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

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

Md Arafatur Rahman\textsuperscript{a,b,}\textsuperscript{*}, Naifees Zaman\textsuperscript{b}, A. Taufiq Asyhari\textsuperscript{c,e}, Fadi Al-Turjman\textsuperscript{d}, Md. Zakirul Alam Bhuiyan\textsuperscript{e}, M.F. Zolkipli\textsuperscript{a}

\textsuperscript{a} Faculty of Computing, University Malaysia Pahang, Gambang 26300, Malaysia
\textsuperscript{b} IBM CoE, Malaysia
\textsuperscript{c} School of Computing and Digital Technology, Birmingham City University, Millennium Point, Birmingham, B4 7XG, UK
\textsuperscript{d} Research Center for AI and IoT, Near East University, 99138 Nicosia, Turkey
\textsuperscript{e} Department of Computer and Information Sciences, Fordham University, New York, NY 10458, USA

\textbf{ARTICLE INFO}

\textbf{Keywords:}
Covid-19
Dynamic clustering
Lockdown
Pandemic

\textbf{ABSTRACT}

The COVID-19 disease has once again reiterated the impact of pandemics beyond a biomedical event with potential rapid, dramatic, sweeping disruptions to the management, and conduct of everyday life. Not only the rate and pattern of contagion that threaten our sense of healthy living but also the safety measures put in place for containing the spread of the virus may require social distancing. Three different measures to counteract this pandemic situation have emerged, namely: (i) vaccination, (ii) herd immunity development, and (iii) lockdown. As the first measure is not ready at this stage and the second measure is largely considered unreasonable on the account of the gigantic number of fatalities, a vast majority of countries have practiced the third option despite having a potentially immense adverse economic impact. To mitigate such an impact, this paper proposes a data-driven dynamic clustering framework for moderating the adverse economic impact of COVID-19 flare-up. Through an intelligent fusion of healthcare and simulated mobility data, we model lockdown as a clustering problem and design a dynamic clustering algorithm for localized lockdown by taking into account the pandemic, economic and mobility aspects. We then validate the proposed algorithms by conducting extensive simulations using the Malaysian context as a case study. The findings signify the promises of dynamic clustering for lockdown coverage reduction, reduced economic loss, and military unit deployment reduction, as well as assess potential impact of uncooperative civilians on the contamination rate. The outcome of this work is anticipated to pave a way for significantly reducing the severe economic impact of the COVID-19 spreading. Moreover, the idea can be exploited for potentially the next waves of corona virus-related diseases and other upcoming viral life-threatening calamities.

1. Introduction

The world population currently faces one of the most life-threatening disasters since the past century with the emergence of a novel Coronavirus variation, commonly referred to as COVID-19 or 2019-nCoV (Backer, Klinkenberg, & Wallinga, 2020). With early confirmed cases of viral pneumonia dated back in December 2019 in Wuhan, China, this virus-induced disease has now affected at least 2.6 million population all over the world (Coronavirus Cases) with the death toll rate at 20% (Coronavirus Cases) within just less than 4–5 months. While the current mortality rate at 14.36% is less than the Coronavirus predecessors such as SARS (at 3.4%) (Roussel et al., 2020) and MERS (at 9.6%) (Jiang, Rayner, & Luo, 2020), the COVID-19 has distinguished characteristics of rapid spreading due to both the viral properties (e.g., long incubation period, asymptomatic effects for the super spreader, droplet carriers) (Shereen, Khan, Kazmi, Bashir, & Siddique, 2020) as well as current socio-demographic globalization and urbanization (ubiquitous mobility and affordable transportation). This had led the world health organization (WHO) to declare the virus as the global pandemic (Fosu & Edunyah, 2020) due to the scale, impact, and pressure to the public health system across the continents.

Given the wide prevalence of the Covid-19 outbreak, scientists and policymakers across the globe have continuously worked to mitigate the evolving impact of Covid-19 with varying measures planned and adopted over different places. The first one is the effort towards vaccine development and vaccination (Cohen, 2020). While this seems to be

\textsuperscript{*} Corresponding authors.
\textit{E-mail address:} arafatur@ump.edu.my (M.A. Rahman).

https://doi.org/10.1016/j.scs.2020.102372
Received 9 May 2020; Received in revised form 29 June 2020; Accepted 30 June 2020
Available online 03 July 2020
2210-6707/ © 2020 Elsevier Ltd. All rights reserved.
widely agreed with consensus among the vast majority of governments, the Covid-19 vaccines are not yet ready and it may take 12–18 months to complete the development (Dunlap, 2020). Even once they are ready, the affordability and availability for population across different economic levels may take some time until extensive deployment. The second measure is relying on the development of herd immunity (Kwok, Florence, Wei, Wong, & Julian, 2020) by allowing the virus to spread in the population until sufficient number of people are infected and develop a self-immune system to the virus. This method has been seen as controversial since it may overload and overburden the already pressurized healthcare systems, resulting in many deaths and fatalities. Therefore, only a few countries (Kwok et al., 2020) are still consistently pursuing this route, despite various calls to revise the policy. The third method is to impose a variety of strict social and physical distancing policies, resulting in “lockdown” situations, in order to contain and/or limit the spread of the virus. It corresponds to isolating physical environments (areas, places, cities, or even countries) from inward and outward traveling as well as imposing and regulating resident mobility within the given areas. The rationale is that the virus has already infected part of the population with an accurate figure of infectious proportion (hidden and asymptotic cases) being not fully understood. This measure has seen the effectiveness in countries such as France (Coronavirus Disease), China (Zhang, Jiang, Yuan, & Tao, 2020) where the curves for confirmed cases and deaths have started flattening.

For a variety of reasons, including its mitigation effectiveness in several countries, the lockdown policies seem to attract a growing popularity across a vast number of governments. A drastic lockdown measure imposes a national-level of mobility restriction with examples being practiced in countries such as Malaysia, Italy, France (Othman, 2020) while a more moderate restriction applies to provincial or regional levels. A major consideration of imposing lockdown in a given area is often due to a steady increase of uncontrollable and newly-appearing. Many countries are usually reluctant initially to directly impose a lockdown since the adverse prospect of the economy due to the restricted mobility (Koresponden, 2020). In many cases, informal labors and daily-earning workers are among the first to feel the impact, which in turn can lead to protests (BBC) and civil unrest (BBC).

In this work, we investigate a data-driven framework to potentially reduce a severe economic impact of the lockdown through improved resident mobility embedded within systematic changing boundary of the lockdown area. We envision near real-time soft lockdown boundaries that exhibit mobility-health risks and can be directly linked with permitted economic activities regulated by the authorities. More specifically, our contributions include the following:

- We devise a data capture framework that integrates public healthcare and telecommunication data records for subsequent epidemic decision making. While the former record has been subject to extensive recent research, utilization of the second record is motivated by gathering the population mobility through their interactions with cellular networks, logged by cellular providers. The cumulative data can be huge (Big Data) due to availability and usage of smart mobile devices all over the population.

- Based on the captured data, we suggest an artificial intelligence (AI) engine to generate a dynamic lockdown map with flexible data-driven updating capability in order to determine the lockdown boundaries.

- Using the infection data as well as mobility data, we partition a given location into a set of lockdown and non-lockdown clusters with soft transition between the extreme cases. The cluster resolution (in terms of area) will depend on the data resolution available at that particular location.

- By first assuming full cooperation from the population during this national crisis, we assess the effectiveness of this dynamic lockdown proposal in terms of the lockdown coverage area, the reduced economic loss as well as associated lockdown-deployment resources, highlighting the promises and potential issues.

- Through simulation, we further examine the impact of non-cooperative segment of the population (a.k.a. noise) to the system effectiveness in terms of average number of contamination, demonstrating that such a noise can significantly change the decision on the size and distribution of the lockdown clusters.

Due to possible cycle of pandemic situations, including corona virus (Yang et al., 2020), we see that such a dynamic lockdown mechanism can assist in devising a suitable strategy, not only in the current situation, but also the future pandemic. This solution will be well fitted in sustainable smart city and society.

The rest of the paper is organized as follows. Section 2 reviews related works with most focus on the non-clinical-related research for COVID-19. Section 3 discusses in details our proposed data-driven framework along with a dynamic clustering algorithm in support for reducing harsh economic consequences. Section 5 describes our simulation to assess the effectiveness of the framework and algorithm as well as discuss the results. Section 6 concludes this work with a summary of key findings.

2. Related works

There are several numbers of works that have been done to battle against the outbreak of COVID-19 and any pandemic situation. Majority of them are inclined to medical domain such as potential vaccines (Ahmed, Quadeer, & McKay, 2020; Prompatchara, Ketloy, & Palaga, 2020), early detection of COVID-19 using X-ray images (Wang & Wong, 2020) and so on. In addition, many technologies have been incorporated to mitigate the adverse impacts of COVID-19. In this section, we discuss that domain below.

Since the outbreak of the COVID-19, several statistical, dynamic and numerical models have been proposed to dissect its escalation dynamics (Li et al., 2020; Tuite & Fisman, 2020; Wu, Leung, & Leung, 2020). In Zhong et al. (2020), the authors have proposed a forecasting model to predict COVID-19 patients based on a plain mathematical model using limited epidemiological information. Even though the epidemiological models help assess the dynamics of escalation, they expect knowledge of certain parameters and rely upon different hypotheses. To surmount constraints of those methods, in Hu, Ge, Jin, and Xiong (2020) the authors have developed an AI model for real-time prediction of new and accruing cases of COVID-19 across China. Besides, this model predicts potential data trend and area of COVID-19 escalation in China and stratify them into different clusters.

These kinds of machine learning models otherwise known as AI-based predictive models are quite handy to classify COVID-19 positive and negative patients and would be helpful to mitigate the spreading of this virus (Rao & Vazquez, 2020), provided that, we utilize appropriately. The main component of constructing an AI-based predictive model is information a.k.a. data or big data. Big data innovation has experienced rapid development and accomplished incredible achievements in various sectors. Big data would be useful to eliminate this kind of pandemic situation and has huge opportunities for harmonizing sustainable development goals. By 2050, it is assumed that two-third of the world population is expected to be urbanized (Kharrazi, Qin, & Zhang, 2016). Advanced big data practices would help to ensure the security and safety of those people and work towards compiling the sustainable development goals. Different national agencies are trying to utilize satellite imagery, mobile phone data, health records as a source to compile statistical indicators such as compiling mobility, road safety, healthcare issues, etc. (MacFeely, 2019). Countries like Taiwan has approached such techniques and mastered the pandemic with the help of big data analytics (Wang, Ng, & Brook, 2020). Moreover, to fight against COVID-19, GIS (Geographic Information Systems) and big data innovations have performed major role in many scenarios such as rapid visualization, spatial tracking, geographical segmentation, balancing of...
resources, which provided exact information to support decision making, assess the effectiveness of this pandemic and control it properly (Zhou et al., 2020). These data are retrieved from different sources and thus, tend to be framed in a structured (e.g., tabular data) or unstructured (e.g., images, videos) form. Handling unstructured data are arduous and sometimes inconvenient for predictive models. Although structured data are flexible to use, AI models are perceptive to extreme values also known as outlier which is common in real life tabular data sets. In some cases, noises (e.g., inconsistency, missing or null values, outlier) caused by behavioral patterns and by distinguished events can further make the prediction task convoluted. To deal with extreme values in time-series datasets, in Lu, Zeng, and Chen (2009), the authors have proposed an algorithm based on Markov switching models that observe both regular and extreme values and output the result accordingly. The outcome of this algorithm outperforms several state-of-the-art methods of detection tasks.

It is worth mentioning that, data collection is another crucial aspect of any data-driven approach and a time-consuming procedure as well. This ruins the purpose and advantages of AI in such pandemic circumstances. To reduce the shortcoming, in Rao and Vazquez (2020) the authors have proposed to collect recent travel history alongside common symptoms of individuals using a cellular phone-based online survey. Such collected data can be fed into an AI framework to speed up the preliminary identification process of COVID-19 contamination. A risk map is designed further to isolate potential high-risk plateau early and abate the chance of expansion. Besides AI, visualization-based solutions are also provided by other researchers. In Jang, Lee, and Kim (2019), the authors have introduced a map-based information system during any contagious disease outbreak that visualizes the map spotting different clusters of risk levels (e.g., normal, caution and danger) and provides statistical analysis. Another major concern during and the post-pandemic situation is an economic perturbation. The timing of COVID-19 is unfortunate and caused countless travel cancellations, shutting down casinos, and so on (Bobdey & Ray, 2020). The consequences of a pandemic typically affect a large portion of the workforce as well. Motivated by this fact, in Orsi and Santos (2009) the authors have enhanced the dynamic inoperability input–output model (DIIM).

Unlike the aforementioned works, in this paper, we present a data-intelligence framework for alleviating adverse economic distress during and post-pandemic situation. The framework is aimed to deal with public data, ensure data security, and construct a real-time dynamic clustering algorithm with the aid of AI.

3. Data-driven framework for mitigating adverse economic impact

A data-driven framework (as depicted in Fig. 1), which alleviates the adverse economic repercussion due to lockdown during the COVID-19 outbreak, is proposed to overcome the drawbacks of conventional lockdown policies. The framework is detailed in this section.

3.1. Proposed framework

The proposed framework has three primary components that are associated with each other. These are: (i) the core telecommunication networks (CTNs), (ii) the COVID-19 control center (CCC), and (iii) the stakeholders. The CTNs are the base of this proposed framework whereas the CCC is responsible for receiving instructions from the stakeholders (e.g. National Health Organization and Health Ministry) upon which, it will accumulate information from the end-users exploiting the CTNs. This information will be further processed inside the CCC and the outcome of this procedure will be shared with stakeholders and end-users through the CTNs. The workflow of these phases is elaborated in detail in the following sub-sections.

3.1.1. Core telecommunication networks (CTNs)

A core communication network will act as a bridge between the CCC and end-users considering its two-way communication characteristic. The healthcare information with geo-location and telecommunication records of end-users will be extracted and further sent to the CCC in this segment. However, other network techniques can also be utilised for developing such communication networks (Abujubbeh, Al-Turjman, & Fahrioglu, 2019; Alabady, Salleh, & Al-Turjman, 2018; Rahman et al., 2019). Moreover, the CTNs will convey further updates or decisions provided by the CCC or stakeholders to the end-users. The network components and architecture are discussed in the following:

Network components: To operate the overall data exchange tasks, three network components will be used, which are elaborated below.

- End device: We choose cell phones as end devices due to their accessibility. Healthcare information (i.e., COVID-19 related features) and telecommunication records (i.e., user location and movement tracking information) will be retrieved from the end-user through a mobile application.
- Base station: A channel or route is a must to transmit data from one location to another. Base stations are those channels to ensure data transmission. In this framework, base stations are telecommunication towers that are connected to the CTNs. With the aid of mobile application, end-users information will reach to nearest telecommunication tower a.k.a. the base station, and further transferred to the CTNs.
- Core networks: This component acts as a backbone connectivity among different communicating terminals that access the services via the base station (access front-end). In this particular framework, three primary functions of the core networks, namely gateways, aggregation and service invocation are extensively leveraged, in addition to other functionalities such as authentication and switching. The gateways function serves as an interfacing layer among different access front-ends (e.g., the base station and the CCC). Aggregation refers to gathering diverse data from end users. On the other hand, service invocation corresponds to provision of services in the form of public broadcast of the dynamic cluster formation and information. These functions require the core networks to have high-capacity communication infrastructure to facilitate data exchange among different end users and stakeholders.

Network architecture: All the network components will be connected based on this architecture to transfer and receive information. The end device will be connected with the base station in a single-hop-communication manner according to the star topology. Finally, the CTNs will store the information into the data center. This architecture has a bidirectional data flow, which are End-device-to-Base-Station-to-CTNs-to-Data-Center and data center-to-CTNs-to-base station-to-end device. The earlier flow can be exploited for data acquisition from the end device, i.e., the people health information related to COVID-19. Based on this information, the suspected COVID-19 cases will be identified using data analytics (DA). The later flow will be exploited to inform people about their lockdown map and other COVID related information from the stakeholders.

3.1.2. COVID-19 control center (CCC)

The CCC holds the core position of this proposed framework. It is accountable for data streaming and generating final decisions. The operation within this unit comprises five sub-phases, which are discussed in the following:

1. Data center: Data transmitting from stakeholders and CTNs will be stored and preserved into the data center. The data transmission approach is designed as two-way communication between the stakeholders and CTNs. The purpose of two-way communication is to
deliver final decisions to the stakeholders and end-users and accumulate data from those zones as well. Besides, depending on the situation, information and messages from stakeholders will be conveyed to end-users through the data center. Once the data are stored properly, the data center will handover this information to the data security for ensuring the privacy of individuals.

2. Data Security: It is an important part that will assure security Triad CIA (i.e., Confidentiality, Integrity and Availability) in order to develop secure data collection and trustworthy networks (Rahman & Assihary, 2020; Tao et al., 2020). The personal healthcare information of a user will be kept confidential (C), cannot be changed by others (I) and the data can be accessed anytime (A). A blockchain enabled healthcare security system named MediChain was presented in Abdullah Al Omar and Md Zakirul Alam Bhuiyan (2019), which can be further enhanced and exploited here. One of the key issues of the MediChain is scalability, which needs to be addressed in the proposed system since a large number of users may involve with this system.

(a) Multi-transnational block: The first feature will facilitate quick and easy processing of data from the client to the data center. Transnational data is not restricted to a limited block size to ensure that multiple transactions within the same block can share the same unique hash code. Thus the time and computing power required to process the data will be exponentially reduced.

(b) Multi-node approach: A multi-node server architecture will ensure that load balancing can be optimized within the network. Stress factors such as an influx of user submissions and server disparity can be mitigated and distributed. Thus, stress points are evened out across the blockchain network.

(c) Lightweight framework: A lightweight web framework can connect multiple servers with its designated clients. A lightweight framework reduces the overhead costs of the blockchain system as well as minimizing latency among the users and service providers.

Once the data security is ensured, these data will be passed to the data analytics unit.

3. Data Analytics: In this sub-phase, a behavioral analysis of aggregated COVID-19 data from the National Health Organisation and CTNs will be executed. Additionally, an AI engine will be devised here to predict potential COVID-19 carriers. The methodology of this predictive model is described below:

(a) Workflow: Our methodology (as depicted in Fig. 2) toward predictive modeling starts with generating hypotheses depending on the problem statement, which, in this case detecting potential COVID-19 patients. We sketch an underlying AI framework to start-off the procedure. As reaping data from the NHO and CTNs is a continuous process, our proposed approach will differentiate them as pre-processed data and current data. Current data will go for the data integration, participate in the exploratory data analysis, and afterward enter into the data preparation section. Whereas, previously processed data will be used for data augmentation to increase the number of training examples in the data preparation section. Once data are processed, we will then extract significant features from the dataset otherwise known as feature engineering. This will help our model to predict more precisely. From that point onward, we will train a deep neural network (NN) model feeding those processed data. The value of hyperparameters will be tuned for optimal solutions and our model will be evaluated further. We will perform this process repeatedly until we locate the desired result. Then, we will check if it satisfies our current objective or not. Negative feedback will force back this procedure to the beginning and it will go through the same processes until the result matches the end goal. We will then deploy the model for real-world usage. Here, two functionalities will be operated. One is the real-time prediction for potential COVID-19 patients, and another one is a data visualization of different scenarios of COVID-19. Moreover, periodical monitoring and debugging will also be available in our proposed workflow to keep updated our model.

(b) ML model architecture: If we do not make any assumption on data there is no reason to choose an ML model over another model, which is known as No Free Lunch (NFL) (Wolpert, 1996) theorem. For some datasets, the support vector machine works fine while the neural network fits well with other types of datasets. No model can ensure to work better in various datasets.
In that case, we must try each of them and evaluate to get the best model which is impossible. Thus, in practice, it is better to make some reasonable assumptions about the data and try out a few reasonable models and find the best amongst them. We will introduce NN to our model, which is an integral part of deep learning. Deep learning is a branch of machine learning algorithms that utilizes multiple layers to dynamically extract higher-level features from the raw input. A deep neural network (DNN) is an artificial neural network (ANN) with numerous layers between the input and output layers. Regardless of a linear or non-linear relationship DNN identifies factual mathematical manipulation to transform the input into the output. The network travels through the layers computing the likelihood of each yield (Dreiseitl & Ohno-Machado, 2002). The reason behind using DNN in our model is that DNN works better with complex data with a huge number of features, generally known as high dimensional data (Zhang, Yang, Chen, & Li, 2018), than basic machine learning algorithms and in our case, there is a possibility of having a higher number of input features. Training examples will be the input for the input layer and the number of hidden layers will be defined according to the problem. We use ReLU (Rectified Linear Unit) (Nwankpa, Ijomah, Gachagan, & Marshall, 2018) as an activation function for hidden layers aside from the last one. As our concern is binary classification thus, we use the sigmoid (Nwankpa et al., 2018) function for the last hidden layer. Finally, we will get the predicted outcome from our model in the output layer.

4. Dynamic clustering: The novelty of this framework is underlined in this sub-phase. A clustering algorithm will be forged here to delineate the lockdown strategy. Unlike other clustering algorithms, the proposed algorithm is dynamic, means shrinking and expanding different clusters will be varied and updated regularly due to the merit of the data. At first, using the infection data and mobility data processed in the Data Analytic sub-phase, a dynamic lockdown cluster will be initialized. This dynamic lock-down cluster will be forwarded to the Decision Engine and remain unchanged until new data arrives from DA. If new information about COVID-19 patients arrives from DA, it will be passed through a decision parameter that checks the similarity of new data with the previous one. If it is similar and no change is required, the lock-down cluster will remain the same. On the other hand, if it finds any dissimilarity the lock-down cluster will be altered according to the information. The algorithm is discussed in detail in Section 4.

5. Decision engine: This phase of CCC will monitor the entire procedure of different sub-phases and account for generating end decisions too. DC will provide the cluster blueprint and Data Center will provide the necessary information and requirements asked by the stakeholders. Taking those into account, the decision engine will conclude a decision and forward it to data center. As data center communicates with both stakeholders and CTNs, it will inform them about the decision.

6. Stakeholders: Stakeholders are the end decision-makers as well as the information providers. Stakeholders could be any Government Organizations such as the Health Ministry, National Health Organization (NHO), or Non-Government Organizations (NGOs). They will provide significant information and requirements for CCC. CCC will process these things and convey them back to the stakeholders. Analyzing those results from CCC, stakeholders will take decisions on what to do depending upon circumstances.

3.2. Complexity and cost consideration

In order to deploy the aforementioned framework, there are several aspects that we need to take into account such as (i) Network Infrastructure Deployment Cost; (ii) System Design Cost; (iii) Ensuring Security Cost. In terms of the first aspect, we do not need any costing as this project will exploit existing telecommunication networks. Moreover, the network utilization cost could be waived due to the support from telecommunication companies to the Government for taking control of this pandemic. For example, many telecommunication companies have already offered free services worldwide. Moving towards the second aspect, as this paper envisions that, the proposed system will be implemented by the Government, we assume they already have data centers to store all the information. We only need to consider the cost of system design which is supposed to be affordable to the Government. To ensure security, a scalable security platform is needed to be developed and the cost will be comparable to the existing security platform. Therefore the implementation cost will not be a big issue and the system complexity should not be high as well.
4. Methodology

From Section 3, we get a clear overview of our proposed model. However, in this paper, we divide our methodology into five phases which are portrayed beneath.

4.1. Phase one: problem definition

As mentioned earlier in Section 2, the conventional lockdown system is insufficiently flexible. The emerging opportunities of “Big Data” seems like potential alternatives. Many researchers have been exploiting this idea and proposing applications that make use of big data and machine learning to minimize the economic catastrophe (Mittal, Goyal, Sethi, & Hemanth, 2019) in different situations. However, a few works have been done to reduce the economic loss in a pandemic situation such as COVID-19 outbreaks. These non-clinical solutions are static and usually one-dimensional, making them less suitable for mitigating the economic impact on different parameters in a pandemic situation. On the contrary, a data-driven dynamic clustering algorithm would be much more efficient than a static lockdown strategy in such scenarios. Like many other approaches, public healthcare information and geographic location can be used for designing a dynamic clustering algorithm in this paper. Machine learning and big data utilization to invent vaccines (Ong, Wong, Huffman, & He, 2020) or any clinical operations are out of the scope of this paper.

4.2. Phase two: proposed algorithm

The clustering algorithm is illustrated in Algorithm 1 where we initialize a lockdown map and we assume it will be updated every day. To start with, we define a set of variables that is depicted in Algorithm 1 – Line #2. Here, TGSA, an ‘N by N’ array, refers to the total grid size of the area where N is the area size in km² and each grid is 1 km². MGSCA is the minimum grid number of clusters. LIASC is the location information of active and suspected cases which will be collected from Core Telecommunication Networks. LP is the total number of days during the lockdown. LCIM is the lockdown cluster information map. We start searching for TGSA from day one until the lockdown period. At first, there will be no lockdown cluster information map, thus we initialize LCIM as zero. After that, as the search progresses, we count the number of patients using a function CNASC and store that in a variable called CPD in Line #7 of Algorithm 1. We then enter the value of CPD into LCIM. In Line #13, we introduce a function named CreateClusteringMap to generate clusters. This function takes three arguments that are lockdown cluster information map, the threshold for the red zone, and the threshold for the orange zone. We define a variable Clusters which is the number of active and suspected cases in all clusters. If this Cluster value surpasses the threshold of the red zone, a red color will be assigned there which means it should be under a hard lockdown zone. If the value of the cluster exceeds the threshold of the orange zone, orange color will be assigned there. If any cluster fails to exceed either of those thresholds mentioned above, the cluster will be marked by a green color indicating a safe zone.

Algorithm 1. Dynamic clustering algorithm

| 1: | Initialization: //Data Sources from Stakeholders and CTHs; We assume that Lockdown Map will be updated Per Day |
| 2: | TGSA[N, N], MGSCA, LIASC, LP, LGIM[N/MGSCA, N/MGSCA], TRZCC, TOZCC; |
| 3: | for Day=1; Day < = LP; Day++ do |
| 4: | LCIM == Zero; |
| 5: | for i = 1; i < M; i + = MGSCA; do |
| 6: | for j = 1; j < N; j + = MGSCA; do |
| 7: | CPD = CNASC(TGSA[[i,i+ MGSCA]]][j], MGSCA), LIASC); |
| 8: | if CPD > 0 then |
| 9: | LCIM[i,j] = CPD; |
| 10: | end if |
| 11: | end for |
| 12: | end for |
| 13: | //Create Clustering Map according to the Red, Orange and Green Zone |
| 14: | Create_Clustering_Map(LCIM, TRZCC, TOZCC) |
| 15: | end |
| 16: | Cluster(Total_Cluster) Number of Active and Suspected Cases in all Clusters. |
| 17: | for cl = 1; cl < Total - Cluster; cl++ do |
| 18: | if Cluster(cl) > TRZCC then |
| 19: | Cluster(cl) will be marked as Red-Zone i.e., Hard lockdown; |
| 20: | else if Cluster(cl) > TOZCC then |
| 21: | Cluster(cl) will be marked as Orange-Zone i.e., Soft lockdown; |
| 22: | else |
| 23: | Cluster(cl) will be marked as Green-Zone i.e., Safe Zone; |
| 24: | end if |
| 25: | end for |
| 26: | } |
| 27: | end for |

4.3. Phase three: analytical model

To analytically support our experiments, we have calculated the R factor often described as the basic reproduction number R Naught (R₀). The parameter R₀ is characterized as the number of cases that expected to happen on average in a homogeneous population due to infection by an individual when the population is susceptible at the start of an epidemic. For example, if a person develops a disease and affects three others, the R₀ is 3. From an epidemiological model called SIIR (Aranson, Brassey, & Mahtani, 2020), we find that R₀ is based on mainly three factors namely susceptibility, infectivity, and removal rate. Susceptibility is the size of the population that is susceptible to disease. Infectivity is the accurate rate of infection and removal rate is the rate of closed cases either by the recovery or deaths. There are many ways to calculate the basic reproduction number. However, a commonly used equation for calculating R₀ is

\[ R_0 = \frac{\beta S}{\gamma} \]  

(1)

where \( \beta \) represents transmission rate and \( \gamma \) is the removal rate. S belongs to the susceptible population. At the start of an epidemic, it is assumed that the whole population is susceptible, thus the value of S is 1. In our experiment, we used the same value when MCO was not imposed. After imposing MCO, we assume 70% of people maintained social distancing and government orders. Thus 30% of the population were susceptible during MCO and the value of S became 0.3. As COVID-19 is an airborne disease, the transmission rate \( \beta \) is high. From Salim et al. (2020) we find the transmission rate in Malaysia during the MCO and non-MCO periods which are 0.58 and 0.69, respectively. From (EVOZI), we took a total of 45 days of data from the 3rd March 2020 to the 17th April 2020 to calculate R₀. We collected the number of closed cases and using the equation below we calculated the removal rate \( \gamma \), i.e.,

\[ \gamma = \frac{\theta_r + \theta_d}{\theta_r} \]  

(2)

where \( \theta_r \) is the total number of recoveries, \( \theta_d \) is the total number of deaths and \( \theta_t \) is the number of total affected people. The mean incubation period of COVID-19 is 6.4 days (Weibull distribution, 95% credible interval (CI): 5.6–7.7) and the range of incubation period is 2.4–15.5 days (Backer et al., 2020). From Fig. 3 we can notice that, the basic reproduction number reached its peak on around 15th day when MCO was not imposed and sharply declined in the MCO period.

The value of this reproduction number depends on many factors with one of which being the mobility of infectious individuals. Towards the end of the paper, we investigate the impact of such a mobility when the movement control order is imposed. For this purpose, our analysis leverages the mathematical findings from the graph theory. More
specifically, by modelling the infection spread as a multiple random walks problem and interacting particle system (Cooper, Frieze, & Radzik, 2009), we inspect the graph cover time, which provides estimation on how quick the infection can spread in the community. Specific results and discussion are given in Section 5.4.

4.4. Phase four: experimental setup

In all the experiments, we have used the same device and MATLAB software for simulation. We have followed the scenarios based on the Malaysian Context. Malaysia has a land area of 329,613 km², comprising of 13 states. However, this total area includes oceans, mountains, forests, etc. where people can hardly be found. Thus, in our simulation, we use 25% of the total land area corresponding to roughly 83,000 km². For simplicity, we assume this portion to be a square shape and divide it into grids. We then set those 13 states as 13 squares with certain area values, a total equivalent to one-fourth of the entire land area of Malaysia. Finally, we generate random numbers to put victims inside those 13 states. To yield results from our simulation we retrieve data from the Malaysian Government Website (Evozi), where key information of this experiment is the COVID-19 active cases. Although we need other data such as the number of suspected cases, their location, etc., we will only highlight the active cases as portrayed in Fig. 4 due to the unavailability of the aforementioned data. Besides, there will be some parameters related to specific coverage areas, namely minimal, moderate, maximum, and six weeks of movement control order (in days), which will play a pivotal role in the simulations. However, these parameters may vary country to country with respect to different factors such as population density, land area, spreading factors etc. From Fig. 4, we can observe that initially, the number of active cases accelerate exponentially and afterward decline linearly. As the incubation period takes time around 3–16 days, the basic reproduction number (as depicted in Fig. 3) justifies the active cases (as depicted in Fig. 4) to accelerate exponentially and then decline eventually. We will see how this phenomenon plays a role in our dynamic lockdown strategy in the following experiments, demonstrated in Section 5.

4.5. Phase five: performance evaluation

In this phase, we have highlighted the impact of dynamic clustering concerning several aspects such as the lockdown coverage area, the reduced economic loss, and also the number of involvements for instance number of militaries, doctors, vaccines, centers, and so on. Finally, we will also examine the “noise” component in a lockdown practice, which corresponds to asymptomatic patients and uncooperative civilians. Along with that, we have validated the results through an analytical model, explained in Section 4.3. Results of these experiments are disclosed in the following section.

5. Results and discussion

In this section, we have performed four major experiments with respect to our proposed algorithm. The outcome of these experiments are discussed below:

5.1. Lockdown coverage area

In the first experiment, we have done performance comparison (as depicted in Fig. 5) to demonstrate the impact of dynamic lockdown and static lockdown concerning a specific coverage area and movement control order (in days). In the dynamic lockdown strategy, we categorize the coverage area into three parts, namely minimal, moderate, and maximum corresponding to five, seven, and nine kilometers, respectively. As expected, in this simulation we can see the number of coverage areas is dynamically evolving due to a reason indicated by Fig. 4. The static strategy contributes all the territory under lockdown whereas the dynamic strategy with maximum coverage expands a bit more than half of the static lockdown area. This captures an interesting trade-off of the dynamic lockdown strategy. Although minimal coverage area cuts down a lot of lockdown spaces and gives people the flexibility to move, the risk factor increases and it has a more chance of community spreading of COVID-19. These will play an important role...
related to the noise, which will be discussed in Section 5.4. On the contrary, the maximum coverage area emphasizes safety, restricting people’s flexible movement.

5.2. Potential reduction of economic loss

Here we will reveal the economic impact of using the dynamic lockdown strategy. We intend to show how we can quantify the losses occurring during the COVID-19 outbreak by utilizing the three approaches of our dynamic lockdown framework. A recent report by the Malaysian Government (Bloomberg), for example, shows an amendment of RM 2.4 billion per day because of this situation. If we use the static lockdown strategy where all the areas are shut down, from Fig. 6, we see that we cannot recuperate anything. However, using our approach, we can save up to 80 percent of economic loss in week one. For maximum, moderate, and minimal coverage areas, we can restore 60%, 70%, and 80% of the losses, respectively. This is reasonable in light of the fact that for the minimal approach, the coverage area is less, and saving can be improved. We further notice that the behavior of this reduced economic loss graph is also dynamic and follows according to Fig. 4.

5.3. Military unit utilization

Even though there are numerous resources contributing to deal with this pandemic circumstance, due to the unavailability of accurate information, in this experiment, we have only highlighted the deployment of the military resources. We have kept the simulation setup analogous as before. However, a few changes have been made to obtain useful findings. The Royal Military (7500 ATM units) is helping the Police in making 405 roadblock checkpoints across Malaysia (Majlis Keselamatan Negara). These involvements are based on schedules with the duration of each being 6 h. For the simplicity of this experiment, we have calculated these involvements as per day. We have randomly put a roadblock which covers around a radius of 3 km. If an infected person presents in that region the roadblock will remain active, otherwise, it will be shifted somewhere else. After doing this simulation, we have delineated the findings in Fig. 7. Herein we notice that our dynamic lockdown strategy has performed outstandingly as expected and subsequently our strategy has managed to reduce the military deployment rate down to practically 67%. It is worth mentioning that, the characteristic of this graph changes dynamically based on a reason given in Fig. 4. This is also logical since more casualties need increasingly military support, and thus, when the active cases are at the peak the requirement for military units is likewise high.

5.4. Effects of noise

Even though our previous experiments have delineated the benefits of using our framework, in this experiment we will show the potential shortcoming of our strategy. As our lockdown strategy works with public data, there could be a possibility of uncooperative civilians. These people might hide any information or not bother to provide any relevant data. Besides, asymptomatic patients can likewise be available in these situations. We model those uncooperative civilians and asymptomatic patients as the “noise”, which can have a devastating effect due to its spreading nature. However, in this experiment, we will show how compelling our framework to deal with this necessarily. According to Bostock (2020), in one cycle a person can contaminate a normal flu up to 1.3 persons. However, in the case of COVID-19, a noise can spread up to 4 persons, which is triple to the rate of normal contamination. In this simulation, we have considered one cycle only and simulated the movement of each noise using a random walk mobility model. Fig. 8 depicts the effects of noise (numbered from 20 to 80) on the average number of contamination for the dynamic lockdown strategy. Note that without any lockdown, 20 noises can spread up to 80 persons whereas, in the dynamic lockdown, as shown in Fig. 8, the contamination drops off to nearly 10 persons with maximum coverage area. With more area being in a locked-down state, the less the possibility of community spreading is occurring. Disregarding the way that as noise goes up the number of passive contacts is likewise high, the
dynamic lockdown strategy appears to be a more reasonable choice than the no lockdown strategy.

An analytical explanation from the noise findings in Fig. 8 can be drawn from the graph theory. By modelling individual grid units as vertices and the mobility pattern as edges, we can cast the infection spread as a multiple random walks problem. By assuming k infectious individuals as noises at the beginning of the random walks and assuming their independent movements in a r-regular graph (r ≥ 3), we can deduce the cover time over the graph given by—with high probability—(Cooper et al., 2009)

\[
C_r(k) \sim \frac{\tilde{c}_r}{k} \log n
\]

(3) for \( k = \alpha(n/\log^2 n) \), and

\[
C_r(k) = O\left( \frac{n}{k} \log n + n \log n \right)
\]

(4) for any other \( k \), where \( \tilde{c}_r = \frac{r^{-1}}{2} \) and \( n \) denotes the number of vertices. The cover time can offer a simplifying view about the speed of the virus spread. As the noises (non-complying individuals) \( k \) increase, the smaller the time required to cover the movement across every vertex in the graph. By a continuity argument, it means that at any given time after this movement, the probability of people residing in the vertices of the graph getting exposed to the virus is amplified with an increase in \( k \). This might explain why the charts in Fig. 8 consistently escalate when the noise value is increased.

Due to our unavailable access to required data, some of these experiments have been performed using randomly simulated data. However, if we have accurate and undistorted data such as the exact location of individuals, we could have derived more meaningful findings from that.

6. Conclusion

In this paper, we have developed a data-driven dynamic clustering framework for mitigating the adverse economic impact of the COVID-19. Our propose framework has been described to have three central components, namely the Data Analytics, Dynamic Clustering, and Data Security. In our analysis, we have mainly focused on Dynamic Clustering and its impact so that both stakeholders and users can infer the benefits. A clustering algorithm has been proposed, which has been simulated extensively across four scenarios to identify the promises and shortcomings of this algorithm. We have noticed that our clustering algorithm significantly improves the relevant performance metrics by almost 50% in the lockdown coverage experiment, 60–80% in the potential reduction of economic loss, and 20% in the military unit utilisation. An actual scheme will be further re-designed by taking into consideration the real data, which will serve as a future direction of this work. Using our proposed framework as a baseline, the future work may include investigation of specific machine learning techniques in which relevant features can be extracted and selected for training the machine learning models.

Conflict of interest

None declared.

Acknowledgment

This work was partially supported by a Research Grant (RDU192215), which is funded by University Malaysia Pahang.

References

A.B.S.K.M.S.R. Abdullah Al Omar, & Md Zakirul Alam Bhuiyan (2019). Privacy-friendly platform for healthcare data in cloud based on blockchain environment. Future

Generation Computer Systems, 95, 511–521.
Ahuja, M., Al-Turjman, F., & Fahroo, M. (2019). Software-defined wireless sensor networks in smart grid: An overview. Sustainable Cities and Society, 51, 101754. https://doi.org/10.1016/j.scs.2019.101754.
Ahmed, S. F., Quadeer, A. A., & McKay, M. R. (2020). Preliminary identification of potential vaccine targets for the covid-19 coronavirus (sars-cov-2) based on sars-cov immunological studies. Viruses, 12(3), 254.
Alabady, S. A., Salleh, M. M. F., & Al-Turjman, F. (2018). Lpe error correction code for iot applications. Sustainable Cities and Society, 42, 663-673. https://doi.org/10.1016/j.scs.2018.01.036.
Arondon, J. K., Brasseye, J., & Mahanti, K. R. (2020). When will it be over?: An introduction to viral reproduction numbers, \( r_0 \) and \( r_e \). The Centre for Evidence-Based Medicine develops, promotes.
Backer, J. A., Klinkenberg, D., & Wallinga, J. (2020). Incubation period of 2019 novel coronavirus (2019-ncov) infections among travellers from Wuhan, China, 20-28 January 2020. Eurosurveillance, 25(5), 2000662.
Corona virus lockdown protest: What’s behind the us demonstrations? (2020). https://www.bbc.com/news/world-us-canada-52359100. http://www.bloomberg.com/news/articles/2020-04-25/malaysia-sees-2-4-billion-trothers-a-day-of-losses-from-lockdown.
Bosch, B. (2020). An intensive-care expert broke down just how contagious the coronavirus is, showing how one person could end up infecting 59,000 in a snowball effect. https://www.businessinsider.my/coronavirus-vs-flu-social-distancing-infections-spread-explainer-video-2020-3?r=US&IR=T.
Cohen, J. (2020). Vaccine designers take first shots at covid-19. https://www.medscape.com/viewarticle/924015.
Cooper, C., Friese, A., & Radzik, T. (2009). Multiple random walks and interacting particle systems. In S. Albers, A. Marchetti-Spaccamela, Y. Matias, S. Nikoletseas, & W. Thomas (Eds.). Automata, languages and programming (pp. 399–410). Berlin, Heidelberg: Springer Berlin Heidelberg.
Coronavirus cases. https://www.worldometers.info/coronavirus/.
Coronavirus disease 2019 (covid-19). https://www.ecdc.europa.eu/sites/default/files/documents/RRA-seventh-update-Outbreak-of-coronavirus-disease-COVID-19.pdf.
Dreisigfeld, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: A methodology review. Journal of Biomedical Informatics, 35(5–6), 352–359.
Dunlap, K. (2020). How long will it take to have a vaccine for covid-19? https://www.clickorlando.com/features/2020/04/02/how-long-will-it-take-to-have-a-vaccine-for-covid-19/.
Evozi, Covid-19: Malaysia outbreak monitor: Live updates. https://www.outbreak.my./
Fosu, G. O., & Edunyah, G. (2020). Flattening the exponential growth curve of covid-19 in Ghana and other developing countries; divine intervention is a necessity. Divine Intervention & A Necessity.
Heidelberg: Springer Berlin Heidelberg.
Hu, Z., Ge, Q., Jin, L., & Xiong, M. (2020). Artificial intelligence forecasting of covid-19 in China. arXiv:2002.07112.
Jang, B., Lee, M., & Kim, J. W. (2019). Peacock: A map-based multivariate disease outbreak information system. IEEE Access, 7, 82966–82969.
Jiang, X., Rayner, S., & Luo, M.-H. (2020). Does sars-cov-2 have a longer incubation period than sars and mers? Journal of Medical Virology.
Kharrazi, A., Qin, H., & Zhang, Y. (2014). Urban big data and sustainable development goals: Challenges and opportunities. Sustainability, 6(12), 2295.
Koresponden, N. (2020). Govt reluctant to implement lockdown due to economic concerns. https://en.tempo.co/read/1320985/govt-reluctant-to-implement-lockdown-due-to-economic-concerns.
Kwok, K. O., Florence, L., Wei, W. I., Wong, S. Y. S., & Julian, T. (2020). Herd immunity-estimating the level required to halt the covid-19 epidemics in affected countries. Journal of Infectious.
Lee, M., Kim, J. W., & Jang, B. (2018). Dove: An infectious disease outbreak statistics visualization system. IEEE Access, 6, 47206–47216.
Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., et al. (2020). Early transmission dynamics in Wuhan, China of novel coronavirus-infected pneumonia. New England Journal of Medicine.
Lu, H.-M., Zeng, D., & Chen, H. (2009). Prospective infectious disease outbreak detection using Markov switching models. IEEE Transactions on Knowledge and Data Engineering, 22(4), 565-577.
Macleod, S. (2020). The big (data) bang: opportunities and challenges for compiling SDG goals: Challenges and opportunities. Sustainable Cities and Society, 521.
Majlis keselamatan negara (rasmi). https://t.me/MKNRasmi/683.
Mittal, M., Goyal, L. M., Sethi, J. K., & Hemanth, D. J. (2019). Monitoring the impact of the pandemic: A network view. International Journal of Computers, Network and ISDN Systems, 4(2), 301–305.
Othman, A. D. P. B. (2020). Countries continue with restrictions. https://www.thestar.com.my/news/nation/2020/04/16/countries-continue-with-restrictions.
Part A: Systems and Humans, 40(2), 359.
Rahman, M. A., & Assihary, T. (2020). Scalable machine learning-based intrusion de-
tection system for iot-enabled smart cities. Sustainable Cities and Society submitted.
https://doi.org/10.1016/j.scs.2019.101754.
Rahman, M. A., Asyhari, A. T., Azad, S., Hasan, M. M., Munaiseche, C. P., & Krisnanda, M. (2019). A cyber-enabled mission-critical system for post-flood response: Exploiting TV white space as network backhaul links. IEEE Access, 7, 100318–100331.

Rao, A. S. S., & Vazquez, J. A. (2020). Identification of covid-19 can be quicker through artificial intelligence framework using a mobile phone-based survey in the populations when cities/towns are under quarantine. Infection Control & Hospital Epidemiology, 1–18.

Roussel, Y., Giraud-Gatineau, A., Jimeno, M.-T., Rolain, J.-M., Zandotti, C., Golson, P., et al. (2020). Sars-cov-2: Fear versus data. International Journal of Antimicrobial Agents, 105947.

Salim, N., Chan, W. H., Mansor, S., Bazin, N. E. N., Amarana, S., Faudzi, A. A. M., et al. (2020). Covid-19 epidemic in malaysia: Impact of lock-down on infection dynamics. medRxiv.

Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., & Siddique, R. (2020). Covid-19 infection: Origin, transmission, and characteristics of human coronaviruses. Journal of Advanced Research.

Tao, H., Bhuiyan, M. Z. A., Rahman, M. A., Wang, T., Wu, J., Salih, S. Q., et al. (2020). Trustdata: Trustworthy and secured data collection for event detection in industrial cyber-physical system. IEEE Transactions on Industrial Informatics.

Tuite, A. R., & Fisman, D. N. (2020). Reporting, epidemic growth, and reproduction numbers for the 2019 novel coronavirus (2019-ncov) epidemic. Annals of Internal Medicine.

Wang, L., & Wong, A. (2020). Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest radiography images. arXiv:2003.09871.

Wang, C. J., Ng, C. Y., & Brook, R. H. (2020). Response to covid-19 in Taiwan: Big data analytics, new technology, and proactive testing. JAMA, 323(14), 1341–1342.

Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. Neural Computation, 8(7), 1341–1390.

Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, China: A modelling study. The Lancet, 395(10222), 689–697.

Yang, Y., Peng, F., Wang, R., Guan, K., Jiang, T., Xu, G., et al. (2020). The deadly Coronaviruses: The 2003 SARS pandemic and the 2020 novel Coronavirus epidemic in China. Journal of Autoimmunity, 109, 102434.

Zhang, Y., Jiang, B., Yuan, J., & Tao, Y. (2020). The impact of social distancing and epicenter lockdown on the covid-19 epidemic in mainland China: A data-driven seir model study. medRxiv.

Zhong, L., Mu, L., Li, J., Wang, J., Yin, Z., & Liu, D. (2020). Early prediction of the 2019 novel coronavirus outbreak in the mainland china based on simple mathematical model. IEEE Access, 8, 51761–51769.

Zhou, C., Su, F., Pei, T., Zhang, A., Du, Y., Luo, B., et al. (2020). Covid-19: Challenges to GIS with big data. Geography and Sustainability.