Survey on the Application of Deep Learning in Extreme Weather Prediction

Wei Fang 1,2,3,* , Qiongying Xue 1,* , Liang Shen 1 and Victor S. Sheng 4

Abstract: Because of the uncertainty of weather and the complexity of atmospheric movement, extreme weather has always been an important and difficult meteorological problem. Extreme weather events can be called high-impact weather, the ‘extreme’ here means that the probability of occurrence is very small. Deep learning can automatically learn and train from a large number of sample data to obtain excellent feature expression, which effectively improves the performance of various machine learning tasks and is widely used in computer vision, natural language processing, and other fields. Based on the introduction of deep learning, this article makes a preliminary summary of the existing extreme weather prediction methods. These include the ability to use recurrent neural networks to predict weather phenomena and convolutional neural networks to predict the weather. They can automatically extract image features of extreme weather phenomena and predict the possibility of extreme weather somewhere by using a deep learning framework.

Keywords: deep learning; extreme weather prediction; recurrent neural network; convolutional neural network

1. Introduction

One minute the sky is clear, and the next minute there will be dark clouds and strong winds; pedestrians have difficulty standing, trees are cut off, and simple single-story buildings are even reduced to rubble. Between 2000 and 2020, the cumulative number of disasters worldwide reached 13,345, and more than 1.5 million people died [1]. When extreme weather occurs, there will be short-term heavy rainfall, thunderstorms, gales, tornadoes, hail, etc. These are all catastrophic weather phenomena. They have the characteristics of suddenness and locality. Although the duration is not long, they are very destructive. At present, there is no effective method for artificial weakening and prevention. Therefore, prevention should be given priority and a combination of prevention and rescue should be achieved. It is very important to strengthen the theoretical research on the severe convective weather system and improve the forecasting level of severe convective weather. With the development of science and technology, radar and geosynchronous satellite are gradually applied to the meteorological field to detect and monitor weather changes. Various weather stations use a weather radar to forecast the occurrence of severe convective weather and monitor its activities. Through the analysis of satellite cloud images taken continuously by meteorological satellites, we can explore and track the occurrence, development, movement, and extinction process of severe convective weather; combined with the analysis of weather situation maps by experienced weather forecasters, the forecast of severe convective weather can be improved level.

As the most important branch of machine learning, deep learning has developed rapidly in recent years and has been applied to various fields. It has attracted widespread...
attention from many countries, and it has a long history of development. Today, deep learning has become increasingly mature, especially in images. Whether in scientific research or application, people are becoming increasingly rational and no longer blindly follow the early trend of deep learning that was considered “omnipotent”. Of course, there are still many problems to be solved in this field, and there are also many interesting and challenging fields to be studied, including in the field of meteorology.

The concept of deep learning was first proposed in 2006 by Hinton, G.E. et al. [2] of the University of Toronto. The principle of deep learning is similar to brain activity, allowing the computer to train the data it has already seen [3]. Deep learning is a mathematical framework for learning representations from data. It emphasizes learning from successive levels. These hierarchical representations are almost always learned through neural network models. The neural network saves the input data of each layer in the weight of that layer. The transformation implemented by each layer is parameterized by its weight. Learning means finding a set of weight values for all layers of the neural network so that the network can accurately correspond to each example input to its goal one to one. We use the loss function to measure the quality of the network output. The input of the loss function is the network predicted value and the true target value, and then the distance value is calculated. Use this distance value as a feedback signal to fine tune the weight value to reduce the current example. This adjustment is done by an optimization program that implements the so-called backpropagation algorithm. This is the core algorithm of deep learning. Aiming at the urban air quality forecast in the era of environmental big data, Yin Wenjun et al. [4] proposed a new method based on deep learning. This method simulates the neural connection structure of the human brain, transforms the feature representation of the data in the original space into a new feature space with semantic features, and then automatically learns the hierarchical feature representation, thereby improving the prediction performance. Similarly, weather forecasting can be seen as a process of big data analysis. Deep learning environments usually require continuous training with a large amount of data, which makes past data a key research point in deep learning. Although a typical computer vision application can process 512 * 512 pixel images [5], many meteorological data, such as digital model output, output data, and satellite data, are global in scale, with a medium-resolution global weather field. The size (about 1 km) is about 40,000 * 20,000 pixels, which is three times more than the order of magnitude that computer vision can handle [5]. Figure 1 shows the relationship among artificial intelligence, machine learning and deep learning.

![Figure 1. Relationship among Artificial Intelligence, Machine Learning and Deep Learning.](image_url)

This article will be able to summarize the previous research, combined with deep learning to make a relatively complete summary of the prediction of extreme weather. This paper is organized as follows: Section 1 outlines the main real learning knowledge in extreme weather prediction; Section 2 focuses on the related models and theories of deep learning in artificial intelligence and the traditional methods of weather prediction;
Section 3 focuses on the depth of artificial intelligence. Learn the existing applications of extreme weather prediction, and finally, Section 4 summarizes the extreme weather prediction method based on deep learning.

2. Traditional Methods

In this section, we will focus on the existing traditional method theories or conclusions of weather prediction related to deep learning methods. There are mainly weather map forecasting methods, statistical forecasting methods, numerical forecasting methods, and single-station forecasting methods. The following will list the current weather forecasting methods commonly used or related weather forecasting knowledge.

2.1. Weather Map Forecast Method

Since the appearance of the telegraph, weather map forecasting methods [6] have existed for more than 100 years. Data can be collected in time to the National Meteorological Center to analyze weather maps. Looking at the weather chart, the low-pressure system is moving. We have learned from the weather map to analyze the weather system and predict its future movement and intensity changes (including generation and elimination), which is the main basis for the weather map forecasting method. We analyze weather maps or other auxiliary maps to timely discover the weather system that causes changes in the weather around us. The correctness of weather map analysis is the premise of the weather map forecasting method. The first step of the weather map forecasting method is to predict the weather, that is, predict the existing weather system and its future trends on the weather map. These include changes in the intensity of the new weather system. In weather forecasting, the simplest method is extrapolation, which assumes the movement and some of the weather system in the future. This method is also called a continuous method. Secondly, in the practice of natural gas prediction, weather forecasters summarize the empirical laws of weather system movement or intensity changes. Rules of thumb also play an important role in weather forecasting. In addition, from the theory of dynamic meteorology, the rules of weather conditions can be derived, and weather forecasters can predict future weather conditions based on these data.

Lou Maoyuan et al. [7] introduced in detail a design idea and application method that can be used to predict cloud and snow weather maps. Based on many years of forecasting practice, they selected the relative humidity, GFS and EC numerical forecast data (usually used for forecasting operations) related to GRAPES cloud forecasting and filtered out the wet layer reflecting the high cloud, thus forming the intuitive profile view. The intuitive profile view can be used to reduce the lack of total cloud volume. Select the temperature element and superimpose it on the graph to form a comprehensive curve of temperature and humidity changes with height. Combined with the predicted points, the “air column method” is used to simulate precipitation, and intuitively judge whether the snowflakes melt during autumn. It can overcome the shortcomings caused by the current station’s use of atmospheric temperature indicators to predict snowfall.

2.2. Weather Forecast Based on Weather Radar

Radar can emit short-wavelength radio waves from the sky-white line. When it encounters weather phenomena in the distant atmosphere such as typhoons, thunderstorms, and rainstorms, its radio waves will be reflected and displayed on the radar screen. Therefore, we can see the overall appearance and internal structure of typhoons, thunderstorms, and rainstorms on the screen. On the weather radar chart, the bright area represents the area monitored by the radar, and the area where the colored blocks appear is the active site of the rainfall system. Generally speaking, the area corresponding to the blue echo means that the area is covered by precipitation clouds, but the rain has not yet appeared; the area covered by the green echo represents light rain in the area; the area covered by the yellow to red echo has moderate to heavy rain; The purple echo area has the highest precipitation intensity. The area is experiencing heavy rain or even heavy rain, and it may
be accompanied by severe weather such as thunder, gale, or even hail. Sokol Z [8] summarized the role of weather radar in rainfall estimation and its application in meteorological and hydrological simulation.

Hu Lijun et al. [9] designed a weather display system based on radar data, which is characterized in that the radar data collector is connected with cloud detection radar, rain detection radar, and wind detection radar. Li S et al. [10] used FY-4A satellites and Doppler weather radar to observe the precipitation characteristics of sudden heavy rain events in the complex terrain of Southwest China. Xie B et al. [11] use satellite radar images for machine learning to estimate losses after natural disasters. They will use satellite radar images and geographic data as input to classify the damage status of individual buildings after a major disaster event. They believe that the application of their damage estimation method in real-world natural disaster events will have great potential to improve social resilience.

In terms of weather radar detection theory and methods, Ge Runsheng and others studied the calibration and measurement methods of various weather radar parameters, and theoretically analyzed the measurement error and accuracy of echo intensity. Researched the radar signal processing method, analyzed the x-band and c-band weather radar’s ability to detect rainfall, theoretically analyzed the error and accuracy of the weather radar measurement, and studied the radar signal processing method [12,13].

For extreme weather forecasts, quality control and preprocessing of radar data can also be performed. Bi Yongheng [14] tried to find a method suitable for attenuation correction of x-band dual-polarization radar during precipitation. They hope to carry out quality control and preprocessing of radar data before calibration and on the basis of analyzing existing calibration methods at home and abroad. They choose and improve the adaptive constraint algorithm as the radar reflectivity attenuation correction method. Goudenhoofdt E et al. [15] estimated the regional frequency of extreme rainfall in Belgium based on radar, and they assessed the potential of single weather radar for 12-year quantitative precipitation estimation. During the period 2005–2016, they compared radar estimates of extreme rainfall within 1 h and 24 h.

2.3. Numerical Prediction Methods

Weather forecasting is always based on the principles of meteorology. With the development of computing technology and detection technology, in addition to the traditional weather map method and mathematical statistics method for forecasting, meteorological radar and satellite detection data are also used in forecasting operations, and numerical forecasting methods have been developed. This method can predict the physical process of the atmosphere by determining the principles of conservation of atmospheric mass, energy, and momentum, and significantly improve the quality of weather forecasts, thereby promoting the objective quantification of weather forecasts [16]. As a discrete calculation model of the earth’s atmosphere, the numerical weather prediction model is a typical nonlinear system with a large amount of calculation. In order to complete all calculations in a shorter time than the actual weather evolution, high-speed computers have become a decisive key technology.

Du Jun et al. [17] believe that the definition of abnormal weather should naturally be based on the local and then climate background (Figure 2). According to this view, the researchers defined the anomalous degree of a weather element as the difference between it and the actual climate average. In addition, considering that the rate of change of weather elements in different places and different seasons is very different (for example, in general, the rate of change is larger in high latitudes and smaller in low latitudes, and larger in winter and larger in summer). In order to facilitate unified comparison, the difference is standardized by the actual climatic standard deviation of the quantity, namely formulas (1) and (2) and called “standardized anomaly” or SA (Standardized Anomaly) [17]:

\[
A_{s-o}(x,t) = \frac{OSB(x,t) - MEAN_{clim}(x,t)}{SD_{lim}(x,t)},
\]  

(1)
The normalized anomaly degree $SA$ is a function of location $x$ and time $t$. For the actual situation $o$, the standardized anomaly degree $A_{s-o}(x,t)$ is calculated with the observed value $OBS(x,t)$ of the element and the climatic mean $MEAN_{clim}(x,t)$ and the climatic standard deviation $SD_{clim}(x,t)$ of the actual atmosphere, that is, formula (1); for the forecast $f$, the standardized anomaly degree $A_{s-f}(x,t)$ is calculated with the forecast value $FCST(x,t)$ instead of the observed value $OBS(x,t)$, that is, formula (2).

$$A_{s-f}(x,t) = \frac{FCST(x,t) - MEAN_{lim}(x,t)}{SD_{clim}(x,t)}, \quad (2)$$

The horizontal resolution and vertical resolution of the global model and the regional model have been significantly improved. At the same time, the physical process of the model has also been improved. The numerical prediction business model has entered the stage of massively parallel computing. Major developed countries and China are committed to developing their own new service numerical forecast models: non-static (multi-scale) integrated models or non-static mesoscale models. A new generation of comprehensive numerical models of weather has begun to operate in some countries. Operational numerical forecasting models are developing in the direction of continuous improvement. With the improvement of model resolution, increasing amounts of attention have been paid to cloud physical processes, surface processes and turbulence processes, parameterization schemes of radiation processes considering aspect factors, and the selection of model vertical coordinates.

Heng BCP et al. [18] invented a data assimilation system for convective-scale numerical weather prediction in Singapore. The background error covariance obtained by them is in a false vertical structure containing higher-level models, which may reduce prediction performance. They found that precipitation forecasts are sensitive to horizontal resolution and lateral boundary conditions. Their observation system tests show that although satellite radiation assimilation has a significant help to the precipitation forecast in the region, it hurts the background temperature and high altitude wind. Compared with the isolated forecast model, the research system is more valuable for the precipitation forecast near the forecast range, but the lead time is longer.

3. Deep Learning Applications

3.1. Application of BP(Back Propagation) Neural Network in Extreme Weather

BP algorithm also played a role in extreme weather forecasting. The topological structure of the BP neural network model includes the input layer, hidden layer, and output
layer [19]. The neural network is mainly composed of three parts, namely: the parameter learning algorithm of network architecture activation function to find the optimal weight. BP algorithm is one of the more widely used parameter learning algorithms [20,21]. BP Neural Network is the concept of error back propagation algorithm proposed by scientists led by Rumelhart and McClelland in 1986. It is a multi-layer feedforward neural network based on training. Since we cannot directly obtain the weight of the hidden layer, can we first adjust the weight of the hidden layer indirectly by obtaining the output result of the output layer and the error of the expected output, the BP algorithm is designed in this way. The basic idea is that the learning process consists of two processes: signal forward propagation (looking for loss) and error back propagation (error back propagation).

According to the basic idea of BP algorithm, the general process of BP algorithm can be obtained:

Forward FP (calculate loss). In this process, we calculate the loss value and actual value between the final output value and the output value according to the input sample and the given initial weight value W and offset the process of back propagation of the item value b. If the loss value is not a given Within the range; otherwise, stop W and B update.

The three-layer error back propagation algorithm model is the most commonly used model in weather forecasting. It is a BP neural network structure diagram composed of N nodes in the input layer, M nodes in the hidden layer, and P nodes in the output layer—see Figure 3.

![Three-layer BP artificial neural network](image)

**Figure 3.** Three-layer BP artificial neural network.

Literature [22] proposed a BP neural network prediction model based on a simulated particle swarm algorithm. The experimental results show that the optimization algorithm has a good pre-calibration effect for network parameter optimization. Literature [23] proposed wavelet analysis and BP neural network wind speed prediction model. The results show that the prediction results of this model are more in line with the actual wind speed change trend and have better generalization ability. Literature [24] proposed an optimized BP neural network prediction model based on error correction. The results show that considering the error factor can improve the accuracy of wind speed forecasting. Wen [25] proposed a neural network wind speed prediction model, which considered meteorological factors, but not only meteorological factors. This method can effectively improve the accuracy of wind speed prediction. The wind speed prediction measurement model is a complex system affected by many factors. This method can only describe parts or fragments of information, and it is difficult to comprehensively and accurately understand its change rules and predict.
At present, weather prediction mainly uses basic physical processes to form mathematical models for prediction research [26]. These traditional linear models are difficult to predict with better prediction accuracy for nonlinear time-series predictions. In recent years, artificial neural networks have been widely used in the field of weather forecasting and have achieved good forecasting results. When predicting problems, combining the neural network and the least square method can make up for the shortcomings of the least square method, and make full use of the advantages of both to accurately predict the average minimum temperature in the future. At the same time, the effective combination of neural network prediction and the least square method can solve the problem of the nonlinear system well. Niu Zhijuan et al. [27] used the basic principles of the BP neural network and the least square method to establish a network model. This model is used to predict extreme temperatures. They convert the nonlinearity into a linear optimization method and calculate the network weights by using the least square method to avoid local minimums. The value problem lies in the training process of the traditional BP network. The experimental results show that the least-squares-optimized BP neural network has a higher prediction accuracy than a single BP neural network when predicting the monthly average minimum temperature, and it reduces the complexity of the prediction method. The fitting and prediction accuracy of the model in this paper is better than the error and accuracy of the single BP neural network, which can better meet the needs of actual prediction. The results show that the BP neural network optimized by the least-squares method has better generalization ability, the prediction of the average minimum temperature is more stable, and the prediction accuracy is higher.

K Dou et al. [28] proposed a long-term weather forecast based on GA-BP neural network. This method uses genetic algorithms to optimize the weights and thresholds of the BP neural network and improves the drawbacks of the BP neural network. For example, it is not sensitive to weights and thresholds. Li B et al. [29] proposed multi-dimensional research on agro-meteorological disasters based on the gray BP neural network. Based on the gray BP neural network, they used the gray correlation analysis method to analyze the correlation between the main agro-meteorological disaster factors and the output of food crops in Henan Province, China.

3.2. Application of Recurrent Neural Network in Extreme Weather

The recurrent neural network is a kind of directed loop formed by connections between units. The artificial neural network is a natural extension of a feedforward neural network in the sequence data processing. The difference is that its input not only includes the input examples currently seen. It also includes the information perceived by the network at the last minute [30]. Using this attribute, information can circulate in the network for any length of time.

The recurrent neural network is a type of pattern used to recognize text, sequence data, such as speech, or digital input time-series data generated by sensors or stock markets. The recursive neural network fully considers the characteristics of the motion state of the time series and improves the self-learning adaptive ability through network learning, so that it can better process sequential data. Recurrent neural networks are suitable for sequence modeling. Accurate forecasting of storm surge can greatly reduce casualties and economic losses and has important practical value. Leisen et al. [31] proposed a storm surge model water increase prediction method based on a regression neural network. In this paper, the storm surge timing data are processed specifically, and a recurrent neural network with a reasonable structure is designed to complete the preprocessing measurement of timing data. Compared with the traditional BP neural network, the recurrent neural network can better deal with the prediction of time series data. This method is applied to the prediction of water increase in Weifang Water Station, and the results show that, compared with BP neural network, the recurrent neural network can obtain better prediction results with smaller errors.
Zhang Shuai et al. [32] studied three models, namely, feedforward neural network model (fuzzy neural network), wavelet neural network model (algorithm), and integrated autoregressive moving average, which were selected to verify and compare recurrent neural network models. Additionally, the long- and short-term memory (LSTM) unit weight and the updated network solution’s time backpropagation algorithm in the long run. The simulation results show that the recurrent neural network model replaces other models, the training result is close to the actual value, and the prediction accuracy is high. As shown in Figure 4, the RNN standard network consists of three layers: input layer, hidden layer, and output layer. An LSTM memory cell is shown in Figure 5. A memory block is mainly composed of four parts: input gate, self-connected neuron, forget gate, and output gate. These gates are used to modulate the interaction between the storage unit itself and the environment. In this way, the state of the storage unit can remain unchanged from one moment to another.

Shi X et al. [33] proposed Convolutional LSTM (ConvLSTM). They built an end-to-end trainable model for precipitation nowcasting problems by stacking multiple ConvLSTM layers and forming a coding prediction structure. By extending the Fully connected LSTM (FC-LSTM), both the input state and the output state have a convolution structure state-to-state transition. In order to establish a good temporal and spatial relationship model, they extended the concept of FC-LSTM to ConvLSTM. Both state transition and state-to-state transition have a convolution structure.
In ConvLSTM, all the inputs $X_1 \ldots X_t$, cell outputs $C_1 \ldots C_t$, hidden state $H_1 \ldots H_t$, and gates $i_t, f_t, g_t$, other 3D tensors in $R^P \times M \times N$, where the first dimension is either the number of measurement (for inputs) or the number of feature maps (for intermediate representations), and the last two dimensions are spatial dimensions ($M$ rows and $N$ columns). The key equations of ConvLSTM are shown as follows [33]:

$$g_t = \tanh \left( W_{xg} * X_t + W_{hg} * H_{t-1} + b_g \right)$$  \hspace{1cm} (3)

$$i_t = \sigma \left( W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \odot C_{t-1} + b_i \right)$$  \hspace{1cm} (4)

$$f_t = \sigma \left( W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \odot C_{t-1} + b_f \right)$$  \hspace{1cm} (5)

$$C_t = f_t \odot C_{t-1} + i_t \odot g_t$$  \hspace{1cm} (6)

$$o_t = \sigma \left( W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \odot C_t + b_o \right)$$  \hspace{1cm} (7)

$$H_t = o_t \odot \tanh (C_t),$$  \hspace{1cm} (8)

where $\sigma$ is sigmoid activation function, $*$ and $\odot$ denote the convolution operator and the Hadamard product, respectively. The use of the input gate $i_t$, forget gate $f_t$, output gate $o_t$, and input-modulation gate $g_t$ controls information flow across the memory cell $C_t$. In this way, the gradient will not be trapped in the memory and disappear quickly.

Han F [34] applied RNN to nowcasting. By using PredRNN (Predictive RNN), by modeling historical radar data, the radar echo within the next hour can be predicted. The purpose of spatiotemporal sequence prediction learning is to generate future images through the learning of historical frames, in which spatial representation and temporal changes are the two key structures of historical frames. The core of this network is a new Spatio-temporal LSTM (ST-LSTM) unit, which simultaneously extracts and memorizes spatial and temporal representations. PredRNN achieves the most advanced prediction performance on three video prediction data sets. It is a more general framework and can be easily extended to other prediction learning tasks through integration with other architectures. Wang Y et al. [35] proposed PredRNN++, which is a recursive network for spatiotemporal prediction learning. In pursuit of powerful short-term video dynamic modeling capabilities, they made our network deeper by using a new loop structure called Causal LSTM and cascaded dual memory. PredRNN with the ST-LSTM is shown in Figure 6 [35].

![Figure 6. PredRNN with the ST-LSTM [35].](image-url)
Where \( C^k_t \) is the temporal memory, and \( M^k_t \) is the spatial memory, where the \( t \) denotes the time step. The superscript \( k \) represents the hidden layer in the overlay causal LSTM network. The current temporal memory is directly dependent on its previous state \( C^k_{t-1} \), and \( f_t \) is a forget gate, \( i_t \) is an input gate, and \( g_t \) is an input modulation gate. The current spatial memory \( M^k_t \) depends on \( M^{k-1}_{t-1} \) in the deep transition path. Update equations of the causal LSTM at the \( k \)th layer can be presented as follows [35]:

\[
\begin{align*}
\begin{pmatrix} g_t \\ i_t \\ f_t \end{pmatrix} &= \begin{pmatrix} \tanh \\ \sigma \\ \sigma \end{pmatrix} W_1 \ast \left[ X_t, H^k_{t-1}, C^k_{t-1} \right] \\
C^k_t &= f_t \odot C^k_{t-1} + i_t \odot g_t \\
\begin{pmatrix} g'_t \\ i'_t \\ f'_t \end{pmatrix} &= \begin{pmatrix} \tanh \\ \sigma \\ \sigma \end{pmatrix} W_2 \ast \left[ X_t, C^k_t, M^{k-1}_t \right] \\
M^k_t &= f'_t \tanh( W_3 \ast M^{k-1}_t ) + i'_t \odot g'_t \\
o_t &= \tanh( W_4 \ast \left[ X_t, C^k_t, M^k_t \right] ) \\
H^k_t &= o_t \odot \tanh( W_5 \ast \left[ C^k_t, M^k_t \right] )
\end{align*}
\]

where \( \ast \) is convolution, \( \odot \) is the element-wise multiplication, \( \sigma \) is the element-wise Sigmoid function, the square brackets denote the series of tensors, and the round brackets denote the system of equations. \( W_{1-5} \) are convolutional filters, where \( W_3 \) and \( W_5 \) are \( 1 \times 1 \) convolutional filters. The final output \( H^k_t \) is co-determined by the dual memory states \( M^k_t \) and \( C^k_t \).

Compared with the shallow neural network model, the recurrent neural network model has better performance. Compared with other models, the recurrent neural network model is better. The recurrent neural network model can better describe the dynamic and nonlinear changes of rainfall under extreme weather. The performance of this model is also better than other models, and it is more suitable for rainfall time series modeling.

### 3.3. Application of Convolutional Neural Network in Extreme Weather

Convolutional neural networks [36–39] are widely used in machine learning applications, such as image classification [40,41], natural language [42], and speech processing [43,44]. Convolutional neural network can extract features directly from the original information without complicated preprocessing; With the combination of network from front to back, features from shallow to deep, removing the limitations of artificial features, convolutional neural network has become a multidisciplinary field in recent years. Convolution neural network has been applied in many meteorological prediction fields. It is also very helpful for extreme weather forecasting. The accuracy of the convolutional neural network algorithm comes at the cost of its huge computational complexity. Convolutional neural networks are usually accelerated on general-purpose processors or graphics processing units. The Convolutional neural network is a deep neural network that can achieve the latest accuracy and high performance for most computer vision tasks. The training of convolutional neural networks requires a lot of calculations and may take several days to complete. Training convolutional neural networks on large data sets require a lot of computation, which has led to a lot of research and development on open source parallel implementations on GPUs. However, there are few studies to evaluate the performance characteristics of those implementations. Amodei D [45] proposed an end-to-end deep learning framework that can be used to predict various extreme weather events. The framework is based on diffusion graph convolutional recursive neural networks and transmission elastic dynamic capture algorithms. The model aims to obtain the temporal and spatial dependence of transportation elasticity from the directed graph at the same time.
In order to achieve this goal, the temporal and spatial prediction based on the topological information of the urban road network; the use of big data in the real world to quantitatively study the spatial and temporal characteristics of urban road elasticity. In terms of calculation, the convolution operation is also used to replace the pooling layer [46], as shown in Figure 7, the convolutional neural network model considers the data of dozens of extreme weather events, and not just case studies of certain disasters.

![Figure 7. Architecture of the DS2 system used to train on both English and Mandarin speech. We explore variants of this architecture by varying the number of convolutional layers from 1 to 3 and the number of recurrent or GRU layers from 1 to 7 [46].](image)

Chen Qiaote et al. [47] invented a time-period telemetry wind speed prediction method, which can predict extreme windy weather conditions. This method is based on convolutional neural networks. It is characterized by first constructing the input characteristic mapping of the model, and then establishing a depth-based volume based on the prediction model of the product neural network, and finally, the prediction model is established based on the time-period telemetry wind speed prediction. The advantage is that the sliding window is used to predict the data structure characteristics of the two-dimensional map from the historical data and the numerical weather forecast model. This form of input data retains the original data sequence information and can participate in the convolution operation to establish a prediction model. The one-dimensional convolutional neural network is used to extract the shallow and local characteristics of adjacent meteorological variables in the time domain, and the two-dimensional convolutional neural network is used to explore the potential and abstract feature information in the shallow local characteristics and return to the prediction data layer to provide effective deep features to improve the overall performance of the model.

The scattered echo images of the new generation of Doppler weather radar are affected by non-rainfall echoes, resulting in a decrease in the accuracy of refined short-term weather forecasts. Shin H C et al. [48] proposed a radar noise image semantic segmentation method based on deep convolutional neural network (DCNN). First, they designed a deep-use convolutional neural network model (DCNNM) to train the data, and extract feature images through the process of forwarding propagation to extract high-dimensional global semantic information and local feature details. Then, use the training error value to iteratively update the network parameters of the backpropagation to achieve the optimization of model collection and convergence effects. Finally, use this model to segment and process weather radar image data. This method can more effectively process extreme
weather data. Experimental results show that this method is effective for weather radar images [49]. Compared with optical flow methods, full convolutional network (FCN), and other methods, this method has a better effect on real echoes and noise echoes in weather radar images. The noise effect recognition accuracy is high, and the image pixel accuracy is high. Wang, S et al. [50] proposes a radar quantitative precipitation estimation algorithm based on the temporal and spatial network model (ST-QPE) and designs the convolutional time-series network QPE-Net8 and the multi-scale feature fusion time-series network QPE-Net22.

3.4. Application of Capsule Neural Network in Extreme Weather

Geoffrey Hinton is one of the pioneers of deep learning and the inventor of classic neural network algorithms such as back propagation. He and his team proposed a brand-new neural network based on a structure called a capsule and also published a dynamic routing algorithm between capsules for training capsule networks.

The artificial neuron outputs a single scalar. The convolutional network uses the convolution kernel to superimpose the calculation results of the same convolution kernel in each region of the two-dimensional matrix to form the output of the convolution layer. The maximum angle in variance achieved by the pooling method, because the maximum pool continues to look for the area of a two-dimensional matrix and the largest number in the selected area, so it meets the activity invariance we want (that is, we slightly adjust the input and the output remains unchanged). In other words, we slightly change the detection of the object in the input image, and we can still detect that the object model pooling layer has lost valuable information, and does not consider the relative spatial relationship between the encoded features, so we should use capsules. In capsule detection, all important information about the characteristic state will be encapsulated as a vector (neurons are scalars).

Capsule network is a novel and promising neural network in the field of deep learning. By encoding features into capsules, establishing partial and overall relationships, it shows good performance in image classification. However, the original capsule network has weak feature extraction capabilities, many training parameters, and has the characteristics of explaining all the content in the image, so it is not suitable for complex background images. To solve the above problems, S. Yang et al. [51] proposed an advanced capsule network RS-Capsnet, which uses Res2Net blocks to extract multi-scale features, uses squeeze (SE) blocks to highlight useful features and suppress useless features. At the same time, the linear combination method between the capsules is adopted to improve the ability of the capsules to express the object to be detected and reduce the number of capsules. In addition, they also proposed a method that first constructs an intermediate capsule that can represent most of the subjects under test, and then uses the intermediate capsule and the main capsule together to construct a classification capsule. Ashesh et al. [52] believe that numerical weather prediction models need to increase computing time and resources, but extreme weather is sometimes difficult to predict. They introduced a data-based framework, based on simulation predictions (using similar patterns from the past), and new deep learning pattern recognition techniques capsule neural networks, CapsNets, and influence-based automatic labeling strategies. Utilizing fully coupled earth from a large collection, the data of the system model, CapsNets, is trained on the large-scale tropospheric circulation model (Z500), the model number is 0–4, depending on the existence of extreme temperature on the North American surface and the geographical area in the next few days. The trained network can only be used Z500 predicts the area of cold wave or heatwave that occurred one to five days ago. Figure 8 shows the Distribution of extreme weather for summer and winter [52].
Two state-of-the-art deep learning techniques are used for pattern recognition: convolutional neural network and more advanced capsule neural network. Compared with traditional image processing techniques, the main advantage of these two methods is that the feature extraction filter of each data set is learned through an algorithm called back propagation, rather than manually designed and specified in advance. Since 2011, Convolutional Neural Network has become a breakthrough method, it has changed the way of image processing, but because CapsNet has the characteristics of equal variables discussed later, it is expected to be better used in our time and space weather data.

4. Conclusions

It can be seen from the introduction of Sections 2 and 3 that the application of deep learning in extreme weather forecasting is extensive and effective and has great potential for further improving the accuracy of forecasting. Through the combination of deep learning and meteorological science, researchers have drawn more conclusions and contributed to better weather forecasting in the future, using deep learning to expand more ways to help predict extreme weather problems.

The following Table 1 summarizes the entire article.

Figure 8. Distribution of extreme weather for summer and winter [52].
Table 1. Summary of deep learning and its application in extreme weather methods.

| Method                        | Specific Method                                                                 | Generalize                                                                                           | Features                                                                                      |
|-------------------------------|---------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| Traditional Methods           |                                                                                  |                                                                                                      |                                                                                                |
| Weather Map Forecast Method   | The first step of the weather map forecasting method is to predict the weather conditions, that is, predict the existing weather system and its future trends on the weather map. | Simple and intuitive                                                                                 |                                                                                                |
| Weather Forecast Based on     | Radar can emit short-wavelength radio waves from the sky-white line. When it encounters weather phenomena in the distant atmosphere such as typhoons, thunderstorms, and rainstorms, its radio waves will be reflected back and displayed on the radar screen. | Weather radar has a wide range of functions, providing necessary weather information for weather forecasting, rocket, missile and spacecraft launch and flight. |                                                                                                |
| Weather Radar                 |                                                                                  |                                                                                                      |                                                                                                |
| Numerical Prediction methods  | Numerical weather forecasting is a set of equations describing fluid mechanics and weather evolution thermodynamics. A large computer performs numerical calculations under certain initial and boundary conditions according to the actual conditions of the atmosphere. This is a way to predict atmospheric movement in a certain period in the future, and weather phenomena. | A typical nonlinear system requires a lot of calculation. In order to complete all calculations in a shorter time than the actual weather evolution. |                                                                                                |
| BP Neural Network             | BP neural network is a multi-layer feed-forward neural network. The main feature of the network is the forward transmission of signals and the backward propagation of errors. | Strong self-study adaptability Generalization and strong fault tolerance                              | Its input not only includes the input examples currently seen. It also includes the information perceived by the network at the last minute. Using this attribute, information can circulate in the network for any length of time. |
| Recurrent Neural Network      | Recurrent neural network is a type of pattern used to recognize text, sequence data. |                                                                                                      | Researchers do not need to care about specific weather conditions, only need to understand the results after training, and realize the encapsulation of feature extraction. |
| Deep learning applications    |                                                                                  |                                                                                                      |                                                                                                |
| Convolutional Neural Network  | From the input of original information, self-learning features, along with the network from front to back, the combination of features, from shallow to deep. The research focus of this subject area. |                                                                                                      |Researchers do not need to care about specific weather conditions, only need to understand the results after training, and realize the encapsulation of feature extraction.|
| Capsule Neural Network        | The artificial neuron outputs a single scalar. The convolutional network uses the convolution kernel to superimpose the calculation results of the same convolution kernel in each region of the two-dimensional matrix to form the output of the convolution layer. | (1) A small amount of data can be used to learn a good representation effect; (2) The way of thinking is closer to the human brain, and the hierarchical relationship of internal knowledge representation is better modeled in the neural network. The intuition behind this capsule is very simple and elegant. |                                                                                                |

5. Discussion

Extreme weather prediction has always been a scientific problem faced by meteorologists all over the world, and it is particularly difficult to predict extreme weather events. At present, the level and technology of monitoring and predicting extreme weather events are relatively limited, and the irregularities and mechanisms of extreme weather events are still unclear. Therefore, in the future, it is necessary to study the statistical characteristics of extreme weather events first and carry out a relatively complete arrangement of the
effective data set, including the frequency and intensity of events, seasonal changes, and the characteristics of annual weather changes. Additionally, study the main factors that cause extreme weather events, combine deep learning and artificial intelligence, and other broader fields to conduct more in-depth research and prediction of extreme weather.

**Author Contributions:** Investigation, W.F., Q.X., L.S. and V.S.S.; Supervision, L.S. and V.S.S.; Writing—original draft, W.F. and Q.X.; Writing—review and editing, W.F. and Q.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (42075007) and the Open Grants of the State Key Laboratory of Severe Weather (No.2021LASW-B19).

**Acknowledgments:** This work was supported by the National Natural Science Foundation of China (Grant No.42075007), the Open Grants of the State Key Laboratory of Severe Weather (No.2021LASW-B19), the Open Project of Provincial Key Laboratory for Computer Information Processing Technology under Grant KJS1935, Soochow University, and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

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

**References and Notes**

1. Xu, Y. Characteristics and Psychological Response in the Post-disaster Era. *Logist. Chin. Gov.* 2020, 68–69. [CrossRef]
2. Hinton, G.E.; Osindero, S.; Teh, Y.W. A fast learning algorithm for deep belief nets. *Neural Comput.* 2006, 18, 1527–1554. [CrossRef] [PubMed]
3. Bengio, Y. Learning Deep Architectures for AI. *Found. Trends® Mach. Learn.* 2009, 2, 1–127. [CrossRef]
4. Yin, W.; Zhang, D.; Yan, J.; Zhang, C.; Li, Y.; Rui, X. Big data air pollution forecast based on deep learning. *China Environ. Manag.* 2015, 7, 46–52.
5. He, J.; Tang, W.; Zhou, Y.; Shen, W. Opportunities and challenges of artificial intelligence in meteorological sciences. *China Inf. Technol.* 2019, 379, 79–81.
6. Song, L.; Li, W. Development of Comprehensive Meteorological Observation System. *Meteorol. Mon.* 2008, 34, 3–9.
7. Lou, M.; Hao, S.; Chen, Y.; Lou, X. Design and application of a weather map for forecasting cloud and snowfall. *Zhejiang Meteorol.* 2019, 4.
8. Sokol, Z.; Szturc, J.; Orellana-Alvear, J.; Popova, J.; Jurczyk, A.; Celleri, R. The Role of Weather Radar in Rainfall Estimation and Its Application in Meteorological and Hydrological Modelling—A Review. *Remote Sens.* 2021, 13, 351. [CrossRef]
9. Hu, L.; Zhuang, K.; Xiang, X.; Yang, H.; Zhao, K. A weather display system and disaster warning system based on radar data [P]. Zhejiang Province: CN20192114217.6 [P]. 2020-08-28.
10. Li, S.; Li, G.; Wang, X.; Li, C.; Liu, H.; Li, G. Precipitation Characteristics of an Abrupt Heavy Rainfall Event over the Complex Terrain of Southwest China Observed by the FY-4A Satellite and Doppler Weather Radar. *Water* 2020, 12, 2502. [CrossRef]
11. Xie, B.; Xu, J.; Jung, J.; Yun, S.; Zeng, E.; Brooks, E.; Dolk, M.; Narasimhalu, L. Machine Learning on Satellite Radar Images to Estimate Damages after Natural Disasters. In *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*, Seattle, WA, USA, 3–6 November 2020; pp. 461–464.
12. Ge, R. *Discussion on the Calibration of Rain Measuring Radar and Z-Value Measurement Error;* Radar Meteorological Collection; Institute of Central Meteorological Bureau: Beijing, China, 1977; pp. 86–94.
13. Lin, Z. *On the Errors of Several Selected Z-I Relations Methods;* Radar Meteorological Collection; Central Meteorological Bureau Research Institute: Beijing, China, 1977; pp. 95–103.
14. Bi, Y.; Liu, J.; Duan, S.; Lu, D.; Su, D.; Chen, Y. Attenuation correction of X-band dual linear polarization weather radar reflectivity. *Atmos. Sci.* 2012, 36, 495–506.
15. Goudenhooft, E.; Delobbe, L.; Willems, P. Regional frequency analysis of extreme rainfall in Belgium based on radar estimates. *Hydrol. Earth Syst. Sci.* 2017, 21, 5385–5399. [CrossRef]
16. Editorial Board of Dictionary of Atmospheric Sciences. *Dictionary of Atmospheric Sciences;* Meteorological Press: Beijing, China, 1994; p. 624.
17. Jun, D.; Grumm, R.H.; Guo, D. The “ensemble anomaly forecasting method” for forecasting abnormally extreme and high-impact weather: Taking Beijing 21 July 2012 as an example. *Atmos. Sci.* 2014, 38, 685–699.
18. Heng, B.C.P.; Tubbs, R.; Huang, X.Y.; Macpherson, B.; Barker, D.; Boyd, D.F.A.; Kelly, G.; North, R.; Stewart, L.; Webster, S.; et al. SINGV-DA: A data assimilation system for convective-scale numerical weather prediction over Singapore. *Q. J. R. Meteorol. Soc.* 2020, 146, 1923–1938. [CrossRef]
19. Du, J. *Neural Network;* Xidian University: Xi’an, China, 2007.
20. Zhou, H.; Chen, Y.; Zhang, S. Ship Trajectory Prediction Based on BP Neural Network. *J. Artif. Intell.* 2019, 1, 29. [CrossRef]
21. Xia, P.; Xu, A.; Tong, L. Analysis and Prediction of Regional Electricity Consumption Based on BP Neural Network. *J. Quantum Comput.* 2020, 2, 25.
22. Chen, W. Research on the location and capacity of electric vehicle charging stations in urban areas based on queuing theory. Logist. Eng. Manag. 2016, 38, 189–191.

23. Xi, J.; Quan, H.; Liu, Z.; Liu, Z.; Xu, B.; Chen, Z.; Zhang, H.; Wu, G. Two-stage optimization method for charging station service radius and scale based on queuing theory. Heilongjiang Electr. Power 2012, 34, 446–449.

24. Zhao, S.; Li, Z.; Dang, L. Optimal location and capacity of electric vehicle charging stations based on urban traffic network information. Electr. Power Autom. Equip. 2016, 36, 8–15.

25. Kala, S.; Paul, D.; Jose, B.R.; Mathew, J.; Nalesh, S. Performance Analysis of Convolutional Neural Network Models. In Proceedings of the 2019 9th International Conference on Advances in Computing and Communication (ICACC), Kochi, India, 6–7 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 22–26.

26. Liu, G.; Mao, N. Research on temperature stochastic model and my country’s temperature option pricing. Math. Stat. Manag. 2008, 27, 959–967.

27. Niu, Z.; Hu, H. Application of BP neural network based on least square method in extreme temperature prediction model. Green Technol. 2015, 17–19. [CrossRef]

28. Dou, K.; Sun, X. Long-Term Weather Prediction Based on GA-BP Neural Network. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Beijing, China, 11–13 December 2020; IOP Publishing: Bristol, UK, 2021; Volume 668, p. 012015.

29. Li, B.; Zhang, S. Multidimensional research on agrometeorological disasters based on grey BP neural network. Grey Syst. Theory Appl. 2020. [CrossRef]

30. Liu, G.; Huang, C.; Ding, J. Artificial neural network model for comprehensive evaluation of water quality. China Environ. Sci. 1998, 18, 514–517.

31. Lei, S.; Shi, Z.; Shi, T.; Gao, S.; Li, Y.; Zhong, S. Storm surge water increase prediction based on recurrent neural network. J. Intell. Syst. 2017, 12, 640–644.

32. Zhang, S.; Wei, Z.; Zhang, Y. Application research of recurrent neural network in rainfall prediction. Water Sav. Irrig. 2017, 63–66+71.

33. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; Woo, W.-C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Adv. Neural Inf. Process. Syst. 2015, 28, 802–810.

34. Han, F. The Application of Recurrent Neural Network to Nowcasting. J. Appl. Appl. Meteorology 2019, 30, 61–69. [CrossRef]

35. Wang, Y.; Gao, Z.; Long, M.; Wang, J.; Yu, P.S. Predrm++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 5123–5132.

36. Sahiner, B.; Chan, H.P.; Petrick, N.; Wei, D.; Helievic, M.A.; Adler, D.D.; Goodsitt, M.M. Classification of mass and normal breast tissue: A convolution neural network classifier with spatial domain and texture images. IEEE Trans. Med Imaging 1996, 15, 598–610. [CrossRef] [PubMed]

37. Chang, L.; Deng, X.; Zhou, M.; Wu, Z.-K.; Yuan, Y.; Yang, S.; Wang, H.-A. Convolutional Neural Network in Image Understanding. Acta Autom. Sin. 2016, 42. [CrossRef]

38. Yang, H.; Wang, F. Semantic segmentation method of weather radar noise image based on deep convolutional neural network. J. Electron. Inf. 2019, 41, 2373–2381.

39. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Commun. ACM 2017, 60, 84–90. [CrossRef]

40. Chen, R.; Pan, L.; Li, C.; Zhou, Y.; Chen, J.; Beckeman, E. An Improved Deep Fusion CNN for Image Recognition. CMC Comput. Mater. Contin. 2020, 65, 1691–1706. [CrossRef]

41. Zhang, X.; Zhao, J.; LeCun, Y. Character-level convolutional networks for text classification. Adv. Neural Inf. Process. Syst. 2015, 28, 649–657.

42. Sainath, T.N.; Mohamed, A.; Kingsbury, B.; Ramabhadran, B. Deep convolutional neural networks for LVCSR. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 8614–8618.

43. Limin, W.; Xuming, H.; Qiang, J. Semi-supervised affinity propagation clustering algorithm based on fireworks explosion optimization. In Proceedings of the 2014 International Conference on Management of e-Commerce and e-Government, Shanghai, China, 31 October–2 November 2014; pp. 273–279.

44. Zhou, Y.T.; Chellappa, R.; Vaid, A.; Jenkins, B.K. Image restoration using a neural network. IEEE Trans. Acoust. Speechand Signal Process. 1988, 36, 1141–1151. [CrossRef]

45. Amodei, D.; Ananthanarayanan, S.; Anubhai, R.; Bai, J.; Battenberg, E.; Case, C.; Casper, J.; Catanzaro, B.; Cheng, Q.; Cheng, G.; et al. Deep speech 2: End-to-end speech recognition in English and Mandarin. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 20–22 June 2016; pp. 173–182.

46. Tao, R.; Zhang, Y.; Wang, L.; Cai, P.; Tan, H. Detection of Precipitation Cloud over the Tibet Based on the Improved U-Net. CMC Comput. Mater. Contin. 2020, 65, 2455–2474. [CrossRef]

47. Chen, Q.; He, C.; Fu, R.; Zhou, Y.; Jin, W. A multi-time wind speed prediction method based on deep convolutional neural network. Zhejiang, China: CN110991690A, 2020-04-10.
48. Shin, H.C.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298. [CrossRef]

49. Wang, Y.; Feng, Y.; Cai, J.; Hu, S. An approach for radar quantitative precipitation estimate based on categorical ZI relations. *J. Trop. Meteorol.* **2011**, *27*, 601–608.

50. Wang, S.; Yu, X.; Liu, L.; Huang, J.; Wong, T.H.; Jiang, C. An Approach for Radar Quantitative Precipitation Estimation Based on Spatiotemporal Network. *Cmc Comput. Mater. Contin.* **2020**, *65*, 459–479. [CrossRef]

51. Yang, S.; Lee, F.; Miao, R.; Cai, J.; Chen, L.; Yao, W.; Kotani, K.; Chen, X. RS-CapsNet: An Advanced Capsule Network. *IEEE Access* **2020**, *8*, 85007–85018. [CrossRef]

52. Chattopadhyay, A.; Nabizadeh, E.; Hassanzadeh, P. Analog forecasting of extreme-causing weather patterns using deep learning. *J. Adv. Modeling Earth Syst.* **2020**, *12*, e2019MS001958. [CrossRef]