



Available online at <http://scik.org>

Commun. Math. Biol. Neurosci. 2022, 2022:118

<https://doi.org/10.28919/cmbn/7761>

ISSN: 2052-2541

SPATIOTEMPORAL CONVOLUTIONAL LSTM WITH ATTENTION MECHANISM FOR MONTHLY RAINFALL PREDICTION

RENALDY FREDYAN, GEDE PUTRA KUSUMA*

Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina Nusantara
University, Jakarta 11480, Indonesia

Copyright © 2022 the author(s). This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract: Climate has always become part of an essential role in human life on earth. Climate change is a scorching topic because of various prediction efforts to predict the future of the earth's climate on human survival. As one institution that deals with climatology in detail, NOAA publishes a journal on the approach used to generate CHIRPS as a spatiotemporal dataset. The data consist of spatial two dimensions and temporal data from 1981 to 2020. The dataset for the experiment contains 480 pictures, which are 480 months or 40 years of data. K-means clustering is used to collect different characteristics of spatial data based on the Within Cluster Sum Square (WCSS) value using the Elbow method. Generalizing four spatial areas of different characteristic data, Authors have proposed a Convolutional Long Short-Term Memory model to carry out and develop by an attention mechanism. A deep learning architecture designed to learn the dependencies of temporal and spatial data exclusively utilizing convolutional layers and recurrent layers. Two issues with convolutional networks' ability to forecast sequences using historical data are addressed by our suggested architecture: (1) break the temporal order throughout the learning process, and (2) demand that the lengths of the input and output sequences be equal. The method's purpose is to improve the prediction for each previous

*Corresponding author

E-mail address: inegara@binus.edu

Received October 1, 2022

method and having lower error than other models such as GRU, LSTM, and LSTM-AT. Still, every model reported in this paper has characteristics of how they learn the pattern of the data and visualize the error rate and performance in every spatial point. The Convolutional LSTM-AT model achieved MAE of 67.97, with training time around 2.5 minutes.

Keywords: rainfall prediction; spatiotemporal data; CHIRPS data; convolutional LSTM-AT model; attention mechanism.

2010 AMS Subject Classification: 68T05, 97R40, 97R30.

1. INTRODUCTION

Rainfall forecast information is one of the crucial analyses to help regulate water resources and often involves several variables since rainfall is part of meteorological phenomena. This prediction is more complicated when dealing with the emergence of climate change in tropical areas such as Indonesia, which lies on the equator with implications from North and South. Furthermore, climate change has affected rainfall patterns, causing several natural disasters such as heavy rains that result in flooding or prolonged absence of precipitation that results in droughts [1]. Drought Management Plans (DMPs) are regulatory instruments that establish priorities among different water uses and define more stringent constraints for access to publicly available water during droughts and reduced water supplies because of climate change vulnerability to drought events. To deal with this problem, rainfall prediction with an excellent and accurate method is needed to anticipate it [2]. Precise rainfall forecasts, both short and long term, have significant benefits in water resource management, flood control, disaster reduction, and agricultural management [3]. However, rainfall is a complicated nonlinear atmospheric system that depends on space and time; besides, many factors can influence rain in the area [4].

Therefore, it is not never convenient to realize the complexity and uncertainty of the predictability of rainfall to produce precise and accurate rainfall forecasts [5]. Forecasting rainfall, the beginning of the rainy season, the duration of the rain, and the end of the rainy season are determined by a monthly period, often using a three-month system known as the SPI (Standardized Precipitation Index) method [6]. In addition, monthly rainfall can provide a more accurate

distribution of the mean intra-year rain when compared to seasonal rainfall [7]. Hence, it is vital to periodically estimate rainfall on a monthly time scale, in which rainfall predictions are usually made using physical-based models and deep learning methods [8]. Climate Hazards Group Infrared Precipitation with Stations, also known as (CHIRPS), is data obtained with specifications such as environmental records, new quasi-global (50 ° S-50 ° N), high resolution (0.05 °), daily, pentadal, and monthly rainfall datasets. These datasets have the spatial surface of the earth and temporal from 1981 to 2020, which are able to visualize the rainfall condition in every place on the land. Scientists developed CHIRPS from various countries to support the United States Agency for International Development Famine Early Warning Systems Network (FEWS NET) [9]. The approach is built using thermal infrared precipitation (TIR), which has been successful in trials like the National Oceanic and Atmospheric Administration's (NOAA) Rainfall Estimate CHIRPS uses Tropical Rainfall Measuring Mission Multi-Satellite Precipitation Analysis version 7 to calibrate global Cold Cloud Duration (CCD) rainfall forecasts. In addition, CHIRPS also employs the current state-of-the-art interpolation measurement approach using an 'intelligent interpolation' approach that can work with anomalies in high-resolution climatology [10].

CHIRPS is one of the unique spatiotemporal data that require specific consideration when utilizing deep learning predictions such as LSTM and GRU. While the deep learning approaches mostly used temporal data to build the model, spatiotemporal has always had a different way when choosing suitable algorithms. Understanding the features is necessary, which has high-dimensional and temporally correlation, which means the data are indexed by up to two dimensions in space and one in time [11]. In general, spatiotemporal data is spatial of the correlation between nearby locations like a photo with the pixel and temporal correlation between adjacent timestamps [12]. To handle spatiotemporal data, Tao *et al.*[3] used attention mechanisms to enhance the prediction model, the *state of the art* of attention mechanism was invented by Bahdanau [13] to improve the accuracy of machine translation algorithms. Another attention model comes from Vaswani *et al.* [14], which is Multi-Head Attention, and some people make the robust algorithm, namely transformers.

Deep Learning fields are always spread over several areas of prediction and classification. In

sequential or time-series data, Recurrent Neural Networks (RNN) and their derivatives maintain some vectors in calculating every neuron using propagated through time [15]. However, RNN has trouble dealing with long data sequences that preside vanishing gradient problems when training using traditional RNN with long legs [16]. LSTM comes with the solution using memory to improve RNN and avoid vanishing gradient, becoming more advanced with modifications such as encoder-decoder, attention mechanism, etc. [17].

In this study, the authors propose Convolutional LSTM with an additional Attention Layer to enhance the accuracy of monthly rainfall prediction using CHIRPS data as a solution to predict rainfall with gridding data to make more accurate forecasting. A hyperparameter is tuned manually with an endless number of models to ensure that it has the same comparison. We compare and analyze each model's loss error and the number of performances according to the evaluation metrics most used in hydrology and deep learning. The results indicate that the proposed Convolutional LSTM-AT model is the best so far. We also analyze the spatial and temporal for interpreting the physical causality of our model.

2. RELATED WORKS

In this section, the authors give reviews of relevant research that can inspire the author to construct the Convolutional LSTM-AT model, including several fundamental studies in rainfall forecasting, sequential data using LSTM-based models, and the exciting method of machine translation, which is an attention-based model.

2.1. Rainfall Forecasting

Seasonal prediction models are commonly used for the prediction of rainfall to make an early warning from the tools or agency of government when hydrological extremes come to attack people. Based on climatologists, climate prediction models can be classified into three approaches first physical or numerical approach, the second empirical or statistical approach, and the third mixing between physical-empirical [18]. Still, rainfall depends on numerous lands, large oceans, and the atmosphere in the top of the processes. On the other hand, Physical models are generally developed based on interpretations of atmospheric processes, but they frequently show weak

predictability in providing good information on annual climate variability [19]. In general, physical-empirical prediction models, which are the most used by climatologists, are developed utilizing the traditional statistical approach. For instance, Zhu and Li [20] 2017 applied the regression method to predict the wet season in East Asia. Li and Wang [21] 2018 studied the forecast capability of summertime highly extreme rainfall days in eastern China by utilizing a stepwise regression model. The ability of the traditional regression-based method is likewise inadequate in forecasting highly nonlinear and nonstationary performance. Therefore, the connection of local climate in a specific area with ocean-atmospheric variables such as SST or sea level pressure cannot be described by employing traditional regression models [22].

2.2. LSTM-Based Methods

Long Short Term Memory (LSTM) is one of the modified versions of recurrent neural networks that have a problem with vanishing gradient, which is designed to resolve the problem of sequential long-distance (time) data reliance by Hochreiter and Schmidhuber [23]. Yuan *et al.* [24] proposed an LSTM network model to build occupancy by simulating energy, operation, and management. ElSaadani *et al.* [25] used the LSTM model to predict soil moisture and fill gaps between the observation. Further, Zhou *et al.* [26] combined the LSTM model and attention mechanism based on machine-translation to recognize skeleton-based abnormal behavior. The conclusions indicated that attention-based LSTM could recognize behavior better than only LSTM Model.

2.3. Attention-based methods

In deep learning, one way to increase accuracy in the model learning process is through attention mechanisms inspired by selective human visuals to choose which information to pay special attention to and which ones to reject. In general, the application of Attention mechanism has been applied in various areas of research and industry, such as machine translation, image captioning, and video motion recognition. Song *et al.* [27] have a proposal related to an end-to-end spatiotemporal attention model to perform recognition and prediction of human action in a video frame. In addition, Chen *et al.* [28] proposed a model of spatial combined with channel attention and image labeling with an additional convolutional neural network, having a good result in their data set. Ding *et al.* [29] proposed spatiotemporal LSTM to predict floods in three basins

in China. Tao *et al.* [3] also proposed LSTM with an attention mechanism to improve monthly rainfall prediction, which performed well in most spatial points. From the above model author was inspired to develop another model, we propose a multi-head attention LSTM to optimize monthly rainfall prediction with spatiotemporal data.

3. STUDY AREA AND DATASET

In this study, Kalimantan Timur was selected as the study area to evaluate and compare the performance of several LSTM models in forecasting monthly rainfall. East Kalimantan is located at 2°33' North Latitude -2°25' South Latitude, 113°44' - 119°00' East Longitude, with a land area of 127,346.92 km² and a sea area of 25,656 km². The occurrence of climate change causes changes in rainfall patterns significantly; this is exacerbated by the event of El Nino, which can bring extreme rain, or it might be because El Nina causes long drought.

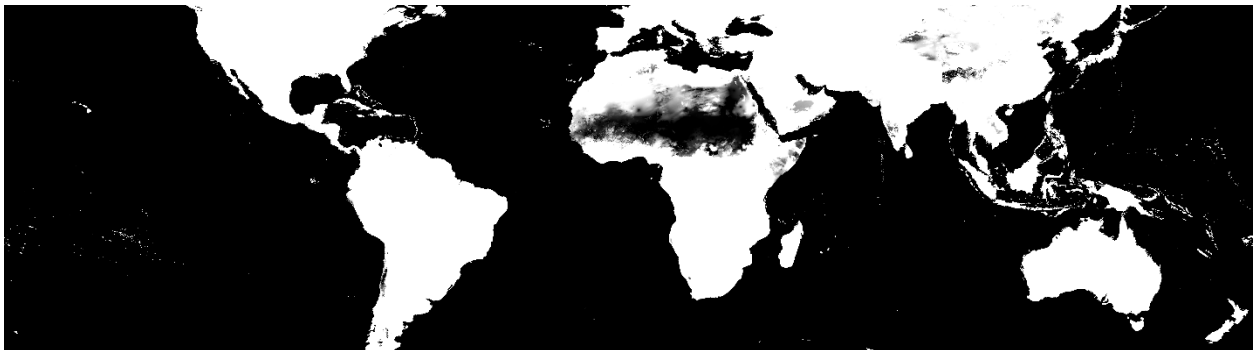


FIGURE 1. Sample data CHIRPS of single monthly rainfall



FIGURE 2. Sample data CHIRPS of single monthly rainfall after cropping in Kalimantan Timur area

MONTHLY RAINFALL PREDICTION

Monthly rainfall data covering January 1980 - December 2020 is CHIRPS data accessible from <https://data.chc.ucsb.edu/products/CHIRPS-2.0/>. The data for the 40 years January 1980 to December 2020 was used as a dataset of this model, as shown in Figure 1, sampling of December 2020. Rainfall data known as CHIRPS is still in the form of worldwide raster data, where the research only focuses on the Kalimantan Timur region, so the data needs to be split. First, a printout of the Kalimantan Timur area is required from <https://tanahair.indonesia.go.id/>. Still, combining the data using the ArcGIS application is necessary because the custom is city and district data. Furthermore, after the data for the East Kalimantan region is obtained, splitting the rainfall data worldwide using the SAGA application is needed. It should be noted that the Split process requires degrees of longitude and degrees of latitude and a grid size that must be adapted to raster data worldwide, which is $0.05^\circ \times 0.05^\circ$, the result can be seen in Figure 2.

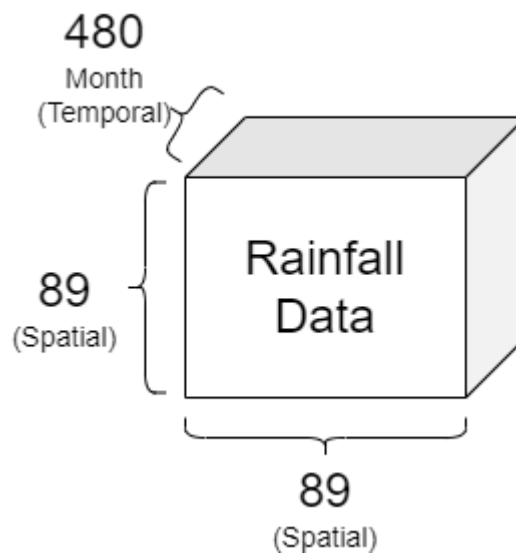


FIGURE 3 Illustrated spatiotemporal rainfall data with all sizes that are high, weight, and deep.

As shown in Figure 2, data visualization has colors black and white which means black has representative sea surface and white island surface. Raster data is one of the best formats of data to represent surface area since raster can keep multi-band of data to create complex spatial conditions. CHIRPS contain a single band to interpret monthly precipitation values without additional variables. It can be seen in Figure 2 that data has three dimensionality of perspective.

As shown in Figure 3, this data includes dimensions 89 x 89 of spatial and 480 of temporal, in this case, monthly data. Having three-dimensional condition make this research more complex since it should be done with a specific method, so the spatial and temporal will not be biased or even removable on that dimension.

4. PROPOSED METHOD

Rainfall is critical in supporting human life; besides, various policies often consider rainfall the main factor. Based on rainfall data, climate classification can be done according to the ratio between the average dry months and the average number of wet months. The dry month occurs when the monthly rainfall is less than 60 mm/month, while the wet month occurs when the monthly rainfall is above 100 mm/month. A humid month occurs between the dry and wet months when the monthly rainfall is between 60-100 mm/month

4.1. Overview

Variational data and models increase with many perspectives to understand data to build the best alternative model. The authors have searched for and understood a literature review to know which is the newest and best model or the strengths and weaknesses of those models. Still, the rainfall forecasting model suffers from predicting rainfall accurately and precisely. Hence, the authors built the proposed model Convolutional LSTM-AT as an alternative solution to optimize monthly rainfall prediction with the spatiotemporal dataset.

4.2. Data Preprocessing

The data preprocessing stage is the data selection stage which aims to obtain relevant data for use. In raw data, missing values are often found, not stored values (misrecording), data sampling that is not good enough, and others. However, because this research does not use raw data but secondary data, preprocessing will be done to process spatial and temporal data. In addition, preprocessing will only be conducted to focus on the data on cells with value, so the cells with no data will not be used.

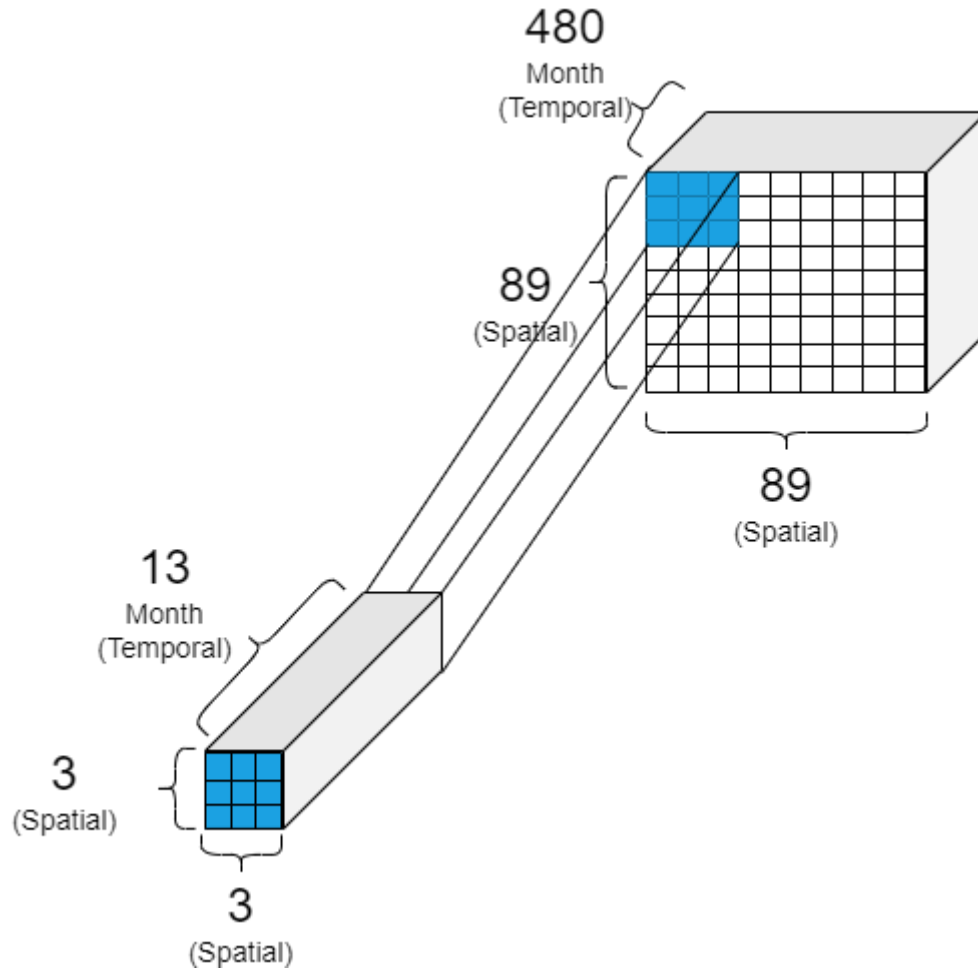


FIGURE 4. Illustrated spatiotemporal data using the sliding window in spatial perspective

In this study, focal operation theory is implemented, a spatial function to calculate the output value of each cell using neighborhood values, like the nearest neighbors' algorithm (K-NN), a machine learning algorithm, as shown in Figure 4 [30]. In addition, this theory is also commonly used in convolution, kernel, and moving windows in deep learning algorithms such as CNN or RNN. Moving Window can be imagined as an arrangement of square cells with a specific size, which in this study is 3 x 3 in size, which shifts its position with certain steps. As the operation is applied to each cell of the moving window, the values in the raster tend to be smoother. It was adopted in this study to smooth the predictive value in spatial conditions.

Spatial-temporal data are generally placed in continuous space, while classical data sets such as images or video data are usually in a discrete area. Spatiotemporal data patterns usually present

very complex spatial and temporal properties, and correlations between data are challenging to explain with traditional methods. Finally, one of the standard statistical assumptions is that the sample is obtained independently. However, this does not apply in spatiotemporal analysis because Spatiotemporal data tend to be highly correlated, so it is impossible to carry out separate studies. As explained earlier, the data used in each time unit (temporal) is 89x89 with a length of 480 temporal, as shown in Figure 4. Hence for modeling, the data is taken spatially with a size of 3x3 for 13 months (temporal), and this data will slice the sliding window along the temporal axis. Moving to the right side with a single step will be implemented in the data, so after the last window on the right area, it will continue by a sliding window in the next row, from left to right. It can be seen in the blue area in Fig. 4 until the end of the spatial data, which is the right bottom side.

4.3. Data Clustering

One of the data mining techniques is clustering to find similarities in character in the group data; this technique is included in traditional machine learning studies and also becoming part of the unsupervised algorithm, which only requires training data without target data [31]. In theory, cluster analysis is one of the tools to group data based on variables or features to maximize the resemblance of characteristics within the cluster and maximize the differences between clusters themselves [32]. The popular algorithm is the K-means clustering algorithm groups data based on the distance between the data and the cluster centroid point obtained through an iterative process [33]. The analysis needs to determine the number of K as input to the algorithm.

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c^j\|^2 \quad (1)$$

Following the Eq. (1), J is the objective function, k is many clusters, n is the number of cases, x is a case in i , and c is the centroid for cluster J itself. In k-means clustering, this distance can be measured using distances: Euclidean distance, Manhattan distance, A-squared Euclidean distance measure, and Cosine distance measure. The choice of this distance measurement method will affect how the algorithm calculates the similarity in the cluster and shape. Nevertheless, some of the problems come when determining the number of k because no theory

states how to choose it very well since the number of k is very essential to searching the cluster. The researcher solves this problem using the Elbow Method, which is obtained by performing a visual assessment of the line graph where the x-axis is the number of K , and the Y-axis is the Within Cluster Sum Square (WCSS) value.

4.4. Convolutional LSTM-AT

LSTM is derivative from the RNN in sequential data study, having three units of gates such as input, output, and forget gate. It allows the gates of LSTM to store and access information or characteristics of the data over a while, dependence to Hochreiter and Schmidhuber [23], mitigating the vanishing gradient problem. The model parameter including all the input is weight W or U and the bias term b , i_t , f_t , o_t , \tilde{C}_t respectively represent input-output forget and memory, the other symbols are h_t σ meaning hidden state and sigmoid activation function, but it always depends on the data, sometime can be changed becoming hyperbolic tangent or ReLU [29] [34] [35].

$$i_t = \sigma(h_{t-1}W^i + x_tU^i + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(h_{t-1}W^c + x_tU^c + b_c) \quad (3)$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{C}_t \quad (4)$$

$$f_t = \sigma(h_{t-1}W^f + x_tU^f + b_f) \quad (5)$$

$$o_t = \sigma(h_{t-1}W^o + x_tU^o + b_o) \quad (6)$$

$$h_t = o_t * \tanh(C_t) \quad (7)$$

The Attention Mechanism is often used to optimize sequence handling models in some deep learning approaches. The preprocessed input sequence is processed by the neural network or a mapping function to generate the raw attention weights. Then with the soft-maxed attention weights and raw input values, they can calculate the final output. According to the value of the weight of attention, the Attention Model can be divided into two types: hard attention and soft

attention. Hard attention refers to selecting a single input data feature, which means the attention weight can only be 0 or 1. Soft attention refers to a weight between 0 and 1, and the range of weight selection is more flexible [36]. Since those several models of attention were invented by Bahdanau *et al.* [13] and Multi-Head Attention by Vaswani *et al.* [14], empirically, additive attention can improve the modal and attention layer's performance and make the unit's weight noticed.

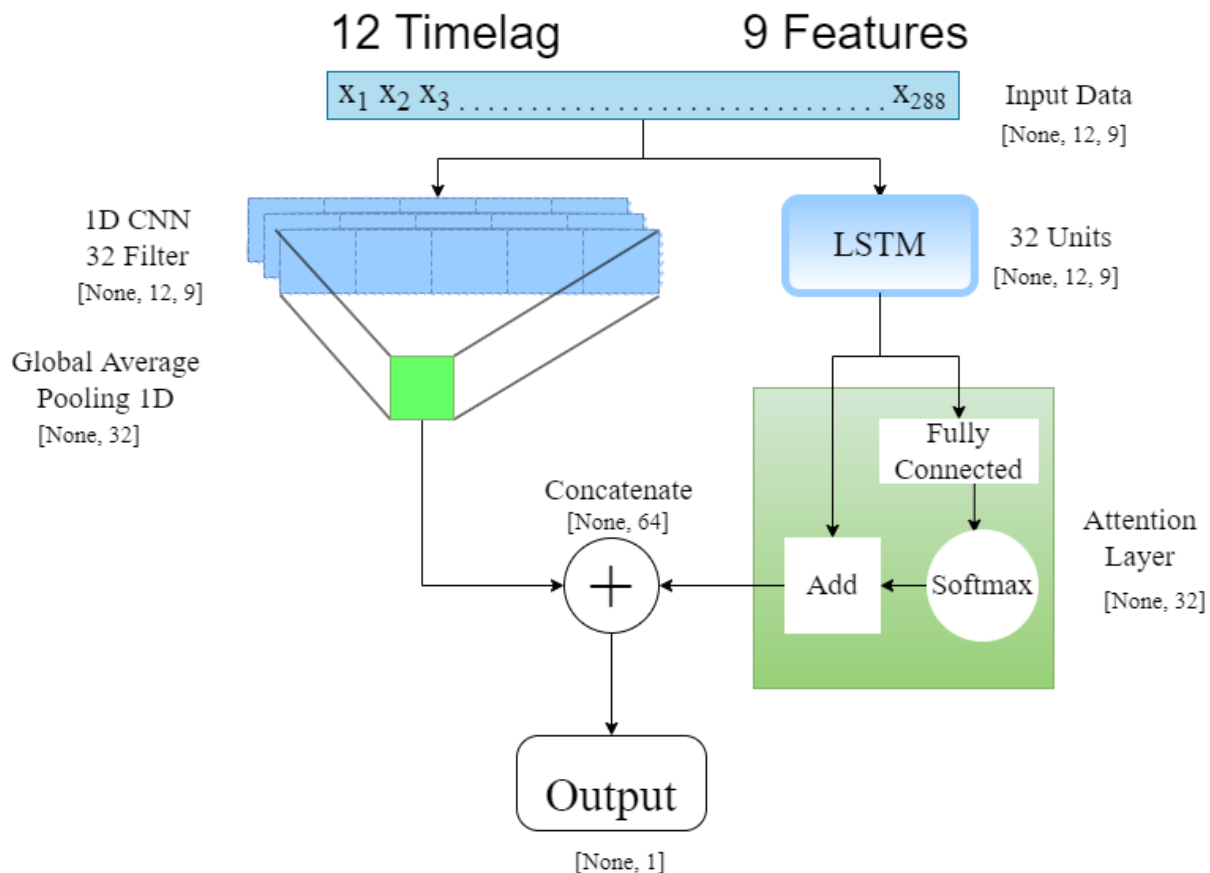


FIGURE 5. Spatiotemporal using Convolutional LSTM Attention Layer based

Modifying the original LSTM with an attention mechanism is necessary to fully utilize the Spatiotemporal input information. The authors take rainfall as input features, and the output of our model is the next n -step rainfall prediction. Spatial and temporal attention weights affect the input and output of LSTM cells [37]. With the help of the Spatiotemporal attention module, the authors were able to dynamically adjust attention weights and improve the performance of LSTM cells

[38]. This model uses Adam's algorithm optimizer [39] to train the model, as shown in Figure 5.

Before training the model, the step that must be conducted is to determine the network architecture, such as deciding how many layers are used, the number of neurons in each layer used, the activation function used, and other parameter values, it can be seen in Table 1. For the input layers based on the features that will be used, 9 spatial features will be used as input neurons; the number 9 comes from 3 x 3 spatial. Then temporal data have 12 timesteps and one time step as the target.

TABLE 1. Optimized Values for Hyperparameters

Characteristic	Specification
Activation Function	<ul style="list-style-type: none"> • Relu for Convolutional 1D • Tanh for LSTM • Softmax for Attention Layer • LeakyRelu for Output Layer
Weight and Bias Initialization	Random (range 0-1)
Learning Rate	<i>Defaults to 0.001</i>
Input Layer	None, 12, 9
Conv1D	None, 10, 9
GlobalAveragePooling1D	None, 32
Attention Layer	None, 32
Concatenate	None, 64
Output Layer	None, 1

4.5. Experimental Design

The fully built model uses five models to compare the proposed model to others. Besides, the whole architecture has explained in Table 1. Postprocessing aims to make better rainfall predictions than "raw" (unprocessed) hydrological simulations. For this aim, it is significant to evaluate the model's performance and compare it with each other to conclude which model is the best. Several metrics are used to evaluate predictions for different wait times. Since accurate and reliable predictions are so crucial during rainfall events, the primary accuracy measure for a

deterministic forecast is the root-mean-square error (RMSE) in equation (8):

$$RMSE = \sqrt{\frac{\sum (y_i - q_i)^2}{n}} \quad (8)$$

Where y_i denotes the i - th time- k prediction of daily rainfall, q_i denotes the observed daily, and n represents the total number of time- k monthly rainfall predictions. Compared with mean absolute error (MAE) metrics, RMSE penalizes significant errors [40], desirable for high rainfall forecasts. Unlike RMSE, which gives a relatively high weight to significant errors, Mean Absolute Error (MAE), a linear statistical measure, is more applicable when the overall impact of errors is proportionate to the increase in error, MAE can be formulated as [40] in equation (9).

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_i - q_i| \quad (9)$$

5. RESULTS AND DISCUSSION

Seven models have been built to forecast rainfall area in Kalimantan Timur, leading by 12 months' time step to predict one month. Those models are:

- RNN: Recurrent Neural Network that allows previous outputs to be used as inputs while having hidden states.
- GRU: Gated recurrent unit (GRU) is a gating mechanism implemented in recurrent neural networks.
- LSTM: Long Short-Term Memory Network is a famous variant of RNN having three gates.
- LSTM-AT: LSTM with additional attention layer.
- Convolutional LSTM-AT: Combination of Convolutional and LSTM with attention layer, as shown in Figure 5.

5.1. Clustering Result

Every spatial point has different statistical distribution, and different models should be trained for different clusters of spatial points with similar characteristics. Because of that, we use K-means clustering to cluster the spatial points. We use the Elbow method to find the optimal

MONTHLY RAINFALL PREDICTION

cluster. Figure 6 shows the best number of clusters that we choose is 4 as a representation of the maximum number of clusters with a significant distance reduction indicator is Cluster 0, Cluster 1, Cluster 2, and Cluster 3. This paper evaluates all clusters as input candidates to build the proposed model that every cluster has own characteristics to generate which spatial become specific cluster. This work it might be the first way to find another clustering.

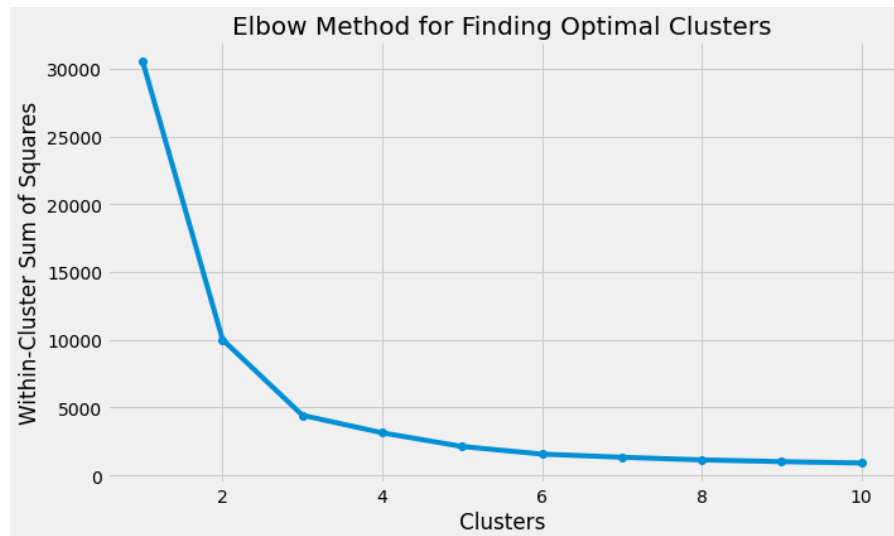


FIGURE 6. Finding Optimal Clusters by using a Within-Cluster Sum of Squares in Elbow Method

TABLE 2. WCSS Value in Every Cluster

Number of Clusters (K)	WCSS
1	30509
2	10014
3	4433
4	3139
5	2132
6	1570
7	1338
8	1144
9	1014
10	914

TABLE 3. Number of Locations per Cluster for K-4

Cluster	Location Point per Cluster
0	779
1	874
2	690
3	1047

In the Elbow method, author is varying the number of clusters (K) from 1 – 10. For each value of K, author is calculating WCSS (Within-Cluster Sum of Square). WCSS is the sum of squared distance between each point and the centroid in a cluster. When author plot the WCSS with the K value, the plot looks like an Elbow. As the number of clusters increases, the WCSS value will start to decrease. WCSS value is largest when $K = 1$. When author analyze the graph author can see that the graph will rapidly change at a point and thus creating an elbow shape. From this point, the graph starts to move almost parallel to the X-axis. The K value corresponding to this point is the optimal K value or an optimal number of clusters. Figure 6 shows the best number of clusters that author choose is 4 as a representation of the maximum number of clusters with a significant distance reduction indicator is cluster 0, cluster 1, cluster 2, and cluster 3. As shown in Table. 2, the number of WCSS can be seen in there same as in Figure. 6. Moreover, Table 3 showed result the clustering location in 4 cluster.

5.2. Convolutional LSTM-AT Result

The first step of this experiment is building a proposed method with LSTM with a modification layer such as an attention mechanism. The challenge of building the model is looking for the best hyperparameter to adjust the number. As shown in Table 1, we use a constant hyperparameter and build all models with the same hyperparameter but different architecture. Furthermore, in the proposed method, we use Relu as an activation function in convolution metrics, tanh in LSTM, softmax in attention layer, and LeakyRelu in output layer. For the learning rate, we use a default 0.001, so the model we run manually one by one.

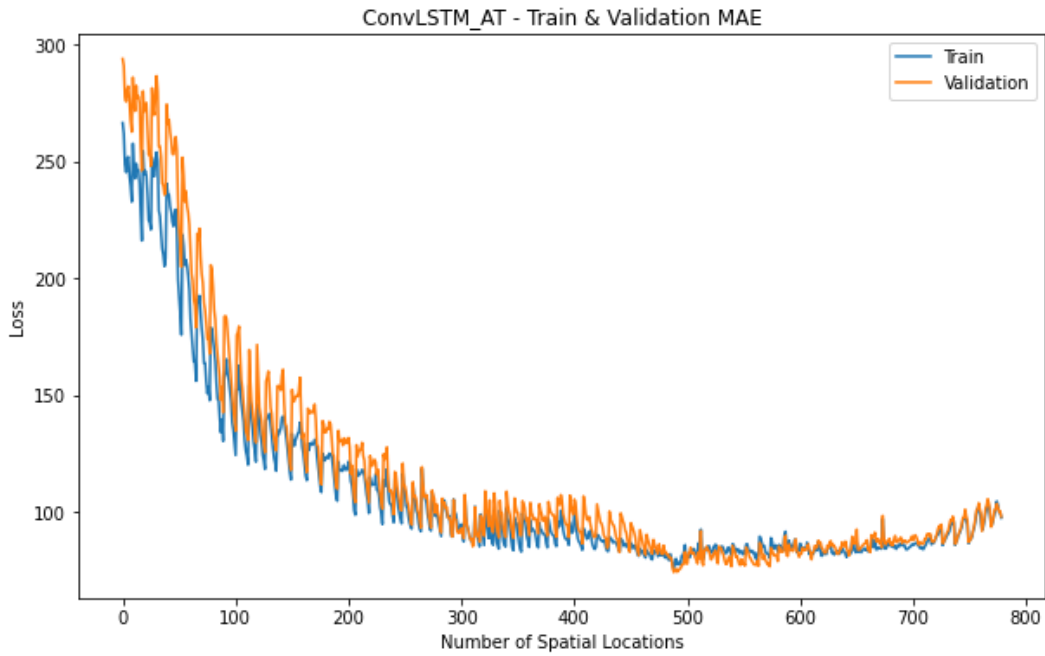


FIGURE 7. Training and Evaluation of ConvLSTM-AT model using MAE

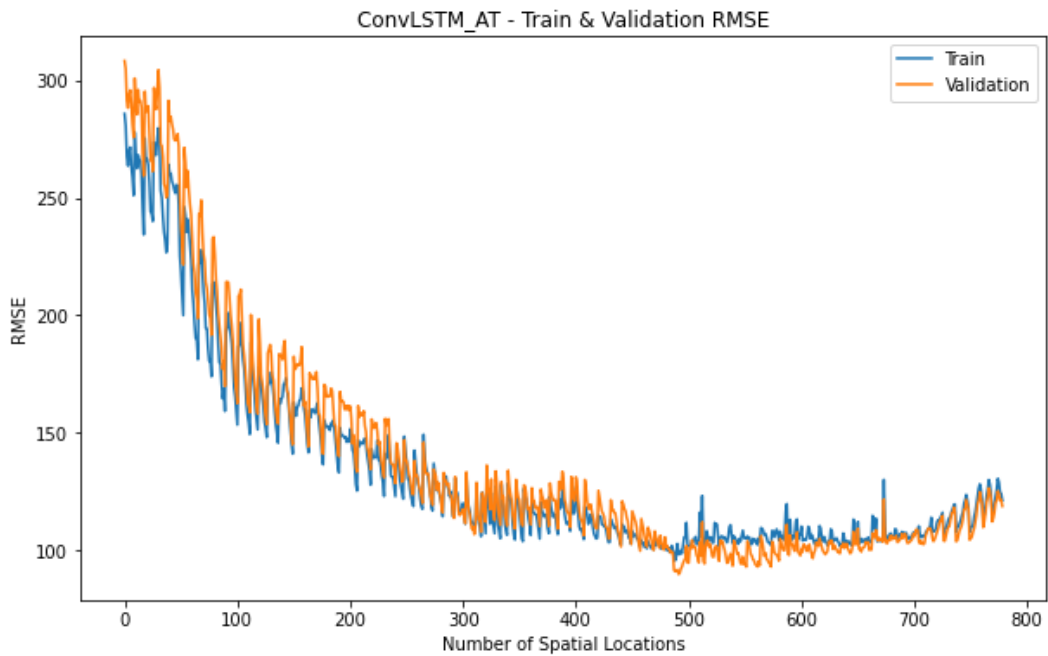


FIGURE 8. Training and Evaluation of ConvLSTM-AT model using RMSE

The proposed model Convolutional-LSTM can be seen in Figure 7 & 8 the performance showed that the performance still best than others method in average value of spatial point. The attention-based models are more accurate and robust than the original LSTM model and reducing

the number of errors significantly. This proves that the proposed method is still the stable to get minimum value of spatial point. Leading to smaller output should be this model perform much better-using data CHIRPS. All the model performances were entirely satisfactory when we see the average of MAE since the average is the testing of all spatial data that have different characteristics. The RMSE and MAE of the predictions from models in experiments is shown in Table 4. On the CHIRPS dataset, the proposed Convolutional LSTM-AT model has lowest error even using maximum value of all spatial target. However, we can infer that the dataset is already split by clustering which makes it harder to see the pattern of the dataset. Thus, the Convolutional-based models can optimally utilize the information contained in the different characteristic and are able to catch the effects of the predictors on the precipitation anomaly under the different time scales. Temporarily, the Attention Mechanism can capture the importance value of different feature states in the data based on past time. Hence, training time can be represented in Figure 9 with RNN takes a long time to train the model while the other models such as GRU, LSTM, LSTM-AT, LSTM-AT (Multi-Layer), LSTM-AM (Multi-Channel), and ConvLSTM-AT are almost similar down to seconds

TABLE 4. Training and Validation Performance for Seven Models

Models	Training		Validation	
	MAE	RMSE	MAE	RMSE
RNN	162.71	193.52	167.04	188.28
GRU	152.82	179.96	164.12	184.77
LSTM	152.67	179.81	163.97	184.61
LSTM-AT	152.28	179.43	163.55	184.20
LSTM-AT Multi-Layer [41]	152.78	179.93	164.10	184.74
LSTM-AM (Multi-Chanel) [3]	120.72	126.60	145.51	145.90
ConvLSTM-AT (ours)	96.43	119.52	108.92	132.41

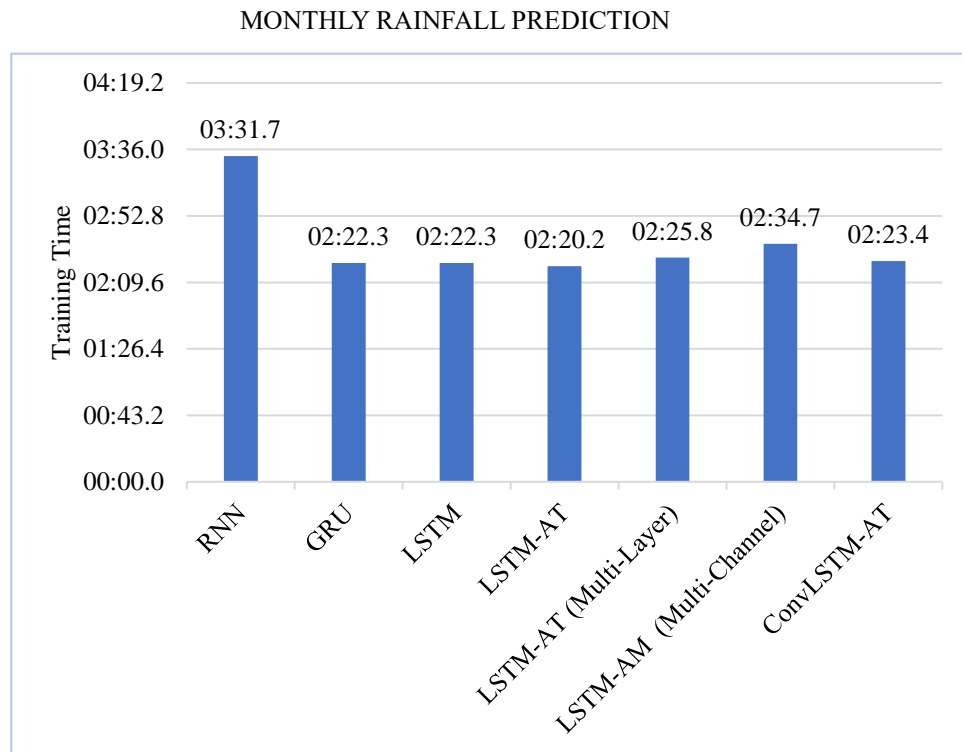


FIGURE 9. Training Time in Different Model

Although the excellent performance is based on the minimum evaluation of models, there are prediction errors in evaluated itself. The errors of MAE shown in Table 5 represent every model learning the data pattern. Because of the effect forecasting architecture and input information, the model increases the number rate gradually. We make analyzing and discuss the performance of models at different metrics measurements. Tables 6 shows that the RMSE of all models of each model that has been built is stable around the time but still makes better before the model stop learning. In this paper, we only pick the rainfall as a candidate for model input, which we believe might be leading to incomplete essential input information. Part of the error was found from the incompleteness of the key to input information. MAE might represent how different data predictions are to actual data in millimeter units, and RMSE is used to analyze performance in standard deep learning linear regression. Therefore, it is worth further exploring how to improve extreme precipitation forecast in the future.

TABLE 5. Evaluation Performance Using MAE

Models	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Weighted Average
RNN	89.02	64.39	117.09	115.52	96.56
GRU	88.14	56.72	118.8	69.16	80.41
LSTM	88.08	57.05	118.5	66.98	79.75
LSTM-AT	87.99	57.41	118.22	64.57	79.02
LSTM-AT Multi-Layer [41]	88.17	58.06	117.98	66.75	79.85
LSTM-AM (Multi-Chanel) [3]	77.70	58.93	84.25	71.93	72.41
ConvLSTM-AT (ours)	74.03	55.38	82.48	64.43	67.97

TABLE 6. Evaluation Performance Using RMSE

Models	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Weighted Average
RNN	110.52	84.5	142.77	146.75	121.56
GRU	109.52	71.66	144.56	88.35	100.35
LSTM	109.46	71.54	144.24	84.5	99.05
LSTM-AT	109.34	72.98	143.95	81.31	98.35
LSTM-AT Multi-Layer [41]	109.56	73.59	143.70	84.30	99.43
LSTM-AM (Multi-Chanel) [3]	96.63	73.61	106.12	89.52	90.43
ConvLSTM-AT (ours)	93.85	68.47	103.92	82.2	85.75

6. CONCLUSION

Rainfall prediction in spatiotemporal data can be challenging, and an attention mechanism can make the Convolutional LSTM-AT model more powerful to represent spatiotemporal data but still open for research to improve the model. The previous work only uses a combination of LSTM and attention layer, so we proposed the model to focus on a statistical model with several feature extractions to consolidate input candidates with a lot of hydrological variables. Since the data is different, we cannot compare from the same perspective, but it still makes sense that our model has a different way of predicting monthly rainfall to determine when the season change,

particularly in Kalimantan Timur, which has tropical climates.

For future work, we will be looking another way to reduce error in the result of the model. The direction may include how to preprocess data and train 3-Dimensional data without losing spatial information. Besides, we will investigate much architecture and develop spatiotemporal approaches. We will consider further improving the performance of the model by utilizing the graph information of area that we predict in the data. Moreover, it will be grated to add flood data augmentation and physical interpretation of model to make prediction more closely with the ground truth.

ACKNOWLEDGEMENT

This work was supported by research grant from the Directorate General of Strengthening for Research and Development, Ministry of Research, Technology, and Higher Education, Republic of Indonesia, with grant numbers 155/E5/PG.02.00.PT/2022, 410/LL3/AK.04/2022, 126/VR.RTT/VI/2022.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

- [1] C.D. Pérez-Blanco, C.M. Gómez, Drought management plans and water availability in agriculture: A risk assessment model for a Southern European basin, *Weather Climate Extremes*. 4 (2014), 11–18.
<https://doi.org/10.1016/j.wace.2014.02.003>.
- [2] Mislan, Haviluddin, S. Hardwinarto, et al. Rainfall monthly prediction based on artificial neural network: A case study in Tenggarong Station, East Kalimantan - Indonesia, *Procedia Computer Sci*. 59 (2015), 142–151.
<https://doi.org/10.1016/j.procs.2015.07.528>.
- [3] L. Tao, X. He, J. Li, et al. A multiscale long short-term memory model with attention mechanism for improving monthly precipitation prediction, *J. Hydrol*. 602 (2021), 126815. <https://doi.org/10.1016/j.jhydrol.2021.126815>.
- [4] X. He, H. Guan, J. Qin, A hybrid wavelet neural network model with mutual information and particle swarm

- optimization for forecasting monthly rainfall, *J. Hydrol.* 527 (2015) 88–100.
<https://doi.org/10.1016/j.jhydrol.2015.04.047>.
- [5] Z. Peng, Q.J. Wang, J.C. Bennett, et al. Seasonal precipitation forecasts over China using monthly large-scale oceanic-atmospheric indices, *J. Hydrol.* 519 (2014), 792–802. <https://doi.org/10.1016/j.jhydrol.2014.08.012>.
- [6] A.M. Bagirov, A. Mahmood, A comparative assessment of models to predict monthly rainfall in Australia, *Water Resour. Manage.* 32 (2018), 1777–1794. <https://doi.org/10.1007/s11269-018-1903-y>.
- [7] S.H. Pour, A.K.A. Wahab, S. Shahid, Physical-empirical models for prediction of seasonal rainfall extremes of Peninsular Malaysia, *Atmos. Res.* 233 (2020), 104720. <https://doi.org/10.1016/j.atmosres.2019.104720>.
- [8] X. Shi, Z. Chen, H. Wang, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting, in: *Proceedings of the 28th International Conference on Neural Information Processing Systems*, 1 (2015), 802–810.
- [9] N.S. Novella, W.M. Thiaw, African rainfall climatology version 2 for famine early warning systems, *J. Appl. Meteorol. Climatol.* 52 (2013), 588–606. <https://doi.org/10.1175/jamc-d-11-0238.1>.
- [10] C. Funk, P. Peterson, M. Landsfeld, et al. The climate hazards infrared precipitation with stations-a new environmental record for monitoring extremes, *Sci. Data.* 2 (2015), 150066.
<https://doi.org/10.1038/sdata.2015.66>.
- [11] G. Gawarkiewicz, K. Chen, J. Forsyth, et al. Characteristics of an advective marine heatwave in the middle atlantic bight in early 2017, *Front. Mar. Sci.* 6 (2019), 712. <https://doi.org/10.3389/fmars.2019.00712>.
- [12] S. Wang, J. Cao, P.S. Yu, Deep learning for spatio-temporal data mining: a survey, *IEEE Trans. Knowl. Data Eng.* 34 (2022), 3681–3700. <https://doi.org/10.1109/tkde.2020.3025580>.
- [13] D. Bahdanau, K.H. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, in: *3rd Int. Conf. Learn. Represent. ICLR 2015*, pp. 1–15, (2015).
- [14] A. Vaswani, N. Shazeer, N. Parmar, Attention is all you need, in: *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 6000–6010, (2017).
- [15] S. Urolagin, N. Sharma, T.K. Datta, A combined architecture of multivariate LSTM with Mahalanobis and Z-Score transformations for oil price forecasting, *Energy.* 231 (2021), 120963.
<https://doi.org/10.1016/j.energy.2021.120963>.

- [16] M. Liu, Y. Huang, Z. Li, et al. The applicability of LSTM-KNN model for real-time flood forecasting in different climate zones in China, *Water*. 12 (2020), 440. <https://doi.org/10.3390/w12020440>.
- [17] Z. Chao, F. Pu, Y. Yin, et al. Research on real-time local rainfall prediction based on MEMS sensors, *J. Sensors*. 2018 (2018), 6184713. <https://doi.org/10.1155/2018/6184713>.
- [18] S. Hastenrath, Climate and climate change | Climate prediction: Empirical and numerical, in: *Encyclopedia of Atmospheric Sciences*, Elsevier, 2015: pp. 26–32. <https://doi.org/10.1016/B978-0-12-382225-3.00322-4>.
- [19] J.F. Vuillaume, S. Herath, Improving global rainfall forecasting with a weather type approach in Japan, *Hydrol. Sci. J.* 62 (2016), 167–181. <https://doi.org/10.1080/02626667.2016.1183165>.
- [20] Z. Zhu, T. Li, Empirical prediction of the onset dates of South China Sea summer monsoon, *Clim. Dyn.* 48 (2016), 1633–1645. <https://doi.org/10.1007/s00382-016-3164-x>.
- [21] J. Li, B. Wang, Predictability of summer extreme precipitation days over eastern China, *Clim. Dyn.* 51 (2017), 4543–4554. <https://doi.org/10.1007/s00382-017-3848-x>.
- [22] E. Lu, H. Chen, J. Tu, et al. The nonlinear relationship between summer precipitation in China and the sea surface temperature in preceding seasons: A statistical demonstration, *J. Geophys. Res. Atmos.* 120 (2015), 12027–12036. <https://doi.org/10.1002/2015jd024030>.
- [23] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.* 9 (1997), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [24] Y. Jin, D. Yan, X. Kang, et al. Forecasting building occupancy: A temporal-sequential analysis and machine learning integrated approach, *Energy Build.* 252 (2021), 111362. <https://doi.org/10.1016/j.enbuild.2021.111362>.
- [25] M. ElSaadani, E. Habib, A.M. Abdelhameed, et al. Assessment of a spatiotemporal deep learning approach for soil moisture prediction and filling the gaps in between soil moisture observations, *Front. Artif. Intell.* 4 (2021), 636234. <https://doi.org/10.3389/frai.2021.636234>.
- [26] K. Zhou, T. Wu, C. Wang, et al. Skeleton based abnormal behavior recognition using spatio-temporal convolution and attention-based LSTM, *Procedia Computer Sci.* 174 (2020) 424–432. <https://doi.org/10.1016/j.procs.2020.06.110>.
- [27] S. Song, C. Lan, J. Xing, et al. An end-to-end spatio-temporal attention model for human action recognition from skeleton data, in: *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)*, pp.

- 4263–4270, (2017). <https://doi.org/10.1609/aaai.v31i1.11212>.
- [28] L. Chen, H. Zhang, J. Xiao, et al. SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, 2017: pp. 6298–6306. <https://doi.org/10.1109/CVPR.2017.667>.
- [29] Y. Ding, Y. Zhu, J. Feng, et al. Interpretable spatio-temporal attention LSTM model for flood forecasting, *Neurocomputing*. 403 (2020), 348–359. <https://doi.org/10.1016/j.neucom.2020.04.110>.
- [30] D. Haynes, P. Mitchell, E. Shook, Developing the raster big data benchmark: A comparison of raster analysis on big data platforms, *ISPRS Int. J. Geo-Inform.* 9 (2020), 690. <https://doi.org/10.3390/ijgi9110690>.
- [31] S. Ramadass, A. Abraham, Cluster Analysis: An Experimental Study, *Int. J. Computer Appl.* 64 (2013), 32–36. <https://doi.org/10.5120/10704-5622>.
- [32] M.Z. Rodriguez, C.H. Comin, D. Casanova, et al. Clustering algorithms: A comparative approach, *PLoS ONE*. 14 (2019), e0210236. <https://doi.org/10.1371/journal.pone.0210236>.
- [33] G.C. Pamuji, H. Rongtao, A comparison study of DBScan and K-means clustering in Jakarta rainfall based on the Tropical Rainfall Measuring Mission (TRMM) 1998-2007, *IOP Conf. Ser.: Mater. Sci. Eng.* 879 (2020) 012057. <https://doi.org/10.1088/1757-899x/879/1/012057>.
- [34] R. Solgi, H.A. Loáiciga, M. Kram, Long short-term memory neural network (LSTM-NN) for aquifer level time series forecasting using in-situ piezometric observations, *J. Hydrol.* 601 (2021), 126800. <https://doi.org/10.1016/j.jhydrol.2021.126800>.
- [35] X.H. Le, H.V. Ho, G. Lee, et al. Application of long short-term memory (LSTM) neural network for flood forecasting, *Water*. 11 (2019), 1387. <https://doi.org/10.3390/w11071387>.
- [36] X. Zhang, X. Liang, A. Zhiyuli, et al. AT-LSTM: An attention-based LSTM model for financial time series prediction, *IOP Conf. Ser.: Mater. Sci. Eng.* 569 (2019) 052037. <https://doi.org/10.1088/1757-899x/569/5/052037>.
- [37] S.F. Tekin, O. Karaahmetoglu, F. Ilhan, I. Balaban, S.S. Kozat, Spatio-temporal weather forecasting and attention mechanism on convolutional LSTMs, (2021). <http://arxiv.org/abs/2102.00696>.
- [38] Y. Bachrach, A. Zukov-Gregoric, S. Coope, et al. An attention mechanism for neural answer selection using a combined global and local view, in: 2017 IEEE 29th International Conference on Tools with Artificial

MONTHLY RAINFALL PREDICTION

- Intelligence (ICTAI), IEEE, Boston, MA, 2017: pp. 425–432. <https://doi.org/10.1109/ICTAI.2017.00072>.
- [39] D.P. Kingma, J.L. Ba, Adam: A method for stochastic optimization, in: 3rd Int. Conf. Learn. Represent. ICLR 2015, pp. 1–15, 2015.
- [40] T. Chai, R.R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature, *Geosci. Model Dev.* 7 (2014), 1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>.
- [41] D. Seng, Q. Zhang, X. Zhang, et al. Spatiotemporal prediction of air quality based on LSTM neural network, *Alexandria Eng. J.* 60 (2021), 2021–2032. <https://doi.org/10.1016/j.aej.2020.12.009>.