Wu F, Yu B, Lv J, et al. C-Reactive Protein Level May Predict the Risk of COVID-19 Aggravation. Open Forum Infect Dis 2020 May;7(5):ofaa153 [FREE Full text] [doi: 10.1093/ofid/ofaa153] [Medline: 32455147]
47. Akgun E, Tuzuner MB, Sahin B, Kilercik M, Kulah C, Cakiroglu HN, et al. Altered molecular pathways observed in naso-oropharyngeal samples of SARS-CoV-2 patients. medRxiv Preprint posted online on May 18, 2020. [FREE Full text] [doi: 10.1101/2020.05.14.20102558]
48. Potì F, Pozzoli C, Adami M, Poli E, Costa LG. Treatments for COVID-19: emerging drugs against the coronavirus. Acta Biomed 2020 May 11;91(2):118-136. [doi: 10.23750/abm.v91i2.9639] [Medline: 32420936]
49. Walls AC, Park Y, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 2020 Apr 16;181(2):281-292.e6 [FREE Full text] [doi: 10.1016/j.cell.2020.02.058] [Medline: 32155444]
50. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell 2020 Apr 16;181(2):271-280.e8 [FREE Full text] [doi: 10.1016/j.cell.2020.02.052] [Medline: 32142651]
51. Zang R, Gomez Castro MF, McCune BT, Zeng Q, Rothlauf PW, Sonnek NM, et al. TMPRSS2 and TMPRSS4 promote SARS-CoV-2 infection of human small intestinal enterocytes. Sci Immunol 2020 May 13;5(47) [FREE Full text] [doi: 10.1126/sciimmunol.abc3582] [Medline: 32404436]

52. Wicik Z, Eyileten C, Jakubik D, Simões SN, Martins Jr DC, Pavão R, et al. ACE2 interaction networks in COVID-19: a physiological framework for prediction of outcome in patients with cardiovascular risk factors. BioRxiv Preprint posted online on October 9, 2020. [doi: 10.1101/2020.05.13.094714]
53. He R, Lu Z, Zhang L, Fan T, Xiong R, Shen X, et al. The clinical course and its correlated immune status in COVID-19 pneumonia. J Clin Virol 2020 Jun;127:104361 [FREE Full text] [doi: 10.1016/j.jcv.2020.104361] [Medline: 32344320]
54. D'Amico F, Baumgart DC, Danese S, Peyrin-Biroulet L. Diarrhea During COVID-19 Infection: Pathogenesis, Epidemiology, Prevention, and Management. Clin Gastroenterol Hepatol 2020 Jul;18(8):1663-1672 [FREE Full text] [doi: 10.1016/j.cgh.2020.04.001] [Medline: 32278065]
55. RECOVERY Collaborative Group, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, et al. Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary Report. N Engl J Med 2020 Jul 17:7273 [FREE Full text] [doi: 10.1056/NEJMoa2021436] [Medline: 32678530]
56. Chaccour C, Hammann F, Ramón-García S, Rabinovich NR. Ivermectin and COVID-19: Keeping Rigor in Times of Urgency. Am J Trop Med Hyg 2020 Jun;102(6):1156-1157 [FREE Full text] [doi: 10.4269/ajtmh.20-0271] [Medline: 32314704]
57. Caly L, Druce JD, Catton MG, Jans DA, Wagstaff KM. The FDA-approved drug ivermectin inhibits the replication of SARS-CoV-2 in vitro. Antiviral Res 2020 Jun;178:104787 [FREE Full text] [doi: 10.1016/j.antiviral.2020.104787] [Medline: 32251768]
58. Sujan M. Use of Ivermectin: Hope held out, caution called for. The Daily Star. 2020. URL: https://www.thedailystar.net/frontpage/news/use-ivermectin-hope-held-out-caution-called-1914041 [accessed 2020-06-26]
59. Al-Tawfiq JA, Al-Homoud AH, Memish ZA. Remdesivir as a possible therapeutic option for the COVID-19. Travel Med Infect Dis 2020 Mar;34:101615 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101615] [Medline: 32145386]
60. Tchesnokov E, Feng J, Porter D, Götte M. Mechanism of Inhibition of Ebola Virus RNA-Dependent RNA Polymerase by Remdesivir. Viruses 2019 Apr 04;11(4):326 [FREE Full text] [doi: 10.3390/v11040326] [Medline: 30987343]
61. Warren TK, Jordan R, Lo MK, Ray AS, Mackman RL, Soloveva V, et al. Therapeutic efficacy of the small molecule GS-5734 against Ebola virus in rhesus monkeys. Nature 2016 Mar 17;531(7594):381-385 [FREE Full text] [doi: 10.1038/nature17180] [Medline: 26934220]
62. NIH Clinical Trial Shows Remdesivir Accelerates Recovery from Advanced COVID-19. 2020. URL: https://www.niaid.nih.gov/news-events/nih-clinical-trial-shows-remdesivir-accelerates-recovery-advanced-covid-19 [accessed 2020-06-26]
63. Clinical trials related to COVID-19. URL: https://clinicaltrials.gov/ct2/results?cond=COVID-19 [accessed 2020-06-26]
64. Hagar M, Ahmed HA, Aljohani G, Alhaddad OA. Investigation of Some Antiviral -Heterocycles as COVID 19 Drug: Molecular Docking and DFT Calculations. Int J Mol Sci 2020 May 30;21(11):3922 [FREE Full text] [doi: 10.3390/ijms21113922] [Medline: 32486229]
65. Grein J, Ohmagari N, Shin D, Diaz G, Asperges E, Castagna A, et al. Compassionate Use of Remdesivir for Patients with Severe Covid-19. N Engl J Med 2020 Jun 11;382(24):2327-2336 [FREE Full text] [doi: 10.1056/NEJMoa2007016] [Medline: 32275812]
66. Wang Y, Zhang D, Du G, Du R, Zhao J, Jin Y, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. Lancet 2020 May 16;395(10236):1569-1578 [FREE Full text] [doi: 10.1016/S0140-6736(20)31022-9] [Medline: 32423584]
67. FDA cautions against use of hydroxychloroquine or chloroquine for COVID-19 outside of the hospital setting or a clinical trial due to risk of heart rhythm problems. URL: https://www.fda.gov/drugs/drug-safety-and-availability/fda-cautions-against-use-hydroxychloroquine-or-chloroquine-covid-19-outside-hospital-setting-or [accessed 2020-05-12]
68. Cannon CP, Braunwald E, McCabe CH, Rader DJ, Rouleau JL, Belder R, et al. Intensive versus Moderate Lipid Lowering with Statins after Acute Coronary Syndromes. N Engl J Med 2004 Apr 08;350(15):1495-1504. [doi: 10.1056/nejmoa040583]
69. Dashti-Khavidaki S, Khalili H. Considerations for Statin Therapy in Patients with COVID-19. Pharmacotherapy 2020 May 04;40(5):484-486 [FREE Full text] [doi: 10.1002/phar.2397] [Medline: 32267560]
70. Castiglione V, Chiriacò M, Emdin M, Taddei S, Vergaro G. Statin therapy in COVID-19 infection. Eur Heart J Cardiovasc Pharmacother 2020 Jul 01;6(4):258-259 [FREE Full text] [doi: 10.1093/ehjcvp/pvaa042] [Medline: 32347925]
71. Coronavirus Response - Active Support for Hospitalised Covid-19 Patients (CRASH-19). 2020. URL: https://clinicaltrials.gov/ct2/show/NCT04343001 [accessed 2020-06-26]
72. Atorvastatin as Adjunctive Therapy in COVID-19 (STATCO19). 2020. URL: https://clinicaltrials.gov/ct2/show/NCT04380402 [accessed 2020-06-26]
73. Deliwala S, Abdulhamid S, Abusalih MF, Al-Qasmi MM, Bachuwa G. Encephalopathy as the Sentinel Sign of a Cortical Stroke in a Patient Infected With Coronavirus Disease-19 (COVID-19). Cureus 2020 May 14;12(5):e8121 [FREE Full text] [doi: 10.7759/cureus.8121] [Medline: 32426200]
74. Alexander S, Armstrong JF, Davenport AP, Davies JA, Faccenda E, Harding SD, et al. A rational roadmap for SARS-CoV-2/COVID-19 pharmacotherapeutic research and development: IUPHAR Review 29. Br J Pharmacol 2020 Nov;177(21):4942-4966 [FREE Full text] [doi: 10.1111/bph.15094] [Medline: 32358833]

75. Ottosen S, Parsley TB, Yang L, Zeh K, van Doorn L, van der Veer E, et al. Antiviral Activity and Preclinical and Clinical Resistance Profile of Miravirsen, a Novel Anti-Hepatitis C Virus Therapeutic Targeting the Human Factor miR-122. Antimicrob Agents Chemother 2014 Nov 10;59(1):599-608. [doi: 10.1128/aac.04220-14]
76. Shi Y, Wang Y, Shao C, Huang J, Gan J, Huang X, et al. COVID-19 infection: the perspectives on immune responses. Cell Death Differ 2020 May 23;27(5):1451-1454 [FREE Full text] [doi: 10.1038/s41418-020-0530-3] [Medline: 32205856]
77. Amanat F, Krammer F. SARS-CoV-2 Vaccines: Status Report. Immunity 2020 Apr 14;52(4):583-589 [FREE Full text] [doi: 10.1016/j.immuni.2020.03.007] [Medline: 32259480]
78. Rosa S, Santos WC. Clinical trials on drug repositioning for COVID-19 treatment. Rev Panam Salud Publica 2020;44:e40 [FREE Full text] [doi: 10.26633/RPSP.2020.40] [Medline: 32256547]
79. Remmelts HH, Meijvis SC, Heijligenberg R, Rijkers GT, Oosterheert JJ, Bos WJW, et al. Biomarkers define the clinical response to dexamethasone in community-acquired pneumonia. J Infect 2012 Jul;65(1):25-31. [doi: 10.1016/j.jinf.2012.03.008] [Medline: 22410382]
80. da Costa DE, Nair AK, Pai MG, Al Khusaiby SM. Steroids in full term infants with respiratory failure and pulmonary hypertension due to meconium aspiration syndrome. Eur J Pediatr 2001 Mar 14;160(3):150-153. [doi: 10.1007/s004310000678] [Medline: 11277374]
81. COVID-19Base 2.0. URL: https://covid-19base.hbku.edu.qa/search [accessed 2020-11-05]
82. COVID-19Base 2.0. URL: http://covid-19base.buet.ac.bd/search [accessed 2020-11-05]
83. COVID-19Base GitHub. URL: https://github.com/JunaedYounusKhan51/COVID-19Base [accessed 2020-11-03]

Abbreviations

ACE2: angiotensin-converting enzyme 2
CORD-19: COVID-19 Open Research Dataset
CRP: C-reactive protein
DNN: deep neural network
DO: Disease Ontology
FDA: Food and Drug Administration
HCV: hepatitis C virus
HGNC: HUGO Gene Nomenclature Committee
miRNA: micro ribonucleic acid
ncRNA: noncoding ribonucleic acid
NIAID: National Institute of Allergy and Infectious Diseases
PDB: Protein Data Bank
POS: part-of-speech
ReLU: rectified linear unit
tf–idf: term frequency–inverse document frequency

Edited by G Eysenbach, C Lovis; submitted 26.06.20; peer-reviewed by A Civit, A Sarafi Nejad; comments to author 21.08.20; revised version received 23.08.20; accepted 06.09.20; published 10.11.20

Please cite as:
Khan JY, Khondaker MTI, Hoque IT, Al-Absi HRH, Rahman MS, Guler R, Alam T, Rahman MS
Toward Preparing a Knowledge Base to Explore Potential Drugs and Biomedical Entities Related to COVID-19: Automated Computational Approach
JMIR Med Inform 2020;8(11):e21648
URL: http://medinform.jmir.org/2020/11/e21648/
doi: 10.2196/21648
PMID: 33055059