Analysis and Prediction of Rainfall with Oceanic Nino Index and Climate Variables Using Correlation Coefficient and Deep Learning

Chayanat Buathongkhue [1] , Kritsana Sureeya [2], Natapon Kaewthong [2]*

[1] *College of Industrial Technology and Management, Rajamangala University of Technology Srivijaya, Nakhon Si Thammarat 80210, Thailand.*

[2] *Faculty of Engineering, Rajamangala University of Technology Srivijaya, Songkhla 90000, Thailand.*

Abstract

This article presents the relationship between the Oceanic Nino Index (ONI) and monthly rainfall on the southern and eastern coast of Thailand, specifically in Narathiwat, Pattani, and Yala provinces, where influences have been commonly observed. This research aims to study the relationship between the Oceanic Nino Index (ONI) and monthly rainfall to develop a model for predicting monthly rainfall. Despite previous related research, there has been no in-depth study on the relationship between the Oceanic Nino Index (ONI) and monthly rainfall in areas adjacent to the sea. The correlation coefficient was used to determine the relationship, revealing that the ONI value is significantly correlated with the amount of rainfall in the current month and the following month. This correlation paved the way for developing a model to predict monthly rainfall. Multiple linear regression, recurrent neural networks, and long short-term memory models were employed for this purpose. The study found that utilizing a recurrent neural network yielded the best prediction efficiency, with Mean Absolute Error (MAE) values of 112.76 mm for Narathiwat province, 81.06 mm for Pattani province, and 97.67 mm for Yala province.

Keywords: Rainfall Prediction; Oceanic Nino Index; Eastern Sea Coast; Deep Learning.

1. Introduction

Global climate change is a problem affecting everyone, with severe floods and droughts resulting from El Niño and La Niña events. It has become more frequent since the 1960s [1] and tends to cause higher salinity levels [2]. Therefore, rainfall forecasts are essential, and highly accurate forecasts can reduce the risk of damage caused by disasters, including floods and droughts. An essential index for measuring climate variation is the Oceanic Nino Index (ONI), which NOAA uses to track the El Niño-Southern Oscillation (ENSO) climate. The index value An ONI index of +0.5 or higher means El Niño is occurring, and an ONI index of -0.5 or lower means La Niña [3-5] is occurring. This variation will affect the climate and become the leading cause of significant impacts on the world's ecosystems and society [6]. This index value can be predicted in advance, both by relevant agencies and from currently available research results [7, 8].

In the past, research and evaluation of impacts caused by ONI included analysis and prediction of rainfall by the ONI index using correlation coefficients and deep learning [9]. In Thailand, more research still needs to be done. There is supporting data that the relative Niño3.4 index is more consistent with impacts on rainfall and will be more helpful in preparing for El Niño and La Niña events in a changing climate [10]. However, in-depth knowledge must be used through a multi-layer artificial neural network tool to study more comprehensively and fully understand the effects of ENSO through deep learning models and complex patterns in the context of climate prediction. Deep learning can model

---

complex climate systems and predict future weather conditions. From the compilation of past statistics, the most common meteorological fields in numerical weather prediction are wind predictions, and the second is rainfall [11]. Many other studies use deep learning to create flood and drought disaster models and provide high accuracy [12–14].

This research studied the Analysis and Prediction of Rainfall with Oceanic Nino Index and Climate Variables Using Correlation Coefficient and Deep Learning in a case study of the southern region of the eastern sea, in the areas of Pattani province, Yala province, and Narathiwat province in Thailand, where it is necessary to know advance rain forecast information for people to use in planting planning because this area is an agricultural area where many types of crops are grown and each of which has different water needs [15]. In some growing seasons, there is not enough water for agriculture. In the areas of Narathiwat, Pattani, and Yala provinces, there are 3 important rivers: the Pattani River [16], the Sai Buri River, and the Kolok River [17]. The Kolok River is the river that separates the border between Thailand and Malaysia, which is the end of the Kolok River in Narathiwat Province. This makes it difficult to predict the amount of water and manage it if the El Niño and La Niña phenomena occur. In some years, there will be a La Nina phenomenon, and there will be much rain in the area. On the other hand, if there is an El Niño phenomenon that year, it causes a low amount of rain. This is why farmers in the area cannot cope with such problems, resulting in damage to agricultural products.

The authors aimed to study the relationship between the Oceanic Niño Index (ONI) and monthly rainfall in Narathiwat, Pattani, and Yala provinces. Using the obtained relationship, we developed a model to predict monthly rainfall using multiple linear regression, recurrent neural networks, and long-term short-term memory, comparing the predictive performance of each model.

2. Related Work

In literature related to the impact of the Oceanic Nino Index on rainfall, we found that Hidayat et al. [18] found that ENSO and IOP significantly impacted rainfall and developed a model for predicting rainfall in Northwestern Java and Makassar in Indonesia from September to November. It was found that one-month forecasts had correlations of approximately 0.72 and 0.80 for the Northwestern Java and Makassar areas, respectively. Similarly, Irwandi et al. [19] studied the effects of El Niño-Southern Oscillation (ENSO) on rainfall in North Sumatra, Indonesia, from 1981 to 2016. They found that if the El Niño phenomenon occurs, the summer will be longer and the rainy season will be shorter, causing a decreasing amount of rainfall during December–February (DJF), June–August (JJA), and August–November (SON) and a decreasing amount of rain by approximately 7% per year. On the other hand, if the La Nina phenomenon occurs, the rainy season will be longer and the summer will be shorter, causing an increase in the amount of rain during December–February (DJF) and June–August (JJA), and the increasing amount of rain will be approximately 6% per year. Ueangsawat et al. [20] presented the impact of ENSO on temperature and rainfall variations in Chiang Mai, the upper northern province of Thailand, from 1988 to 2011. From the study, they found that ENSO significantly affects rainfall, with the highest El Niño effects typically occurring between March and September. In contrast, La Niña usually has its highest impact from January to December.

From the previous research on rainfall prediction, we found that various traditional mathematical models were used, such as multiple linear regression [21, 22]; and fuzzy logic [23, 24] with data imports such as temperature, humidity, pressure, and wind speed. However, in the past few years, machine learning (ML) and artificial neural networks have become increasingly popular in hydrology prediction, water quality prediction, business, and others; it is a self-learning model. It is a branch of artificial intelligence that can find non-linear relationships between inputs and outputs without using relationships between physical variables. Machine learning can be used for rainfall prediction [25, 26], runoff prediction [27, 28], salinity or water quality [29–31], and flood prediction [32, 33]. ML has many advantages, such as low complexity, low cost, and many models to choose from according to the suitability of the research. For this reason, ML is used to predict rainfall. For example, Htike & Khalifa [34] used annual and monthly rainfall data for rainfall forecasting. They trained Focused Time Delay Neural Networks (FTDNN). When trained on annual rainfall data, the FTDNN model gave the highest prediction accuracy. Hong [35] used a support vector machine to predict hourly rainfall using rainfall data obtained from August 1985 to August 1997, divided into three phases: training (300 hours), verification (206 hours), and testing (167 hours). The study found high prediction accuracy. Sivapragasam et al. [36] developed a rainfall and runoff prediction model in Singapore using singular spectrum analysis (SSA) with a support vector machine (SVM). The study found that the correlation coefficient is 0.70. Li et al. [37] developed a rainfall prediction model for 1, 3, and 6 months using wavelet analysis and support vector machines together in Qilian, Yeniugou, and Tuole stations of the Qilian Mountains, China. It was found that Qilian Station had correlation coefficient values of 0.915, 0.917, and 0.916, respectively.

However, deep learning is currently being used to predict rainfall. There are many types of deep learning, such as convolutional neural networks (CNN) [38], recurrent neural networks [39], and long-short-term memory [40]. Deep learning has the advantage of handling extensive, complex data. This is difficult for traditional machine learning algorithms to process, which is why there is now more and more research on deep learning. For example, Van et al. [41] applied a 1D convolutional neural network to predict runoff from rainfall using 1996–2011 data in Chau Doc and Can

Tho, Vietnam. The ReLU activation function was chosen from the study. The study found that the correlation coefficient value was more significant than 0.9. Haidar and Verma [42] also applied CNN to predict rainfall in Innisfail, Australia, using data from 2001–2012, which included imported weather variables such as the Southern Oscillation Index (SOI), Nino 1.2, Nino 3.0, Nino 3.4, Nino 4.0, Dipole Mode Index (DMI), and interdecadal.

3. Study Area and Data Set

This study focuses on the eastern coast of Thailand. It consists of Narathiwat province, which has an area of approximately 447,500 ha; Pattani province, which has an area of approximately 194,000 ha; and Yala province, which has an area of approximately 452,100 ha, as shown in Figure 1.
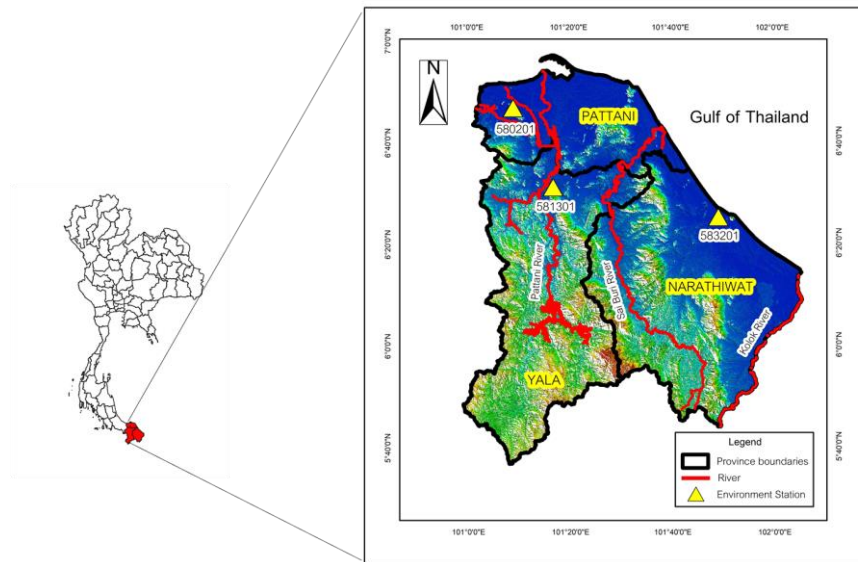

Figure 1. Location of this study

The climate data used, such as rainfall, temperature, relative humidity, amount of water evaporation, wind direction, and wind speed, came from measurement stations of the Southern-East Coast Meteorological Centre Thailand, consisting of 3 stations: Narathiwat province, Pattani province, and Yala province, with monthly data from 1994 to 2023. Table 1 shows this study's basic statistical information and data units. This study combines the ONI with climate data to analyze and create a model for monthly rainfall prediction. Table 1 shows the identifying climate variables for the prediction of monthly rainfall.

Table 1. Variable and statistical basis of environmental data

Location	Variable Parameter	Unit	Maximum	Average	Minimum	Standard Deviation
-	x_1: Oceanic Nino Index					
Narathiwat	x_2: Humidity	%	89.70	80.93	73.87	3.07
	x_3: Rain	mm	1324.60	235.61	0.40	248.25
	x_4: Temperature	°C	30.25	27.76	25.28	0.99
	x_5: Evaporated water	mm	177.00	120.77	56.10	23.71
	x_6: Wind direction	degree	207.67	106.18	34.19	33.72
	x_7: Wind speed	m/s	14.63	5.49	3.98	0.73
Pattani	x_8: Humidity	%	91.42	81.00	72.20	3.41
	x_9: Rain	mm	1036.00	166.60	0.00	165.60
	x_{10}: Temperature	°C	30.54	28.40	26.21	0.89
	x_{11}: Evaporated water	mm	181.60	125.32	57.90	22.90
	x_{12}: Wind direction	degree	280.33	161.51	55.81	68.05
	x_{13}: Wind speed	m/s	12.86	6.48	3.16	1.69
Yala	x_{14}: Humidity	%	89.85	81.22	70.80	3.48
	x_{15}: Rain	mm	923.05	201.66	0.00	163.57
	x_{16}: Temperature	°C	30.37	28.27	25.77	0.99
	x_{17}: Evaporated water	mm	228.60	116.66	49.80	26.18
	x_{18}: Wind direction	degree	250.67	149.65	64.00	54.78
	x_{19}: Wind speed	m/s	12.01	6.48	2.28	1.73

Figure 2 shows monthly ONI in the Pacific Ocean from 1994 to 2023. ONI values generally range from -2.5, indicating a strong La Nina, to 2.5, indicating a strong El Nino.
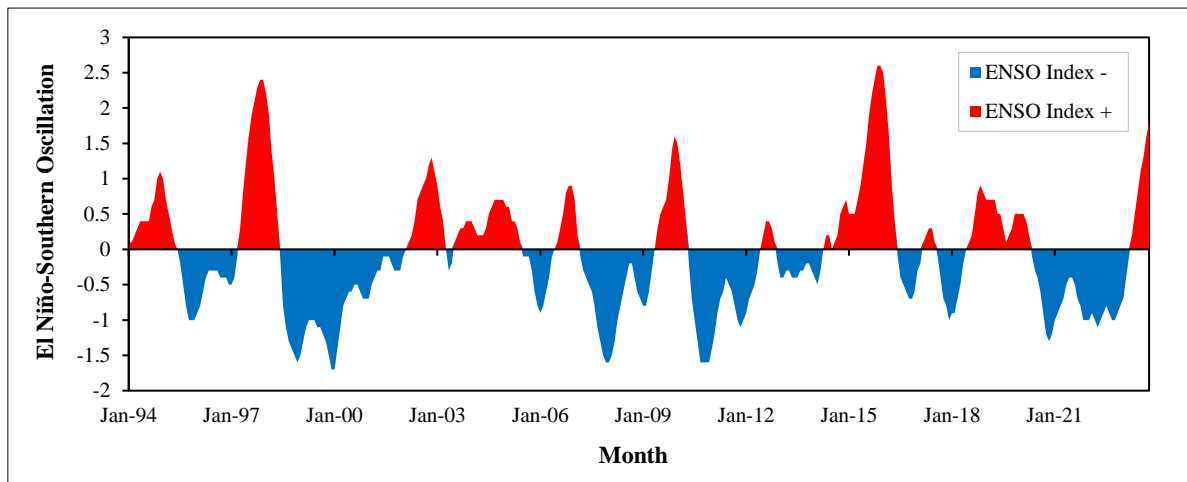

Figure 2. ONI in the Pacific Ocean

4. Methodology and Prediction Model

This section explains the methodology, which is divided into 3 experiments. It also describes the selection of models to develop a monthly rainfall prediction system and evaluate the model's performance.

4.1. Methodology

This section is divided into three sub-studies. In the first study, monthly average rainfall from 1994 to 2023 is analyzed and compared to monthly average rainfall during El Niño and La Niña events. This study aims to determine how the average monthly rainfall from 1994 to 2023 compares to the monthly average rainfall if El Niño or La Niña phenomena occur. The second study examined the influence of the ONI on monthly rainfall. The study used the correlation coefficient equation [43] as in Equation 1. The study's objective is to determine how the ONI in each month affects the amount of rain in the future. For example, the ONI in January, equal to +1.5, affects the rainfall in January less than in February. The results of such studies lead to the selection of input variables for the prediction model. (feature selection), as shown in Figure 3.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{1}$$

where r is Correlation Coefficient, x_i is value of the x-variable, $\bar{x}$ is mean of the values of the x-variable, y_i is value of the x-variable, and $\bar{y}$ is mean of the values of the x-variable.
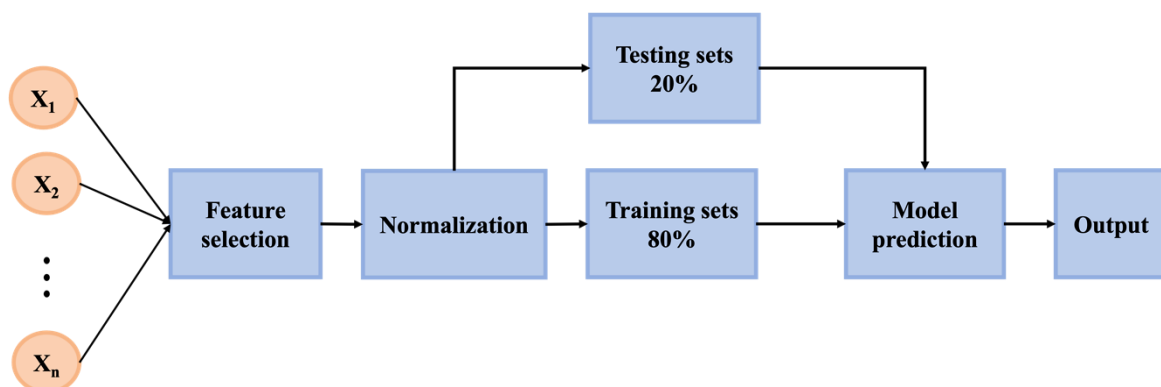

Figure 3. Block diagram of the modelling processes

The correlation coefficient result has a value from -1 to 1 [44], with a negative value indicating a negative relationship between the variables, meaning that if variable 1 increases in value, variable 2 decreases. On the other hand, if the calculation result is positive, it indicates a positive relationship among the variables, meaning that an increase in variable 1 gives a likewise results in an increase in variable 2. However, the relationship between an angular variable and a variable with values is linear. For example, the relationship between the Oceanic Nino Index and the wind direction requires using Equation 2 Circular-Linear Correlation [45].

$$r^2 = \frac{r_{xc}^2 + r_{xs}^2 - 2r_{xc}r_{xs}r_{cs}}{1 - r_{cs}^2} \tag{2}$$

where r^2 is circular – linear correlation Coefficient, r_{xc} is correlation coefficient of x and $\cos \propto$, r_{xs} is correlation coefficient of x and $\sin \propto$, and r_{cs} is correlation coefficient of $\cos \propto$ and $\sin \propto$.

The second study is part of the final study, which involves the development of a model to predict monthly rainfall. The results from the second study are used to describe how, in the future, the ONI value will affect monthly rainfall. Once known, the data will be rearranged again before normalizing all data such as ONI, rainfall, temperature, relative humidity, amount of water evaporation, wind direction, and wind speed using Equation 3 [46].

$$x_{normalized} = \frac{x - x_{minimum}}{x_{maximum} - x_{minimum}} \tag{3}$$

where, $x_{i,normalized}$ is the different climate variables obtained. Normalization, which has a value between 0 and 1. x_i is climate variables. $x_{i,maximum}$ and $x_{i,minimum}$ are the maximum and minimum values of the climate variables, respectively.

After normalization, all data from 1994 to 2023 were split into 2 parts: data for the training model and data for the test model, which were divided into 80/20 percent, respectively, as shown in Figure 3. Then, this information was entered into a model for predicting monthly rainfall. In this study, 3 models have been used, and the first is a traditional model with multiple linear regression. The second and third models are neural networks for prediction values in time series, namely recurrent neural networks and long-short-term memory. Recurrent neural networks and long-short-term memory models are popular because they are suitable for developing systems to predict data in the time domain.

4.2. Multiple Linear Regression

Multiple linear regression is a method of analyzing data to find relationships between dependent variables and independent variables. This method relies on a linear relationship between variables for use in prediction. The multiple linear regression equation shows the relationship between the dependent and independent variables. It can be written as an equation below.

$$y_i = \beta_0 = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_n x_n + \epsilon \tag{4}$$

where y_i is dependent variable, β_0 and β_i are intercept and coefficient of regression respectively, x_i represent a independent variable, and ϵ is the error term [47].

4.3. Recurrent Neural Network

Recurrent neural networks (RNNs) are artificial neural networks that use past and present data to build models [48]. RNNs can find results that will occur in the future. Information from the previous hidden state or sequence *t-1* and information in sequence *t* will be used to predict the output at *t*. The architecture of a recurrent neural network can be seen in Figure 4 [49]. This is the reason RNNs are suitable for time series prediction. Equations 5 and 6 show equations for calculating the output of RNNs.

$$h_t = ReLU(W_h h_{t-1} + W_x x_t + b_i) \tag{5}$$

$$y_t = tanh(W_0 h_t + b_0) \tag{6}$$

Equations 5 and 6 are Equations for calculating output of RNNs [49], h_t is the hidden state of the RNNs cell. W_h and W_x are the weight matrices. b_i is the bias vector for the hidden state. x_t represents current input. In this paper, ReLU function was used as activation function for hidden state because it had a simpler function than other activation functions. W_0 and b_0 are the weight matrices and bias vector for the cell output, and hyperbolic tangent function is used for the activation function for the cell output.
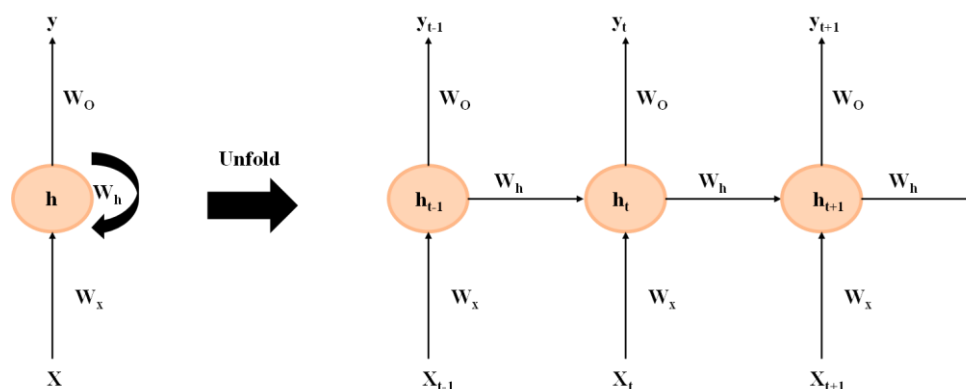
Figure 4. Architecture of recurrent neural network

4.4. Long Short-Term Memory

Long Short Term Memory (LSTM) is an artificial neural network with memory capabilities. LSTM is suitable for processing and predicting events with relatively long sequences [50]. LSTM was developed from the Recurrent Neural Network (RNN) model. LSTM's working principle is that it can store each node's "state" or information. The LSTM technique states that a particular function (gate) controls the data that comes into each node, consisting of the forget gate, input gate, and output gate [51]. It can be seen in Figure 5.
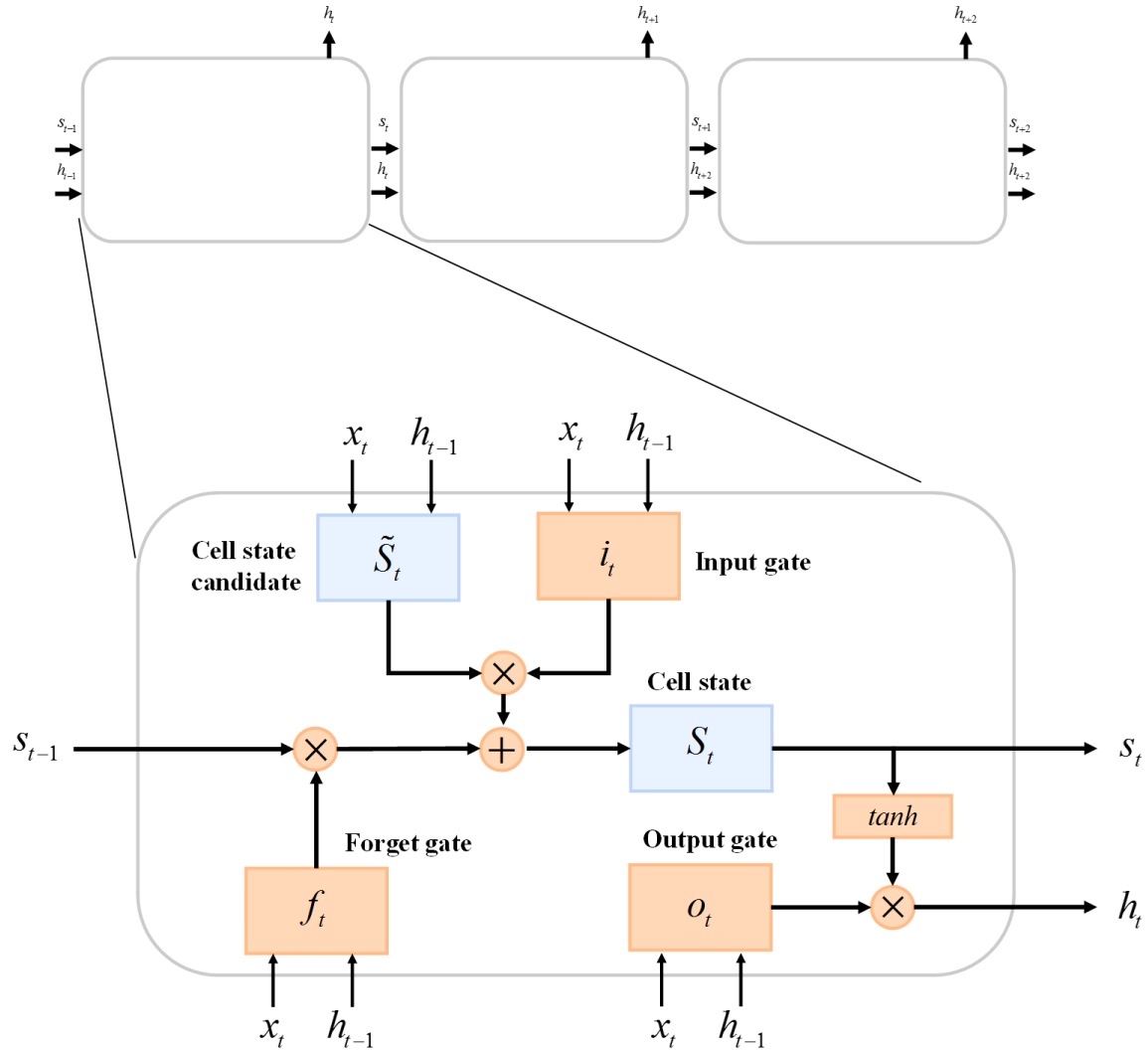

Figure 5. Architecture of long short term memory

The forget gate is a gate that determines whether data entering the cell state should be kept or discarded. The data that is kept is evaluated from the input to that node, combined with the calculated results of the previous node via the sigmoid function. It can be described as Equation 7.

$$f_t = sigmoid(w_{f,x}x_t + w_{f,h}h_{t-1} + b_f) \tag{7}$$

where, f_t is forget gate (a value between 0 and 1), $W_{f,x}$ and $W_{f,h}$ are weight matrices , h_{t-1} is the output value of the previous cell state , x_t is the input value , and b_f is the bias vector.

The input gate is a gate that is responsible for receiving new input data and then recording it at each node. The procedure is divided into two parts. The first part is an inspection of the cell state update. When receiving input data with the sigmoid function, the controller calls the input gate to choose whether to update the cell state or not. The second part is that if the input gate chooses to update the cell state, a candidate value ($\widetilde{s_t}$) is created with a hyperbolic tangent function. It can be described as Equations 8 and 9.

$$i_i = sigmoid(w_{i,x}x_t + w_{i,h}h_{t-1} + b_i) \tag{8}$$

$$\widetilde{s_i} = \tanh(w_{\tilde{s},x}x_t + w_{i,h}h_{t-1} + b_i) \tag{9}$$

where , i_t and $\tilde{s}_t$ are input gate and candidate value respectively, $W_{i,x}$, $W_{i,h}$, $W_{\tilde{s},x}$, and $W_{\tilde{s},h}$ are weight matrices, h_{t-1} is the output value of the previous cell state , x_t is the input value , and b_i is the bias vector.

The output gate is the gate that is responsible for prepare to export data. It works with a previously calculated cell state process, in which the sigmoid function will determine data from the cell state. Then bring the cell state value into the hyperbolic tangent function (the resulting value is 1 or -1). After that, the value obtained from the hyperbolic tangent function will be calculated with the output value obtained from the Sigmoid function. It can be described as Equations 10–12.

$$o_i = sigmoid(w_{o,x}x_t + w_{o,h}h_{t-1} + b_o) \tag{10}$$

$$s_t = f_t \times s_{t-1} + i_t \times \tilde{s}_t \tag{11}$$

$$h_t = o_t \times tanh\,(s_t) \tag{12}$$

where , i_t, $\tilde{s}_t$, and h_t are output gate, cell states, and output values respectively, $W_{o,x}$ and $W_{o,h}$ are weight matrices, h_{t-1} is the output value of the previous cell state , x_t is the input value , and s_{t-1} is previous cell state value.

For all 3 models mentioned above. It was tested in six different scenarios. It is the import of ONI, rainfall, temperature, relative humidity, amount of water evaporation, wind direction, and wind speed data from the past 1-6 months to predict the amount of rainfall in the next month. For example, import data from January - April to predict monthly rainfall in May. This can be explained, as shown in Figure 6.
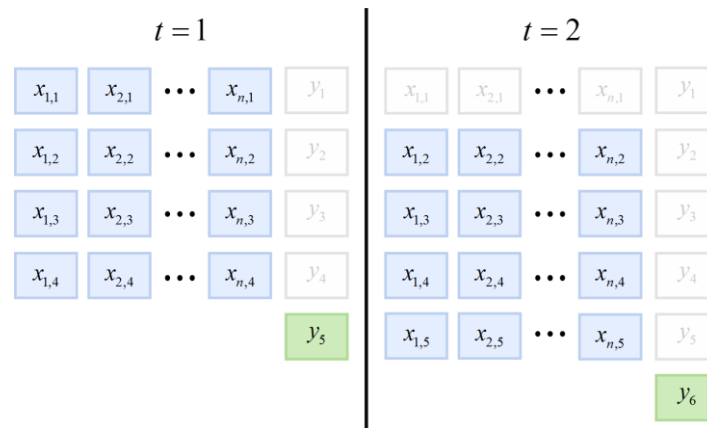

Figure 6. Input data to predict monthly rainfall

In this study, statistical indicators were used as criteria to evaluate model performance: mean absolute error (MAE) [31], correlation coefficient (r) [52]. According as in Equations 13 and 14.

$$MAE = \frac{\sum_{i=1}^{n}|R_{obs}-R_{Pred}|}{n} \tag{13}$$

$$r = \frac{\sum_{i=1}^{n}(R_{Obs} - \bar{R}_{Obs})(R_{Pred} - \bar{R}_{Pred})}{\sqrt{\sum_{i=1}^{n}(R_{Obs} - \bar{R}_{Obs})^2}\sqrt{\sum_{i=1}^{n}(R_{Pred} - \bar{R}_{Pred})^2}} \tag{14}$$

where MAE and r are mean absolute error and correlation coefficient, respectively. R_{Obs} and R_{Pred} are monthly rainfall obtained from observed and monthly rainfall obtained from predicted.

MAE is a measurement of the average magnitude of the error. It shows an error between the simulated values calculates by the model and the observed rainfall values. MAE ranges from 0 to ∞. Lower values are better. The r measures the linear relationship between two variables. Its values range between -1 and 1. If r is close to 1, both variables are highly correlated and have the same direction.

5. Result and Discussion

5.1. Preliminary Analysis Between Rainfall and El Niño-Southern Oscillation

The amount of rain in the area was analyzed by comparing it with weather oscillations in the southern hemisphere (El-Nino Southern Oscillation: ENSO). In the analysis, the measuring stations of the Department of Meteorology, which consist of 583201 stations in Narathiwat province, 580201 stations in Pattani province, and 581301 stations in Yala province, were chosen to be used.

In Figure 7, the accumulated monthly rainfall at all 3 stations was consistent. When the studied month had a La Nina phenomenon, the monthly accumulated rainfall was higher than the monthly average; when the studied month had an El Nino phenomenon, the monthly accumulated rainfall was lower than the monthly average. However, in some months, from May to October, the monthly rainfall did not correspond to La Nina or El Nino phenomena.
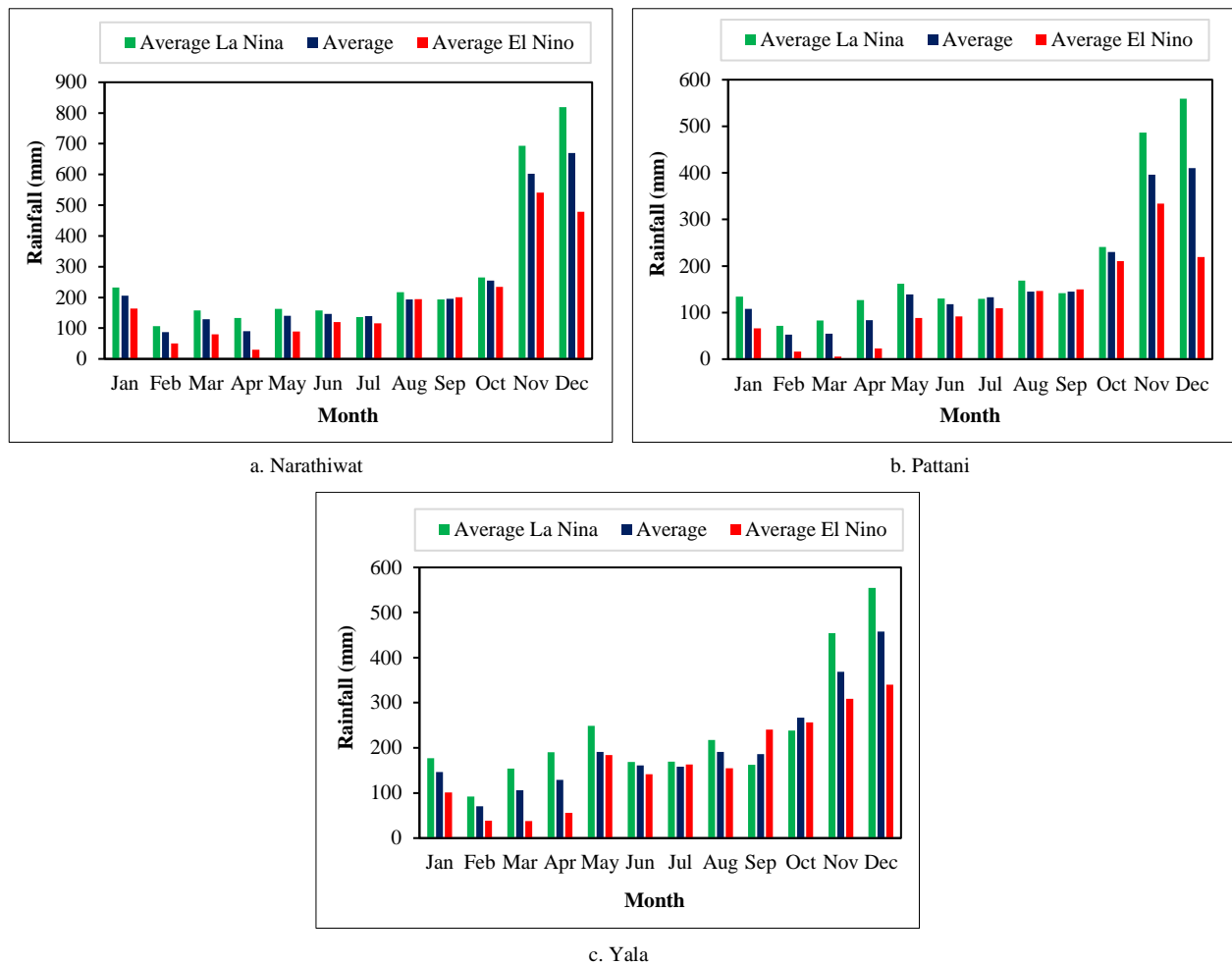

a. Narathiwat

b. Pattani

c. Yala

Figure 7. Average rainfall vs Average rainfall in the El Nino/ La Nina

Table 2 shows the average monthly rainfall in the event of an El Nino or La Nina phenomenon compared to the average monthly rainfall from 1994 to 2023. In Narathiwat province, when El Nino occurred during the summer, the average monthly rainfall was approximately 67.3% less than the average monthly rainfall of June in 1994–2023. On the other hand, when the La Nina phenomenon occurred, the average monthly rainfall was approximately 47.4% greater than the average monthly rainfall in June. When the El Nino phenomenon occurred in the rainy season, the average monthly rainfall in February was less than that rainfall from 1994–2023 by approximately 28.4%. If the La Nina phenomenon occurs during the rainy season, the average monthly rainfall in February was more than 22.3%. Likewise, in Pattani province, when El Nino occurred during the summer, the average monthly rainfall was approximately 90.1% less than the average monthly rainfall from 1994 to 2023 in March.

On the other hand, when the La Nina phenomenon occurred, the average monthly rainfall was approximately 51.2% greater than the average monthly rainfall in April. When the El Nino phenomenon occurred in the rainy season, the average monthly rainfall in February was less than that from 1994–2023, which is approximately 69.4%. If the La Nina phenomenon occurs during the rainy season, the average monthly rainfall in February was more than that of approximately 37.0%. Similarly, in Yala province, when El Nino occurs during the summer, the average monthly rainfall was approximately 64.4% less than the average monthly rainfall from 1994 to 2023 in March. On the other hand, if the La Nina phenomenon occurred, the average monthly rainfall would be approximately 4.7% greater than the average monthly rainfall in April. When the El Nino phenomenon occurred in the rainy season, the average monthly rainfall in February was less than that from 1994 to 2023, which is approximately 45.5%. When the La Nina phenomenon occurred in the rainy season, the average monthly rainfall in February was more than average, at approximately 30.2%.

Table 2. Percentage of rainfall when La Nina or El Nino occurs compared to average monthly rainfall

Season	Month	Narathiwat		Pattani		Yala	
		La Nina	El Nino	La Nina	El Nino	La Nina	El Nino
Summer	Mar	12.7	-20.3	51.0	-90.1	44.7	-64.4
	Apr	22.3	-41.8	51.2	-72.7	47.2	-56.8
	May	21.6	-38.2	16.5	-36.2	30.1	-3.7
	Jun	47.4	-67.3	10.2	-21.9	4.9	-12.1
	Jul	16.4	-36.0	-2.2	-17.8	7.1	3.5
	Aug	8.2	-17.7	16.1	0.7	13.7	-19.2
Rainy	Sep	-2.1	-16.9	-2.1	3.3	-12.8	28.9
	Oct	12.1	0.5	4.5	-8.5	-10.7	-4.1
	Nov	-1.5	2.5	22.9	-15.5	23.2	-16.2
	Dec	4.1	-7.7	36.5	-46.5	21.2	-25.8
	Jan	15.0	-10.2	24.1	-38.6	21.1	-30.9
	Feb	22.3	-28.4	37.0	-69.4	30.2	-45.5

5.2. Enso Lag Time Selection

Table 3 shows the correlation coefficient between the ONI and the climate variables. If the ONI in the current month is calculated to find its relationship with the climate variables in the current month, the calculation results have the strongest relationship. If the ONI values from the past month are calculated to find the relationship, it can be noted that the correlation coefficient values are not significantly different from those calculated using the ONI for the current month. However, if the ONI values from more than 2 months are used, the correlation coefficient calculation results decrease accordingly.

Table 3. Correlation coefficient between Oceanic Nino Index and climate variable at lag times

Parameter	Lead Time (Month)						
	0	1	2	3	4	5	6
x_2	-0.101	-0.108	-0.087	-0.102	-0.067	-0.054	-0.044
x_3	-0.143	-0.146	-0.118	-0.136	-0.096	-0.085	-0.080
x_4	0.147	0.150	0.119	0.139	0.097	0.073	0.051
x_5	0.152	0.156	0.138	0.148	0.129	0.121	0.109
x_6	0.086	0.088	0.107	0.093	0.116	0.105	0.078
x_7	-0.146	-0.127	-0.097	-0.110	-0.079	-0.059	-0.046
x_8	-0.166	-0.176	-0.164	-0.174	-0.146	-0.127	-0.106
x_9	-0.193	-0.191	-0.151	-0.175	-0.126	-0.108	-0.087
x_{10}	0.208	0.206	0.163	0.190	0.137	0.112	0.089
x_{11}	0.157	0.161	0.150	0.159	0.140	0.127	0.110
x_{12}	0.017	0.027	0.076	0.049	0.101	0.110	0.107
x_{13}	0.078	0.087	0.102	0.098	0.101	0.096	0.096
x_{14}	-0.170	-0.183	-0.179	-0.187	-0.161	-0.140	-0.118
x_{15}	-0.187	-0.191	-0.164	-0.183	-0.137	-0.113	-0.090
x_{16}	0.186	0.183	0.144	0.168	0.117	0.091	0.069
x_{17}	0.158	0.155	0.145	0.155	0.133	0.120	0.112
x_{18}	0.046	0.057	0.103	0.076	0.126	0.138	0.132
x_{19}	0.125	0.115	0.102	0.110	0.095	0.085	0.075

However, the relationship between ONI values and monthly rainfall in Narathiwat, Pattani, and Yala provinces (x3, x9, and x15) had a correlation coefficient between -0.143 and -0.193. The results differed from Ueangsawat et al. [20], who also studied the relationship between ONI values and monthly rainfall in Chiang Mai province, northern Thailand, far from the studied area of the current study. The correlation coefficients were between -0.093 and 0.766, demonstrating the strong relationship between ONI values and monthly rainfall in the area. The correlation coefficient values were higher in their study because they combined the ONI value with the average monthly rainfall for 3 months to calculate the correlation coefficient. On the contrary, the current study calculated the correlation coefficient directly without considering the 3-month average rainfall. The correlation coefficient calculation results in Table 3 were used to develop a model for predicting monthly rainfall values. From Table 3, it can be concluded that the month-to-month ONI values will be used to import the model to predict next month's monthly rainfall.

5.3. Rainfall Prediction Model

To develop a model for monthly rainfall prediction, 3 models were used for comparison: multiple linear regression, recurrent neural networks, and long-term short-term memory. Then, the experiment was divided into 6 scenarios to import Oceanic Nino Index data and climate variable conditions from 1-6 months. Figure 8–10 shows the model prediction results for all 3 provinces: Narathiwat, Pattani, and Yala. The correlation coefficient and mean absolute error were chosen to evaluate the model's performance, which results in an efficiency evaluation, as shown in Table 4.

a. Input data for the past 1 month

b. Input data for the past 2 months

c. Input data for the past 3 months

d. Input data for the past 4 months

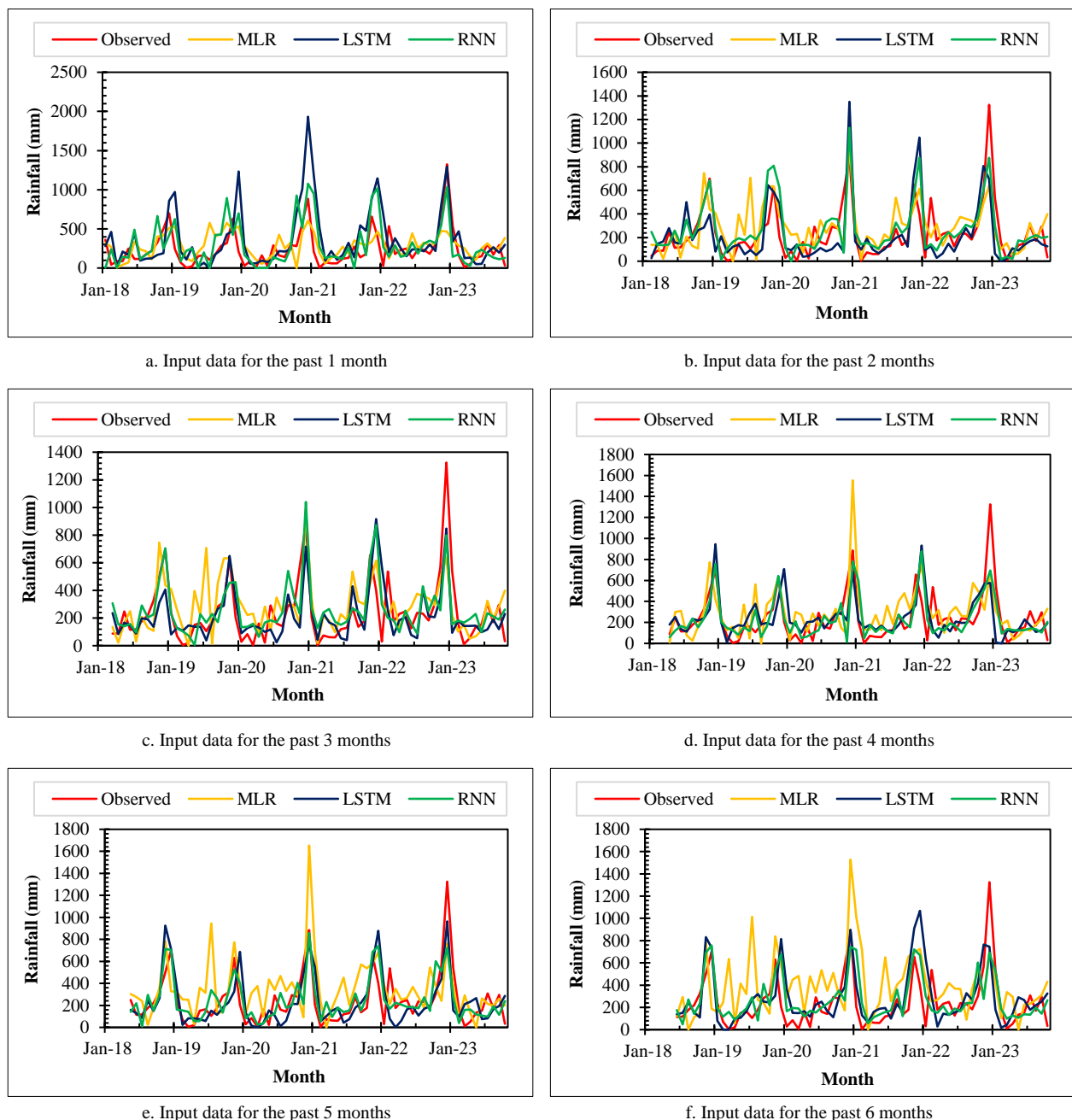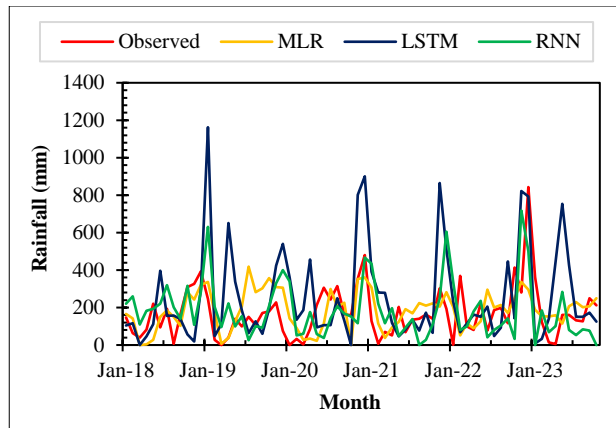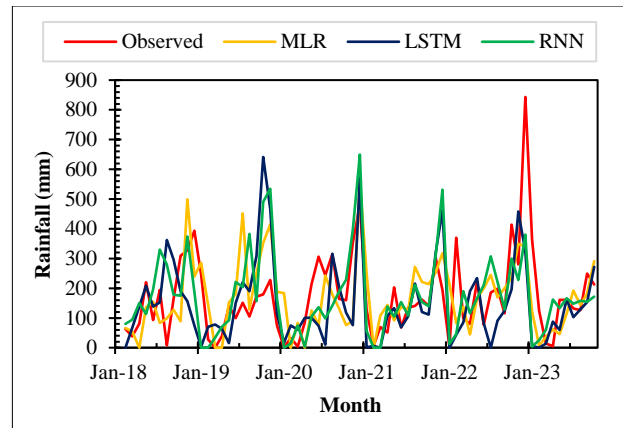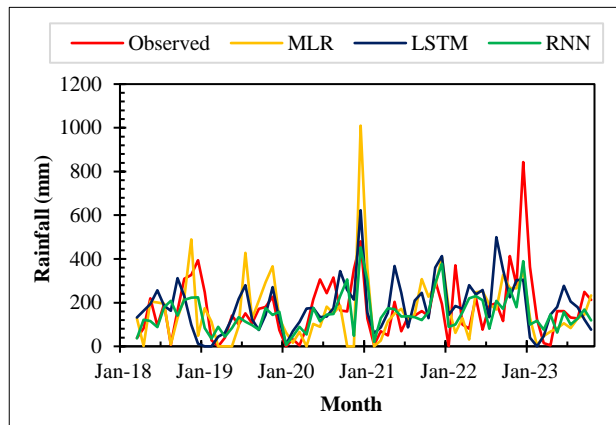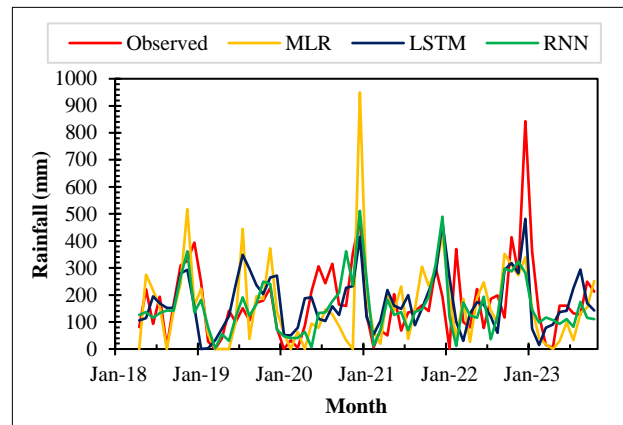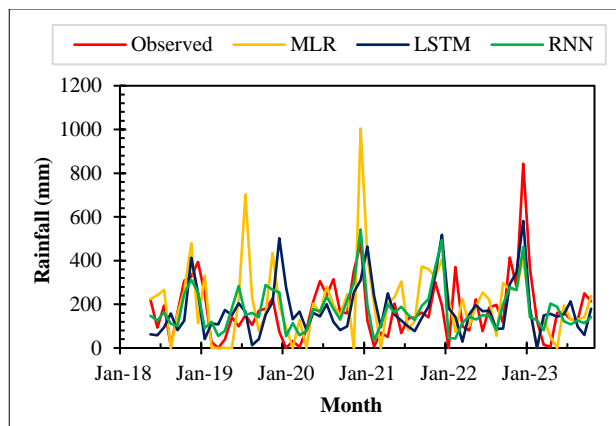e. Input data for the past 5 months

f. Input data for the past 6 months

Figure 8. Comparison of rainfall prediction results in Narathiwat Province

a. Input data for the past 1 month
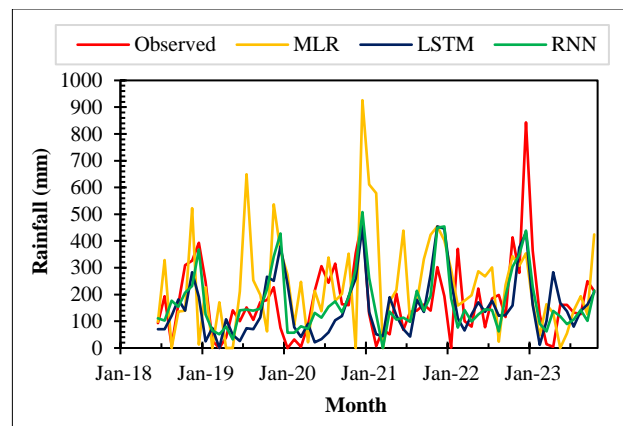
b. Input data for the past 2 months

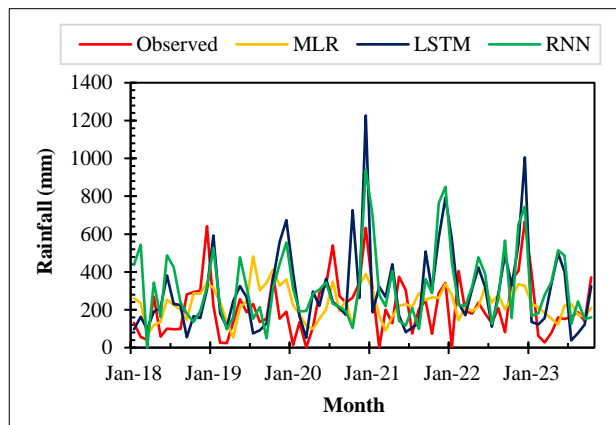c. Input data for the past 3 months
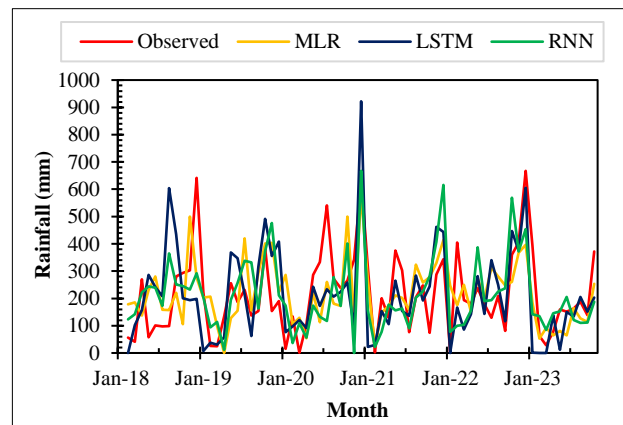
d. Input data for the past 4 months

e. Input data for the past 5 months

f. Input data for the past 6 months

Figure 9. Comparison of rainfall prediction results in Pattani Province

a. Input data for the past 1 months
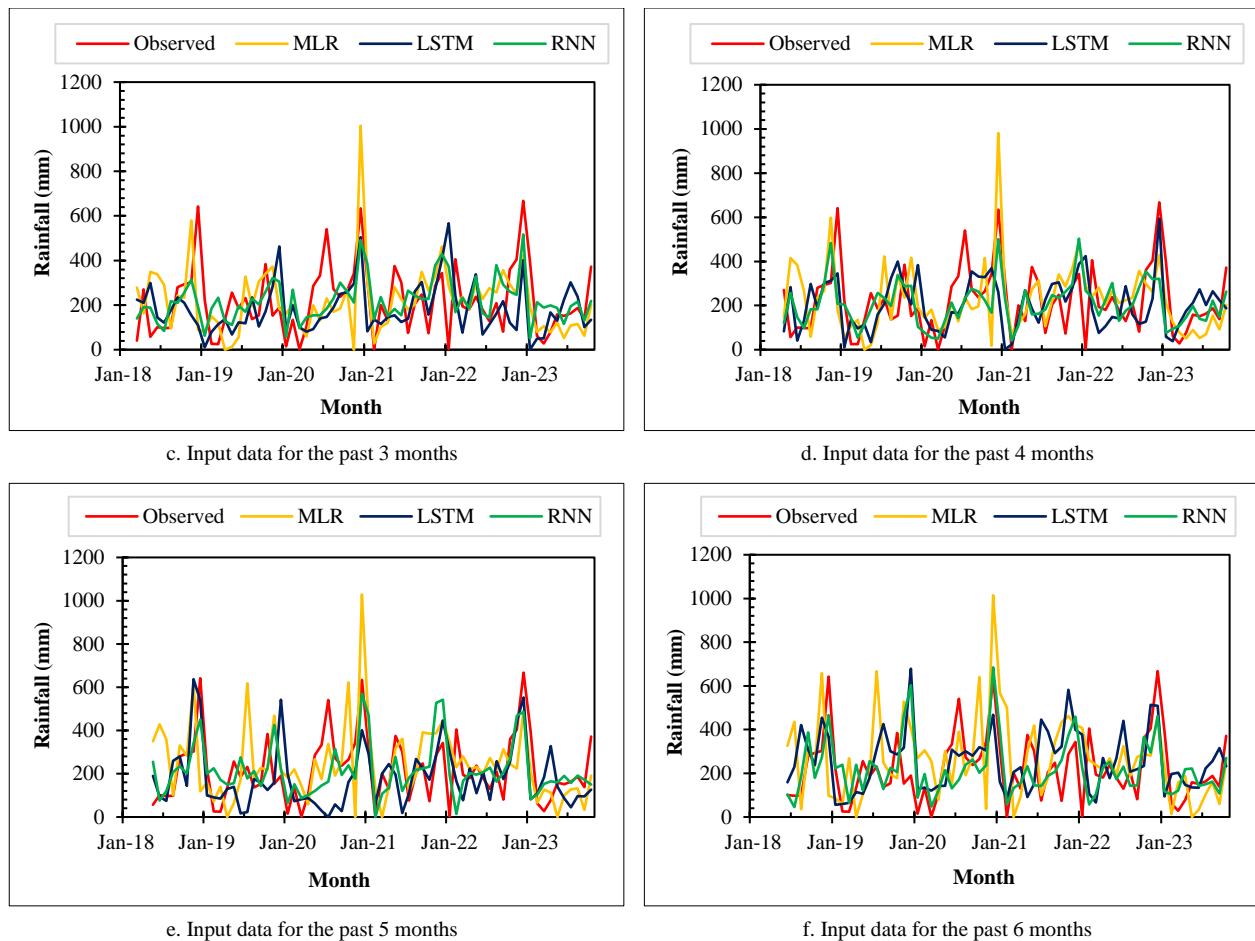
b. Input data for the past 2 months

c. Input data for the past 3 months

d. Input data for the past 4 months

e. Input data for the past 5 months

f. Input data for the past 6 months

Figure 10. Comparison of rainfall prediction results in Yala Province

Table 4. Performance comparison

Data for past (Month)	Performance Criteria	Narathiwat			Pattani			Yala		
		MLR	LSTM	RNN	MLR	LSTM	RNN	MLR	LSTM	RNN
1	r	0.49	0.61	0.59	0.42	0.31	0.25	0.45	0.41	0.30
	MAE (mm)	158.33	209.93	172.85	100.16	189.26	133.94	110.94	178.99	189.10
2	r	0.58	0.66	0.74	0.49	0.37	0.47	0.48	0.53	0.47
	MAE (mm)	152.20	133.36	113.55	99.53	107.68	101.18	107.92	115.21	113.11
3	r	0.58	0.66	0.71	0.47	0.34	0.55	0.40	0.19	0.38
	MAE (mm)	153.03	124.00	120.46	105.60	114.25	83.73	125.65	127.34	110.07
4	r	0.58	0.55	0.69	0.49	0.49	0.48	0.43	0.33	0.49
	MAE (mm)	163.77	137.03	112.76	108.66	96.72	89.05	125.72	124.51	97.67
5	r	0.50	0.69	0.74	0.41	0.40	0.62	0.40	0.39	0.53
	MAE (mm)	198.34	133.55	114.08	124.80	111.47	81.06	137.59	124.99	99.83
6	r	0.34	0.62	0.67	0.23	0.44	0.59	0.31	0.31	0.54
	MAE (mm)	250.43	145.87	126.05	162.58	101.28	87.01	159.77	134.75	97.76

Table 4 shows the results of the evaluation of the efficiency of the model. We found that when multiple linear regression was used to predict monthly rainfall for Narathiwat province, importing data from the past 3 months had the best prediction efficiency with an *MAE* value of 153.20 mm. When using Long Short Term Memory to predict monthly rainfall, we found that data from the past 3 months had to be imported into the model to achieve the best performance with an *MAE* value of 124.20 mm. However, when using a recurrent neural network for prediction, it was found that importing data from the past 4 months would make the model more efficient. The best prediction obtained from the calculation was an *MAE* value of 112.76 mm. We found that when multiple linear regression is used to predict monthly rainfall in Pattani province, importing data from the past 2 months has the best prediction efficiency with an *MAE* value of 99.53 mm. If using Long Short Term Memory to predict monthly rainfall, we found that data from the past 4 months

had to be imported into the model to achieve the best performance with an *MAE* value of 96.72 mm. However, if using a recurrent neural network for prediction, it was found that importing data from the past 5 months would make the model more efficient. In the best prediction, an *MAE* value of 81.06 mm. For Pattani province, we found that when multiple linear regression is used to predict monthly rainfall, importing data from the past 2 months, the prediction has the best efficiency with an *MAE* value of 107.02 mm. When using long short term memory to predict monthly rainfall, we found that data from the past 2 months had to be imported into the model to achieve the best performance with an *MAE* value of 115.21 mm. However, when using a recurrent neural network for prediction, it was found that importing data from the past 4 months would make the model more efficient, with the best prediction having a *MAE* value of 96.67 mm. Figure 11 shows a scatter plot of all 3 provinces, comparing the 3 best models of each model.
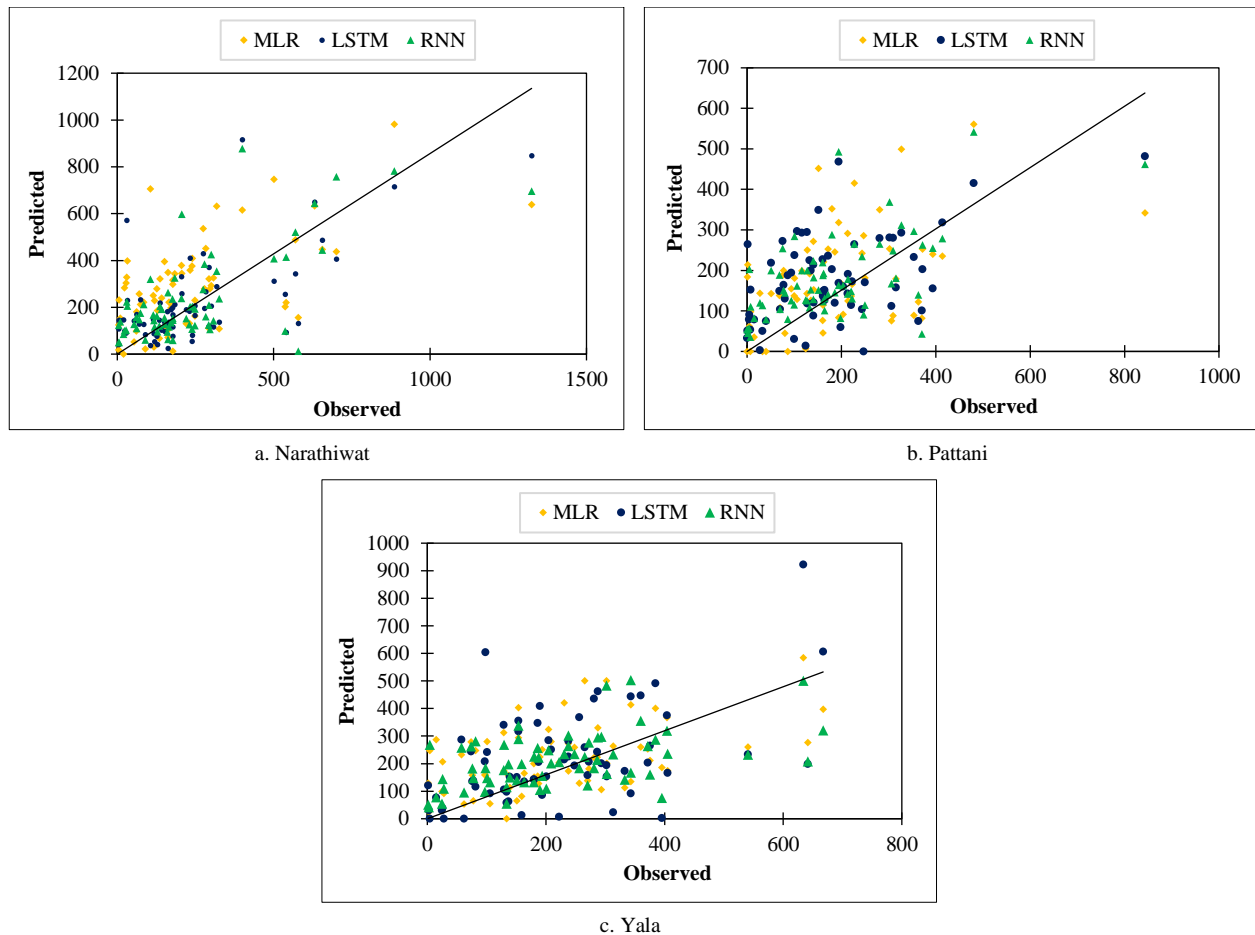

a. Narathiwat

b. Pattani

c. Yala

Figure 11. Scatter plot of all 3 provinces using the best model

The analysis of the results of evaluating the model's efficiency found that using multiple linear regression to make predictions would give the best prediction results by importing data from the past 2−3 months. Errors increase if more than 3 months of data are imported. However, when importing data from the past more than 3 months, recurrent neural networks and long-short-term memory have better prediction performance than multiple linear regression because recurrent neural networks use the previous time output for the next output. Long-short-term memory has similar prediction characteristics to recurrent neural networks, but long-short-term memory has memory [52]. However, long-term memory performance is better than recurrent neural networks if the data for training has a long sequence. Therefore, in this study, recurrent neural networks had better prediction performance than long short term memory and multiple linear regression.

6. Conclusion

This study examines the influence of the Oceanic Nino Index (ONI) on monthly climate data from 1994 to 2023 in the coastal regions of Thailand, specifically in the provinces of Narathiwat, Pattani, and Yala. It was observed that the ONI values had a significant impact on rainfall patterns in all three provinces and other climate variables. The ONI exhibited a substantial influence on monthly rainfall, affecting not only the current month but also the subsequent month; for instance, an ONI value in January would influence rainfall in both January and February. Notably, this relationship was explicitly identified in the southern-eastern region adjacent to the Gulf of Thailand, indicating potential variability in other parts of Thailand.

Compared to prior research by Ueangsawat et al. [20], who investigated the relationship between the 3-month average ONI value and the 3-month average rainfall in Chiang Mai province, Thailand, the correlation coefficients ranged from -0.093 to 0.766. This variance could be attributed to differing analytical methodologies between the studies. This study sought to develop a model for predicting monthly rainfall by leveraging the identified relationship between ONI and climate data. Three models were employed: multiple linear regression, which is relatively less complex and straightforward, and two models suitable for time-domain data prediction: recurrent neural networks and long short-term memory networks. It was found that multiple linear regression yielded better prediction results when incorporating data from the past 2–3 months. However, when utilizing historical data beyond this timeframe, recurrent neural networks and long short-term memory networks outperformed multiple linear regression, as they are adept at handling sequential data. Overall, the recurrent neural network exhibited superior performance in predicting monthly rainfall, yielding predictions of 114.08 mm, 81.06 mm, and 97.67 mm for Narathiwat, Pattani, and Yala provinces, respectively. Nonetheless, further data refinement is necessary to enhance the accuracy of monthly rainfall forecasts.

7. Declarations

7.1. Author Contributions

Conceptualization, B.C. and K.N.; methodology, S.K. and K.N.; software, S.K.; validation, B.C., K.N., and S.K.; formal analysis, B.C. and K.N.; investigation, B.C. and K.N.; resources, B.C. and K.N.; data curation, S.K.; writing—original draft preparation, B.C.; writing—review and editing, K.N.; visualization, S.K.; supervision, B.C. and K.N.; project administration, K.N.; funding acquisition, K.N. All authors have read and agreed to the published version of the manuscript.

7.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

7.3. Funding

7.4. Acknowledgements

7.5. Conflicts of Interest

The authors declare no conflict of interest.

8. References

[1] Cai, W., Ng, B., Geng, T., Jia, F., Wu, L., Wang, G., Liu, Y., Gan, B., Yang, K., Santoso, A., Lin, X., Li, Z., Liu, Y., Yang, Y., Jin, F. F., Collins, M., & McPhaden, M. J. (2023). Anthropogenic impacts on twentieth-century ENSO variability changes. Nature Reviews Earth and Environment, 4(6), 407–418. doi:10.1038/s43017-023-00427-8.

[2] Cheng, L., Abraham, J., Trenberth, K. E., Boyer, T., Mann, M. E., Zhu, J., Wang, F., Yu, F., Locarnini, R., Fasullo, J., Zheng, F., Li, Y., Zhang, B., Wan, L., Chen, X., Wang, D., Feng, L., … Lu, Y. (2024). New Record Ocean Temperatures and Related Climate Indicators in 2023. Advances in Atmospheric Sciences, 41(6), 1068–1082. doi:10.1007/s00376-024-3378-5.

[3] Glantz, M. H., & Ramirez, I. J. (2020). Reviewing the Oceanic Niño Index (ONI) to Enhance Societal Readiness for El Niño's Impacts. International Journal of Disaster Risk Science, 11(3), 394–403. doi:10.1007/s13753-020-00275-w.

[4] L'Heureux, M. L., Tippett, M. K., Wheeler, M. C., Nguyen, H., Narsey, S., Johnson, N., Hu, Z. Z., Watkins, A. B., Lucas, C., Ganter, C., Becker, E., Wang, W., & Di Liberto, T. (2024). A Relative Sea Surface Temperature Index for Classifying ENSO Events in a Changing Climate. Journal of Climate, 37(4), 1197–1211. doi:10.1175/JCLI-D-23-0406.1.

[5] Prasetyo, Y., & Nabilah, F. (2017). Pattern Analysis of El Nino and la Nina Phenomenon Based on Sea Surface Temperature (SST) and Rainfall Intensity using Oceanic Nino Index (ONI) in West Java Area. IOP Conference Series: Earth and Environmental Science, 98(1), 12041. doi:10.1088/1755-1315/98/1/012041.

[6] Varotsos, C., Sarlis, N. V., Mazei, Y., Saldaev, D., & Efstathiou, M. (2024). A Composite Tool for Forecasting El Niño: The Case of the 2023–2024 Event. Forecasting, 6(1), 187–203. doi:10.3390/forecast6010011.

[7] Silva, K. A., de Souza Rolim, G., & de Oliveira Aparecido, L. E. (2022). Forecasting El Niño and La Niña events using decision tree classifier. Theoretical and Applied Climatology, 148(3–4), 1279–1288. doi:10.1007/s00704-022-03999-5.

[8] Wang, G. G., Cheng, H., Zhang, Y., & Yu, H. (2023). ENSO analysis and prediction using deep learning: A review. Neurocomputing, 520, 216–229. doi:10.1016/j.neucom.2022.11.078.

[9] Bouach, A. (2024). Artificial neural networks for monthly precipitation prediction in north-west Algeria: a case study in the Oranie-Chott-Chergui basin. Journal of Water and Climate Change, 15(2), 582–592. doi:10.2166/wcc.2024.494.

[10] Van Oldenborgh, G. J., Hendon, H., Stockdale, T., L'Heureux, M., Coughlan De Perez, E., Singh, R., & Van Aalst, M. (2021). Defining El Nio indices in a warming climate. Environmental Research Letters, 16(4), 44003. doi:10.1088/1748-9326/abe9ed.

[11] Bochenek, B., & Ustrnul, Z. (2022). Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives. Atmosphere, 13(2), 180. doi:10.3390/atmos13020180.

[12] Haggag, M., Siam, A. S., El-Dakhakhni, W., Coulibaly, P., & Hassini, E. (2021). A deep learning model for predicting climate-induced disasters. Natural Hazards, 107(1), 1009–1034. doi:10.1007/s11069-021-04620-0.

[13] Kumar, V., Azamathulla, H. Md., Sharma, K. V., Mehta, D. J., & Maharaj, K. T. (2023). The State of the Art in Deep Learning Applications, Challenges, and Future Prospects: A Comprehensive Review of Flood Forecasting and Management. Sustainability, 15(13), 10543. doi:10.3390/su151310543.

[14] Zhang, Y., Xie, D., Tian, W., Zhao, H., Geng, S., Lu, H., Ma, G., Huang, J., & Choy Lim Kam Sian, K. T. (2023). Construction of an Integrated Drought Monitoring Model Based on Deep Learning Algorithms. Remote Sensing, 15(3), 667. doi:10.3390/rs15030667.

[15] Apipattanavis, S., Ketpratoom, S., & Kladkempetch, P. (2018). Water Management in Thailand. Irrigation and Drainage, 67(1), 113–117. doi:10.1002/ird.2207.

[16] Maprasit, S., Pradabphetrat, P., Madmanang, R., Sathawong, S., Boonkaew, R., & Suksaroj, C. (2021). Physical-Chemical Properties Relationship of Pattani River and Implication for Water Quality Monitoring Study and Academic Service. Journal of Physics: Conference Series, 1835(1), 12112. doi:10.1088/1742-6596/1835/1/012112.

[17] Nur Amyliyana Wan Faizurie Zaidee, W., Shakir Mohd Saudi, A., Khairul Amri Kamarudin, M., Ekhwan Toriman, M., Juahir, H., Fahmy Abu, I., Mahmud Nur Zahidah Shafii, M., Nizam, K., & Elfithri, R. (2018). Flood Risk Pattern Recognition Using Chemometric Techniques Approach in Golok River, Kelantan. International Journal of Engineering & Technology, 7(3.14), 75. doi:10.14419/ijet.v7i3.14.168655.

[18] Hidayat, R., Ando, K., Masumoto, Y., & Luo, J. J. (2016). Interannual Variability of Rainfall over Indonesia: Impacts of ENSO and IOD and Their Predictability. IOP Conference Series: Earth and Environmental Science, 31(1), 12043. doi:10.1088/1755-1315/31/1/012043.

[19] Irwandi, H., Pusparini, N., Ariantono, J. Y., Kurniawan, R., Tari, C. A., & Sudrajat, A. (2018). The Influence of ENSO to the Rainfall Variability in North Sumatra Province. IOP Conference Series: Materials Science and Engineering, 335(1), 12055. doi:10.1088/1757-899X/335/1/012055.

[20] Ueangsawat, K., Nilsamranchit, S., & Jintrawet, A. (2015). Fate of ENSO Phase on Upper Northern Thailand, a Case Study in Chiang Mai. Agriculture and Agricultural Science Procedia, 5, 2–8. doi:10.1016/j.aaspro.2015.08.001.

[21] Ahmed, H. A. Y., & Mohamed, S. W. A. (2021). Rainfall Prediction using Multiple Linear Regressions Model. 2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), Khartoum, Sudan. doi:10.1109/iccceee49695.2021.9429650.

[22] Shaker Reddy, P. C., & Sureshbabu, A. (2019). An Enhanced Multiple Linear Regression Model for Seasonal Rainfall Prediction. International Journal of Sensors, Wireless Communications and Control, 10(4), 473–483. doi:10.2174/2210327910666191218124350.

[23] Agboola, A., Gabriel, A., Aliyu, E., & Alese, B. (2014). Development Of A Fuzzy Logic Based Rainfall Prediction Model. International Journal of Engineering & Technology, 3(4), 427–435.

[24] Janarthanan, R., Balamurali, R., Annapoorani, A., & Vimala, V. (2021). Prediction of rainfall using fuzzy logic. Materials Today: Proceedings, 37, 959–963. doi:10.1016/j.matpr.2020.06.179.

[25] Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. A. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. Machine Learning with Applications, 7, 100204. doi:10.1016/j.mlwa.2021.100204.

[26] Rahman, A., Abbas, S., Gollapalli, M., Ahmed, R., Aftab, S., Ahmad, M., Khan, M. A., & Mosavi, A. (2022). Rainfall Prediction System Using Machine Learning Fusion for Smart Cities. Sensors, 22(9), 3504. doi:10.3390/s22093504.

[27] Mohammadi, B. (2021). A review on the applications of machine learning for runoff modeling. Sustainable Water Resources Management, 7(6), 98. doi:10.1007/s40899-021-00584-y.

[28] Singh, A. K., Kumar, P., Ali, R., Al-Ansari, N., Vishwakarma, D. K., Kushwaha, K. S., Panda, K. C., Sagar, A., Mirzania, E., Elbeltagi, A., Kuriqi, A., & Heddam, S. (2022). An Integrated Statistical-Machine Learning Approach for Runoff Prediction. Sustainability, 14(13), 8209. doi:10.3390/su14138209.

[29] Hu, Z., Zhang, Y., Zhao, Y., Xie, M., Zhong, J., Tu, Z., & Liu, J. (2019). A Water Quality Prediction Method Based on the Deep LSTM Network Considering Correlation in Smart Mariculture. Sensors, 19(6), 1420. doi:10.3390/s19061420.

[30] Melesse, A. M., Khosravi, K., Tiefenbacher, J. P., Heddam, S., Kim, S., Mosavi, A., & Pham, B. T. (2020). River water salinity prediction using hybrid machine learning models. Water (Switzerland), 12(10), 1–21. doi:10.3390/w12102951.

[31] Rajabi-Kiasari, S., & Hasanlou, M. (2020). An efficient model for the prediction of SMAP sea surface salinity using machine learning approaches in the Persian Gulf. International Journal of Remote Sensing, 41(8), 3221–3242. doi:10.1080/01431161.2019.1701212.

[32] Mosavi, A., Ozturk, P., & Chau, K. (2018). Flood Prediction Using Machine Learning Models: Literature Review. Water, 10(11), 1536. doi:10.3390/w10111536.

[33] Motta, M., de Castro Neto, M., & Sarmento, P. (2021). A mixed approach for urban flood prediction using Machine Learning and GIS. International Journal of Disaster Risk Reduction, 56, 102154. doi:10.1016/j.ijdrr.2021.102154.

[34] Htike, K. K., & Khalifa, O. O. (2010). Rainfall forecasting models using focused time-delay neural networks. International Conference on Computer and Communication Engineering (ICCCE'10), Kuala Lumpur, Malaysia. doi:10.1109/iccce.2010.5556806.

[35] Hong, W. C. (2008). Rainfall forecasting by technological machine learning models. Applied Mathematics and Computation, 200(1), 41–57. doi:10.1016/j.amc.2007.10.046.

[36] Sivapragasam, C., Liong, S. Y., & Pasha, M. F. K. (2001). Rainfall and runoff forecasting with SSA-SVM approach. Journal of Hydroinformatics, 3(3), 141–152. doi:10.2166/hydro.2001.0014.

[37] Feng, Q., Wen, X., & Li, J. (2015). Wavelet Analysis-Support Vector Machine Coupled Models for Monthly Rainfall Forecasting in Arid Regions. Water Resources Management, 29(4), 1049–1065. doi:10.1007/s11269-014-0860-3.

[38] Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2022). A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. IEEE Transactions on Neural Networks and Learning Systems, 33(12), 6999–7019. doi:10.1109/TNNLS.2021.3084827.

[39] Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. Neural Computation, 31(7), 1235–1270. doi:10.1162/neco_a_01199.

[40] Van Houdt, G., Mosquera, C., & Nápoles, G. (2020). A review on the long short-term memory model. Artificial Intelligence Review, 53(8), 5929–5955. doi:10.1007/s10462-020-09838-1.

[41] Van, S. P., Le, H. M., Thanh, D. V., Dang, T. D., Loc, H. H., & Anh, D. T. (2020). Deep learning convolutional neural network in rainfall-runoff modelling. Journal of Hydroinformatics, 22(3), 541–561. doi:10.2166/hydro.2020.095.

[42] Haidar, A., & Verma, B. (2018). Monthly Rainfall Forecasting Using One-Dimensional Deep Convolutional Neural Network. IEEE Access, 6, 69053–69063. doi:10.1109/ACCESS.2018.2880044.

[43] Van Viet, L. (2021). Development of a new ENSO index to assess the effects of ENSO on temperature over southern Vietnam. Theoretical and Applied Climatology, 144(3–4), 1119–1129. doi:10.1007/s00704-021-03591-3.

[44] Asuero, A. G., Sayago, A., & González, A. G. (2006). The correlation coefficient: An overview. Critical Reviews in Analytical Chemistry, 36(1), 41–59. doi:10.1080/10408340500526766.

[45] Mardia, K. V. (1976). Linear circular correlation coefficients and rhythmometry. Biometrika, 63(2), 403–405. doi:10.2307/2335637.

[46] Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. Applied Soft Computing, 97, 105524. doi:10.1016/j.asoc.2019.105524.

[47] Uyanık, G. K., & Güler, N. (2013). A Study on Multiple Linear Regression Analysis. Procedia - Social and Behavioral Sciences, 106, 234–240. doi:10.1016/j.sbspro.2013.12.027.

[48] Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. International Journal of Forecasting, 37(1), 388–427. doi:10.1016/j.ijforecast.2020.06.008.

[49] Medsker, L. R., & Jain, L. (2001). Recurrent neural networks. Design and Applications, 5(64-67), 2.

[50] Lin, Y., Yan, Y., Xu, J., Liao, Y., & Ma, F. (2021). Forecasting stock index price using the CEEMDAN-LSTM model. North American Journal of Economics and Finance, 57, 101421. doi:10.1016/j.najef.2021.101421.

[51] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669. doi:10.1016/j.ejor.2017.11.054.

[52] Apaydin, H., Feizi, H., Sattari, M. T., Colak, M. S., Shamshirband, S., & Chau, K. W. (2020). Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting. Water, 12(5), 1500. doi:10.3390/w12051500.