Proceedings of the 10th World Congress on Civil, Structural, and Environmental Engineering (CSEE 2025)
Barcelona, Spain - April, 2025
Paper No. ICGRE 153
DOI: 10.11159/icgre25.153

A Comparative Analysis of Deep Learning Approaches for Rainfall Forecasting in Taiwan

Alok Kumar Sharma[1] and Sung-Chi Hsu[2, *]
[1]Department of Computer Science and Information Engineering
Chaoyang University of Technology, Taichung, Taiwan (R.O.C)
rbaloksharma@gmail.com
[2]Department of Construction Engineering
Chaoyang University of Technology, Taichung, Taiwan (R.O.C)
schsu@mail.cyut.edu.tw

Abstract **–** Forecasting rainfall is critical for agriculture, urban planning, and disaster management. This study evaluated the performance of three models: LSTM, CNN+LSTM, and BiLSTM, for forecasting rainfall in Hualien City, situated on Taiwan's eastern coast. The dataset utilized was sourced from the Department of Atmospheric Sciences at Chinese Culture University, covering the period from 1998 to 2018. This comprehensive dataset includes measurements such as datetime, temperature, humidity, air pressure, wind direction, and wind speed, providing a robust foundation for predictive modelling. The study's findings demonstrated that the BiLSTM model significantly outperformed the other models, with an MSE of 6.21, an MAE of 0.56, and an RMSE of 2.49. These findings underscore the BiLSTM's superior ability to identify temporal dependencies and handle the complexities of atmospheric data compared to the simpler LSTM and hybrid CNN+LSTM models. This study improves our understanding of deep learning applications in meteorological forecasting and demonstrates the efficacy of the BiLSTM model in managing the intricacies of time-series data processing.

Keywords**:** Rainfall Prediction, Deep learning, LSTM, Convolutional Neural Network, BiLSTM

1. Introduction

Rainfall forecasting is very important in managing water resources, planning agriculture, preparing for disasters, and protecting the environment, especially for Taiwan, which is located in a highly prone area to extreme weather conditions. Given the geographical location of Taiwan, with its topography, it has endowed the island with a highly variable climate, making it prone to typhoons, monsoons, and rainfall that is highly irregular [1]. Severe flooding, landslides, and other weather hazards frequently plague these climatic conditions, necessitating accurate and timely rainfall forecasts to mitigate associated risks.

The intricate dynamics of the atmosphere, involving millions of interacting variables like temperature, pressure, humidity, wind, and topography, make rainfall forecasting extremely challenging [2]. On the contrary, rainfall forecasting, which relies heavily on statistical models and numerical weather prediction techniques in most areas, has considerably improved over these years [3]. However, these models often struggle with the non-linear nature of meteorological data and the long-term dependencies inherent in weather systems. This limitation has led to growing interest in the application of machine learning (ML) and deep learning (DL) models, which have demonstrated promising results in addressing the challenges associated with non-linearity and long-range dependencies in time-series data. DL models involving RNNs and CNNs have recently emerged as very powerful tools for time-series forecasting in several domains, such as weather prediction. Among these, the LSTM network, CNN+LSTM hybrid model, and BiLSTM models have gained much attention because they have an evident capability to model complex temporal and spatial relations in data. Compared with traditional methods, these methods, on most applications, provide much more precise and robust forecasts owing to their intrinsic ability to learn the pattern from history without explicit feature engineering.

Several studies have compared the performance of various ML and DL models for rainfall forecasting. For instance, Yao et al. [4] employed a CNN+LSTM and GRU model to predict rainfall, and the results showed that the CNN+LSTM performed well. Aderyani et al. [5] evaluated the performance of ML and DL models on rainfall data, specifically comparing support vector regression (SVR), LSTM, and CNN, with the results showing that LSTM achieved the best performance. V et al. [6] utilized a BiLSTM model and compared its performance against RNN, LSTM, and a combination of RNN+LSTM for rainfall prediction. The results indicated that BiLSTM delivered superior performance.

This study aims to evaluate and compare the performance of LSTM, CNN+LSTM, and BiLSTM models in forecasting rainfall in Taiwan. Each of them adds its own unique strengths to time-series forecasting modelling, and their relative performances have important implications for the improvement of rainfall forecasting over a region like Taiwan with highly complicated weather. This research will identify which of the models fits best in capturing the temporal and spatial dependencies in rainfall data and hence contribute to the general understanding of how machine learning techniques can help improve weather-forecasting systems.

2. Method

2.1. Data and Data Processing

This study utilized rainfall data from Hualien, Taiwan, acquired from the Department of Atmospheric Sciences at Chinese Culture University. The dataset contains a total of 173,657 observations that were recorded between 1998 and 2018. It includes a collection of six critical meteorological parameters: wind speed, wind direction, air pressure, temperature, and humidity. This study executed a data cleansing procedure that entailed the elimination of duplicates and null values to guarantee the precision of the analysis. Furthermore, we detected other instances when the wind direction was inaccurately documented as 999. These entries were eliminated to preserve the dataset's integrity and enhance the reliability of the meteorological evaluation. After data elimination, the size is 173,586.

2.2. Training and Test Dataset

This study divided the data into three distinct segments: training, validation, and testing, aiming to evaluate the effectiveness of the model under different conditions. The last year's data was specifically reserved for testing to simulate real-world application and ensure the model's performance on recent, unseen data. The remaining dataset was used for training, with 20% of it set aside for validation during the training phase. This approach allows for continuous model assessment and tuning, helping to prevent overfitting and ensuring that the model generalizes well to new data. Finally, the study employed the testing dataset to rigorously evaluate the model's predictive accuracy and robustness, ensuring that the results are reliable and replicable in practical scenarios.

2.3. Models

2.3.1. LSTM

The LSTM [7] neural network architectures belong to a subclass of Recurrent Neural Networks (RNN) architectures used mainly for handling and modelling sequential and time series data. The LSTMs constitute a clear restructuration of RNN's constant folds because they can effectively use a memory cell and therefore escape the vanishing gradient problem commonly experienced with RNNs. This makes it particularly advantageous for tasks like language modelling, speech recognition, and sequence prediction, which require the utilization of long-term dependencies. Three types of gates, namely the input gate, forget gate, and output gate, are utilized in LSTMs to prevent the influx of information. The input gate directs the addition of new information to the system, the forget gate eliminates unwanted information from the array of cells, and the output gate determines the new output. This gating mechanism enables the information systems of LSTMs to maintain their current state for an extended period while discarding irrelevant data. This study utilized two LSTM layers with 50 units each and employed the Adam optimizer.

Fig. 1 shows the LSTM architecture, where, X_t represents the input at a time step, h_t represents the output, C_t represents the cell state and $\hat{C}_t$ represents the internal cell state. Three different gates are identified: f_t, - i_t, and o_t, which stand for forget gate, input gate, and output gate, respectively. And these gates are established upon very clear fundamental principles: The forget gate tells the cell which information needs to be forgotten, or deleted, within the internal cell state. The input gate tells the cell regarding what new information should be added to the internal cell state. The output gate basically determines the final result, or exactly what the cell is outputting, being some sort of improved version of the internal cell state.

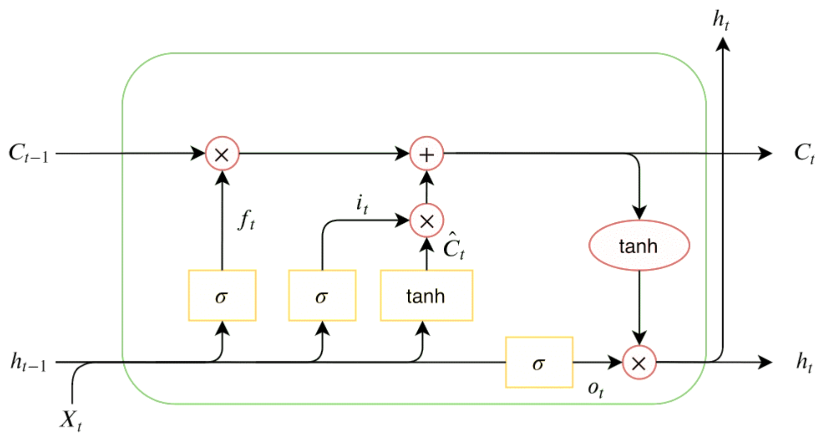

Fig. 1. LSTM Architecture

2.3.2. CNN + LSTM

The CNN+LSTM architecture combines convolutional neural networks (CNNs) and long short-term memory networks (LSTMs) to efficiently process data with both spatial and temporal characteristics. This model is particularly useful for applications like video analysis, speech processing, and time-dependent sensor data prediction [8]. This study combines specially designed LSTM layers with a selection of convolutional layers to enhance the extraction and processing of temporal features in the data. The model includes three convolutional layers, of which the first has 16 filters, the second 32 filters, and the third has 64 filters. Each Conv1D layer uses a kernel size of 2, followed by the MaxPooling1D layer to carry out dimensionality reduction while retaining the most significant features. After the convolutional layers, the model uses two LSTM layers with 50 units each, intended for capturing temporal dependencies. We set only the first of these LSTM layers to return sequences for stacking purposes. At the top, the architecture ends with a fully connected, dense output layer, as shown in Fig.2. The model optimizes for predictive accuracy during training by using the Adam optimizer and a mean squared error loss function.

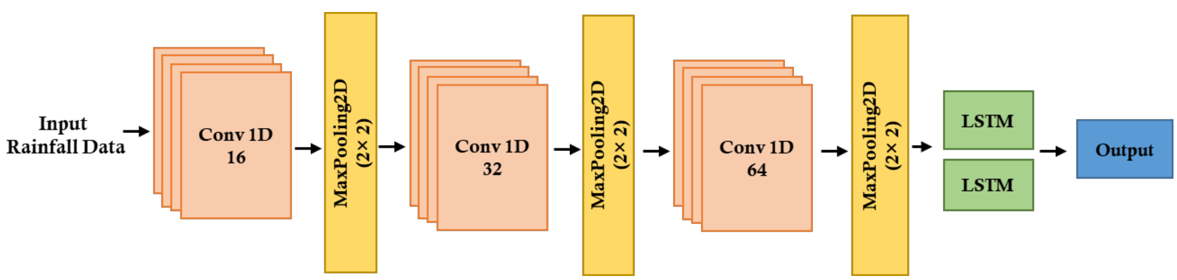

Fig. 2. CNN + LSTM Architecture

2.3.3. BiLSTM

BiLSTM [9] extends the standard LSTM architecture by adding two parallel LSTM layers: one processing the input sequence in the forward direction and the other in the backward direction. This method allows the model to remember information from the past, and the information from the time steps ahead of a certain time step is really important for most tasks that require context from the preceding and succeeding time steps. The ability of BiLSTM to perform dual processing is highly beneficial in various applications, including Natural Language Processing (NLP), where a single word can rely on both medieval Northern words and modern Northern words. In time series analysis, Bi-LSTM can enhance the results and spatial strategic relationships by considering temporal relationships in both time periods [10]. This is crucial for disease prediction from sensor data, as it incorporates data from both the past and the future. This study utilized two BiLSTM layers with 50 units each and employed the Adam optimizer. BiLSTM architecture shown in Fig. 3, where BiLSTM model displaying the input and output layers. The red arrows indicate the backward sequence direction, and the green arrows show the forward sequence direction.

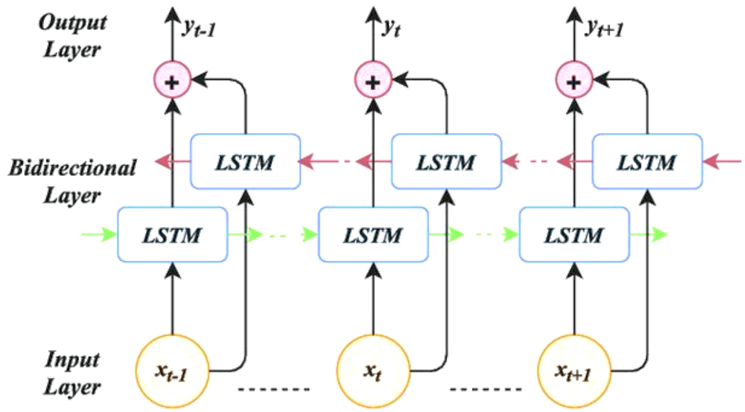

Fig. 3. BiLSTM Architecture

2.4. Evaluation Matrices

This research evaluated the model's performance using three metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). In equations (1), (2), and (3), γ represents actual value and γ' represents the predicted value.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\gamma - \gamma')^2 \tag{1}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\gamma - \gamma'| \tag{2}$$

$$RMSE = \sqrt{\sum_{i=1}^{N}\frac{(\gamma - \gamma')^2}{N}} \tag{3}$$

3. Results

Table 1 provides a comprehensive statistical summary of six meteorological variables collected from 173,586 observations. The mean air pressure is recorded at 1011.15, with a standard deviation (SD) of 6.61, indicating a relatively consistent distribution around the mean. The variance of air pressure is 43.73, and its kurtosis is 1.47, suggesting a distribution slightly more peaked than a normal distribution. Humidity has a mean value of 77.33%, with an SD of 10.26, reflecting moderate variability. The kurtosis of -0.15 indicates a near-normal distribution, with minor deviations. Temperature exhibits a mean of 23.72°C and an SD of 4.57, with a variance of 20.91 and a negative kurtosis of -0.63, indicating a flatter distribution compared to a normal curve.

Wind speed averages 2.94 m/s, with a larger SD of 1.82 and kurtosis of 6.35, pointing to a highly peaked distribution and occasional extreme values, as evidenced by a maximum wind speed of 42.7 m/s. Wind direction, with a mean of

176.89° and an SD of 108.34, shows substantial variability in directionality. The variance in wind direction is 12009.93, and the kurtosis is -0.06, which indicates a distribution close to normal. Rainfall has a relatively low mean of 0.25 mm, with significant variability (SD 1.79), and a maximum value of 96 mm, reflecting rare but intense rainfall events. The kurtosis of 388.46 suggests a highly leptokurtic distribution, with infrequent but extreme values dominating the data.

Overall, the data set reflects a variety of distribution shapes across different meteorological parameters, with specific interest in the higher variability and kurtosis observed in wind speed and rainfall, indicating the presence of extreme weather conditions in these variables. This summary provides essential insights into the variability and distribution characteristics of these key meteorological factors, which could have significant implications for forecasting models and climate studies.

Table 1: Descriptive statistics of rainfall data

Variable	Count	Mean	STD	Min	25%	50%	75%	Max	Variance	Kurtosis
Air Pressure	173586	1011.15	6.61	946.8	1006.5	1011	1015.9	1031.9	43.73	1.47
Humidity	173586	77.33	10.26	26	71	78	85	100	105.35	-0.15
Temperature	173586	23.72	4.57	8	20.3	24.2	27.2	36.1	20.91	-0.63
Wind Speed	173586	2.94	1.82	0	1.7	2.5	3.8	42.7	3.31	6.35
Wind Direction	173586	176.89	108.34	0	50	210	250	360	12009.93	-0.06
Rainfall (MM)	173586	0.25	1.79	0	0	0	0	96	3.22	388.46

Table 2 provides a performance comparison of three deep learning models: LSTM, CNN + LSTM, and BiLSTM, based on evaluation metrics including MSE, MAE, and RMSE. The LSTM model demonstrates an MSE of 6.53, an MAE of 0.57, and an RMSE of 2.56, indicating a reasonable performance in minimizing prediction errors. The CNN + LSTM model shows a slight improvement over the LSTM model, with a lower MSE of 6.40 and RMSE of 2.53, while maintaining a comparable MAE of 0.56. This reduction in error values suggests that integrating convolutional layers with LSTM enhances the model's ability to extract spatial and temporal patterns in the data.

The BiLSTM model outperforms the other two models, achieving the lowest MSE of 6.21, an RMSE of 2.49, and a MAE of 0.56. The bidirectional structure of the BiLSTM enables the model to leverage both forward and backward temporal information, which likely contributes to the reduction in overall error metrics. These results indicate that the BiLSTM model is more effective in minimizing predictive errors compared to both the LSTM and CNN + LSTM models. Thus, the bidirectional architecture proves advantageous for this specific task, resulting in consistently lower error values across the metrics evaluated.

Table 2: Performance metrics of LSTM, CNN+LSTM, and BiLSTM Model

Model	MSE	MAE	RMSE
LSTM	6.53	0.57	2.56
CNN + LSTM	6.40	0.56	2.53
BiLSTM	6.21	0.56	2.49

Figs. 4, 5, and 6 compare actual rainfall with the predictions of three models: LSTM, CNN + LSTM, and BiLSTM from January 1, 2018, to April 1, 2018. In Fig. 4, the LSTM model shows predictions that generally follow the timing of actual rainfall events but significantly underestimate the intensity, especially during periods of heavy rainfall, such as mid-February and mid-March. Fig. 5, using the CNN + LSTM model, presents similar results, but there is a slight improvement in capturing the intensity of rainfall compared to the LSTM model. However, it still struggles to match the peaks of the actual rainfall. In Fig. 6, the BiLSTM model performs better than the other two models. It more accurately predicts the timing and intensity of rainfall events, although it still underestimates the magnitude during the highest rainfall periods. Overall, while all three models are able to capture the general trend of rainfall occurrences, the BiLSTM model demonstrates the most promising performance in terms of aligning with actual rainfall values, though underestimation of peak intensities persists across all models.

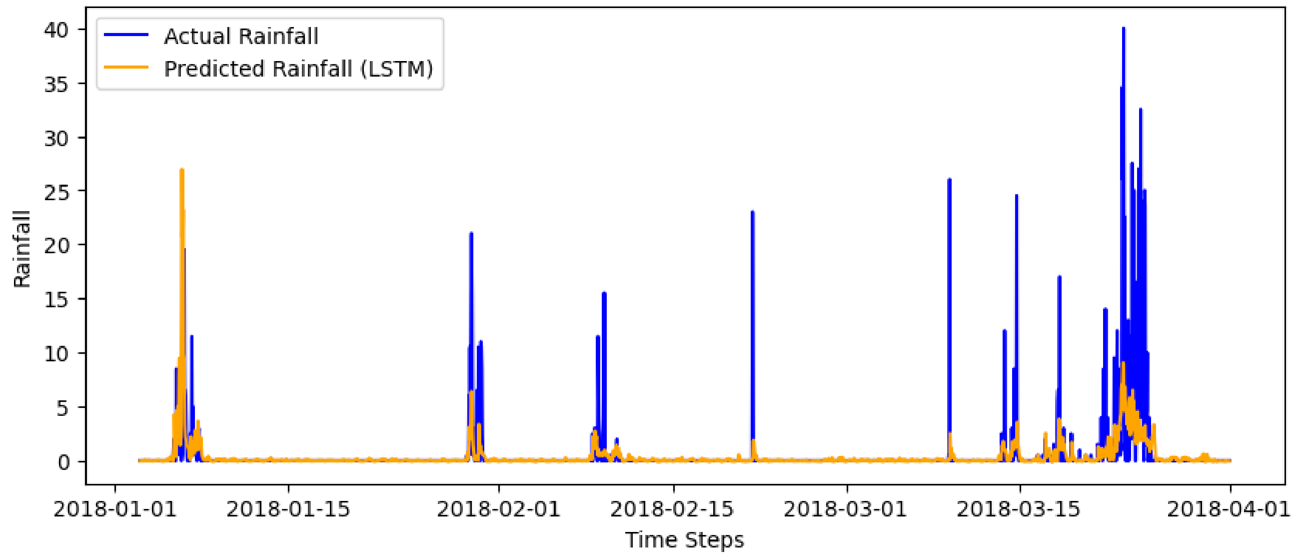

Fig. 4: Comparison between actual and predicted rainfall using the LSTM model.

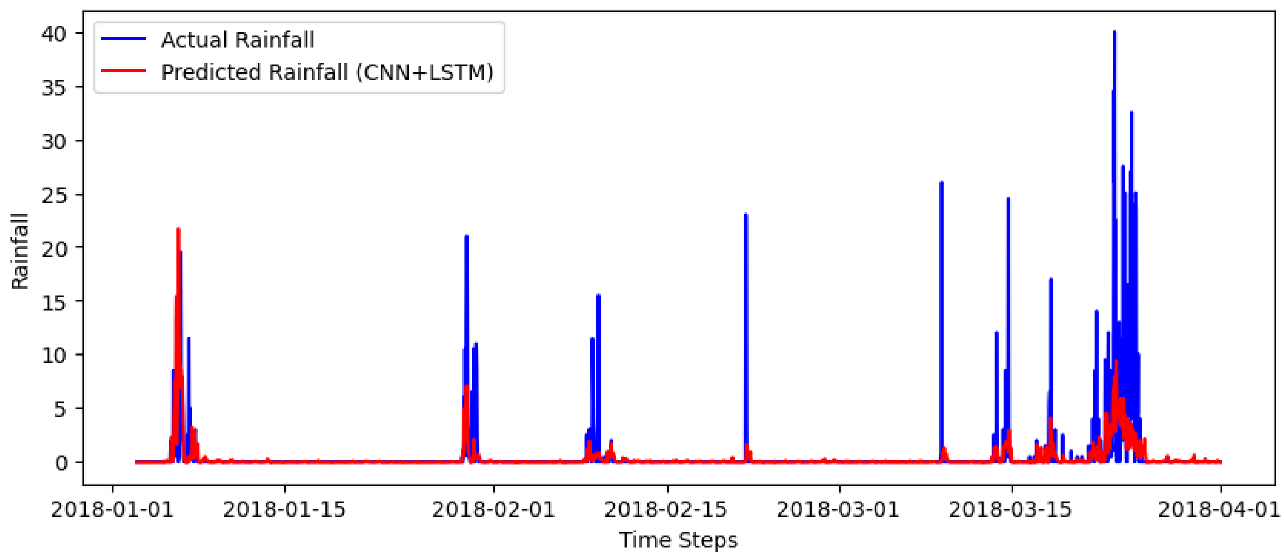

Fig. 5: Comparison between actual and predicted rainfall using the CNN+LSTM model.

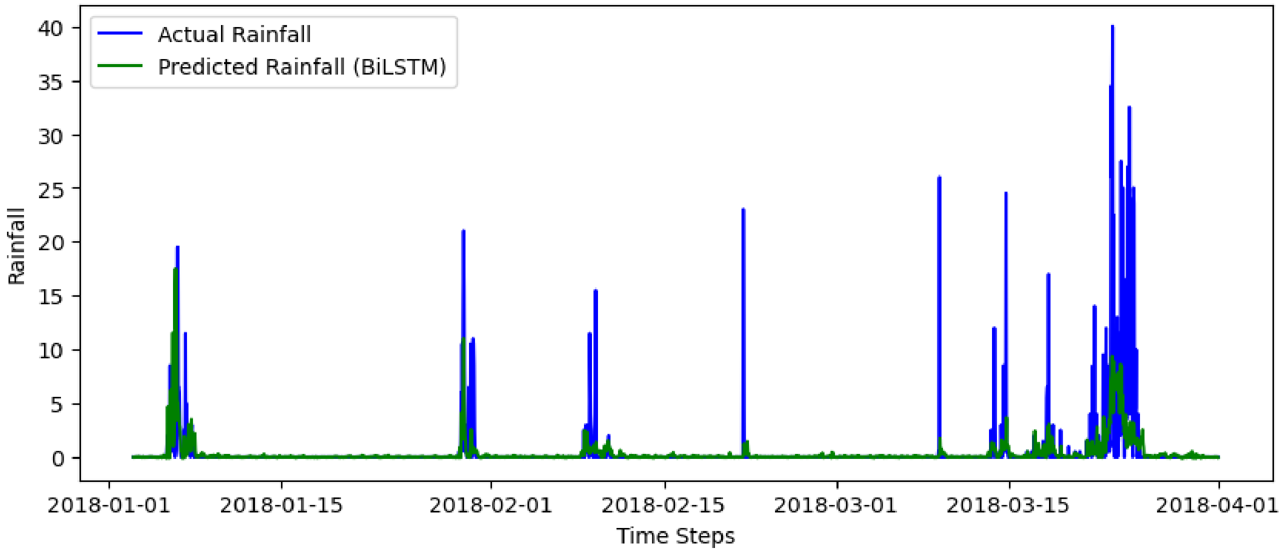

Fig. 6: Comparison between actual and predicted rainfall using the BLSTM model.

4. Conclusion

This study successfully evaluated the performances of the three deep learning models, LSTM, CNN+LSTM, and BiLSTM, in the rainfall forecast in Hualien, Taiwan. The outcomes revealed that a bidirectional structure BiLSTM model outperforms both the LSTM and CNN+LSTM models on all three important metrics: MSE, MAE, and RMSE. The better performance of BiLSTM is due to the fact that it can learn the past and future temporal dependencies, which are crucial in the time-series forecast of complex meteorological data. While all the models have represented the general trend of rainfall occurrences, the effectiveness of the BiLSTM in reducing predictive errors is reflected by its lower MSE and RMSE values. However, each of these three models shared the same problem when they tended to underestimate the peak rainfall events, particularly at times when extreme weather conditions are present. This means that while the BiLSTM model yields a significant improvement, there is room for further refinement in predicting the magnitude of rare but severe rainfall occurrences.

The findings of this study demonstrated how advanced deep learning models can enhance the performance of rainfall forecasting for a region with complex weather conditions, such as Taiwan. An improved forecast model may allow for the following: better disaster management and water resource planning, through appropriate agricultural decisions that can reduce extreme weather risks. Such research could be directed either at further optimization of the model or at investigations into hybrid architectures that combine strengths of various deep learning methods towards further improved forecasting.

References

[1] C.-H. Chang, M.-K. Chung, S.-Y. Yang, C.-T. Hsu, and S.-J. Wu, “A Case Study for the Application of an Operational Two-Dimensional Real-Time Flooding Forecasting System and Smart Water Level Gauges on Roads in Tainan City, Taiwan,” *Water*, vol. 10, no. 5, p. 574, Apr. 2018, doi: 10.3390/w10050574.

[2] N. Thanvisitthpon, K. Kallawicha, and H. J. Chao, “Introduction to meteorology, weather, and climate,” in *Health and Environmental Effects of Ambient Air Pollution*, Elsevier, 2024, pp. 303–329. doi: 10.1016/B978-0-443-16088-2.00004-1.

[3] M. Troin, R. Arsenault, A. W. Wood, F. Brissette, and J. Martel, “Generating Ensemble Streamflow Forecasts: A Review of Methods and Approaches Over the Past 40 Years,” *Water Resour. Res.*, vol. 57, no. 7, Jul. 2021, doi: 10.1029/2020WR028392.

[4] Z. Yao, Z. Wang, D. Wang, J. Wu, and L. Chen, “An ensemble CNN-LSTM and GRU adaptive weighting model based improved sparrow search algorithm for predicting runoff using historical meteorological and runoff data as

input," *J. Hydrol.*, vol. 625, p. 129977, Oct. 2023, doi: 10.1016/j.jhydrol.2023.129977.

[5] F. R. Aderyani, S. Jamshid Mousavi, and F. Jafari, "Short-term rainfall forecasting using machine learning-based approaches of PSO-SVR, LSTM and CNN," *J. Hydrol.*, vol. 614, p. 128463, Nov. 2022, doi: 10.1016/j.jhydrol.2022.128463.

[6] C. M. V., C. P., H. S. M., and A. S., "A Deep Learning Ensemble Model for Short-Term Rainfall Prediction," in *2022 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, IEEE, Mar. 2022, pp. 135–138. doi: 10.1109/WiSPNET54241.2022.9767163.

[7] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: a deep learning approach for short-term traffic forecast," *IET Intell. Transp. Syst.*, vol. 11, no. 2, pp. 68–75, Mar. 2017, doi: 10.1049/iet-its.2016.0208.

[8] W. Lu, J. Li, Y. Li, A. Sun, and J. Wang, "A CNN-LSTM-Based Model to Forecast Stock Prices," *Complexity*, vol. 2020, pp. 1–10, Nov. 2020, doi: 10.1155/2020/6622927.

[9] A. Aziz Sharfuddin, M. Nafis Tihami, and M. Saiful Islam, "A Deep Recurrent Neural Network with BiLSTM model for Sentiment Classification," in *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, IEEE, Sep. 2018, pp. 1–4. doi: 10.1109/ICBSLP.2018.8554396.

[10] I. K. Ihianle, A. O. Nwajana, S. H. Ebenuwa, R. I. Otuka, K. Owa, and M. O. Orisatoki, "A Deep Learning Approach for Human Activities Recognition From Multimodal Sensing Devices," *IEEE Access*, vol. 8, pp. 179028–179038, 2020, doi: 10.1109/ACCESS.2020.3027979.