International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
IJIREV
Volume 7 Issue 20 (March 2025) PP. 259-272
DOI 10.35631/IJIREV.720016

INTERNATIONAL JOURNAL OF INNOVATION AND INDUSTRIAL REVOLUTION (IJIREV)
www.ijirev.com

FROM OUTBREAKS TO ENDEMIC: ANALYZING THE EVOLUTION OF COVID-19 CLUSTERS IN MALAYSIA

Zanariah Idrus[1], Nurul Husna Mahadzir[2*], Ahmad Afif Ahmarofi [3], Daniyal Rosli[4], Hariz Zamzuri[5], Muhammad Al Hanis Ahmad Nazri[6], Muhammad Luqman Mohd Asni [7]

[1] Integrated Simulation & Visualization Research Interest Group Simulation, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Kedah
Email: zanaidrus@uitm.edu.my
[2] Integrated Simulation & Visualization Research Interest Group Simulation, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Kedah
Email: husnamahadzir@uitm.edu.my
[3] Integrated Simulation & Visualization Research Interest Group Simulation, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Kedah
Email: ahmadafif@uitm.edu.my
[4] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Kedah
Email: daniyalrosli@gmail.com
[5] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Kedah
Email: harizzamzuri20@gmail.com
[6] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Kedah
Email: alhanis1605@gmail.com
[7] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Kedah
Email: luqmanasni@gmail.com
[*] Corresponding Author

Article Info:

Article history:
Received date: 14.01.2025
Revised date: 23.01.2025
Accepted date: 17.02.2025
Published date: 20.03.2025

To cite this document:

Idrus, Z., Mahadzir, N. H., Ahmadrofi, A. A., Rosli, D., Zamzuri, H., Nazri, M. A. H. A., & Asni, M. L. M. (2025). From Outbreaks To Endemic: Analyzing The Evolution Of Covid-19 Clusters In Malaysia.

Abstract:

From 2020 to 2023, Malaysia experienced multiple waves of COVID-19, with clusters playing a significant role in transmission dynamics. Understanding these clusters is crucial for public health, as their characteristics and management significantly impact outbreak control. Unfortunately, existing regional research often focuses on national-level data, neglecting the insights hidden within cluster-level analysis. In these regards, this research aims to bridge this gap by comprehensively analyzing COVID-19 clusters in Malaysia from 2020 to 2023 through the development of a data dashboard using Microsoft Power BI. To fulfil that, the high-risk areas, the trend of COVID-19 clusters, the active time of each cluster, and the total cases of COVID-19 in each cluster have been identified. As a result, the developed dashboard reveals that the community cluster has the highest death toll of over 500 people while the highest number of COVID-19 infection cases has been recorded by the workplace group (over 300,000 cases) followed by the community group

International Journal of Innovation and Industrial Revolution EISSN: 2637-0972 IJIREV
Volume 7 Issue 20 (March 2025) PP. 259-272
DOI 10.35631/IJIREV.720016

International Journal of Innovation and Industrial Revolution, 7 (20), 259-272.

DOI: 10.35631/IJIREV.720016

(100,000 cases). It can be induced that community clusters often see higher mortality rates despite the lower number of cases because they often affect more vulnerable populations such as the elderly compared to workplace clusters which usually involve younger working-age individuals. This knowledge will empower policymakers, healthcare professionals, and local communities to tailor their efforts to mitigate future waves and reduce the burden of COVID-19 in Malaysia. Consequently, by bridging the gap in research and focusing on the granular level of clusters, the project aspires to pave the way for a more data-driven and localized approach to pandemic management in Malaysia and across the globe.

Keywords:

COVID-19 Clusters, Pandemic, Prediction, Machine Learning, Data Visualization

Introduction

The outbreak of the COVID-19 pandemic in late 2019 indicates a significant shift in the trajectory of world history. Large-scale changes were felt worldwide, forcing countries to deal with unexpected difficulties on several fronts. Healthcare institutions were forced to engage in an unexpected struggle to fulfil the rising demands for resources and medical services. Economies were torn apart by lockdowns and limitations that required quick and significant adjustments (Llc & Segall, 2023). Masks, social distance, and remote interactions have become a normal daily practice. Ready for adaptation and resiliency, Malaysia has led the way in this international conflict. Similar to other countries worldwide, Malaysia faced significant challenges in its efforts to control and lessen the extensive effects of this extremely infectious virus. Healthcare personnel worked nonstop to treat the ill and carry out life-saving procedures, challenging the country's healthcare system (Awang et al., 2022). Creating and implementing policies meant to protect public health while causing the least damage to society.

In many countries, machine learning was used to predict infectious disease outbreaks such as influenza, insect bite illness, and food contamination (Kamarudin et al., 2021). A wide range of approaches were used in this quickly changing state, each was carefully designed to handle certain aspects of the problem. Among these, managing and identifying clusters were crucial to the group's attempt to decelerate the virus's spread. Authorities quickly and effectively manage epidemics by identifying isolated illnesses and implementing focused actions. In addition to protecting vulnerable areas, this calculated strategy was crucial in preventing COVID-19 from spreading widely and eventually, it prevented casualties and reduced the burden on the nation's healthcare system (Vaishya et al., 2020). Accurate prediction and proactively avoiding possible infections remains a major issue, even with concentrated efforts to identify high-risk areas for COVID-19 cluster growth. The current approaches depend on epidemiological models and historical data, thus, lacking in recognizing new trends or unforeseen elements influencing the formation of clusters. This limitation presents a significant risk since it may result in irregular outbreaks requiring resource-intensive control measures. Moreover, variations in the socioeconomic conditions and healthcare accessibility among various places may make it more challenging to identify and successfully manage these high-risk areas (Vahedi et al., 2021). Due to this, a more flexible and dynamic strategy is needed to improve the accuracy of locating and addressing possible cluster hotspots (Razzak et al., 2020). Therefore, these research objectives are to identify the high-risk areas of COVID-19 cluster

International Journal of Innovation and Industrial Revolution IJIREV
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 259-272
DOI 10.35631/IJIREV.720016
formation and prevent potential infection, analyze the trend of COVID-19 clusters and the active time of each cluster, and monitor the total cases of COVID-19 in each cluster.

The dataset was obtained from the Kaggle website and in the analysis, the targeted subset of the data is employed to address specific key objectives. The primary focus of the study is to track the trends in COVID-19 cases and assess the impact of vaccination campaigns in Malaysia. The 'district' attribute within the dataset will play a pivotal role in this analysis, allowing for a detailed examination of case trends in specific regions, including Tawau, Titiwangsa, and Kulim. By concentrating on relevant attributes and districts, the research aims to offer meaningful insights into the dynamics of the pandemic, exploring potential correlations between vaccination rates and the number of reported cases. This study explains how vaccination efforts influence the overall trajectory of COVID-19 situations in different districts across Malaysia presenting using charts and data visualizations.

Public health authorities, government officials, and medical professionals involved in COVID-19 response tasks are among the users' targets. The results and suggestions of this study are useful to epidemiologists, researchers, and analysts who focus on infectious diseases and cluster investigations. Moreover, the insights presented could prove advantageous to public health initiative-focused organizations and local community leaders, especially in areas designated as high-risk for the emergence of COVID-19 clusters. In summary, Malaysia's reaction to the COVID-19 epidemic is evidence of the country’s determination, creativity, and united dedication to public health. By carefully putting methods into practice, such as identifying and managing clusters, Malaysia showed that it could adapt to previously unheard-of difficulties while preserving the wealth and well-being of its people.

Literature Review

COVID-19 Pandemic Outbreak

The COVID-19 pandemic has substantially affected human health and socio-economic, thus shifting the emergence of prediction and visualization in healthcare. On 11 March 2020, The World Health Organization (WHO) announced COVID-19 as a pandemic, and by 31 December 2021, COVID-19 had spread over to 237 countries causing over 286 million confirmed cases and approximately 5.43 million deaths. Meanwhile, in ASEAN countries, the most infected with COVID-19 are Indonesia, Philippines, Malaysia, and Singapore, with Indonesia having 1,718,575 cases and 47 218 deaths while the least country is Brunei with 330 cases and 3 deaths (Purnomo et al., 2022).

The ecological epidemiological studies assumed that disease transmission had become extensively widespread in the community and was caused by the mixing patterns of individuals within the environment relevant to the growth of populations, human mobility, and interactions (Ganasegeran et al., 2021). Proactive measures were taken to prevent the COVID-19 virus outbreak including enlarging quarantines, promoting social distancing, enhancing healthcare infrastructure, and permitting only necessary goods to leave the home (Sayeed et al., 2024). ASEAN countries have taken several steps to manage the spread of COVID-19 such as national lockdowns among all ASEAN member countries, local city lockdowns, rapid tests, and social distancing (Arnakim & Kibtiah, 2021;Tantrakarnapa et al., 2022).

International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 259-272
DOI 10.35631/IJIREV.720016
The COVID-19 pandemic represents one of the most substantial global public health crises in history. The outbreak revealed the weakness in healthcare systems, particularly in resource-limited countries, while highlighting the importance of coordinated global responses (Khorram-Manesh et al., 2024). To mitigate the spread, governments executed various strategies, including lockdowns, travel restrictions, and mass vaccination campaigns. The pandemic underlined the critical role of science, data, and international in collaborating with public health while highlighting the need for improvement and readiness to tackle future pandemics.

COVID-19 in Malaysia

In Malaysia, the first COVID-19 case was detected on 25 January 2020 and was officially announced as an outbreak in March 2020 (Jayaraj et al., 2023). Each country has a different cluster classification and categories requiring different regulation methods. Malaysia manages the outbreaks by clustering patients according to seven clusters which are the community, workplace, high-risk, religious, detention center, education, and import clusters. These cluster categorization and management strategies were introduced to leverage the COVID-19 pandemic (Cheong et al., 2022).

The Community Clusters are clustered in community activities such as at home, at large communal residences, and during festivals, funerals, receptions, and weddings (Ang et al., 2023). Focusing on community transmission was significant, especially during the initial waves of the pandemic. Contact tracing is one of the most important tools for successfully tackling community transmission. Another method is the Movement Control Order (MCO), which reduces the reproduction rate R0 from 3.6 to 0.3 (Azhar et al., 2020). Aggressive control measures, such as rapid tests for targeted people and a social distancing policy, were crucial and successful in controlling community clusters (Azit et al., 2022;Arnakim & Kibtiah, 2021) Workplace Clusters have high transmission rates, therefore, industrial areas prompted mass testing and geospatial mapping using tools like ArcGIS to identify hotspots and guide containment measures. For instance, geospatial animation maps were used to monitor cases in Selangor's industrial areas (Talib et al., 2023). High-risk clusters are elderly populations and individuals with comorbidities. Strict monitoring in nursing homes and hospitals minimized outbreaks. Religious clusters are patients infected during religious activities. Major outbreaks originated from large religious gatherings, such as the religious event in Sri Petaling, contributing to 47% of early cases. Religious leaders collaborated with the Ministry of Health (MOH) to promote compliance with public health measures in controlling the spread (Tan et al., 2021). Crowded detention centers posed significant challenges. To manage outbreaks, enhanced hygiene protocols and mass screenings were implemented (Azit et al., 2022).

Education clusters are schools and universities with multiple clusters. Temporary closures and hybrid learning models were introduced to limit exposure (Mustapha et al., 2024). Import clusters are travellers entering Malaysia subjected to strict quarantine protocols and testing to prevent new variants from entering the community (Ang et al., 2023).

The COVID-19 pandemic in Malaysia highlighted the critical role of clusters in driving the spread of the virus. These clusters, categorized into community, workplace, religious, custodial settings, and high-risk groups, served as focal points for targeted public health interventions. Malaysia's proactive approach, including Movement Control Orders (MCOs), targeted screening, and rigorous contact tracing, effectively curtailed the spread of the virus.

International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
IJIREV
Volume 7 Issue 20 (March 2025) PP. 259-272
DOI 10.35631/IJIREV.720016
Community clusters, often linked to familial and social gatherings, underlined the necessity of public compliance with Standard Operating Procedures (SOPs). Similarly, the management of workplace and custodial clusters emphasized the need for strict measures in high-contact environments, such as factories and detention centers. Socio-environmental remains as critical factor in alleviating future outbreaks. Thus, policies such as national lockdowns, local-city lockdowns, restrictions on community activities, quarantines, social distancing, rapid tests, and vaccinations taken in Malaysia and other countries such as Thailand and Indonesia have effectively reduced the spreading of COVID-19.

Prediction Using Data Visualization
The role of new technologies in managing a pandemic situation is unquestionable in particular forecasting and prediction. Machine learning has been proven to be one of the effective methods to predict possible disease patterns including COVID-19 (Sayeed et al., 2024). In similar work, predictive analytics increase human knowledge of identifying the disease-transmission pattern. By using observational data, it establishes forecasting of future occurrences and provides awareness of the pandemic (Gawande et al., 2025). A machine-learning model is suggested to predict COVID-19 cases and intensive care unit (ICU) requirements (Podder & Mondal, 2020).

Data visualization has become one of the vital digital data representations, which includes statistical tools, cartographic tools, and dashboards that combine multiple analyses on a single screen. It simplified complex datasets and the COVID-19 pandemic has prompted the improvement of interactive and enhancement tools in data visualization allowing better presentable data (Grandi & Bernasconi, 2021). Dashboards have gained popularity as a method of exploring and monitoring data, facilitating simple outlines of several charts for users (Shin et al., 2024). Interactive dashboards in healthcare support the visualization of complex datasets and presenting data in an accessible and interpretable format, dashboards assist healthcare professionals (Sedek & Annuar, 2024).

The Johns Hopkins University combines data on COVID-19, real-time infection rates, deaths, and recoveries into a Forecasting Global Data Aggregation Dashboard. Its visualization tools produce trend predicting and identify hotspot locations (Dong et al., 2020). In China, dashboards combined with Artificial Intelligence (AI) models and case data have accurately forecasted infection trends. Successfully, a hybrid AI model, integrating long short-term memory (LSTM) and Natural Language Processing (NLP), predicted infection rates with over 90% accuracy. Furthermore, the predicted results are in line with the actual epidemic development trend. It proves that transparency, and efficiency in releasing data are significant for determining a modernized epidemic prevention system. (Zheng et al., 2020). In the United States of America, Google Trends data was integrated into dashboards for real-time predicting cases in Washington, DC. Predicted trends significantly correlated with actual cases (Wang et al., 2023). In another research, Google Trends data, combined with the historical time series produced case predictions without using complicated mathematical modeling and reduced the algorithm complexity (Pan et al., 2020).

Overall, the evidence presented in this section suggests that predicting, especially in health authorities, can reduce the spread of disease and provide early prevention action. Modern technology has contributed to improving humans' lives and using AI is a significant advantage in fighting COVID-19. AI uses real-time data analysis that can provide updated information

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 259-272
DOI 10.35631/IJIREV.720016

and be helpful in the prevention of COVID-19. In addition, prediction enables the identification of the probable sites of infection, the influx of the virus, and the need for beds, and healthcare professionals during this crisis. AI is helpful for future virus and disease prevention, with the help of previous mentored data over data established at different times. It identifies traits, causes, and reasons for the spread of infection. In the future, this will become an important technology and provide a preventive measure to fight against other epidemics and pandemics. Data visualization such as a dashboard and Google Trends contributes valuable insight and assists decision-makers and healthcare professionals in decision-making. Although all countries have entered the endemic phase, initiatives must be driven to support government and health authorities in preparing for any future epidemic. In the future, AI will play a vital role in providing more predictive and preventive healthcare.

Methodology

This section presents the methodology used in this research and is separated into four phases which are Data Collection, Data Preprocessing, Exploratory Data Analysis, and Data Visualization as depicted in Figure 1.

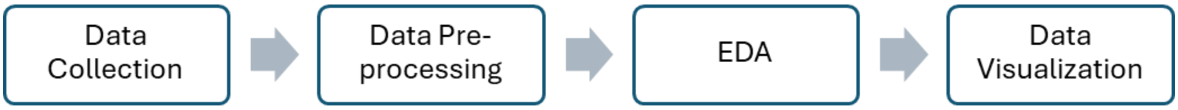


Figure 1: Research Activities

Data Collection

The dataset was collected from https://www.kaggle.com/datasets/mcpenguin/malaysia-covid19 and encompasses a comprehensive set of information, comprising 14 attributes and 70,000 records. It contains detailed information on COVID-19 in Malaysia, including the number of cases in different states and vaccinations. It gives a clear picture of how the virus is spreading across the country and provides insights into the progress of the vaccination campaign, including the distribution of vaccines and coverage across different age groups.

Data Preprocessing

Data preprocessing for the COVID-19 dataset involves several crucial steps to ensure the dataset is clean, organized, and ready for data analysis. In the initial phase of data cleaning, attention is directed towards handling missing values in key columns including New Cases, Total Cases, Active Cases, Tests, ICU, Deaths, and Recovered. The strategy involves detailed identification and subsequent resolution of missing data, through the imputation methods and the removal of incomplete records. The activities include replacing the missing values with the mean, median, or mode, depending on the distribution of the data. Additionally, rows and columns with excessive missing values have been removed to avoid bias.

Concurrently, a thorough outlier detection process has been executed on numeric columns such as New Cases, Total Cases, Active Cases, Tests, ICU, Deaths, and Recovered. Decisions on handling outliers are made, considering approaches like capping or further investigation to maintain data integrity. Outliers in numeric columns are detected using statistical methods, such as identifying values beyond 1.5 times the interquartile range (IQR) or outside ±3 standard deviations from the mean, as well as visualization tools like boxplots and histograms. Domain knowledge is also applied to recognize data points that deviate significantly from expected ranges. Once detected, outliers are handled through various strategies, including capping

International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 259-272
DOI 10.35631/IJIREV.720016

extreme values at set thresholds and investigating the cause of anomalies to determine their validity. These steps cooperatively ensure that the dataset is robust and reliable, minimizing bias while preserving data integrity for further analysis.

Exploratory Data Analysis (EDA)

EDA commences with the calculation and review of descriptive statistics for numeric columns encompassing New Cases, Total Cases, Active Cases, Tests, ICU, Deaths, and Recovered. This statistical overview sets the stage for comprehensive insights. Concurrently, data visualization techniques, such as line charts, bar graphs, or heatmaps, are deployed to explore trends, patterns, and potential correlations within the COVID-19 data and vaccination information.

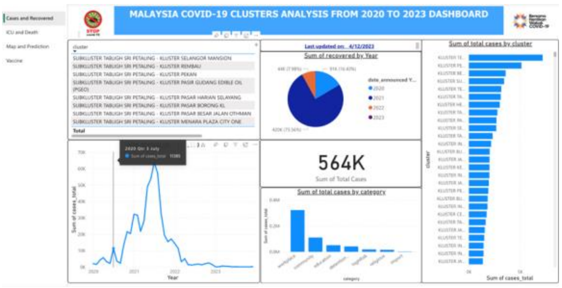
Data Visualization

This phase provides readers with relevant information; and data visualization, which entails summarizing and displaying enormous amounts of data in clear, concise visualizations. In this project, a variety of network charts were used to demonstrate collaboration ties between each attribute to uncover the narrative, which is hidden behind the data. Microsoft Power BI was selected because of its interactive and compelling features in analyzing data and customizing visualization.

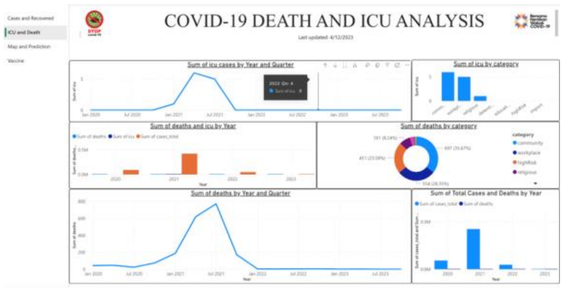
Results and Discussion

Dashboard Design

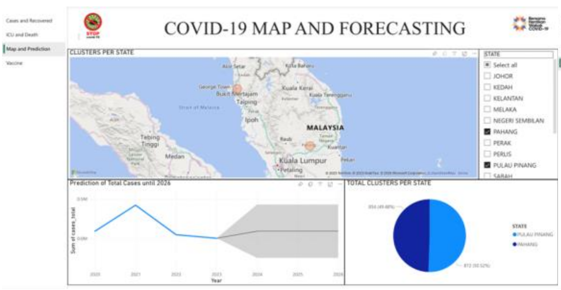
The developed dashboard consists of four pages as illustrated in Figure 2 below.



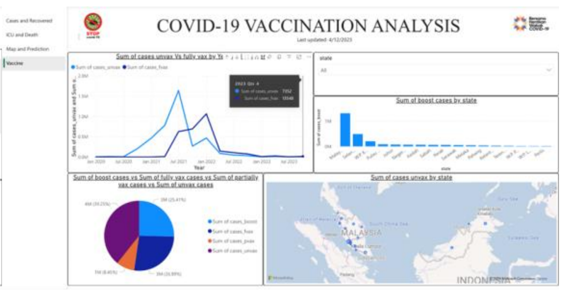
Cases and Recovered



ICU and Death



Map and Forecasting



Vaccination Analysis

Figure 2: Dashboard Design

This dashboard provides an in-depth analysis of COVID-19 in Malaysia, likely spanning from 2020 to 2023, and is divided into four sections, each highlighting different aspects of the pandemic. The first section, "Cases and Recovered," presents trends and statistics on COVID-

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 259-272
DOI 10.35631/IJIREV.720016

19 cases and recoveries through visuals such as a line chart depicting daily or cumulative case trends, pie charts summarizing percentages (potentially by state or category), and a bar chart listing clusters or areas with the highest cases. This section aims to give an overview of the pandemic's progression and recovery trends.

The second section, "ICU and Death," focuses on severe cases, ICU admissions, and fatalities. It uses line charts to illustrate ICU and death trends, complemented by pie or bar charts showing distributions, such as by age group or region, offering insights into the critical impact of the pandemic on healthcare systems and mortality.

The third section, "Map and Forecasting," examines the geographic spread and future projections of COVID-19 cases. It includes a map of Malaysia highlighting affected regions, a line chart comparing actual versus forecasted cases, and pie charts potentially showing demographic or regional breakdowns. This section is designed to monitor spatial patterns and predict future case trends.

Finally, the "Vaccination Analysis" section analyzes vaccination efforts and their outcomes, with visuals such as line charts tracking vaccination progress over time, bar or pie charts detailing vaccine distribution or coverage by category or region, and a map illustrating vaccination rates across Malaysia. This section evaluates the effectiveness and reach of vaccination campaigns. Together, these sections provide a comprehensive view of the pandemic's progression, its impact on public health, and mitigation efforts.

Trend of COVID-19 Cases in Malaysia

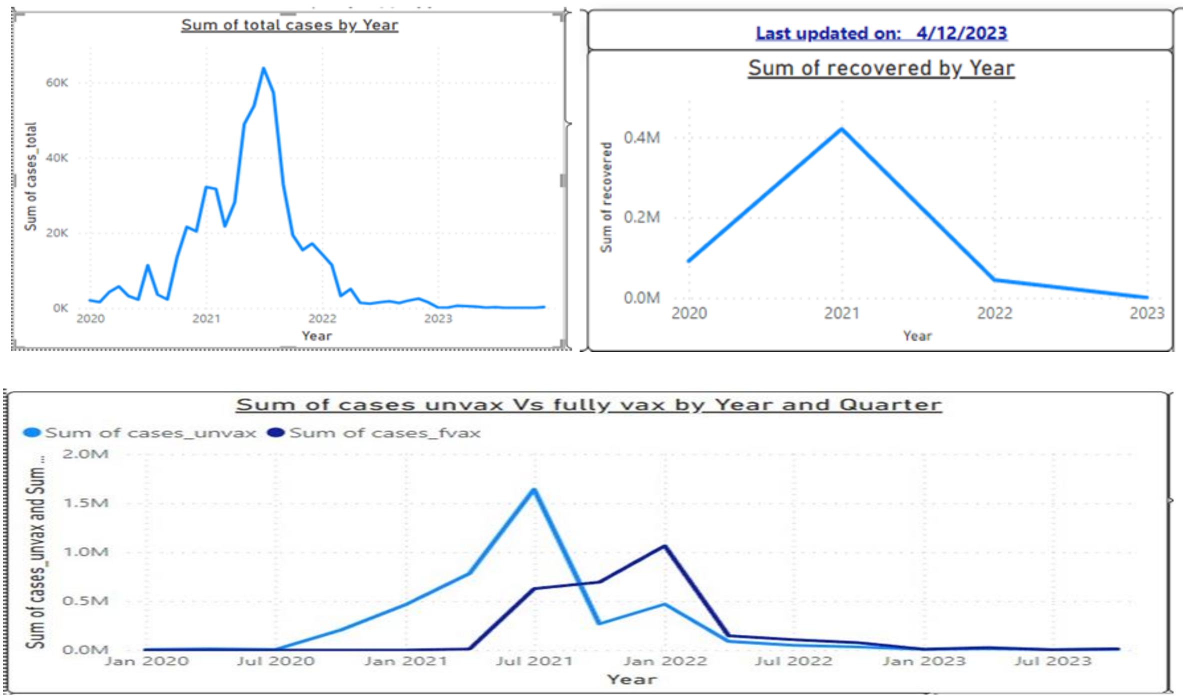


Figure 3: COVID-19 Trends

Figure 3 displays three distinct line charts illustrating the total number of COVID-19 cases over four years, from 2020 to 2023, in Malaysia. The top-left chart outlines the total number of COVID-19 cases by year, while the chart on the right depicts the total number of recovered patients per year. From the visual representation, it is evident that the number of COVID-19

International Journal of
Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 259-272
DOI 10.35631/IJIREV.720016

cases started rising in mid-2020, reaching its peak around mid-2021. Subsequently, there was a gradual decline, with a significant drop observed in 2023 compared to the peak in 2021. The bottom chart focuses on the number of cases among vaccinated and unvaccinated patients. This trend is influenced by several external factors, including policy changes, economic conditions, and societal behaviours. Policy interventions such as vaccination campaigns played a significant role, with the decline in cases among the unvaccinated population after 2021 reflecting the widespread rollout of vaccines. Lockdown measures and social distancing efforts likely contributed to controlling transmission during certain periods, while relaxation of these measures led to spikes in cases, as seen in 2021. Changes in testing policies also affected the trends, with increased testing during surges capturing more cases and reduced testing during low transmission periods potentially underreporting cases. Additional measures like mask mandates and travel restrictions further helped mitigate the spread of the virus.

Economic factors also influenced the trends, as pressures to reopen economies and increase mobility, particularly in job-related activities, contributed to spikes in transmission. Economic gaps, such as unequal access to healthcare, may have impacted the spread of cases, especially among unvaccinated populations in low-income regions. Behavioral and social factors, including vaccine hesitancy and pandemic fatigue, also played a role. By mid-2021, reduced compliance with preventive measures, coupled with increased public gatherings for seasonal or cultural events, likely contributed to localized surges in cases. Furthermore, the emergence of more transmissible variants like Delta and Omicron fuelled significant increases in transmission, although vaccination efforts mitigated severe outcomes, particularly among the fully vaccinated population.

Overall, these trends highlight the interplay between policy decisions, economic pressures, social behavior, and the biological evolution of the virus. The eventual decline in cases reflects the effectiveness of vaccination campaigns and public health measures in managing the pandemic over time.

Covid 19 Clusters

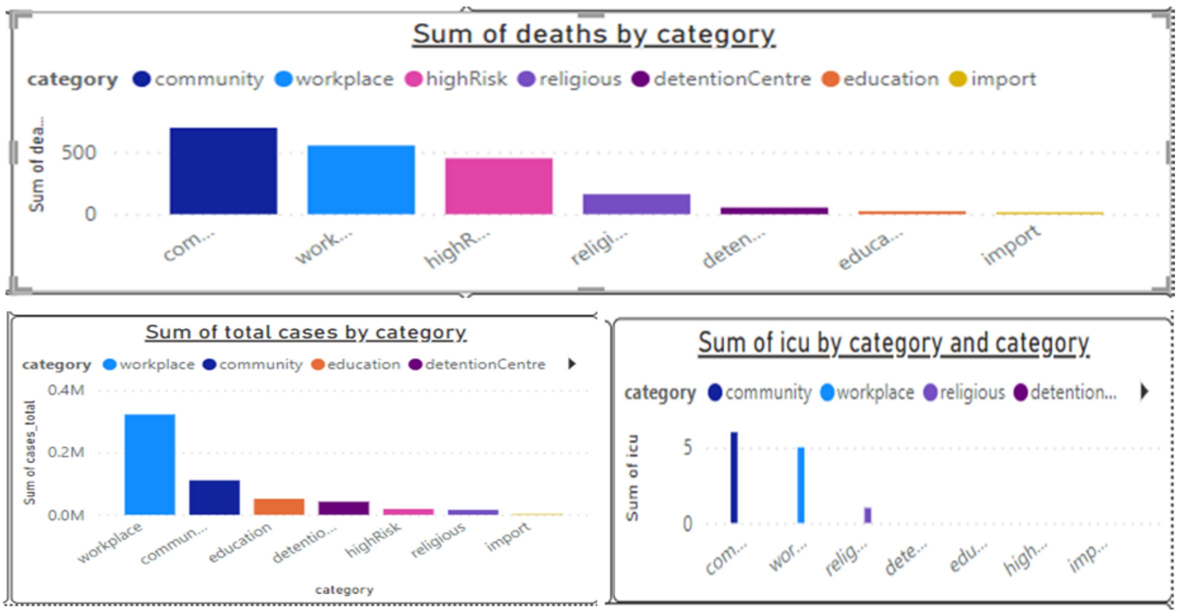


Figure 4: Covid 19 Clusters

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 259-272
DOI 10.35631/IJIREV.720016

Figure 4 provides an overview of the categories to which all COVID-19 clusters in Malaysia belong. These categories include community, workplace, high-risk, religious, detention center, education, and import clusters. The bottom left chart indicates that workplace clusters have the highest number of cases, followed by community clusters, education clusters, and the import cluster, which has the fewest cases. The initial surge in workplace cases is attributed to the lack of early warnings when COVID-19 first hit Malaysia. The Malaysian government's restriction on travel has kept import cluster cases relatively low.

Contrastingly, the top charts reveal that community clusters have the highest death count, followed by workplace clusters. This trend is mirrored in the bottom right chart, which displays the category with the most cases of ICU-warded patients. The inference drawn is that clusters in the community category pose greater challenges, likely due to a larger spread, making them harder to contain compared to workplace clusters. Consequently, this leads to reduced death and ICU cases in the workplace category.

The COVID-19 clusters in Malaysia exhibit unique characteristics compared to global trends, particularly in the distribution of cases, deaths, and ICU usage across various categories such as workplaces, communities, religious gatherings, detention centers, and high-risk groups. Workplace clusters in Malaysia accounted for the largest share of total cases, reflecting the country's industrial and manufacturing-based economy, where densely populated settings like factories and construction sites facilitated transmission. This dominance of workplace clusters stands out compared to global trends, where community spread typically plays a larger role. Community clusters in Malaysia, however, showed a disproportionately high contribution to deaths, suggesting that while they were not the primary driver of total cases, they significantly impacted mortality. This may indicate challenges in reaching vulnerable populations or delays in healthcare access within these clusters.

Prediction of COVID-19 Cases in Malaysia

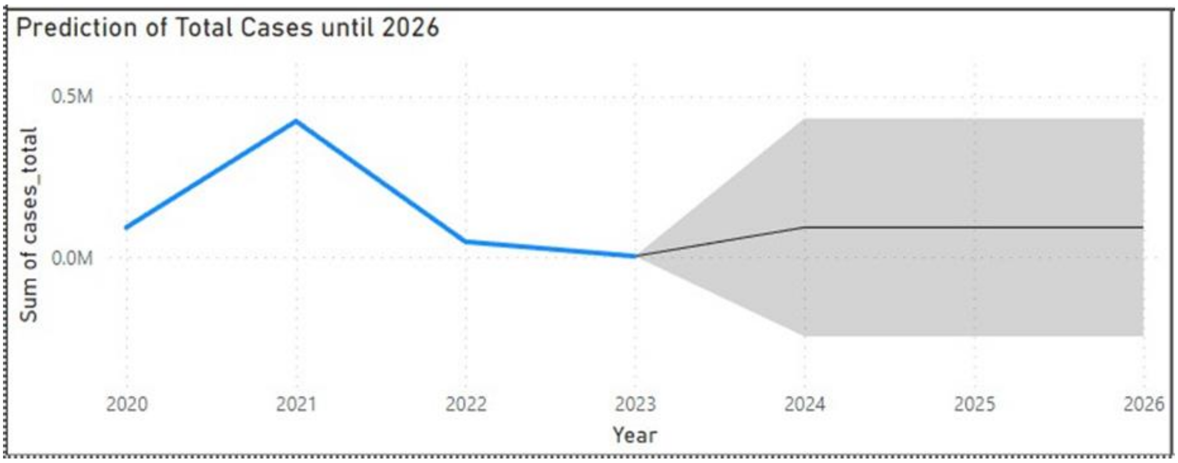


Figure 5: Prediction of Covid 19 Cases

In Figure 5 above, the line map shows the total number of cases in Malaysia from 2020 until 2023 with added features of prediction using Microsoft Power BI feature to predict the total number of cases up until 2026. The data for the COVID-19 cases in Malaysia that were used are not complete until the end of 2023 thus may cause miscalculation by the prediction

International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 259-272
DOI 10.35631/IJIREV.720016
algorithm. Using our data which is updated until the 4th of December 2023, the number of COVID-19 cases in Malaysia is 2369 for the year 2023.

The forecast feature use is set up with a 95% confidence interval which was based on a previous study. (Alabdulrazzaq et al., 2020). The forecast feature predicted that in 2024, the number of cases throughout the year will sum up to 92022 cases. The same count is predicted from the year 2025 until 2026. The reason for our project to only predict up until 2026 is due to the constant number of total predicted cases for upcoming years starting from 2024 which is 92022. The predicted number of cases for 2024 compared to the current cases in 2023, shows a difference by a large margin. Although it may seem impossible for cases to increase by that much in 2024 compared to the relatively low number of cases in 2023, history however begs to differ.

In post-COVID-19 conditions among adults in Malaysia following the Omicron wave, more than 95% of the adult population in Malaysia were fully vaccinated against COVID-19 by the end of 2021 like many other countries globally, Malaysia was not spared by the Omicron variant onslaught which led to the increase in cases during 2022 (Keng Tok et al., 2024). This showed that the Coronavirus is a constantly mutating virus that may penetrate through our vaccine-boosted antibody by each variant, thus always posing a constant threat to the health of Malaysians and can be the leading cause of the increase of COVID-19 cases in Malaysia in 2024 and years after that.

In addition, the prediction model for COVID-19 cases in Malaysia, as shown in the chart, presents certain limitations that should be considered when interpreting the results. One significant limitation is the potential for incomplete or insufficient data for the year 2023. If the data for 2023 is partial, it could skew the model's forecast for future years, as the trends observed in 2023 might not fully reflect the actual situation. Incomplete data can lead to inaccuracies in the model's ability to capture seasonal patterns, changes in transmission dynamics, or the impact of new interventions, such as updated vaccination campaigns or public health policies.

Another limitation is the uncertainty inherent in prediction models, especially for a long-term period up to 2026. External factors such as the emergence of new variants, changes in population behaviour, government policy shifts, or unforeseen global health crises can significantly alter the trajectory of cases, making predictions less reliable. Additionally, if the model relies heavily on historical trends without accounting for recent changes, such as improvements in healthcare infrastructure or vaccination rates, it may fail to adapt to evolving circumstances.

Conclusion

The research that was done has shown to be beneficial in terms of offering a deeper understanding of the COVID-19 situation in Malaysia. This research identifies high-risk areas where COVID-19 clusters emerge through detailed observation and analysis. This information is essential because it makes it possible to manage and reduce the virus's transmission in a more proactive approach. One key aspect highlighted in the research is the ability to analyze the trend of COVID-19 clusters. By understanding how these clusters evolve, authorities and healthcare sectors gain valuable information to make informed decisions. This could involve implementing specific measures in response to emerging patterns or identifying periods of heightened risk, enabling a more strategic allocation of resources. The data generated by the research is beneficial for immediate decision-making and offers a foundation for continuous monitoring. Authorities can keep track of the total number of COVID-19 cases within each cluster, providing a real-time and accurate picture of the situation. This allows for the adaptation of strategies as the situation evolves, ensuring a dynamic and effective response to the ongoing challenges posed by the pandemic in Malaysia. Overall, the research contributes significantly to the collective efforts of authorities and healthcare sectors in managing and combating the COVID-19 pandemic in Malaysia. The development of the data dashboards in this research provides almost real-time visualization, thus helping decision-makers assess certain scenarios and operational efficiency. Furthermore, it is suggested that the Malaysian government should disseminate transparent data through dashboards between public health agencies and healthcare facilities to the public. Consequently, disease surveillance systems can be strengthened and able to detect new variants and outbreaks quickly in the future.

Acknowledgment
We would like to acknowledge and extend our gratitude to the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM) Kedah Branch for providing the opportunity to conduct this research.

References
Alabdulrazzaq, H., Alenezi, M. N., Rawajfih, Y., & Alghannam, B. A. (2020). *On the accuracy of ARIMA based prediction of COVID-19 spread*. *January*.

Ang, Z. Y., Balqis-Ali, N. Z., Jailani, A. S., Kong, Y. L., Sharif, S. M., & Fun, W. H. (2023). COVID-19 clusters in Malaysia: characteristics, detection methods and modes of early transmission. *Western Pacific Surveillance and Response Journal : WPSAR*, *14*(4), 1–11. https://doi.org/10.5365/wpsar.2023.14.4.1058

Arnakim, L. Y., & Kibtiah, T. M. (2021). Response of ASEAN member states to the spread of COVID-19 in Southeast Asia. *IOP Conference Series: Earth and Environmental Science*, *729*(1). https://doi.org/10.1088/1755-1315/729/1/012100

Awang, H., Mahmud, N., Wahab, A., Abdul Rashid, N., Abd Rahman, M. A., & Embong, K. (2022). Situational Analysis of COVID-19 among Healthcare Workers in Terengganu State of Malaysia. *European Journal of Medical and Health Sciences*, *4*(2), 63–66. https://doi.org/10.24018/ejmed.2022.4.2.1271

Azhar, Z. I., Chen, X. W., Mohamad, M., Ahmad Saman, M. S., Isa, M. R., & Ismail, N. (2020). COVID-19 Review: An Epidemiological Perspective and Malaysian Scenario in Handling the Pandemic (January – May 2020). *Journal of Clinical and Health Sciences*, *5*(1), 26. https://doi.org/10.24191/jchs.v5i1.9002

Azit, N. A., Mohd Suan, M. A., Omar, N., Dali, N., Romli, M., Md Yusof, M. A., Ahmad, M., Ibrahim, M. Z., & Abdul Rahman, S. (2022). Epidemiological Investigation of a COVID-19 Community Cluster in Kedah, Malaysia. *International Journal of Travel Medicine and Global Health*, *10*(1), 10–15. https://doi.org/10.34172/ijtmgh.2022.03

Cheong, Y. L., Ghazali, S. M., Che Ibrahim, M. K. bin, Kee, C. C., Md Iderus, N. H., Ruslan, Q. binti, Gill, B. S., Lee, F. C. H., & Lim, K. H. (2022). Assessing the Spatiotemporal Spread Pattern of the COVID-19 Pandemic in Malaysia. *Frontiers in Public Health*, *10*(March), 1–11. https://doi.org/10.3389/fpubh.2022.836358

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, *20*(5), 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1

Ganasegeran, K., Jamil, M. F. A., Ch'ng, A. S. H., Looi, I., & Peariasamy, K. M. (2021). Influence of population density for covid-19 spread in malaysia: An ecological study. *International Journal of Environmental Research and Public Health*, *18*(18). https://doi.org/10.3390/ijerph18189866

Gawande, M. S., Zade, N., Kumar, P., Gundewar, S., Weerarathna, I. N., & Verma, P. (2025). The role of artificial intelligence in pandemic responses: from epidemiological modeling to vaccine development. *Molecular Biomedicine*, *6*(1), 1–25. https://doi.org/10.1186/s43556-024-00238-3

Grandi, S., & Bernasconi, A. (2021). Geo-online explanatory data visualization tools as crisis management and communication instruments. *Proceedings of the ICA*, *4*(December), 1–8. https://doi.org/10.5194/ica-proc-4-41-2021

Hashim, J. H., Adman, M. A., Hashim, Z., Mohd Radi, M. F., & Kwan, S. C. (2021). COVID-19 Epidemic in Malaysia: Epidemic Progression, Challenges, and Response. *Frontiers in Public Health*, *9*(May), 1–19. https://doi.org/10.3389/fpubh.2021.560592

Jayaraj, V. J., Chong, D. W. Q., Wan, K. S., Hairi, N. N., Bhoo-Pathy, N., Rampal, S., & Ng, C. W. (2023). Estimating excess mortalities due to the COVID-19 pandemic in Malaysia between January 2020 and September 2021. *Scientific Reports*, *13*(1), 1–11. https://doi.org/10.1038/s41598-022-26927-z

Kamarudin, A. N. A., Zainol, Z., Kassim, N. F. A., & Sharif, R. (2021). Prediction of COVID-19 cases in Malaysia by using machine learning: A preliminary testing. *2021 International Conference of Women in Data Science at Taif University, WiDSTaif 2021*. https://doi.org/10.1109/WIDSTAIF52235.2021.9430222

Keng Tok, P. S., Kang, K. Y., Ng, S. W., Rahman, N. A., Syahmi, M. A., Pathmanathan, M. D., Appannan, M. R., Peariasamy, K. M., & Sivasampu, S. (2024). Post COVID-19 condition among adults in Malaysia following the Omicron wave: A prospective cohort study. *PLoS ONE*, *19*(1 January), 4–11. https://doi.org/10.1371/journal.pone.0296488

Khorram-Manesh, A., Goniewicz, K., & Burkle, F. M. (2024). Unleashing the global potential of public health: A framework for future pandemic response. *Journal of Infection and Public Health*, *17*(1), 82–95. https://doi.org/10.1016/j.jiph.2023.10.038

Llc, S. K. J. H., & Segall, R. S. (2023). *Post-Pandemic Analysis of the Broader Impact of COVID-19 on the World ' s Economy , Health and Education Department of Information Systems & Business Analytics Department of Mathematics and Computer Science*. *Wmsci*, 286–293.

Mustapha, R., Awang, H., Mahmud, M., Paris, T. N. S. T., Zakaria, J., & Ibrahim, N. (2024). Online Learning in Malaysia amidst COVID-19 Pandemic (2020-2022): A Bibliometric Analysis. *Open Journal of Social Sciences*, *12*(03), 278–292. https://doi.org/10.4236/jss.2024.123020

Pan, Z., Nguyen, H. L., Abu-Gellban, H., & Zhang, Y. (2020). Google Trends Analysis of COVID-19 Pandemic. *Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020*, *1*, 3438–3446. https://doi.org/10.1109/BigData50022.2020.9377852

Podder, P., & Mondal, M. R. H. (2020). Machine learning to predict COVID-19 and ICU requirement. *Proceedings of 2020 11th International Conference on Electrical and Computer Engineering, ICECE 2020*, 483–486. https://doi.org/10.1109/ICECE51571.2020.9393123

Purnomo, E. P., Agustiyara, Nurmandi, A., Dewi, A., Rosa, E. M., Bayu, A. H., & Erviana, R. (2022). ASEAN Policy Responses to COVID-19 Pandemic: Adaptation and Experimentation Policy: A Study of ASEAN Countries Policy Volatility for COVID-19 Pandemic. *SAGE Open*, *12*(1). https://doi.org/10.1177/21582440221082145

Volume 7 Issue 20 (March 2025) PP. 259-272
DOI 10.35631/IJIREV.720016
Razzak, M. I., Imran, M., & Xu, G. (2020). Big data analytics for preventive medicine. In *Neural Computing and Applications* (Vol. 32, Issue 9). Springer London. https://doi.org/10.1007/s00521-019-04095-y

Sayeed, M. S., Hishamuddin, S. N., & Song, O. T. (2024). Covid-19 forecasting model based on machine learning approaches: a review. *Bulletin of Electrical Engineering and Informatics*, *13*(6), 4335–4345. https://doi.org/10.11591/eei.v13i6.7304

Sedek, K. A., & Annuar, K. (2024). *Leveraging Big Data and Interactive Dashboards for Enhanced Lung Cancer Risk Leveraging Big Data and Interactive Dashboards for Enhanced Lung Cancer Risk Assessment*. *October*.

Shin, S., Na, I., & Elmqvist, N. (2024). *Drillboards: Adaptive Visualization Dashboards for Dynamic Personalization of Visualization Experiences*. 1–13. http://arxiv.org/abs/2410.12744

Talib, N., Zulkifli, N. F., Saad, N. M., & Zaki, N. A. M. (2023). Covid-19 Geospatial Animation Map of Workplace Cluster in Selangor, Malaysia. *Journal of Sustainability Science and Management*, *18*(12), 56–65. https://doi.org/10.46754/jssm.2023.12.006

Tan, M. M., Musa, A. F., & Su, T. T. (2022). The role of religion in mitigating the COVID-19 pandemic: the Malaysian multi-faith perspectives. *Health Promotion International*, *37*(1). https://doi.org/10.1093/heapro/daab041

Tantrakarnapa, K., Bhopdhornangkul, B., & Nakhaapakorn, K. (2022). Influencing factors of COVID-19 spreading: a case study of Thailand. *Journal of Public Health (Germany)*, *30*(3), 621–627. https://doi.org/10.1007/s10389-020-01329-5

Vahedi, B., Karimzadeh, M., & Zoraghein, H. (2021). Spatiotemporal prediction of COVID-19 cases using inter- and intra-county proxies of human interactions. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-26742-6

Vaishya, R., Javaid, M., Khan, I. H., & Haleem, A. (2020). Artificial Intelligence (AI) applications for COVID-19 pandemic. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, *14*(4), 337–339. https://doi.org/10.1016/j.dsx.2020.04.012

Y. Wang, W. Shi, Y. Sun and C. -H. Yeh, "A Novel Framework to Forecast COVID-19 Incidence Based on Google Trends Search Data," in IEEE Transactions on Computational Social Systems, vol. 11, no. 1, pp. 1352-1361, Feb. 2024, doi: 10.1109/TCSS.2023.3255256.

Zheng, N., Du, S., Wang, J., Zhang, H., Cui, W., Kang, Z., Yang, T., Lou, B., Chi, Y., Long, H., Ma, M., Yuan, Q., Zhang, S., Zhang, D., Ye, F., & Xin, J. (2020). Predicting COVID-19 in China Using Hybrid AI Model. *IEEE Transactions on Cybernetics*, *50*(7), 2891–2904. https://doi.org/10.1109/TCYB.2020.2990162