

Spatiotemporal Prediction of Radar Echoes Based on ConvLSTM and Multisource Data

Mingyue Lu ^{1,*}, Yuchen Li ¹, Manzhu Yu ² , Qian Zhang ³, Yadong Zhang ⁴, Bin Liu ⁵ and Menglong Wang ¹

¹ Geographic Science College, Nanjing University of Information Science & Technology, Nanjing 210044, China

² Department of Geography, The Pennsylvania State University, State College, PA 16802, USA

³ School of Management Engineering, Xi'an University of Finance and Economics, Xi'an 710100, China

⁴ Puer Simao District Meteorological Bureau, Pu'er 665099, China

⁵ Institute of Bei-Stars Geospatial Innovations (Nanjing) Pty Ltd., Nanjing 210000, China

* Correspondence: lumingyue@nuist.edu.cn

Abstract: Accurate and timely precipitation forecasts can help people and organizations make informed decisions, plan for potential weather-related disruptions, and protect lives and property. Instead of using physics-based numerical forecasts, which can be computationally prohibitive, there has been a growing interest in using deep learning techniques for precipitation prediction in recent years due to the success of these approaches in various other fields. These deep learning approaches generally use historical composite reflectivity (CR) at the surface level to predict future time steps. However, other relevant factors related to the potential motion and vertical structure of the storm have not been considered. To address this issue, this research proposes a multisource ConvLSTM (MS-ConvLSTM) model to improve the accuracy of precipitation forecasting by incorporating multiple data sources into the prediction process. The model was trained on a dataset of radar echo features, which includes not only composite reflectivity (CR), but also echo top (ET), vertically integrated liquid (VIL) water, and radar-retrieved wind field data at different elevations. Experiment results showed that the proposed model outperformed traditional methods in terms of various evaluation metrics, such as mean absolute error (MAE), mean squared error (MSE), probability of detection (POD), false alarm rate (FAR), and critical success index (CSI).

Keywords: precipitation forecasting; deep learning; radar echo; multisource; spatiotemporal correlation



Citation: Lu, M.; Li, Y.; Yu, M.; Zhang, Q.; Zhang, Y.; Liu, B.; Wang, M. Spatiotemporal Prediction of Radar Echoes Based on ConvLSTM and Multisource Data. *Remote Sens.* **2023**, *15*, 1279. <https://doi.org/10.3390/rs15051279>

Academic Editor: Kenji Nakamura

Received: 28 December 2022

Revised: 12 February 2023

Accepted: 23 February 2023

Published: 25 February 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

Precipitation forecasting, which involves predicting the amount and type of precipitation (such as rain, snow, sleet, or hail) likely to occur in a specific location at a specific time, is vital for individuals and organizations. Accurate precipitation forecasts can help individuals and organizations make informed decisions about transportation, agriculture, construction, and other weather-related activities. Precipitation forecasts can also help governments and organizations prepare for and respond to natural disasters such as floods or landslides, which can have severe consequences for public safety and property. Based on timescales, precipitation forecasting can include nowcasting, medium-term, or long-term forecasts [1]. Among these time scales, short-term precipitation forecasting (within the next few hours) is more critical, as it can provide warning information for meteorological events such as hail, squall lines, thunderstorms, and rainstorms.

Most of the existing precipitation forecasting methods have been primarily physics-based or statistical, which leverage the intensity and movement patterns observed in radar echoes from the past time steps [2,3]. Observed by Doppler weather radar (DWR), these radar echo images reflect the atmospheric convection process, such as velocity spectrum width, radial velocity, and reflectivity, over the target radar station [4]. Physics-based numerical weather prediction (NWP) models, such as High-Resolution Rapid Refresh

(HRRR), use mathematical equations to simulate the physics of the atmosphere and are used to make forecasts of weather conditions over a range of time scales. Such physics-based models can be computationally prohibitive, especially when increasing spatial and temporal resolutions for forecasting. In addition, the prediction time period for short-term precipitation forecasting (up to a couple of hours ahead) is much shorter than the NWP model's spin-up time (~6 h), which might result in unreliable predictive outcomes. Probabilistic precipitation forecasting methods use statistical models to estimate the likelihood of different precipitation scenarios occurring but require the accurate estimation of advection-related equations, which are not robust across different spatiotemporal ranges. Ensemble forecasting approaches involve combining the output of multiple NWP models or other forecasting methods to create a composite forecast and can provide a range of possible precipitation scenarios and account for the uncertainty inherent in weather forecasting. The predictive capability of ensemble-based methods depends heavily on each involved NWP model and can be even more computationally expensive.

Deep learning approaches have recently been used for precipitation forecasting, primarily due to breakthroughs in various fields, particularly in image processing and natural language processing [5]. These approaches generally involve using neural networks to analyze radar products and make spatiotemporal predictions about future precipitation. Deep neural networks that combine convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been used to consider spatial and temporal dependence for precipitation forecasting [6]. Convolutional Long Short-Term Memory (ConvLSTM) networks, combining CNNs and long short-term memory (LSTM) networks, were used for imminent precipitation forecasting and showed an improved predictive performance by utilizing spatiotemporal features of radar echo images [7]. Similarly, a Convolutional Gated Recurrent Unit (ConvGRU) model was proposed to address the blurring effects of extrapolated radar images [8]. A Predictive Coding Network (PredNet) was proposed to introduce the skip structure and hollow convolution to enhance the predictive capability performance of short-term forecasting [9,10]. A Trajectory Gated Recurrent Unit (TrajGRU) model that combines the concepts of trajectory modeling and gated recurrent units (GRUs) was proposed to process the temporal structure of radar products and model the movements of precipitation patterns over time [11]. Pysteps is an open-source, community-driven Python library that supports various input/output file formats, and implements multiple optical flow methods and advanced random generators to generate set proximity prediction [12].

Another primary type of deep neural network used for precipitation forecasting relies on Generative Adversarial Networks (GANs), which consist of two networks: a generator network and a discriminator network. GANs are trained by having the generator network try to generate synthetic radar echo data similar to a given dataset, and the discriminator network tries to distinguish the synthetic data from the real data. A deep generative model was proposed to incorporate temporal, spatial, and sample consistency within the GAN-based model structure to address the blurry forecast effects at longer lead times. Similarly, four GAN-based models were compared in extensive experiments to validate the effectiveness of adversarial regularizations to improve their predictive capability regarding storm structure, shape, and position.

The problem of precipitation forecasting that leverages deep learning techniques and radar products can be formally described as follows. In space, P measurements of a dynamical system at a given time over a spatial region with an $M \times N$ grid can be considered as a tensor $x \in \mathbb{R}^{P \times M \times N}$. From the temporal point of view, a sequence of tensors, $x_1, x_2, \dots, x_t$, can be obtained by collecting measurements at fixed time intervals over time t . Accordingly, this nowcasting problem can be formulated as follows:

$$\hat{y}_t = \underset{y_t}{\operatorname{argmax}} p(y_t | x_{t-K+1}, \dots, x_t) \quad (1)$$

where $\hat{y}_t$ represents the predicted value, $\{x_{t-K+1}, \dots, x_t\}$ is the historical measurement sequence data of length K , and y_t is the real value obtained from historical data.

The abovementioned deep learning approaches are limited because they usually use composite reflectivity as their single data input to extrapolate into future time steps without considering vertical structures and intensities of the storm. Common challenges for precipitation forecasting were also identified by the abovementioned approaches, including the difficulty of accurately capturing storm structures, shapes, and positions in predictions and maintaining predictive accuracy at a longer lead time. Potential improvements regarding these challenges involve the integration of multiple data sources into the training process, especially including radar products that reflect both horizontal and vertical structure and intensity of the precipitation pattern.

In this study, we propose to integrate radar echo features—including not only composite reflectivity (CR), but also echo top (ET), vertically integrated liquid (VIL) water, and radar-retrieved wind field data at different elevations—into a multisource ConvLSTM (MS-ConvLSTM) model to capture the multivariate correlations and spatiotemporal movements of precipitation patterns. Specifically, we extend from the original structure of ConvLSTM to include a 3D convolution auxiliary mechanism that incorporates radar echo features. The composite reflectivity (CR) is the main feature, whereas the echo tops (ET), vertically integrated liquid (VIL) water data, and radar-retrieved wind data are auxiliary features. This auxiliary mechanism combines the characteristics of the auxiliary channel with those of the main channel, which helps to strengthen interdependence and improve predictive capability. In addition, this auxiliary mechanism integrates long-term channel information with short-term spatiotemporal information, generating enhanced characteristics for the input sequence.

The following sections of this article are arranged in the following structure. Section 2 describes the experimental design and the MS-ConvLSTM model in detail. Section 3 describes the study area and experimental data and presents the experimental results used to evaluate the model. Section 4 provides a conclusion and charts a way forward.

2. Materials and Methods

2.1. ConvLSTM Structure

ConvLSTM is a variant of the long short-term memory (LSTM) network that uses convolutional layers instead of fully connected layers in the input-to-state and state-to-state transitions [7]. The equations for a ConvLSTM cell can be written as follows:

$$\left\{ \begin{array}{l} \text{Input gate : } i_t = \sigma(W_f[H_{t-1}, X_t] + b_i) \\ \text{Forget gate : } f_t = \sigma(W_f[H_{t-1}, X_t] + b_f) \\ \text{Output gate : } O_t = \sigma(W_o[H_{t-1}, X_t] + b_o) \\ \text{New memory cell : } \tilde{C}_t = \tanh(W_c[H_{t-1}, X_t] + b_c) \\ \text{Updated memory cell : } C_t = f_t * c_{t-1} + i_t * \tilde{C}_t \\ \text{Output : } H_t = O_t * \tanh(C_t) \end{array} \right. \quad (2)$$

where X_t is the input at time t , H_t is the output at time t , C_t is the memory cell at time t , and i_t , f_t , O_t , and $\tilde{C}_t$ are the input, forget, output, and new memory cell gates, respectively. The weights W and biases b are learned during training. The symbol $*$ denotes a convolution operation, and σ is the sigmoid activation function. The function $\tanh$ is the hyperbolic tangent activation function.

An encoder–decoder structure is used to effectively predict spatiotemporal sequences by learning both spatial and temporal features of the input data (Figure 1). The encoder module of the model consists of three downsampling layers followed by three long short-term memory (LSTM) layers. The downsampling layers reduce the size of the input image features through convolution and extract spatial features from the image. The LSTM layers, on the other hand, learn the temporal features of the input sequence, which in this case is a sequence of radar echo images. The decoder module, on the other hand, consists of three upsampling layers followed by three LSTM layers. The upsampling layers increase the size of the image features through convolution and help guide the update of lower-level

features based on higher-level features. The LSTM layers learn the sequence features of the images and output the predicted radar echo images.

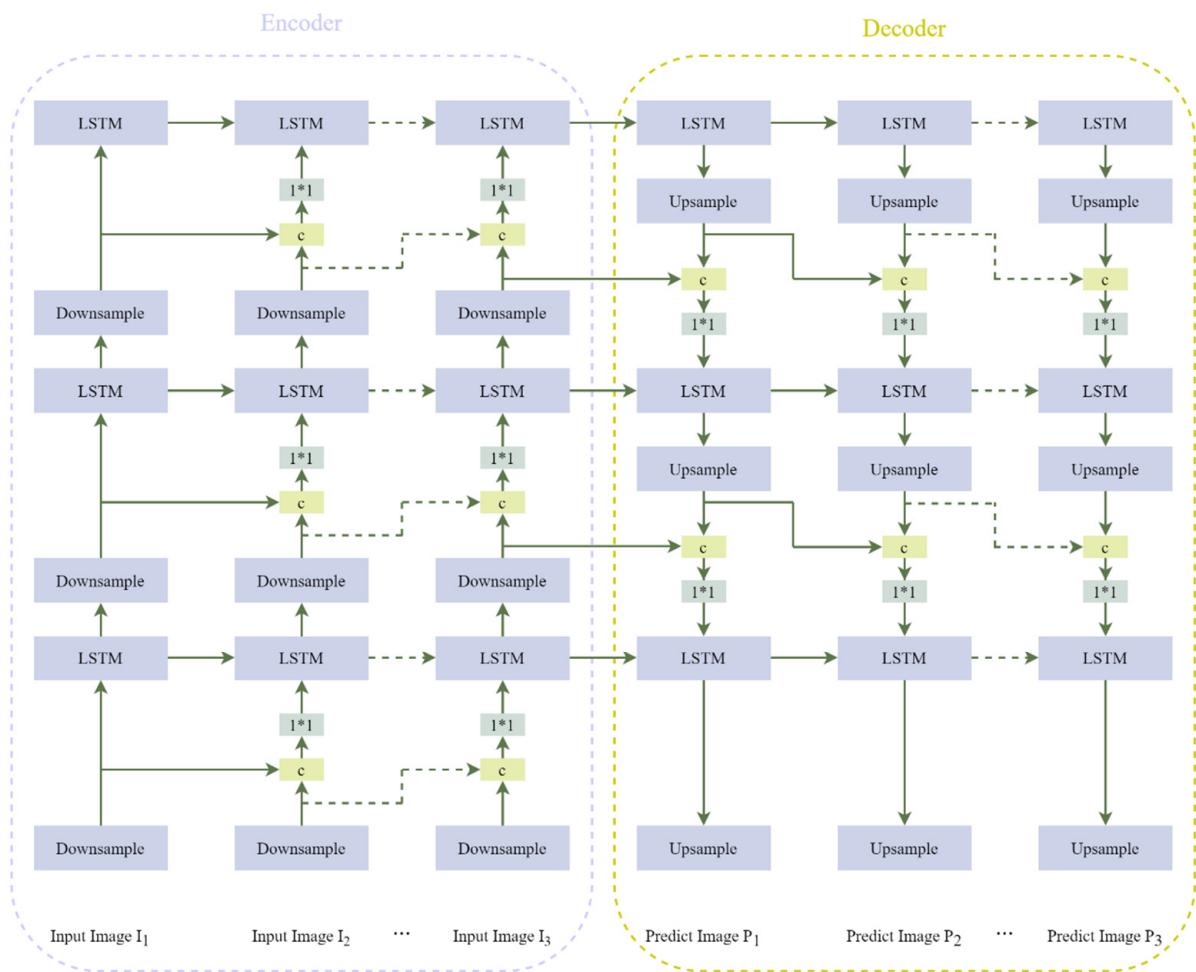


Figure 1. Architecture of the ConvLSTM network in the model.

2.2. MS-ConvLSTM

Here we describe the proposed model, MS-ConvLSTM, that is used to extrapolate radar echo information from radar data. The model uses a three-dimensional convolutional neural network to extract features synchronously from the spatial and temporal dimensions of the image streams, in order to capture the motion information between images. The model is based on the idea of a sequence-to-sequence prediction method (SEQ2SEQ) and uses an encoder–decoder structure, with an encoding network for reading the input sequence and a decoding network for predicting the output vector [13]. The model is composed of multiple ConvLSTM layers stacked to form the encoder–decoder structure. In this research, the model is trained on a sequence of 20 sets of radar product images, each representing 6 min of data. The first 10 sets of radar echo images are used as inputs, and the CR images in the last 10 sets of images are used as the expected outputs. The model is trained to predict the CR images based on the input radar echo images and other data such as ET, VIL, and wind field data.

2.2.1. Three-Dimensional Convolutional Operation

Convolutional neural networks (CNNs) are effective for image-processing tasks because they integrate the process of feature extraction and classification in a single network structure and reduce computational complexity through value-sharing and pooling operations [14]. However, traditional two-dimensional CNNs cannot accurately model the

time dimension in a radar frame sequence. To address this problem, the use of a three-dimensional CNN with a three-dimensional convolution kernel is proposed. This allows the network to analyze data by considering the cube formed by stacking multiple adjacent frames together [15]. This also allows the CNN to capture both spatial and temporal information in the sequence, improving its ability to model the data.

The two-dimensional convolution operates on a single image, while the three-dimensional convolution operates on a cube of stacked images. The three-dimensional (3D) convolution is an extension of traditional two-dimensional convolution that operates in the time dimension as well as the spatial dimensions. When applied to a sequence of images, the 3D convolution is able to capture both spatial and temporal information in the data. As a result, the output of the 3D convolution is a 3D spatiotemporal fusion structure, even if the input data consists of fewer than 3D cubes of images. Accordingly, the calculation formula of the 3D convolution is as follows (3):

$$f_{lj}^{xyz} = \sigma \left(b'_j + \sum_m \sum_{p=0}^{P_l-1} \sum_{q=0}^{Q_l-1} \sum_{r=0}^{R_l-1} W_{ljm}^{pqr} v_{(l-1)m}^{(x+p)(y+q)(z+r)} \right) \quad (3)$$

where f_{lj}^{xyz} indicates the output result of this convolutional layer, and $v_{(l-1)m}^{(x+p)(y+q)(z+r)}$ represents the input data data sample. $(l-1)m$ represents the feature map with serial number m in layer $l-1$; x and y represent the length and width of the spatial dimension of input samples, respectively, and z represents the length of input samples on the time axis; p , q and r respectively represent the specific size of the convolution operation in the above three dimensions. W_{ljm}^{pqr} is the weight parameter whose coordinate is (l, j, m) in the feature map; P_l , Q_l , R_l represents the size of the convolution kernel; b'_j represents the bias parameter of the feature map; and $\sigma(\cdot)$ is the activation function adopted by the neural network.

2.2.2. MS-ConvLSTM Architecture

The proposed MS-ConvLSTM network structure consists of three ConvLSTM layers. With each ConvLSTM layer containing three downsampling layers, the MS-ConvLSTM results in a total of nine layers in the model (Figure 2). The layers are as follows:

1. Input: The model receives two inputs: a main channel and an auxiliary channel. The main channel consists of a sequence of 10 consecutive frames of radar images, with each frame having a spatial size of 459×459 . The auxiliary channel consists of an auxiliary data sequence corresponding to the main channel information, with the same spatial size of 459×459 .
2. ConvLSTM1 with 3D Convolution: To extract features, a spatiotemporal convolution is conducted, which involves using 16 different 3D convolution kernels, each with a size of $5 \times 5 \times 5$. The size of 5×5 in the spatial dimensions represents the size of the convolution kernel in the x and y dimensions of the image. The length of 5 in the temporal dimension represents the number of time steps that the convolution kernel considers when performing the convolution. The convolution operation is performed on the input data, which consists of a sequence of 10 images with a spatial size of 459×459 . The 3D convolution kernels are applied to the data, sliding across the spatial and temporal dimensions and performing a dot product between the kernel values and the input data at each position. This results in a feature map with eight times the number of channels as the input data.
3. Pooling: After the convolution operation, the model performs a pooling operation. Pooling is often used in CNNs to reduce the size of feature maps, which can reduce the computational complexity of the model and improve its ability to generalize to new data. Pooling can also help the model to be invariant to small translations in the input, as the pooling operation aggregates values within a certain region and is not sensitive to the exact positions of the values within that region. The pooling operation involves downsampling the feature maps by taking the average or maximum value

within a certain region. In this case, the model performs a downsampling operation with a unit of 2×2 in the spatial domain and a downsampling with a unit of 2 in the time domain. This reduces the spatial and temporal resolutions of the feature maps, resulting in the third layer of the model. The specific parameters and configurations used for these operations, such as the size of the convolution kernels and the size of the downsampling operation, can affect the performance of the model and have been determined based on experiments.

4. ConvLSTM2 with 3D Convolution: To further extract features, 3D convolution is conducted, which involves using 32 different 3D convolution kernels, each with a size of $5 \times 5 \times 5$. This results in double the number of feature maps in the third layer.
5. Pooling: A 2×2 downsampling operation is applied to the spatial domain of each feature map in the fifth layer, and a subsampling operation with a sampling unit of 2 is applied to the time domain.
6. ConvLSTM3 with 3D Convolution: A 3D convolution operation is applied using 48 different convolution kernels with a size of $4 \times 4 \times 4$, resulting in a feature map that is 1.5 times the size of the fifth layer. This is followed by a downsampling operation with a size of $2 \times 2 \times 2$ to obtain the seventh layer of the model.
7. Pooling: This is followed by a downsampling operation with a size of $2 \times 2 \times 2$ to obtain the eighth layer of the model.
8. Fully connected classification: After three data feature processing steps using ConvLSTM, the model uses a traditional three-layer fully connected classifier using a softmax activation function for the final radar echo extrapolation. The ninth layer of the model consists of the feature map of a 1×1 convolution kernel, which is fully connected to all the feature maps in the eighth layer. This is defined as the input layer of the softmax classifier, and the number of nodes in the middle hidden layer is 96.
9. Prediction: The model takes a series of radar echo images and processes them using ConvLSTM1, ConvLSTM2, and ConvLSTM3, and finally makes a prediction about the next 10 consecutive frames. The prediction of the model is then established based on this processing.

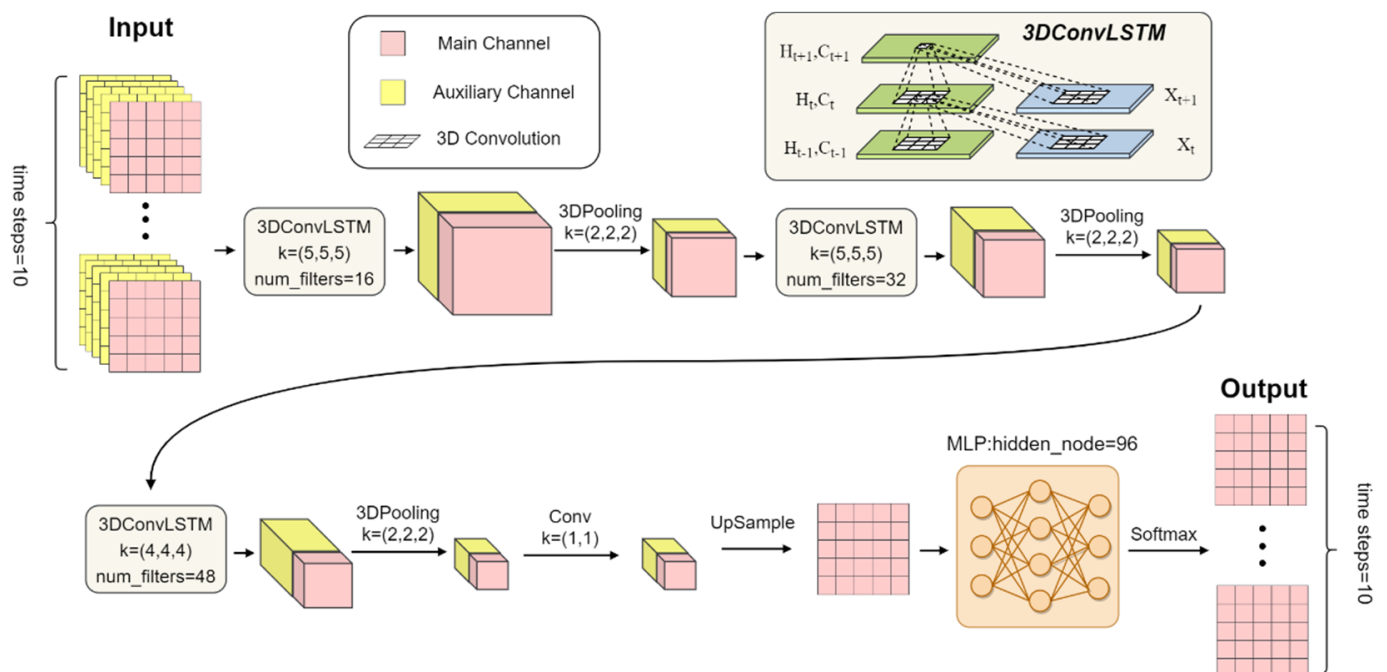


Figure 2. MS-ConvLSTM model structure diagram.

2.2.3. Evaluation Metrics

We use the following metrics to evaluate the performance of a binary classification model based on a predetermined threshold. In meteorological services, it is important to pay attention to different rainfall intensities, so the model was evaluated based on its overall performance at different radar echo levels. In the experiment, three thresholds are applied: 10 dBZ, 20 dBZ, and 30 dBZ. If the radar echo value is greater than a given threshold, it is 1; otherwise, it is 0.

True positive (*TP*) is the number of times the model correctly predicted that an event would occur (i.e., predicted the radar echo would exceed the threshold). False positive (*FP*) is the number of times the model incorrectly predicted that an event would occur (i.e., predicted the radar echo would exceed the threshold, but it actually did not). True negative (*TN*) is the number of times the model correctly predicted that an event would not occur (i.e., predicted the radar echo would not exceed the threshold). False negative (*FN*) is the number of times the model incorrectly predicted that an event would not occur (i.e., predicted the radar echo would not exceed the threshold, but it actually did). The probability of detection (*POD*) is a measure of the model's ability to correctly predict events. The false alarm rate (*FAR*) is a measure of the model's tendency to predict events that do not actually occur. The critical success index (*CSI*) is a measure of the overall performance of the model [16]. The calculation formulas of *POD*, *FAR*, and *CSI* are described as follows:

$$\begin{cases} POD = \frac{TP}{TP+FN} \\ FAR = \frac{FP}{TP+FP} \\ CSI = \frac{TP}{TP+FN+FP} \end{cases} \quad (4)$$

Conventional metrics—mean absolute error (MAE) and mean squared error (MSE)—were also used. Balanced mean absolute error (B-MAE) and balanced mean squared error (B-MSE) were also calculated to account for the imbalanced frequencies of different reflectivity magnitudes [11]. B-MAE is calculated as the MAE of the model's predictions, weighted by the frequency of each class or value. The weight for each class or value is set according to its frequency, with higher weights assigned to less frequent classes or values. In order to better apply the effect of the model in this paper, this article calculates the sum of MAE, MSE, B-MAE, and B-MSE (unit: dBZ) of each pixel pair of the predicted image and the ground truth radar echo image. Its calculation unit is the radar echo intensity unit dBZ. The B-MAE and B-MSE scores obtained are calculated as:

$$B\text{-MAE} = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^{459} \sum_{j=1}^{459} w_{n,i,j} |x_{n,i,j} - \hat{x}_{n,i,j}| \quad (5)$$

$$B\text{-MSE} = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^{459} \sum_{j=1}^{459} w_{n,i,j} (x_{n,i,j} - \hat{x}_{n,i,j})^2 \quad (6)$$

where N is the number of radar echo images and $w_{n,i,j}$ is the weight corresponding to the (i, j) pixel. $W(x)$ is set as 1, 2, and 5 according to reflectivity, respectively. More weight is allocated for larger reflectivity.

2.2.4. Model Parameter and Experiment Design

Considerations have been taken in designing and training the proposed MS-ConvLSTM model, including initialization of the model's parameters and softmax activation function. Initialization of the model's parameters is an important step in the training process, as it can affect the performance and convergence of the model. In this research, we utilized the Xavier initializer, which is a method for initializing the parameters to ensure that the input and output distributions of each layer are similar, which can help to avoid issues such as the output values of the activation function tending to zero or the gradient variance being affected by the number of outputs of the layer [17]. The softmax activation function is used in this study, and it maps the input values to a probability distribution over the different classes. It takes a vector of input values and produces a corresponding vector of output

values between 0 and 1, with each value representing the probability of the input belonging to a particular class. By normalizing the output values in this way, the model can choose the class with the highest probability as the output of the classifier.

In this study, the module uses a three-layer MS-ConvLSTM network with 5×5 cores in each layer. The first layer has 64 hidden states, while the second and third layers have 96 hidden states. To validate the proposed model's ability to predict radar echoes, experiments were conducted to compare with the original ConvLSTM model. The original learning rate is set to 0.0001, and the batch size is 4. The mean squared error is used as the loss function. The Adam optimization algorithm is used to minimize the loss function by training the optimization parameters. Adam is a first-order optimization algorithm that updates the neural network weights iteratively based on training data and is an effective optimizer in many cases. Before training, all the radar echo images are normalized to the range [0, 1]. The training and testing are performed using PyTorch on an NVIDIA GeForce RTX 3080 GPU.

2.3. Study Area and Materials

2.3.1. Study Area

The study area is Ningbo, a subprovincial city located in northeastern Zhejiang Province, China (Figure 3). This region has a temperate and humid subtropical monsoon climate with four distinct seasons. It has a complex and variable climate, with high rates of severe convective weather. The southwest part of the region is steep, while the northeast is low-lying.

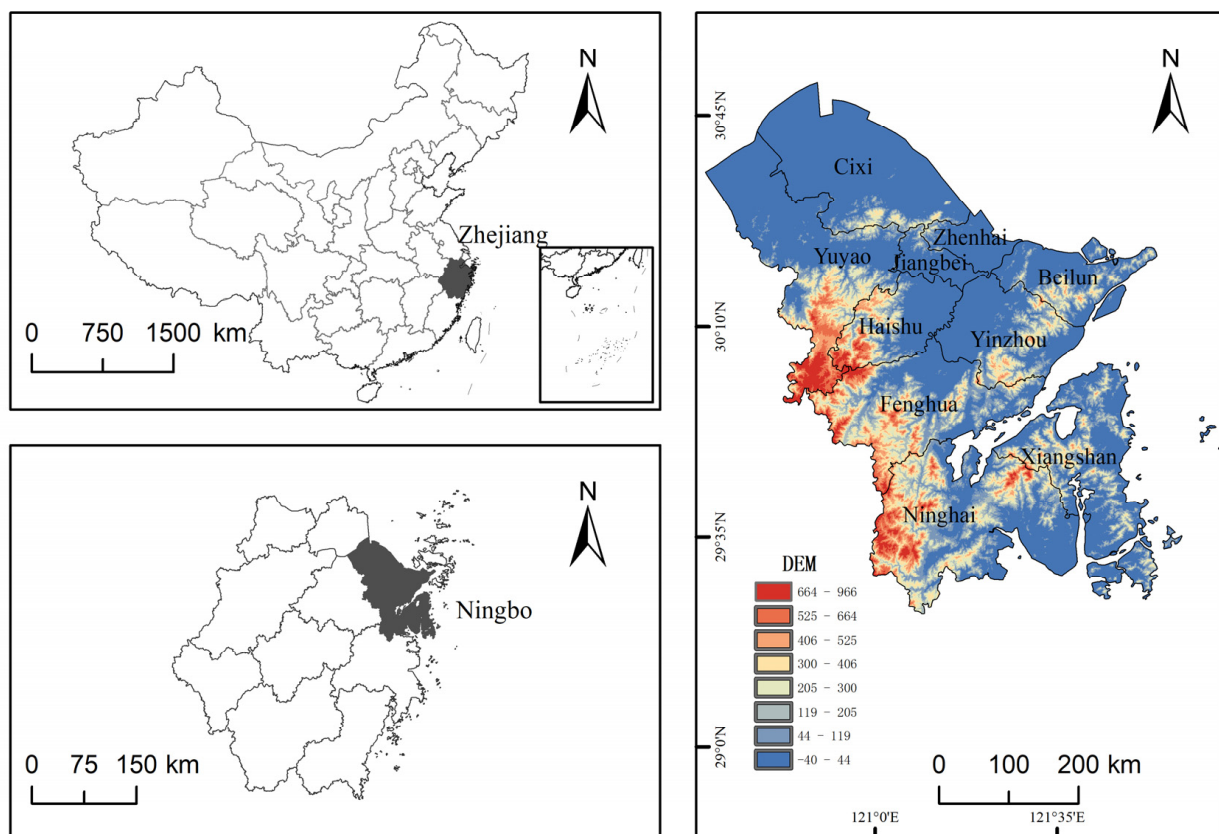


Figure 3. Geographical locations of Zhejiang Province in China (upper left), Ningbo in Zhejiang (lower left), and digital elevation model (DEM) in Ningbo (right).

2.3.2. Data Description

Data used in this research were obtained from the Ningbo Meteorological Bureau using an S-band Doppler weather radar (DWR) system at wavelengths of 8–15 cm [18]. DWR

is a piece of equipment that is widely used in meteorological observation; it can measure the position, intensity, and velocity of cloud droplets, raindrops, and other particles in the air, as well as some air molecules. This allows it to determine the corresponding weather conditions and internal structures. Weather radar data has a high spatial and temporal resolution of 1 km/6 min, and can provide information on the location, intensity, and motion of rainfall particles [19]. These data are used to study the correlation between the height and echo strength of the radar echo top and the occurrence of intense convective weather [20].

Multiple radar products are included in this dataset. Composite reflectivity (CR) is the maximum value of the fundamental reflectivity that occurs in a vertical column within the radar umbrella [21]. It can be used to predict the future movement of radar echoes. Echo top (ET) is the height of the top of the radar echo, measured from mean sea level, at the highest elevation angle at which the reflectivity coefficient is greater than or equal to 18 dBZ (adjustable threshold). ET is an important indicator of the intensity of convective weather and indirectly reflects the intensity of vertical updrafts in clouds [22]. Vertically integrated liquid (VIL) water content is a parameter obtained from volume-scanning radar representing the atmospheric water content that can be measured by classical (C- or S-band) weather radars. It is a useful indicator for detecting severe storms and may be useful for short-term rainfall prediction methods [23]. Radar-retrieved wind field data is used to calculate the average radial velocity of precipitation targets in each volume through the Doppler effect. It can detect atmospheric structure and determine wind direction [24,25]. In this study, a two-layer average of radar-retrieved wind field data is used to achieve computational efficiency. They are all important product data of radar and have high correlation with radar echo to a certain extent. Among them, CR, ET, and VIL data can reflect echo characteristics, and wind data have an important indicative role for the trend of radar echo. Therefore, this article selects these four types of data to predict radar echo.

2.3.3. Training Data Preparation

The weather radar data used in this study were collected from 2018 to 2020. These data were preprocessed to match the time and space of the radar data. The time matching process involved aligning the radar product data within the corresponding 6-min time based on the time of the CR radar product. The spatial matching process involved aligning the radar product data based on longitude and latitude. To ensure data quality, the data were also filtered to retain only radar products with reflectivity greater than 10 dBZ. The resulting dataset was used to construct a summer radar echo feature dataset for Ningbo in 2018, 2019, and 2020. This dataset was split into a training set of 21,984 sequences, a validation set of 2671 sequences, and a test set of 2773 sequences, each containing 20 radar images for a period of 6 min. The input data for the model consists of the radar product dataset (CR, ET, VIL, and wind field data), with CR as the main channel data and ET, VIL, and wind field data as auxiliary channel information. The output of the model is the future time CR image.

3. Results

In this experiment, the model was used to predict radar echo maps for the next 10 time steps based on the radar echo maps of the previous 10 time steps. This means that the radar echo data for the next hour was predicted using the historical data for the previous hour. To better understand the forecast performance, two cases from the test set were selected for analysis. These cases represent the general features of microscale and mesoscale weather and contain regions of strong echoes. They cover the creation, progression, and dissipation phases of convective processes.

The overall performance scores comparing ConvLSTM and MS-ConvLSTM are demonstrated in Table 1, where the MS-ConvLSTM model has better extrapolation performance than the ConvLSTM model. It is important to consider the specific context in which these models are being used and the goals of the prediction task when interpreting the results. In

general, lower values for the evaluation metrics indicate better performance, but this can depend on the specific cases and the relative importance of different types of errors.

Table 1. Comparison of the prediction results of the two models. The best results are marked in bold.

Model	MAE	MSE	B-MAE	B-MSE
ConvLSTM	7511	2973	14,647	5729
MS-ConvLSTM	7038	2667	14,501	5613

Based on the ground truth and prediction results in Figure 4, it seems that both the ConvLSTM model and the MS-ConvLSTM model can forecast the development of radar echoes at 6 min. At 30 min, some echo details are missing, but both models can capture the change in the strong echo position and movement. The MS-ConvLSTM model appears to be more effective at forecasting the strongest echo region than the ConvLSTM model. After 30 min, the predicted image is based on the echo features obtained from the model due to the long extrapolation time, and some strong echo areas can still be recognized. However, compared to the ground truth, the edge details of the radar echoes of different intensities in the model-predicted radar echo maps are progressively missing.

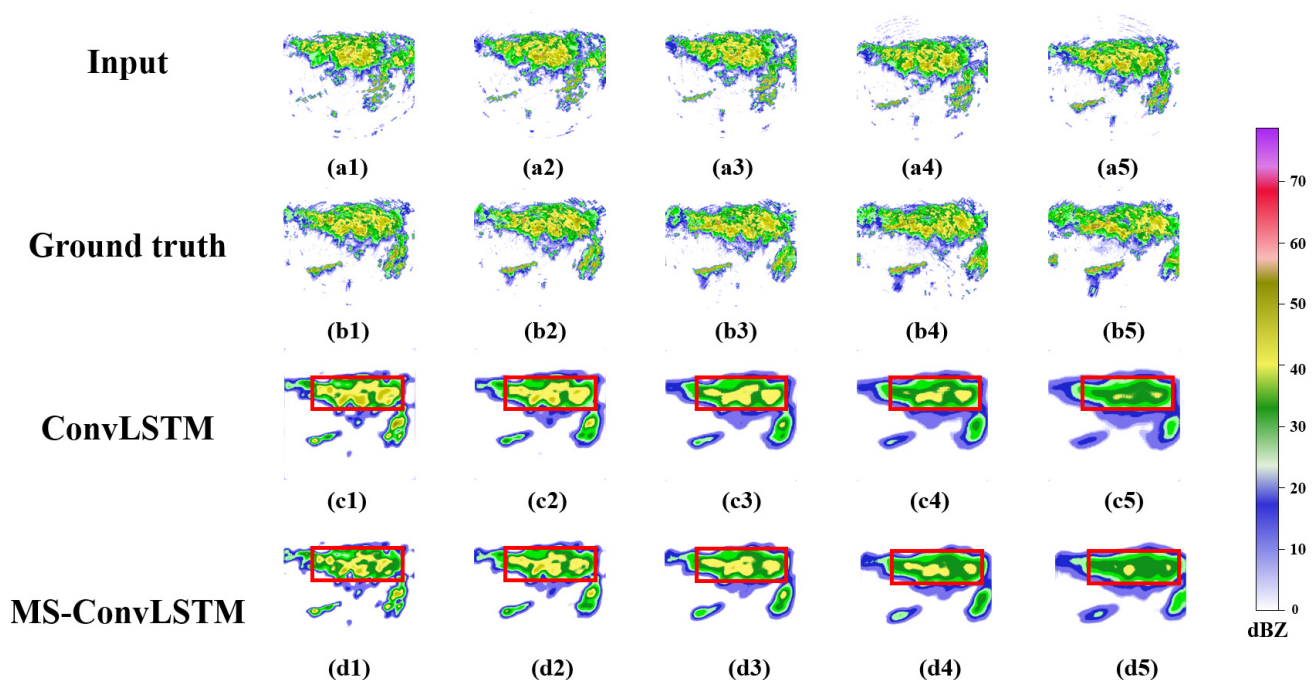


Figure 4. The visualization results of test sample 1: (a1–a5) represent part of the input of the model; (b1–b5) the ground real radar echo charts of 6, 18, 30, 42 and 60 min; and (c1–c5) and (d1–d5) are the corresponding prediction results of ConvLSTM and MS-ConvLSTM, respectively.

To quantitatively compare the predictions of test sample 1, we provide the POD, FAR, and CSI values of the two models at 6, 30, and 60 min, respectively (Table 2). The POD and CSI values decrease, and the FAR value increases as the prediction time increases. The MS-ConvLSTM model has better prediction performance than the ConvLSTM model at all intensities and time intervals, based on the average scores.

Table 2. The probability of detection (POD) values, false alarm rate (FAR) values, and critical success index (CSI) of the two models in test sample 1 at 6 min, 30 min, and 60 min. The best results are marked in **bold**.

Model	Time (min)	POD				FAR				CSI			
		dBZ Threshold				dBZ Threshold				dBZ Threshold			
		10	20	30	avg	10	20	30	avg	10	20	30	avg
ConvLSTM	6	0.8915	0.8581	0.5806	0.7767	0.0953	0.0994	0.2125	0.1357	0.8222	0.8004	0.6048	0.7425
	30	0.8302	0.7372	0.4409	0.6694	0.1682	0.1718	0.2397	0.1932	0.7105	0.6804	0.3870	0.5926
	60	0.7540	0.5659	0.1280	0.4826	0.1886	0.1963	0.3628	0.2492	0.6415	0.5337	0.1104	0.4286
MS-ConvLSTM	6	0.9002	0.8778	0.7228	0.8336	0.0655	0.0735	0.1403	0.0931	0.8391	0.8031	0.5303	0.7242
	30	0.8411	0.7720	0.4620	0.6917	0.1279	0.1510	0.2367	0.1719	0.7487	0.7165	0.4041	0.6231
	60	0.7978	0.6357	0.1646	0.5327	0.1531	0.1842	0.4284	0.2552	0.6972	0.5918	0.1568	0.4818

Based on the ground truth and prediction results shown in Figure 5, both the ConvLSTM model and the MS-ConvLSTM model can accurately forecast the location of intense radar echoes at 6 min for a test sample with weaker total reflectivity compared to test sample 1. As in test sample 1, the MS-ConvLSTM model appears to be more effective at predicting the position of the strongest echo, although some echo details are missing at 15 min. After 30 min, the predictions lose more details due to the long extrapolation time, and the strong echo areas are difficult to recognize in the corresponding prediction maps. The prediction maps also combine the regions into one wide echo, creating a stronger and larger echo region than the ground truth map. This flaw may need to be addressed in future work.

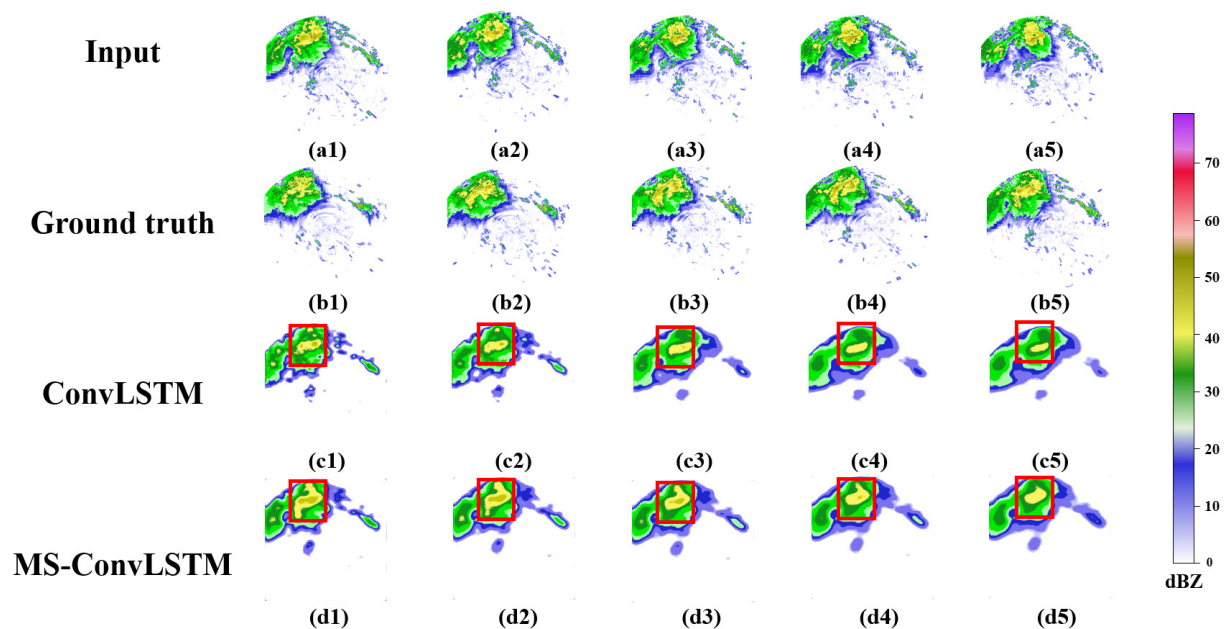


Figure 5. The visualization results of test sample 2: (a1–a5) represent part of the input of the model; (b1–b5) are the ground real radar echo images of 6, 18, 30, 42 and 60 min; and (c1–c5) and (d1–d5) are the corresponding prediction results of ConvLSTM and MS-ConvLSTM, respectively.

Table 3 shows the POD, FAR, and CSI values for the test sample 2 prediction framework acquired by the two models at 6, 30, and 60 min, respectively. The MS-ConvLSTM model had better predictive scores than the ConvLSTM model at all forecast times (6, 30, and 60 min). It also seems that both POD and CSI values decrease with increasing forecast time and rainfall intensity, while FAR values increase. This suggests that the models become

less accurate as the forecast time increases and the intensity of the event increases, and that the MS-ConvLSTM model is better at avoiding false alarms than the ConvLSTM model.

Table 3. The probability of detection (POD) values, false alarm rate (FAR) values, and critical success index (CSI) for test sample 2 at 6 min, 30 min, and 60 min. The best results are marked in **bold**.

Model	Time (min)	POD				FAR				CSI			
		dBZ Threshold				dBZ Threshold				dBZ Threshold			
		10	20	30	avg	10	20	30	avg	10	20	30	avg
ConvLSTM	6	0.8616	0.8213	0.5604	0.7477	0.1161	0.2123	0.3772	0.2352	0.7361	0.7182	0.5386	0.6637
	30	0.8020	0.7121	0.4251	0.6464	0.1865	0.3720	0.6384	0.3989	0.6205	0.5732	0.3203	0.5046
	60	0.7825	0.5480	0.1346	0.4883	0.2115	0.4975	0.7628	0.4906	0.6115	0.5223	0.1087	0.4141
MS-ConvLSTM	6	0.8802	0.8636	0.6282	0.7906	0.0928	0.1807	0.3437	0.2057	0.7491	0.7143	0.4674	0.6402
	30	0.8331	0.7402	0.4517	0.6750	0.1680	0.3531	0.6367	0.3859	0.6487	0.6878	0.3914	0.5759
	60	0.7946	0.6194	0.1427	0.5189	0.1931	0.3741	0.7075	0.4249	0.6172	0.5973	0.1582	0.4575

Based on multisource data, we propose a deep learning network named MS-ConvLSTM for radar echo extrapolation. MS-ConvLSTM can effectively achieve 0–1 h radar echo extrapolation. Two radar echo extrapolation cases demonstrate that our MS-ConvLSTM model is proved to be effective through qualitative and quantitative analysis.

The MS-ConvLSTM model has a strong feature learning ability, which can effectively realize the fusion of multisource data. Among these auxiliary data, ET can help the model to better learn the change characteristics of radar echo, VIL can help the model understand the variability properties of the water vapor content in the study area, and wind field data can add some dynamic promotion or constraint. This work is an attempt to realize the comprehensive application of multisource data. Deep learning has strong feature extraction ability, so it will become a promising method of radar echo extrapolation.

4. Discussion

This research proposes the MS-ConvLSTM model that extends from the original ConvLSTM using a combination of 3D convolutions and auxiliary mechanisms. In order to extrapolate the radar echo more effectively, we incorporate an auxiliary data channel into the network, which enables the model to incorporate multiple data sources into the prediction process to improve the accuracy of radar echo extrapolation. We have conducted extensive experiments to verify the effectiveness of the proposed MS-ConvLSTM. The model is trained on a large dataset of multisource radar data, and experimental results show that it has competitive prediction accuracy and can delay the loss of predictive radar information to some extent. The visualization results and quantitative metric score analysis show that the MS-ConvLSTM model can efficiently capture the temporal and spatial features of radar echo maps, and that the predicted radar echoes more closely resemble the ground truth than the ConvLSTM model. The MS-ConvLSTM model has lower values of B-MAE and B-MSE, and significantly lower values of MAE and MSE, compared to the ConvLSTM model. This suggests that the MS-ConvLSTM model has a lower error rate and is more accurate than the ConvLSTM model. The MS-ConvLSTM model also performs better on the POD, FAR, and CSI scores, indicating that it has better predictive capability and effectiveness than the ConvLSTM model.

5. Conclusions

This study proposed an MS-ConvLSTM model based on 3D convolutions and auxiliary mechanisms for the spatiotemporal prediction of radar echoes. It showed that incorporating multisource data could improve the accuracy of radar echo extrapolation of the proposed model compared to traditional ConvLSTM.

Despite the improved prediction accuracy of the MS-ConvLSTM, many shortcomings are still evident in the prediction stages. Many prediction details are gradually lost, and the best input sequence prediction performance under different thresholds is unstable. In addition, most CNN-based deep learning technologies have an over-smoothing problem when dealing with image problems because convolution is essentially an aggregation method. When the convolution kernel takes a specific value, it is a smoothing operation. In future work, we will conduct experiments with different convolution kernels to alleviate the problem of over-smoothing. There is potential for refinement by combining a meteorology physical conceptual model with extrapolation, and the model's generalizability could be evaluated by applying it to larger datasets in the future. Factors such as terrain and temperature could potentially be incorporated into the forecast in the future.

Author Contributions: Conceptualizations, M.L., M.Y., Y.Z. and Q.Z.; methodology, M.L., Q.Z., M.Y., Y.Z. and B.L.; software, Y.L. and M.Y.; validation, M.L.; formal analysis, Y.L.; investigation, Y.L. and Q.Z.; resources, M.L. and M.Y.; data curation, Y.L.; writing—original draft preparation, Y.L., M.L. and M.Y.; writing—review and editing, M.L., M.Y. and B.L.; visualization, M.W.; supervision, M.L. and Q.Z.; project administration, Y.L.; funding acquisition, M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was supported by the NSFC Project (41871285).

Data Availability Statement: Some data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: Many thanks to the reviewers for their valuable comments.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Zaytar, M.A.; El Amrani, C. Sequence to sequence weather forecasting with long short-term memory recurrent neural networks. *Int. J. Comput. Appl.* **2016**, *143*, 7–11.
2. Morris, L.W.; Christopher, D.; Wei, W.; Kevin, W.M.; Joseph, B.K. Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model. *Weather Forecast.* **2008**, *23*, 407–437.
3. Cong, W.; Ping, W.; Di, W.; Jinyi, H.; Bing, X. Nowcasting multicell short-term intense precipitation using graph models and random forests. *Mon. Weather Rev.* **2020**, *148*, 4453–4466.
4. Houze, R.A., Jr.; Rutledge, S.A.; Biggerstaff, M.I.; Smull, B.F. Interpretation of Doppler weather radar displays of midlatitude mesoscale convective systems. *Bull. Am. Meteorol. Soc.* **1989**, *70*, 608–619. [[CrossRef](#)]
5. Jing, J.; Li, Q.; Ma, L.; Chen, L.; Ding, L. REMNet: Recurrent Evolution Memory-Aware Network for Accurate Long-Term Weather Radar Echo Extrapolation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4109313. [[CrossRef](#)]
6. Nitish, S.; Elman, M.; Ruslan, S. Unsupervised learning of video representations using LSTMs. *PLMR* **2015**, *37*, 843–852.
7. Xingjian, S.; Hao, W.; Dit-Yan, Y.; Zhourong, C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; pp. 802–810.
8. Lin, T.; Xutao, L.; Yunming, Y.; Pengfei, X.; Yan, L. A generative adversarial gated recurrent unit model for precipitation nowcasting. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 601–605.
9. Sato, R.; Kashima, H.; Yamamoto, T. Short-term precipitation prediction with skip-connected prednet. In Proceedings of the International Conference on Artificial Neural Networks, Rhodes, Greece, 4–7 October 2018; Springer: Cham, Switzerland, 2018; pp. 373–382.
10. Rane, R.P.; Szügyi, E.; Saxena, V.; Ofner, A.; Stober, S. Prednet and predictive coding: A critical review. In Proceedings of the 2020 International Conference on Multimedia Retrieval, Dublin, Ireland, 8–11 June 2020; pp. 233–241.
11. Shi, X.; Gao, Z.; Lausen, L.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; Woo, W.-C. Deep learning for precipitation nowcasting: A benchmark and a new model. *arXiv* **2017**, arXiv:1706.03458.
12. Pulkkinen, S.; Nerini, D.; Pérez Hortal, A.A.; Velasco-Forero, C.; Seed, A.; Germann, U.; Foresti, L. Pysteps: An open-source Python library for probabilistic precipitation nowcasting (v1. 0). *Geosci. Model Dev.* **2019**, *12*, 4185–4219. [[CrossRef](#)]
13. Kedong, Z.; Yaping, L.; Wenbo, M.; Feng, L. LSTM enhanced by dual-attention-based encoder-decoder for daily peak load forecasting. *Electr. Power Syst. Res.* **2022**, *208*, 107860. [[CrossRef](#)]
14. Feltus, C. Learning Algorithm Recommendation Framework for IS and CPS Security: Analysis of the RNN, LSTM, and GRU Contributions. *Int. J. Syst. Softw. Secur. Prot.* **2022**, *13*, 36. [[CrossRef](#)]

15. Li, Y.; Zhang, H.; Shen, Q. Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network. *Remote Sens.* **2017**, *9*, 67. [[CrossRef](#)]
16. Lin, C.; Vasić, S.; Kilambi, A.; Turner, B.; Zawadzki, I. Precipitation forecast skill of numerical weather prediction models and radar nowcasts. *Geophys. Res. Lett.* **2005**, *32*, L14801. [[CrossRef](#)]
17. Datta, L. A survey on activation functions and their relation with xavier and he normal initialization. *arXiv* **2020**, arXiv:2004.06632.
18. Binetti, M.S.; Campanale, C.; Massarelli, C.; Uricchio, V.F. The Use of Weather Radar Data: Possibilities, Challenges and Advanced Applications. *Earth* **2022**, *3*, 157–171. [[CrossRef](#)]
19. Smith, T.M.; Elmore, K.L.; Dulin, S.A. A damaging downburst prediction and detection algorithm for the WSR-88D. *Weather Forecast.* **2004**, *19*, 240–250. [[CrossRef](#)]
20. Usharani, B. ILF-LSTM: Enhanced loss function in LSTM to predict the sea surface temperature. *Soft Comput.* **2022**. [[CrossRef](#)]
21. Sun, F.; Li, B.; Min, M.; Qin, D. Deep Learning-Based Radar Composite Reflectivity Factor Estimations from Fengyun-4A Geostationary Satellite Observations. *Remote Sens.* **2021**, *13*, 2229. [[CrossRef](#)]
22. Lakshmanan, V.; Hondl, K.; Potvin, C.K.; Preignitz, D. An Improved Method for Estimating Radar Echo-Top Height. *Weather Forecast.* **2013**, *28*, 481–488. [[CrossRef](#)]
23. Boudevillain, B.; Andrieu, H. Assessment of vertically integrated liquid (VIL) water content radar measurement. *J. Atmos. Ocean. Technol.* **2003**, *20*, 807–819. [[CrossRef](#)]
24. Altube, P.; Bech, J.; Argemí, O.; Rigo, T.; Pineda, N.; Collis, S.; Helmus, J. Correction of Dual-PRF Doppler Velocity Outliers in the Presence of Aliasing. *J. Atmos. Ocean. Technol.* **2017**, *34*, 1529–1543. [[CrossRef](#)]
25. Miller, M.L.; Lakshmanan, V.; Smith, T.M. An automated method for depicting mesocyclone paths and intensities. *Weather Forecast.* **2013**, *28*, 570–585. [[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.