Bulletin of Electrical Engineering and Informatics
Vol. 9, No. 6, December 2020, pp. 2396~2403
ISSN: 2302-9285, DOI: 10.11591/eei.v9i6.2621

An open toolbox for generating map of actively confirmed SARS-CoV-2 or COVID-19 cases in Vietnam

Duc Chung Tran
Computing Fundamental Department, FPT University, Hoa Lac Hi-Tech Park, Vietnam

Article Info

Article history:

Received Mar 21, 2020
Revised May 29, 2020
Accepted Jun 30, 2020

Keywords:

Confirmed case
Country map
COVID-19
Novel coronavirus
SARS-CoV-2

ABSTRACT

The recent outbreak of novel coronavirus, SARS-CoV-2 or COVID-19, discovered in late 2019, being continued to spread across regions worldwide, has resulted in 1,914,916 "confirmed" cases with up to 123,010 deaths, as in situation report –85 by World Health Organization (WHO). Most of the developed disease monitoring and tracking tools currently available only present the reported cases up to country-level and not detail down to provincial- or state-, city- level within the countries. This is insignificant for supporting activities in quickly reducing and preventing the spread of the disease within a certain country because further detail potential infectious locations are not provided for people to avoid traveling or passing by there. Thus, this work presents an open toolbox for generating map of actively "Confirmed" cases in a country, i.e., Vietnam, given a dataset containing their statuses and current locations, detail down to provincial-or state-, city-level. The newly released algorithm reduced approximately 24.41% of processing time of the preceding one. In addition, the algorithm can be easily extended for supporting other countries given suitable datasets.

Corresponding Author:

Duc Chung Tran,
Computing Fundamental Department,
FPT University,
Hoa Lac Hi-Tech Park, Hanoi, 155300, Vietnam.
Email: chung.tranduc89@gmail.com

1. INTRODUCTION

The recent outbreak of coronavirus, SARS-CoV-2 or COVID-19 [1-7], discovered in late 2019, being continued to spread across regions worldwide [8-11], resulted in 1,914,916 "confirmed" cases with up to 123,010 deaths [12]. The disease is spreading extremely quickly resulting in exponentially infected cases all over the world [13-16]. As a result, catastrophic consequences happen to the world's economics [13, 17-20]. Since the detection of the virus, researchers worldwide have paid rapid attentions and put great efforts in addressing plenty of socioeconomic consequences. In [15] presented an approach for predicting the trend of the pandemic in Italy with the main purpose for developing strategies for public health. The time series data from 22th Jan 2020 to 16th Mar 2020 were used by the model named eSIR. The model utilizes different intervention measures' effects of dissimilar period in the prediction. In addition, Markov Chain Monte Carlo methods were used to estimate the basic reproductive number. Working on different approach, [14] discussed the severity of acute respiratory syndrome of infected patients and the challenges of this pandemic. Based on the research, the mean incubation period was found to be approximately 7 days while the basic reproduction number was between 2.24 and 3.58 [14]. Therefore, patients should be monitored closely. Aside from the pneumonia as the most manifestation of the virus, [4] suggested that extra-pulmonary symptoms such as initial cerebrovascular shall manifest the virus. In order to prepare for a quick response to COVID-19, UW Medicine presented an impatient response plan for high-quality palliative care [8] while Australian authority kept continuous updates of the situation in its formal reports [3]. In another research, [13] discussed the implication for physiotherapy topic with respect to global impact. It was stated that the clinical health system has plenty of difficulties to deal with the surge; and at global level, workers are among the ones to have least access to such medical care system. Meanwhile, it was studied that pregnant shall expose to greater risks than normal people when infected by the virus [2]. For supporting fast detection of the virus, i.e., under 30 minutes, [9] presented an approach by reversing transcription-loop-mediated (RT-LAMP) isothermal amplification.

It is seen from the above analysis that various research topics were studied since the pandemic of COVID-19. In order to effectively prevent the spread of the disease, it is critically important for people in a specific country to be aware of the potential infectious locations (i.e., provinces or states or cities) to avoid traveling or passing by there. There have been multiple attempts to present the global map of the infectious condition. However, currently there is only country-level data in reports of WHO [12], real-time dashboard COVID-19 tracking [1] as shown in Figure 1, dashboard in [1, 21], information in [22-31] as shown in Figures 2, 3. Meanwhile, data in [32] had provincial level for Vietnam, however they are incomplete, only for some cities such as Hanoi, Vinh Phuc, Nha Trang and Ho Chi Minh. Thus, the provided information is not sufficient for offering essential warnings for people living in the country to be aware of the potential infectious areas and avoid to move there for preventing the spread of COVID-19. Here, it should be noted that the potential infectious areas are defined as those the actively "confirmed" cases have passed by recently which is detailed down to district or state level.

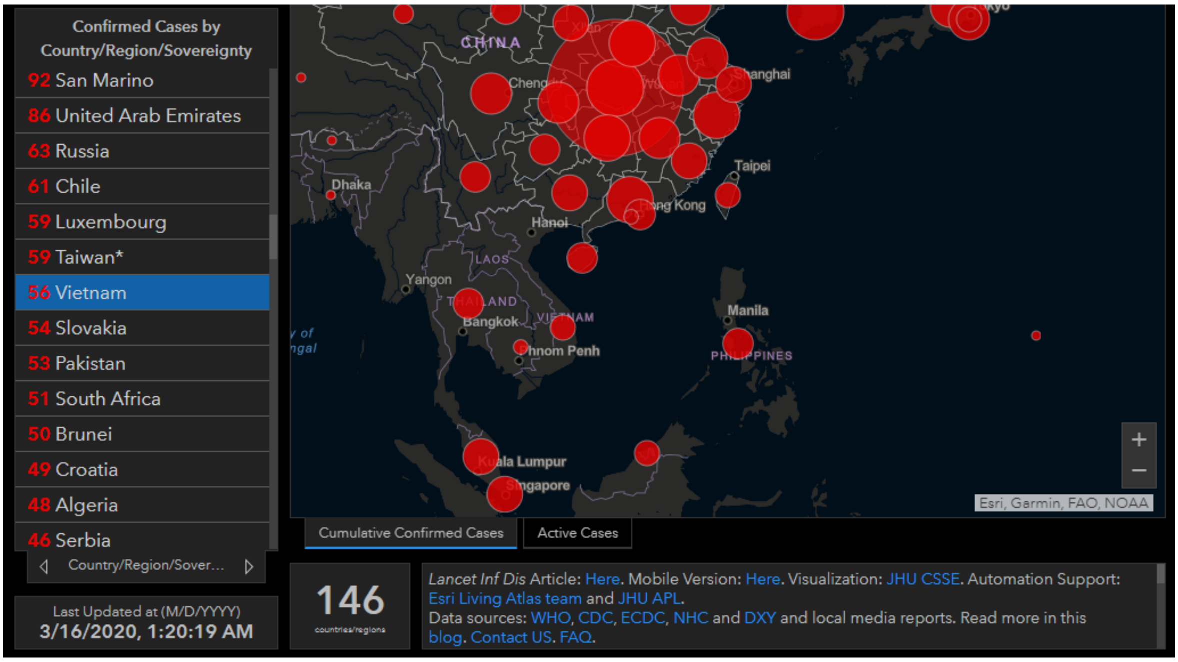


Figure 1. Confirmed COVID-19 cases in Vietnam, without detail locations, at 2 AM, 16 March 2020, GMT +7 [1]

Based on the analysis, it is found that the obtained country-level information [1, 22-25] is only useful for statistics analysis purpose. Whereas in deeper look, state- or city-level are much more useful and meaningful for the living people in the country. The key reason is that without knowing information from a city with newly "confirmed" cases, many people may pass by the city or the infectious area which can spread the virus. Thus, for the safety of the country, it is important to keep track of all "confirmed" cases and provide timely information to warn citizen for self-avoiding or preventing the spread of this deadly virus. Therefore, this work addresses the current issue of the country-level COVID-19 information in the existing disease monitoring platforms, particularly for the case of Vietnam. Below are the contributions of this research paper:

a. It does take into consideration of deeper information level, to be particular, state- or city-level in developing the "confirmed" cases map.

b. The demonstration research for Vietnam; it can be easily extended to other countries since the complete source code, demonstrated kernel [33] and sample dataset [34] were released publicly.

c. The solution (the first open toolbox for generating SARS-CoV-2 or COVID-19 map of actively confirmed cases in Vietnam) is completely free since it does not utilize any commercial cloud application programming interface (API) services from common providers such as Google, Microsoft. Thus, this enables wide use of the work even in undeveloped countries with low incomes where access to such premium information is limited.

Country, Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	Active Cases	Serious, Critical	Tot Cases/ 1M pop
Hungary	103	+18	4		7	92	6	11
Vietnam	94	+3			17	77	2	1.0
Bosnia and Herzegovina	93	+4	1	+1	2	90	1	28
Faeroe Islands	92	+12			3	89		1,883
Andorra	88	+13			1	87	2	1,139

Figure 2. 94 Confirmed COVID-19 cases in Vietnam, without detail locations, at 2 AM, 22 March 2020, GMT +7 [22]

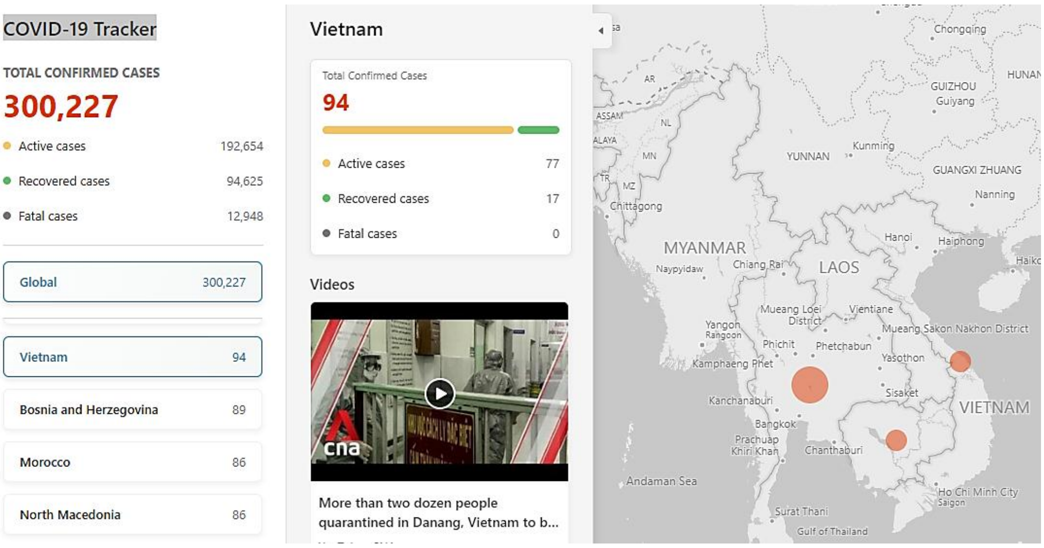


Figure 3. 94 Confirmed COVID-19 cases in Vietnam, without detail locations, at 2 AM, 22 March 2020, GMT +7 [24]

2. RESEARCH METHOD

2.1. Data preparation

In this work, in order to gereate a map of the actively "Confirmed" cases in Vietnam, the input data were manually processed from reliable sources of information including release news of Ministry of Health (MOH), Vietnam [35], local news websites such as VnExpress.net [36], DanTri.com.vn [37], ThanhNien.vn [38], and TuoiTre.vn [39]. Here, it should be noted that the news from local websites were also sourced and summarized from MOH thus, they had equivalent information publishing credential as the MOH.

Although the dataset contained various information fields, in order to generate the actively "Confirmed" cases map, the information from column "Case" (which contains patient identification, anonymous abbreviation), "Current Location" (which contains the currently district/city/state that the patients are located), "Confirmed" and "Recovered" statuses were used. When a case is "Confirmed", his status will be marked as 1, else it will be marked as 0. When a case is "Recovered", his status will be marked as 1, else it will be left unmarked. The patients who had been recovered were excluded from the current red-flagging focus locations. The "Current Location" records were searchable from an open map platform namely OpenStreetMap [40]. The data after processing were made available on Kaggle platform [34]. This opens opportunity for researchers worldwide to explore the dataset for their research on COVID-19 as well.

2.2. The developed algorithm

In this work, the algorithm is written in Python language and released publicly on Kaggle platform in [33]. In this algorithm as shown in Figure 4, three libraries are installed namely: Calmap [41] (for displaying actively "confirmed" cases map), Requests [42] (for working with HTTP, HTTPS requests and responses, obtaining data from Kaggle dataset), GeoPy [43] (for retrieving latitude, longitude information of patients' locations). The remaining esstential libraries such as Pandas are naturally provided by Kaggle.

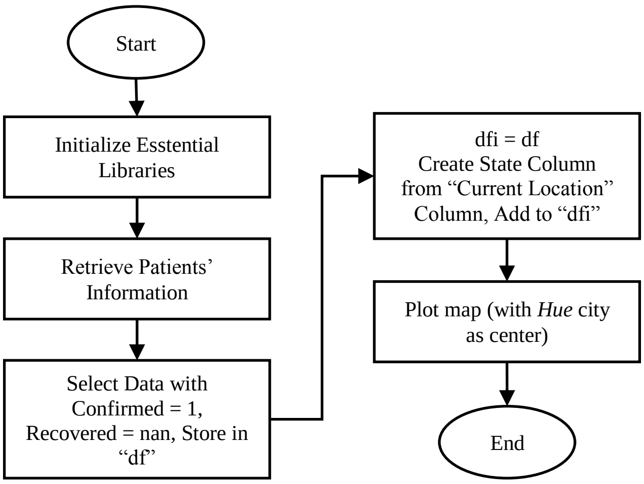


Figure 4. Algorithm's flowchart

To start the process, the algorithm first imports the essential libraries for processing data. It then retrieves patient dataset from [34] and stores in a dataframe called "**df**". The dataframe are simplified to utilize four columns described in section 2.1. For having data of only actively "Confirmed" cases, patients with state *Confirmed = 1*, and *Recovered = nan* are selected and stored in a dataframe called "**dfi**". In order to be able to display the accumulated actively "Confirmed" cases at state level, an additional column named "State" is derived from the data available in "Current Location" column. The data transition from original input to ***df*** then ***dfi*** is illustrated in Figure 5.

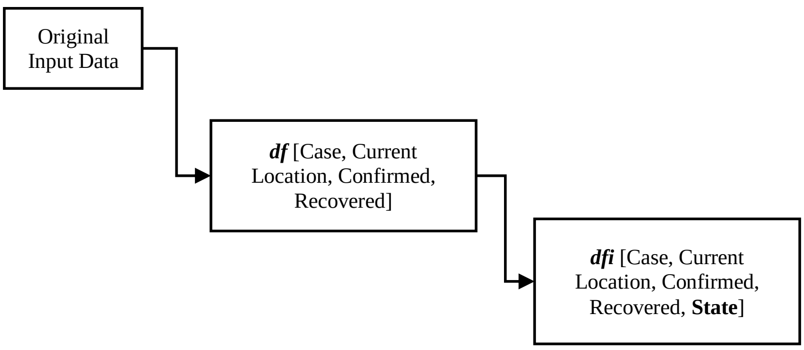


Figure 5. Dataframe transition from original input data to ***df*** then ***dfi***

Because the address stored in "Current Location", by nature, has different field lengths, sometimes, detail down to city-level. For ensuring correct state-level information in "State" column, the following code portion as shown in Figure 6 is used. Here, the current location having format of "***city, state, country***" are separated by using comma delimiter "**,**". The state information is stored in the second-to-the-last element of the resulting separation process (`state = s[len(s) - 2]`). Since there is a space character preceding to the state name, this character is stripped from the variable **state** forming the needed information to fill up "State" column.

```
for i in range(nrow):
    addr = dfState.iloc[i,0]
    if (str(addr) == 'nan'):
        print('index = ', i, ' addr = ', addr, ' -> no
address')
                                else:
        s = addr.split(',')    #delimiter = ','
        state = s[len(s) - 2]  #get province / state
        state = state.strip()
        dfState.at[i,'State'] = state
        #print('state = ', dfState.iloc[i,1])
```

Figure 6. Algorithm to retrieve "state" from "current location"

Finally, the algorithm loops through all rows in dataframe "**dfi**" and searches for the current latitude, longitude of each patient, then adds them to the plotting map. Because there shall be a location having more than one "Confirmed" cases, multiple searches of the same location is not necessary. Thus, to implement one-time search only for each location, algorithm in Figure 7 is used.

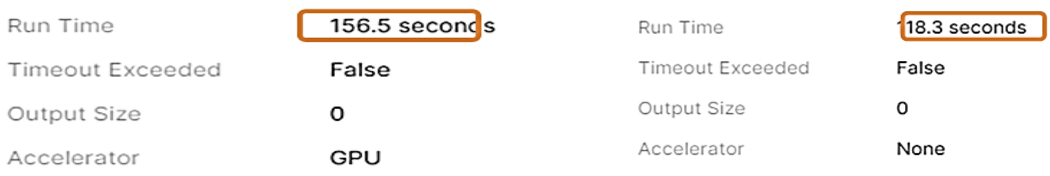
```
for i in range(0, nrow):
    time.sleep(1) #delay 1s to avoid #except OSError as err: #
timeout error
    addr = dfi.iloc[i,1]
                          print(i,"-", addr)
    if (str(addr) == 'nan'):
        print('no address')
    else:
        #search if address appear once, lat long are available
already, then no need to retrieve lat long again
        alreadyExist = 0
        foundPos = 0
        for j in range(0, i-1):
            addr2 = dfi.iloc[j,1]
            if (str(addr) == str(addr2)):
                alreadyExist = 1
                foundPos = j
                break
        if (alreadyExist):
            lat = dfi.iloc[foundPos,5]  #col Lat
            long = dfi.iloc[foundPos,6] #col Long
        else:
            try:
                lat, long = latlongGet(addr)
                print(lat, long)
            except:
                print('')

            if(lat):
                dfi.iloc[i,5] = lat         #col Lat
            if(long):
                dfi.iloc[i,6] = long        #col Long
        #print(lat, long)
```

Figure 7. Reducing multiple redundant latitude, longitude searches of "Current Location"

For each location in dataset at row i^{th}, a time sleep of 1 second is added to ensure stability of information transferred between client and the licence-free map server. If there is no address information, the algorithm prints a log indicating there is no address in the input field, else, it starts the search for latitude and longitude. Two intermediate variables called **alreadyExist** and **foundPos** are created. The former is used to indicate that the location details already exist while the latter is used to locate the position (row number) of the details stored in dataframe ***dfi***. An internal dataframe search is performed to find if the address appears in the preceding rows j^{th} ranging from **0** to ***i-1***. If the address is found, the algorithm sets **alreadyExist = 1** and **foundPos = j**, then breaks the loop, directly retrieves latitude and longitude information to fill up the i^{th} rows in the respective columns. If the address is not found, the latitude and longitude are retrieved by sending a query to the map server. By having this approach, multiple redundant queries between client and map server is greatly reduced, thus improving significantly processing time: reducing from 156.5 seconds [44] to 118.3 seconds [33] (approximately 24.41% reduction). Here it should be noted that, in order to position the country, i.e., Vietnam, into the center of the map, coordinates of Hue city, Vietnam is used as the map center. The minimum working example of the algorithm can be found in [33].

3. RESULTS AND DISCUSSION

Figure 8 presents the improvement of approximately 24.41% in processing time when plotting actively "Confirmed" cases map by reducing redundant latitude, longitude queries to map server when running the algorithm shown in Figure 7.



(a) (b)

Figure 8. Algorithm processing time reducing from (a) 156.5 seconds (version 8 of 8) [44] to (b) 118.3 seconds (version 8 of 8) [33]

In addition, Figure 9 illustrates an output of the developed toolbox which reveals an insight into the potentially infectious areas within Vietnam. The time taken to run the developed toolbox is 118.3 seconds (version 8 of 8). In this figure, one easily observes that the COVID-19 spread to major center parts of North, Middle, and South Vietnam. The states that were in between had no indication of "Confirmed" cases. Thus, people should be more cautious when travelling in locations flagged in "Red". When mousing over the highlighted location, information of actively "Confirmed" cases and "Recovered" cases are displayed in detail.

(a) (b)

Figure 9. (a) Generated map (locations) of the 77 actively "Confirmed" cases in Vietnam, as of data on 21 March 2020. Note: the 17 "Recovered" cases were not used to generate this map, (b) Detailed location of one actively "Confirmed" case in Hai Duong Province (Thanh Mien District)

4. CONCLUSION

In conclusion, this work has presented a generic toolbox for creating a map of actively “Confirmed” cases in a country, i.e., Vietnam. It is evident that the algorithm’s processing time improved significantly (approximately 24.41%) by reducing redundant latitude and longitude queries sent to the map server. Although the tested dataset is for Vietnam, the algorithm can utilize similar datasets from any other countries for plotting the COVID-19 maps across those countries. Thus, it is definitely possible for the algorithm to generate all locations of the actively “Confirmed” cases in timely manner for any country given a suitable dataset. Future work will take into consideration of applying the toolbox to other nearby countries.

ACKNOWLEDGEMENTS

The author would thank FPT University for supporting this research.

REFERENCES

[1] E. Dong, H. Du, and L. Gardner, “An interactive web-based dashboard to track COVID-19 in real time,” *he Lancet infectious diseases*, vol. 20, no. 5, pp. 533-534, Feb. 2020.
[2] Y. Liu, H. Chen, K. Tang, and Y. Guo, “Clinical manifestations and outcome of SARS-CoV-2 infection during pregnancy,” *The Journal of infection*, 2020, doi: 10.1016/j.jinf.2020.02.028.
[3] C. of A. as represented by the D. of Health, “COVID-19, Australia: Epidemiology Report 5: Reporting week ending 19:00 AEDT 29 February 2020. COVID-19 National Incident Room Surveillance Team,” *Communicable Diseases Intelligence*, 2020, doi: https://doi.org/10.33321/cdi.2020.44.20.
[4] A. Sharifi-Razavi, N. Karimi, and N. Rouhani, “COVID 19 and Intra cerebral hemorrhage: Causative or Coincidental,” *New microbes and new infections*, vol. 35, no. C, 2020.
[5] T. P. Velavan and C. G. Meyer, “The COVID- 19 epidemic,” *Tropical medicine & international health*, vol. 25, no. 3, pp. 278–280, Mar. 2020.
[6] A. Wilder-Smith, C. J. Chiew, and V. J. Lee, “Can we contain the COVID-19 outbreak with the same measures as for SARS?,” *The Lancet Infectious Diseases*, vol. 20, no. 5, pp. e102-e107, May 2020.
[7] H. A. Rothan and S. N. Byrareddy, “The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak,” *Journal of Autoimmunity*, vol. 109, 2020.
[8] J. Fausto *et al.*, “Creating a Palliative Care Inpatient Response Plan for COVID19-The UW Medicine Experience,” *Journal of Pain and Symptom Management*, vol. 60, no. 1, pp. e21-e26, July 2020.
[9] L. E. Lamb, S. N. Bartolone, E. Ward, and M. B. Chancellor, “Rapid Detection of Novel Coronavirus (COVID19) by Reverse Transcription-Loop-Mediated Isothermal Amplification,” *SSRN Electron. J.*, 2020.
[10] A. Remuzzi and G. Remuzzi, “COVID-19 and Italy: what next?,” *The Lancet*, vol. 395, no. 10231, pp. 1225-1228, 2020.
[11] E. Shim, A. Tariq, W. Choi, Y. Lee, and G. Chowell, “Transmission potential and severity of COVID-19 in South Korea,” *International Journal of Infectious Diseases*, vol. 93, pp. 339-344, April 2020.
[12] W. H. Organization, “Coronavirus disease 2019 (COVID-19) Situation Report – 85,” 2020. [Online]. Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200415-sitrep-86-covid-19.pdf?sfvrsn=c615ea20_2. [Accessed: 16-Apr-2020].
[13] M. D. Landry, L. Geddes, A. P. Moseman, J. P. Lefler, S. R. Raman, and J. van Wijchen, “Early Reflection on the Global Impact of COVID19, and Implications for Physiotherapy,” *Physiotherapy*, vol. 107, pp. A1-A3, 2020.
[14] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang, and P.-R. Hsueh, “Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges,” *International Journal of Antimicrobial Agents*, vol. 55, no. 3 , Mar. 2020.
[15] J. Wangping *et al.*, “Extended SIR Prediction of the Epidemics Trend of COVID-19 in Italy and Compared with Hunan, China,” *SSRN Electron. J.*, 2020, doi: 10.2139/ssrn.3556691.
[16] P. Kinross *et al.*, “Rapidly increasing cumulative incidence of coronavirus disease (COVID-19) in the European Union/European Economic Area and the United Kingdom, 1 January to 15 March 2020,” *Eurosurveillance*, vol. 25, no.11, pp. 1-5, 2020.
[17] K. Kingsly and K. Henri, “COVID-19 and Oil Prices,” *SSRN Electron. J.*, 2020, doi: 10.2139/ssrn.3555880.
[18] W. J. McKibbin and R. Fernando, “The Global Macroeconomic Impacts of COVID-19: Seven Scenarios,” *SSRN Electron. J.*, 2020, doi: 10.2139/ssrn.3547729.
[19] L. Lin, “Solidarity with China as it holds the global front line during COVID-19 outbreak,” *Journal of Travel Medicine*, vol. 27, no. 3, pp. 1-3, Mar. 2020.
[20] S. Semple and J. W. Cherrie, “Covid-19: Protecting Worker Health,” *Annals of Work Exposures and Health*, vol. 64, no. 5, pp. 461–464, Mar. 2020.
[21] DXY, “COVID-19 Global Pandemic Real-time Report,” 2020. [Online]. Available at: https://ncov.dxy.cn/ncovh5/view/en_pneumonia. [Accessed: 29-Mar-2020].
[22] Worldometers, “COVID-19 CORONAVIRUS PANDEMIC,” 2020. [Online]. Available at: https://www.worldometers.info/coronavirus/#countries. [Accessed: 22-Mar-2020].
[23] HERE, “Tracking Coronavirus COVID-19,” 2020. [Online]. Available at: https://app.developer.here.com/coronavirus/. [Accessed: 22-Mar-2020].
[24] Microsoft, “COVID-19 Tracker,” 2020. [Online]. Available at: https://www.bing.com/covid. [Accessed: 22-Mar-2020].

[25] U. of Washington, “Novel Coronavirus (COVID-19) Infection Map,” 2020. [Online]. Available at: https://hgis.uw.edu/virus/. [Accessed: 22-Mar-2020].
[26] U. of Virginia, “COVID-19 Surveillance Dashboard 1.1.3,” 2020. [Online]. Available at: https://nssac.bii.virginia.edu/covid-19/dashboard/. [Accessed: 22-Mar-2020].
[27] W. H. Organization, “Novel Coronavirus (COVID-19) Situation,” 2020. [Online]. Available at: https://who.maps.arcgis.com/apps/opsdashboard/index.html#/c88e37cfc43b4ed3baf977d77e4a0667. [Accessed: 22-Mar-2020].
[28] E. C. for D. P. and Control, “Situation update worldwide, as of 21 March 2020,” 2020. [Online]. Available at: https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases. [Accessed: 22-Mar-2020].
[29] K. F. FOUNDATION, “COVID-19 Coronavirus Tracker – Updated as of March 20, 2020,” 2020. [Online]. Available at: https://www.kff.org/global-health-policy/fact-sheet/coronavirus-tracker/. [Accessed: 22-Mar-2020].
[30] S. C. M. Post, “Coronavirus: the new disease Covid-19 explained,” 2020. [Online]. Available at: https://multimedia.scmp.com/infographics/news/china/article/3047038/wuhan-virus/index.html?src=article-launcher. [Accessed: 22-Mar-2020].
[31] E. StoryMaps, “Mapping the novel coronavirus outbreak,” 2020. [Online]. Available at: https://storymaps.arcgis.com/stories/4fdc0d03d3a34aa485de1fb0d2650ee0.
[32] HealthMap, “HealthMap | Novel Coronavirus (COVID-19),” 2020. [Online]. Available at: https://www.healthmap.org/covid-19/. [Accessed: 29-Mar-2020].
[33] D. C. Tran, “Vietnam SARS-CoV-2 | COVID-19 Infographics v2,” 2020. [Online]. Available at: https://www.kaggle.com/chungtranduc89/vietnam-sars-cov-2-covid-19-infographics-v2.
[34] D. C. Tran, “Vietnam SARS-CoV-2 | COVID-19 Data,” 2020. Kaggle, doi: 10.34740/KAGGLE/DSV/1098604. [Online]. Available at: https://www.kaggle.com/chungtranduc89/vietnam-sarscov2-covid19-data.
[35] M. of Health, “Ministry of Health Portal,” 2020. [Online]. Available at: https://moh.gov.vn/.
[36] VnExpress, “VnExpress,” 2020. [Online]. Available at: https://vnexpress.net/.
[37] DanTri, “DanTri,” 2020. [Online]. Available at: https://dantri.com.vn/.
[38] ThanhNien.vn, “ThanhNien.vn,” 2020. [Online]. Available at: https://thanhnien.vn/.
[39] TuoiTre.vn, “TuoiTre.vn,” 2020. [Online]. Available at: https://tuoitre.vn/.
[40] OpenStreetMap, “OpenStreetMap,” 2020. [Online]. Available at: https://www.openstreetmap.org.
[41] T. Marvin and V. Martijn, “Calmap,” 2020. [Online]. Available at: https://pypi.org/project/calmap/.
[42] K. Reitz, “requests,” 2020. [Online]. Available at: https://pypi.org/project/requests/.
[43] G. Contributors, “GeoPy,” 2020. [Online]. Available at: https://geopy.readthedocs.io/en/stable/.
[44] D. C. Tran, “Vietnam SARS-CoV-2 | COVID-19 Infographics,” 2020. [Online]. Available at: https://www.kaggle.com/chungtranduc89/vietnam-sars-cov-2-covid-19-infographics. [Accessed: 16-Apr-2020].

BIOGRAPHY OF AUTHOR

Tran Duc Chung completed his bachelor degree in Electrical & Electronic Engineering with major in Instrumentation & Control from PETRONAS University of Technology, Malaysia. In 2018, at the same university, he completed his Ph.D. degree in the area of WirelessHART networked control system. He had been working for several companies namely Intel Technology Sdn. Bhd. Malaysia, R&D Centre-Sony EMCS Sdn. Bhd. Malaysia, Dasan Zhone Solutions Vietnam, FPT Technology Research Institute, Hanoi, Vietnam. Currently he is with FPT University, Hanoi, Vietnam, researching in emerging technologies and topics including but not limited to Natural Language Processing and Generation (NLP, NLG), and Applied Artificial Intelligence.