




Article

Flood Forecasting Using Hybrid LSTM and GRU Models with Lag Time Preprocessing

Yue Zhang ¹, Zimo Zhou ¹, Jesse Van Griensven Thé ², Simon X. Yang ^{1,*} and Bahram Gharabaghi ^{1,*}

¹ School of Engineering, University of Guelph, 50 Stone Road East, Guelph, ON N1G 2W1, Canada; zhang26@uoguelph.ca (Y.Z.); zzhou17@uoguelph.ca (Z.Z.)

² Lakes Environmental, 170 Columbia St. W, Waterloo, ON N2L 3L3, Canada

* Correspondence: syang@uoguelph.ca (S.X.Y.); bgharaba@uoguelph.ca (B.G.)

Abstract: Climate change and urbanization have increased the frequency of floods worldwide, resulting in substantial casualties and property loss. Accurate flood forecasting can offer governments early warnings about impending flood disasters, giving them a chance to evacuate and save lives. Deep learning is used in flood forecasting to improve the timeliness and accuracy of flood water level predictions. While various deep learning models similar to Long Short-Term Memory (LSTM) have achieved notable results, they have complex structures with low computational efficiency, and often lack generalizability and stability. This study applies a spatiotemporal Attention Gated Recurrent Unit (STA-GRU) model for flood prediction to increase the models' computing efficiency. Another salient feature of our methodology is the incorporation of lag time during data preprocessing before the training of the model. Notably, for 12-h forecasting, the STA-GRU model's R-squared (R²) value increased from 0.8125 to 0.9215. Concurrently, the model manifested reduced root mean squared error (RMSE) and mean absolute error (MAE) metrics. For a more extended 24-h forecasting, the R² value of the STA-GRU model improved from 0.6181 to 0.7283, accompanied by diminishing RMSE and MAE values. Seven typical deep learning models—the LSTM, the Convolutional Neural Networks LSTM (CNNLSTM), the Convolutional LSTM (ConvLSTM), the spatiotemporal Attention Long Short-Term Memory (STA-LSTM), the GRU, the Convolutional Neural Networks GRU (CNNGRU), and the STA-GRU—are compared for water level prediction. Comparative analysis delineated that the use of the STA-GRU model and the application of the lag time pre-processing method significantly improved the reliability and accuracy of flood forecasting.

Keywords: flood forecasting; water level prediction; STA-LSTM; STA-GRU



Citation: Zhang, Y.; Zhou, Z.; Van Griensven Thé, J.; Yang, S.X.; Gharabaghi, B. Flood Forecasting Using Hybrid LSTM and GRU Models with Lag Time Preprocessing. *Water* **2023**, *15*, 3982. <https://doi.org/10.3390/w15223982>

Academic Editor: Akira Kawamura

Received: 9 October 2023

Revised: 11 November 2023

Accepted: 13 November 2023

Published: 16 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As urbanization and climate change intersect, the flood risks escalate [1–7]. One primary reason is the swifter water runoff from surfaces that are impervious to water absorption [8]. This phenomenon is closely tied to land-use patterns, which play a pivotal role in flood predictions. The surge in urbanization contributes to the proliferation of these impervious surfaces, amplifying rainwater runoff. In tandem, factors like the dwindling of vegetation and forests, shifts in agricultural land management, and modifications to rivers and wetlands all influence the volume and velocity of water flow. Given these interdependencies, rendering accurate flood predictions mandates an integrative approach, taking into account these land-use dynamics alongside other pertinent data [9–11]. These flooding events are more than mere natural phenomena; they pose grave threats to human safety and have the potential to cause significant economic damages, especially in regions more susceptible to inundation [12–14]. Recognizing the gravity of these threats, governments have heavily invested in early flood warning and forecasting systems [15]. These systems do more than just signal potential dangers; they are critical assets in both safeguarding lives and substantially reducing property damage by facilitating the timely implementation of preventive protection measures such as sandbags [16,17].

Traditional methods of assessing flood risks, while foundational, are no longer adequate on their own. The paradigm has shifted towards predictive models that can proactively alert communities about impending flood threats [18,19]. An essential feature of these systems is the provision of varying lead times, which are invaluable for both managing and preemptively addressing the risks associated with imminent flood events and other related disasters [20,21].

But how do these systems work, and what makes them so effective? They are meticulously designed to provide insights into the expected scale, onset, locale, and potential repercussions of a flood event [22–24]. These predictions are not based on guesswork; they are underpinned by data diligently collected throughout the year from strategically placed sensors in water basins, inclusive of lakes and rivers, as well as from flood deterrent structures like dams, dikes, and embankments. Moreover, purpose-built infrastructures for flood prediction and monitoring play a pivotal role in data collection, emphasizing that the quality of the dataset is directly proportional to the forecasting model's efficacy [25,26].

In the realm of flood prediction, three variables stand out in their significance: precipitation, river flow and water levels. The data on rainfall offer insights into its intensity and duration, which in turn affects the volume of water flowing into the river system [27–29]. Concurrently, the river's current water level acts as a barometer for its capacity to accommodate incoming water surges. An accurate flood prediction hinges on a nuanced understanding of the dynamics between these two factors. As soon as the soil's moisture levels or the river's capacity reach critical thresholds, flood risks amplify [30–33]. Through continuous monitoring and data analysis of both precipitation and river water levels, these sophisticated forecasting models can identify and highlight patterns indicative of potential flood events.

1.1. The Flood Prediction Models and Lag Time Preprocessing

Long Short-Term Memory (LSTM) is a form of Recurrent Neural Network (RNN) intended to address the long-term dependency issue encountered by RNNs when processing extensive sequence data [34,35]. In recent years, LSTM has achieved considerable success across multiple domains, including natural language processing, speech recognition, and time series prediction [36].

LSTM exhibits significant potential in flood forecasting, a prototypical time series prediction problem [37–39]. This process requires the handling and understanding of continuous meteorological and hydrological data (such as precipitation, river water levels, soil moisture, etc.), forming the basis for future flood prediction. The unique internal structure of LSTM, capable of processing and memorizing long-term sequential dependencies, makes it an ideal candidate for solving such problems [40–44].

The application of the LSTM model in flood prediction continues to evolve. Initial research primarily focused on employing LSTM to model and predict rainfall and river water levels at individual sites [45]. As deep learning technology advanced, researchers began exploring more complex models, such as integrating Convolutional Neural Networks (CNN) with LSTM, to handle meteorological and hydrological data across multiple geographical locations, thereby further enhancing the accuracy and timeliness of flood predictions [46–48].

It is crucial to note that flood prediction is not only a data-driven problem, but also requires understanding and consideration of various complex influencing factors such as geography, climate, and human activities (ref. [49]). Currently, hybrid models are receiving significant attention because they enhance the generalizability, and stability of single models. Although LSTM, the Convolutional Neural Networks LSTM (CNN-LSTM), the Convolutional LSTM (ConvLSTM) and other deep learning models have shown tremendous potential in flood prediction, ongoing optimization and improvements of these models are needed in practice to better address the various challenges inherent in flood prediction [47,50–53]. Moreover, since the flooding prediction dataset contains not only time series but also spatial series, with a large amount of data over a long span of

time, the attention mechanism can assist the model to deal with long-term dependencies more effectively. The attention mechanism can help the model extract useful information from the input data more efficiently, thereby improving the accuracy of prediction. The spatiotemporal Attention LSTM (STA-LSTM) model has been used in flood forecasting and has achieved good results, as demonstrated in Table 1 [46,49,54].

Table 1. The performance of flood prediction models.

Reference	Model Name	Applicable to Spatiotemporal Data	Maximum Prediction Duration	Model Performance
Liu et al. (2023) [55]	RNN	No	12 h	$MSE = 0.936, RMSE = 0.124$
Dehghani et al. (2023) [53]	CNN	Yes	6 h	$NSE = 0.68\sim 0.74$
Liu et al. (2023) [55]	LSTM	No	12 h	$MSE = 0.942, RMSE = 0.109$
Dehghani et al. (2023) [53]	ConvLSTM	Yes	6 h	$NSE = 0.965\sim 0.986$
Zhang et al. (2022) [46]	CNNLSTM	Yes	24 h	$MAE = 3.52, MSE = 85.43$
Zhang et al. (2022) [46], Ding et al. (2020) [49]	STA-LSTM	Yes	24 h	$MAE = 2.88, MSE = 63.92, R_2 = 0.78\sim 0.96$

Prediction models should not only ensure accuracy but also strive for the highest possible computational efficiency. Developed in 2014, the Gated Recurrent Unit (GRU) is a prediction model founded on similar principles as the LSTM model [56]. While GRU and LSTM research findings share similarities, key distinctions exist between the two models. GRU, for example, boasts a superior numeration ability, signifying its capacity to effectively capture and retain essential information over longer sequences [57,58]. This ability is vital for tasks involving long-term dependencies, where the model must take into account past information to make precise predictions.

The GRU model incorporates a gating mechanism that enables the model to selectively update its hidden state based on the input data [59]. In particular, GRU employs an update gate, combining the roles of LSTM forget and input gates. This combination simplifies the architecture, and reduces the number of parameters, leading to computational efficiency and quicker training times. Furthermore, GRU's streamlined design, merging the cell state and hidden state, fosters efficient information flow within the model. This architecture empowers the GRU to capture relevant information and discard unnecessary details, rendering it particularly suitable for tasks involving sequential data analysis. In the last two years, both the GRU and the Convolutional Neural Networks GRU (CNNGRU) models have been explored for their utility in flood prediction. The GRU model has proven to be more effective for short-term flood forecasts compared to LSTM [57]. While the CNN-GRU model has shown promise in flood prediction, enhancements in its performance for long-term forecasting are still necessary [60,61].

In hydrology, the lag time is the catchment response time between the rainfall and the runoff response [62]. With the increase in urban land taking over the previously rural land, infiltration rates can decrease and have adverse effects on flood risk for people living in the vicinity of a flood zone. Accurate modeling of flood events contribute to improved watershed management and mitigation of potential flood hazards [63,64].

The lag time is defined as the delay between the time a rainfall event over a watershed begins until runoff reaches its maximum peak [65,66]. The lag time of a catchment indicates the speed at which the river will react to increased precipitation and can be influenced by several parameters. These are slope, length and roughness of flow path, size of the basin, soil type and land use [67]. The estimation of lag time can be determined both empirically using formulas and by use of hydrological data [68]. This method utilizes data from an upstream precipitation station and a downstream flow monitoring site. The lag time of the stream is ascertained from the time difference between the peak precipitation and the peak runoff. Various studies have proposed the use of both hydrological data and empirical equations and have achieved success [69].

Furthermore, in the field of flood forecasting, the integration of spatiotemporal data is commonly adopted to enhance prediction accuracy. Within this context, the temporal

delay between upstream and downstream hydrological stations emerges as a critical factor. This delay is attributed to a combination of factors, including the river's natural flow rate, channel morphology and length, topographic gradients, and human interventions [70,71]. In the preprocessing of spatiotemporal data, we can determine the specific delay between the upstream and downstream by examining historical data, focusing on the time difference between peak values observed at each upstream station and the target downstream station [66,72,73]. The lag time of a catchment plays a significant role in stream flow model performance. With the addition of lag time to the stream flow-driven applications, the accuracy of the models' travel time accuracy will improve significantly [67].

1.2. Contribution

When employing deep learning methods, gradient vanishing is a common issue, which can be even worse when dealing with long sequence data. It leads to minute weight updates, thus causing the network to learn very slowly, or even fail to learn. LSTM and GRU models may serve as a good choice to attenuate such an issue. While the STA-LSTM model has been successfully employed in flood forecasting and yielded satisfactory results, such a model demands large computational efforts to perform the training and is often time-consuming.

Based on the above observation, our research augmented time series prediction with spatial information to improve forecasting capabilities. To overcome challenges in handling spatiotemporal datasets, the data are preprocessed before training, in which stage the lag time between rainfall volumes and the target station is determined, and the lag time between each hydrological station and the target station is determined. To better extract the features of data that contain both temporal and spatial information, the attention mechanism is used to deal with long-term dependencies effectively. Then based on the STA-LSTM model, the spatiotemporal Attention GRU (STA-GRU) model is constructed to reduce the model complexity and improve the computational efficiency. Flood forecasting models with high computational efficiency are capable of providing more timely warnings, thereby facilitating the faster implementation of emergency measures and mitigating the impact of disasters. Compared with the STA-LSTM, STA-GRU has similar mechanisms in data processing but a much simpler model architecture, and therefore comparable performance can be achieved while less computational effort is needed. Finally, the performance of seven models is compared, that is, Convolutional Neural Networks LSTM (CNN-LSTM), Convolutional LSTM (Conv-LSTM), spatiotemporal Attention LSTM (STA-LSTM), Convolutional Neural Networks GRU (CNN-GRU), and spatiotemporal Attention GRU (STA-GRU). These hybrid models synergize the unique strengths of their individual components, aiming to intricately capture the spatiotemporal dynamics inherent in flood prediction.

2. Materials and Methods

Originating from Orangeville, the Credit River winds its way through the landscapes of southern Ontario, Canada, meandering through towns such as Brampton, before gracefully merging with Lake Ontario in Mississauga. Complementing the river's natural allure, the surrounding areas boast multi-functional parks and verdant open spaces, inviting enthusiasts for activities ranging from fishing and hiking to immersive wildlife observation. However, with all its serene beauty, the Credit River is not without its perils. In times of torrential rain or during the spring melt, its tranquil waters can surge, posing flood threats. As a cautionary note, those residing or venturing near its banks are advised to be vigilant, heeding local weather updates and flood advisories.

Located in Mississauga, Credit River's station 02HB029 plays a pivotal role in flood forecasting for this bustling metropolitan area. Given that the Credit River courses directly through the heart of downtown Mississauga, accurately predicting the discharge in the southern part of the river is vital for safeguarding both lives and property. Although the real-time rainfall monitoring network in the Credit River watershed is limited, one precipitation monitor is situated near station 02HB025. With an aim to strike a balance

between simplicity and precision in the flood forecasting system, we have incorporated the data from this precipitation monitor into the purview of our manuscript. Ideally, a flawless early flood forecasting system would harmonize the objectives of governmental bodies, affected residents, and the insurance sector, facilitating a shared understanding of flood loss implications. Considering the escalating trend of insured catastrophic losses annually, it is imperative that a highly accurate early flood forecasting system is available.

Figure 1 shows that stations 02HB025, 02HB018, 02HB001, 02HB013, and 02HB031 are strategically located in the headwaters of the Credit River Watershed. These are positioned upstream of the vital station 02HB029, which is nestled in the flood-sensitive regions of downtown Mississauga, close to the basin of the Credit River watershed. Rainfall station 25 is located near station 02HB025, while rainfall station 18 is situated close to water station 02HB018. Both of them are upstream of the vital station 02HB029.

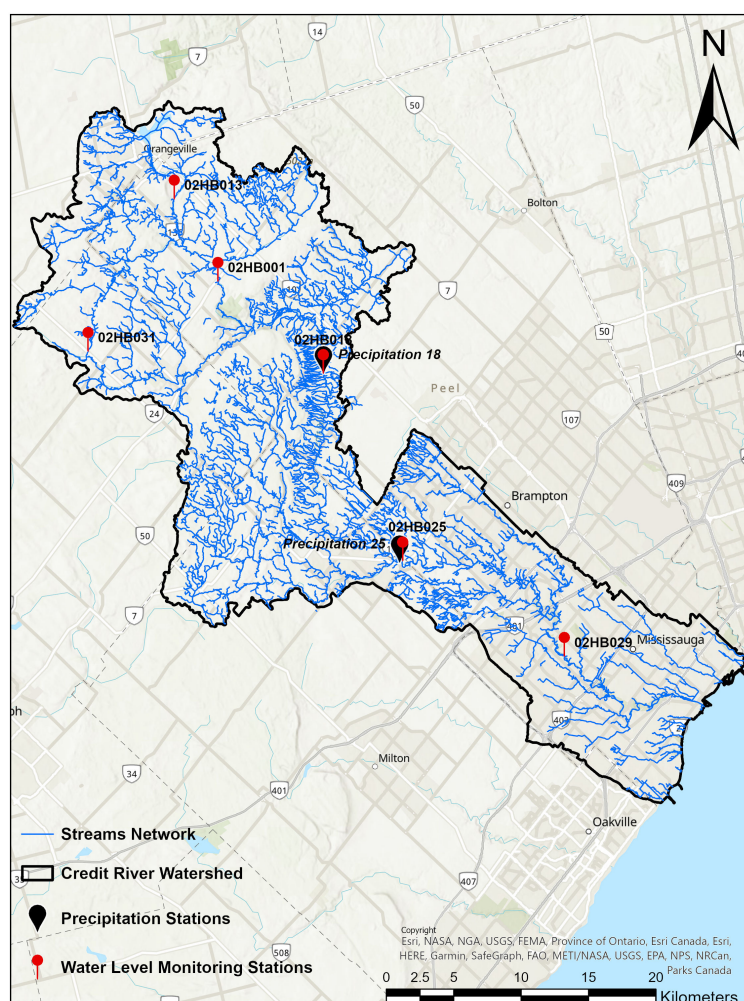


Figure 1. The network of real-time hydrometric monitoring stations in the Credit River Water.

Our hydrological prediction models have exhibited exceptional performance on spatiotemporal data, prompting our endeavor to further enhance their capabilities. Recognizing the notable success of the STA-LSTM in flood prediction using this type of data, the research attempted to further bolster the model's generalization capability and computational speed. To gain a comprehensive understanding, we have embarked on a comparison of STA-LSTM performance against other models, namely LSTM, CNN-LSTM, ConvLSTM, GRU, CNNGRU, and STA-GRU, specifically in the realm of flood forecasting with spatiotemporal data. The rationale behind spatially coupling LSTM and GRU-based models lies in their superior proficiency in handling spatiotemporal series data sequences, particularly when contrasted with their traditional counterparts.

Taking into account the urbanization levels and the expanse of the Credit River watershed, the catchment’s response time generally oscillates between three and eight hours. This variance is contingent upon the nature and duration of the rainfall event, ranging from abrupt yet intense summer thunderstorms to more prolonged rainfalls paired with snowmelt during spring. Flood warnings for the Credit River watershed cater to diverse users and objectives. Among these are the mobilization of operational teams and emergency responders, alerting the public about the specifics of the impending event, and, in severe instances, initiating evacuation and emergency protocols. In light of these requirements, our models were trained and tested for both 12-h and 24-h forecast scenarios, with subsequent evaluations of their accuracy.

2.1. The Correlation of Water Level, Discharge and Precipitation

The variables under consideration present a distinct positive correlation, as illustrated in Figure 2. The correlation coefficients, ranging from 0 to 1, further underline this observation. Such a trend indicates a deep-seated interconnectedness and mutual influence among the watershed stations, suggesting that changes or events in one station might resonate in others. This interrelation is not merely an interesting observation but holds practical implications. Precisely due to this pronounced correlation, utilizing these data as test or benchmark datasets for evaluating model performance gains increased weight. A model that can accurately predict under such conditions of high interrelatedness is likely to be robust and reliable.

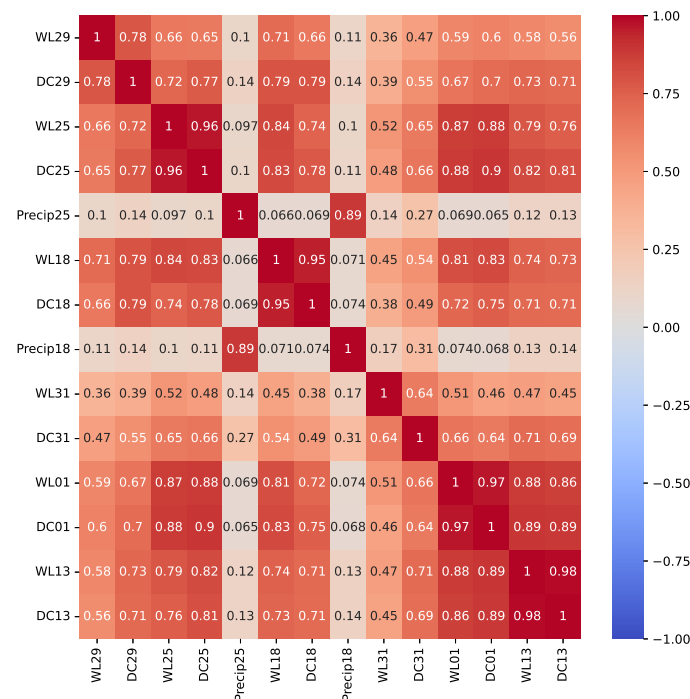


Figure 2. Matrix plot of correlation between the precipitation, water level and discharge.

2.2. The Water Level Lag Time between Each Station

The lag time between the upstream rainfall station and the target water station is ascertained by plotting the lag time graphs for nine different flood events. The average lag time between the rainfall events and runoff responses is used in this study. Similarly, the lag time between the upstream water level station and the target water level station is also determined. Additionally, the distances between the upstream water level station and the target water level station are measured, as shown in Table 2.

Table 2. The Lag time between each upstream water station and station 29.

Station No.	Average Lag Time (h)	Euclidean Distance (km)
02HB025	5	13.9
02HB018	7	27.6
02HB001	8	37.9
02HB031	9	41.9
02HB013	12	44.7

As the Euclidean distance increases, the lag time also tends to increase. Such a data preprocessing approach aims to ensure optimal correlation between each upstream station and the target station 02HB029, enhancing the predictive model’s performance.

2.3. Theoretical Background of the Models and Performance Metrics

In deep learning research for flood forecasting, common models like LSTM, GRU, CNNLSTM, ConvLSTM, and CNNGRU have been widely adopted. These models integrate temporal characteristics with convolutional features to process spatiotemporal data. LSTM and GRU emphasize capturing long-term sequence patterns, while CNNLSTM and CNNGRU combine the feature extraction capabilities of convolutional neural networks with the temporal modeling strengths of recurrent networks. In contrast, the STA-LSTM and STA-GRU models, which are more intricate in structure and specifically designed to capture spatiotemporal relationships, have not yet been extensively utilized. To cater to our spatiotemporal dataset, we have made adaptive modifications to these existing STA-LSTM and STA-GRU models, enhancing their efficacy in flood prediction.

2.3.1. STA-LSTM Model

The STA-LSTM model is tailored for spatiotemporal analyses. It is adept at processing datasets that intertwine time and spatial elements. While maintaining the foundational LSTM elements such as the forget, input, and output gates shown in Figure 3, the STA-LSTM integrates advanced structures like convolutional layers or attention mechanisms to discern spatial patterns more effectively as Figure 4.

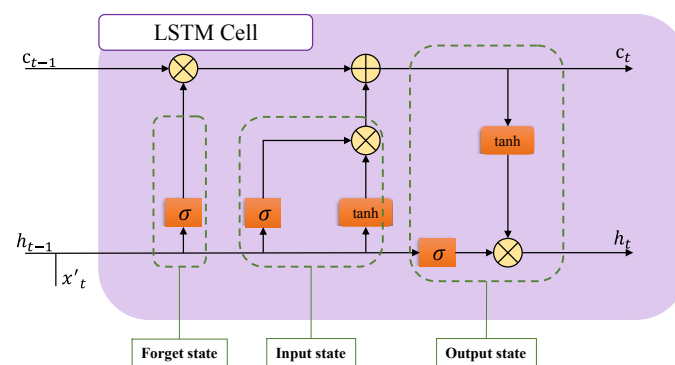


Figure 3. LSTM cell structure.

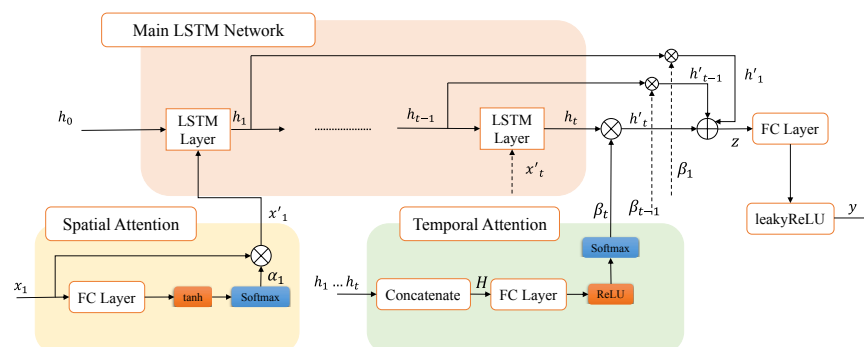


Figure 4. STA-LSTM model structure.

The main output of the STA-LSTM model is given as

$$h'_t = \beta_t \cdot h_t \tag{1}$$

$$z = \sum_{i=1}^t h'_i \tag{2}$$

$$y = \text{leakyReLU}(W_t \cdot z) \tag{3}$$

where β_t is the result of the temporal attention part, z represents the summation of h_1 to h_t , y shows the output of the model, and W_t is the weight. The Leaky ReLU [74] activation function is used before output.

In the Temporal Attention (TA) part, the equations are provided as follows:

$$H = \text{concate}(h_1 \dots h_t) \tag{4}$$

$$\{\beta_1, \dots, \beta_t\} = \text{softmax}(\text{ReLU}(W_{TA} \cdot H)) \tag{5}$$

where H as the concatenation of hidden states h_1 to h_t , and W_{TA} is the weight. The softmax activation function is defined as $\text{softmax}(x)_i = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}$ and ReLU activation function is defined as $\text{ReLU} = \max(0, x)$.

In the Spatial Attention (SA) part, the given equations are

$$S_t = \tanh(W_{SA} \cdot x_t) \tag{6}$$

$$\alpha_t = \text{softmax}(S_t) \tag{7}$$

where W_{SA} is the weight, and α_t represents the result after applying the softmax operation. The softmax activation function and tanh activation function are utilized in the result range of $(-1, 1)$. The resulting S_t will lie in the range between -1 and 1 .

For the LSTM cell, the equations are provided as

$$x'_t = \alpha_t \odot x_t \tag{8}$$

$$c_t = \sigma[W_f \cdot (x'_t, h_{t-1})] \odot c_{t-1} + \sigma[W_i \cdot (x'_t, h_{t-1})] \odot \tanh[W_g \cdot (x'_t, h_{t-1})] \tag{9}$$

$$h_t = \sigma[W_o \cdot (x'_t, h_{t-1})] \odot \tanh(c_t) \tag{10}$$

where x_t means input matrix, x'_t represents the x_t is modulated, \odot means the Hadamard product, c_t means cell state (long memory), and h_t means hidden state (short memory).

2.3.2. STA-GRU Model

The STA-GRU model is meticulously crafted for spatiotemporal data processing too. Suited for complex datasets with overlapping spatial and temporal attributes. It retains the fundamental GRU mechanisms, notably the reset and update gates, ensuring effective sequence dependency tracking. Moreover, to augment its spatial pattern comprehension, STA-GRU may integrate sophisticated elements like FC layers or attention frameworks, as shown in Figure 5.

Additionally, methods such as Grid Search and Random Search were employed for the optimization of the model's hyperparameters, to further enhance the model's performance and reliability. In the STA-GRU model, a GRU cell is used in place of the LSTM cell from STA-LSTM. The structure of the GRU cell is illustrated below as Figure 6.

The GRU cell is provided as follows:

$$u_t = \sigma(W_u \cdot [x'_t, h_{t-1}]) \tag{11}$$

$$r_t = \sigma(W_r \cdot [x'_t, h_{t-1}]) \tag{12}$$

$$h_t = (1 - u_t) \odot h_{t-1} + u_t \odot \tanh(W \cdot [x'_t, r_t \odot h_{t-1}]) \tag{13}$$

where the result of the update gate is u_t , the result of the reset gate is r_t , and W_u , W_r and W are the weights of the update gate, reset gate and cell state, respectively. Then the input

data for each GRU cell comprises a 1×14 vector. This vector comprises water level and discharge for stations 01, 13, 18, 25, 29, 31 and precipitation at stations 18 and 25 values.

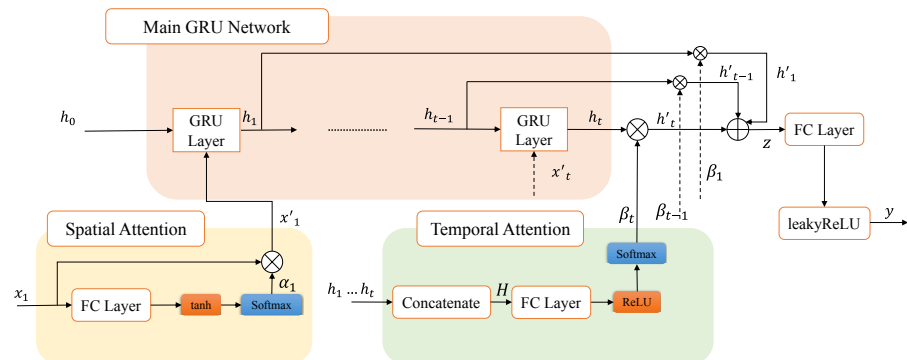


Figure 5. STA-GRU model structure.

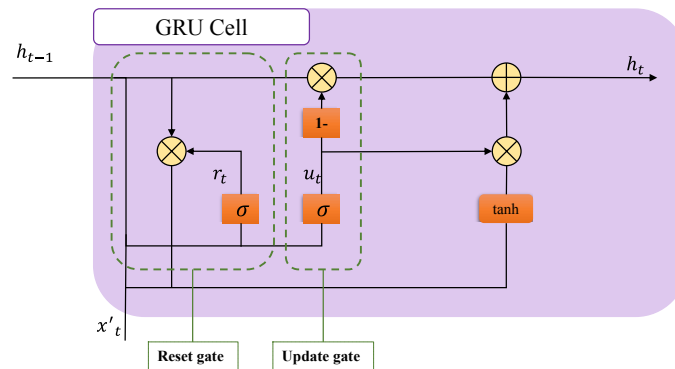


Figure 6. GRU cell structure.

2.4. Performance Metrics

Evaluating the performance of flood prediction models involves a crucial decision in selecting the appropriate metrics. The combination of Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-square provides a comprehensive model assessment [75].

1. RMSE emphasizes large errors by squaring the differences, making the model sensitive to significant deviations in predicting flood quantities, thus ensuring robustness and accuracy. The formula of RMSE is given as

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \tag{14}$$

where O_i is the observation value, P_i is the prediction value, and n is the number of observations/predictions.

2. MAE assigns equal weight to each error, aiding in evaluating the model’s average predictive precision in general scenarios. The MAE can be represented by the following equation

$$MAE = \frac{\sum_{i=1}^n |O_i - P_i|}{n} \tag{15}$$

3. R-square offers a measure of how well the model explains the variability in flood flow, where higher R-square values indicate better capability to account for observed fluctuations, enhancing the model’s interpretability and reliability. R-square is defined by

$$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \tilde{O}_i)^2} \tag{16}$$

where \tilde{O}_i is the average of the observation value.

Considering these three metrics collectively, they provide a wealth of information from different angles—RMSE and MAE focus on error magnitude and mean accuracy, while R-square emphasizes model explanatory power—resulting in a well-rounded evaluation that helps accurately gauge and refine the performance of flood prediction models.

3. Results and Discussion

3.1. Description of Validation Case

During the model training, the test data comprise 20% of the total data, while the training data constitute 80% of the total data and the validation data make up 10% of the training data. The batch size is set to 128, the learning rate is from 0.001 to 0.0001 and the number of epochs is 200. The training time for the STA-GRU model averages about 3 s per epoch, in contrast to the STA-LSTM model, which takes roughly 5 s per epoch. This result unequivocally demonstrates a notable enhancement in computational efficiency when employing GRU models. As shown in Figure 7, a loss trajectory was employed to scrutinize the effects of two data processing strategies on the performances of different models. The graphs distinctly illustrate that before the ‘lag time’ preprocessing, the training loss curve of the STA-GRU model has better numerical stability than the STA-LSTM model, and the validation loss curve of the STA-GRU model has greater accuracy and fitting. Moreover, after undergoing ‘lag time’ preprocessing, both the training loss and validation loss curves exhibit superior performance in comparison to data not subjected to this preprocessing. Both training loss and validation loss are pivotal metrics in evaluating the proficiency of machine learning models, with lower loss values indicating enhanced predictive accuracy and generalization capabilities.

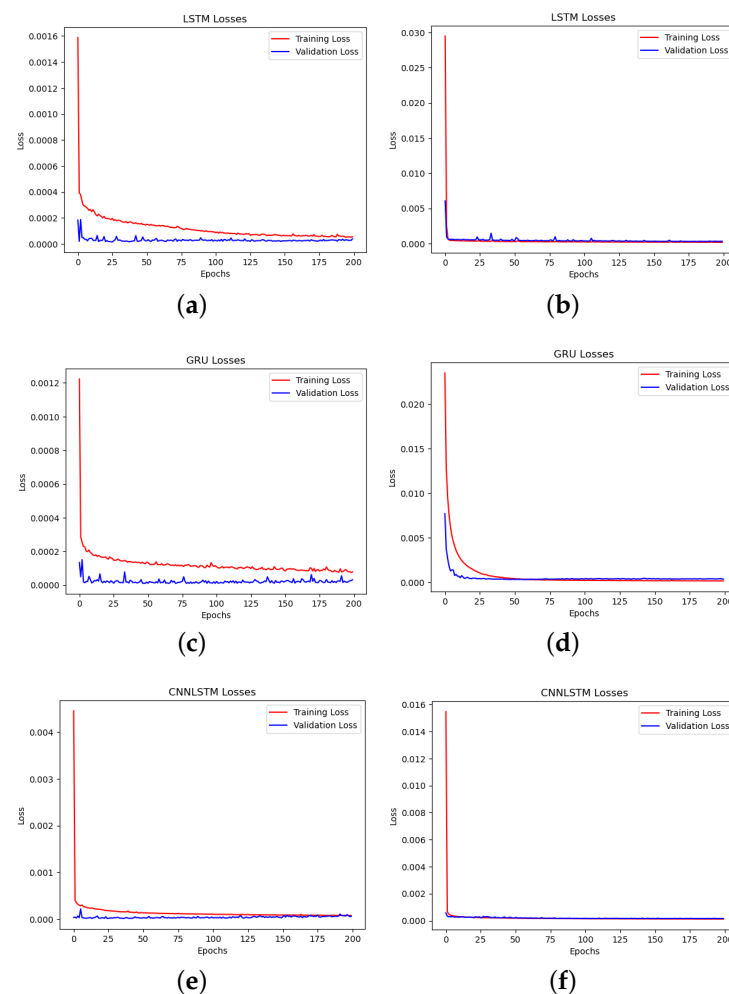


Figure 7. Cont.

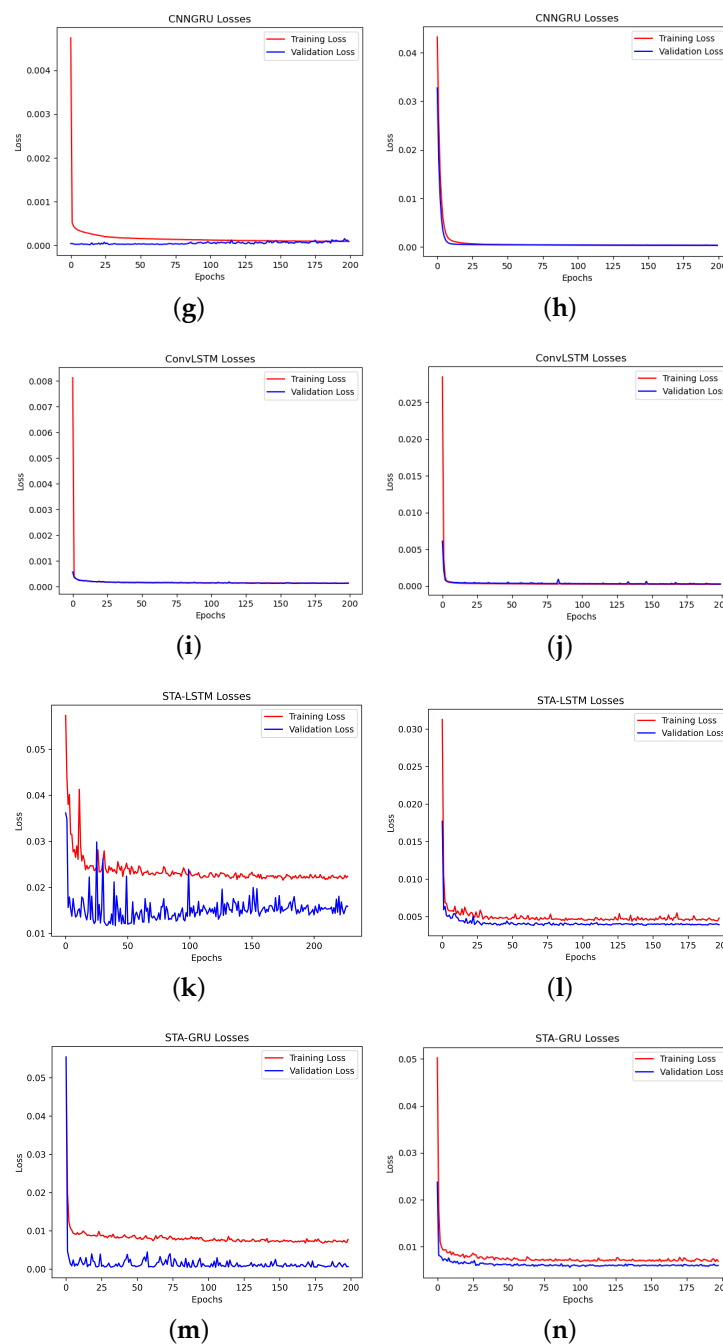


Figure 7. Performance of the LSTM, GRU, CNNLSTM, CNNGRU, ConvLSTM, STA-LSTM, and STA-GRU models during training and validation error. (a,c,e,g,i,k,m) Before lag time; (b,d,f,h,j,l,n) After lag time.

The ‘lag time’ preprocessing might have captured spatiotemporal dependencies or other salient features within the data, enabling the model to learn the data’s inherent structures and patterns more effectively. In contrast, data not subjected to this preprocessing may lack these essential cues, leading to challenges in model fitting and consequently manifesting higher loss values during both training and validation phases. In summation, ‘lag time’ preprocessing evidently furnishes the model with a richer and more accurate data representation, thereby bolstering its fitting and generalization prowess.

3.2. Discussion of Results

In this study, we have employed a range of advanced sequential models for the task of time series forecasting. These models include LSTM, GRU, CNNLSTM, CNNGRU, ConvLSTM, STA-LSTM, and STA-GRU, and their performance metrics have been evaluated across various prediction time intervals, encompassing Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R-square).

In light of the collated results, a discernible trend emerges irrespective of whether the data undergo lag time preprocessing. As the prediction time interval lengthens, both RMSE and MAE values manifest a progressive increase, whereas the R-square values exhibit a decline. This pattern accentuates that, over extended prediction time horizons, the predictive efficacy of models tends to wane, leading to a broadening of prediction errors.

From a holistic perspective, as seen Table 3, the STA-GRU model and STA-LSTM model consistently excel in longer-term forecasts, with the majority of their R^2 values comfortably surpassing the 0.8 benchmark. For initial performance, STA-LSTM, ConvLSTM, and STA-GRU emerge as front runners for the 6-h forecast, all boasting an impressive R^2 value of 0.93–0.94. This suggests that these models capture the immediate temporal dependencies in the data with remarkable precision. Then at the 12-h midpoint, the STA-GRU, STA-LSTM, and ConvLSTM continue to dominate with commendable R^2 values of 0.81, 0.81, and 0.80, respectively. This accentuates their stability in medium-term forecasting. Extending the forecast to 24 h, STA-GRU maintains its supremacy with an R^2 of 0.62—the highest value of the evaluated models. On the other end of the spectrum, CNNLSTM lags, registering the lowest R^2 of 0.54. This positions the STA-GRU model as a relatively stable long-term forecaster.

Table 3. The proposed models' performance statistics before the lag time.

Hourly	Algorithm	RMSE	MAE	R^2
6	LSTM	0.0623	0.0309	0.9001
6	GRU	0.0589	0.0278	0.9107
6	CNNLSTM	0.0620	0.0292	0.9012
6	CNNGRU	0.0573	0.0275	0.9158
6	ConvLSTM	0.0513	0.0243	0.9323
6	STA-LSTM	0.0503	0.0229	0.9385
6	STA-GRU	0.0464	0.0228	0.9445
12	LSTM	0.0939	0.0435	0.7734
12	GRU	0.0911	0.0431	0.7865
12	CNNLSTM	0.0954	0.0481	0.7660
12	CNNGRU	0.0931	0.0433	0.7780
12	ConvLSTM	0.0864	0.0408	0.8080
12	STA-LSTM	0.0833	0.0407	0.8106
12	STA-GRU	0.0832	0.0405	0.8125
24	LSTM	0.1332	0.0757	0.5461
24	GRU	0.1255	0.0658	0.5971
24	CNNLSTM	0.1322	0.0673	0.5528
24	CNNGRU	0.1262	0.0652	0.5925
24	ConvLSTM	0.1241	0.0641	0.6061
24	STA-LSTM	0.1227	0.0631	0.6143
24	STA-GRU	0.1220	0.0625	0.6181

Moreover, the most pronounced dip in performance is observed in the models between the 12th and 24th-h forecasts—a decline of 0.2. This might hint at challenges the models face in accommodating certain temporal shifts or cyclic patterns beyond the 6-hour mark. As we progress through the forecast horizon, certain models witness a steeper attrition in performance. Case in point, the CNNLSTM's R^2 value plummets from 0.90 (at 6 h) to 0.55 (at 24 h)—a descent markedly steeper than the STA-GRU's slide from 0.93 to 0.61. Typically, a slower decline in a model's R^2 over the forecast period is emblematic of its generalization prowess and stability. Gauging from the data at hand, STA-GRU and STA-LSTM emerge as frontrunners in this regard. In addition, The foundational GRU and LSTM models show perceptible performance disparities compared to their advanced counterparts like STA-

GRU and STA-LSTM. This indicates the tangible benefits brought about by sophisticated features such as spatial attention.

In flood prediction models, based on our spatiotemporal dataset, accounting for the lag time between upstream hydrological and precipitation stations and downstream target stations is crucial to further enhance the model's long-term predictive accuracy. This stems from the intrinsic spatiotemporal dynamics of hydrological processes, wherein a clear time lag exists between precipitation events and subsequent river level elevations. By accounting for this lag time between upstream and downstream stations, the model can achieve a more precise data alignment, bolster the capture of causal relationships, factor in the influences of terrain and soil conditions, and encapsulate the dynamic characteristics of flood events. Furthermore, the inclusion of lag time furnishes the model with enhanced spatiotemporal sequence features, facilitating a deeper contextual understanding and thereby significantly enhancing prediction accuracy, as shown in Table 4. Therefore, after undergoing lag time preprocessing, it is evident that as the prediction horizon extends from 6 h to 12 h, and further to 24 h, the R-value of each predictive model decreases by approximately 0.1 less than that of models without lag time preprocessing.

Table 4. The proposed models' performance statistics after the lag time.

Hourly	Algorithm	RMSE	MAE	R ²
6	LSTM	0.0456	0.0243	0.9466
6	GRU	0.0520	0.0290	0.9304
6	CNNLSTM	0.0482	0.0299	0.9402
6	CNNGRU	0.0499	0.0272	0.9359
6	ConvLSTM	0.0405	0.0213	0.9578
6	STA-LSTM	0.0399	0.0203	0.9590
6	STA-GRU	0.0382	0.0199	0.9646
12	LSTM	0.0644	0.0353	0.8935
12	GRU	0.0643	0.0351	0.8936
12	CNNLSTM	0.0677	0.0372	0.8821
12	CNNGRU	0.0652	0.0324	0.8907
12	ConvLSTM	0.0631	0.0332	0.8974
12	STA-LSTM	0.0553	0.0318	0.9214
12	STA-GRU	0.0526	0.0291	0.9288
24	LSTM	0.1165	0.0600	0.6525
24	GRU	0.1150	0.0607	0.6637
24	CNNLSTM	0.1178	0.0575	0.6453
24	CNNGRU	0.1154	0.0569	0.6592
24	ConvLSTM	0.1134	0.0550	0.6713
24	STA-LSTM	0.1052	0.0548	0.7164
24	STA-GRU	0.1039	0.0534	0.7232

Upon comparing the performance metrics of various flood prediction models with and without lag time preprocessing, it becomes evident that preprocessing substantially bolsters the efficacy of all models. Our initial findings, prior to the implementation of lag time preprocessing, indicated that our predictive model exhibited RMSE and MAE values comparable to those reported by Liu et al. (2023), Dehghani et al. (2023), and Ding et al. (2020) [49,53,55]. However, a salient discovery of this study is the significant enhancement in prediction accuracy and model performance observed after applying lag time preprocessing to spatiotemporal data, even when operating under identical forecast durations. Furthermore, in a bid to augment computational efficiency and extend the prediction horizon to 24 h, our developed STA-GRU model demonstrated superior performance compared to existing models documented in the literature. These findings not only affirm the pivotal role of lag time preprocessing in improving the precision of spatiotemporal data predictions but also highlight the potential of the STA-GRU model in flood forecasts.

In our research, we employed a bar chart to juxtapose the effects of two data processing methodologies on the R² values of our models. In Figure 8, the orange bars represent data subjected to a 'lag time' preprocessing, while the blue bars symbolize data that were not

processed in this manner. The R^2 , or the coefficient of determination, is a statistical metric used to quantify the goodness of fit of a regression model, with its value ranging between 0 and 1. A value closer to 1 indicates superior predictive prowess of the model. As the prediction timeline extended, the R^2 values derived from the 'lag time' preprocessed data consistently surpassed those from the non-preprocessed data, with this disparity widening over time. This suggests that 'lag time' preprocessing not only enhances the overall goodness of fit of the model but also accentuates its advantages in long-term forecasting scenarios. This offers robust theoretical support for future data preprocessing endeavors, signifying that in certain applications, 'lag time' preprocessing could be a pivotal step, especially when extended forecasting is requisite.

In summary, the STA-GRU model exhibited superior performance over other models in each predictive time frame on data without "lag time" preprocessing. However, following the "lag time" preprocessing of the data, the STA-GRU model not only sustained its comparative advantage but also achieved a higher performance with reduced forecast error statistics. This demonstrates the STA-GRU model's outstanding adaptability and efficiency when dealing with spatiotemporal data collected from a network of real-time hydrometric stations for rapid response flood early warning applications.

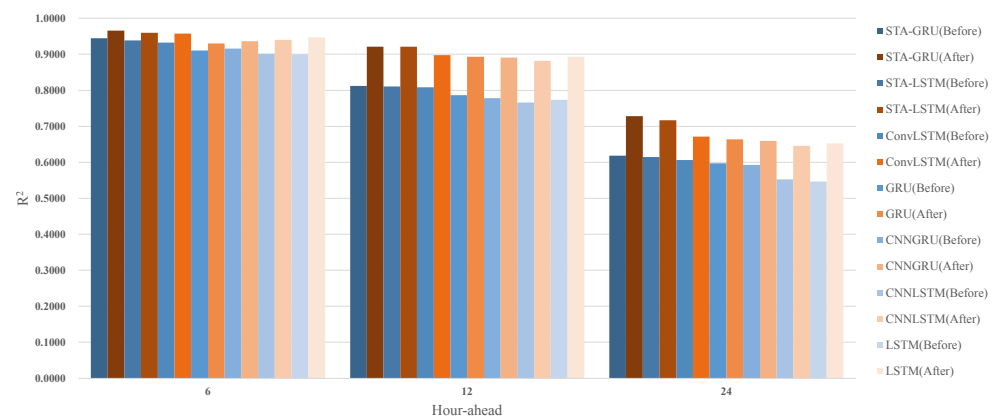


Figure 8. Comparing the R-square of each model before and after handling lag time.

4. Conclusions

The realm of flood forecasting has greatly benefited from the integration of deep learning techniques, which have emerged as transformative tools for enhancing prediction accuracy. In this context, our study examined the performance of several deep learning models to improve flood forecasting. These models included Long Short-Term Memory (LSTM) and its spatial derivatives such as CNNLSTM, ConvLSTM, and STA-LSTM, as well as the Gated Recurrent Unit (GRU) and its associated models like CNNGRU and STA-GRU. These models were methodically compared, analyzing their capacities to process complex hydrological data and forecast floods. Given the geographic and climatic influences on floods, a comprehensive approach to data analysis and modeling is essential. By harnessing spatial information and integrating it with time series data, we determined a more holistic flood prediction model. Among our key results, it is found that models incorporating the spatiotemporal attention mechanism, like the STA-LSTM and STA-GRU, exhibit an enhanced ability to manage long-term dependencies. Particularly, the STA-GRU model improves computational efficiency while maintaining prediction performance at a level not lower than that of the STA-LSTM model. Elevating computational efficiency is crucial in the context of flood forecasting, as it allows the predictive system to rapidly process and analyze extensive datasets, thereby enabling swift responses and real-time surveillance of flood incidents. This not only contributes to the prompt issuance of warnings but also facilitates the contemporaneous updating of the predictive models, enhancing the accuracy of the alerts. Furthermore, when the datasets are preprocessed with lag time, the R^2 value of the STA-GRU model increases from 0.6181 to 0.7232, RMSE decreases from 0.1220 to 0.1039, and MAE reduces from 0.0625 to 0.0534. These results indicate that the prediction

performance of the STA-GRU model is enhanced. The STA-GRU model and STA-LSTM model prioritize significant data and offer more accurate flood predictions, capturing intricate spatiotemporal patterns and making them potential frontrunners in the quest to advance flood forecasting systems.

Author Contributions: Conceptualization: Y.Z.; methodology: Y.Z. and Z.Z.; software: J.V.G.T., Y.Z. and Z.Z.; validation: Y.Z.; formal analysis: Y.Z.; investigation: Y.Z.; resources: Y.Z. and B.G.; data curation: Y.Z.; writing—original draft preparation: Y.Z., B.G., J.V.G.T. and S.X.Y.; writing—review and editing: B.G., J.V.G.T. and S.X.Y.; supervision: B.G. and S.X.Y.; project administration: B.G., S.X.Y. and J.V.G.T.; funding acquisition: B.G. and J.V.G.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) Alliance Grant #401643.

Data Availability Statement: The datasets used in this study are in public domain and are available for download from the Water Survey of Canada (for the Discharge and Water level data) and Environment Canada (for Precipitation data).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Jalili Pirani, F.; Najafi, M.R. Multivariate analysis of compound flood hazard across Canada's Atlantic, Pacific and Great Lakes coastal areas. *Earths Future* **2022**, *10*, e2022EF002655. [[CrossRef](#)]
- Lin, H.; Mo, R.; Vitart, F.; Stan, C. Eastern Canada flooding 2017 and its subseasonal predictions. *Atmosphere-Ocean* **2019**, *57*, 195–207. [[CrossRef](#)]
- Ebtehaj, I.; Bonakdari, H. A reliable hybrid outlier robust non-tuned rapid machine learning model for multi-step ahead flood forecasting in Quebec, Canada. *J. Hydrol.* **2022**, *614*, 128592. [[CrossRef](#)]
- Zadeh, S.M.; Burn, D.H.; O'Brien, N. Detection of trends in flood magnitude and frequency in Canada. *J. Hydrol. Reg. Stud.* **2020**, *28*, 100673. [[CrossRef](#)]
- Taraky, Y.M.; Liu, Y.; McBean, E.; Daggupati, P.; Gharabaghi, B. Flood risk management with transboundary conflict and cooperation dynamics in the Kabul River Basin. *Water* **2021**, *13*, 1513. [[CrossRef](#)]
- Burn, D.H.; Whitfield, P.H. Changes in floods and flood regimes in Canada. *Can. Water Resour. J./Revue Can. Resour. Hydriques* **2016**, *41*, 139–150. [[CrossRef](#)]
- Saurav, K.; Shrestha, S.; Ninsawat, S.; Chonwattana, S. Predicting flood events in Kathmandu Metropolitan City under climate change and urbanisation. *J. Environ. Manag.* **2021**, *281*, 111894.
- Lin, J.; Zhang, W.; Wen, Y.; Qiu, S. Evaluating the association between morphological characteristics of urban land and pluvial floods using machine learning methods. *Sustain. Cities Soc.* **2023**, *99*, 104891. [[CrossRef](#)]
- Ward, P.J.; de Ruyter, M.C.; Mård, J.; Schröter, K.; Van Loon, A.; Veldkamp, T.; von Uexkull, N.; Wanders, N.; AghaKouchak, A.; Arnbjerg-Nielsen, K.; et al. The need to integrate flood and drought disaster risk reduction strategies. *Water Secur.* **2020**, *11*, 100070. [[CrossRef](#)]
- Nguyen, H.D.; Fox, D.; Dang, D.K.; Pham, L.T.; Viet Du, Q.V.; Nguyen, T.H.T.; Dang, T.N.; Tran, V.T.; Vu, P.L.; Nguyen, Q.H.; et al. Predicting future urban flood risk using land change and hydraulic modeling in a river watershed in the central Province of Vietnam. *Remote. Sens.* **2021**, *13*, 262. [[CrossRef](#)]
- Kumar, V.; Azamathulla, H.M.; Sharma, K.V.; Mehta, D.J.; Maharaj, K.T. The state of the art in deep learning applications, challenges, and future prospects: A comprehensive review of flood forecasting and management. *Sustainability* **2023**, *15*, 10543. [[CrossRef](#)]
- Bubeck, P.; Otto, A.; Weichselgartner, J. Societal impacts of flood hazards. In *Oxford Research Encyclopedia of Natural Hazard Science*; Oxford University Press: Oxford, UK, 2017.
- Shah, A.A.; Ajiang, C.; Gong, Z.; Khan, N.A.; Ali, M.; Ahmad, M.; Abbas, A.; Shahid, A. Reconnoitering school children vulnerability and its determinants: Evidence from flood disaster-hit rural communities of Pakistan. *Int. J. Disaster Risk Reduct.* **2022**, *70*, 102735. [[CrossRef](#)]
- Jonkman, S.N. Global perspectives on loss of human life caused by floods. *Nat. Hazards* **2005**, *34*, 151–175. [[CrossRef](#)]
- Emerton, R.E.; Stephens, E.M.; Pappenberger, F.; Pagano, T.C.; Weerts, A.H.; Wood, A.W.; Salamon, P.; Brown, J.D.; Hjerdt, N.; Donnelly, C.; et al. Continental and global scale flood forecasting systems. *Wiley Interdiscip. Rev. Water* **2016**, *3*, 391–418. [[CrossRef](#)]
- Cloke, H.L.; Pappenberger, F. Ensemble flood forecasting: A review. *J. Hydrol.* **2009**, *375*, 613–626. [[CrossRef](#)]
- Moore, R. Real-time flood forecasting systems: Perspectives and prospects. *Floods Landslides Integr. Risk Assess.* **1999**, 147–189.
- Acharya, A.; Prakash, A. When the river talks to its people: Local knowledge-based flood forecasting in Gandak River basin, India. *Environ. Dev.* **2019**, *31*, 55–67. [[CrossRef](#)]

19. Kaur, B.; Szentimrey, Z.; Binns, A.D.; McBean, E.A.; Gharabaghi, B. Urban flood susceptibility mapping using supervised regression and machine learning models in Toronto, Canada. In Proceedings of the AGU Fall Meeting Abstracts, Online, 17 December 2020; Volume 2020, p. NH012-07.
20. Borga, M.; Anagnostou, E.; Blöschl, G.; Creutin, J.D. Flash flood forecasting, warning and risk management: The HYDRATE project. *Environ. Sci. Policy* **2011**, *14*, 834–844. [[CrossRef](#)]
21. Nauman, C.; Anderson, E.; Coughlan de Perez, E.; Kruczkiewicz, A.; McClain, S.; Markert, A.; Griffin, R.; Suarez, P. Perspectives on flood forecast-based early action and opportunities for Earth observations. *J. Appl. Remote. Sens.* **2021**, *15*, 032002. [[CrossRef](#)]
22. Lawford, R.; Prowse, T.; Hogg, W.; Warkentin, A.; Pilon, P. Hydrometeorological aspects of flood hazards in Canada. *Atmosphere-Ocean* **1995**, *33*, 303–328. [[CrossRef](#)]
23. Alfieri, L.; Salamon, P.; Pappenberger, F.; Wetterhall, F.; Thielen, J. Operational early warning systems for water-related hazards in Europe. *Environ. Sci. Policy* **2012**, *21*, 35–49. [[CrossRef](#)]
24. Merz, B.; Kuhlicke, C.; Kunz, M.; Pittore, M.; Babeyko, A.; Bresch, D.N.; Domeisen, D.I.; Feser, F.; Koszalka, I.; Kreibich, H.; et al. Impact forecasting to support emergency management of natural hazards. *Rev. Geophys.* **2020**, *58*, e2020RG000704. [[CrossRef](#)]
25. Jain, S.K.; Mani, P.; Jain, S.K.; Prakash, P.; Singh, V.P.; Tullios, D.; Kumar, S.; Agarwal, S.; Dimri, A. A Brief review of flood forecasting techniques and their applications. *Int. J. River Basin Manag.* **2018**, *16*, 329–344. [[CrossRef](#)]
26. Kim, G.; Barros, A.P. Quantitative flood forecasting using multisensor data and neural networks. *J. Hydrol.* **2001**, *246*, 45–62. [[CrossRef](#)]
27. Brocca, L.; Melone, F.; Moramarco, T. Distributed rainfall-runoff modelling for flood frequency estimation and flood forecasting. *Hydrol. Process.* **2011**, *25*, 2801–2813. [[CrossRef](#)]
28. Toth, E.; Brath, A.; Montanari, A. Comparison of short-term rainfall prediction models for real-time flood forecasting. *J. Hydrol.* **2000**, *239*, 132–147. [[CrossRef](#)]
29. Hapuarachchi, H.; Wang, Q.; Pagano, T. A review of advances in flash flood forecasting. *Hydrol. Process.* **2011**, *25*, 2771–2784. [[CrossRef](#)]
30. Yin, D.; Xue, Z.G.; Bao, D.; RafieeiNasab, A.; Huang, Y.; Morales, M.; Warner, J.C. Understanding the role of initial soil moisture and precipitation magnitude in flood forecast using a hydrometeorological modelling system. *Hydrol. Process.* **2022**, *36*, e14710. [[CrossRef](#)]
31. Li, Y.; Grimaldi, S.; Walker, J.P.; Pauwels, V.R. Application of remote sensing data to constrain operational rainfall-driven flood forecasting: A review. *Remote. Sens.* **2016**, *8*, 456. [[CrossRef](#)]
32. Piadeh, F.; Behzadian, K.; Alani, A.M. A critical review of real-time modelling of flood forecasting in urban drainage systems. *J. Hydrol.* **2022**, *607*, 127476. [[CrossRef](#)]
33. Taraky, Y.M.; Liu, Y.; Gharabaghi, B.; McBean, E.; Daggupati, P.; Shrestha, N.K. Influence of headwater reservoirs on climate change impacts and flood frequency in the Kabul River Basin. *Can. J. Civ. Eng.* **2022**, *49*, 1300–1309. [[CrossRef](#)]
34. Oruh, J.; Viriri, S.; Adegun, A. Long short-term memory recurrent neural network for automatic speech recognition. *IEEE Access* **2022**, *10*, 30069–30079. [[CrossRef](#)]
35. Li, W.; Kiaghadi, A.; Dawson, C. Exploring the best sequence LSTM modeling architecture for flood prediction. *Neural Comput. Appl.* **2021**, *33*, 5571–5580. [[CrossRef](#)]
36. Kumar, A.; Bhatia, A.; Kashyap, A.; Kumar, M. LSTM Network: A Deep Learning Approach and Applications. In *Advanced Applications of NLP and Deep Learning in Social Media Data*; IGI Global: Hershey, PA, USA, 2023; pp. 130–150.
37. Iparraguirre-Villanueva, O.; Guevara-Ponce, V.; Ruiz-Alvarado, D.; Beltozar-Clemente, S.; Sierra-Liñan, F.; Zapata-Paulini, J.; Cabanillas-Carbonell, M. Text prediction recurrent neural networks using long short-term memory-dropout. *Indones. J. Electr. Eng. Comput. Sci.* **2023**, *29*, 1758–1768. [[CrossRef](#)]
38. Hayder, I.M.; Al-Amiedy, T.A.; Ghaban, W.; Saeed, F.; Nasser, M.; Al-Ali, G.A.; Younis, H.A. An Intelligent Early Flood Forecasting and Prediction Leveraging Machine and Deep Learning Algorithms with Advanced Alert System. *Processes* **2023**, *11*, 481. [[CrossRef](#)]
39. Granata, F.; Di Nunno, F. Neuroforecasting of daily streamflows in the UK for short-and medium-term horizons: A novel insight. *J. Hydrol.* **2023**, *624*, 129888. [[CrossRef](#)]
40. Le, X.H.; Ho, H.V.; Lee, G.; Jung, S. Application of long short-term memory (LSTM) neural network for flood forecasting. *Water* **2019**, *11*, 1387. [[CrossRef](#)]
41. Boopathi, S. Deep Learning Techniques Applied for Automatic Sentence Generation. In *Promoting Diversity, Equity, and Inclusion in Language Learning Environments*; IGI Global: Hershey, PA, USA, 2023; pp. 255–273.
42. Tabrizi, S.E.; Xiao, K.; Thé, J.V.G.; Saad, M.; Farghaly, H.; Yang, S.X.; Gharabaghi, B. Hourly road pavement surface temperature forecasting using deep learning models. *J. Hydrol.* **2021**, *603*, 126877. [[CrossRef](#)]
43. Li, J.; Yuan, X. Daily Streamflow Forecasts Based on Cascade Long Short-Term Memory (LSTM) Model over the Yangtze River Basin. *Water* **2023**, *15*, 1019. [[CrossRef](#)]
44. Zou, Y.; Wang, J.; Lei, P.; Li, Y. A novel multi-step ahead forecasting model for flood based on time residual LSTM. *J. Hydrol.* **2023**, *620*, 129521. [[CrossRef](#)]
45. Jia, P.; Cao, N.; Yang, S. Real-time hourly ozone prediction system for Yangtze River Delta area using attention based on a sequence to sequence model. *Atmos. Environ.* **2021**, *244*, 117917. [[CrossRef](#)]

46. Zhang, Y.; Gu, Z.; Thé, J.V.G.; Yang, S.X.; Gharabaghi, B. The Discharge Forecasting of Multiple Monitoring Station for Humber River by Hybrid LSTM Models. *Water* **2022**, *14*, 1794. [[CrossRef](#)]
47. Moishin, M.; Deo, R.C.; Prasad, R.; Raj, N.; Abdulla, S. Designing deep-based learning flood forecast model with ConvLSTM hybrid algorithm. *IEEE Access* **2021**, *9*, 50982–50993. [[CrossRef](#)]
48. Yao, Z.; Wang, Z.; Wang, D.; Wu, J.; Chen, L. 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.* **2023**, *625*, 129977. [[CrossRef](#)]
49. Ding, Y.; Zhu, Y.; Feng, J.; Zhang, P.; Cheng, Z. Interpretable spatiotemporal attention LSTM model for flood forecasting. *Neurocomputing* **2020**, *403*, 348–359. [[CrossRef](#)]
50. Li, P.; Zhang, J.; Krebs, P. Prediction of flow based on a CNN-LSTM combined deep learning approach. *Water* **2022**, *14*, 993. [[CrossRef](#)]
51. Khorram, S.; Jehbez, N. A Hybrid CNN-LSTM Approach for Monthly Reservoir Inflow Forecasting. *Water Resour. Manag.* **2023**, *37*, 4097–4121. [[CrossRef](#)]
52. Yang, Y.; Xiong, Q.; Wu, C.; Zou, Q.; Yu, Y.; Yi, H.; Gao, M. A study on water quality prediction by a hybrid CNN-LSTM model with attention mechanism. *Environ. Sci. Pollut. Res.* **2021**, *28*, 55129–55139. [[CrossRef](#)]
53. Dehghani, A.; Moazam, H.M.Z.H.; Mortazavizadeh, F.; Ranjbar, V.; Mirzaei, M.; Mortezaei, S.; Ng, J.L.; Dehghani, A. Comparative evaluation of LSTM, CNN, and ConvLSTM for hourly short-term streamflow forecasting using deep learning approaches. *Ecol. Inform.* **2023**, *75*, 102119. [[CrossRef](#)]
54. Wu, Y.; Ding, Y.; Zhu, Y.; Feng, J.; Wang, S. Complexity to forecast flood: Problem definition and spatiotemporal attention LSTM solution. *Complexity* **2020**, *2020*, 7670382. [[CrossRef](#)]
55. Liu, Y.; Yang, Y.; Chin, R.J.; Wang, C.; Wang, C. Long Short-Term Memory (LSTM) Based Model for Flood Forecasting in Xiangjiang River. *KSCE J. Civ. Eng.* **2023**, *27*, 5030–5040. [[CrossRef](#)]
56. Shewalkar, A.; Nyavanandi, D.; Ludwig, S.A. Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU. *J. Artif. Intell. Soft Comput. Res.* **2019**, *9*, 235–245. [[CrossRef](#)]
57. Gao, S.; Huang, Y.; Zhang, S.; Han, J.; Wang, G.; Zhang, M.; Lin, Q. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. *J. Hydrol.* **2020**, *589*, 125188. [[CrossRef](#)]
58. Zhao, Z.; Yun, S.; Jia, L.; Guo, J.; Meng, Y.; He, N.; Li, X.; Shi, J.; Yang, L. Hybrid VMD-CNN-GRU-based model for short-term forecasting of wind power considering spatiotemporal features. *Eng. Appl. Artif. Intell.* **2023**, *121*, 105982. [[CrossRef](#)]
59. Cho, M.; Kim, C.; Jung, K.; Jung, H. Water level prediction model applying a long short-term memory (lstm)-gated recurrent unit (gru) method for flood prediction. *Water* **2022**, *14*, 2221. [[CrossRef](#)]
60. Pan, M.; Zhou, H.; Cao, J.; Liu, Y.; Hao, J.; Li, S.; Chen, C.H. Water level prediction model based on GRU and CNN. *IEEE Access* **2020**, *8*, 60090–60100. [[CrossRef](#)]
61. Hua, G.; Wang, S.; Xiao, M.; Hu, S. Research on the Uplift Pressure Prediction of Concrete Dams Based on the CNN-GRU Model. *Water* **2023**, *15*, 319. [[CrossRef](#)]
62. Hood, M.J.; Clausen, J.C.; Warner, G.S. Comparison of Stormwater lag times for low impact and traditional residential development 1. *JAWRA J. Am. Water Resour. Assoc.* **2007**, *43*, 1036–1046. [[CrossRef](#)]
63. Gericke, O.; Smithers, J. Direct estimation of catchment response time parameters in medium to large catchments using observed streamflow data. *Hydrol. Process.* **2017**, *31*, 1125–1143. [[CrossRef](#)]
64. Berne, A.; Delrieu, G.; Creutin, J.D.; Obled, C. Temporal and spatial resolution of rainfall measurements required for urban hydrology. *J. Hydrol.* **2004**, *299*, 166–179. [[CrossRef](#)]
65. Perdikaris, J.; Gharabaghi, B.; Rudra, R. Reference time of concentration estimation for ungauged catchments. *Earth Sci. Res* **2018**, *7*, 58–73. [[CrossRef](#)]
66. Langridge, M.; Gharabaghi, B.; McBean, E.; Bonakdari, H.; Walton, R. Understanding the dynamic nature of Time-to-Peak in UK streams. *J. Hydrol.* **2020**, *583*, 124630. [[CrossRef](#)]
67. Seyam, M.; Othman, F. The influence of accurate lag time estimation on the performance of stream flow data-driven based models. *Water Resour. Manag.* **2014**, *28*, 2583–2597. [[CrossRef](#)]
68. Adeyi, G.; Adigun, A.; Onyeocha, N.; Okeke, O. Unit hydrograph: Concepts, estimation methods and applications in hydrological sciences. *Int. J. Eng. Sci. Comput.* **2020**, *10*, 26211–26217.
69. Barbero, G.; Costabile, P.; Costanzo, C.; Ferraro, D.; Petaccia, G. 2D hydrodynamic approach supporting evaluations of hydrological response in small watersheds: Implications for lag time estimation. *J. Hydrol.* **2022**, *610*, 127870. [[CrossRef](#)]
70. Oliveira Santos, V.; Costa Rocha, P.A.; Scott, J.; Thé, J.V.G.; Gharabaghi, B. A New Graph-Based Deep Learning Model to Predict Flooding with Validation on a Case Study on the Humber River. *Water* **2023**, *15*, 1827. [[CrossRef](#)]
71. Elkurdy, M.; Binns, A.D.; Bonakdari, H.; Gharabaghi, B.; McBean, E. Early detection of riverine flooding events using the group method of data handling for the Bow River, Alberta, Canada. *Int. J. River Basin Manag.* **2022**, *20*, 533–544. [[CrossRef](#)]
72. Langridge, M.; McBean, E.; Bonakdari, H.; Gharabaghi, B. A dynamic prediction model for time-to-peak. *Hydrol. Process.* **2021**, *35*, e14032. [[CrossRef](#)]
73. Soltani, K.; Ebtehaj, I.; Amiri, A.; Azari, A.; Gharabaghi, B.; Bonakdari, H. Mapping the spatial and temporal variability of flood susceptibility using remotely sensed normalized difference vegetation index and the forecasted changes in the future. *Sci. Total. Environ.* **2021**, *770*, 145288. [[CrossRef](#)]

74. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the 30th International Conference on International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; Volume 30, p. 3.
75. Zhang, Y.; Pan, D.; Van Griensven, J.; Yang, S.X.; Gharabaghi, B. Intelligent flood forecasting and warning: A survey. *Intell. Robot.* **2023**, *3*, 190–212. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.