1. Thanpitcha Atiwanwong[1], 2. Adirek Jantakun[1], 3. Adisak Sangsongfa[2]

Rajamangala University of Technology Isan (1), Rajabhat Mu-Ban Chombueng University (2)
ORCID: 1.0000-0001-9388-0390; 2.0000-0001-8474-9315; 3. 0000-0002-6231-4000

doi:10.15199/48.2023.07.17

Comparative Analysis of Optimization Value Between Artificial Neural Network and Long Short-Term Memory for Prediction Particulate Matter (PM2.5) in Bangkok Thailand

***Abstract**. The crisis that the general public is worried about is particulate matter as small as 2.5 microns, which is invisible to the naked eye, causing a great lack of awareness of health hazards. One of the key goals and visions of government leaders around the world is to tackle PM2.5 particulate matter, but without measurements, reports and predictions, how will it lead to emission reduction and remedial steps? Therefore, the prediction of PM2.5 is considered as the main factor that will help to reduce the pollution of PM2.5. So, Neural networks have been widely used in predictive research, but the problem is What type of neural network would be most suitable for predicting the value of PM2.5? In this research, the predictions were compared between Artificial Neural Network (ANN) and Long Short -Term Memory (LSTM) using values measured from the performance test results with accuracy. The results showed that when the values of similar hyperparameters were given different results, the average ANN accuracy is 91.1460%. The average accuracy LSTM is 96.8496%. The values obtained from the comparison clearly show that for the prediction of PM2.5, the LSTM neural network was significantly more suitable than the ANN neural network.*

***Streszczenie.** Kryzys, który niepokoi opinię publiczną, to pył zawieszony o wielkości zaledwie 2,5 mikrona, który jest niewidoczny gołym okiem, powodując ogromny brak świadomości zagrożeń dla zdrowia. Jednym z kluczowych celów i wizji przywódców rządów na całym świecie jest rozwiązanie problemu pyłu zawieszonego PM2,5, ale bez pomiarów, raportów i prognoz, w jaki sposób doprowadzi to do redukcji emisji i działań zaradczych? Dlatego prognoza PM2,5 jest uważana za główny czynnik, który pomoże zmniejszyć zanieczyszczenie PM2,5. Tak więc sieci neuronowe były szeroko stosowane w badaniach predykcyjnych, ale problem polega na tym, jaki typ sieci neuronowej byłby najbardziej odpowiedni do przewidywania wartości PM2,5? W tym badaniu porównano przewidywania między sztuczną siecią neuronową (ANN) a pamięcią długokrótkoterminową (LSTM) przy użyciu wartości zmierzonych z wynikami testu wydajności z dużą dokładnością. Wyniki pokazały, że przy różnych wartościach podobnych hiperparametrów średnia dokładność ANN wynosi 91,1460%. Średnia dokładność LSTM wynosi 96,8496%. Uzyskane z porównania wartości jednoznacznie wskazują, że do predykcji PM2,5 sieć neuronowa LSTM okazała się znacznie bardziej odpowiednia niż sieć neuronowa ANN. (**Analiza porównawcza wartości optymalizacji między sztuczną siecią neuronową a długoterminową pamięcią dla przewidywania cząstek stałych (PM2,5) w Bangkoku w Tajlandii**)*

Keywords: Artificial Neural Network, Long Short-Term Memory, Comparative Neural Network
Słowa kluczowe: Sztuczna sieć neuronowa, pamięć krótkotrwała, porównawcza sieć neuronowa

Introduction

Environmental researchers and the United Nations Environment Program (UNEP) reveal that PM2.5 air pollution is strongly associated with global warming. The United State Environmental Protection Agency has established a standard for fine particulate matter in the air that is harmful to human health by using PM (Particulate Matters) PM 2.5 dust, also known as “Final dust”. Particles are particulate matter particles in the air with a diameter of 2.5 microns [1]. A study by the Institute for Health and Evaluation, University of Washington [2], found that air pollution is a common contributor to many diseases due to the presence of a wide range of chemicals, from irritants to carcinogens. cause disease, including COPD, cerebrovascular disease, ischemic heart disease, lung cancer, and acute infectious diseases, lower respiratory tract for ozone gas is a lung irritant. This makes the lungs easily infected, so it is a contributing factor to chronic obstructive pulmonary disease.

Based on the importance of the dangers caused by PM2.5 dust mentioned above, there are various types of research related to PM2.5 dust management such as PM2.5 dust prediction, PM2.5 dust classification, PM2.5 dust generation forecast, etc. From the research data related to PM2.5 dust pollution, most found that it is research related to prediction or forecasting related to PM2.5 dust by using different methods but aiming in the same direction. However, efforts have been made to predict or predict the cause of PM2.5 dust generation in order to adopt appropriate methods to prevent or determine the cause of dust generation. Therefore, the research example can be given as follows. Xianghong Wang et. al. [3] presented a research paper on predicting environmental floating dust and PM2.5 using an artificial neural network. The BP neural network design was used for the optimization of PM2.5 dust. The results were compared between a normal BP neural network and a BP neural network using a genetic algorithm for parameter optimization. The results were found to be better than the conventional BP neural network. Usually, the test accuracy is 96%. Sachit Mahajan et. al. [4] presented a research paper on improving the accuracy and efficiency of PM2.5 dust forecasting using an integrated neural network cluster model. This research uses a combination of Artificial Neural Networks (ANN) with a linear combination function and activation function. The experiments and evaluation are done using Airbox devices data from 557 stations deployed all over Taiwan. The results showed that the accuracy was satisfactory and the PM2.5 dust value could be predicted with less error. Ian G. McKendry [5] research on PM2.5 dust forecasting, evaluation of artificial neural networks for fine particulate pollution (PM10 and PM2.5) forecasting. The researchers used multi-layer perceptron (MLP) artificial neural network (ANN) models compared with traditional multiple regression (MLR) models for daily maximum and average O3 and particulate matter (PM10 and PM2.5) forecasting. Yanlin Qi [6] presented the research that used a hybrid model based on deep learning methods that integrate Convolutional neural network and Long Short-Term Memory networks (GC-LSTM) to model and forecast the spatiotemporal variation of PM2.5 concentrations. Researchers made random division on spatiotemporal blocks (each block consists of a T graph in time ordering), rather than original observations. The researcher proposed this methodology can be used for concentration forecasting of different air pollutants in the future. Dewen Seng et. al. [7] presented their research which is a spatiotemporal prediction of air quality based on an LSTM neural network.

This research is based on long short-term memory (LSTM), a comprehensive prediction model with multi-output and multi-index of supervised learning (MMSL) was proposed. From the dataset, the air quality time series datasets collected from 35 air quality monitoring stations in Beijing from January 1, 2016, to December 31, was7, were used to validate the performance of the model compared with other baseline models and the two most advanced models. The results showed that overall, the performance of the developed model was MAE, and RMSE better than the base model such as SVM, ARMA, and MMSL. Moreover, some researchers have done research on creating highly efficient but inexpensive tools to collect PM2.5 dust. Thanpitcha Atiwanwong and Saweth Hongprasit [8] have done research on a low-power real-time pollution monitoring system using ESP LoRa to collect PM2.5 dust, it can perform more efficiently compared to expensive tools. The results from the experiment show that PM2.5 dust can be measured no differently from expensive instruments and can report PM2.5 dust in real life. time through the webpage which makes it possible to know the PM dust in real-time. The observations from the above-mentioned research on PM2.5 particulate matter are particularly important as it is a matter of importance for human health and most studies attempt to predict the incidence of PM2.5 particulate matter. More than that, efforts are being made to forecast the occurrence of PM2.5 dust in order to find ways to prevent the harmful effects of PM2.5 dust on humans. Syed Ahsin Ali Shah et. al. [9] presented a research paper on the forecasting of PM2.5 particulate matter A hybrid model for forecasting of particulate matter concentrations based on multiscale characterizations and machine learning (ML) techniques using machine multi-method techniques. learning combined with multiscale characterization, for example, random forest, support vector regressor, k-nearest neighbours, feed-forward neural network, and AdaBoost. The algorithm used by the researchers used the empirical mode decomposition. The developed algorithm compared with the standard ML method found that the RMSE value of PM2.5 was 4.81 and MAE was 3.02, which was effective in predicting the occurrence of PM2.5 particulate matter in order to remind people to protect themselves.

According to the above-mentioned research studies on the prediction or forecasting of PM2.5 dust, it was found that most researchers used neural networks for their research. In particular, use ANN or LSTM for prediction or forecasting of PM2.5 dust. Therefore, it can be concluded initially that neural networks are well suited to use in dust prediction or forecasting. But more importantly, the question arises as to which ANN and LSTM neural networks are most suitable for predicting or forecasting PM2.5 dust better? An artificial neural network (ANN) [10] is the creation of a computer with a model that simulates the way the human brain works, or is it to enable the computer to think and remember, or to use the neural network to enable the computer to recognize and understand human language. The working principle of a neural network is that in a computer, a neural network consists of the inputs that come to the network and multiply that input by the weight of each network. Compare that to the threshold which is defined. Therefore, ANN is to make computers recognize things such as images, handwriting, data for use in forecasting or forecasting the weather, stock forecasts. An example of research using ANN for PM2.5 dust forecasting is as follows: X. Feng and et. al. [11] presented his research on Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation. The research findings were to use a hybrid model of air mass analysis and wavelet conversion to improve the accuracy of PM2.5 dust forecasting. Daily average concentration in two days in advance Data used for forecasting were obtained from 13 air pollution monitoring stations in Beijing, Tianjin and Hebei provinces (Jing-Jin-G area). The results showed that the root mean square error (RMSE) of hybrid models can be reduced by an average of 40 percent. In particular, a high day's PM2.5 value was most predictable on a basis of approximately 90% average based on wavelet decay and detection rate (DR). This high average was achieved through the development of this model in conjunction with the ANN dust prediction, which is highly interoperable. [12] presented research using a backpropagation neural network that optimizes parameters by using an evolutionary algorithm (Genetic Algorithms). The results show that neural networks optimized by genetic algorithms have better performance in predicting PM2.5 mass concentrations, thereby improving the accuracy of prediction results and reducing error rates etc. From the research example above, the researcher has chosen ANN for PM2.5 dust forecasting, but the ANN neural network may not be a very good choice because some researchers do not use the neural network. The original neural network was optimized for parameters for better forecasting results. In the other way, Long Short-Term Memory (LSTM) [13] is a type of Recurrent Neural Network (RNN). That is a neural network that reuses its previous output and is a type of neural network designed for sequence processing. This is a behavior required in complex problem domains like machine translation, speech recognition, and more. It can be said that the LSTM is a deep learning model created to simulate the human memory model with limited memory capacity. When new events enter the memory, the brain chooses to Accept or not accept new events into memory according to the importance of events and when the brain chooses to accept new events that are important to be stored in the memory system. However, it is necessary to have some past events that are forgotten in order not to fill the memory. However, the advantages of LSTM can be applied in a wide range of applications such as image captioning, machine translation, language modeling, handwriting generation and question answering chatbots, etc. It demonstrates its ability and variety of applications. Because LSTM brings together the advantages of RNN [14].

From the research mentioned above, it was found that the prediction or forecasting of PM2.5 dust is most widely used by researchers, whether it is ANN or LSTM. However, from the previous research, the use of both neural networks was not much different. Thus, it was observed that which type of neural network yielded better prediction results. Therefore, in this research, the researcher chose to use the prediction of PM2.5 dust in Bangkok, Thailand as the data for prediction and used both ANN and LSTM neural networks to make predictions to compare the predictions of the neural networks. Which type gives the most suitable prediction result? The results obtained from the prediction will determine which neural network between ANN and LSTM is more suitable for predicting PM2.5 dust. This will enable the best neural network to be developed in the future.

2. Related Work

A. Long Short-Term Memory Network: LSTM

In the description of LSTM, it is necessary to mention Recurrent Neural Network (RNN) first because LSTM is actually a type of RNN but LSTM is an RNN [15]. More details have been added to address some of the weaknesses of the RNN. In order to understand the main idea of RNN, it is created to work with sequential data such as video (sequence of images) or text (sequence of words). However, in order to understand the way of working RNN is easier Let's give an example of doing sequential data in the beginning. Compared to human reading, which may have a question about how to read a book, it is sequential data. It can be easily explained that the sequence of words is that when we read a book, we will read one word at a time from left to right (for Thai or English). So that we can know what the sentence we are reading is related to, we mix the story from what we have read (let's call it the hidden state or previous state) and the word we just read (which is input data or the word we are reading at that time). It helps us to understand the meaning of the sentence in the right part that we are reading at that moment. So, RNN uses the same principle of neural modulation. the original network to be able to add the previous state or knowledge to the input The new data that comes in to understand the data that has been read over and over. Therefore, the two important parts of the RNN are 1) the previous hidden state and 2) the input data at that time. However, RNNs have problems with data sequences being too long, which when concatenated can cause insufficient memory. Therefore, unnecessary words should be forgotten, but the problem is how to choose the words to forget, what words should be remembered? As a solution to RNN's problem with long sequences of data, the use of Long Short-Term Memory (LSTM) was proposed by H. Sepp and S. Jurgen [16]. In LSTM, it's like a type of RNN, but it's slightly more detailed: RNN is like a neural network with simple memory inside to save the previous hidden state. M. Chovatiya and et. al. [17] The function of remembering the long periods of time data is the core function of LSTM, it has a performance to record old data and current input data by internal memory, and it can show when to write, forget (delete), or allow to read. As shown, the operation can be seen in Figure 1.

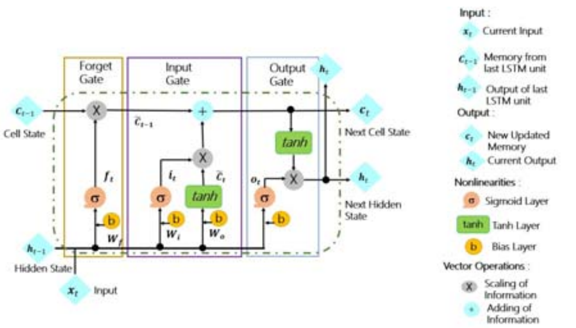

Fig.1. The structure of LSTM

Form Figure 1, Greff, K and et. al. [18] the function of LSTM can be explained as follows: Cell state is a storage state of memory cell in LSTM and Gate. It is a data flow controller, an analog value that controls when to read, write, or forget. It is like a gateway to see when data should be allowed to flow in, flow out or disappear (forget). To create this, forget gate, we consider the input data attributed to the previous hidden state (RNN formula) to make a decision by using the sigmoid function as the deciding equation.

$$(1) \qquad f_t = \sigma \left(W_{xf} x_t + W_{hf} h_{t-1} + W_{cf} c_{t-1} + b_f\right)$$

When a new input data arrives, the question is whether to update the cell state or not. This question is governed by a so-called input gate, here it still uses the sigmoid function to decide whether to allow updates or not. This uses the input data value that came with the hidden state before it according to the equation:

$$(2) \qquad i_t = \sigma \left(W_{xi} + x_t + W_{hi} h_{t-1} + b_i\right)$$

If there is a need to adjust the value that needs to be modified, what is the best value to use? This cycle uses what is called input modulation gate handled by the equation, similar to input gate, but uses a tanh function instead of the resulting value. will be regarded as cell state candidate according to the following equation:

$$(3) \qquad g_t = tanh(W_{xc}\, x_t + W_{hc} h_{t-1} + b_c)$$

Where tanh (Tanh Function), or full name is Hyperbolic Tangent Activation Function, is a function that solves many disadvantages of Sigmoid function, but the shape is the same S. Given the data from forget gate, input gate and input modulation gate, we combine all three to modulate cell state, so the equation for combining everything is as follows equation:

$$(4) \qquad C_t = f_t \cdot C_{t-1} + i_t \cdot g_t$$

From the RNN, the hidden state must be removed at the time of t or h_t, which at the time of t_1 this LSTM takes this h_t value (by the formula). So, the word read in it is to allow external devices to read h_t or not (actually like setting a certain permission that can be seen or not) or is it? It is kept unexpressed and does not return h_t as a result. Here there will be output gate to help decide. It will continue to use the same formula with forget gate and input gate. Use the sigmoid function with the previous hidden state value and input data that work together as shown below equation:

$$(5) \qquad O_t = \sigma(W_{xo} + x_t + W_{ho} h_{t-1} + b_o)$$

The value that will be exported to h_t for the next sequence will get the following equation:

$$(6) \qquad h_t = o_t \cdot tanh(c_t)$$

If output gate gives $\boldsymbol{o_t}$ a value of $\boldsymbol{O}$, then $\boldsymbol{h_t}$s value is $\boldsymbol{O}$, meaning nothing is sent, meanwhile, if $\boldsymbol{o_t}$ is 1, we calculate ht and send it outside. In other words, it allows other devices to see the $\boldsymbol{h_t}$ value [19].

B. Artificial Neural Network: ANN

Neural network are algorithms that are based on brain activity and are used to simulate complex patterns and predictions to solve problems. Artificial neural networks are deep learning methods that emerge from the concept of biological neural networks in the human brain. The development of the ANN was the result of an attempt to simulate the functioning of the human brain. The ANN function is very similar to the function of biological neural networks, although they are not identical. Whereas ANN's algorithm accepts only numeric data and structured data.

The architecture of a neural network has three layers: input layer, hidden layer. However, if the neural network has multiple layers, it is also known as MLP (Multi-Layer Perceptron), as shown in Figure 2. Therefore, it is possible to think of a hidden layer, also known as a "distillation layer", which extracts some of the most relevant patterns from the input and sends them to the next layer for further analysis. It increases and improves network performance by recognizing only the most important information from the input and eliminating redundant information. So, the activation function is important for two reasons: first, it allows you to turn on your computer, and second. This model captures the non-linear relationship between inputs and contributes to a more convenient input-to-output conversion as shown in Figure 3.

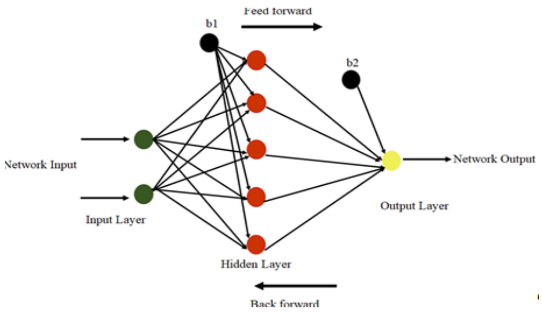

Fig.2. Artificial Neural Network Architecture

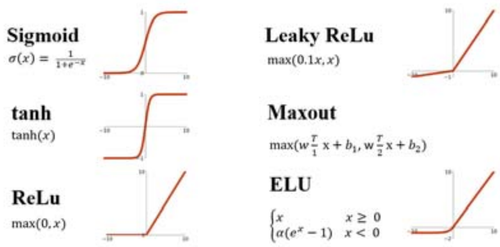

Fig.3. The Activation Function of ANN

From Figure 3, it can be described as follows. The “optimal value of W (weight)” that minimizes errors in predictions is crucial to successful modelling Backpropagation algorithm does this by converting ANN to a learning algorithm using learning from errors itself. In order to focus on the optimization of the neural network, the technique of gradient descent is used to quantify the error in predicting the outcome. To find the optimal value for W, try adjusting the W value slightly and examine the effect on the prediction error. Finally, those W values are chosen as the ideal because further changes to W do not reduce errors [20].

3. Methodology

From the research framework shown in Figure 4, the main research objective is to design the most suitable neural network parameters to obtain the best value by the neural network designed. A functional comparison method was used between ANN and LSTM neural networks with identical parameters tuned to obtain the optimal neural network. that will be further developed to use the most suitable artificial neural network for forecasting the value of PM2.5 dust. However, the conceptual framework for research can be divided into the following sections:

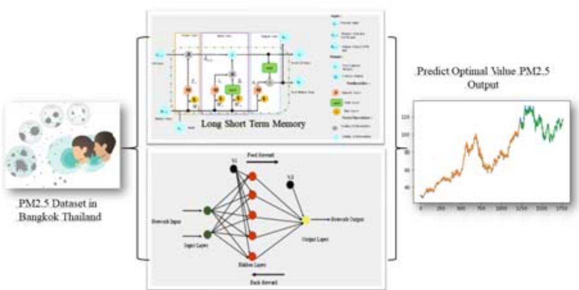

Fig.4. Framework of Comparative the Optimal Value Between ANN and LSTM for Prediction PM2.5

(1) Data Preparation:

The data preparation process is divided into 3 steps:

- Data cleaning is the process of extracting information from a database that has been disturbed by unwanted nuisances, resulting in information that can be used in further predictive processes such as hourly incomplete data. Nonconforming data, etc., remove these data from the database, resulting in information that can be used in the future.
- Selection of PM2.5 particulate matter data obtained from the Pollution Control Department at 12 stations from air monitors in Bangkok, the researcher selected all 7 stations to be representatives of air monitoring stations in Bangkok. Based on the location criteria of the weather measurement stations spread over the Bangkok area as a selection criterion. All stations used historical data from August 2019–December 2021. The total duration of each station is 2 years and 5 months, details are shown in Figure 5. List of all 7 weather stations in Bangkok that are representative stations.

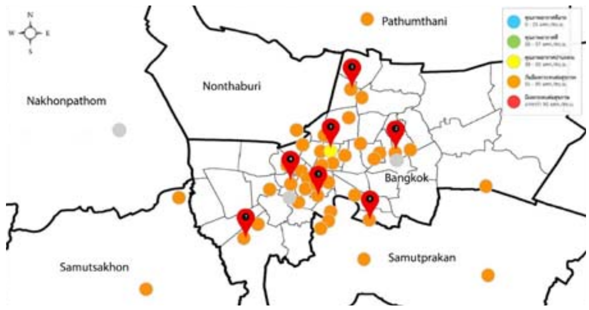

Fig.5. Map of 7 representative weather stations in Bangkok

1) Chulalongkorn Hospital Station (ST01)
2) Khlong Chan Community Housing Station (ST02)
3) Public Relations Department Station (ST03)
4) MEA Substation Thonburi Station (ST04)
5) Nonsiwitthaya School Station (ST05)
6) Bangna Meteorological Department Station (ST06)
7) Station on Kanchanaphisek Road, Bang Khun Thian District (ST07)

- Data transformation is the process of formatting non-uniform data into the same format so that it can be used in further prediction steps, e.g., String data must be converted to integer, etc.

Where all the seven stations in the above use a total of 20,000 records and use the station on the roadside Din Daeng Housing Authority totaling 40,000 records. Figure 6 shows a sample of the data used in the research.

1	Y/M/D	Hour	PM2.5
2	190820	1700	9
3	190820	1800	5
4	190820	1900	3
5	190820	2000	8
6	190820	2100	13
7	190820	2200	12
8	190820	2300	14
9	190820	2400	13
10	190821	100	18
11	190821	200	20
12	190821	300	17
13	190821	400	7
14	190821	500	6
15	190821	600	10

Fig.6. Example PM2.5 dataset from Khlong Chan Community Housing Station in Bangkok

(2) Measuring the efficiency of the neural network:

In this research, the researcher chose to measure the efficiency of the neural network from two types: accuracy, mean absolute error (MAE) and mean squared error (MSE), as detailed below.

- Prediction accuracy is the ratio of the number of prediction accuracy to the total number of predictions. The equation is as follow:

$$(7)\quad prediction\ accuracy = \frac{correct\ prediction}{total\ prediction} x\ 100\%$$

- Mean Absolute Error (MAE) Mean absolute error is the mean of the absolute difference between the forecast value and the If the MAE value is small, then the model can estimate close to the real value. The equation is as follows:

$$(8)\qquad MAE = \frac{\sum_{i=1}^{n}|y_i - i|}{n}$$

- Mean Squared Error (MSE) is a measure of error by square the error value and bringing it to its mean. To measure the accuracy of this method, the smaller the value obtained, the more accurate the model will be. The equation is as follows:

$$(9)\qquad MSE = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y_i} - \boldsymbol{\hat{y}_i})^2$$

Where $\boldsymbol{y_i}$ = The actual value used in the prediction test.
$\boldsymbol{\hat{y}_i}$ = This is the value that the model predicted out of the test.
n = This was all used in the test.

- The equation uses to train the Artificial Neural Network to give the requirement to know if the change increases or decreases a given weight. Therefore, this research will use two equations, Stochastic Gradient Descent (SGD) and ADAM equations as follows:

$$(10)\qquad Gradient\ descent\ in\ W_{t+1} = W_t - \alpha\frac{\partial L}{\partial W_t}$$

Where W_t = It represents the value of the parameter at the current time.
$\frac{\partial L}{\partial W_t}$ = It represents the gradient or slope caused by the change in loss when changing W_t, and since gradient has the opposite direction to descending, the equation is updated with a negative value of gradient.
α = It instead of learning rate.

The equation for calculating momentum is

$$(11)\qquad m_{new} = \beta_1 * m_{old} - (1 - \beta_1 * \frac{\partial(Loss)}{\partial(W_{old})}$$

The equation for calculating the cache to adjust the learning rate is

$$(12)\ cache_{new} = \beta_2 * cache_{old} + (1 - \beta_2 * (\frac{\partial(Loss)}{\partial(W_{old})})^2$$

Where $\beta_1$and β_2 are usually defined as 0.9 and 0.99, respectively.

But from the analysis, it was found that both the momentum term and cache term calculated in this way have a small bias caused by m_0 and the default $cache_0$ is set to 0, causing the initial bad symptoms in both terms to be corrected before the bias according to the equation.

$$(13)\qquad \acute{m}_{new} = \frac{m_{new}}{1-\beta_2^t}$$

$$(14)\qquad ca\acute{c}he_{new} = \frac{cache_{new}}{1-\beta_2^t}$$

$$(15)\qquad ADAM\ in\ W_{new} = W_{old} - \frac{\alpha}{\sqrt{ca\acute{c}he_{new} + \epsilon}} * \acute{m}_{new}$$

$$(16)\qquad SGD\ in\ W = W - (lr \times W \bullet grad)$$

Where W = It is the Weight that is about to be updated; lr = It is Learning Rate; W • grad = It is the Gradient or Slope of that W per Loss.

From the equations used to measure the efficiency of all neural networks from the PM2.5 dust data set at 7 stations, it can be concluded to compare the predictive efficiency of 2 neural networks. The comparison results are shown in Table 1 in the next section of the experimental results.

The hyperparameters used in the ANN and LSTM neural networks for PM2.5 dust prediction for comparison of optimal results are described below:

To compare the experimental results of the two neural networks, it is possible to distinguish which neural network provides better prediction of PM2.5 dust values by measuring the accuracy, MAE and MSE values respectively. Therefore, the researcher has set the values of hyperparameters that are not different in both neural networks as detailed below. Starting from the hyperparameters of the LSTM contain input shape = 120 nodes, dense = 32 and in the hidden layer both layers contain dense = 32, the output layer has the value = 1. While the learning rate is used to control each step of the learning as lr = 0.00001, optimizer = adam, loss = MAE and MSE respectively. Therefore, the hyperparameters configuration of LSTM can be summarized as shown in Figure 7 as follows. At the same time, the ANN hyperparameter is assigned the same value as that of LSTM, as detailed below: the value of inputshape = 120 and the value of dense = 32 in both hiddenlayers are defined using activate function=relu, values of dropout rate = 0.2 and outputlayer = 1 without activate function, optimizer = adam, loss = mae and mse respectively. The hyperparameters configuration of LSTM can be summarized as shown in Figure 8 as follows.

```
#Make MultiLayer LSTM Model
model = keras.Sequential()
model.add(layers.LSTM(32, input_shape=(120, 1), return_sequences=True))
model.add(layers.LSTM(32, return_sequences=True))
model.add(layers.LSTM(32))
model.add(layers.Dense(1))
model.summary()

Model: "sequential_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm_1 (LSTM)               (None, 120, 32)           4352

 lstm_2 (LSTM)               (None, 120, 32)           8320

 lstm_3 (LSTM)               (None, 32)                8320

 dense_3 (Dense)             (None, 1)                 33

=================================================================
Total params: 21,025
Trainable params: 21,025
Non-trainable params: 0
```

Fig.7. The summarized of LSTM for prediction of PM2.5

```
model = keras.Sequential()
model.add(layers.Flatten(input_shape=(120, 1)))#120,11
model.add(layers.Dense(32, activation='relu'))
model.add(layers.Dense(32, activation='relu'))
model.add(layers.Dropout(0.2))
model.add(layers.Dense(1)) # Regression -> No Need for Activation
model.summary()

Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 flatten (Flatten)           (None, 120)               0

 dense (Dense)               (None, 32)                3872

 dense_1 (Dense)             (None, 32)                1056

 dropout (Dropout)           (None, 32)                0

 dense_2 (Dense)             (None, 1)                 33

=================================================================
Total params: 4,961
Trainable params: 4,961
Non-trainable params: 0
```

Fig.8. The Summarized of ANN for prediction of PM2.5

In all of the methodological design steps mentioned above, the results of the work resulting from the comparison of the two neural networks are presented in full detail in the experimental results section.

4. Experimental Result

Based on the design of both neural networks and the data using the data cleaning process mentioned in the methodology, the researchers divided the data obtained at each station to be used for learning and testing. Starting from random data, it is divided into 70% for learning and 30% for testing. The data characteristics can be graphed as shown in Figure 9. To process the work of both neural networks, the researchers used a Lenovo computer, intel core i7 gen 10th GPU, speed 2.60 GHz, 32 GB ram, and OS Microsoft Windows 11 Pro. The processing time of the neural network was found that the LSTM took about 15 minutes and the ANN took about 10 minutes to perform 150 times. While ANN does not have a reverse process, the time utilization of ANN is less than that of LSTM. To measure the efficiency of the two neural networks to compare the optimal values to predict the value of PM2.5 dust. To find the most suitable neural network to predict PM2.5 dust values, the researchers used MSE, MAE, and ACC measurements. The loss function was defined as MAE and MSE, respectively. Therefore, Figure 10 and Figure 11 show an example. The MAE curves of the LSTM and ANN, tend to decrease the error rate with the number of cycles of both operations. In determining the parameters of the Optimizers in the Artificial Neural Network training, the researcher wanted to know how a change in weight gain or decrease affects the increase or decrease of the loss function. Comparisons were made between ADAM and SGD to obtain optimal results of PM2.5 dust predictions. The equation for both SGD and ADAM the equations 15 and 16.

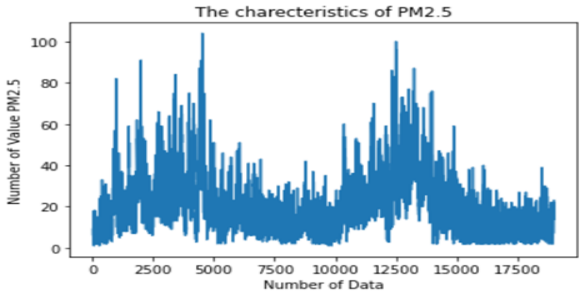

Fig.9. The data set of PM2.5 for prediction

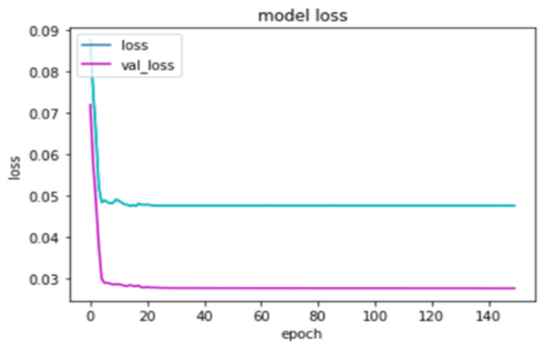

Fig.10. MAE graph of loss function in LSTM

To make the conclusions of the experiments easy to understand, the researchers showed the results of the processing experiments for the ANN using the ADAM optimizer. The results were both learning and testing for PM2.5. A total of 150 cycles consisted of the accuracy value, the MSE value and the MAE value respectively. While the LSTM showed the same experimental results as the ANN, an optimizer of the SGD type was added as a comparison between ADAM and SGD, which would lead to the conclusion of which optimizer yields the optimal value as shown in Table 1. From the performance values shown in Table 1, the comparative accuracy results between ANN and LSTM neural networks can be summarized in Figure 12 as follows.

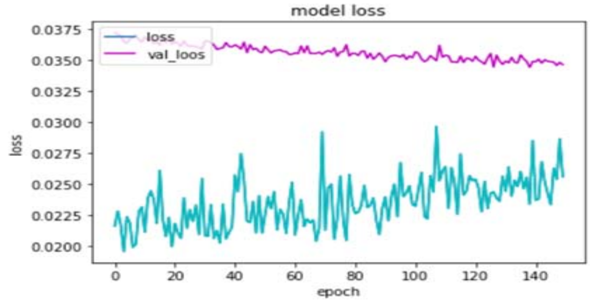

Fig. 11. MAE graph of loss function in ANN

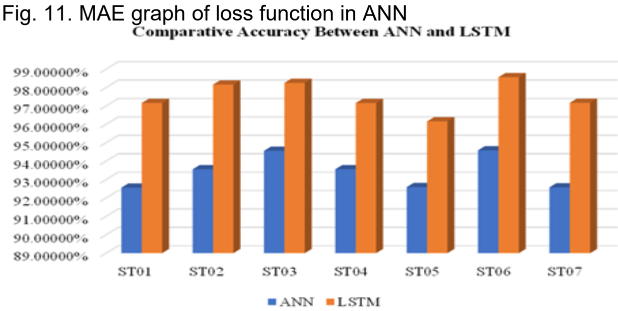

Fig.12. Accuracy comparison graph between ANN and LSTM neural networks

5. Conclusion

Based on this research, the researcher presents a comparison of PM2.5 dust prediction results between ANN neural networks and LSTM neural networks to determine which neural networks yield results. Optimal prediction whereby the researcher assigns similar hyperparameter values and performs learning and testing on the same number of epochs (150 epochs). It was found that Fig. 10 shows a graph that the trend of MAE values of LSTM neural networks has decreased steadily as well as being consistent. The ANN neural network type tends to decrease, but the value is not consistent. The LSTM average is 0.0720, the MSE is 0.0072 in the ADAM optimizer, and for the SGD optimizer the MAE is 0.0777, and the MSE is 0.00750. While the average ANN neural network MAE is 0.13380, MSE is 0.03370. The average accuracy of the LSTM neural network
for training was 97.7005%, the testing was 96.8496%, and the ANN neural network for training was 93.4860%, the testing was 91.1460%. From the values shown, the average value of LSTM accuracy was higher than that of the ANN for both learning and testing. The training accuracy values were 97.7005% for LSTM, 93.4860% for ANN, and testing values were 96.8496% for LSTM, 91.1460% for ANN. Moreover, the LSTM's MSE and MAE values show fewer errors than ANN. From the results of the experiments and all the conclusions obtained, it was found that the LSTM neural network was more predictive of PM2.5 dust than the ANN neural network. The pseudoscience takes longer to process than the ANN, which explains that the LSTM structure reverts to reprocessing, meaning it can remember what has been processed, which can take some time, but the results are accurate. However, the limitation of LSTM is that the data used must be sequential, and PM2.5 dust data is sequential. Moreover, in the LSTM neural network, the ADAM optimizer was found to provide better error and accuracy values than the SGD optimizer. Therefore, it was concluded that if the researcher wanted to predict the value of PM2.5 dust in which the data was sequenced, an LSTM-based neural network and ADAM-type optimizer should be used, as reported in all studies. It is well-proven.

Table 1. The table summarizers the results of the comparison of the accuracy, MSE and MAE performance between ANN and LSTM neural networks.

	Model Neural Network (Adam)				Model LSTM (Adam)				Model LSTM (SGD)			
	Train	Test	Error		Train	Test	Error		Train	Test	Error	
	Accuracy	Accuracy	MSE	MAE	Accuracy	Accuracy	MSE	MAE	Accuracy	Accuracy	MSE	MAE
ST01	92.5606%	91.5652%	0.0504	0.2057	97.1606%	96.8582%	0.0084	0.0857	97.1606%	97.2882%	0.0084	0.0856
ST02	93.5603%	90.5271%	0.0315	0.1989	98.1603%	97.2882%	0.0073	0.0976	98.1603%	97.2271%	0.0073	0.0776
ST03	94.5603%	91.5271%	0.0196	0.1250	98.2503%	96.2853%	0.0068	0.0601	98.1603%	96.2891%	0.0094	0.0969
ST04	93.5605%	90.5205%	0.0156	0.0980	97.1581%	97.2882%	0.0080	0.0620	97.1603%	96.2882%	0.0090	0.0719
ST05	92.5870%	90.8501%	0.0576	0.1221	96.1607%	96.2783%	0.0081	0.0784	97.1607%	96.2783%	0.0071	0.0780
ST06	94.5881%	92.0194%	0.0189	0.0987	98.5605%	97.2870%	0.0065	0.0661	98.2055%	97.2870%	0.0065	0.0661
ST07	92.5765%	90.5241%	0.0658	0.1267	97.1685%	96.8976%	0.0067	0.0771	97.1605%	96.2876%	0.0067	0.0857
Dindang	93.8942%	91.6342%	0.0099	0.0951	98.9845%	96.6130%	0.0055	0.0485	98.9941%	97.1131%	0.0054	0.0596

6. Recommendation

From the conclusions mentioned above, it was found that the LSTM type neural network used in this study, although giving a good output value, the accuracy was not very high. In addition, when the results were plotted Figure 10 overfitting graphs showed good learning but may not be able to predict results effectively. Because there is a possibility that the parameters are not optimal. Similarly, ANN neural network results when plotted graphs Figure 11 perform worse than LSTM neural network plots which are underfitting, indicating that the model cannot be learned from the training dataset. Specifically, the parameter configuration should be consistent with the input data and can be optimized accordingly with the input data. Therefore, the researcher will take this opinion for further research. It uses an algorithm that can optimize the parameters of the LSTM to obtain optimal results. It is possible to use a heuristic algorithm which is swarm intelligence to optimize the hyperparameters of an LSTM neural network to achieve the best output, which used the PM2.5 dataset for prediction in the next times.

Acknowledgement

This research was supported by the Faculty of Engineering, Rajamangala University of Technology, Isan, Khon Kaen Campus, Khon Kaen, Thailand.

Authors: *Thanpitcha Atiwanwong, D.Eng. candidate., Rajamangala University of Technology Isan, Khon Kaen Campus, Thailand, E-mail: thanpitcha.at@rmuti.ac.th; Asst. Prof. Dr. Adirek Jantakun, Ph.D., Rajamangala University of Technology Isan, Khon Kaen Campus,Thailand, E-mail: adirek.ja@rmuti.ac.th; Dr. Adisak Sangsongfa, Ph.D.,Rajabhat Mu-Ban Chombueng University, Thailand, E-mail: adisaksan@mcru.ac.th*

REFERENCES

[1] T. Airveda, What is pm2.5 and why is it important?, *Airveda,* (2017), [Online]. Available: https://www.airveda.com/ |blog/what-is-pm2-5-and-why-is-it-important

[2] *Ambient air pollution,* World Health Organization: Health Impact, Institute for Health and Evaluation, University of Washington, (2020).

[3] Wang, Xianghong, and Baozhen Wang. Research on prediction of environmental aerosol and PM2. 5 based on artificial neural network. *Neural Computing and Applications* 31.12 (2019): 8217-8227.

[4] Mahajan, Sachit, et al. Improving the accuracy and efficiency of PM2. 5 forecast service using cluster-based hybrid neural network model. *IEEE Access 6* (2018): 19193-19204.

[5] I. G. McKendry, Evaluation of artificial neural network for fine particulate (pm10 and pm2.5) forecasting, *Journal of the Air and Waste Management Association*, (2002) vol. 52, 1096–1101.

[6] Y. Qi, Q. Li, H. Karimian, and D. Liu, A hybrid model for spatiotemporal forecasting of pm2.5 based on graph convolutional neural network and long short-term memory, *Science of The Total Environment*, (2019), vol. 664, 02.

[7] D. Seng, Q. Zhang, X. Zhang, G. Chen, and X. Chen, Spatiotemporal prediction of air quality based on lstm neural network, *Alexandria Engineering Journal*, (2020), vol. 60, 12.

[8] T. Atiwanwong and S. Hongprasit, A low-power real-time pollution monitoring system using esp lora, *Mahasarakham International Journal of Engineering Technology*, (2020), vol. 6, no. 1, 36–40.

[9] Shah, Syed Ahsin Ali, et al. A hybrid model for forecasting of particulate matter concentrations based on multiscale characterization and machine learning techniques. *Mathematical Biosciences and Engineering 18.3* (2021): 1992-2009.

[10] IBM, Neural networks, *IBM Cloud Education*, (2020), August 17, accessed on 2022-5-30. [Online]. Available: https://www.ibm.com/cloud/learn/neuralnetworks

[11] Feng, Xiao, et al. Artificial neural networks forecasting of PM2. 5 pollution using air mass trajectory based geographic model and wavelet transformation. *Atmospheric Environment 107* (2015): 118-128.

[12] Asquero, Advantages and disadvantages of artificial neural networks, *Deep Learning*, (2020), [Online]. Available: https://www.asquero.com/article/advantages-and-disadvantages-of-artificial-neural-networks/

[13] S. Saxena., Introduction to long short-term memory (lstm). *Analytics Vidhya*, (2021), [Online]. Available: https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/

[14] GeeksforGeeks., Understanding of lstm networks. *why Learn Data Science*, (2021), [Online]. Available: https://www.geeksforgeeks.org/understanding-of-lstm-networks/

[15] Y. Jiao, Z. Wang, and Y. Zhang., Prediction of air quality index based on LSTM. *in Proceedings of 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference*, (2019), ITAIC, 17–20. http://dx.doi.org/10.1109/ITAIC.2019.8785602.

[16] H. Sepp and S. Jurgen., Long short term memory. *Neural Computation*, (1997), vol. 9, no. 8, 1735-1780. http://dx.doi.org/10.1162/neco.1997.9.8.1735

[17] M. Chovatiya, A. Dhameliya, J. Deokar, J. Gonsalves, and A. Mathur., Prediction of dengue using recurrent neural network. *in Proceedings of the International Conference on Trends in Electronics and Informatics*, (2019), vol. 2019–April, 926–929. http://dx.doi.org/10.1109/ICOEI.2019.8862581

[18] Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J., LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, (2017), 28 (10), 2222-2232. http://dx.doi.org/10.1109/TNNLS.2016.2582924

[19] C. Olah., Understanding lstm networks Neural Networks, (2022), [Online]. Available: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

[20] G. Singh, Introduction to Artificial Neural Networks. *Analytics Vidhya*, (2021), [Online]. Available: https://www.analyticsvidhya.com/blog/2021/09/introduction-to-artificial-neural-networks/