https://doi.org/10.48048/tis.2023.6884

A Hybrid LSTM and MLP Scheme for COVID-19 Prediction: A Case Study in Thailand

Peerapol Kompunt[1], Suparat Yongjoh[1], Phet Aimtongkham[2], Paisarn Muneesawang[1], Kiatichai Faksri[3,4] and Chakchai So-In[2,4,*]

[1]Department of Electrical and Computer Engineering, Faculty of Engineering, Naresuan University, Phitsanulok 65000, Thailand
[2]Applied Network Technology (ANT), Department of Computer Science, College of Computing, Khon Kaen University, Khon Kaen 40002, Thailand
[3]Department of Microbiology, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand
[4]Research and Diagnostic Center for Emerging Infectious Diseases, Khon Kaen University, Khon Kaen 40002, Thailand

([*]Corresponding author's e-mail: chakso@kku.ac.th)

Received: 20 March 2023, Revised: 25 April 2023, Accepted: 27 April 2023, Published: 1 August 2023

Abstract

After the COVID-19 epidemic, Thailand was affected in a variety of ways, with the most obvious being the economic downturn and the huge impact on health, including the loss of medical and human resources to combat the epidemic. However, Thailand still lacks analysis and prediction tools required to prepare for future epidemic situations. Therefore, we present development models for predicting the spread of the COVID-19 epidemic. In particular, the application of a long short-term memory (LSTM) and multilayer perceptron (MLP) model was investigated to predict new cases, total cases, new deaths, and total deaths. There are a total of 77 provinces in Thailand. The data used in this trial were obtained from the Department of Disease Control (DDC) of the Thai government. The modeling employed 2 types of data: dynamic (time series) and static. There were 2 phases: 1) the LSTM was used to manipulate time series data and 2) the MLP model was used to manipulate static data. Then, the models were merged for further analysis. We evaluated the performance of the combined model, yielding an accuracy of 99.72 % based on R^2 values, higher than the values obtained for state-of-the-art methods. In addition, the prediction results can be further combined with GIS data in each province and displayed via an easy-to-use web application for mapping.

Keywords: COVID-19, Prediction, Long short-term memory (LSTM), Multilayer perceptron (MLP)

Introduction

The current COVID-19 epidemic has had a worldwide impact on economics, public health, transportation, tourism, education, and even the daily lives of human beings. The beginning of the COVID-19 epidemic occurred in December 2019 in the city of Wuhan [1]. As the capital of Hubei Province in the People's Republic of China, Wuhan is the most populous city in central China, with over 19 million people [2]. On December 30, 2019, the National Health Commission of the People's Republic of China issued an official notice of idiopathic pneumonia, and due to the large population of Wuhan, the infection rate rose dramatically in a very short period. The infection caused a large number of deaths, and SARS-CoV-2, also known as COVID-19 [3], was transmitted from person to person through the air according to the National Health Commission of the People's Republic of China and the World Health Organization (WHO). The top 5 symptoms of people infected with COVID-19 are fever, cough, tiredness, shortness of breath, and headaches [4], and in patients with underlying diseases such as heart disease and diabetes, other symptoms may be present. Different countries have established different policies and methods of prevention for the COVID-19 epidemic.

According to statistics on the COVID-19 global epidemic by the World Health Organization [5], the top 5 countries with the most cumulative cases in the world are the USA, China, India, France, and Germany. The USA is the country with the most cumulative cases (102,977,369) and cumulative deaths (1,120,529). The top 5 most infected countries in Asia were China, India, Japan, South Korea, and Turkey, which are located in the same continent as Thailand. In Asia, China is the country with the most cumulative

cases, with 99,240,488 cumulative cases and 120,912 cumulative deaths, and Thailand is ranked 13th in most cumulative cases and 11th in cumulative deaths.

The COVID-19 outbreak in Thailand started on January 12, 2020 [6]. The first case was confirmed for a Chinese tourist who entered Thailand within 2 weeks. On January 31, 2020, a taxi driver who had never had a history of traveling abroad but had a history of driving taxis serving Chinese customers was confirmed to have contracted the virus; at this time, the number of infected continued to increase. In fact, Thailand has experienced 5 waves of the COVID-19 epidemic. The spread of COVID-19 has affected many aspects of life in Thailand, such as the economy, tourism [7], the health system, employment conditions, and the environment. The severity of the economic impact has been largely dependent on 2 factors: 1) the duration of restrictions on people's movement and economic activities in the country and 2) the effectiveness of the financial response during the crisis. Focusing on health spending to control the spread of the disease and supporting income for the households most affected by the pandemic has reduced the likelihood of an economic recession.

In the first wave of the outbreak, the impact of the measures taken by the government was notable. Lockdowns were issued, with government spending to support the economy. The office of the National Economics and Social Development Council (NESDC) estimates that the outbreak of COVID-19 affected the employment of workers in 9 sectors of production, covering more than 8 million individuals, not including the impact on millions of workers in the agricultural sector or on workers in the tourism and related service sectors, such as the restaurant and hotel trades, who were severely affected [8]. Following the main COVID-19 outbreak, there has been much violence since April 2021. The World Bank estimated that Thailand's economic growth in 2021 declined from the forecasted 3.4 to 2.2 %. In addition, employment was severely affected. especially in the service and business sectors [9].

Even after 2 years since COVID-19 virus reached Thailand, deaths still occur daily. There have now been 32,755 deaths nationally, with a total confirmed count of 4,682,470 new cases, or 839 cases per day (updated on 30 September 2022) [10]. Consequently, awareness of the virus quickly became widespread, and the decision was made to focus on disease prevention, surveillance, and data analysis. Initially, Thailand lacked tools, including deep learning models, for analyzing or predicting COVID-19 epidemic trends.

Thus, in this research, we propose a COVID-19 analysis tool with a novel architecture to analyze and predict COVID-19 trends by applying deep learning techniques to assess statistical data related to COVID-19 in each of the 77 provinces in Thailand. The tool combines long short-term memory (LSTM) [11] and a multilayer perceptron (MLP) model [12] to effectively analyze different types of data. There are 3 main modules, as shown in **Figure 1**: Data management, prediction modeling, and visualizing predictions. To enhance the ability to display data, we introduced a geographic information system (GIS) integrated with the model. Notably, data can be displayed in map form, and analyses can be conducted at the province level in Thailand. The contributions are as follows:

1) A new architecture that can be used as a tool to predict the COVID-19 epidemic situation is proposed.

2) Deep learning is applied to create predictive models with LSTM for time series data and MLPs for static data.

3) A tool used for displaying predictions using a web application in the form of a map of each province in Thailand is developed to show the prediction results in detail.

Literature review

The various research databases in Thailand were searched to find studies involving the application of deep learning methods to analyze the COVID-19 situation, and few studies that included mapping were identified, even though the epidemic has been ongoing for 3 years.

In the published research, many methods are used to predict the spread of COVID-19, including mathematical models such as parametric models to estimate the number of deaths due to COVID-19. One paper described the application of a suspicious-infected-death (SpID) model with notable parameter dimensionality and strong coupling; this approach was applied to estimate the number of deaths in Turkey [13]. In addition, image processing techniques are now being applied to help predict the rate of transmission risk of COVID-19 and its spread in human organs, in addition to the number of infected individuals; for example, chest X-ray images from infected individuals have been used to calculate the risk of lung infection [14,15].

By combining Internet of Things (IoT) devices and deep learning, people can be tracked, and the spread of infection can be predicted in a systematic way, such as by using BLE (Bluetooth Low Energy) to

track people's behavior and assess social distancing to reduce the spread of disease [16]. The economy also plays an important role in driving the progress of COVID-19 research at the national scale. Notably, each country worldwide has implemented different measures to support the national economy. Another method for analyzing these problems is to use artificial intelligence (AI) [17-19].

Ensemble machine learning of the factors that influenced COVID-19 across 3142 US counties was performed in [20]. Factors affecting COVID-19 infection were analyzed using the following features: county-scale physical and mental health levels, environmental pollution levels, access to health care, demographic characteristics, vulnerable population scores, and other epidemiological metrics. After collecting the relevant data, the correlations among the features and county trends were mapped, and 5 resulting metrics were explored: COVID-19 cases at day 25, total COVID-19 cases, deaths at day 100, total deaths, and quantile of the day of the first case. In terms of machine learning, a superlearner ensemble and a marginal prediction method were selected to develop a model with the following predictors: demographics, health resource availability, health risk factors, social vulnerability, and COVID-19-related information. The total number of COVID-19 cases to date, with an R^2 value of 0.87, had the greatest effect on outcomes based on tests of the cross-validated superlearner risk across COVID-19 trials, followed by the day of the first cases, COVID-19 cases at day 25, total deaths, and deaths at day 100. Additionally, the strongest predictors were ethnicity, accessibility to public transportation, and the level of disease prevention.

In another study, the COVID-19 epidemic trend was analyzed by combining LSTM and Markov methods [21]. A model for predicting COVID-19 transmission was developed to establish preventive measures and enhance advanced preparation for outbreaks. The study involved the use of LSTM in modeling, and the LSTM model has since been improved through coupling with a recurrent neural network (RNN). An LSTM model was trained using cumulative COVID-19 patient data from 4 countries: The US, Britain, Brazil, and Russia. The country model was trained for 50 epochs, with an average accuracy of 76 % and an average percentage error of 0.152. However, as the amount of data used in modeling increases, LSTM models become biased and yield higher errors. Therefore, to solve these problems, a Markov model was applied. After testing the LSTM-Markov model, accuracies of 96, 94, 97 and 98 % were obtained for the countries mentioned above. With an average accuracy of 96 % and an average percentage error of 0.038, it was concluded that incorporating the Markov model improved the accuracy of LSTM. Notably, up to a 60 % increase in model accuracy was observed, and the percentage error was reduced by 75 %.

In another study, a COVID-19 prediction analysis using artificial intelligence procedures was performed, and ArcGIS spatial analyst was applied, with a case study of Iraq [22]. Based on real infection data from Iraq, a model was established and tested to predict the spread of COVID-19. The total numbers of cases, recoveries, and deaths were calculated for a sample of 38,000 people. The data were divided into 26,600 cases, including 70 % in a randomized format to train the model. The remaining 30 % of cases were divided into half for model testing and verification. The model involved 3 types of artificial intelligence neural networks (AINNs), namely, a radial basis function (RBF), fuzzy c-means (FCM) clustering, and a nonlinear autoregressive network with exogenous inputs (NARX). The metric used to measure accuracy was the mean absolute percentage error (MAPE). Based on tests of the models, the accuracy of the NARX, FCM, and RBF models reached 84.8, 83.8 and 81.4 %, respectively. Overall, the method achieved 81.6 % accuracy, with an average error of 5.74 %.

A method for projecting the criticality of COVID-19 transmission in India using GIS and machine learning methods was proposed in [23]. A model used to predict the incidence of COVID-19 infection based on criticality index (CI) assessment was presented for use in lockdown analysis in India. It is critical to examine strategies for issuing and ceasing lockdowns because they affect many aspects of the economy and directly affect the population. The data used in this study were obtained for COVID-19 patients and spanned cumulative confirmed, deceased, and recovered individuals. The following machine learning techniques were used in the modeling process: Gaussian process regression (GPR), a support vector machine (SVM), and a decision tree (DT). The model performance and accuracy were evaluated based on the root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), and coefficient of determination (R^2). Based on model testing, modeling using the GPR network yielded the highest accuracy for all techniques, with 95 % accuracy, while the DT and SVM produced 94 and 93 % accuracy, respectively.

An epidemic analysis of COVID-19 in Italy based on spatiotemporal geographic information and Google Trends was performed in [24]. The COVID-19 epidemic in Italy has affected people's health and resulted in many deaths. A machine learning model was established to analyze the infection situation with spatiotemporal geographic information. The data used in the modeling were divided into the following types: Global outbreak data, age group data, and Google Trends (GTs). The GT dataset contained a total of

40 keywords related to infection factors. These data were processed with selected machine learning algorithms, e.g., a Bayesian network, a random forest, ad hoc clustering, C4.5, an SVM, k-nearest neighbors, and an ANN.

In the above study, the use of certain keywords in the filtering process had a considerable influence on the epidemic assessment results. The first round of testing resulted in the selection of 14 of the most effective keywords. The accuracy of all algorithms was compared in the first round of testing, and AdaBoost was the most accurate. After choosing AdaBoost for model construction, retraining was performed to improve performance and identify the most relevant keywords. The experimental results showed that the accuracy of model training was 93.75 % and the test accuracy was 95.24 %. The most correlated keywords were as follows: Mask, pneumonia, thermometer, Italian National Institute of Health (ISS), disinfection, and disposable gloves. Consequently, a Markov model was used to reduce the error of the LSTM model. After testing the LSTM-Markov model, accuracies of 96, 94, 97 and 98 % were obtained for the countries mentioned above. With an average accuracy of 96 % and an average percentage error of 0.038, it was concluded that incorporating Markov model to improve the accuracy of the LSTM model yielded a 60 % increase in model accuracy and reduced the error rate by 75 %.

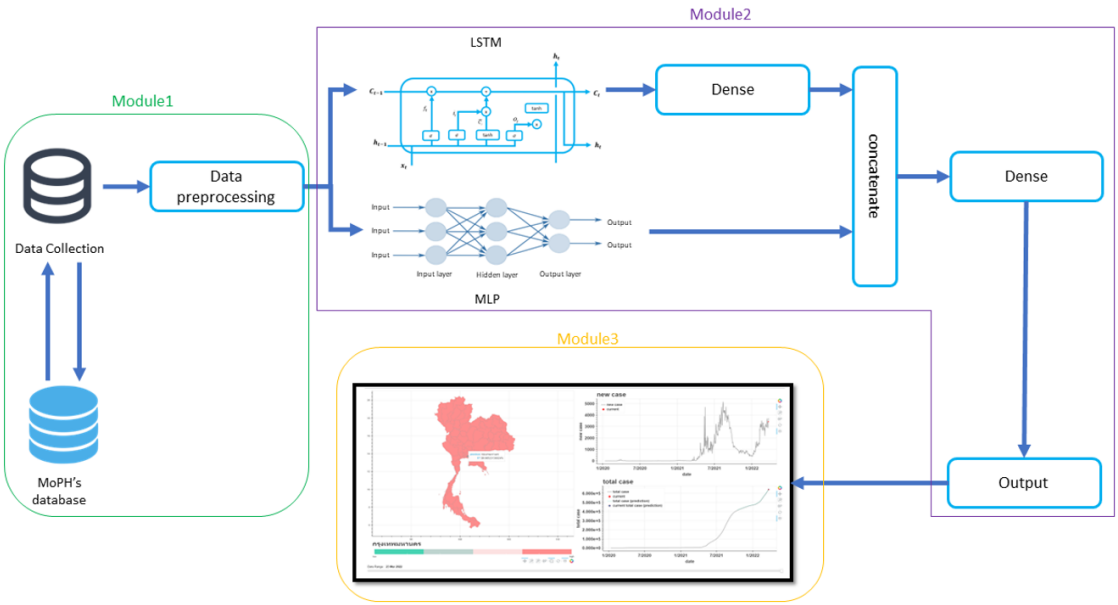

Figure 1 The system architecture of the 3 main modules of the system.

Figure 1 shows the overall architecture of the proposed system, consisting of 3 modules, namely, data management, prediction modeling, and the display of predictions. Note that **Table 1** shows the notations used throughout the paper.

Table 1 Notation.

Notation	Definition
f_t	The forget gate
σ	The sigmoid function
W_f, W_c, W_i, W_0	The weight of matrices
h_{t-1}	The output of the previous cell state (at time t-1)
$x_t, \widetilde{C}_t$	The input value of the cell state at time t
b_f, b_c, b_i, b_0	Bias
i_t	The input gate
$tanh$	Tanh function
o_t	The output gate

Module 1: Data management

Module 1 is used to manage all data. This module retrieves data from the MoPH database via an API provided by the Thai government. The data retrieved span new cases, total cases, new deaths, and total deaths. After the data are retrieved through the API, all the data are combined into a data framework and preprocessed before they are input into the model.

Data collection

For the experiment, we collected data from the Thai government, which included new cases, total cases, new deaths, total deaths, and time series of records from each province in Thailand. There are 77 provinces in Thailand, and the data were collected from the first to the third wave of the pandemic in Thailand, or from 12/01/2020 until now (phases 1 - 2 were from 12/01/2020 - 31/03/2021, and phase 3 spans 01/04/2021-present). Therefore, there are approximately 50-thousand records (77 provinces × 700+ days) in the example shown in **Figure 2**.

The data are available from the API as follows:
https://covid19.ddc.moph.go.th/api/Cases/round-1to2-by-provinces and
https://covid19.ddc.moph.go.th/api/Cases/report-round-3-y21-line-lists-by-province.

	txn_date	province_en	new_case	total_case	new_case_excludeabroad	total_case_excludeabroad	new_death	total_death	Unnamed: 9
0	12-01-20	Krabi	0	0	0	0	0	0	NaN
1	13-01-20	Krabi	0	0	0	0	0	0	NaN
2	14-01-20	Krabi	0	0	0	0	0	0	NaN
3	15-01-20	Krabi	0	0	0	0	0	0	NaN
4	16-01-20	Krabi	0	0	0	0	0	0	NaN
...	...	...	...	...	...	...	...	...	...
55050	23-09-22	Mae Hong Son	0	6085	0	6083	0	79	NaN
55051	24-09-22	Mae Hong Son	1	6086	1	6084	0	79	NaN
55052	25-09-22	Mae Hong Son	0	6086	0	6084	0	79	NaN
55053	26-09-22	Mae Hong Son	0	6086	0	6084	0	79	NaN
55054	27-09-22	Mae Hong Son	0	6086	0	6084	0	79	NaN

55055 rows × 9 columns

Figure 2 Example of time series of records from 77 provinces collected from the Thai government.

Data preprocessing

In the data preprocessing stage, the null values were checked; notably, null data affect the training processes and the accuracy of the results. The data were then filtered by province and sorted in ascending order to ensure no shuffling in the time series.

Module 2: Prediction modeling

After the data were preprocessed, they were reformatted into a data framework in preparation for model construction. We used 2 models to establish different data formats. The LSTM model was created to handle dynamic time series data.

Moreover, an MLP model was created to manipulate data in a static format. After processing with both models, the outputs of both models were concatenated to produce the final output, and a plot was output to the 3rd module to display the predictions with a web application.

Long short-term memory

In the first experiment, we used the LSTM model as the primary model because it is able to efficiently manipulate sequential or time series data, such as that we collected from various sources. Information about the COVID-19 epidemic situation was obtained from the Department of Disease Control in Thailand.

The LSTM model was developed from a recurrent neural network (RNN). An RNN uses the computed output (results) from the previous nodes as the input data for the next node. Each node of the RNN has 2 inputs: The input data for that node and the output of the previous node(s). Over time, more stable and efficient LSTM models have been developed. LSTM retains the “status” of each node or the corresponding information to maintain temporal continuity and consider historical data. The LSTM technique includes gates that control the incoming data at each node, such as a memory gate layer, an input gate layer, and an output gate layer.

Forget gate layer

The forget gate is responsible for determining whether data entering the cell state should be retained or discarded. The information that is retained is evaluated based on the input data entering the node and the result. The result of the forget gate layer is a value between 0 and 1 and is computed considering the input from the previous node and a sigmoid function, as shown in Eq. (1). h_{t-1} is the previous hidden state.x_t is the current input, and the addition of both is based on the sigmoid function, which converts the output value. W_f is the weight matrix that connects the inputs to the hidden layer, $b_f$and is the bias.

$$f_t = \sigma\ (w_f \bullet [h_{t-1}, x_t] + b_f) \tag{1}$$

Input gate layer

The input gate receives new input data and then retains or "writes" that data to each node. The input gate process is divided into 2 parts. The first part involves updating the cell state. When an input is received, a sigmoid control function is used to determine whether to update the cell state. A sigmoid layer called the input gate layer is used to determine which values will be updated. A tanh layer is used to generate a vector of new candidate values that might be added to the state, as in Eqs. (2) and (3). i_t is a new input called the input gate input. W_i, W_c is the weight matrix that connects the inputs to the hidden layer. h_{t-1} is the previous hidden state, and x_t is the current input. Both variables use the sigmoid function, which yields an output value. b_i and b_c are the bias terms. $C\tilde{t}$ is the candidate cell state at time *t*.

$$i_t = \sigma\ (W_i \bullet [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\widetilde{C_t} = tanh\ \bullet (W_c \bullet [h_{t-1}, x_t] + b_c) \tag{3}$$

Output gate layer

The output gate prepares the output data. The output data are viewed through the lens of the cell state following certain computational processes. The sigmoid function is used to determine which data in the cell state to output; then, the cell state value is entered into the tanh function (to determine if the output is 1 or -1), and the value obtained from the tanh function is calculated based on the output value obtained from the sigmoid gate. The desired output is obtained as shown in Eq. (4). o_t is the cell state information multiplied by the output information. h_{t-1} is the previous hidden state. W_0 is the weight matrix that connects the inputs to the hidden layer. x_t is the current input and is assessed with the sigmoid function, which yields an output value. b_0 is the bias.

$$o_t = \sigma(W_0 \bullet [h_{t-1}, x_t] + b_0) \tag{4}$$

Multilayer perceptron

Multilayer perceptron (MLP): An MLP is a mathematical model or computer model that processes data based on connected computing. The basic neural network of an MLP consists of 3 main parts, as shown in **Figure 3**: The input layer, the hidden layer, and the output layer.

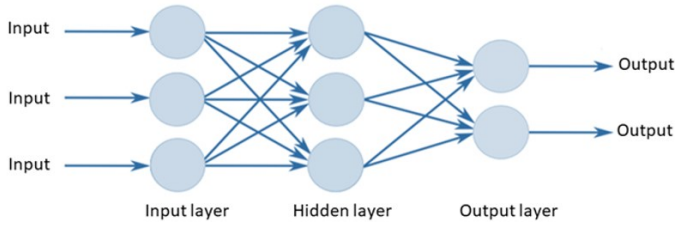

Figure 3 The structure of a multilayer perceptron.

The input is passed from sublayer to sublayer in the hidden layer. The activation function is used to determine when nodes in the previous layer send a signal (output). Each layer includes a different function. Before sending the data to the output layer, the incoming data determined to be linearly separable are processed. It may be necessary to use more than 1 hidden layer to convert the data into a linear layer and obtain a satisfactory output. Each hidden layer contains 1 or more neurons. The neurons are responsible for processing the inputs and outputs, with 1 neuron being able to read more than 1 value. A single value is called a single input perceptron, and any value greater than 1 is called a multiple-input perceptron. Each

neuron includes a summation function and an activation function. The MLP infrastructure is a form of a neural network with multiple layers, which is appropriate for complex data. A supervised learning process and a backpropagation process are used. In learning the process, there are 2 types of operations: Forward passing and backward passing.

The collected and preprocessed data are processed with a noncomplex LSTM model with an LSTM layer and a dense layer. The structure of this model, including model adjustment, is shown in **Figures 4** and **5**. The LSTM (only) model is duplicated to establish 8 models for different predictions: Predictions of new cases, total cases, new deaths, and total deaths with and without model adjustment.

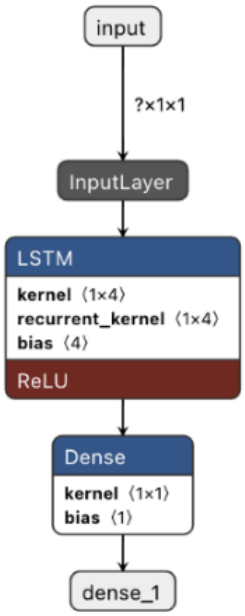

Figure 4 A visualization of the LSTM model.

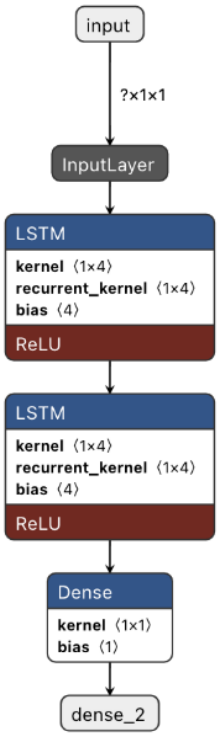

Figure 5 A visualization of LSTM model adjustment.

Tools and libraries

The Python programming language was used in conjunction with machine learning libraries to develop the proposed tools, libraries, and environment in this study.

Training and testing

The preprocessed data were separated into training and testing sets. At each iteration, data were filtered by province and ordered by date, and the filtered data were split into a training set and a test set at a ratio of 80:20. For the training and test sets, independent variables were separated from dependent variables to train each network.

For the combined LSTM and MLP model, the same Thailand government data as that used for the LSTM alone were used for evaluation (the adjusted and nonadjusted models for predicting new cases, total cases, new deaths, and total deaths). The number of hidden layers in the LSTM and MLP was set to 10 in each model. We assigned 1 neuron to each layer, and a batch size of 64 was used to determine the number of neurons. **Figure 6** shows the structure of the LSTM + MLP model.

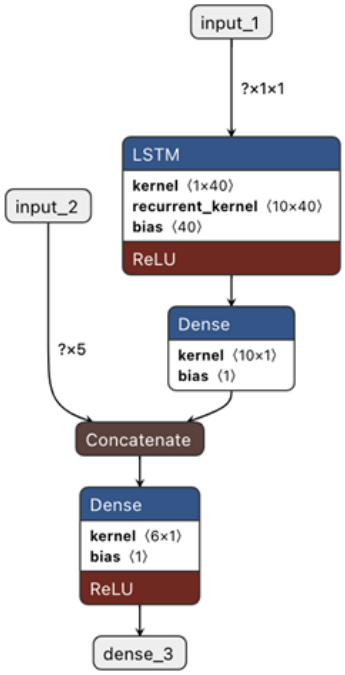

Figure 6 A visualization of the combined LSTM and MLP model.

Module 3: Displaying predictions.

When the results from modeling are plotted, the results are displayed in a web application created specifically for displaying model results. The coordinates of each province in Thailand are shown, and in each province, the results of the model (prediction results) are illustrated in graphical and report forms with reference to the obtained R^2 values. To perform retrospective analyses of past situations, we added a feature that supports the analysis of results from the first case of confirmed infection in Thailand, as shown in **Figure 7**.

To create the map of Thailand, we used the Bokeh library in Python, and the map was combined with the COVID-19 prediction results. The latitude and longitude of all provinces in Thailand were used for clarity; for example, Bangkok has a latitude of 13.7278956 and a longitude of 100.52412349999997. When the user selects any province on the map, a pop-up shows the forecasted result and the corresponding accuracy (R^2). The map and example pop-up are shown in **Figure 8**.

If users want to obtain detailed prediction results, a graph display can be used in conjunction with the map display of Thailand (**Figure 9**). The predictions of new cases, total cases, new deaths, and total deaths can be displayed in graph form. The results are displayed for a given date selected by the user, and the first confirmed case of infection in Thailand is used as a historical reference for all visualizations.

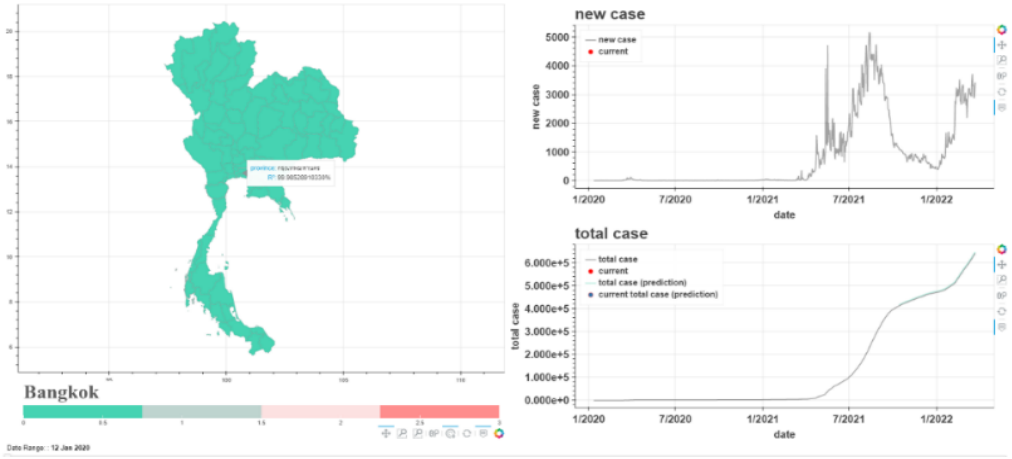

Figure 7 An example of a web application that shows the prediction results and uses a sliding bar for visualizing the spread of COVID-19 on the first day in Thailand.

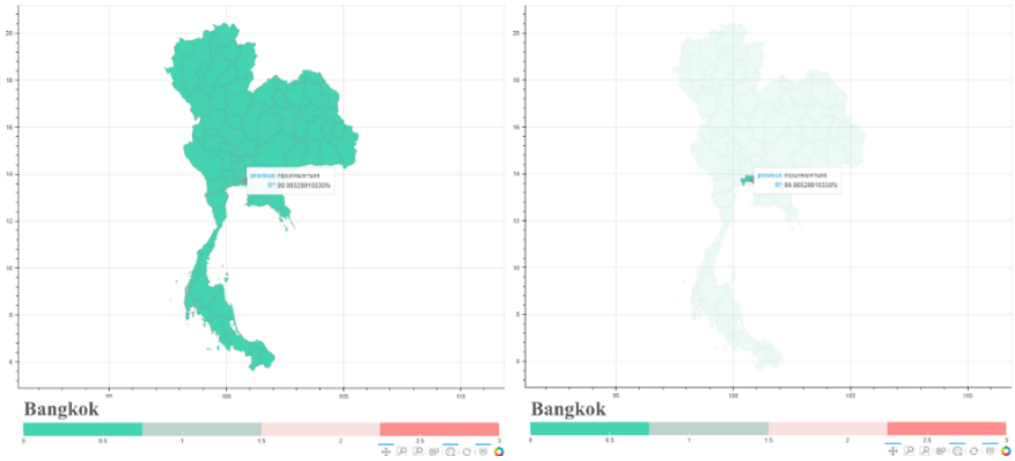

Figure 8 Map of Thailand, including latitude and longitude information.

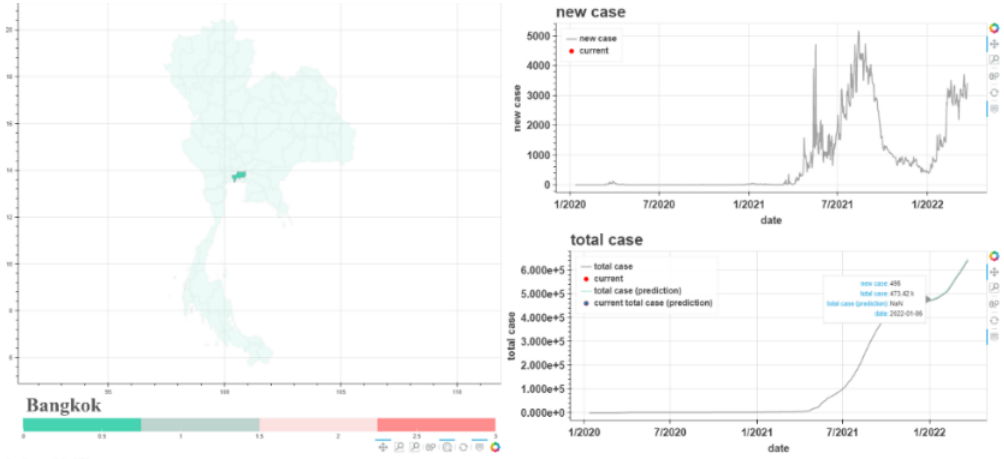

Figure 9 The prediction results for each province in Thailand.

For retrospective analysis of the COVID-19 epidemic in Thailand since the first confirmed case (on January 12, 2020), we created a slider bar to help users look back and obtain historical statistics, and they can select the time period they want to display. The forecast can also be obtained from this tab.

We divided our experimental process into 2 main cases based on the model structure: LSTM alone and LSTM combined with MLPs. For each of the main models (total cases), new cases were processed to obtain predictions. For accuracy comparison, adjustments were made. Finally, a total of 12 combinations were obtained for model evaluation, as shown in **Figure 10**. The modeling process for GIS data from Thailand was optimized, and the results were integrated for trend analysis and planning.

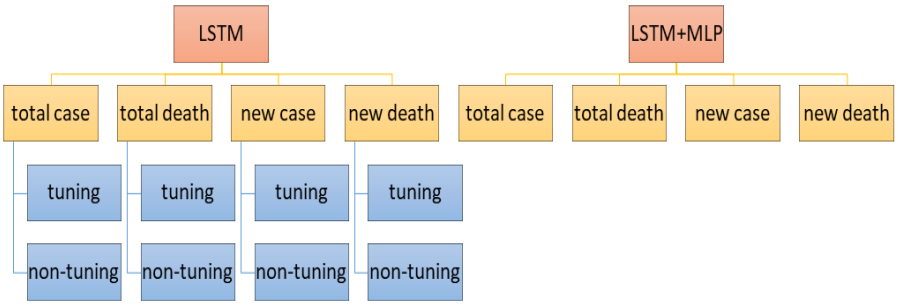

Figure 10 The 12 evaluated models.

Results and discussion

The models were evaluated and compared using different metrics, e.g., R^2, MSE and MAE, to find the optimal model for use in prediction. **Figures 11 - 16** show accuracy metric comparisons between 2 different models (adjusted vs. nonadjusted) in the same prediction scenario (total cases and total deaths) for each province.

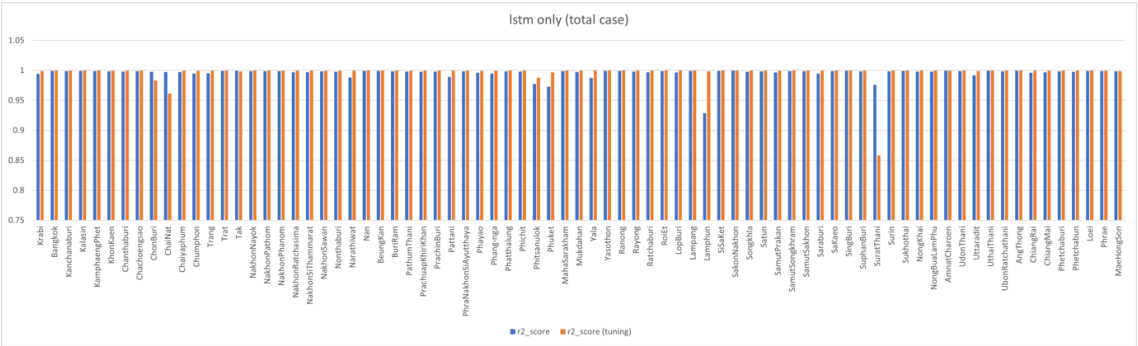

Figure 11 Comparison of the R^2 results of the nonadjusted and adjusted LSTM-only models (total cases) for all provinces.

Based on a comparison of the predictions of total cases obtained with the LSTM-only models (adjusted vs. nonadjusted) and the R^2 values obtained for the 77 provinces in the figure, the overall accuracy of the adjusted LSTM model is better than that of the nonadjusted model. The blue bar graph is the result for the nonadjusted LSTM model, and the orange bar graph is the result for the adjusted LSTM model.

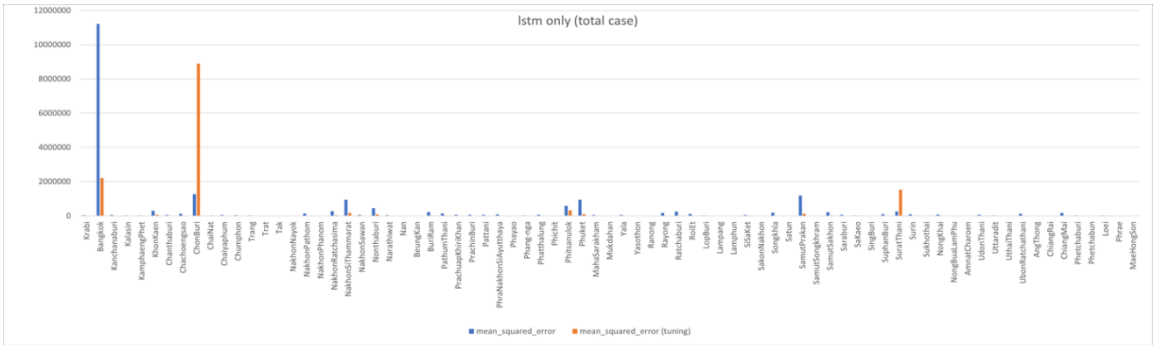

Figure 12 Comparison of the MSE results of the nonadjusted and adjusted LSTM-only models (total cases) for all provinces.

Figure 12 shows a comparison between the predictions of total cases with the LSTM-only models based on MSE obtained for each province. Notably, the overall error of the adjusted model is less than that of the nonadjusted model.

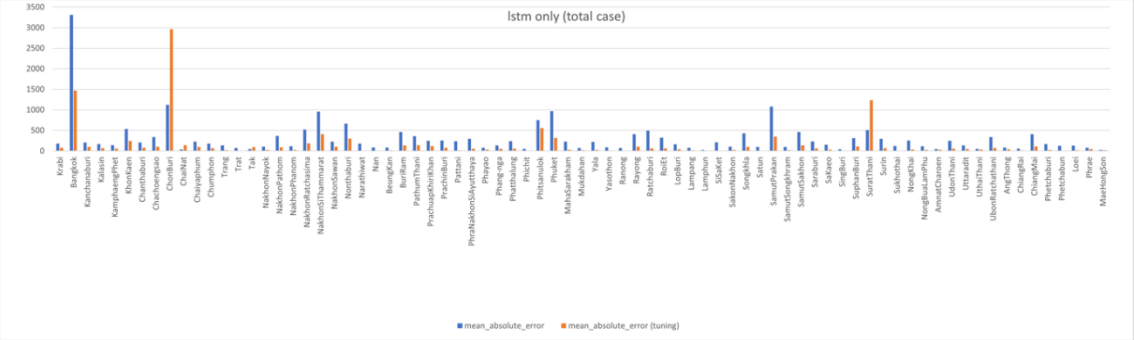

Figure 13 Comparison of the MAE results of the nonadjusted and adjusted LSTM-only models (total case) for all provinces.

Figure 13 shows a comparison between the LSTM-only models (prediction of the total number of cases) based on the MAE for each province. The overall accuracy of the adjusted LSTM model is better than that of the nonadjusted model.

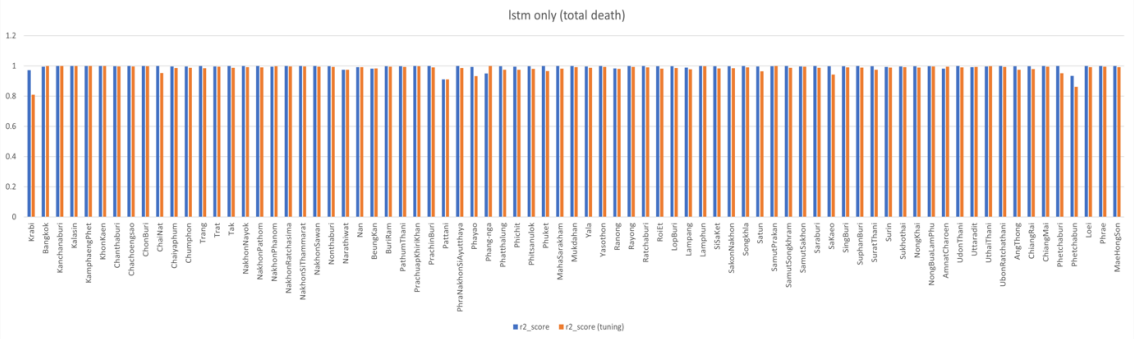

Figure 14 Comparison of the R^2 results of the nonadjusted and adjusted LSTM-only models (total deaths) for all provinces.

Figure 14 depicts a comparison of the total number of death predictions for the LSTM-only models (adjusted versus nonadjusted) based on R^2 for each of the 77 provinces. From the figure, the overall accuracy of the adjusted LSTM model is better than that of the nonadjusted LSTM model. The blue bar graph shows the results of the nonadjusted LSTM model, and the orange bar graph shows the results for the adjusted LSTM model.

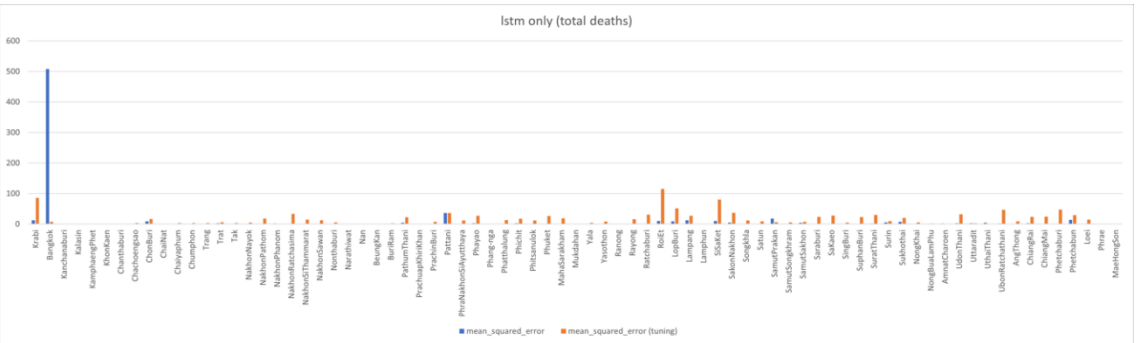

Figure 15 Comparison of the MSE results of the nonadjusted and adjusted LSTM-only models (total deaths) for all provinces.

Figure 15 shows a comparison of the total death predictions of the LSTM-only models based on the MSE for each province. The results show that the overall accuracy of the nonadjusted LSTM model is better than that of the adjusted model. In other words, the overall error of the nonadjusted model is lower than that of the adjusted model. This difference is due to the data collection approach used in each case. As shown in the figure, for provinces with high numbers of people (residents or tourists), such as Bangkok or Samutprakarn, the accuracy of the adjusted model improves with the amount of data available.

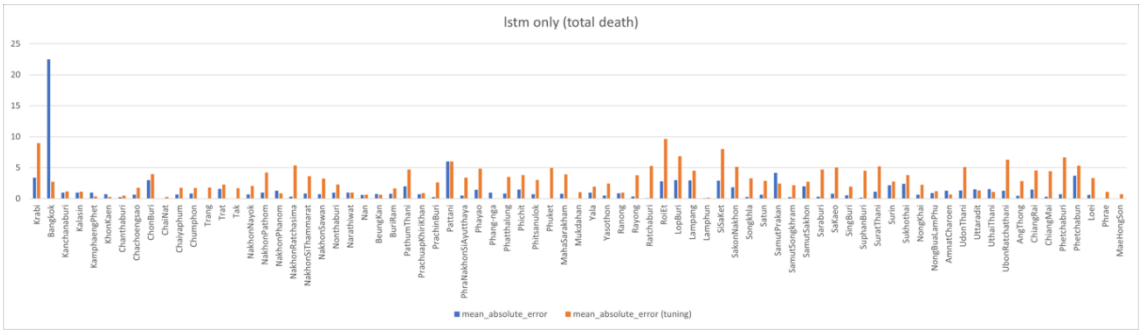

Figure 16 Comparison of the MAE results of the nonadjusted and adjusted LSTM-only models (total deaths) for all provinces.

Figure 16 shows a comparison of the total death predictions of the LSTM-only models based on the MAE for each province. The result shows that the overall error of the adjusted model is higher than that of the nonadjusted model due to the data collection process used.

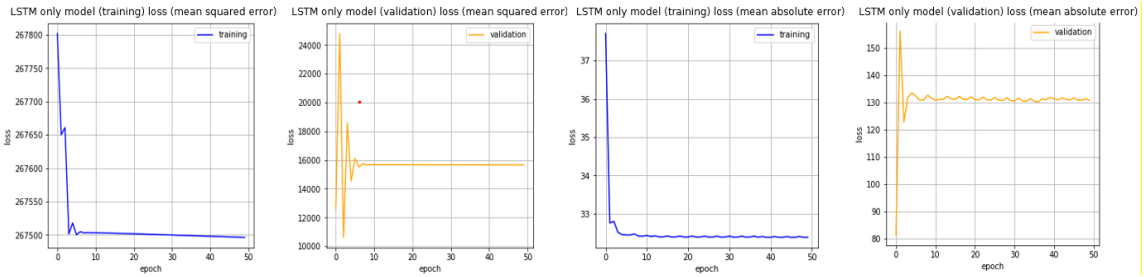

Figure 17 LSTM-only model loss in the prediction of total cases.

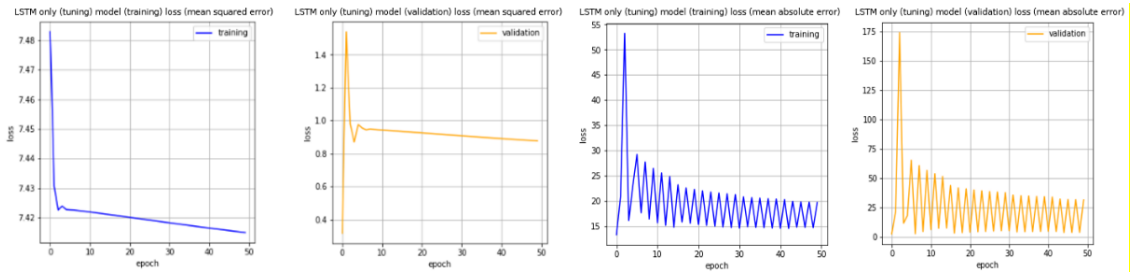

Figure 18 LSTM-only model loss in the prediction of total deaths.

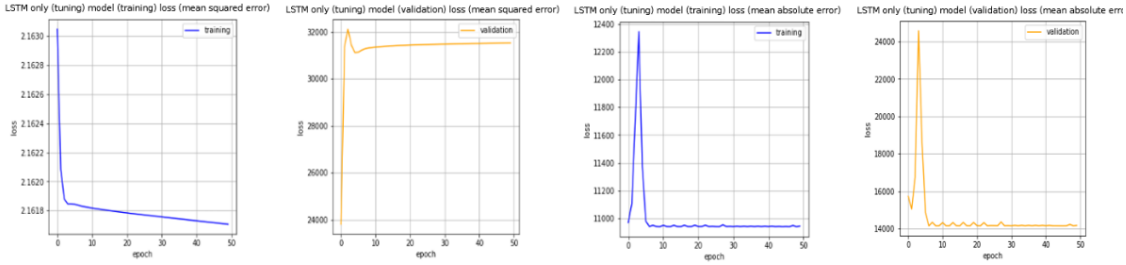

Figure 19 LSTM-only model (adjusted) loss for the prediction of total cases.

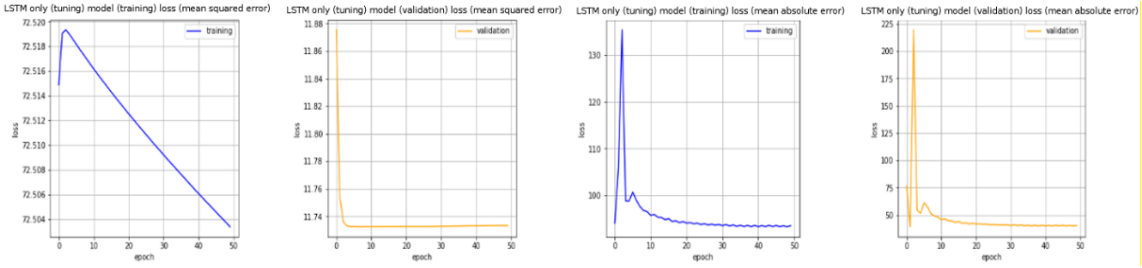

Figure 20 LSTM-only model (adjusted) loss for the prediction of total deaths.

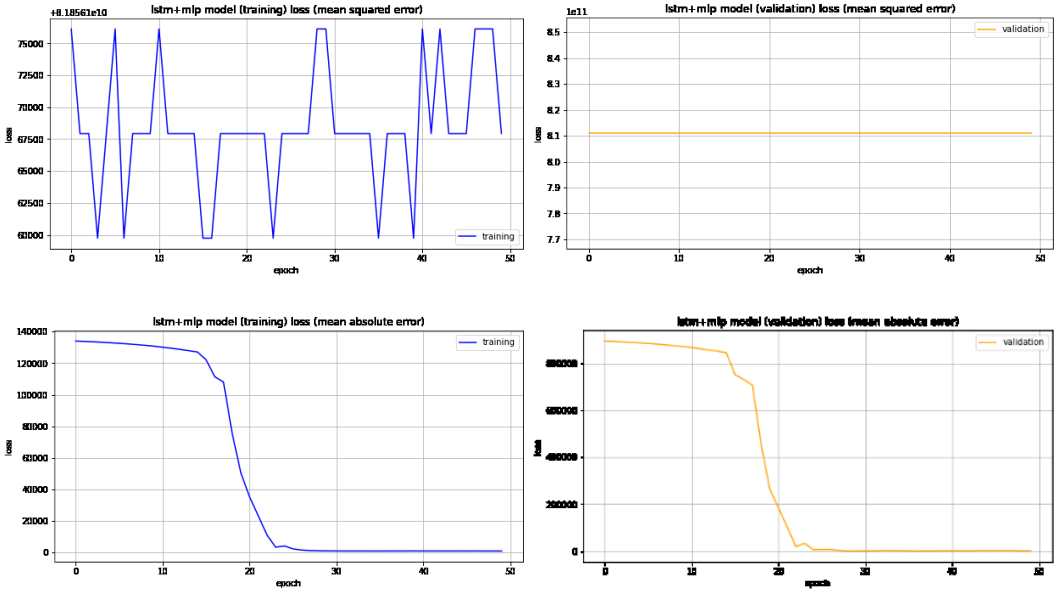

Figure 21 LSTM+MLP loss for the prediction of total cases.

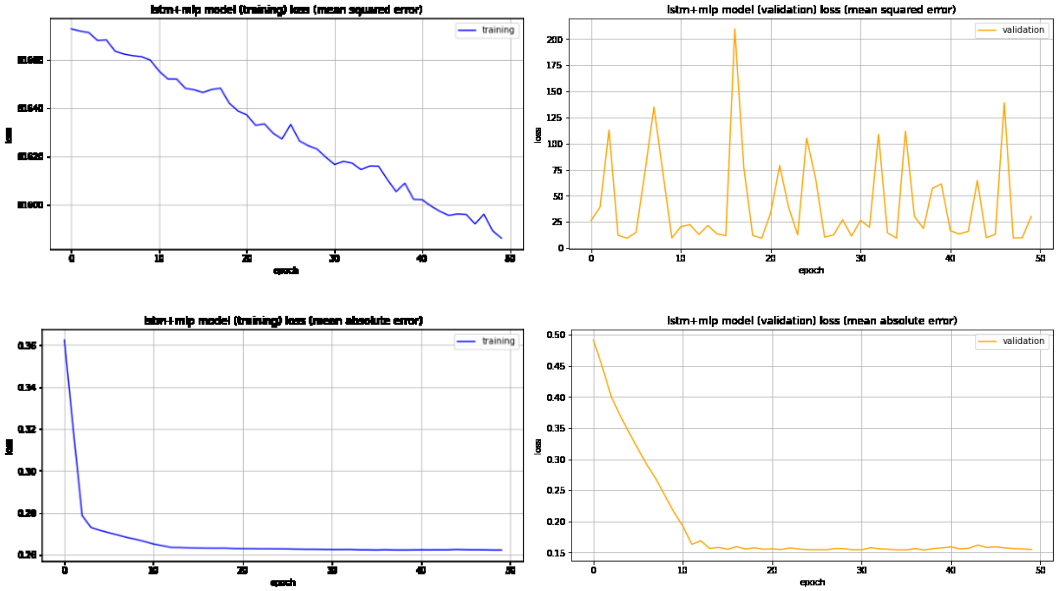

Figure 22 LSTM+MLP loss for the prediction of total deaths.

Figures 17 - 22 show the training and validation loss for all models. Specifically, **Figure 17** shows the training and validation loss based on the MSE and MAE for the nonadjusted LSTM-only model (prediction of total cases).

The training and validation loss based on the MSE and MAE of the nonadjusted LSTM-only model (prediction of total deaths) is shown in **Figure 18**.

Figure 19 shows the training and validation loss based on the MSE and MAE for the adjusted LSTM-only model (prediction of total cases).

The training and validation losses of the adjusted LSTM-only model (total deaths) based on the MSE and MAE are shown in **Figure 20**. Additionally, the training and validation loss of the adjusted LSTM+MLP model (total cases) based on the MSE and MAE is shown in **Figure 21**. The training and validation of the LSTM+MLP model (total deaths) based on the MSE and MAE is shown in **Figure 22**.

Table 2 Results of the experiments.

Models	Prediction	R^2	Mean squared error	Mean absolute error
LSTM (only)	total cases	0.9959	27,3328.0636	300.9691
LSTM (only)	total deaths	0.9944	9.5235	1.4257
LSTM (only) (adjusted)	total cases	0.9970	18,1981.4418	150.1811
LSTM (only) (adjusted)	total deaths	0.9826	16.0260	3.0948
LSTM+MLP	total cases	0.9980	11,421.3598	100.6024
LSTM+MLP	total deaths	0.9964	8.0124	1.0406

From the table of experimental results of the models used to predict total cases and total deaths (**Table 2**), the LSTM model without adjustment yielded an average R^2 of 0.9915, and the average R^2 after adjustment was 0.9898. The most accurate model in this experiment was LSTM+MLP, with an R^2 of 0.9972, and the average MSE and MAE values were 5714.6861 and 50.8215, respectively. From the experimental results, the adjusted LSTM (only) model outperforms the nonadjusted model in the prediction of total cases based on every metric. Finally, LSTM+MLP outperformed all models based on all metrics. Since LSTM+MLP yielded the highest accuracy, we adopted the LSTM+MLP approach for final predictions.

The reason why LSTM+MLP is more accurate than LSTM+MLP is because the vaccination data that were added through the MLP, making the model capable of handling more diverse data. However, features should be added to provide a variety of dimensions to reasonably consider other factors contributing to the spread of COVID-19.

Based on experiments with different models used to predict COVID-19 trends, LSTM is appropriate for time series analysis, and MLPs are best for static data. The experimental results can be used to support plans and goals that have been set. However, the inaccuracy of some models in some provinces, such as Bangkok or Samutprakarn, may hinder further application. Additionally, data for other factors that could affect the spread of COVID-19 were not considered. The inclusion of these data could the model more comprehensive. Thus, we are actively attempting to collect continuous information in Thailand to further enhance modeling and mapping capabilities.

When comparing our presented model with other state-of-the-art models, our model's performance is superior (highest accuracy). However, each method may have different precision that may depend on the variety of the dataset used and various feature selection methods. In particular, our model is most suitable for COVID-19 case data in Thailand.

Conclusions

To create a tool to predict the spread of COVID-19 by using deep learning techniques with GIS in all 77 provinces of Thailand, we proposed an integrated LSTM and MLP model. LSTM was used to process dynamic time series, and the MLP model was applied to static data. The results showed that the best-performing model was LSTM+MLP, with an accuracy of 99.72 %, a 3.72 % improvement over other state-of-the-art methods, as shown in **Table 3**. When compared to all 5 state-of-the-art models, the developed models are the most accurate. However, based on the different conditions in different study areas and the datasets used, the best method may vary in a given scenario.

Table 3 Comparison of the results of state-of-the-art methods.

Method	Model & Algorithm	Accuracy	Embedded GIS
[20]	• SUPER LEARNER ENSEMBLE	87 %	YES
[21]	• LSTM • MARKOV	96 %	YES
[22]	• RBF • FCM • NARX	81.6 %	YES
[23]	• GPR • SVM • DECISION TREE	95 %	YES
[24]	• BAYESIAN • ADABOOST • SVM • KNN • RANDOM FOREST • C4.5 • ANN	95.24 %	YES
The proposed model	LSTM+MLP	99.72 %	YES

Our work has a key limitation: The dataset is limited to enhance model efficiency. It is necessary to obtain a larger dataset and more features from a reliable source for comprehensiveness modeling and analysis. The proposed tool can help people analyze data and obtain daily reports from the agencies involved in monitoring the COVID-19 situation in Thailand; moreover, it is a decision-making tool that can be used to prevent, manage, and plan for different COVID-19 scenarios. We can use the established model to analyze COVID-19 trends, and we hope that our research will benefit people influenced by the COVID-19 epidemic. In addition, we expect that we will be able to use the developed model to anticipate or analyze data from potential pandemic outbreaks, including those that may arise from monkeypox, acute hepatitis, and H3N8 bird flu or noncommunicable diseases (NCDs). However, future model implementations need to consider the compatibility between the data and model.

Acknowledgements

The author would like to thank the National Broadcasting and Telecommunication Commission (NBTC), Thailand, and the Research and Diagnostic Center for Emerging Infectious Diseases (RCEID), Khon Kaen University, Thailand for supporting this paper.

References

[1] JH Zeng, YX Liu, J Yuan, FX Wang, WB Wu, JX Li, LF Wang, H Gao, Y Wang, CF Dong, YJ Li, XJ Xie, C Feng, L Liu. First case of COVID-19 complicated with fulminant myocarditis: A case report and insights. *Infection* 2020; **48**, 773-7.

[2] R Strugnell and W Nancy. Sustained neutralising antibodies in the Wuhan population suggest durable protection against SARS-CoV-2. *Lancet* 2021; **397**, 1037-9.

[3] YC Li, B Wan-Zhu and H Tsutomu. The neuroinvasive potential of SARS-CoV2 may play a role in the respiratory failure of COVID-19 patients. *J. Med. Virol*. 2020; **92**, 552-5.

[4] Y Zoabi, S Deri-Rozov and Noam Shomron. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. *NPJ Digit. Med.* 2021; **4**, 3.

[5] WHO Coronavirus (COVID-19) Dashboard, Available at: https://covid19.who.int/, accessed December 2022.

[6] K Tantrakarnapa, B Bhopdhornangkul and K Nakhaapakorn. Influencing factors of COVID-19: A case study of Thailand. *J. Publ. Health* 2022; **30**, 621-7.

[7] B Leerapan, P Kaewkamjornchai, R Atun and MS Jalali. How systems respond to policies: Intended and unintended consequences of COVID-19 lockdown policies in Thailand. *Health Pol. Plann*. 2022; **37**, 292-3.

[8] K Pornpatanapaisansakul. *FAQ Issue 171: Labor market digital transformation*. Bank of Thailand, Bangkok, Thailand, 2020.
[9] N Anusonphat and C Poompurk. New normal ways for the survival and adaptation of community enterprise after COVID-19 crisis of Thailand. *J. MCU Nakhondhat* 2022; **9**, 1-19.
[10] Ministry of Public Health. The report of COVID-19 situation in Thailand, Available at: https://ddc.moph.go.th/covid19-dashboard/, accessed December 2022.
[11] S Montaha, S Azam, AKMRH Rafid, Z Hasan, A Karim. TimeDistributed-CNN-LSTM: A Hybrid approach combining CNN and LSTM to classify brain tumor on 3D MRI scans performing ablation study. *IEEE Access* 2022; **10**, 60039-59.
[12] Z Cheng, Z Yi and T Chengkai. Solving monocular sensors depth prediction using MLP-based architecture and multi-scale inverse attention. *IEEE Sensor. J.* 2022; **22**, 16178-89.
[13] O Tutsoy, Ş Çolak, A Polat and Kemal Balikci. A novel parametric model for the prediction and analysis of the COVID-19 casualties. *IEEE Access* 2020; **8**, 193898-906.
[14] Y Zhu, S Wang, S Wang, Q Wu, L Wang, H Li, M Wang, M Niu, Y Zha and J Tian. Mix contrast for COVID-19 mild-to-critical prediction. *IEEE Trans. Biomed. Eng.* 2021; **68**, 3725-36.
[15] S Tabik, A Gómez-Ríos, JL Martín-Rodríguez, I Sevillano-García, M Rey-Area, D Charte, E Guirado, JL Suárez, J Luengo, MA Valero-González, P García-Villanova, E Olmedo-Sánchez and F Herrera. COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on chest X-ray images. *IEEE J. Biomed. Health Informat*. 2020; **24**, 3595-605.
[16] C Moremada et al. Energy efficient contact tracing and social interaction-based patient prediction system for COVID-19 pandemic. *J. Comm. Network*. 2021; **23**, 390-407.
[17] D Vekaria, A Kumari, S Tanwar and N Kumar. *ξboost*: An AI-based data analytics scheme for COVID-19 prediction and economy boosting. *IEEE Internet Things J*. 2020; **8**, 15977-89.
[18] N Saeed, A Bader, TY Al-Naffouri and MS Alouini. When wireless communication responds to COVID-19: Combating the pandemic and saving the economy. *Front. Comm. Network.* 2020; **1**, 566853.
[19] OA Ehiorobo. Strategic agility and AI-enabled resource capabilities for business survival in post-COVID-19 global economy. *Int. J. Informat. Bus. Manag.* 2020; **12**, 201-13.
[20] D McCoy, W Mgbara, N Horvitz, WM Getz and Alan Hubbard. Ensemble machine learning of factors influencing COVID-19 across US counties. *Sci. Rep.* 2021; **11**, 11777.
[21] R Ma, X Zheng, P Wang, H Liu and C Zhang. The prediction and analysis of COVID-19 epidemic trend by combining LSTM and Markov method. *Sci. Rep.* 2021; **11**, 17421.
[22] BM Yahya, FS Yahya and RG Thannoun. COVID-19 prediction analysis using artificial intelligence procedures and GIS spatial analyst: A case study for Iraq. *Appl. Geomat.* 2021; **13**, 481-91.
[23] FM Khan, A Kumar, H Puppala, G Kumar and R Gupta. Projecting the criticality of COVID-19 transmission in India using GIS and machine learning methods. *J. Saf. Sci. Resilience* 2021; **2**, 50-62.
[24] B Niu, R Liang, S Zhang, H Zhang, X Qu, Q Su, L Zheng and Q Chen. Epidemic analysis of COVID-19 in Italy based on spatiotemporal geographic information and Google trends. *Transbound. Emerg. Dis.* 2021; **68**, 2384-400.