**A Partitioning-based Approach for Clustering COVID-19 Drugs and Co-Medication for Safe Use**

**Ahmad Alqurneh1, Aida Mustapha1\*, Nurfadhlina Mohd Sharef2**

1 Faculty of Computer Science and Information Technology,

Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, MALAYSIA

2 Faculty of Computer Science and Information Technology,

Universiti Putra Malaysia, 43000 Serdang, Selangor, MALAYSIA

\*Corresponding Author

DOI: https://doi.org/10.30880/ijie.2020.12.05.028

Received 17 May 2020; Accepted 30 June 2020; Available online 30 June 2020

**Abstract:** Drugs Interaction (https://covid19-druginteractions.org/) is a website that allows users to select a COVID- 19 drug among the nine popular drugs with one or more drug from a different class co-medication to check for any interaction between the chosen drugs. Beyond the interaction, however, the website does not advise if the resulting class of co-medication is safe to use or otherwise along with one of the nine COVID drugs. Therefore, there exist a need to use an unsupervised clustering approach to group the COVID drugs and respective co-medications that are safe to use in the absence of interaction and vice versa. This paper focused on application of the Partitioning Around Medoid (PAM) clustering algorithm to categorize the combinations as safe or otherwise. The resulting clusters are then measured using the Silhoutte value and presented.

**Keywords:** Clustering, unsupervised learning, COVID-19, drugs

# Introduction

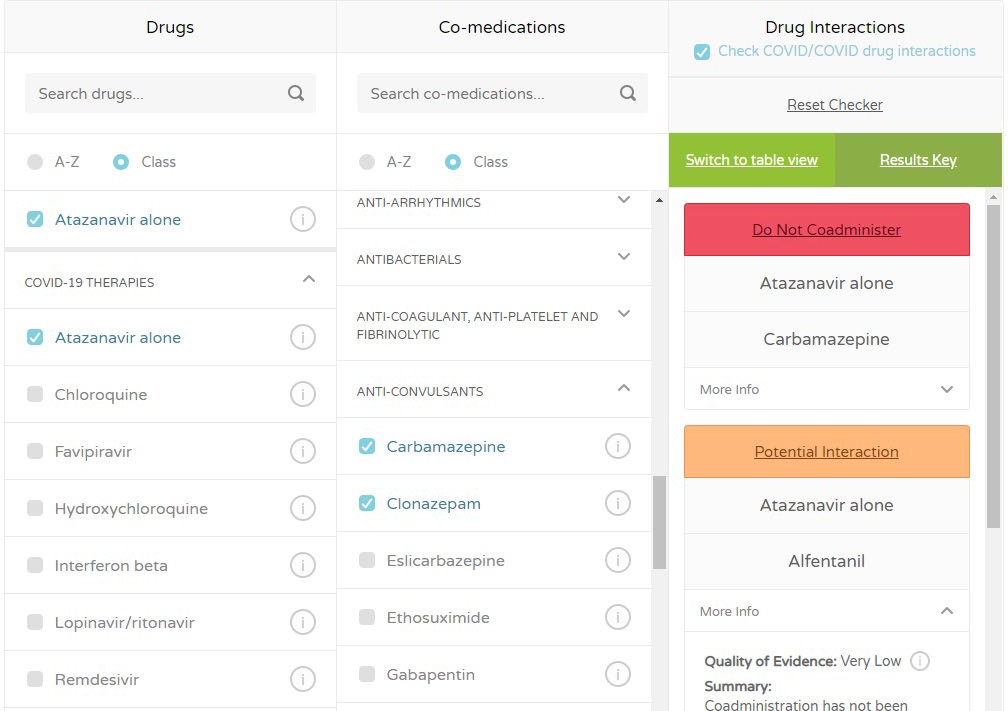
Since the crisis of coronavirus started in the late 2019, the pathogenic test in medical lab becomes essential for suspected patients even though it is considered time consuming. In solving epidemic prediction and detection, Artificial Intelligence (AI) algorithms such as the Convolutional Neural Networks (CNN) have been widely used across many areas for screening COVID-19. With a small dataset from CT images of patients with COVID-19, the CNN model has been compared against radiologists’ analysis and showed much higher accuracy and sensitivity [1]. Deep learning algorithm has also been applied on CT images to differentiate common pneumonia from COVID-19 disease [2]. Deep learning algorithms are more focused on chest CT images, and required quite large datasets where time is a critical issue.

AI algorithms have also been used to detect the size, lengths and ending time of COVID-19 across China cities using stacked auto-encoder [3]. In [4], COVID-19 detection was performed via speech recognition through a smartphone application, whereby the application recognizes the sound of individual cough patterns and check if it matches with the cough features produced in a diseased person. Within communication, AI techniques have been used in tools such as chatbots to represent human operators in emergency hotlines to receive patients’ reports on their health status if they are affected with COVID-19 [5, 6, 7]. According to [5], computer-based approaches are two-fold: First is yes/no online questionnaires and these are limited in their informative value. Second is the general-purposed symptom checkers such as Symptoma [5] that takes in list of symptoms from patients. Work in [6] presented mobile phone-based surveys in cities under quarantine [6] to supply AI models with sufficient data in order to determine individual’s health standing [7].

However, in areas where traditional data collection is not applicable, active learning is proposed to train AI machines using on-the-fly data [8]. Data mining techniques such as anomaly detection (AD) or outlier detection have been used to deal with changes in data over time and to detect uncommon features from cross-population data such as the CT scans

and Chest X-rays (CRRs). Cross-population train/test data gathered from one country are then can assist in the automation detection of the virus in other country in the future [8]. In monitoring the pandemic, AI-based methods are used for real time forecasting of the confirmed cases [9]. It has also taken a huge role in charities management as to determine the optimal time and place to spend limited supply of protective equipment [9].

There are several online applications that allow the checking of drug interaction and their implications on patients to or potential side effects such as WebMD (<https://www.webmd.com/interaction-checker/default.htm>) and Drugs.com (<https://www.drugs.com/drug_interactions.html>). Drug-drug interactions (DDIs) are ubiquitous, harmful and a leading cause of morbidity and mortality. Primary care physicians are increasingly challenged with identifying and preventing DDIs. An aging population, growth in polypharmacy, widespread use of supplements, and the rising opioid abuse epidemic increase the demand of online DDIs service. The Drugs Interaction (<https://covid19-druginteractions.org/>) is a an example of website that allows users to select a COVID-19 drug among the nine popular drugs with one or more drug from a different class co-medication to check for any interaction between the chosen drugs. However, the interaction checker does not indicate if the resulting class of co-medication is safe or which one of the nine COVID drug has better interaction with the co-medications. Fig. 1 shows a running example from the COVID-19 Drug Interactions website.



### Fig. 1 - Running example from COVID-19 Drug Interactions website [10]

High quality, safe care requires that DDIs be recognized and avoided. DDIs also support drug repositioning or repurposing. This paper addressed this need by proposing a clustering approach to group the COVID drugs and respective co-medications that are safe to use in the absence of interaction and vice versa. The results from the clustering approach can help the physician, pharmacist and public to take precautions on DDIs usage. Clustering is an AI task that focuses on dividing a population of multivariate data points into several groups. In theory, data points that are in the same group should have similar properties and/or features, while data points in different groups should have highly dissimilar properties and/or features. We can use clustering analysis to gain some valuable insights from our data by seeing what groups the data points fall into when we apply a clustering algorithm. There are five types of clustering methods which are partitioning, hierarchical, fuzzy, density and model-based clustering. This paper presents the investigation of a partitioning algorithm called Partition Around Medoid (PAM) that subdivides the data sets into a set of k groups.

The remaining of this paper proceeds as follows. Section 2 presents the methodology along with the dataset and clustering algorithm, Section 3 presents the results, and finally Section 4 concludes the paper.

# Material and Method

This paper is set to study and analyze the interaction relationship between popular COVID-19 drugs and the set of co- medications drugs from 25 different classes of therapies provided in the Drugs Interaction website. A partitioning-based approach to clustering drugs and co-medication related to the COVID-19 disease is proposed to group COVID-19 drugs and its co-medication based on their level of safe use or otherwise. The experiments will be based on a standard clustering

methodology as shown in Fig. 2. The clustering experiments will be performed for both COVID-19 drugs and the co- medications, separately.

COVID-19

Dataset Pre-processing

PAM Clustering Approach

Silhouette Method Evaluation to find best *k* value

Comparison with Real Cases from the Website

# Dataset

### Fig. 2 - Clustering methodology

The fast and wide spread of COVID-19 datasets is incredible as there are many datasets are recently available either freely or with restriction. Most of these datasets are difficult to access or to deal with to carry out information extraction or retrieval as most of these datasets deal with drugs. In other datasets, the size is small with incomplete patients’ data [11]. It is important to mention that most of the literatures related to COVID-19 diseases are in preprints due to the sensitivity nature of the corona crisis. The availability of such datasets is important to encourage more studies related to the COVID-19 pandemic.

The dataset used in this study is sourced from the set of nine (9) popular COVID-19 drugs (Table 1) and 25 classes of 474 comedications (Table 2) provided in the online interaction checker in [10]. This means, 25 x 9 pairs can be checked exhaustively. For example, if a user were to know if the Remdesivir antiviral can be co-administered with drugs that belong to a certain class, the user needs to browse through each drug in that particular class. Therefore PAM is expected to be able to cluster the drugs and co-medications automatically.

### Table 1 - List of COVID-19 drugs

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **COVID-19 class** | **No.** | **COVID-19 class** |
| 1 | ATV | 6 | HCLQ |
| 2 | LPV/r | 7 | RBV |
| 3 | RDV | 8 | TCZ |
| 4 | FAVI | 9 | IFN- |
| 5 | CLQ |  |  |

**Table 2 - List of COVID-19 co-medication drugs**

|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **Co-medication class** | **No.** | **Co-medication class** |
| 1 | Anesthetics and Muscle Relaxants | 14 | Anxiolytics-Hypnotics-Sedatives |
| 2 | Analgesics | 15 | Beta Blockers |
| 3 | Antiarrhythmics | 16 | Bronchodilators |
| 4 | Antibacterials | 17 | Calcium Channel Blockers |
| 5 | Anti-coagulant, anti-platelet | 18 | Contraceptives-HRT |
| 6 | Anticonvulsants | 19 | Gastrointestinal Agents |
| 7 | Antidepressants | 20 | anti-emetics |
| 8 | Anti-diabetics | 21 | Hypertension |
| 9 | Antifungals | 22 | mmunosuppressants |
| 10 | Antipsychotics – Neuroleptics | 23 | Inotropes & Vasopressors |
| 11 | Antivirals – COVID-19 | 24 | Lipid Lowering Agents |
| 12 | Antivirals-HCV DAAs | 25 | Steroids |
| 13 | Oseltamivir |  |  |

Given the list of COVID-19 drugs and their co-medication, Drugs Interaction [10] provides interaction information, which indicates whether a co-medication should be co-administered or otherwise with any of the COVID-19 drugs. The interaction result is categorized as one of the four outputs shown in Table 3.

### Table 3 - Category of interaction between the COVID-19 drugs and co-medication



|  |  |  |
| --- | --- | --- |
| **Interaction category symbol** | **Interaction category (pre- processed value)** | **Interaction Output** |
|  | 1 | These drugs should not be co-administered. |
|  | 2 | Potential interaction. May require close monitoring, alteration of drug dosage or timing of administration. |
|  | 3 | Potential interaction likely to be of weak intensity. |
|  | 4 | No clinically significant interaction expected. |

The dataset is prepared in the form of a 25 rows X 9 columns where the rows represent the co-medication drugs and the columns represent the COVID-19 drugs. The cells are the category of interaction. The goal of the clustering is to identify the set of co-medication classes which have similar interaction type. Final dataset is shown in Table 4.

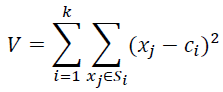
### Table 4 - Excerpt of dataset from [anesthetics & muscle class]

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Features** | **ATV** | **LPV/r** | **RDV** | **FAVI** | **CLQ** | **HCLQ** | **RBV** | **TCZ** | **IFN-** |
| Alcuronium | 4 | 4 | 4 | 3 | 4 | 4 | 4 | 4 | 4 |
| Bupivacaine | 2 | 2 | 4 | 4 | 4 | 4 | 4 | 3 | 4 |
| Cisatracurium | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Desflurane | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Dexmedetomidine | 4 | 2 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Enflurane | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Ephedrine | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Etidocaine | 2 | 2 | 4 | 4 | 4 | 4 | 4 | 3 | 4 |

Based on Table 4, the dataset shows more values for no significant interactions (interaction category = 4). This indicates that if a user were to conduct pairwise interaction checking, most results will indicate that there is no interaction between the drugs. This also states a need to computationally cluster these drugs to identify safe combination of co- medication drugs. A clustering approach using the Partitioning Around Medoid (PAM) algorithm is therefore proposed to cluster the unlabeled dataset.

# PAM Algorithm

This paper focused on application of the PAM (also called as *K*-Medoids) algorithm [12]. A medoid can be defined as a point in a cluster, whose dissimilarities with all the other points in the cluster is minimum. Basically the algorithm works by choosing *k* datapoints as centers, hence the name *k*-medoids. PAM is set to find a local minimum to the objective function as shown in Eq. 1.

 (1)

Based on Eq. 1, the algorithm for PAM can be described as follows:

1. Initialize: Select *k* random points out of the 𝑆𝑆𝑖𝑖 data points as the medoids.
2. Associate each data point to the closest medoid by using any common distance metric methods.
3. While the cost decreases:

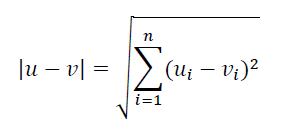
For each medoid 𝑐𝑐𝑖𝑖 in 𝑆𝑆𝑖 , for each data *o* point which is not a medoid:

* 1. Swap *m* and *o*, associate each data point to the closest medoid, recomputes the cost.
  2. If the total cost is more than that in the previous step, undo the swap.

# Evaluation Metrics

To evaluate the performance of the clustering algorithms, the optimal value of the resulting clusters will be measured using the Silhouette analysis. The Silhouette analysis validates the consistency within [cluster of data](https://en.wikipedia.org/wiki/Cluster_analysis) by measuring the separation distance between the resulting clusters. The silhouette value is a similarity value between an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette ranges from −1 to +1, where a high value indicates that the object is well matched to its own cluster (the sample is far away from neighboring clusters), a value indicates that the sample is on or very close to the decision boundary between two neighboring clusters, and negative values indicates that those samples might have been assigned to the wrong cluster. In general, if many points have a low or negative value, then the clustering configuration may have too many or too few clusters.

In this paper, the silhouette value is calculated using the Euclidean distance, which is the distance between two elements u and v equals to the length of the segment connecting the elements. The [distance](https://en.wikipedia.org/wiki/Distance) metric as shown in Eq. 2.



(2)

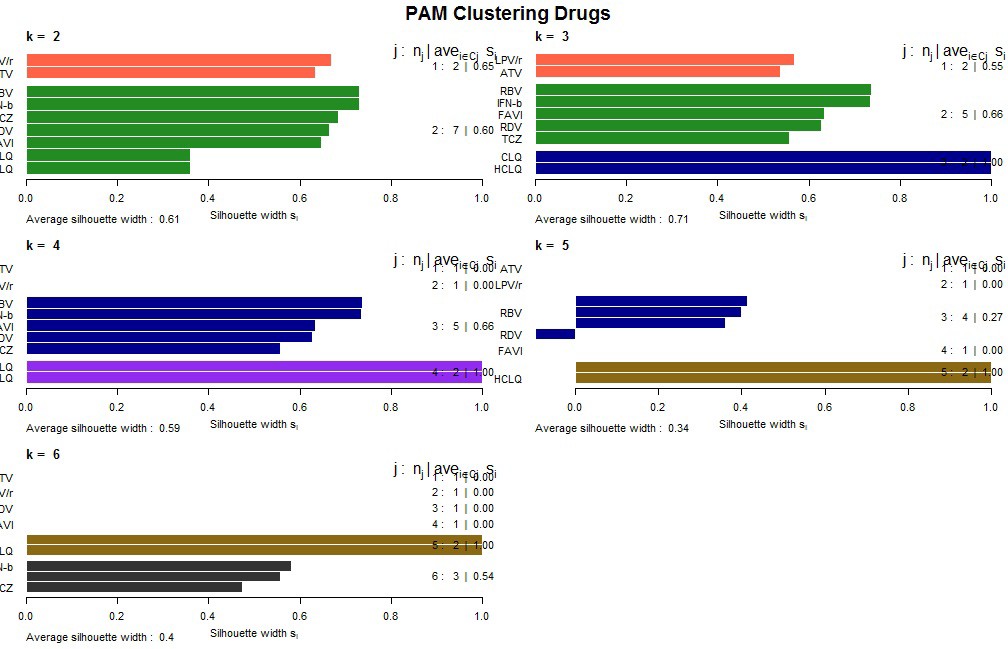
# Results and Discussion

The Partitioning Around Medoid (PAM) algorithm was applied to both COVID-19 drugs and the co-medications.

Next, the clustering experiment will produce Silhouette plot to further analysis.

# Clustering COVID-19 Drugs

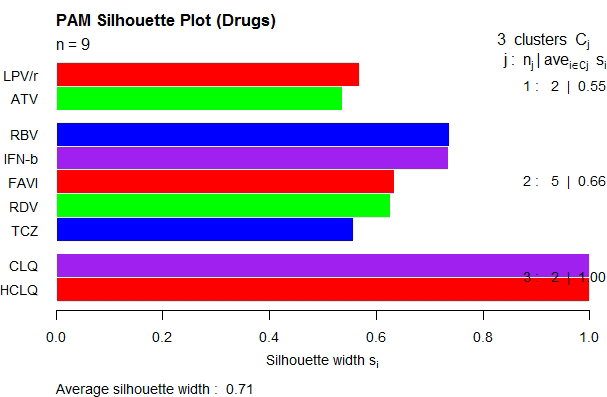
Applying PAM on the nine drugs set, we get the following silhouette width values at different *k* values. The “goodness” of the given value of *k* will be assessed with [silhouette method](https://en.wikipedia.org/wiki/Silhouette_(clustering)) as shown Fig. 3.



### Fig. 3 - PAM analysis for *k* = 2 to 6 for COVID-19 drugs

To find the best cluster (k) value, the silhouette method indicated that the best element grouping in the cluster is at

0.71 with k = 3. This is shown in Fig. 4.

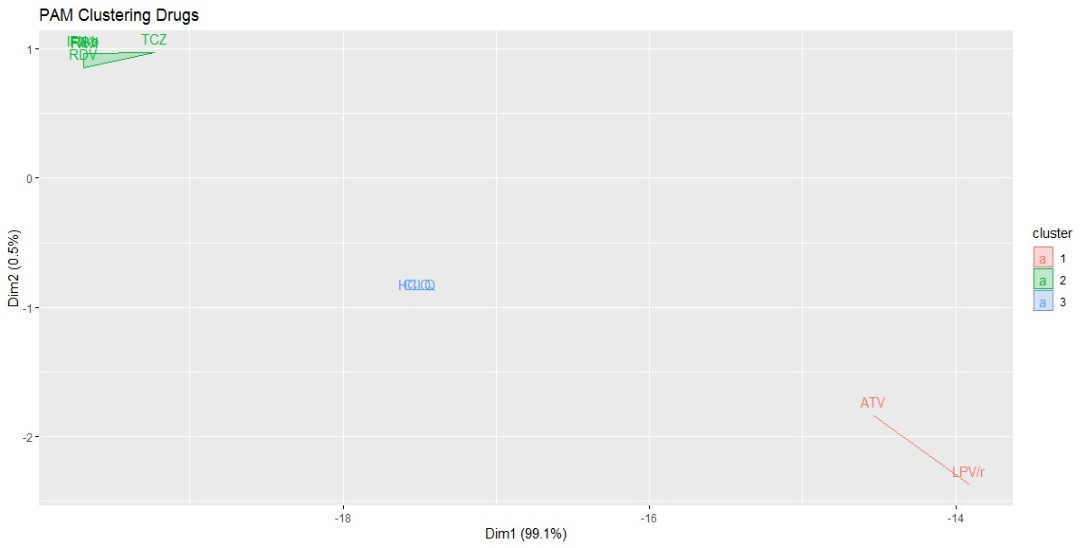


### Fig. 4 - Silhouette plot for COVID-19 drugs

Next, based on the silhouette plot, the cluster is validated by showing how well an observation is clustered by estimating the average distance between clusters. Therefore, PAM produced three clusters of drugs as shown in Table 5 and Fig. 5.

### Table 5 - Resulting clusters by PAM algorithm

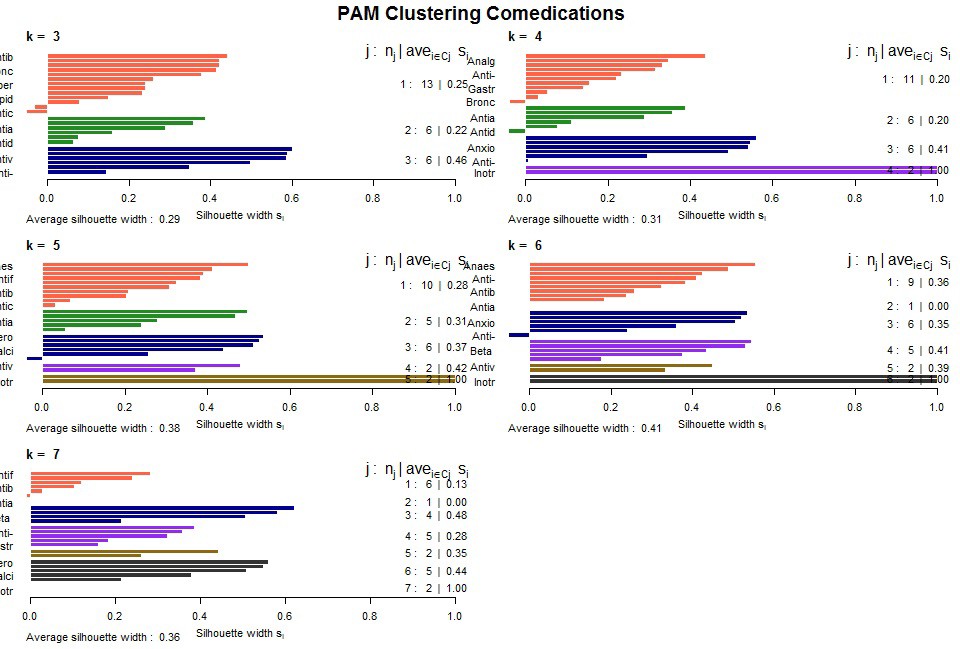
|  |  |  |  |
| --- | --- | --- | --- |
| **No.** | **PAM Cluster** | **Interaction Category** | **Safety Level** |
| C1 | {ATV, LPV/r}  (in order) | 2.82.7 | Low  (More interactions) |
| C2 | {**RDV\***, (IFN-, RBV, FAVI), TCZ}  (in order) | 43.8 | High |
| C3 | {CLQ, HCLQ}  (in order) | 3.4 | Medium  Low (Mild interactions) |



**Fig. 5 - Resulting clusters for COVID-19 drugs**

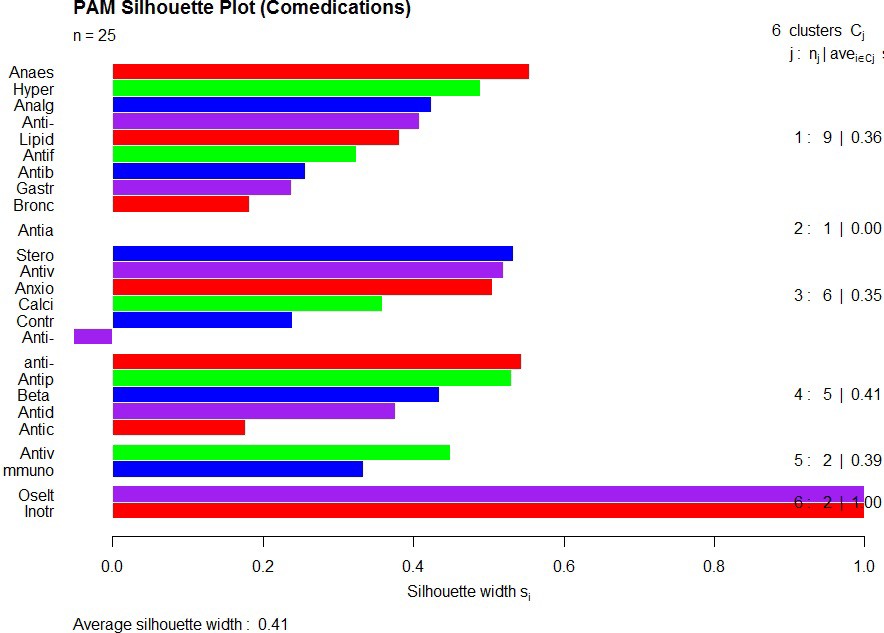
**3.2 Clustering Co-Medications**

Next, PAM was applied on the 25 co-medication classes (contains 474 drugs) and measured the silhouette width values at different *k* values. The “goodness” of the given value of *k* for co-medication drugs is shown in Fig. 6.



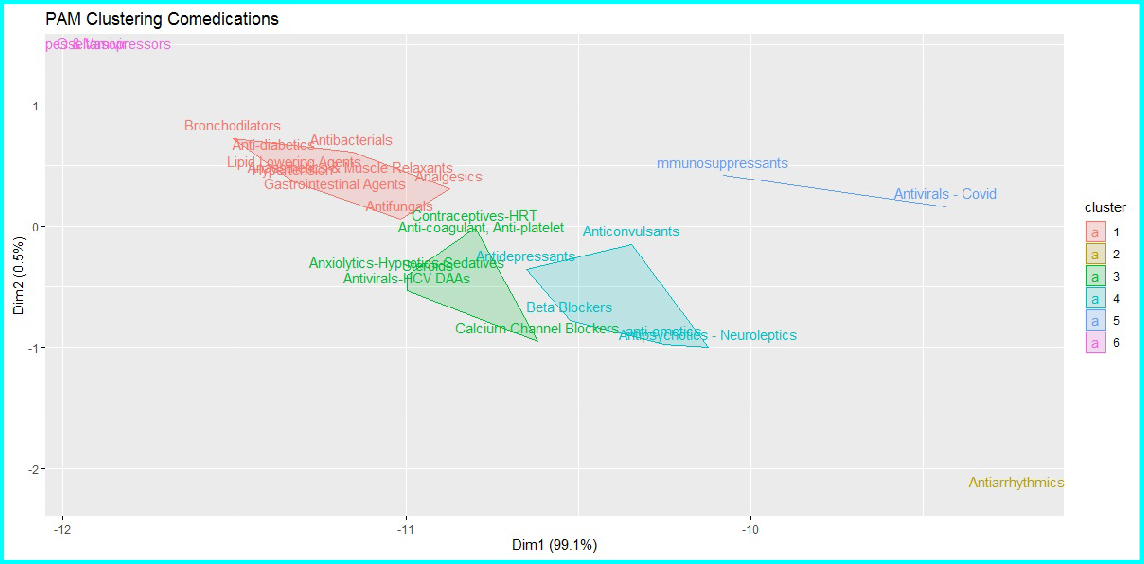
### Fig. 6 - PAM analysis for *k* = 3 to 7 for co-medication drugs

To find the best cluster (*k*) value, the silhouette method is used to get best silhouette width, which in turn, indicate the best element grouping in the cluster at 0.41 when *k* = 6 as shown in Fig. 7.



### Fig. 7 - Silhouette plot for co-medication drugs

Fig. 8 shows PAM clustering. Left side of the plot compromises the category-1 of co-medications, clusters C1, C3, and C6, where the right side compromise the other category-2 C2, C4, and C5. Clearly, C6 represent the extreme of first category and, C1 is slightly better than C3. In Contract, C2 represent the extreme of the other category, and C5 is slightly better than C4. Evaluation of the clusters with internet checker [10], PAM clusters and the level of safety of C19 co- medications are shown in Table 6. The order of the comedication in their set is based on their distance in their cluster. Refer to the silhouette cluster validation, it is noted here that [anti-coagulant, anti-platelet] is the only outlier in cluster 3 with negative silhouette value at 0.052. This indicates it should be in its neighboring cluster C1. However, C3 is still in the left side category.



### Fig. 8 - Resulting clusters for co-medication drugs

**Table 6 - Resulting clusters by PAM algorithm**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **PAM Cluster: Co-medication Class** | | | **Interaction Category** | **Safety Level** |
|  | C1 | [Anaesthetics & Muscle Relaxants, Hypertension, Analgesics, Anti-diabetics, Lipid  Lowering Agents, Antifungals, Antibacterials, Gastrointestinal Agents, Bronchodilators] | 3.83.7 | Higher |
| Left Side Cluster Category 1 |  |  |  |
| C3 | [Steroids, Antivirals-HCV DAAs, Beta Blockers, Calcium Channel Blockers, Contraceptives-HRT]  [Anti-coagulant, Anti-platelet\*] = outlier | 3.63.5 | High |
|  | C6 | [Oseltamivir, Inotropes & Vasopressors] | 4 | Extreme High |
|  | C2 | [Antiarrhythmics] | 2.9 | Extreme Low |
| Right Side Cluster Category 2 | C4 | [Anti-emetics, Antipsychotics – Neuroleptics, Beta Blockers, Antidepressants, Anticonvulsants] | 3.3 | Low |
| C5 | [Immunosuppressants, Antivirals – Covid] | 3.33.1 | Lower |

* 1. **Conclusions**

Drugs interaction is an important precaution to ensure that medication is conducted safely, especially in the case of critical health case such as COVID-19. However, manual browsing of the COVID-19 and co-medication drugs interaction is inefficient. This scenario is also consistent in situations where high variations of drugs are available. This paper presents the application of PAM algorithm and Silhouette analysis as effective safe drug interaction clustering. PAM is a partitioning algorithm that identifies the best clustering number. An experiment using PAM and Silhouette analysis on a dataset that consists of nine (9) COVID-19 drugs and 474 co-medication that belong to a different 25 classes identifies that there are four (4) clusters of COVID-19 drugs and six (6) cluster of co-medication classes that can have the same interaction characteristics. Since Remdesivir (RDV) is almost announced globally as Covid19 (C19) official treatment, this study examined the interaction safety level of nine C19 drugs including (RDV) used along with these co-medications. The study concluded that not only RDV is the only safe C19 drug, but there are 5 more drugs are highly safe can be used with these co-medications, in particular IFN-. On the other hand, while CLQ and HCLQ are marginally accepted drugs, ATV and LPV/r proved to be inappropriate with the majority of comedication classes. In terms of co-medications, only two classes showed complete safe level with C19 drugs named as Oseltamivir and (Inotropes & Vasopressors). While clustering analysis is most popular for segmentation tasks such [13], the future work from this study is to extend the drug interaction analysis using other data mining techniques such as classification [14], and sequential pattern analysis

[15].

# Acknowledgement

This paper is supported by Research Fund E15501, Research Management Centre, Universiti Tun Hussein Onn Malaysia.

# References

1. Wang, S.,Kang, B. & Ma, J., Zeng, X., Xiao, M., Jia, G., Mengjiao, C., Yang, J., Li, Y., Meng, X., & Xu, B. (2020). A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). doi: <https://doi.org/10.1101/2020.02.14.20023028>
2. Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., Bai, J., Lu, Y., Fang, Z., Song, Q., Cao, K., Liu, D., Wang, G., Xu, Q., Fang, X., Zhang, S., Xia, J., & Xia, J. (2020). Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT. Radiology. [doi.org/10.1148/radiol.2020200905](https://doi.org/10.1148/radiol.2020200905)
3. Hu, Z., Ge, Q., Jin, L., & Xiong, M. (2020). Artificial Intelligence Forecasting of Covid-19 in China. arXiv preprint arXiv:2002.07112, 2020b.
4. Iqbal, M. (2020). Active Surveillance for COVID-19 through artificial intelligence using concept of real-time speech-recognition mobile application to analyse cough sound. DOI: [10.31219/osf.io/cev6x.](https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.31219%2Fosf.io%2Fcev6x)
5. Martin, A., Nateqi, J., Gruarin, S., Munsch, N., Abdarahmane, I., & Knapp, B .(2020). An artificial intelligence- based first-line defence against COVID-19: digitally screening citizens for risks via a chatbot. [doi:10.1101/2020.03.25.008805v3.](https://www.biorxiv.org/c%20ntent/10.1101/2020.03.25.008805v3)
6. Srinivasa Rao, A., & Vazquez, J. (2020). Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone–based survey when cities and towns are under quarantine. Cambridge University Press, 1-5. doi:10.1017/ice.2020.61
7. Onoja, A. (2020). A Proposed Machine Learning approach for Monitoring Individual's Health Status on Corona virus (COVID19) cases. doi: 10.13140/RG.2.2.14001.74081.
8. Santosh, KC. (2020). AI-Driven Tools for Coronavirus Outbreak: Need of Active Learning and Cross-Population Train/Test Models on Multitudinal/Multimodal Data. Journal of Medical Systems. 44. doi: 10.1007/s10916-020- 01562-1.
9. Johnstone, S. (2020) A Viral Warning for Change. COVID-19 Versus the Red Cross: Better Solutions Via Blockchain and Artificial Intelligence. University of Hong Kong Faculty of Law Research Paper No. 2020/005.
10. Covid19-Drug Interactions. (2020). The University of Liverpool. <https://www.covid19-druginteractions.org/>
11. Bullock, J., Luccioni, A., Pham, K., Lam, C., & Luengo-Oroz, M. (2020). Mapping the Landscape of Artificial Intelligence Applications against COVID-19. arXiv preprint arXiv:2003.11336
12. Kaufman, L. & Rousseeuw, P.J. (1987). Clustering by Means of Medoids. Statistical Data Analysis Based on the Norms and Related Methods, edited by Y. Dodge, North-Holland, 405–416.
13. Marisa, F., Ahmad, SSS, Yusof, ZIM, Hunaini, F. & Aziz, TMA. (2019). Segmentation Model of Customer Lifetime Value in Small and Medium Enterprise (SMEs) using K-Means Clustering and LRFM Model. International Journal of Integrated Engineering, 11(3):169-180.
14. Ali, SA., Sulaiman, N., Mustapha, A., Mustapha, N. (2009). K-Means Clustering to Improve the Accuracy of Decision Tree Response Classification. Information Technology Journal, 8(8):1256-1262.
15. Jamali, S., & Navaei, YD. (2016). A Two-level Product Recommender for E-commerce Sites by Using Sequential Pattern Analysis. International Journal of Integrated Engineering, 8(1):15-24.