Forecasting Mobility Trends in Southeast Asia during the Coronavirus (Covid-19) Pandemic by Machine Learning Approaches

Mekruksavanich, S.¹ and Jitpattanakul, A.²

¹Department of Computer Engineering, School of Information and Communication Technology University of Phayao, Phayao, Thailand, E-mail: sakorn.me@up.ac.th ²Intelligent and Nonlinear Dynamic Innovations Research Center, Department of Mathematics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand E-mail: anuchit.j@sci.kmutnb.ac.th

DOI: https://doi.org/10.52939/ijg.v17i5.2007

Abstract

The novel coronavirus (COVID-19) was declared as the 2019-20 coronavirus pandemic by the World Health Organization (WHO) in March 2020. The COVID-19 virus has rapidly spread nationwide and internationally and caused 188 countries to report more than ten million cases of individuals contracting COVID-19. Typically, the virus is conveyed from person to person via respiratory droplets produced by coughing and sneezing. The time period between exposure and onset of symptoms is typically between two and fourteen days, and on average five days. The COVID-19 pandemic has affected many businesses relating to transportation including tourism, import-export commerce, the aviation business, and so forth. Governmental intervention in each country has had an impact on mobility trends depending on the degree of restriction such as social distancing, sharing mobility, and public transport. A COVID-19 surveillance system is one of the principal methods used for detecting COVID-19 epidemics, using short-period monitoring. However, while these networks present information on the activities of COVID-19, acquiring completed surveillance data from every medical station is profusely difficult due to many factors. This research aims to propose a performance model of machine learning approaches for COVID-19 pandemic forecasting of mobility trends in each country in Southeast Asia. Spatial data and non-spatial data are used to build the machine learning models. The experiments conducted showed that the model gave a forecasting accuracy in walking and driving mobility of 94.40% and 92.00%, respectively. The proposed forecasting model was developed to be of benefits to health authorities in the planning and administration of a suitable strategy to make decisions concerning transportation planning in each country.

1. Introduction

Coronaviruses are a large family of respiratory viruses. Several viral epidemics have occurred in the previous two decades, including the severe acute respiratory syndrome coronavirus (SARS-CoV) in 2002–2003 and H1N1 influenza in 2009. In 2012, the Middle East respiratory syndrome coronavirus (MERS-CoV) was detected for the first time in Saudi Arabia (Letko and Munster, 2020). Originally, these viruses were passed from animals to humans. SARS, for example, was transferred to people via civet cats, while MERS was transmitted to humans via a species of camel.

The new coronavirus (COVID-19) is an infectious disease caused by the coronavirus 2 (SARS-CoV 2) that causes severe acute respiratory syndrome. The disease was originally diagnosed and a global outbreak was declared in December 2019 in the Chinese province of Wuhan. The COVID-19 virus has rapidly spread across the country and

resulting in the World Health worldwide, Organization's (WHO) declaration of the 2019-20 coronavirus pandemic in March 2020. This WHO notification was prompted by the increasing number of cases, deaths, and impacted nations. By 26 April 2020, 114 nations had reported that almost three million people had caught COVID-19, the viruscaused sickness. Almost 200,000 individuals had perished. Around 80% of people infected displayed minor respiratory infections, though nearly half developed pneumonia. Another 15% developed serious sickness, while 5% required critical care. In comparison to SARS, the COVID-19 outbreak has a geographic significantly broader spread. Additionally, the number of cases reported outside of China nearly quadrupled within a two-week period while the number of afflicted countries nearly tripled (Wu et al., 2020 and Wu and McGoogan, 2020). The global instances of COVID-

19 as reported by the Johns Hopkins University Center for Systems Science and Engineering (CSSE) (Dong et al., 2020) are depicted in Figure 1. As a result, the global pandemic of COVID-19 has posed a significant threat to public health and the global economy.

The COVID-19 virus affects individuals differently. Fever, cough, fatigue, shortness of breath, and loss of smell are all common signs of COVID-19 infection (Yapici-Eser et al., 2021 and Larsen et al., 2020). While the majority of cases are mild, some develop into viral pneumonia, multiorgan failure, or cytokine storm (Verity et al., 2020). Less frequently, muscle soreness, sputum production, and sore throat occur. Typically, the virus is conveyed from person to person via respiratory droplets produced by coughing and sneezing. The time interval between exposure and the start of symptoms is typically between two and fourteen days, with an average of five days. Numerous factors can influence the transmission of COVID-19 viruses, including climate conditions (such as temperature and humidity), population density, medical quality, and public transportation.

Southeast Asia's governments reacted swiftly to reports of the coronavirus spreading in China, demonstrating that many had learned valuable lessons from the 2003 SARS pandemic. Thailand, Singapore, and a number of other countries quickly established screening, contact tracing, and quarantine processes for suspected cases. Vietnam's government and numerous individual airlines suspended flights from China, whereas the majority of ASEAN states only implemented travel screening. The sudden increase in cases in Indonesia, Malaysia, the Philippines, and Thailand that began in mid-March indicates a shift in the trajectory of the outbreak away from a steady and controlled increase toward a more significant outbreak. While the delay in widespread local infection illustrates the limited effectiveness of the region's initial approach, Thailand and its ASEAN neighbors now face a much more complex challenge as they reluctantly accept social isolation and containment initiatives in an attempt to prevent the spread.

Over the last few years, the ongoing difficulty in global public health surveillance has been in determining the danger posed by infectious disease outbreaks with an increased understanding of their natural geographic distribution. The bulk of spatial epidemiology data continues to rise, and extracting valuable intelligence from this information has become a priority. They exhibit similar qualities of big data in terms of volume, velocity, diversity, value, and authenticity. Spatial epidemiology data are critical for addressing the extensive data and health analytics difficulties inherent in digital epidemiology (Elliott et al., 2000). The volume and velocity with which transportation and mobility data are generated now significantly exceed the scales at which they were acquired, processed, and evaluated at the turn of the twentieth century. The combination of new digitization paradigms such as the Internet of Things (IoT), the proliferation of Smart Cities (Zanella et al., 2014 and Syed et al., 2021), the dramatic decline in data storage costs in the silicon market, the latest advancements in wireless technologies, and the widespread use of sensors or personal devices has low-cost significantly improved the human ability to gain a more detailed understanding of the world. To decrease the spread of COVID-19 epidemics, it is critical to limit or minimize travel, as well as the requirements for travel in general. This is especially true for non-essential traffic.



Figure 1: COVID-19 Southeast Asia cases on April 27, 2020 International Journal of Geoinformatics, Vol. 17, No. 5, October 2021 ISSN 2673-0014 (Online) / © Geoinformatics International The impact of government intervention on mobility patterns varies according on the degree of restrictions imposed by social distancing, shared mobility, and public transportation. Thus, this research aims to provide a performance model for machine learning (ML) techniques to anticipate the COVID-19 pandemic movement in each Southeast Asian country. ML models are constructed using both spatial and non-spatial data. The proposed forecasting model was developed to assist health authorities in planning and administering an appropriate approach for transportation planning decisions in each country. The remainder of this article is organized as follows: Section 2 establishes the necessary prerequisites, such as data preparation and ML. Section 3 discusses the research methods for predicting mobility patterns using ML methodologies using spread data from the coronavirus (COVID-19) pandemic. In Section 4, we provide our findings and discuss them. Finally, Section 5 presents conclusions for this research.

2. Related Works

Machine learning is utilized in a variety of sectors, including medicine, to forecast disease and its outcomes (Mekruksavanich, 2017). In medicine, accurate diagnosis and therapy at the appropriate time are critical to success. If the treatment has a high risk of mistakes, it may result in multiple deaths. As a result, researchers have begun developing uses of artificial intelligence for medical therapy. The process is exacerbated by the fact that the researchers must select the appropriate tool: this is a matter of life and death (Al-Turaiki et al., 2016).

ML set a precedent in the realm of health care with this accomplishment. Machine learning techniques are used to comprehend, analyze, and forecast the output of massive datasets (Mekruksavanich et al., 2020) (Mekruksavanich and Jitpattanakul, 2021). These machine learning algorithms were used to classify samples into treatment groups and to identify illness signs. Hospitals rely on machine learning to manage administrative operations and treat infectious diseases (Chouhan et al., 2020 and Köse et al., 2015). ML approaches have already been utilized to treat cancer, pneumonia, diabetes, Parkinson's disease, rheumatoid arthritis, and neuromuscular disorders, amongst other ailments; they provide above 90% accuracy in prediction and forecasting (Quinlan, 1993). COVID-19, the pandemic infection, is the consequence of a lethal virus that has claimed the lives of millions of people worldwide. This infection is incurable. On the basis of WHO and CDC-defined symptoms, machine learning algorithms have been employed to predict

whether patients are infected with the virus (Chandra and Verma, 2020). Additionally, machine learning is utilized to diagnose the disease using X-ray pictures. For example, chest pictures of patients can be utilized to determine whether they have the COVID-19 infection (Köse et al., 2015). Additionally, social distancing may be tracked using machine learning; with this method, we can protect ourselves from COVID-19 (Chouhan et al., 2020).

Numerous machine learning algorithms are utilized to foresee and predict future events (Mekruksavanich and Jitpattanakul, 2021a). Support vector machines. linear regression, logistic regression, Naive Bayes, decision trees, K-nearest neighbor, and neural networks are all examples of machine learning approaches used for forecasting (multilayer perceptron) (Sharmila et al., 2017 and Er et al., 2010). Each technique is distinct in its own way and is applied differently depending on the accuracy outcomes. For prediction, the model with the highest accuracy during the model evaluation procedure is considered. Similarly, we identified and deployed artificial neural networks to estimate mobility trends in each country in Southeast Asia during the initial spread of COVID-19 in 2020, as they produced the most accurate results of any classifier or forecasting method we investigated.

3. Proposed Methodology

The proposed framework for anticipating mobility trends during the COVID-19 pandemic is composed of four stages: data collection, data analysis, model training, and forecasting mobility using the ML model. These steps must be integrated into a forecasting system, with the system flow described by a generic model, as seen in Figure 2.



Figure 2: The proposed framework of forecasting mobility trends

The following sections discuss each stage of the process presented by the research for forecasting mobility trends during the COVID-19 pandemic.

3.1 Data Collection

Data collection process refers to a group of operations that are performed in order to:

- collect data acquired and sampled from signals,
- convert those samples into digital values that can be used and modified by the computer,
- and feed that data into the processing flow of a specific data system.

Using machine learning methodologies, two datasets are collected for analysis and the construction of a forecasting model in this study. The first dataset is COVID-19 data from Johns Hopkins University's Center for System Science and Engineering (CSSE) (Dong et al., 2020). This COVID-19 dataset has evolved into a go-to resource for experts and the general public interested in assessing the global spread of the virus. The data are aggregated on a country-by-country basis.

The second dataset is given by Apple, which is sharing aggregated mobility data from Apple Maps with public health officials to help them better understand how people move in their areas. Apple's mobility trend reports have been charting the diversity of human mobility in countries and cities globally since January 2020, using location data from Apple's maps products. Mobility trend reports in major cities and 63 nations or regions are included in the data. Figure 3 illustrates two examples of mobility trend datasets (driving mobility trends and walking mobility trends) used in this research study. Then, we collected quantitative data of mobile operating system market share for each country in Southeast Asia from (StatCounter, 2021) and the Southeast Asia population data from (Worldometer, 2021). These quantitative data are shown in Table 1.



Figure 3: Two types of mobility trend dataset: a) driving mobility trends, and b) walking mobility trends

Country	Population in 2020	iOS Market Share
Cambodia	16,718,965	9.78%
Indonesia	273,523,615	8.78%
Malaysia	32,365,999	23.85%
Philippines	109,581,078	10.80%
Singapore	5,850,342	27.16%
Thailand	69,799,978	25.98%
Vietnam	97,338,579	33.54%

Table 1: Size of mobile operating system market share of each country

Table 2: The overall number of confirmed cases, active cases, recoveries, and deaths

Total	Numbers
Confirmed cases	49,925
Active cases	35,757
Recoveries	12,497
Deaths	1,671



-reb 16-reb 24-reb 1-mai o-mai 12-mai 16-mai 24-mai 50-mai 5-Api 10-Api 10-Api

a) Overall number of cases





Figure 4: The overall number of confirmed cases, active cases, recoveries, and deaths in Southeast Asian locations

3.2 Data Analytics

Typically, data analytics models are characterized as descriptive analytics. The purpose of descriptive analytics is to attempt to summarize and explain the acquired data. This is accomplished by reducing useless data and inferring regularities in the form of groups or fitting statistical distributions. The objective of predictive analytics is to discover a relationship between a set of observable features or predictors and a target variable. Predictive models are capable of forecasting the value of an unknown target variable based on new input data that is not necessarily identical to any of the historical examples used to build the model. Prescriptive analytics uses optimization techniques, expert systems, and mathematical programming to propose the optimal course of action or decision to make from a set of alternatives. The COVID-19 pandemic data and mobility trend data during the initial spreading period are evaluated in this research to interpret some patterns through statistical Covid19 Stripes: Reported deaths (new cases per day)



calculations as part of the descriptive analytics. Then, the artificial neuron network approach from among the ML methodologies is chosen to develop a forecasting model for COVID-19 pandemic movement trends in each of the Southeast Asian countries.

The COVID-19 epidemic in Southeast Asia is examined in this study via the examination of seven countries: Thailand, Vietnam, Singapore, Malaysia, the Philippines, and Cambodia. Indonesia, According to a rolling total maintained by the United States' Johns Hopkins University, these Southeast Asian countries had reported approximately 30,000 cases at the time of writing, with a total of 1,144 fatalities. Table 2 and Figure 4 detail the overall number of confirmed cases, active cases, recoveries, and deaths in this location. The overall number of confirmed and death cases during the research study is shown in Figure 5. The cumulative number of confirmed cases from January to April 26, 2020 is depicted in Figure 6.

Covid19 Stripes: Confirmed cases (new cases per day)



b) Death cases





Covid19: Confirmed cases (cumulative) as of April 26, 2020



Figure 6: Cumulative confirmed cases since April 26, 2020

3.3 Model Training

Artificial neural networks (ANNs) are computational structures that, through an iterative learning process, could learn the links between a collection of input and output variables. Although these networks employ simple computations such as addition and multiplication, they are capable of tackling complicated, non-linear problems (Bishop, 1995). Once trained, a network can be used to predict a variable of interest using an independent dataset, typically with minimal change. The primary component of ANNs is a layer of neurons that are fully connected to the next layer via a set of weights (edges). Each ANN is composed of at least one input layer, one output layer, and one hidden layer. The simplest type of ANN is referred to as a perceptron, and serves as the fundamental building block of neural networks. Each input is multiplied by a corresponding weight in a perceptron and then aggregated using a mathematical procedure called neuron activation. The output is subsequently computed using another function. ANNs are a collection of layers formed by the stacking of perceptrons. For instance, if the inputs to the *i*th perceptron in a network are denoted by x_{1i} , ..., x_{ni} , assuming that a summation function is used to calculate the outputs (denoted by z_i), that is:

$$z_i = \sum_{j=1}^m x_{ij} W_i ij + b_i$$

Equation 1

where *n* is the number of inputs; *m* is the number of neurons in the current layer; w_{ij} is the weight of the *j*th neuron (*j*th input to the *i*th cell), and b_i is a bias term.

3.4 Forecasting Modeling

Different supervised ANN architectures were trained to model in this step. The dates and number of days used to train the ANNs differed significantly among countries. The initial date was January 1, 2020 for all scenarios. The final dates and the number of days of the training periods were May 27, 2020 and 148 days, respectively.

The trained ANNs were of the multilayer perceptron type, with a feedforward architecture and a hidden layer, and were constructed using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) iterative technique for addressing optimization issues. The processing was carried out on a machine equipped with two Intel i7 - 8700 Processors (12 M Cache, 4.60 GHz), 16 GB of RAM, six cores, and 64-bit Windows 10 Pro. The BFGS memoryless quasi-Newton algorithm has been successfully utilized to reduce mistakes in artificial neural networks. When calculating the Hessian matrix is complicated or time intensive, the quasi-Newton approach is utilized. When compared to the gradient descent, this approach converges quickly.

4. Results and Discussions

In this study, we introduced an ANN model to forecast and provide a quantitative overview of the mobility trends status of COVID-19 in Southeast Asia during a five-month period (January 1 to May 27, 2020). Take note that utilizing artificial intelligence in epidemiological investigations is a new method. Table 3 shows the accuracy performance of mobility trend forecasting in the walking category. The average accuracy of walking mobility was around 94.40%. The best accuracy was on mobility trend predictions in the Philippines with 97.90%. However, the minimum accuracy was 87.40% for mobility trend predictions in Singapore. Table 4 summarizes the forecasting accuracy for mobility trends in the driving category. Driving mobility accuracy was approximately 92.20% on average. With 99.80% accuracy, the Indonesian mobility trends were predicted. However, the minimal accuracy was 71.60% for predicting Singapore's mobility trends.

Country	Mobility	Accuracy
Cambodia	Walking	0.968
Indonesia	Walking	0.973
Malaysia	Walking	0.923
Philippines	Walking	0.979
Singapore	Walking	0.874
Thailand	Walking	0.932
Vietnam	Walking	0.963
Average		0.944

Table 3: Accuracy performance of mobility trend forecasting in the walking category

Country	Mobility	Accuracy
Cambodia	Driving	0.869
Indonesia	Driving	0.998
Malaysia	Driving	0.989
Philippines	Driving	0.954
Singapore	Driving	0.716
Thailand	Driving	0.984
Vietnam	Driving	0.933
Average		0.920

Table 4: Accuracy performance of mobility trend forecasting in the driving category

5. Conclusion

We present an ANN prediction model in this paper that utilizes a multilayer perceptron neural network with a feedforward design. This model is used as artificial intelligence to estimate COVID-19 movement trends in Southeast Asia over a fivemonth timeframe. The proposed model is assessed in terms of its accuracy in estimating the number of walking and driving mobility incidents, as well as the number of occurring instances, during the COVID-19 scenario period. The proposed model is highly effective at forecasting mobility trends within 148 days (January 1 to May 27, 2020). According to the promising results achieved by the multilayer perceptron neural network model, the suggested model benefits health authorities by assisting them in planning and administering an appropriate strategy for transportation planning decisions in each country.

Acknowledgement

This research project was supported by the Thailand Science Research and Innovation Fund and the University of Phayao (Grant No. FF64-UoE008).

References

- Al-Turaiki, I., Alshahrani, M. and Almutairi, T., 2016, Building Predictive Models for MERS-CoV Infections Using Data Mining Techniques. *Journal of Infection and Public Health*, Vol. 9(6), 744-748.
- Bishop, C. M., 1995, *Neural Networks for Pattern Recognition*. USA: Oxford University Press, Inc.
- Chandra, T. B and Verma, K., 2020, Pneumonia Detection on Chest X-ray Using Machine Learning Paradigm, *Proceedings of 3rd International Conference on Computer Vision and Image Processing*, 21–33. DOI:10.1007/978-981-32-9088-4 3.
- Chouhan, V., Singh, S. K., Khamparia, A., Gupta, D., Tiwari, P., Moreira, C., Damaševičius, R. and de Albuquerque, V. H. C., 2020, A Novel Transfer Learning-Based Approach for

Pneumonia Detection in Chest X-ray Images. Applied Sciences, Vol. 10(2), 1-17, DOI:10.3390/app10020559.

- Dong, E., Du, H. and Gardner, L., 2020, An Interactive Web-Based Dashboard to Track Covid19 in Real Time. *The Lancet Infectious Diseases*, Vol. 20(5), 533–534.
- Elliott, P., Wakefield, J., Best, N. and Briggs, D., 2000, *Spatial Epidemiology: Methods and Applications*.

Oxford: Oxford University Press.

- Er, O., Yumusak, N. and Temurtas, F., 2010, Chest Diseases Diagnosis Using Artificial Neural Networks. *Expert Systems with Applications*, Vol. 37(12), 7648–7655.
- Köse, U., Güraksın, G. E. and Deperlioğlu, M., 2015, Diabetes Determination Via Vortex Optimization Algorithm Based Support Vector Machines. *Medical Technologies National Conference (TIPTEKNO)*, 1–4.
- Larsen, J. R., Martin, M. R., Martin, J. D., Kuhn, P. and Hicks, J. B., 2020, Modeling the Onset of Symptoms of Covid19. *Frontiers in Public Health*, Vol. 8, https://doi.org/10.3389-/fpubh.2020.00473.
- Letko, M. and Munster, V., 2020, *Studying Evolutionary Adaptation of MERSCoV*. New York, NY: Springer US, 3–8.
- Mekruksavanich, S., 2017, Forecasting the Spread of Seasonal Influenza Epidemics by Neural Networks with Spatial Data. *International Journal of Geoinformatics*, Vol. 13(1), DOI:https://doi.org/10.52939/ijg.v13i1.1003.
- Mekruksavanich, S., Jitpattanakul, A., Youplao, P. and Yupapin, P., 2020, Enhanced Hand-Oriented Activity Recognition Based on Smartwatch Sensor Data Using LSTMs. *Symmetry*, Vol. 12, 1-19, https://doi.org/10.3390/sym12091570.
- Mekruksavanich, S. and Jitpattanakul, A., 2021, Biometric User Identification Based on Human Activity Recognition Using Wearable Sensors: An Experiment Using Deep Learning Models. *Electronics*, Vol. 10, 308. https://doi.org/10.3390/electronics10030308.

- Mekruksavanich, S. and Jitpattanakul, A., 2021a, LSTM Networks Using Smartphone Data for Sensor-Based Human Activity Recognition in Smart Homes. *Sensors*, Vol. 21, 1-25, https://doi.org/10.3390/s21051636.
- Quinlan, J. R., 1993, C4.5: Programs for Machine Learning. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Sharmila, L., Dharuman, C. and Venkatesan, P., 2017, Disease Classification Using Machine Learning Algorithms-A Comparative Study. *International Journal of Pure and Applied Mathematics*, Vol. 114(6), 1–10.
- StatCounter, 2021, Mobile Operating System Market Share Thailand (Aug 2020 - Aug 2021). https://gs.statcounter.com/os-marketshare/mobile/thailand/#monthly-202008-202108.
- Syed, A. S., Sierra-Sosa, D., Kumar, A. and Elmaghraby, A., 2021, IoT in Smart Cities: A Survey of Technologies, Practices and Challenges. *Smart Cities*, Vol. 4(2), 429–475.
- Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-Dannenburg, G., Thompson, H., Walker, P. G. T., Fu, H., Dighe, A., Griffin, J. T., Baguelin, M., Bhatia, S., Boonyasiri, A., Cori, A., Cucunubá, Z., FitzJohn, R., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Laydon, D., Nedjati-Gilani, G., Riley, S., van Elsland, S., Volz, E., Wang, H., Wang, Y., Xi, X., Donnelly, C. A., Ghani, A. C.

and Ferguson, N. M., 2020, Estimates of the Severity of Coronavirus Disease 2019: A Model-Based Analysis, *The Lancet Infectious Diseases*, Vol. 20(6), 669–677.

Worldometer, 2021, *South-Eastern Asia Population*. https://www.worldometers.info/worldpopulation/south-eastern-asia-population.

- Wu, J. T., Leung, K. and 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*, Vol. 395, 689– 697.
- Wu, Z. and McGoogan, J. M., 2020, Characteristics of an Important Lessons from the Coronavirus Disease 2019 (COVID19) Outbreak in China: Summary of a Report of 72314 Cases from the Chinese Center for Disease Control and Prevention. JAMA, Vol. 323(13), 1239–1242.
- Yapici-Eser, H., Koroglu, Y. E., Oztop-Cakmak, O., Keskin, O., Gursoy, A. and Gursoy-Ozdemir, Y., 2021, Neuropsychiatric Symptoms of Covid19 Explained by Sarscov2 Proteins' Mimicry Of Human Protein Interactions. *Frontiers in Human Neuroscience*, Vol. 15, DOI: 10.3389/fnhum.2021.656313.
- Zanella, A., Bui, N., Castellani, A., Vangelista, L. and Zorzi M., 2014, Internet of things for Smart Cities. *IEEEInternet of Things Journal*, Vol. 1(1), 22–32.