Enhancing spatiotemporal weather forecasting accuracy with 3D convolutional kernel through the sequence to sequence model

Renaldy Fredyan[1], Karli Eka Setiawan[1], Kelvin Asclepius Minor[2]
[1]Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia
[2]Mathematics Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Article Info

Article history:

Received Jun 29, 2024
Revised Nov 25, 2024
Accepted Dec 2, 2024

Keywords:

Convolutional layer
Data CHIRPS
Rainfall analysis
Seq2Seq model
Spatiotemporal forecasting

ABSTRACT

Accurate weather forecasting is important when dealing with various sectors, such as retail, agriculture, and aviation, especially during extreme weather events like heat waves, droughts, and storms to prevent disaster impact. Traditional methods rely on complex, physics-based models to predict the Earth's stochastic systems. However, some technological advancements and the availability of extensive satellite data from beyond Earth have enhanced meteorological predictions and sent them to Earth's antennae. Deep learning models using this historical data show promise in improving forecast accuracy to enhance how models learn the data pattern. This study introduces a novel architecture, convolutional sequence to sequence (ConvSeq2Seq) network, which employs 3D convolutional neural networks (CNN) to address the challenges of spatiotemporal forecasting. Unlike recurrent neural network (RNN)--based models, which are time-consuming due to sequential processing, 3D CNNs capture spatial context more efficiently. ConvSeq2Seq overcomes the limitations of traditional CNN models by ensuring causal constraints and generating flexible length output sequences. Our experimental results demonstrate that ConvSeq2Seq outperforms traditional and modern RNN-based architectures in both prediction accuracy and time efficiency, leveraging historical meteorological data to provide a robust solution for weather forecasting applications. The proposed architecture outperforms the previous method, giving new insight when dealing with spatiotemporal with high density.

Corresponding Author:

Renaldy Fredyan
Computer Science Department, School of Computer Science, Bina Nusantara University
Kebon Jeruk Raya Street 27, Jakarta, Indonesia
Email: renaldy.fredyan@binus.ac.id

1. INTRODUCTION

In situations of severe natural occurrences like heat waves (high temperatures), droughts, and storms, resource planning depends critically on weather forecasting. It also affects decision-making in retail markets, agriculture, aviation, and other industries since bad weather lowers corporate revenues [1]. Meteorological variable forecasts are becoming increasingly accurate due to technical advancements over time. Nevertheless, conventional forecasting needs sophisticated, physics-based models to forecast the weather because of the stochastic demeanor of the Earth systems, which are controlled by physical laws [2]. Historical data is available. Hence, researchers can create deep-learning models that can forecast the weather more accurately [3]. Shi *et al.* [4] suggested that the convolutional long short-term memory (ConvLSTM) architecture uses a radar echo dataset for precipitation forecasting in order to overcome the sequence prediction issue. To concurrently learn incoming data's temporal and spatial context and project the future sequence, they merged the convolution operator—used by the convolutional neural network (CNN)—with a recurrent neural network (RNN). Even while ConvLSTM architecture has been seen as a possible method to develop prediction models for geoscience data [5], fresh prospects have surfaced thanks to current developments in deep learning. The long short-term memory (LSTM) unit was enhanced in several studies to memorize spatial-temporal information.

RNN-based architectures are well-suited for multi-step forecasting tasks involving spatiotemporal data due to their ability to predict long sequences while maintaining temporal order (causal constraint) [6]. However, these models rely on information from previous time steps to generate outputs, leading to lengthy training times. To address this limitation, we propose a novel architecture that exclusively utilizes 3D CNN for spatiotemporal forecasting. CNNs are highly effective in capturing spatial context and have demonstrated state-of-the-art performance in image classification with 2D kernels [7]. Recent advancements have extended the application of CNNs, such as using 1D kernels for tasks like machine translation [8], which enables the extraction of temporal patterns in sequences. Applications like video analysis, action recognition [9], and climate event detection [10] highlight the potential of 3D CNN models. However, CNN-based approaches face two key challenges for multi-step forecasting: they are unable to produce output sequences longer than the input and disrupt temporal order by incorporating future information during temporal reasoning [11]. We introduce convolutional sequence-to-sequence (ConvSeq2Seq) network, a spatiotemporal prediction model explicitly designed for multi-step forecasting tasks to address these limitations. As far as we know, ConvSeq2Seq is the first 3D CNN-based architecture developed as an end-to-end trainable model that adheres to the causal constraint while allowing the prediction of output sequences of flexible lengths—unrestricted by the length of the input sequence.

Through experimental evaluations, we assessed the predictive accuracy and time efficiency of ConvSeq2Seq in comparison to RNN-based architectures. Using meteorological datasets such as climate hazard group infrared precipitation satellite (CHIRPS) that combine satellite data and in situ station [12] measurements, the proposed architecture matches or outperforms existing techniques. This study contributes in two significant ways. First, it introduces variations of the ConvSeq2Seq architecture that satisfy the causal constraint. One approach involves adapting causal convolution within 3D convolutional layers, while another applies a novel technique that reverses sequences deliberately. Second, to enable longer output sequences, we developed a temporal generator block featuring an innovative use of transposed convolutional layers.

2. RELATED WORKS

Historical data regarding temperature, precipitation, and other meteorological variables have been used to forecast the weather using several statistical and machine-learning approaches [13]. Times series analysis is traditionally handled statistically using auto-regressive integrated moving averages (ARIMA) [14]. Other research has also used artificial neural networks (ANN) for time series prediction in meteorological data, including temperature readings [15]. Using LSTM networks in particular, many writers have been developing novel deep learning-based methods recently to enhance time series forecasting performance [16]. Applying LSTM designs effectively includes traffic flow analysis [17], landslide displacement prediction [18], petroleum production [19], and sea surface temperature forecasting [20]. However, spatial relationships in the data are not captured by these methods (which are directed at time series).

Spatiotemporal deep learning algorithms effectively address both geographic and temporal dimensions. Shi *et al*. [4] treat weather forecasting as a sequence-to-sequence problem, utilizing sequences of 2D radar maps as both input and output. They introduce the ConvLSTM architecture to create an end-to-end model for precipitation nowcasting, integrating convolutional operations into the LSTM network to capture spatial patterns. Similarly, Kim *et al*. [21] employ ConvLSTM for predicting severe climatic events, framing their task as a sequence-based problem using storm density map sequences as input. Souto *et al*. [22] propose a spatiotemporal-aware ensemble approach leveraging ConvLSTM, while Setiawan *et al*. [23] work by incorporating a novel LSTM unit that uniformly handles temporal and spatial variations in its memory pool. Wang *et al*. [24] also enhance memory functionality by introducing non-stationarity modeling within the LSTM unit. Although these approaches combine LSTM and CNN for climate and weather-related tasks, our model adopts a purely CNN-based methodology, avoiding the hybrid strategy of merging LSTM with CNN.

A few studies have been done on applying spatiotemporal convolutions for action recognition and video analysis. Tran *et al.* [25] demonstrate that factorizing the 3D convolutional kernel into distinct and consecutive spatial and temporal convolutions increases accuracy by comparing multiple spatiotemporal designs employing just 3D CNN. Limitation of factorized 3D CNN as well as 3D CNN Tran *et al.* [25] violates the temporal order by lacking a causal requirement. The 3D convolution, as Tran *et al.* [25], is factorized by Singh and Cuzzolin [26] and Cheng *et al*. [27]. The causal constraint in temporal learning for action recognition tasks is addressed by Singh and Cuzzolin [26] using a recurrent convolution unit technique; the causal constraint is satisfied by Cheng *et al.* [27] by using causal convolution in discrete and parallel spatial and temporal convolutions. However, with a different implementation, we likewise used a factorized 3D CNN. We use a full CNN technique; similarly, we provide a novel way not to break the temporal order and do not employ parallel convolutions when adopting a causal convolution. After Xu *et al*. [28] successfully captured spatial correlation in pictures, they proposed a technique to estimate vehicle pollution emissions by independently collecting temporal and spatial correlation using 2D CNN. Mudigonda *et al*. [29] identify severe climatic events by using a 3D CNN in an encoder-decoder architecture.

3. METHOD
3.1. Data

Gridded rainfall time series with daily frequency and a 0.05° geographic resolution are produced by combining satellite images and in situ station data in the CHIRPS dataset. In this work, we performed interpolation to shrink the grid size to 50×50 using a records sample from January 1981 to December 2020. Figure 1 displays the coverage area, which is 127,346.92 km^2 on land and 25,656 km^2 on water, employed in our studies, from 2°33′ North Latitude -2°25' South Latitude, 113°44'–119°00' East Longitude. In keeping with Shi *et al.*'s methodology [4], we set the input sequence length to five, *i.e.*, the next set of grids is predicted using the previous five grids. For the CHIRPS dataset (*https://www.chc.ucsb.edu/data/chirps*), thus, the input data shapes for the deep learning architectures are 5×50×50×1. Here, 1 denotes the single channel (like a grayscale picture), 5 is the forecasting sequence length, and 32 and 50 is the number of latitudes and longitudes utilized to build the spatial grid for every dataset.

From the rainfall dataset, we produced 13,960 grid sequences. After that, non-overlapping training, validation, and test sets were created from both datasets in proportions of 60%, 20%, and 20%, respectively. We have used rainfall datasets in our experimental assessment because of their importance as main meteorological variables. Studying their spatiotemporal representation improves our knowledge of long-term climate variability and is essential for short-term forecasting. However, the proposed architecture is adaptable and may be used for other meteorological variables or domains, provided that the training data can be organized as described in the subsequent section.

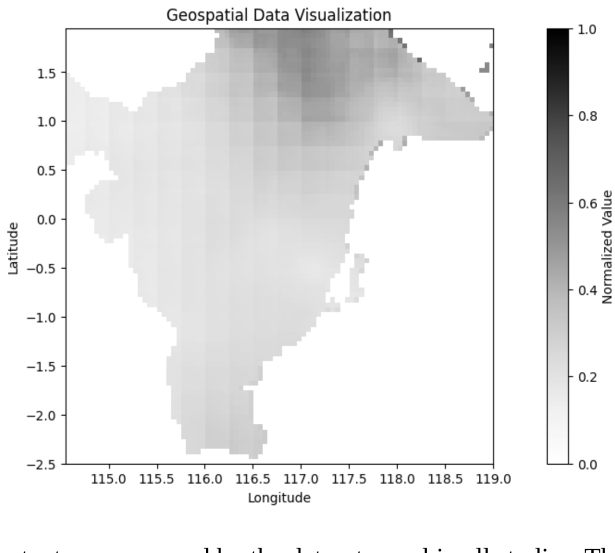

Figure 1. The geographical extent encompassed by the datasets used in all studies. The grid below represents the chosen sequence for December 2020, showing the recorded rainfall levels

3.2. Proposed models

The 3D convolutional layer through Seq2Seq model is a comprehensive deep neural network designed to learn and predict patterns that occur in both space and time. This network is especially beneficial in industries like weather forecasting, where these patterns are frequently observed. Our methodology allows for the prediction of multi-step sequences without incorporating the anticipated outcome back into the input sequence. Our proposed deep learning framework is comprehensively illustrated in Figure 2.

The majority of weather forecasting techniques employ a combination of 2D CNN and LSTM to learn spatial and temporal representations. However, our approach exclusively employs 3D convolutional layers to acquire both spatial and temporal contexts. ConvSeq2Seq, in contrast to the conventional convolution employed in certain 3D CNN architectures, guarantees that it does not depend on future knowledge during temporal learning, a critical requirement for task prediction. Our network architecture's ability to accommodate variable output sequence lengths is critical. This suggests that it is capable of predicting a significant number of future time steps, regardless of the input sequence's fixed duration. In the following section, we offer additional details regarding the components that constitute our architecture.

We employ a factorized 3D kernel that is inspired by the $R(2+1)\,D$ network introduced by Tran *et al.* [25] instead of a traditional kernel for 3D convolutional layers, where the kernel size is defined by *d* in the spatial dimensions (H and W) and t in the temporal dimension (T). The authors introduce a factorized kernel, denoted as 1-d-d and t-1-1, in their work. This kernel partitions the convolution procedure of a single layer into two distinct operations: a spatial convolution and a temporal convolution. We employ an alternative methodology in our innovative design, which involves the non-sequential execution of operations within each convolutional layer. The factorized kernels are partitioned into two groups, which leads to discrete learning abilities for each. The temporal block employs the t-1-1 kernel in its layers to understand temporal relationships in a specific manner. Conversely, the spatial block encapsulates spatial dependencies by employing a 1-d-d kernel.

The kernel decomposition employed in ConvSeq2Seq has the advantage of increasing the number of nonlinearities in the network, in contrast to the complete 3D kernel used in conventional convolutions. This is accomplished by incorporating supplementary activation functions between factorized convolutions, which leads to greater complexity in the patterns that can be represented. Our recommended solution is flexible over the (2+1) D block. This is due to the fact that the temporal and spatial units may have varying numbers of layers, which facilitates their optimization.

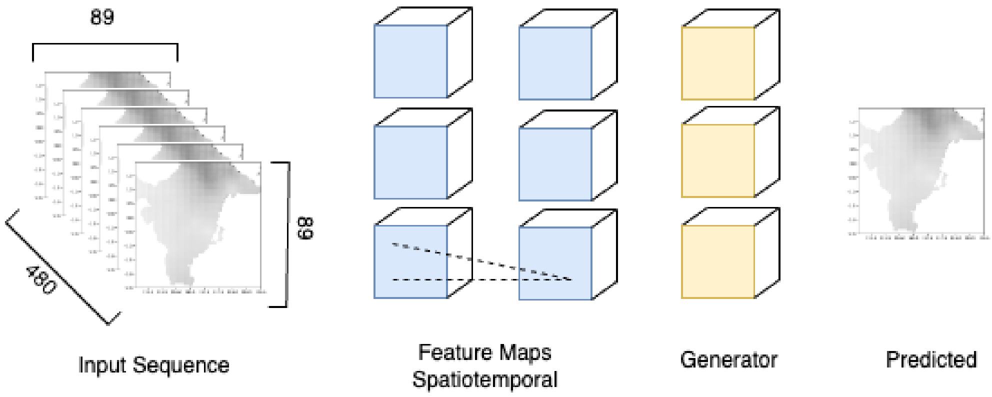

Figure 2. ConvSeq2Seq architecture

3.3. Evaluation matrix

Postprocessing is sought to obtain better rainfall forecasts than "raw" (unprocessed) hydrological models. For this purpose, it is important to assess the performance of the models and contrast them to choose the best one. A number of measures are used to assess forecasts for various wait durations. The equation's root-mean-square error (RMSE) is the main accuracy metric for a deterministic forecast because precise and trustworthy forecasts are so important during rainfall events.

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

where q_i is the observed daily rainfall, y_i is the i^{th} time-k forecast of daily rainfall, and n is the total number of time-k monthly rainfall predictions. RMSE penalizes more substantial mistakes for high rainfall projections than mean absolute error (MAE) measures. Where the total effect of mistakes is proportional to the increase in error, MAE, a linear statistical measure, is more useful than RMSE, which assigns a comparatively large weight to big errors.

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

The mean squared error (MSE) quantifies the average of the squared discrepancies between the observed and estimated values. The metric measures the proximity of the predictions to the actual results, where smaller values imply superior performance of the model. Where: n is the number of observations, and q_i represents the actual value. The variable y_i represents the value that is being forecasted.

$$MSE = \frac{1}{n}\sum_{j=1}^{n}(y_i - q_i)^2$$

R-squared is a statistical metric that quantifies the percentage of the variation in the dependent variable that the independent variables can account for in a regression model. The range of values is from 0 to 1, where greater numbers indicate a stronger match. Where: q_i represents the current value The variable y_i represents the expected value. The term $\overline{q_i}$ represents the average of the actual numbers. n represents the total count of observations. The numerator in this formula indicates the aggregate of squared errors in the forecast, while the denominator is the overall variance in the data. A value of 1 for R^2 signifies a complete match, while a value of 0 shows that the model fails to account for any of the variations in the response data around its average.

$$R^2 = 1 - \frac{\sum_{j=1}^{n}|y_i - q_i|}{\sum_{j=1}^{n}|y_i - \overline{q_i}|}$$

4. RESULTS AND DISCUSSION

It is conceivable for a Seq2Seq model to have a high R2 value but relatively low MAE, MSE, and RMSE values for a number of different reasons. There is a possibility that Seq2Seq is a useful method for identifying severe variances or outliers in the data that have a substantial impact on the R2 score. A high R2 number suggests that the model is effective in explaining big fluctuations in the data as a whole, but if these outliers are not deleted or controlled efficiently, they have the potential to interfere with the MAE, MSE, and RMSE values, which are more sensitive to absolute error than R2. Sequence-to-Sequence (Seq2Seq) models often effectively capture temporal connections and the complexity of time series data. Additionally, this may enable the model to make correct forecasts at the per-data point prediction level, resulting in increased prediction errors (worsening MAE, MSE, and RMSE). This might be a consequence of the model's ability to produce predictions that follow the data trend well overall (higher R2). Both the magnitude of the data and the absolute error significantly impact the MAE, MSE, and RMSE. Small inaccuracies in forecasts may result in big increases in these values, even if the model normally adjusts to the pattern of the data. This is the case when the data is huge in size or has a high degree of variance.

A complete image of rainfall is provided by CHIRPS data, which combines observations from satellites and ground stations. This picture may be beneficial for models that are used to anticipate or evaluate weather-related problems such as drought. On the basis of the abundant and intricate characteristics of the CHIRPS data, the following are a few probable causes for the disparate performance of the models shown in the table. Seq2Seq models can more readily explain temporal (time) and spatial (space) fluctuations in CHIRPS data than CNN or CNN-LSTM models. This is because the Seq2Seq model takes into account both factors simultaneously. This may be the reason why R2 is high (the model explains a significant amount of the variance in the data), but the absolute errors (MAE, MSE, and RMSE) are also high because of the difficulty of the model in capturing specific features of the local environment, as can be seen in Table 1. It is possible for rainfall statistics to have a significant amount of variation, depending on the location and the time period. Models that can capture overall trends may only sometimes be successful when forecasting precise values at particular times and places, which might result in larger value errors. Because the Seq2Seq model is able to adjust to the overall trend, it is influenced by extreme values in the MAE, MSE, and RMSE calculation errors. For instance, if there are outliers in the rainfall data, such as very uncommon heavy rainfall, the model is able to adapt to the general trend. The different models' architectural complexity and particular operations account for the disparities in memory use and training time observed, as seen in Table 2. Convolutional neural networks (CNNs), often need a large amount of memory because of its many convolutional layers, which record spatial hierarchies in input, such as pictures. Still, given their simple layer structure that permits parallel data processing, their training time per epoch is quite efficient. Longer training periods and higher memory consumption are features of CNN-LSTM models, which mix convolutional and LSTM layers. The model becomes more complicated and resource-intensive as LSTM layers capture temporal relationships and CNN levels extract spatial data. Comparing ConvLSTM models to distinct CNN and LSTM layers, ConvLSTM models handle spatiotemporal input directly by combining convolution processes with LSTM units.

Table 1. Performance data for rainfall forecasting predicting the next five observations (5 → 5) using the previous five observations (grids)

Model	MAE	MSE	RMSE	R^2
CNN	0.038	0.004	0.068	0.024
CNN-LSTM	0.036	0.004	0.066	0.047
ConvLSTM	0.037	0.004	0.065	0.058
ConvSeq2Seq	0.073	0.034	0.186	0.965

Table 2. Efficiency analysis of model memory usage and training time

Model	Memory usage (MB)	Training time (s)	Training time/epoch (s)
CNN	2005.07	16.56	0.82
CNN-LSTM	2232.02	25.72	0.80
ConvLSTM	1457.76	10.94	0.54
ConvSeq2Seq	565.07	1037.70	20.75

To demonstrate how closely the model's predictions match the actual values, the 3D visualization in Figure 3 contrasts the actual and projected data. Plots of the data show values, latitude, and longitude. The real data plot on the left displays a complicated surface with clear value fluctuations over several geographic areas. This surface is tried to be duplicated in the projected data plot (right). Though the actual and anticipated data are comparable, there are differences in certain places that point to places where the model's predictions differ from the real values. Figures 4 and 5 show how well two distinct models—ConvSeq2Seq and ConvLSTM—predict rainfall throughout a five-month test set of the CHIRPS data. The ConvSeq2Seq model's projected and actual rainfall maps are compared in Figure 3 for each of the five months. Whereas the anticipated maps indicate the model's forecasts, the ground truth maps reflect the actual rainfall that has been measured.

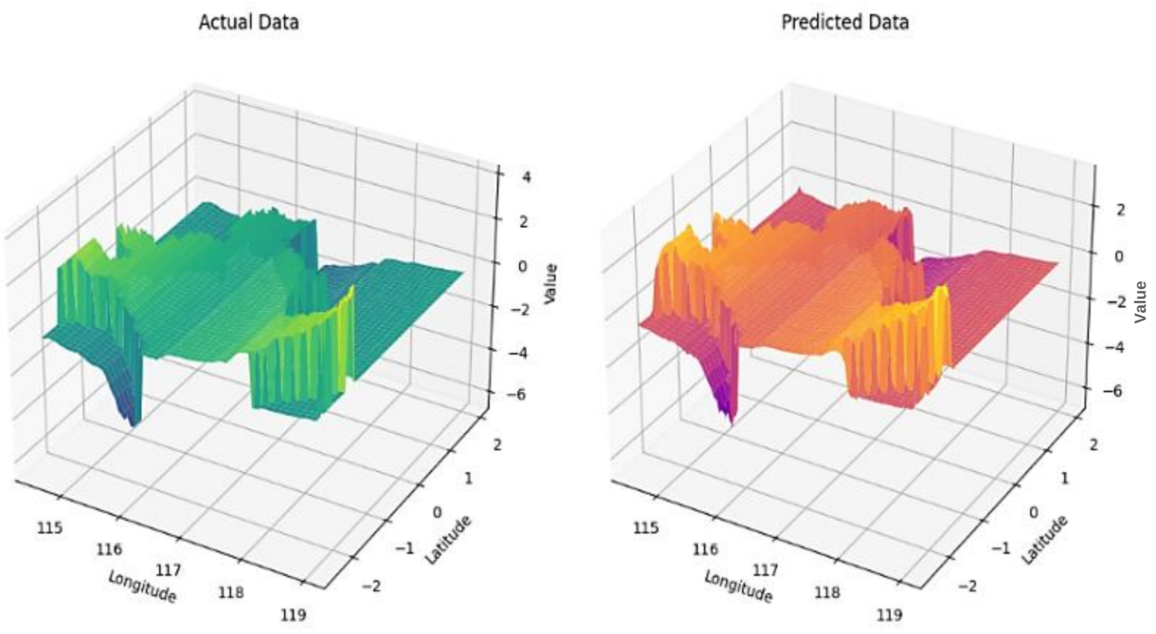

Figure 3. 3D Visualization of actual vs predicted data

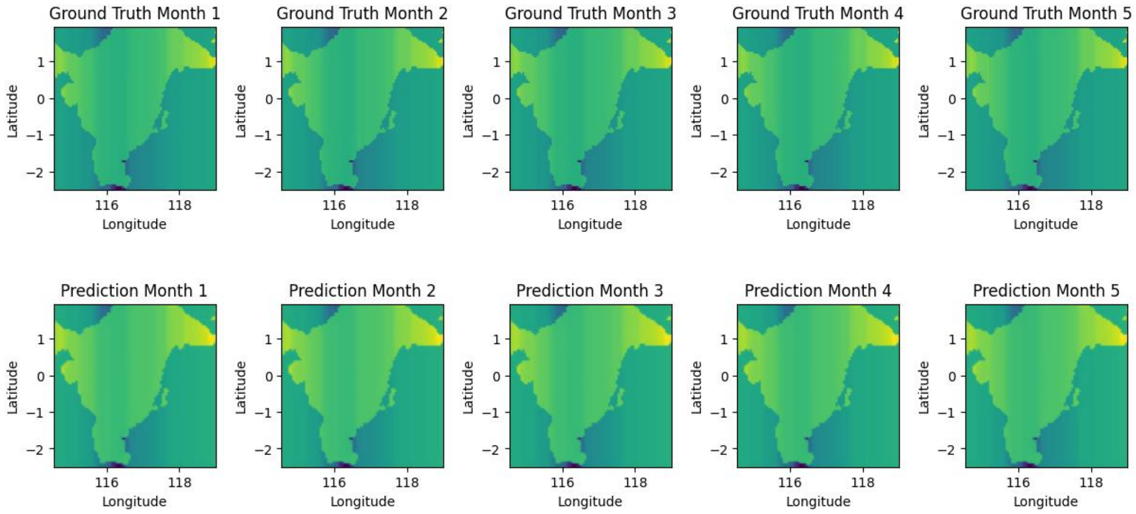

Figure 4. An example of predicting rainfall on a test set of the CHIRPS dataset using ConvSeq2Seq

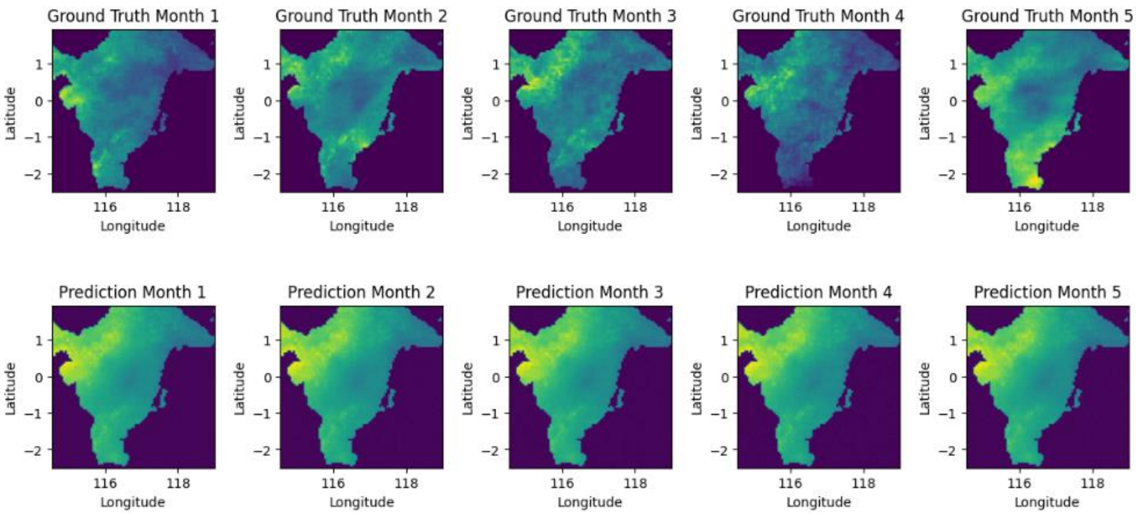

Figure 5. An example of predicting rainfall on a test set of the CHIRPS dataset using ConvLSTM

5. CONCLUSION

The study has shown that the effective utilization of hardware significantly impacts machine learning models' training duration and resource consumption. By using parallel processing and high-performance graphics processing units (GPUs) to optimize hardware use, training time and operating expenses may be cut. More extended training periods and greater memory use are typical of more complicated architectures, as the table comparing several models (CNN, CNN-LSTM, ConvLSTM, and ConvSeq2Seq) demonstrates. The spatiotemporal data processing efficiency of models such as ConvLSTM leads to improved memory use and training time performance. The 3D visualizations comparing actual and predicted data show the difficulties of precisely modeling real-world data. These differences result from overfitting or underfitting, hyperparameter tuning, feature selection, training data constraints, and model complexity. It is clear from comparing the ConvSeq2Seq and ConvLSTM models for rainfall prediction on the CHIRPS dataset that, while both models capture broad patterns of rainfall, the ConvLSTM model does a better job of precisely capturing geographic distribution and intensity.

ACKNOWLEDGEMENTS

This work is supported by Bina Nusantara University as a part of Bina Nusantara University's BINUS Research for Early Career Researchers entitled "*Analisis Spasial-Temporal dengan LSTM untuk*

Meningkatkan Prediksi Curah Hujan Bulanan" with contract number: 069A/VRRTT/III/2024 and contract date: March 18, 2024.

REFERENCES

[1] S. F. Tekin, O. Karaahmetoglu, F. Ilhan, I. Balaban, and S. S. Kozat, "Spatio-temporal weather forecasting and attention mechanism on convolutional LSTMs," *ArXiv*, pp. 1–13, 2021.

[2] J. A. Brotzge *et al.*, "Challenges and opportunities in numerical weather prediction," *Bulletin of the American Meteorological Society*, vol. 104, no. 3, pp. E698--E705, Mar. 2023, doi: 10.1175/BAMS-D-22-0172.1.

[3] R. Fredyan and K. E. Setiawan, "An investigation into improving el niño-southern oscillation prediction based on temporal transformer architecture," *Communications in Mathematical Biology and Neuroscience*, 2024, doi: 10.28919/cmbn/8371.

[4] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W. Woo, "Convolutional LSTM network: a machine learning approach for precipitation nowcasting," *Advances in neural information processing systems*, vol. 28, 2015.

[5] C.-A. Diaconu, S. Saha, S. Gunnemann, and X. Xiang Zhu, "Understanding the role of weather data for earth surface forecasting using a ConvLSTM-based model," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2022, pp. 1361–1370, doi: 10.1109/CVPRW56347.2022.00142.

[6] Y. Ding, Y. Zhu, J. Feng, P. Zhang, and Z. Cheng, "Interpretable spatio-temporal attention LSTM model for flood forecasting," *Neurocomputing*, vol. 403, pp. 348–359, Aug. 2020, doi: 10.1016/j.neucom.2020.04.110.

[7] C.-F. R. Chen *et al.*, "Deep analysis of CNN-based spatio-temporal representations for action recognition," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2021, pp. 6161–6171, doi: 10.1109/CVPR46437.2021.00610.

[8] C. Wang, K. Cho, and J. Gu, "Neural machine translation with byte-level subwords," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 05, pp. 9154–9160, Apr. 2020, doi: 10.1609/aaai.v34i05.6451.

[9] M. C. Schiappa *et al.*, "A large-scale robustness analysis of video action recognition models," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2023, pp. 14698–14708, doi: 10.1109/CVPR52729.2023.01412.

[10] J. Diez-Sierra and M. del Jesus, "Long-term rainfall prediction using atmospheric synoptic patterns in semi-arid climates with statistical and machine learning methods," *Journal of Hydrology*, vol. 586, p. 124789, Jul. 2020, doi: 10.1016/j.jhydrol.2020.124789.

[11] Y. Jin, D. Yan, X. Kang, A. Chong, H.-- Sun, and S. Zhan, "Forecasting building occupancy: a temporal-sequential analysis and machine learning integrated approach," *Energy and Buildings*, vol. 252, p. 111362, Dec. 2021, doi: 10.1016/j.enbuild.2021.111362.

[12] Y. Zhang *et al.*, "Evaluation and comparison of precipitation estimates and hydrologic utility of CHIRPS, TRMM 3B42 V7 and PERSIANN-CDR products in various climate regimes," *Atmospheric Research*, vol. 265, p. 105881, Jan. 2022, doi: 10.1016/j.atmosres.2021.105881.

[13] "Spatiotemporal convolutional LSTM with attention mechanism for monthly rainfall prediction," *Communications in Mathematical Biology and Neuroscience*, vol. 118, no. 1, pp. 1–25, 2022, doi: 10.28919/cmbn/7761.

[14] M. M. H. Khan, N. S. Muhammad, and A. El-Shafie, "Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting," *Journal of Hydrology*, vol. 590, p. 125380, Nov. 2020, doi: 10.1016/j.jhydrol.2020.125380.

[15] S. Mehdizadeh, "Assessing the potential of data-driven models for estimation of long-term monthly temperatures," *Computers and Electronics in Agriculture*, vol. 144, pp. 114–125, Jan. 2018, doi: 10.1016/j.compag.2017.11.038.

[16] R. Fredyan, M. R. N. Majiid, and G. P. Kusuma, "Spatiotemporal analysis for rainfall prediction using extreme learning machine cluster," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 13, no. 6, pp. 2240–2248, Dec. 2023, doi: 10.18517/ijaseit.13.6.18214.

[17] S. Narmadha and V. Vijayakumar, "Spatio-temporal vehicle traffic flow prediction using multivariate CNN and LSTM model," *Materials Today: Proceedings*, vol. 81, pp. 826–833, 2023, doi: 10.1016/j.matpr.2021.04.249.

[18] Q. Ge, H. Sun, Z. Liu, and X. Wang, "A data-driven intelligent model for landslide displacement prediction," *Geological Journal*, vol. 58, no. 6, pp. 2211–2230, Jun. 2023, doi: 10.1002/gj.4675.

[19] I. Kumar, B. K. Tripathi, and A. Singh, "Attention-based LSTM network-assisted time series forecasting models for petroleum production," *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106440, Aug. 2023, doi: 10.1016/j.engappai.2023.106440.

[20] A. Barth, A. Alvera-Azcárate, C. Troupin, and J.-M. Beckers, "DINCAE 2.0: multivariate convolutional neural network with error estimates to reconstruct sea surface temperature satellite and altimetry observations," *Geoscientific Model Development*, vol. 15, no. 5, pp. 2183–2196, Mar. 2022, doi: 10.5194/gmd-15-2183-2022.

[21] S. Kim *et al.*, "Deep-hurricane-tracker: tracking and forecasting extreme climate events," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2019, pp. 1761–1769, doi: 10.1109/WACV.2019.00192.

[22] Y. M. Souto, F. Porto, A. M. Moura, and E. Bezerra, "A spatiotemporal ensemble approach to rainfall forecasting," in *2018 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2018, pp. 1–8, doi: 10.1109/IJCNN.2018.8489693.

[23] K. E. Setiawan, R. Fredyan, and I. N. Alam, "Enhancing el niño-southern oscillation prediction using an attention-based sequence-to-sequence architecture," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 6, pp. 7057-7066, Dec. 2024, doi: 10.11591/ijece.v14i6.pp7057-7066.

[24] Y. Wang *et al.*, "PredRNN: a recurrent neural network for spatiotemporal predictive learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 2208–2225, Feb. 2023, doi: 10.1109/TPAMI.2022.3165153.

[25] D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun, and M. Paluri, "A closer look at spatiotemporal convolutions for action recognition," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 6450–6459, doi: 10.1109/CVPR.2018.00675.

[26] G. Singh and F. Cuzzolin, "Recurrent convolutions for causal 3D CNNs," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Oct. 2019, pp. 1456–1465, doi: 10.1109/ICCVW.2019.00183.

[27] C. Cheng, C. Zhang, Y. Wei, and Y.-G. Jiang, "Sparse temporal causal convolution for efficient action modeling," in *Proceedings of the 27th ACM International Conference on Multimedia*, Oct. 2019, pp. 592–600, doi: 10.1145/3343031.3351054.

[28] Z. Xu, Y. Cao, and Y. Kang, "Deep spatiotemporal residual early-late fusion network for city region vehicle emission pollution prediction," *Neurocomputing*, vol. 355, pp. 183–199, Aug. 2019, doi: 10.1016/j.neucom.2019.04.040.

[29] M. Mudigonda *et al.*, "Deep learning for detecting extreme weather patterns," *Deep Learning for the Earth Sciences*, pp. 161–185, 2021, doi: 10.1002/9781119646181.ch12.

BIOGRAPHIES OF AUTHORS

Renaldy Fredyan received a master's degree in computer science from Bina Nusantara University in Indonesia in 2023. He is currently employed by Indonesia's Bina Nusantara University as a lecturer in the Department of Computer Science. His areas of interest in research include time series, machine learning, pattern recognition, and computer vision. He can be contacted at email: renaldy.fredyan@binus.ac.id.

Karli Eka Setiawan received an S.Si. degree in mathematics from Padjadjaran University in 2015 and an M.Kom. degree in computer science from Bina Nusantara University in 2023. He achieved outstanding grades of excellence, summa cum laude, and best graduate in his master's study. His research interests cover artificial intelligence, computer vision, data science, data security, blockchain, and mathematical modeling. He is an associate lecturer specialist S2 in the School of Computer Science at Bina Nusantara University. He also has experience as a credit analyst at a conventional bank, which is part of a state-owned enterprise. He can be contacted at email: karli.setiawan@binus.ac.id.

Kelvin Asclepius Minor is a mathematics lecturer at Bina Nusantara University (BINUS). Driven by a strong curiosity to understand how things work, he aims to contribute to the organization's growth by leveraging his knowledge and focusing on detailed descriptions, precise specifications, and factual insights. His passion for engaging directly with his environment aligns with his interest in learning systems and understanding the actions required to achieve tangible outcomes. He can be contacted at email: kelvin.minor@binus.ac.id.