Computational and Bioinformatics Approaches for Identifying Comorbidities of COVID-19 Using Transcriptomic Data

Shudeb Babu Sen Omit*1, Md Mohiuddin2, Salma Akhter3, Md. Hasan Imam1, A. K. M. Mostofa Kamal Habib4, Syed Mohammad Meraz Hossain5, Nitun Kumar Podder6

1Institute of Information Technology, Noakhali Science and Technology University, Noakhali, 3814, Bangladesh
2Department of Mechanical Engineering, Chittagong University of Engineering and Technology, Chattogram, 4349, Bangladesh
3Department of Environmental Science and Disaster Management, Noakhali Science and Technology University, Noakhali 3814, Bangladesh
4National Skills Development Authority, Dhaka, Bangladesh
5Department of Information and Communication Technology, Dhaka, Bangladesh
6Department of Computer Science and Engineering, Pabna University of Science and Technology, Pabna, 6600, Bangladesh

*Corresponding author: Shudeb Babu Sen Omit, Email: omit.shen@gmail.com

ABSTRACT: Comorbidity is the co-existence of one or more diseases that occur concurrently or after the primary disease. Patients may have developed comorbidities for COVID-19 that cause harm to the patient’s organs. Besides, patients with existing comorbidities are at high risk, since mortality rates are strongly influenced by comorbidities or former health conditions. Therefore, we developed a computational and bioinformatics model to identify the comorbidities of COVID-19 utilizing transcriptome datasets of patient’s whole blood cells. In our model, we employed gene expression analysis to identify dysregulated genes and curated diseases from Gold Benchmark databases using the dysregulated genes. Subsequently, Tippett’s Method is used for COVID-19’s P-value calculation, and according to the P-value, Euclidean distances are calculated between COVID-19 and the collected diseases. Then the collected diseases are ordered and clustered based on the Euclidean distance. Finally, comorbidities are selected from the top clusters based on a comprehensive literature search. Applying the model, we found that acute myelocytic leukemia, cancer of urinary tract, body weight changes, abdominal aortic aneurysm, kidney neoplasm, diabetes mellitus, and some other rare diseases have correlation with COVID-19 and many of them reveal as comorbidity. Since comorbidities are in conjunction with the primary disease, thus similar drugs and treatments can be used to recover both COVID-19 and its comorbidities by further research. We also proposed that this model can be further useful for detecting comorbidities of other diseases as well.

KEYWORDS: COVID-19, Comorbidity Identification, Transcriptomic Data, Tippett’s Method, Euclidean Distance

1. Introduction

COVID-19 is an infectious disease first reported in December 2019 in Wuhan, Hubei Province, China, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Through direct contact or droplets from the infected person’s respiratory system, COVID-19 is spread between individuals [1], [2]. Consequently, COVID-19 has spread to almost all countries around the world, and on 11 March 2020, the World Health Organization (WHO) declared it a global pandemic. The WHO estimates that COVID-19 causes more than 180 million confirmed cases and 4 million deaths worldwide. This death is often caused by the underlying comorbidities of COVID-19, and the severity of the disease is increased by these existing conditions, as the comorbidities are correlated with morbidity and mortality [3] — [5]. In addition, COVID-19 indirectly impacts damage to organ systems and organs of the human body, which is also resulting in death [6] — [9]. It is therefore vital to develop a computation-based approach that can detect the comorbidities of COVID-19 so that early protection, prevention, and treatment can be implemented.

There have been several studies performed to identify COVID-19 comorbidities. In [10], the authors found that hypertension, diabetes, cardiovascular diseases, and
chronic kidney disease are the most common comorbidities of COVID-19. In [11], cardiovascular disease and chronic respiratory disease are the riskiest comorbidities observed by the authors. Another study performed in [12] and the authors showed that hypertension, diabetes, obstructive pulmonary disease, cardiovascular disease, and cerebrovascular disease are the risk factors associated with COVID-19. Moreover, according to [13], hypertension, chronic cardiovascular disease, diabetes, and chronic cerebrovascular disease have a high percentage in COVID-19 patients. Besides, in [14], the authors reported hypertension, diabetes mellitus, cerebrovascular, cardiovascular disease, kidney disease, etc. as comorbidities of COVID-19. A similar analysis by the authors of [15] also reported that hypertension, diabetes, obesity, asthma, etc. are the comorbidities in hospitalized COVID-19 patients.

We found that a considerable number of works identifying the comorbidities of COVID-19 are based on literature-driven meta-analyses and some of them are discussed in the literature review section. Therefore, we have performed a computational and bioinformatics-based approach to detect COVID-19’s comorbidities.

Our analysis was based on transcriptomic data derived from blood samples of COVID-19 patients and healthy individuals. Then we analyzed the dataset with the GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r/) tool and identified up and down-regulated genes. As a result of these dysregulated genes, we compiled a list of associated diseases from Gold Benchmark databases. Then the P-value of COVID-19 is calculated, as described in the following sections. After that, Euclidean distance is used to measure how close the collected diseases are to COVID-19. The Euclidean distances are calculated between COVID-19’s P-value and the Gold Benchmark diseases’ P-value. Consequently, the short Euclidean distances mean more proximity of diseases with COVID-19. Next, we organized the collected diseases in small to large orders according to the Euclidean distance. Then the diseases of similar Euclidean distance are clustered from top to bottom of the organized list. In our final step, we selected the comorbidity from the top clusters.

The main contributions of our paper are as follows:

- As a result of conducting our analysis on genetic data, we avoided the influences of literature-driven type biases.
- As we have used a transcriptomic dataset of the whole genome from the blood sample, some novel findings can potentially be revealed as the comorbidity of COVID-19.
- Our pipeline is capable of generating automated results provided input dataset.
- Finally, if the transcriptomic dataset exists, it is possible to apply the technique swiftly to different diseases.

The rest of the paper is structured as follows: Section II describes related previous works and their shortcomings; Section III depicts the proposed pipeline outlining our methodology and analysis. Section IV discusses the outcomes and validation. Section V concludes our paper and introduces future extensions of the work.

2. Literature Review

According to the World Health Organization, there have been 170,426,245 confirmed cases, including 3,548,628 deaths due to the COVID-19 pandemic globally as of 1 June 2021. The number of infected people and deaths is increasing rapidly and this mortality, infection, and severity are getting more dangerous because of existing comorbidities and underlying health conditions [16] — [18]. Hence it is important to identify the comorbidities or risk factors to reduce severe COVID-19 progression. Some COVID-19 comorbidities associated research works are presented as follows:

A meta-analysis of the published global literature was conducted by the authors of [19] by combining clinical data with comorbidity details, they were able to identify significant COVID-19 comorbidities. They found chronic kidney disease, type-2 diabetes, malignancy, hypertension, cardiovascular disease, etc. as the most significant comorbidities associated with COVID-19; working with published articles and aggregated clinical cohort data there is a possibility of biases.

By reviewing the literature the authors of [20] selected angiotensin-converting enzyme 2 (ACE2), transmembrane protease serine 2 (TMPRSS2), and basigin (BSG) as viral attack receptors and obtained PPI data from bioinformatics and system biology databases. They identified 59 proteins that interact with SARS-CoV-2 proteins and using the identified proteins they collected diseases named as comorbidities from DisGeNet [21], OMIM [22], ClinVar [23], and PheGenI [24] databases. They also selected 79 diseases from a list of diseases that are associated with 30 proteins selected from 59 proteins and the selections are not clear.

In [25], the authors explored the gene expression patterns from 30 array-based acute, chronic, and infectious gene expression datasets by meta-analysis. They observed the genes of the diseases are correlated and using literature mining they identified 78 genes which include receptors, and proteases of SARS-CoV-2 pathogenesis. Their findings also indicated that COVID-19 is susceptible to leukemia, nonalcoholic fatty liver disease, psoriasis, diabetes, and pulmonary arterial hypertension as comorbidities, through co-expression analysis of genes, pathways enrichment analysis, and significant enrichment of pathways.

Using data from preexisting diagnoses and hospitalized COVID-19 patients, the authors of [26] identified COVID-19 risk factors using the UK Biobank.

Using logistic regression model, they estimated risks for each comorbidity from the existing comorbidities and found that males, people of black ethnicity, and no educational qualification people have a higher risk of COVID-19.

A study by the authors of [27] attempted to relate COVID-19 to breast cancer, colon cancer, kidney cancer, liver cancer, bladder cancer, prostate cancer, thyroid cancer, and lung cancer. Their findings identified dysregulated genes, pathways, gene ontologies, and associations and interactions between cancers and COVID-19. They marked the cancers as comorbidities based on their association and interaction with COVID-19. Anteriorly they selected the diseases, then COVID-19 is associated with the selected diseases.

In [28], the authors selected five illnesses: kidney, liver, diabetes, lung, and cardiovascular illness from the literature as comorbidities and examined their potential association with COVID-19. First, they identified the common genes of the comorbidities, then they conducted phenotype enrichment analysis and pathway enrichment analysis using the shared genes; they also identified common pathways between COVID-19 and the comorbidities and investigated the association of the pathways with COVID-19 itself.

By the authors of [29], laboratory-confirmed COVID-19 studies were meta-analyzed. In COVID-19 patients, the spread of comorbidities was evaluated, and the risk factor was identified. They found that older and male patients are more susceptible to COVID-19; also, they marked hypertension, diabetes, cardiovascular and followed by respiratory system diseases as prevalent risk factors from the literature.

Based on pooled proportion, the authors of [30] performed a systematic review and meta-analysis of comorbidities prevalence, severity, and mortality concerning age, gender, and geographic areas. They observed a high prevalence of comorbidities in the USA, the highest severity in Asia with the highest mortality in Europe and Latin American region; chronic kidney disease is highly responsible for disease severity and cerebrovascular accident, chronic kidney disease, and cardiovascular disease for mortality of COVID-19 patients. By reviewing the articles, they also got higher severity in female patients and mortality in older and male patients.

We analyzed the COVID-19 dataset in a systematic way to figure out the comorbidity.

P-value combination is very common in bioinformatics [31] and there are diverse methods for combining P-values from multiple statistical tests such as Brown’s Method [31], Fisher’s Method [32], Truncated Product Method [33], Stouffer’s Method [34], Tippett’s Method [35], Bonferroni Method [36], Weighted Z-test [37], Z-transform Test [38], Liptak Test [39], Sidak Test [40], Simes’ Test [40], Gamma Distribution [41] etc. For different situations and datasets, different methods are powerful and suitable [41]; in our case, Tippett’s Method is decent, and using Tippett’s Method we got COVID-19’s P-value. In later sections, a discussion about COVID-19’s P-value is there and after getting the P-value we did our further analysis.

3. Methods and Analysis

3.1. Covid-19 Dataset and Geo2r Analysis

For our investigation, we collected the dataset from National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/) with accession number GSE166552 which is a microarray dataset prepared from blood samples of COVID-19 patients and healthy people. Initially, we analyzed the gene expression dataset with a web-based GEO2R tool, setting force normalization [42], precision weight (vooma) [43], and Benjamini & Hochberg (False discovery rate) [44] and comparing the transcriptomic profile of infected and control samples incorporating GEOquery [45], limma [46], and umap [47] R packages. The dataset comprises columns such as gene ID, adjusted P-value, P-value, Log2 fold change (logFC), gene sequence, and gene symbol. It was observed that certain cells within the gene symbol column were identified as missing values. Consequently, data related to these missing gene symbols were filtered out during the preprocessing phase. After preprocessing the dataset, we worked on the following analysis.

3.2. Dysregulated Genes and Disease Collection

To identify the comorbidities, statistically significant dysregulated (up-regulated and down-regulated) genes (DGs) are selected as follows:

\[
DGs = \begin{cases} 
\text{Up-regulated, } P \text{-value } \leq 0.05 \text{ & log FC } \geq 1.0 \\
\text{Down-regulated, } P \text{-value } \leq 0.05 \text{ & log FC } \leq 1.0 
\end{cases}
\] (1)

Afterwards, more than 20,000 diseases associated with the selected DGs were obtained from several Gold Benchmark databases, including DisGeNET [21], ClinVar 2019 [23], PheWeb 2019 [48], dbGap [49], OMIM Disease [50], OMIM Expanded [50], Rare Diseases AutoRF ARCHS4 Predictions [51], Rare Diseases GeneRIF ARCHS4 Predictions [51], Rare Diseases GeneRIF Gene Lists [51], Rare Diseases AutoRF Gene Lists [51] and GWAS Catalog 2019 [52].

3.3. Covid-19’s P-value Calculation

The dataset used in this study was derived from COVID-19 patients and comprises a large number of genes with individual P-values. Our objective is to obtain a single P-value from the dataset representing the P-value of COVID-19. Thus, we combined all the genes’ P-values to get a single P-value that is considered COVID-19’s P-value using Tippett’s Method as follows:

\[
p_c = 1 - (1 - \min(p_1, p_2, p_3, \ldots, p_i))^k
\] (2)
Where $\min$ is the minimum function of $P_1, P_2, P_3, ..., P_n$ as well as $P_1, P_2, P_3, ..., P_n$ denote separate $P$-values and $k$ is the total number of $P$-values of the genes of the dataset.

3.4. Euclidean Distance and Disease Order

In our investigation, we calculated the Euclidean distance between the $P$-value of COVID-19 and all the other individual diseases that were collected using the dysregulated genes. The distance is the absolute value difference of the $P$-values as follows:

$$E_d = |X_c - Y|$$  \hspace{1cm} (3)

where $X_c$ denotes the COVID-19’s $P$-value and $Y = Y_1, Y_2, Y_3, ..., Y_n$ denotes the certain $P$-value of the collected disease. After getting the Euclidean distance of all other collected diseases from COVID-19, we sorted the collected diseases in ascending order depending on the Euclidean distance.

3.5. Clustering and Comorbidities Selection

In the collected diseases we found that some diseases have the same $P$-value and as a result, the same Euclidean distance is found for those diseases. We clustered the disease based on the same Euclidean distance of the sorted diseases list from top to bottom maintaining the ascending order. Finally, we select the comorbidities of COVID-19 from the top six clusters.

3.6. Our Proposed Model as An Algorithm

**Algorithm:** The algorithmic pseudocode to generate the workflows for identification of the comorbidities of COVID-19 as follows:

**Input:** Microarray dataset with six samples where three samples from COVID-19 patients and another three samples from healthy people.

**Output:** Related comorbidities of COVID-19.

- Assign patient’s samples to the Case and healthy people’s samples to the Control group in the GEO2R tool.
- Analyze the dataset using GEO2R by setting force normalization, precision weights, and the Benjamini & Hochberg algorithm.
- Preprocess the dataset after analyzing it by GEO2R.
- Identify up-regulated genes applying $P$-value $\leq 0.05 \& \log FC \geq 1.0$.
- Identify down-regulated genes applying $P$-value $\leq 0.05 \& \log FC \leq 1.0$.
- Enter the identified dysregulated (up-regulated and down-regulated) genes into the Gold Benchmark databases and gather a list of diseases.
- Calculate COVID-19’s $P$-value using Tippett’s method.
- Calculate Euclidean distance from COVID-19 to all other diseases on the list.

- Sort all diseases of the list based on the Euclidean distance in ascending order.
- Cluster all the diseases on the list based on the same Euclidean distance.
- Select correlated comorbidities from the top clusters.

The workflow steps of our proposed model are shown in Figure 1.

4. Results and Discussion

After analyzing, preprocessing, and filtering we got 35,011 differentially expressed genes with $P$-value, adjusted $P$-value, and fold change value. Then we evicted 1,421 up-regulated genes using a statistical threshold of $P$-value $\leq 0.05$ and absolute $\log_2$ fold change ($\log FC$) $\geq 1.0$. Also, we got 2,138 down-regulated genes using $P$-value $\leq 0.05$ and $\log_2$ fold change ($\log FC$) $\leq 1.0$; a total of 3,559 dysregulated genes were identified. To attain the result, we uploaded a total of 3,559 dysregulated genes in the stated method and analysis section’s eleven Gold Benchmark databases of Enrichr [53] which is a web-based gene set enrichment analyzing tool.

In our investigation, we calculated the Euclidean distance between cluster 1 and COVID-19 is $0.301041817$ by combining the $P$-values of all 35,011 genes from our filtered dataset using Tippett’s method. Next Euclidean distances among the $P$-value of COVID-19 and 22,820 diseases are enumerated individually and then the total of 22,820 diseases are rearranged in ascending order using the computed Euclidean distances.

Table 1 shows the primary disease and its $P$-value, top collected diseases, $P$-values of the top collected diseases, and Euclidean distance from the primary disease to the top collected diseases from left to right. After sorting the collected disease, we got prominent ear, acute myelocytic leukemia, adducted thumb, cancer of urinary tract, increased appetite, malignant neoplasm of vulva, myelodysplastic syndrome with isolated del(5q), nonnuclear polymorphic congenital cataract, chromosome 12 12p trisomy, chromosome 15q trisomy, chromosome 9 monosomy 9p, body weight changes, abdominal aortic aneurysm, kidney neoplasm, diabetes mellitus, and other diseases from top of the total diseases.

Many diseases have the same Euclidean distance from COVID-19 and diseases with similar Euclidean distances are retained in the same cluster.

Table 2 represents cluster number, Euclidean distance between COVID-19 and the clusters or clusters’ disease, and the disease names in the clusters from left to right. Euclidean distance between cluster 1 and COVID-19 is $0.000149032$; in the same way, Euclidean distances are
We clustered the entire 22,820 diseases based on the same Euclidean distance and took the top six clusters. From Figure 2, we see that in cluster no. 1 there is only one disease prominent ear; in cluster 2, only acute myelocytic leukemia is present; cluster 3 contains ten diseases and they are adducted thumb, cancer of urinary tract, increased appetite, malignant neoplasm of vulva, myelodysplastic syndrome with isolated del(5q), nonnuclear polymorphic congenital cataract, chromosome 12 12p trisomy, chromosome 15q trisomy, chromosome 9 monosomy 9p and body weight changes. Again, cluster 4, cluster 5, and cluster 6 hold abdominal aortic aneurysm, kidney neoplasm, and diabetes mellitus.

Finally, we selected comorbidity from our top six clusters. Figure 3 shows that acute myelocytic leukemia is selected as comorbidity from cluster 2, from cluster 3 cancer of urinary tract and body weight changes are selected as comorbidity; abdominal aortic aneurysm, kidney neoplasm, and diabetes mellitus are selected from cluster 4, cluster 5, and cluster 6; only cluster 1 is not considered to select comorbidity as we have not found a relationship between COVID-19 and prominent ear in global research.

We developed a computational and bioinformatics framework to identify the comorbidities of COVID-19 and our model is suitable to obtain the comorbidities of other diseases quickly. As far as we know there is no other such approach to ascertain comorbidities of diseases numerically; there are clinical tests, meta-analysis, and cross-checking tests to determine comorbidities but they are time-consuming and take much exertion. Acute myelocytic leukemia [54] — [57], cancer of urinary tract [58], [59], body weight changes [60], [61], abdominal aortic aneurysm [62], [63], kidney neoplasm [64] — [66], and diabetes mellitus [67] — [70] are the outcomes of our research where cancer of urinary tract and kidney neoplasm are related to bladder cancer [71], [72]. Moreover, bladder cancer is a post COVID-19 condition [59]. So, there is a chance of ensuing cancer of urinary tract and kidney neoplasm in COVID-19 recovered patients.
Body weight changes (weight gain or weight loss) is also a post COVID-19 condition [60], [73] and excess weight gain or overweight causes obesity [74]. What we found in our outcomes is also the same in the published literature; thus, we validated our work.

Clinical decision-making and outcomes are significantly affected by comorbidities, which influence treatment strategies and patient management. Research indicates that comorbid conditions complicate the treatment landscape by influencing drug interactions, therapeutic efficacy, and patient compliance [75]. Additionally, comorbidity can complicate the prognosis of primary illnesses, increasing healthcare expenditures [76]. Early detection of comorbidities enables healthcare providers to implement preventive measures and targeted interventions to arrest disease progression and improve patient health [77]. It is also essential to assess and optimize patients with comorbidities before surgery to reduce complications during surgery [78]. Therefore, it is imperative to identify accurate comorbidities early in clinical practice to optimize treatment strategies, improve patient management, and improve health outcomes. Comorbidities that are the result of our work are shown in Figure 4.

Table 1: Summary of results along with the P-value of COVID-19, top collected disease, their P-values and Euclidean distance between COVID-19 and the top collected disease

<table>
<thead>
<tr>
<th>Primary Disease</th>
<th>P-value of COVID-19</th>
<th>Top Collected Disease</th>
<th>Collected Disease’s P-value</th>
<th>Euclidean Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>COVID-19</td>
<td>0.301041817</td>
<td>Prominent Ear</td>
<td>0.300892784</td>
<td>0.000149032</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Acute Myelocytic Leukemia</td>
<td>0.301192317</td>
<td>0.0001505</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Adducted Thumb</td>
<td>0.301209965</td>
<td>0.000168149</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cancer of Urinary Tract</td>
<td>0.301209965</td>
<td>0.000168149</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Increased Appetite</td>
<td>0.301209965</td>
<td>0.000168149</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Malignant Neoplasm of Vulva</td>
<td>0.301209965</td>
<td>0.000168149</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Myelodysplastic Syndrome with Isolated del(5q)</td>
<td>0.301209965</td>
<td>0.000168149</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nonnuclear Polymorphic Congenital Cataract</td>
<td>0.301209965</td>
<td>0.000168149</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chromosome 12 12p Trisomy</td>
<td>0.301209965</td>
<td>0.000168149</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chromosome 15q Trisomy</td>
<td>0.301209965</td>
<td>0.000168149</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chromosome 9 Monosomy 9p</td>
<td>0.301209965</td>
<td>0.000168149</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Body Weight Changes</td>
<td>0.301209965</td>
<td>0.000168149</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Abdominal Aortic Aneurysm</td>
<td>0.30024875</td>
<td>0.000756942</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kidney Neoplasm</td>
<td>0.3002432</td>
<td>0.000798617</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diabetes Mellitus</td>
<td>0.301850763</td>
<td>0.000808946</td>
</tr>
</tbody>
</table>

Table 2: Disease in each cluster and Euclidean distance from COVID-19 to clusters or clusters’ disease

<table>
<thead>
<tr>
<th>No.</th>
<th>Euclidean distance</th>
<th>Diseases in the cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>0.000149032</td>
<td>Prominent Ear</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>0.0001505</td>
<td>Acute Myelocytic Leukemia</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>0.000168149</td>
<td>Adducted Thumb</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cancer of Urinary Tract</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Increased Appetite</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Malignant Neoplasm of Vulva</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Myelodysplastic Syndrome with Isolated del(5q)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nonnuclear Polymorphic Congenital Cataract</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chromosome 12 12p Trisomy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chromosome 15q Trisomy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chromosome 9 Monosomy 9p</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Body Weight Changes</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>0.000756942</td>
<td>Abdominal Aortic Aneurysm</td>
</tr>
<tr>
<td>Cluster 5</td>
<td>0.000798617</td>
<td>Kidney Neoplasm</td>
</tr>
<tr>
<td>Cluster 6</td>
<td>0.000808946</td>
<td>Diabetes Mellitus</td>
</tr>
</tbody>
</table>
5. Conclusion

COVID-19’s comorbidities increase the disease severity and mortality as well as influence organ damage of the patients; often the patients die due to their main comorbidities. Because of that, it is exigent for medical practitioners to find out the comorbidities of diseases, and in our study, we developed a pipeline that can detect comorbidities of COVID-19 from genetic datasets in a computational and bioinformatics way. We identified DGs and diseases related to the DGs using the bioinformatics approach as well as detected risk factor diseases linked with COVID-19 as comorbidities from the gathered diseases in a systematic computational way.

In this research, one of the main challenges for us is to find the accurate P-value of COVID-19. We observed that the more accurate COVID-19’s P-value gives the more accurate comorbidities as a result. In the future, we will apply other alternative methods to derive the exact P-value of COVID-19. We suggest that our developed approach will assist in the diagnosis of comorbidities of other diseases early if the genetic dataset is available and thus this model will help healthcare systems to reduce the comorbidity diagnosis cost of a disease.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The assistance for this project is provided by the ICT Division of the Government of the People’s Republic of Bangladesh, under Grant number 22FS15614.

References


Available at: https://covid19.who.int/ (Accessed: June 22, 2021).


M.E. Dolan, D.P. Hill, G. Mukherjee, M.S. McAndrews, E.J.


Copyright: This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-SA) license (https://creativecommons.org/licenses/by-sa/4.0/).