


# Machine Learning Methods for Weather Forecasting: A Survey

Huijun Zhang <sup>1,\*</sup> , Yaxin Liu <sup>1</sup>, Chongyu Zhang <sup>1</sup> and Ningyun Li <sup>2</sup>

<sup>1</sup> China Huaneng Clean Energy Research Institute, Beijing 102209, China; yx\_liu@qny.chng.com.cn (Y.L.); cy\_zhang2@qny.chng.com.cn (C.Z.)

<sup>2</sup> Beijing Big Data Center, No. 3 Courtyard, Liuzhuang Road, Tongzhou District, Beijing 101117, China; lny14@tsinghua.org.cn

\* Correspondence: hj\_zhang@qny.chng.com.cn

**Abstract:** Weather forecasting, a vital task for agriculture, transportation, energy, etc., has evolved significantly over the years. Comprehensive surveys play a crucial role in synthesizing knowledge, identifying trends, and addressing emerging challenges in this dynamic field. In this survey, we critically examines machine learning (ML)-based weather forecasting methods, which demonstrate exceptional capability in handling complex, high-dimensional datasets and leveraging large volumes of historical and real-time data, enabling the identification of subtle patterns and relationships among weather variables. Research on specific tasks such as global weather forecasting, downscaling, extreme weather prediction, and how to combine machine learning methods with physical principles are very active in the current field. However, several unresolved or challenging issues remain, including the interpretability of models and the ability to predict rare weather events. By identifying these gaps, this research provides a roadmap for advancing machine learning-based weather forecasting techniques to complement and enhance weather prediction results.

**Keywords:** machine learning; weather forecasting; deep learning; survey



Academic Editors: Erick G. Sperandio Nascimento, Taciana Toledo De Almeida Albuquerque and Prashant Kumar

Received: 13 December 2024

Revised: 5 January 2025

Accepted: 8 January 2025

Published: 14 January 2025

**Citation:** Zhang, H.; Liu, Y.; Zhang, C.; Li, N. Machine Learning Methods for Weather Forecasting: A Survey. *Atmosphere* **2025**, *16*, 82. <https://doi.org/10.3390/atmos16010082>

**Copyright:** © 2025 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

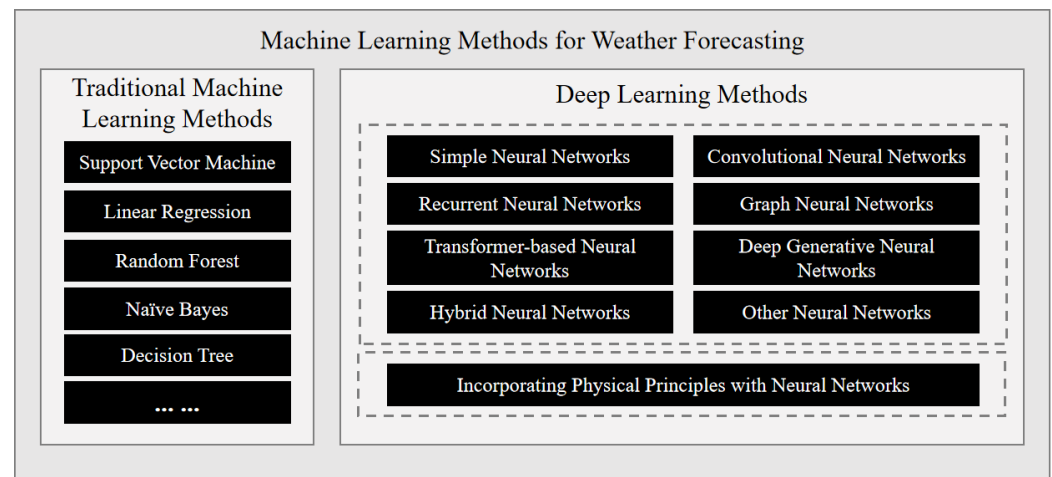
Weather forecasting plays a crucial role in modern society, influencing a wide range of sectors, from agriculture and transportation to energy and public safety [1–9]. Accurate weather forecasts allow for better planning and decision-making, helping industries mitigate risks, optimize resources, and enhance operational efficiency [10]. For example, in the field of renewable energy, accurate weather forecasts can ensure the safety and stability of power systems, benefiting wind and solar power prediction, electricity market transactions, etc. [11]. Furthermore, accurate weather forecasting is essential for disaster preparedness and response, enabling authorities to issue timely warnings for hurricanes, floods, or heatwaves, potentially saving lives and minimizing economic damage [12]. As dramatic climate changes are coming [13], the importance of precise and timely weather forecasts continues to grow, highlighting the need for advanced prediction methods to address these challenges.

Traditional weather forecasting primarily relies on Numerical Weather Prediction (NWP) models [14], which use mathematical equations based on atmospheric physics to simulate and predict weather patterns. However, NWP models face several challenges, including limited understanding of physical mechanisms, difficulties in extracting valuable insights from vast amounts of observational data, and the need for substantial computational resources [15].

Machine learning techniques have gained significant popularity across a wide range of domains due to their ability to handle large and complex datasets, identify intricate

patterns, and make accurate predictions. Machine learning has been widely used in fields such as natural language processing, computer vision, and signal processing, serving industries and scenarios such as biological sciences, transportation, energy, and electricity. This widespread success has naturally extended to weather forecasting, where machine learning models are leveraged to enhance accuracy and efficiency [16]. The ability of ML algorithms to deal with complex, high-dimensional datasets, as well as to learn from large amounts of historical and real-time data [17], allows them to recognize subtle patterns and relationships in weather variables that may not be captured by traditional models.

As shown in Figure 1, machine learning methods for weather forecasting can be divided into two categories, namely traditional machine learning methods and deep learning methods.



**Figure 1.** A categorization of machine learning methods for weather forecasting.

Traditional machine learning methods, such as support vector machines (SVM), linear regression (LR), random forests (RF), etc. have been applied successfully to simple weather forecasting tasks. For instance, SVM and LR have been employed for classifying weather conditions and predicting continuous variables like temperature or precipitation by learning from historical data [18–20]. While these models work well for specific tasks such as classification or regression based on simpler datasets, they struggle with capturing more complex relationships in weather data, particularly when large-scale, high-dimensional inputs are involved [21].

Deep learning methods have gained increasing attention in weather forecasting due to their ability to handle large-scale, high-dimensional datasets and model complex nonlinear relationships. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), etc. are widely used for various weather forecasting tasks, e.g., precipitation nowcasting and forecasting, global and regional multi-factor weather forecasting, and extreme weather events detection. These models have shown promising results in handling the high complexity of weather systems and are increasingly being used to enhance weather forecasting.

This survey provides a comprehensive overview of machine learning applications in weather forecasting, covering a range of approaches from traditional machine learning methods to advanced deep learning techniques. It also introduces and catalogs commonly used datasets and evaluation metrics, offering guidance for future researchers. Furthermore, this paper discusses the challenges as well as the potential for future developments in this field.

By reviewing the current state-of-the-art machine learning methods and their contributions to weather forecasting, this survey provides insights into the evolving landscape of weather forecasting, highlighting the strengths and limitations of these emerging techniques. Through this, we aim to illustrate the potential of machine learning to revolutionize

weather forecasting practices, paving the way for more accurate and timely predictions in the face of a changing global climate.

## 2. A Categorization of Machine Learning Methods for Weather Forecasting

### 2.1. Traditional Machine Learning Methods

Driven by the development of artificial intelligence, major traditional machine learning techniques, such as Support Vector Machine (SVM) and Random Forest (RF), have been widely used in tasks like geoscientific classification, meteorological change, and anomaly detection [22–24]. Focusing on the weather forecasting field, traditional machine learning techniques were applied in weather classification, temperature prediction, wind speed prediction, etc.

Aiming to estimate the weather through predictive analysis, Biswas et al. [25] presented a classifier approach for predicting weather conditions, demonstrating how Naive Bayes and the Chi-square algorithm could be applied for classification. The system functioned as a web application with an effective graphical user interface. Two key functions, i.e., classification (training) and prediction (testing), were performed, with results indicating that these data mining methods were sufficient for weather forecasting. However, the system could only predict the class label (Good/Bad) as weather forecasting results. Another study assessed four classification algorithms, i.e., K Nearest Neighbor (KNN), Gaussian naive Bayes (Gaussian-NB), Gradient Boosting Classifier, and Support Vector Classifier—for predicting weather conditions [18]. Ensemble learning was also explored to improve accuracy. KNN and hybrid ensemble outperformed others in precision and recall, highlighting the potential of machine learning techniques for enhancing weather forecasting. Anderson et al. [26] used random forests to build a statistical model for high-resolution climate predictions, focusing on energy flux and precipitation. By leveraging low-resolution simulations, the approach reduced the need for high-resolution data. It found that global precipitation is more sensitive to resolution changes than model parameters.

Holmstrom et al. [19] investigated machine learning methods for weather forecasting, focusing on maximum and minimum temperatures over seven days using data from the previous two days. They applied a linear regression model and a variant of a functional regression model to capture weather trends. Although both models were outperformed by professional forecasting services, the gap narrowed for longer-term forecasts. The linear regression model performed better, suggesting that two days of data were insufficient for the functional model to detect significant trends. Extending the data window to four or five days could enable the functional model to surpass the linear model. Similarly, Era et al. [27] employed regression algorithms to forecast air temperature. It evaluated Decision Tree, AdaBoost, Random Forest, and Gradient Boosting using meteorological data from 2015 to 2019. The results indicated that the Random Forest Regressor achieved the highest accuracy. Sun et al. [28] compared statistical methods, including linear regression, ML, and DL, with various strategies against numerical models for short-term wind speed forecasts over the Pearl River Estuary (2018–2021). RF and SVM using the direct strategy performed best, and efforts are underway to improve strong wind prediction and ensemble calibration. Zhang et al. [29] introduced a LightGBM-based method for winter precipitation-type prediction using data from 32 national weather stations in northern China (1997–2018). The proposed model achieved an accuracy of 0.83 and an HSS of 0.69, outperforming the ECMWF benchmark. Performance was particularly good at altitudes up to 800 m. Narang et al. [20] focused on improving All Indian Summer Monsoon Rainfall (AISMR) forecasting using recent machine learning techniques. They applied statistical machine learning models, including Linear Regression (LR), Support Vector Regression (SVR), XGBoost, Autoregressive Integrated Moving Average Model (ARIMA), Seasonal AutoRegressive Inte-

grated Moving Average Model (SARIMA), etc. Kühnlein et al. [30] explored using random forests for estimating rainfall rates based on cloud properties from Meteosat Second Generation data. The model accounts for non-linear relationships between predictors and rainfall, focusing on extra-tropical cyclones and various precipitation processes. Rainfall rates were assigned to classified rain areas, using variables like cloud top temperature and water vapor differences. The model showed strong performance with a correlation coefficient of 0.5 for the hourly rainfall rate and a high correlation coefficient of 0.78 for the 8-hourly total rainfall rate, indicating effective prediction under daytime, nighttime, and twilight conditions.

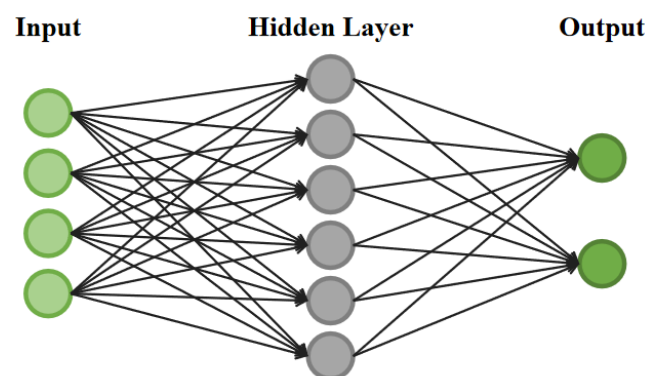
## 2.2. Deep Learning Methods

Traditional machine learning methods rely on manually hand-crafted, domain-specific features to capture temporal dynamics. However, these techniques often concentrate on single variables, missing the interactions among them [16]. In addition, manual feature selection carries risks of error and requires substantial domain expertise. Machine learning models also struggle with large-scale, high-dimensional data, often lacking generalization and being prone to overfitting [21].

Deep learning, however, employs multilayer neural networks that automatically identify complex nonlinear patterns and spatiotemporal dynamics, reducing the reliance on manual feature engineering [31]. These methods are capable of identifying intricate patterns, and achieving superior prediction accuracy and generalization across multiscale and complex environments [32]. Therefore, deep learning methods are well-suited to processing large-scale and high-dimensional weather data, and more and more researchers have paid attention to utilizing deep learning in Earth and atmospheric sciences. In this section, we will introduce various kinds of deep learning methods applied in weather forecasting, including simple neural network, convolutional neural network (CNN), recurrent neural network (RNN), Transformer-based neural network, generative deep learning methods, and other deep learning methods, as well as physics-informed neural networks.

### 2.2.1. Simple Neural Network

In the early days, some simple neural networks were utilized for weather forecasting related tasks, such as temperature forecasting and precipitation nowcasting. An illustration of a simple one-layer neural network is shown in Figure 2.



**Figure 2.** An illustration of a simple one-layer neural network.

Schizas et al. developed an artificial neural network (ANN) model [33] to forecast minimum temperature using meteorological data from Larnaca Airport, Cyprus. They utilized synoptic observations recorded every three hours and extracted various weather elements and additional inputs like astronomical day length and previous minimum temperature. The model was trained and evaluated on 141 days of data from winter

and spring 1984. Baboo et al. employed Back Propagation Neural Network (BPN) for temperature prediction [34]. The BPN method effectively approximated various functions and captured complex temperature influences. The results of real-time data showed that BPN forecasts outperformed both numerical model guidance and official local weather service predictions. Abhishek et al. [35] explored the applicability of ANN in developing reliable nonlinear predictive models for weather analysis, and compared the performance of forecasting maximum temperature over a year with various transfer functions, hidden layers, and neurons. Hossain et al. [36] compared Stacked Denoising Auto-Encoders (SDAE) with a standard multilayer feed-forward network in predicting air temperature using historical data in Northwestern Nevada. The experimental results showed that SDAE performed better in this task and using related weather variables could benefit air temperature prediction. Liu et al. used a stacked auto-encoder to simulate 30 years of hourly weather data [37], leveraging big data principles to enhance prediction accuracy. By automatically learning features through deep models and avoiding traditional assumptions, the approach captured hidden correlations and prevented overfitting, resulting in higher accuracy for time series forecasting. Yonekura et al. developed a deep learning model for very short-term local weather forecasting using dense weather station data [38]. The architecture included point and tensor prediction models to enhance spatial coverage. The experimental results showed the model outperformed state-of-the-art methods like XGBoost and SVM using real observed data.

Grimes et al. presented a neural network for daily rainfall estimation in African river basins using Cold Cloud Duration imagery and numerical weather model data [39]. It showed improved performance over standard methods, especially for higher rainfall amounts, enhancing hydrological forecasting. Capacci et al. investigated the use of ANNs for precipitation diagnostics in the Nimrod forecasting system [40]. ANN processing of MODIS data enhanced precipitation analysis compared to traditional methods. The research identified optimal MODIS channel combinations and proposed future applications with SEVIRI data.

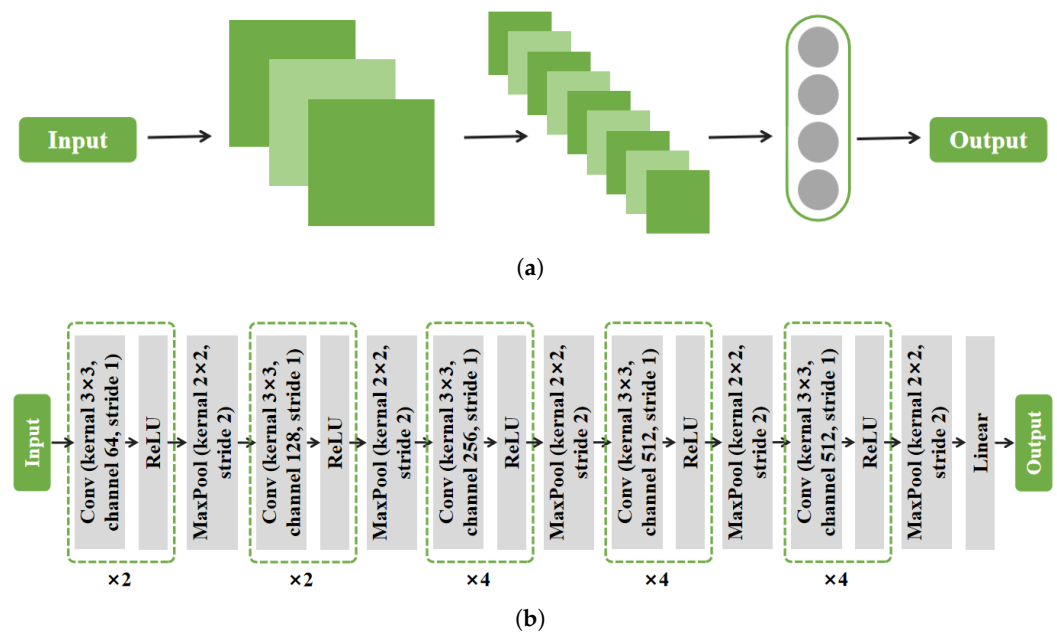
Rivolta et al. focused on nowcasting rainfall using geostationary satellite imagery and neural networks [41]. The process consisted of forecasting the infrared radiance of Meteosat 7 30 min to 1 h in advance and then estimating the rainfall field using a microwave-based calibration algorithm. The accuracy of the method is evaluated by means of error metrics and illustrated with a case study of a rainfall event in central Italy.

However, simple neural networks have limited performance when dealing with complex tasks, which has led to other studies based on following various types of neural networks.

### 2.2.2. Convolutional Neural Network

Since CNNs have demonstrated exceptional ability in efficiently extracting spatiotemporal patterns in various tasks (e.g., video prediction [42,43] and action recognition [44,45]), they have emerged as a powerful tool for weather forecasting, which similarly requires the ability to learn complex spatiotemporal features. As shown in Figure 3, an illustration of a typical CNN framework is provided.





**Figure 3.** Illustrations of CNN. (a) An illustration of a simple CNN; (b) a typical CNN namely VGG-19 [46].

Liu et al. introduced deep CNN models (1D-CNN, 2D-CNN, and 3D-CNN) for predicting temperature and wind speed using spatiotemporal weather data [47]. Applied to datasets from the Netherlands, Belgium, and Denmark, the models demonstrated improved prediction accuracy for multiple locations by learning shared representations through convolutional operations, outperforming traditional neural networks. Weyn et al. developed deep CNN models to predict fundamental meteorological fields over the Northern Hemisphere using historical data without physical processes [48]. The CNNs outperformed basic models at lead times of up to 3 days and could forecast weather system intensity changes. Although not outperforming operational models, they showed promise for efficient ensemble forecasting. Further, a study proposed the STConvS2S model [49], a deep learning architecture using only convolutional layers to learn both spatial and temporal data dependencies. It performed well in forecasting air temperature and rainfall in South America. The model was 23% more accurate and five times faster in training than RNN-based models. Ali et al. [50] employed convolutional neural networks (CNNs) to evaluate the impact of geopotential height at various tropospheric levels on the predictability of extreme surface temperature ( $t_{2m}$ ). They analyzed temperature forecasts for the continental United States at lead times ranging from 1 to 30 days, using geopotential heights at Z100, Z200, Z500, Z700, and Z925 hPa as inputs to the CNN. Gao et al. introduced the Multi-Scale Large Kernel Spatiotemporal Attention Neural Network (MSLKSTNet) [51] to model temperature variations across different temporal and spatial scales. The model consists of three main modules: a feature encoder, a multi-scale spatiotemporal translator, and a feature decoder. Its core component, Multi-scale Spatiotemporal Attention (MSSTA), captures spatial features at multiple scales and their evolution over time. Some studies used UNet [52], a fully convolutional network, as a backbone and improved it to perform better in spatiotemporal meteorological prediction tasks. Zhang et al. developed a UNet-based fully convolutional neural network called STD-UNet [53] to predict refined grid wind speed and direction in China. The model effectively extracts spatiotemporal characteristics and fits wind data using a hybrid loss function. The model successfully predicted vector wind speed in Northwest China and North America using data from Northeast China. Efficiency tests indicated that STD-UNet required only 26 ms for 24-h forecasts, making it an effective tool for wind power centers in ultra-short-term and short-term deployment planning. Xu et al. introduced Hierarchical UNet (HU-Net) [54], which combines

global and local pattern extraction using adaptive Fourier neural operators, self-attention, and convolution operations. Another study introduced the spatiotemporal meteorological prediction UNet (STMP-UNet) [21], which combines an attention-optimized spatial UNet and a multiscale temporal pyramid gating module (TPGM) to capture meteorological features. Trained on Northeast China data, it effectively predicted Pacific conditions, demonstrating transportability. STMP-UNet achieved five-day weather predictions in just 38 ms, showcasing its superior efficiency.

Precipitation nowcasting and forecasting as well as rainfall application are also widely explored by researchers using CNN-based methods. Qiu et al. proposed a multi-task convolutional neural network [55] for short-term rainfall prediction using weather features from multiple observation sites. The model effectively extracted features and utilized correlations between sites, marking the first use of multi-task learning for this purpose. Experimental results demonstrated significant improvements over several baseline models, including ECMWF. A convolutional deep belief network [56] was proposed for predicting rainfall runoff, addressing the challenges of locality and nonlinearity in hydrology. Tested in Guangdong's Luo River Basin, the model demonstrated superior prediction accuracy compared to the traditional Xinanjiang rainfall runoff model, especially for 1-day, 3-day, and 5-day forecasts. Haidar et al. [57] introduced a deep convolutional neural network for predicting monthly rainfall in eastern Australia, marking its first application in this context. The CNN outperformed the Australian Community Climate and Earth-System Simulator (ACCESS) and a multi-layered perceptron (MLP) in key metrics, demonstrating promise for agricultural and industrial applications. A modified Deep Learning Weather Prediction using Cubed Spheres model (MDLWP-CS) [58] was proposed for spatiotemporal weather prediction, specifically targeting precipitation using a multivariate setup. Using 2 m surface air temperature as input, MDLWP-CS demonstrated improved fidelity compared to linear regression and matched the Global Forecast System (GFS) in daily precipitation prediction accuracy with a one-day lag. Shreya et al. [59] explored the use of deep learning for higher-resolution precipitation nowcasting, specifically targeting  $1 \text{ km} \times 1 \text{ km}$  high-resolution 1-hour short-term precipitation predictions. They approached forecasting as an image-to-image translation problem, employing the widely used UNET convolutional neural network. Their results showed that this method outperformed three traditional models: optical flow, persistence, and NOAA's one-hour HRRR numerical nowcasting predictions. Pan et al. [60] introduced a convolutional neural network model to enhance statistical downscaling (SD) methods for daily precipitation prediction, using variables directly resolved by atmospheric dynamics equations. This approach serves as an alternative to traditional precipitation parameterization schemes, offering improved accuracy over raw model precipitation products. A study explored forecasting seasonal to sub-seasonal rainfall using CNNs [61]. This study combined ECMWF SEAS5 seasonal forecasts with convolutional neural networks (CNNs) to enhance total monthly regional rainfall predictions for Great Britain. The CNN was trained using mean sea-level pressure and 2 m air temperature forecasts, supervised by benchmark rainfall data from CEH-GEAR [62]. Results indicated that the CNN outperformed ECMWF predictions across all lead times.

Gradually, CNNs are also being utilized for more comprehensive, higher-resolution weather forecasts. Scher et al. [63] examined the efficacy of data-driven weather forecasting systems, particularly focusing on convolutional neural networks (CNNs). They explored replacing standard convolution operations with spherical convolution to account for Earth's curvature and incorporated hemisphere-specific properties into the network structure. Testing on the Weatherbench dataset [64], they found that both methods improved forecast skill, with their combination yielding the best results. Clare et al. [65] used a neural network to predict full probability density functions for weather forecasting, offering probabilistic

outputs rather than single values. Trained on WeatherBench data, the model predicted geopotential and temperature up to five days ahead. It achieved accuracy comparable to complex models while providing uncertainty and skill metrics for forecasts. Klein et al. introduced the Dynamic Convolutional Layer, a new deep network layer that generalized conventional convolutional layers and was applied to short-range weather prediction [66]. Unlike traditional layers that used fixed filters learned during training, the dynamic convolutional layer featured filters that adapted to each input during testing. This adaptation was made possible by a learned function that mapped inputs to the corresponding filters. Their results indicated that the dynamic convolutional layer outperformed existing baseline methods in short-range weather forecasting. Scher et al. evaluated CNN-based methods as an alternative for predicting weather forecast uncertainty based on the large-scale atmospheric state at initialization [67]. For new weather situations, the model assigns a scalar confidence value to medium-range forecasts, indicating whether predictability is higher or lower than usual for that time of year. While their approach showed lower skill than ensemble forecast models, it proved computationally efficient and outperformed several non-numerical alternatives.

For higher-resolution weather forecasts and downscaling approaches, CNN provided efficient abilities. Cho et al. developed DeU-Net [68], a statistical downscaling and bias correction method for Tmax and Tmin forecasts from the Global Data Assimilation and Prediction System over South Korea, using U-Net and spatial interpolation. DeU-Net outperformed the dynamical downscaling model and SVR-based models, achieving the highest spatial correlation and lowest RMSE for next-day forecasts. Soares et al. [69] assessed four CNN architectures for predicting temperatures and precipitation using ERA5 data [70], comparing results with Iberia01 data [71]. They then downscaled CMIP6 ESGCMs with these CNNs and evaluated them against Iberia01. The study produced a CNN-based, multi-model ensemble of 21st-century climate projections for the Iberian Peninsula at 0.1° resolution under four SSP scenarios. Xu et al. proposed the tensor-based long- and short-range convolution (TLS-Conv) [72] to address challenges in modeling multiple weather condition correlations and spatial dependencies. They introduced the tensor-based long- and short-range convolution for multiple weather prediction (TLS-MWP) model, which outperformed previous methods in extensive real-world experiments. Rodrigues et al. [73] presented and assessed a strategy utilizing a deep neural network to derive high-resolution representations from low-resolution weather forecasts. By adopting a supervised learning approach, they enabled the automatic generation of labeled data. Their findings demonstrated substantial improvements over conventional methods while keeping the strategy lightweight enough to operate on modest computer systems.

Additionally, CNN could also be applied to better detect extreme weather events. Racah et al. presented a semi-supervised extreme weather events prediction method [74]. They constructed a multichannel spatiotemporal CNN architecture, leveraging temporal information and unlabeled data to achieve better representation learning and improve the localization predicting of extreme weather events. Following, a study applied deep learning to detect extreme climate events, using a CNN classification system with Bayesian hyper-parameter optimization [75]. The model achieved 89–99% accuracy in identifying tropical cyclones, atmospheric rivers, and weather fronts, offering an effective alternative to traditional, subjective threshold-based methods for climate pattern detection. A recent study introduced an end-to-end rainstorm forecasting model using a gated convolutional encoder-decoder network [76]. The model combined physical principles with spatiotemporal patterns and effectively captured these patterns producing predictions that were both physically consistent and spatially coherent.

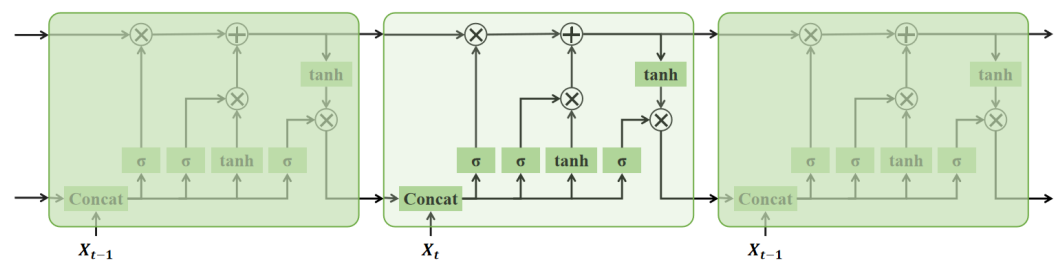
Nevertheless, CNN-based methods still face challenges, such as difficulty in handling teleconnections between distant grid points [77,78] and declining performance over longer



forecast periods [21]. These limitations have spurred the development of RNN-based, GNN-based, and Transformer-based approaches, as well as hybrid models that integrate CNNs with other techniques to enhance long-term prediction performance and long-distance dependence.

### 2.2.3. Recurrent Neural Network

RNN, as well as the improved RNN structures GRU and LSTM, excel at sequential prediction tasks [79]. An illustration of LSTM is shown in Figure 4. Since weather forecasting can be defined as a time-sequential prediction task, RNN-based methods were explored.



**Figure 4.** An illustration of LSTM [80].

Samy et al. [81] proposed a deep learning approach using LSTM to predict blizzards, focusing on recognizing weather patterns through sequential data analysis. Tested on real-time Polar Region datasets, it uses bivariate attributes like wind speed and air pressure, aiming to enhance the accuracy of extreme weather event prediction. Chkeir et al. developed an LSTM encoder-decoder model to nowcast extreme rain and wind speed using data from radar, lightning, Global Navigation Satellite System (GNSS), and weather stations [82]. Evaluated using four dataset configurations for 1-hour nowcasting with 10 min intervals, results showed over 90% detection accuracy for extreme wind and strong performance in detecting extreme rain within 30 min. Zenkner et al. employed two Bi-LSTM models in TensorFlow to predict the next 24 and 72 h based on the previous 120 h [83]. One model focused on temperature predictions at Kew Gardens, while the other predicted air temperature and relative humidity at Heathrow. Yan et al. [84] combined Multitask Learning with LSTM networks to enhance the prediction of hydrological variables across multiple variables using the ERA5-Land dataset. The multitask model outperformed single-task models in the prediction of hydrological variables, showing significant improvement for first-layer soil moisture as well as for evapotranspiration. The model also exhibited better stability in tropical monsoon regions for the prediction of hydrological variables, confirming the benefits of multitask learning in hydrological predictions. Li et al. introduced the Residual-EnDecode-Feedforward Attention Mechanism-Long Short-Term Memory (REDF-LSTM) model [85] to improve soil moisture prediction accuracy. This model integrated feedforward attention mechanisms with an encoder-decoder framework, leveraging LSTM's strengths in handling complex data sequences. Ding et al. proposed a hybrid VMD-LSTM-rolling model [86] for predicting non-stationary wave heights. Time series data were generated using a rolling method, followed by decomposition, training, and prediction for each series. The predictions were then combined to produce the final forecast of significant wave height. Results demonstrated that the VMD-LSTM-rolling model outperformed other methods and is applicable in practice.

However, individual RNN structures struggle to recognize geospatial information, leading to decreased performance in specific scenarios. Therefore, spatiotemporal RNN methods, such as ConvLSTM [87], TrajGRU [88], PredRNN [89], and PredRNN++ [90] were utilized.

ConvLSTM was first proposed to address precipitation nowcasting [87]. They framed it as a spatiotemporal sequence forecasting problem and then proposed ConvLSTM, an extended FC-LSTM with convolutional structures. Experiments showed that ConvLSTM

performed well in capturing spatiotemporal correlations. Following, Shi et al. introduced the Trajectory GRU (TrajGRU) [88], enhancing the ConvLSTM framework by enabling active learning of location-variant structures for recurrent connections. In addition, they established a benchmark for precipitation nowcasting, comprising a large-scale real-world dataset from the Hong Kong Observatory. Additionally, they presented a novel training loss function and a comprehensive evaluation protocol. This study facilitated related research and assessed the current state of the art in precipitation nowcasting. Around the same time, Wang et al. introduced PredRNN [89], a recurrent network for predictive learning of spatiotemporal sequences, which models modular visual dynamics using decoupled memory cells. These cells operate independently in a zigzag memory flow, enhancing communication across RNN layers. A memory decoupling loss prevents redundancy, and a new curriculum learning strategy aids long-term dynamics learning. Based on PredRNN, Zhao et al. [91] introduced a model system architecture that combines multiple spatiotemporal data sources with attention mechanisms, integrating it with the above PredRNN model for precipitation forecasting on the ERA5 dataset. This Att-PredRNN model, using global meteorological and ECMWF reanalysis data, predicted 24-hour precipitation distribution and significantly improved spatial and temporal correlation capture, outperforming other models in key metrics like RMSE, CSI, POD, and FAR. Kim et al. [92] compared PredRNN and a token-mixing model for precipitation forecasting, providing a quantitative guideline for applying advanced machine learning approaches. Using data from ERA5 and RAR databases, eight atmospheric variables across four pressure levels were employed to train and assess the models. Wang et al. [93] applied ConvLSTM, PredRNN, and PredRNN++ for precipitation nowcasting using GOES-13 infrared imagery over the central U.S. The models, trained on  $64 \times 64$  pixel grids and validated over 2012, demonstrated superior performance compared to traditional optical flow techniques.

Ma et al. proposed PrecipLSTM [94], integrating two modules: the spatial local attention memory module, which captures spatial relationships, and the time difference memory module, which captures temporal variations. Combined with PredRNN, PrecipLSTM effectively models spatiotemporal dependencies. Experiments on four radar datasets showed it outperformed previous methods with fewer parameters. Further, to address the issue of forecast blurring in AI-based precipitation nowcasting, Ma et al. proposed DB-RNN [95], reducing blurring via adversarial and gradient losses. DB-RNN was compatible with various RNNs, such as ConvLSTM, TrajGRU, PredRNN, PredRNN++, etc. Karevan et al. proposed a 2-layer spatiotemporal stacked LSTM model [79] for weather forecasting. The first layer used independent LSTM models per location, and the second layer combined their hidden states. Results demonstrated that incorporating spatial information improved the model's prediction performance in most scenarios. Wang et al. proposed a novel deep uncertainty quantification method with the RNN-based encoder-decoder structure and a novel negative log-likelihood error loss function for weather forecast [96]. In addition, this method contains an information fusion mechanism to incorporate NWP knowledge. Evaluated on a public dataset collected from weather stations in Beijing, China, It outperformed NWP by 47.76% accuracy, demonstrating state-of-the-art results. Naz et al. [97] presented an enhanced convolutional long short-term 2D (ConvLSTM2D) architecture for precipitation nowcasting. The proposed method incorporates time-distributed layers, which facilitate parallel Conv2D processing on each individual image input, thereby enhancing the analysis of spatial characteristics. Subsequently, the ConvLSTM2D layer is utilized to encapsulate spatiotemporal features, thereby boosting the model's predictive capabilities and computational efficiency. Tekin et al. [98] introduced the Weather Model for rapid and accurate long-term spatial predictions of high-resolution spatiotemporal weather data. Built on a stacked ConvLSTM network, it incorporates attention and context matcher mechanisms

to capture spatial patterns and preserve long-term dependencies, outperforming baseline models like ConvLSTM and U-Net.

Overall, RNN-based especially spatiotemporal RNN-based methods demonstrated competitive results in precipitation nowcasting and forecasting. However, these RNN-based methods are still constrained by the recurrent structure; that is, the inherent sequential calculation method hinders the parallelization of training, resulting in relatively high time complexity and slow training efficiency [99].

#### 2.2.4. Transformer-Based Neural Network

The Transformer, proposed by Vaswani et al. [99], relies solely on an attention mechanism, as shown in Figure 5. It combines the strengths of CNNs and RNNs, effectively capturing global dependencies while enabling greater parallelization. Since Transformer was proposed and successfully applied in lots of tasks facing to massive data, weather forecasting models based on Transformer have received great attention and development.

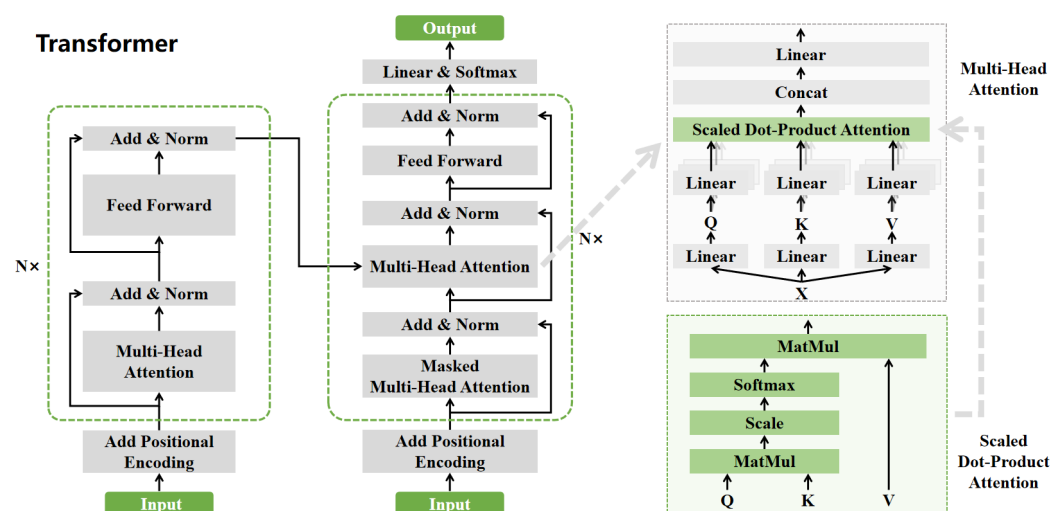


Figure 5. An illustration of Transformer [99].

Jin et al. developed the Preformer [100], a streamlined Transformer framework for precipitation nowcasting. Using an encoder-translator-decoder architecture, the Preformer captures spatiotemporal dependencies, models dynamics, and forecasts precipitation. Despite its simplicity and minimal parameters, it achieved state-of-the-art performance without complex structures or strategies. Bai et al. introduced Rainformer [101], a novel framework with two key components: a global feature extraction unit using window-based multi-head self-attention and a gate fusion unit that balances local and global features. Rainformer's architecture improved accuracy, especially for high-intensity rainfall. Gao et al. proposed Earthformer [102], a space-time Transformer designed for Earth system forecasting, based on a Cuboid Attention mechanism. This method decomposes data into cuboids and applies self-attention at the cuboid level, efficiently connecting them with global vectors. Experiments on real-world benchmarks for precipitation nowcasting and ENSO forecasting demonstrated its effectiveness. Another study proposed a Swin Transformer-based deep learning model to map fundamental variables from NWP models to precipitation maps [103]. This model effectively extracted features from meteorological variables and showed promising results in estimating heavy precipitation during strong convective events in two case studies. Li et al. tackled the challenges of quantitative precipitation nowcasting (QPN) by proposing a new lightweight physics-informed transformer model, LPT-QPN [104]. Using vertical cumulative liquid water content data, the model employed novel transformer modules and multihead squared attention to capture long-term precipitation dynamics while reducing

complexity. Experimental results showed LPT-QPN outperformed other models, particularly in high-intensity areas and longer lead times. Facing extreme weather events, Civitarese et al. used the Temporal Fusion Transformer (TFT) model [105] to forecast seasonal extreme precipitation. They forecast weekly quantiles of maximum daily precipitation up to six months ahead. Comparing TFT with climatology and ECMWF SEAS5 forecasts, results showed TFT outperformed both, especially at six-month lead times, capturing deviations from normal that climatology missed.

In addition, transformer-based methods have been more widely applied to global weather forecasts. Bi et al. introduced Pangu-Weather [106], a robust AI-driven system for medium-range global weather forecasting. This system utilizes the Swin Transformer as its backbone and features a three-dimensional Earth-Specific Transformer (3DEST), incorporating Earth-specific absolute positional biases related to latitude, longitude, and height to capture irregular spatial components. Additionally, Pangu-Weather implements a hierarchical temporal aggregation strategy, training four models for different forecast intervals: 1-hour, 3-hour, 6-hour, and 24-hour, which are optimized using a greedy algorithm to minimize prediction iterations. Results from the ECMWF ERA5 reanalysis data demonstrated that Pangu-Weather excels in both deterministic and extreme weather forecasts, operating 10,000 times faster than the Operational IFS. Chen et al. developed a large-scale global medium-range weather forecast model called FengWu [107]. This model was built using multi-modal and multi-task deep learning methods, incorporating multi-modal neural networks and multi-task automatic equalization weights to address the representation and interactions of various atmospheric variables. Fengwu generated high-precision global forecasts for the next 10 days in just 30 seconds. Gan et al. introduced W-MRI [108], a multi-output residual integration model for global weather forecasting. Incorporating a residual mechanism, W-MRI effectively extracts meteorological features, captures internal relations, and provides fast, accurate forecasts for variables like surface wind speed and precipitation through a specialized residual network. Aiming at longer-term weather forecasting, Chen et al. proposed a cascaded deep learning weather forecast system (called FuXi) that provides a 15-day global weather forecast [109]. FuXi was an autoregressive model, consisting of three modules, i.e., Cube Embedding, U-Transformer, and fully connected layer. It used the weather parameters of the first two time steps as input to predict the weather parameters of the next time step (one time step is 6 h) and used an autoregressive multi-step loss to reduce the cumulative error over long lead times. Experimental results showed that FuXi achieved 15-day forecasts.

Furthermore, researchers strived to more accurately simulate atmospheric dynamics at finer scales. Fengwu-GHR, an updated version of the Fengwu model, introduces the Spatial Identical Mapping Extrapolate (SIME) method to apply the prior knowledge learned from a pretrained low-resolution model to high-resolution analysis [110]. Additionally, the Decompositional and Combinational Transfer Learning (DCTL) algorithm is employed to better capture fine-scale weather patterns. Fengwu-GHR is the first AI-based weather forecasting model to operate at a  $0.09^\circ$  resolution and outperforms IFS-HRES, the current state-of-the-art numerical weather prediction (NWP) model, in 91.3% of the 320 target variables. Zhong et al. adapted transformer-based models, including SwinIR [111] and Uformer [112], to downscale 2 m temperature and 10m wind speed over Eastern Inner Mongolia, covering the region from  $39.6\text{--}46^\circ\text{N}$  latitude and  $111.6\text{--}118^\circ\text{E}$  longitude [113].

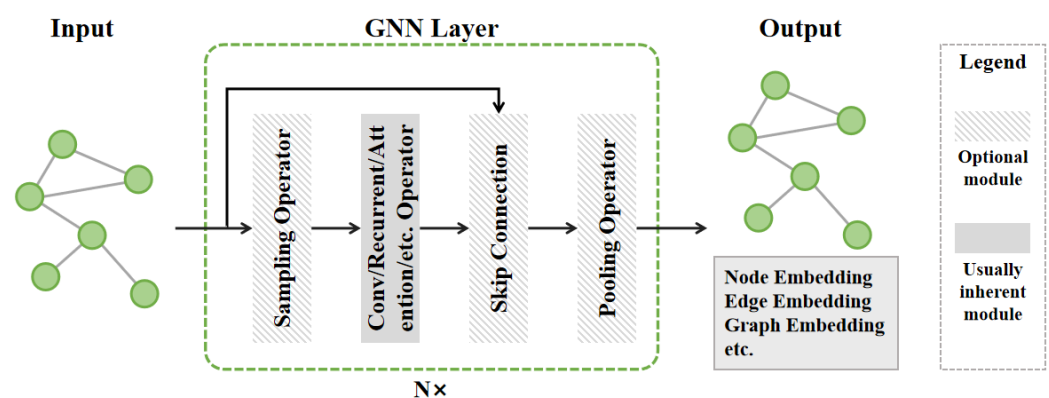
Additionally, longer-term forecasts, e.g., subseasonal forecast and climate analysis were also explored by researchers based on Transformer structures. Zhao et al. proposed a teleconnection-informed transformer that utilized the pretrained Pangu model and incorporated a teleconnection-informed temporal module to enhance predictability over an extended range [114]. Adjusting only 1.1% of the Pangu model's parameters, their

method improved predictions for multiple atmospheric variables, highlighting the role of teleconnections in future weather forecasting. Another study introduced ClimaX [115], a versatile deep learning model for weather and climate science, capable of training on heterogeneous datasets across various spatiotemporal tasks. By extending the Transformer architecture and using self-supervised learning on CMIP6 datasets, ClimaX showed superior performance in weather forecasting and climate projections compared to existing models, even with lower resolution and compute resources. Targeted at addressing various prediction tasks, a large-scale foundation model called Aurora was developed for atmospheric modeling [116]. Aurora features a Multi-scale 3D Swin Transformer U-Net backbone, along with 3D encoders and decoders based on Perceiver [117]. With a model size of 1.3 billion parameters and pretraining on over a million hours of diverse datasets, Aurora excels at capturing complex atmospheric patterns and structures. Consequently, Aurora outperforms state-of-the-art numerical weather prediction models in 10-day global weather forecasting at a spatial resolution of  $0.1^\circ$ .

Although the Transformer architecture and its core self-attention mechanism have demonstrated competitive performance, as weather forecasts are requested at finer spatiotemporal scales and extended to longer time frames, the associated computational costs pose a significant challenge for researchers [118].

#### 2.2.5. Graph Neural Network

Inspired by the effectiveness of GNN-based methods applied in learning the complex dynamics modeled by partial differential equations [119–121], GNN-based methods were explored to model weather dynamics and achieve weather forecasting. An illustration of the general design pipeline for a GNN model is shown in Figure 6.



**Figure 6.** An illustration of the general design pipeline for a GNN model [122].

Keisler introduced a data-driven method for global weather forecasting using graph neural networks [123], capable of advancing the 3D atmospheric state by six h. Trained on ERA5 reanalysis and GFS forecast data, the model outperformed prior data-driven approaches in predicting metrics like Z500 and T850, matching operational models at 1-degree scales. It also integrated live operational forecasts from GFS.

Lam et al. proposed a GNN-based method called GraphCast [124]. GraphCast was a cutting-edge AI model designed for medium-range weather forecasting with exceptional accuracy. It employed a graph neural network (GNN) as its foundational architecture and featured an encoding-processing-decoding structure. The encoder transformed data from latitude and longitude grids into multi-grid graph nodes to extract relevant features. The processor consisted of a 16-layer deep GNN that learned the information transfer process and updated multiple grid nodes. Finally, the decoder utilized a directed GNN to convert the learned multi-grid features back to the latitude and longitude grid, integrating this with



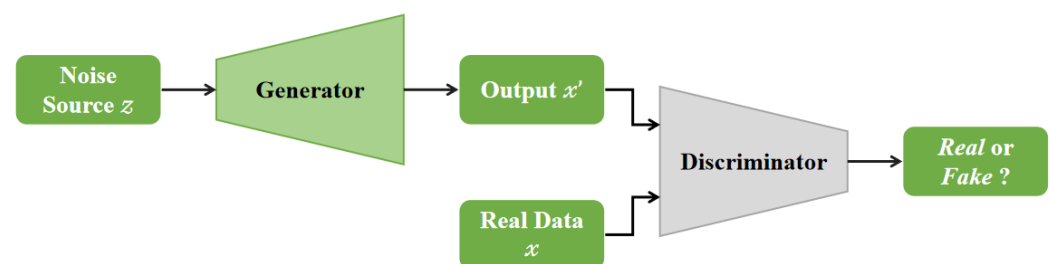
the input data to produce the output. GraphCast was able to forecast weather conditions up to 10 days in advance, achieving higher accuracy and speed than the European Centre for Medium-Range Weather Forecasts' High-Resolution Forecast (HRES) system, which was considered the industry standard.

Considering dynamical spatiotemporal changes in weather forecasting, Wei et al. introduced the Dynamic Climate Pattern Graph Recurrent Network (DCPGRN) [125] to tackle the challenges of weather forecasting caused by non-linearities and spatiotemporal correlations. Integrating a topology learner and domain knowledge into a spatiotemporal framework, DCPGRN captures optimal topology via using graph structures. Evaluations of real-world datasets showed significant performance improvements over existing algorithms, with insights from seven city stations demonstrating the model's interpretability and advantages. Another study proposed a wind speed prediction framework [126], utilizing a Graph Attention Network (GAT) for spatial feature extraction and an Informer model for training predictions. Results indicated that the framework achieved superior performance compared to traditional deep learning models, effectively aggregating data from multiple stations. In addition, Xu et al. focused on station-level weather forecasting and proposed the Dynamic Graph Former (DGFormer) model using a spatial-temporal GNN framework [127]. DGFormer combines a topology learner and domain knowledge within the STGNN structure for efficient integration with earth systems. Tested on short-term and medium-range weather predictions, DGFormer outperformed existing methods and demonstrated its advantages through detailed city-specific analyses and graph visualizations.

However, as the layers in GNNs increase, node features often become more alike, causing over-smoothing [128]. In weather forecasting, this reduces the ability to detect and differentiate subtle regional climatic variations. Moreover, GNNs are highly dependent on the input data's graph structure, making them sensitive to noise and inaccuracies, which can degrade model performance [21].

#### 2.2.6. Deep Generative Methods

Benefit from the effective detail generation capability, deep generative methods, e.g., Variational Auto-Encoder (VAE) [129], Generative Adversarial Network (GAN) [130], and diffusion model [131] have been utilized to enhance weather forecast resolution and restore finer-scale details of meteorological variables. An illustration of a typical GAN is shown in Figure 7.



**Figure 7.** An illustration of a typical GAN [132].

Ravuri et al. [133] introduced a deep generative model that improves probabilistic nowcasting by generating realistic, spatiotemporally consistent predictions up to 90 min ahead. Their method overcame the limitations in predicting medium-to-heavy rain events and producing sharp forecasts at longer lead times. Evaluations by over 50 expert meteorologists confirmed the model's superiority in 89% of cases for accuracy and usefulness. The model's predictions remained sharp, avoiding blurring while providing valuable probabilistic forecasts, outperforming existing methods in both resolution and lead time. Martín et al. [134] explored deep-learning diffusion models for super-resolution of weather data,

enhancing spatial resolution and detail of meteorological variables. Using SR3 and ResDiff architectures, the study proposed a methodology to transform low-resolution weather data into high-resolution outputs. The NowcastNet model [135] comprises a deterministic evolution network and a stochastic generative network to forecast extreme precipitation. Particularly, physical first principles was incorporated into the NowcastNet framework. The evolution network employed a two-path U-Net backbone to learn the nonlinear physical evolution of the precipitation processes and makes the predictions at 20 km scale. Meanwhile, the generative network, which features a U-Net encoder-decoder structure, simulates chaotic dynamics and delivers the final forecast at a 1- to 2-kilometer scale. FuXi-Extreme [136] updated the original FuXi model by incorporating a denoising diffusion probabilistic module to restore finer-scale details of surface meteorological variables. When compared with ECMWF-HRES and FuXi, FuXi-Extreme shows superior capability in forecasting extreme weather parameters as well as tracking and predicting the intensity of tropical cyclones. Chen et al. introduced FuXi Subseasonal-to-Seasonal (FuXi-S2S) [137], a machine learning model inspired by VAE for global daily mean forecasts up to 42 days. Trained on 72 years of ECMWF ERA5 data, FuXi-S2S outperformed existing models, enhancing predictions for precipitation and the Madden–Julian Oscillation (MJO). Zhong et al. introduced FuXi-ENS [138], an advanced machine learning model for 6-hourly global ensemble weather forecasts up to 15 days. Designed to address limitations in traditional ensemble prediction systems (EPS), FuXi-ENS operates at a high spatial resolution of  $0.25^\circ$  and incorporates 5 atmospheric and 13 surface variables across multiple pressure levels. Utilizing VAE with a loss function combining CRPS and KL divergence, it improves the incorporation of flow-dependent perturbations. FuXi-ENS outperformed the ECMWF in 98.1% of CRPS comparisons, highlighting its potential to significantly enhance ensemble weather forecasting. Price et al. developed GenCast [139], a probabilistic weather model based on diffusion models that outperformed the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble (ENS) in 97.2% of evaluated cases. Trained on reanalysis data, GenCast produced 15-day global forecasts at high resolution in just 8 min, improving predictions of extreme weather and wind power generation.

Generative deep models still generally have problems such as complex training processes, slow training speed, and difficult-to-explain mathematical processes [140]. However, there have been many attempts to optimize these problems [141,142], and the deep generative method for weather forecasts is still worth further exploration.

#### 2.2.7. Hybrid Deep Learning Models and Other Deep Learning Methods

The deep neural network methods introduced above for weather forecasting have their own advantages and limitations. A comparison is detailed in Table 1. Naturally, the researchers explored the integration of multiple types of neural networks within a single solution to combine their strengths.

Volkovs et al. proposed a hybrid deep learning method for weather forecasting [143]. They used a CNN-based UNet architecture, implanting a LSTM module processing time series. In addition, they primed the model states with atmospheric data to achieve better results. Sønderby et al. [144] leveraged CNNs and RNNs to forecast precipitation. The researchers introduced MetNet, a neural network that forecasted precipitation up to 8 h ahead with 1 km spatial resolution and 2 min temporal resolution, achieving low latency. It generated probabilistic precipitation maps and employed axial self-attention to aggregate global context from large input patches covering one million square kilometers. Following MetNet, Ali et al. present MetNet-2 [145], which offers a forecast range of up to 12 h with a frequency of 2 min and a spatial resolution of 1 km. To capture sufficient input context, it utilizes observations from a  $2048 \text{ km} \times 2048 \text{ km}$  area and incorporates novel neural network architectural elements

to effectively process this extensive context. Jin et al. [146] proposed a spatiotemporal inference network for precipitation prediction using multimodal meteorological data. They designed a spatiotemporal-aware convolutional layer (STACnv) to extract specific features and a ConvLSTM-based encoder-decoder framework to capture temporal dynamics. Experiments on three datasets demonstrated the effectiveness of STIN in precipitation nowcasting. A study developed a deep learning method using a Long-Term Recurrent Convolutional Network for nowcasting extreme weather events by combining radar reflectivity videos and lightning data to predict severe thunderstorms [147]. The model, framed as a classification task, utilized an ensemble approach with value-weighted skill scores, validated against radar data in Liguria, Italy. Chen et al. [148] proposed a CNN-LSTM model combining 3D and 2D CNNs with LSTM to capture spatiotemporal relationships in typhoon formation. By analyzing atmospheric and oceanographic variables, the model improved prediction accuracy compared to traditional numerical and machine learning methods. Experiments demonstrated its superiority across three datasets.

**Table 1.** A comparison of the introduced deep neural network for weather forecasting.

Methods	Advantages	Disadvantages
Simple Neural Network	Deep learning utilizes multilayer neural networks to automatically capture complex nonlinear patterns and spatiotemporal dynamics, minimizing the need for manual feature engineering typically required in traditional machine learning approaches.	Simple neural networks struggle with handling complex tasks effectively.
CNN	CNNs excel at efficiently capturing spatiotemporal patterns in weather forecasting.	CNN-based methods face challenges in capturing teleconnections between distant grid points and maintaining performance over extended forecast periods.
RNN	Designed to capture dependencies and relationships across time steps, RNNs are particularly well-suited for processing sequential information.	RNN-based methods are limited by their recurrent structure, as the sequential nature of their computations restricts parallelization during training, leading to higher time complexity and reduced training efficiency.
Transformer	Transformer architecture integrates the advantages of CNNs and RNNs, efficiently capturing global dependencies while supporting enhanced parallelization.	The associated computational cost is still a significant challenge.
GNN	GNN-based methods are adept at modeling weather dynamics by representing complex atmospheric interactions as graph structures, allowing them to effectively capture spatial dependencies and relationships between different regions.	GNNs rely heavily on the graph structure of the input data, making them vulnerable to noise and inaccuracies in the graph representation, which can adversely affect model performance.
Deep generative method	Deep generative methods are capable of probabilistic forecasts.	Generative deep models often face challenges such as a complicated training process, slow training speed, and a lack of interpretability in their mathematical mechanisms.

In addition to the combination of CNNs and RNNs, researchers also tried to combine Transformer with CNNs and RNNs. Lin et al. developed StHCFFormer [149], a multivariate spatiotemporal hybrid convolutional attention network for ocean weather prediction. The model combines spatial and temporal attention to capture dynamic correlations and uses homoscedasticity uncertainty loss for balanced predictions. Tested on the ERA5 dataset, StHCFFormer outperformed state-of-the-art methods in accuracy. Hu et al. proposed the Swin Transformer-based Variational recurrent neural network (SwinVRNN) [150] for ensemble weather forecasting. By integrating a deterministic SwinRNN with a perturbation module based on the Variational Auto-Encoder paradigm, it learned multivariate Gaussian distributions, improving forecast accuracy and ensemble performance compared to traditional perturbation methods. Wang et al. [151] developed STPF-Net, a Transformer-implanted recurrent neural network for longer-term precipitation forecasting, addressing error accumulation and limited spatial context in deep learning models. Using a multi-tier structure and transformer-based modules, STPF-Net improved 6-hour and 12-hour forecasts. Tests on southeastern China radar data showed it outperformed benchmark models, though challenges remained in storm initiation prediction.

In addition, there are some other deep neural networks that have been applied to weather forecasting. For example, Zhang et al. [152] introduced a novel precipitation forecasting method called DBNPF (Deep Belief Network for Precipitation Forecast), which uses deep belief networks to simulate human brain connectivity. By applying a Gaussian kernel for data transformation and a back-propagation network for fine-tuning, the approach maps original data features into a new semantic space through dimensionality reduction. It captures hidden information and identifies patterns in time-series data. The model uses seven environmental factors as input to predict 24-hour precipitation. Experiments conducted with data from Zunyi, Guizhou Province, validated its feasibility and demonstrated that this deep learning approach improves forecast accuracy compared to traditional machine learning methods.

### 2.3. Incorporating Physical Principles with Neural Networks

Inspired by the success of Fourier Neural Operators (FNOs) in solving complex partial differential equations [153], researchers have explored operator learning-based methods to model the intricate dynamics of weather and solve the partial differential equations governing atmospheric dynamics, simulating traditional numerical methods. FourCastNet [154,155], a data-driven deep learning simulator for Earth systems, was introduced to weather forecasting tasks via the Adaptive Fourier Neural Operator (AFNO) [156]. Utilizing the AFNO, it handles continuous, high-resolution inputs and can produce real-time weather forecasts up to a week ahead. This model outperforms existing numerical weather prediction methods in speed and accuracy for global and mesoscale forecasts, achieving five orders of magnitude in performance. Following, to address the limitations of Fourier Neural Operators (FNOs) in spherical geometries, which cause artifacts due to the assumption of flat geometry, Bonev et al. [157] introduced Spherical Fourier Neural Operators (SFNOs) for learning operators on spherical geometries. SFNOs effectively model long-range dependencies in spatiotemporal data. Applied to forecasting atmospheric dynamics, SFNOs demonstrated stable year-long autoregressive rollouts with physically plausible results. This advancement holds the potential for improving machine learning-based climate simulations, aiding in climate change response efforts. Around the same time, Lin et al. [158] modeled global weather dynamics using partial differential equations (PDEs) in spherical space, solved via Green's function and spherical convolution. They proposed the Spherical Neural Operator (SNO) combining spherical and vanilla convolutions to capture global and local correlations, respectively, developing an effective global weather prediction model.

In addition to operator learning, Kochkov et al. proposed the neural general circulation model (NeuralGCM) [159], which employs a differentiable dynamical core to simulate large-scale atmospheric dynamics and a neural network-based learned physics module to model unresolved physical processes, such as clouds, radiation, and precipitation. The framework of NeuralGCM also includes learned encoder and decoder components, which transform the pressure coordinates of the ERA5 dataset into the sigma coordinates used by the dynamical core. NeuralGCM has demonstrated performance comparable to state-of-the-art machine learning methods for 1- to 15-day weather forecasts, and it also shows the capability to provide accurate and stable decadal climate simulations.

In addition to employing neural networks modeling physical principles, considering meteorological domain knowledge in neural networks is another valuable way [160]. DGFormer [127], proposed by Xu et al., incorporated domain-specific knowledge into the STGNN framework to enable efficient integration with the Earth system. Manepalli et al. [161] approached emulation as an image-to-image translation task, aiming to convert gridded meteorological variables into snow–water equivalent (SWE) grids. The authors incorporated domain knowledge into the conditional GAN via a custom-designed loss function.

Overall, in the field of scientific computing including weather forecasting, how to combine physical principles with neural networks is still an emerging research direction.

### 3. Discussion

#### 3.1. Dataset

In machine learning, data are one of the most important factors in determining the quality of the results. Fortunately, the field of meteorology is extremely rich in data, and the storage and transfer rates of data are growing rapidly. In addition, there have been studies on constructing datasets and evaluation criteria that are more suitable for use in deep learning, e.g., WeatherBench [64]. Therefore, we summarized a list of the most commonly used datasets in existing machine learning-based weather forecasting work to help subsequent researchers perform their work better.

ERA5 [70]: The ERA5 reanalysis is a global meteorological dataset provided by ECMWF, which contains a comprehensive record of the global atmosphere, land surface, and ocean waves from 1950 to today. ERA5 is built on the IFS Cy41r2, which became operational in 2016. As a result, ERA5 benefits from ten years of advancements in model physics, core dynamics, and data assimilation. The time resolution of ERA5 is 1 hour and the spatial resolution is  $0.25^\circ \times 0.25^\circ$ , which is significantly higher than the 80 km for ERA-Interim and can capture much more intricate details of atmospheric phenomena. Moreover, ERA5 provides an uncertainty estimate from an ensemble, with data available every three hours at half the horizontal resolution. ERA5 contains an ensemble system of one control and nine perturbed members which provide background-error estimates for the deterministic HRES Data Assimilation system.

ERA5 includes upper-air parameters on 37 pressure levels from 1000 to 1hPa, a great many parameters on the surface level, and additional two-dimensional fields. Compared with radiosonde and PILOT data prior to assimilation, ERA5 shows an enhanced alignment for temperature, wind, and humidity in the troposphere, whereas no such improvement is observed in the stratosphere. And compared to independent buoy data, ERA5 shows a much-improved fit for ocean wave height. And ensemble spread and average data can be directly accessed. The dataset is updated in a timely manner. One day of reanalysis products is added every day, but the ERA5 dataset has a delay of 5 days. In addition, one month of the preliminary data is replaced by final data that has undergone more rigorous quality checks every month, but it has a delay of 2 to 3 months.

WeatherBench [64]: WeatherBench is a benchmark dataset for data-driven weather forecasting. It focuses on global medium-range (roughly 2 days to 2 weeks) prediction. The dataset is obtained by reducing the spatial resolution of ERA5 through regridding, and it contains data with spatial resolutions of  $5.625^\circ$  ( $32 \times 64$  grid points),  $2.8125^\circ$  ( $64 \times 128$  grid points), and  $1.40625^\circ$  ( $128 \times 256$  grid points) and temporal resolution of one hour from 1979 to 2018. In addition, the dataset contains  $z$ ,  $t$ ,  $q$ ,  $r$ ,  $u$ ,  $v$ ,  $vo$ , and  $pv$  of 13 vertical levels, including 50, 100, 150, 200, 250, 300, 400, 500, 600, 700, 850, 925, and 1000 hPa. And it includes several two-dimensional fields: 2 m temperature, 10 m wind, total cloud, precipitation, and the top-of-atmosphere incoming solar radiation. In addition to the time-varying fields, the dataset also contains several time-invariant fields, which are the land-sea mask, the soil type, and orography.

WeatherBench2 [162]: WeatherBench2 is a benchmark for the next generation of data-driven global weather models. The dataset is an update to WeatherBench and designed to accelerate progress in data-driven weather modeling. WeatherBench2 dataset includes ERA5, ERA5 Forecasts, Climatology, IFS HRES, IFS ENS, and the IFS ENS mean. ERA5 is used as the ground truth dataset. ERA5 Forecasts is a dataset of 10-day hindcasts initialized from ERA5 states at 00/12 UTC with the same IFS model version used to create ERA5.



For WeatherBench2, some variables in ERA5 Forecasts are downloaded to compute the headline scores for the year 2020. Notably, the data are available at 6-hour intervals for up to 5 days lead time, and at 12-hour intervals from 5 to 10 days lead time. And it offers a comparable baseline for AI models initialized using ERA5 and evaluated against it. The climatology is used to compute certain skill scores, especially ACC and SEEPS, and as a baseline forecast. And a probabilistic version of the climatology is generated by considering each of the 30 years spanning from 1990 to 2019 as an ensemble member. IFS HRES is the main baseline in WeatherBench2, which is derived from operational forecasts generated using ECMWF's IFS model. IFS HRES is evaluated by standard verification metrics and is generally considered the foremost in global, medium-range weather predictions. IFS HRES has been operated with a horizontal resolution of  $0.1^\circ$  since 2016. For WeatherBench2, the IFS HRES forecasts at  $t = 0$  are used as ground truth for evaluating IFS forecasts, rather than ERA5. The reason is that assessing IFS forecasts against ERA5 would introduce a non-zero error at the initial time step ( $t = 0$ ), unfairly disadvantaging the IFS when compared to AI models that are trained and validated using ERA5. IFS ENS is run by ECMWF and consists of a control run and 50 perturbed members. IFS ENS is run at  $0.1^\circ$  resolution and initiated at 00 and 12 UTC, with a forecast horizon extending to 15 days. Moreover, the IFS ENS mean is chosen as a baseline, which is computed by simply averaging over the 50 members.

**Weather2K [163]:** Weather2K is a multivariate spatiotemporal benchmark dataset for meteorological forecasting based on real-time observations from ground weather stations. The dataset includes observation data from 2130 ground weather stations throughout China, covering an area exceeding 6 million square kilometers. Currently, Weather2K includes Weather2K-R, Weather2K-N, and Weather2K-S. Among them, Weather2K-R is open source, including data from 1866 ground weather stations. Weather2K-N is the remaining unopened data, and Weather2K-S is a special version, which includes data from 15 representative ground weather stations distributed in different regions. In addition, Weather2K can be obtained from January 2017 to August 2021, and the temporal resolution is one hour. The factors contain three time-invariant constants: latitude, longitude, and altitude to provide position information and 20 important near-surface meteorological variables. Moreover, all ground weather stations use CST time (UTC +8 h), where the meteorological observation instruments and sensors for data collection are united. Therefore, Weather2K has the advantages of the real-time availability and reliability of the data to be applied to different tasks of weather forecasting.

**CONUS404 [164]:** CONUS404 is a high-resolution hydroclimate reanalysis dataset over the conterminous United States from October 1979 to September 2021, which has a 4 km spatial resolution and hourly time scales. The dataset is created using the WRF Model version 3.9.1 by dynamically downscaling of ERA5, and it is appropriate for forcing hydrological models and conducting hydroclimate scientific analyses over the conterminous United States. The dataset covers an area of  $5472 \text{ km} \times 4064 \text{ km}$ , which contains 1368 (in the east–west direction)  $\times$  1016 (in the north–south direction) grid points. And the dataset has 51 stretched vertical levels with the uppermost layer at 50 hPa. The finest vertical grid spacing is near the surface, at approximately 50 m, and gradually increases to roughly 700 m towards the top. In addition, some studies have demonstrated that CONUS404 has a high degree of accuracy in the spatiotemporal distribution of precipitation and temperature by better resolving fine-scale weather events such as summer mesoscale convective systems (MCSs) and winter orographic precipitation [165]. But the downward solar radiation appears to be overestimated at the surface by approximately 5%, while the snowpack is underestimated by roughly 15% [164]. This bias in radiation seems to be associated with an inadequate representation of nonprecipitating clouds in WRF.

LandBench 1.0 [166]: LandBench 1.0 is a benchmark dataset for AI land surface variable prediction. The dataset is collected from various sources, mainly including the ERA5-Land reanalysis dataset and ERA5 reanalysis dataset, SoilGrid, and the SMSC and MODIS datasets. ERA5-Land reanalysis dataset is a comprehensive and coherent global dataset, which spans multiple decades from 1950, has a high spatial resolution of  $0.1^\circ$  with  $3600 \times 7200$  grid points, and can offer precise support for future time. By integrating observations sourced from across the globe employing physical laws, the ERA5-Land reanalysis dataset is created and has shown superior accuracy compared to other datasets of land surface variables that rely on model outputs, remote sensing, and reanalysis [167]. LandBench 1.0 collected from the ERA5-Land reanalysis dataset includes soil moisture and soil temperature variables from layer 1 to layer 4, surface latent and sensible heat fluxes, skin temperature, albedo, leaf area index, snow cover, snow depth water equivalent, surface radiation, and runoff. And LandBench 1.0 collected from the ERA5 reanalysis dataset contains some atmospheric variables, which are used as the main forcing data for land surface variables and include precipitation, 2 m temperature, specific humidity,  $u$ ,  $v$ , and surface pressure. In addition, LandBench 1.0 includes constant variables, which include clay, sand, silt, soil water capacity, and vegetation type. Currently, LandBench1.0 can provide 0.5, 1, 2, and 4 degree resolution daily global data from 1979–2020.

CEH-GEAR [62]: CEH-GEAR is a 1 km daily and monthly rainfall dataset for Great Britain and Northern Ireland from 1890 to 2012. The goal of the CEH-GEAR dataset is to generate temporally consistent areal rainfall data over the longest possible period to support hydrological modeling. The dataset contains the Met Office compiled historical weather observations for the UK, specifically daily and monthly rainfall totals, which were collected from a national network of rain gauges. Due to the uneven geographic distribution of precipitation monitoring in the UK, some regions had adequate rain gauge coverage even in the early 20th century (e.g., London, Somerset, and the West Midlands), while others had very sparse gauge density (e.g., Scotland, South Central England, Wales, and the East and North of England). Consequently, caution is advised when using CEH-GEAR prior to 1961, as the data quality may vary significantly both temporally and spatially. Moreover, the CEH-GEAR provides both rainfall grids and minimum distance grids, which indicate the distance to the nearest gauge used to calculate rainfall for each grid cell. Users are strongly encouraged to utilize the minimum distance grids, particularly for data prior to 1961, in order to assess the suitability of the data for their specific applications. In addition, to prevent a downward bias in the gridded rainfall estimates, it is necessary to normalize the rain gauge rainfall totals before interpolation, with the most appropriate variable for this being the AAR. For Great Britain, the version used was the Met Office 1 km grid for the 1961–1990 standard period. This was developed by deriving AAR values for a 10 km grid using monthly data from approximately 13,100 rain gauges. These values were then gridded at a 1 km resolution using a bicubic spline interpolation method. For Northern Ireland, the 1 km grid of long-term average rainfall for 1961–1990 from Met Éireann was used. This dataset, covering the entire island of Ireland, was derived from rain gauge observations through regression analysis.

Iberia01 [71]: Iberia01 is a daily precipitation and temperature dataset over the Iberian Peninsula from 1971 to 2015 at  $0.1^\circ$  regular and  $0.11^\circ$  rotated resolutions. Iberia01 is the first gridded dataset of daily precipitation and temperatures for the entire Iberian Peninsula and can be seen as an update to the PT02 and IB02 datasets. The construction of Iberia01 is based on a dense network of thousands of stations from AEMET, IPMA, and APA between 1951 and 2015. To ensure consistency with previous datasets, the final network was established through a quality control process, which mandates that stations have at least 15 years of data for precipitation (or 40 years for temperature) between 1951 and

2015, with no more than 10% of yearly data missing. The resulting observational network includes 3486 stations for precipitation and 275 stations for temperature. Moreover, the number of stations has significantly decreased over the past two decades, particularly for precipitation. As a result, the gridded product is not suitable for historical trend analysis, as the changing number of stations could lead to biased outcomes. Additionally, between 2009 and 2014, there were very few precipitation stations in Portugal, so caution should be exercised when interpreting results from this period. For this reason, the reference climate period of 1971–2000 was recommended to use for this dataset. Overall, the station distribution across the Iberian Peninsula is fairly homogeneous, with good representation of orographical gradients, particularly for precipitation. Thus, the elevation was incorporated as a covariate in the interpolation process to model and reflect these gradients, a factor that was not included in the original PT02 and Spain02 datasets. In addition, the Iberia01 was built applying the area-averaged three-dimensional interpolation method, which is an area-averaged method based on ordinary kriging and three-dimensional thin plate splines.

We summarize these datasets in Table 2.

**Table 2.** Typical datasets for AI-based weather forecast.

Datasets	Time Coverage	Time Resolution	Spatial Coverage	Spatial Resolution	Variable
ERA5	1950–present	hourly	global	0.25°	Many including upper-air parameters on 37 pressure levels, 2 m temperature, 10 m wind.
WeatherBench	1979–2018	hourly	global	5.625°, 2.8125°, 1.40525°	Many including z, t, q, r, u, v, vo and pv of 13 vertical levels, 2 m temperature, 10 m wind.
WeatherBench2	1959–2023 (ERA5), 2018–2022 (IFS ENS), 2018–2020 (ERA5 forecast), etc.	hourly (ERA5), 6 hourly (climatology), etc.	global	0.25° (ERA5), 0.1° (IFS HRES), etc.	Many including upper-air parameters on pressure levels, 10 m u, 10 m v.
Weather2K	2017.7–2021.8	hourly	China	2130 ground weather stations	20 near-surface meteorological variables including 2.5 m temperature, 10 m wind speed, wind direction, etc.
CONUS404	1979.10–2021.9	hourly	the United States	4 km	Many including z, t, q, r, u, v, radiation.
LandBench 1.0	1979–2020	daily	global	0.5°, 1°, 2°, 4°	Land surface variables from ERA5-Land, atmospheric variables from ERA5, static variables.
CEH-GEAR	1890–2012	daily, monthly	Great Britain, Northern Ireland	1 km	Rainfall.
Iberia01	1971–2015	daily, monthly	Iberian Peninsula	0.1° regular, 0.11° rotated	Precipitation, temperature.

### 3.2. Evaluation Metrics

Model evaluation is also a key component of AI-based weather forecasting. Many metrics are used to quantify the skillfulness of models, including deterministic forecast metrics and ensemble forecast metrics.

The RMSE quantifies the magnitude of the differences between forecasts and the ground truth for a given variable indexed by  $j$  and a given lead time  $t$ . In general, RMSE is defined as the mean latitude-weighted RMSE over all forecasts [64,106,124,154]:

$$RMSE = \frac{1}{N_{forecasts}} \sum_i^{N_{forecasts}} \sqrt{\frac{1}{N_{lat}N_{lon}} \sum_j^{N_{lat}} \sum_k^{N_{lon}} L(j)(f_{i,j,k} - t_{i,j,k})^2} \quad (1)$$

where  $f$  is the model forecast and  $t$  is the ground truth.  $L(j)$  is the the latitude-weighting factor for the latitude at the  $j$  th latitude index, as follows:

$$L(j) = \frac{\cos(\text{lat}(j))}{\frac{1}{N_{\text{lat}}} \sum_j^{N_{\text{lat}}} \cos(\text{lat}(j))} \quad (2)$$

The ACC measures the correlation between the differences in forecasts from climatology and the differences in the ground truth from climatology. The latitude-weighted ACC is defined as follows [64]:

$$\text{ACC} = \frac{\sum_{i,j,k} L(j) f'_{i,j,k} t'_{i,j,k}}{\sqrt{\sum_{i,j,k} L(j) f_{i,j,k}^2 \sum_{i,j,k} L(j) t_{i,j,k}^2}} \quad (3)$$

where the prime  $'$  denotes the difference to the climatology. In general, climatology is defined as the long-term average of the weather states  $y_{j,k} = \frac{1}{N_{\text{time}}} \sum t_{j,k}$ .

CRPS and SSR can be used to evaluate the performance for ensemble weather forecast. Mathematically, CRPS is defined as follows [106]:

$$\text{CRPS} = \int_{-\infty}^{+\infty} [F(f_{i,j,k}) - \mathbb{I}(t_{i,j,k} \leq z)] dz \quad (4)$$

where  $F(\cdot)$  denotes the cumulative distribution function of the forecast distribution and  $\mathbb{I}(\cdot)$  is an indicator function that equals 1 if the statement is true and 0 otherwise.

SSR is obtained by dividing *Spread* by RMSE [106]. *Spread* can be calculated as follows:

$$\text{Spread}(i) = \sqrt{\frac{1}{N_{\text{lat}} N_{\text{lon}}} \sum_j^{N_{\text{lat}}} \sum_k^{N_{\text{lon}}} L(j) \cdot \text{var}(f_{i,j,k})} \quad (5)$$

where  $\text{var}(\cdot)$  represents the variance in the ensemble dimension. The SSR is calculated using the spread values averaged over all forecasts and RMSE. A perfectly reliable ensemble should yield an SSR of 1.0.

In addition, some categorical verification metrics are used to evaluate the performance of model precipitation forecasts. CSI can be used to quantitatively assess the forecast accuracy for precipitation amounts above a certain threshold [168]. The higher the CSI, the better the forecast performance.

$$\text{CSI} = \frac{H}{H + M + F} \quad (6)$$

HSS measures the proportion of correct forecasts after removing those that are correct purely by random chance [169]. A higher HSS indicates better performance.

$$\text{HSS} = \frac{2(H \cdot \text{CR} - F \cdot M)}{(H + M)(M + \text{CR}) + (H + F)(F + \text{CR})} \quad (7)$$

where  $H$ ,  $F$ ,  $M$ , and  $\text{CR}$  indicate the samples of hit, false alarm, miss, and correct rejection, respectively. They are defined in Table 3.

**Table 3.** All possible outcomes for categorical forecasts of binary events.

Events	Observed	Not Observed
Predicted	H	F
Not Predicted	M	CR

POD, POFD, and FAR can be also utilized for further assessments. A higher POD and lower POFD and FAR indicate better performance [169].

$$POD = \frac{H}{H + M} \quad (8)$$

$$POFD = \frac{F}{F + CR} \quad (9)$$

$$FAR = \frac{F}{F + H} \quad (10)$$

### 3.3. Challenges

#### 3.3.1. Explainability

Explainability is very important in weather forecasting since the transparent ML model helps in manual analysis of results, reliability judgment, causal mining, and even scientific knowledge discovery. However, explainability is still a common potential weakness in machine learning. Many ML models, especially DL models, consist of multiple layers with millions or even billions of parameters, which interact in highly nonlinear ways, making it difficult to trace how the model arrives at a specific prediction. A straightforward solution is to use a simple linear model [170], but this approach will obviously sacrifice predictive performance.

This challenging problem has attracted the attention of researchers. The two main improvement ideas are physics-informed neural networks (PINNs) and explainable artificial intelligence (XAI) techniques. Integrating physical principles into ML models can help ensure the forecasting adheres to established physical laws. Some recent state-of-the-art work, e.g., operator learning and meteorological domain knowledge implanting, has been introduced in Section 2.3. XAI is an active emerging field, receiving increased attention. Aiming at the challenges from analyzing DL models, there are mainly four types of practical methods in XAI [171], i.e., interpretable local surrogates [172], occlusion analysis [173], integrated gradients [174], and layerwise relevance propagation [175]. Methods for explaining DL models have been applied in geosciences. For example, Ebert-Uphoff et al. [176] and Lee et al. [177] predicted various meteorological values from satellite imagery and attributed predictions to specific input features.

#### 3.3.2. Extreme Weather Events

Machine learning models often face difficulties in accurately predicting extreme values, primarily due to (1) the scarcity of training samples for such rare events and (2) the inherent challenges associated with extrapolation and inference beyond the observed data range [178]. In addition, most existing global weather forecasting models minimize mean-squared error or mean absolute error, optimizing for average forecasts rather than extremes and assuming symmetric forecast error distributions.

Therefore, attending to better recognizing extreme weather events using ML-based methods, more effort is needed for researchers. Optimizing the objective loss function or refining the model design could help enhance the robustness of predictions for all relevant variables, even at or beyond the boundaries of their training range. For example, Fuxi-Extreme [136] added a diffusion module on the basic of Fuxi [109] to enhance finer-scale details, achieving better performance in extreme total precipitation, 10-meter wind speed, and 2-meter temperature. Furthermore, Watson et al. [179] proposed several straightforward measures that can be implemented to help assess the capability of machine learning models to deliver reliable forecasts for extreme events. Olivetti et al. [178]



proposed a workflow in extreme event prediction, providing a general foundation for the following researchers.

### 3.3.3. Integration with Existing Systems

Implementing ML methods into existing weather forecasting systems can offer a powerful approach to enhance performance by combining the strengths of data-driven techniques and physics-based models. ECMWF has implemented several large weather models online (<https://charts.ecmwf.int/>, accessed on 1 December 2024) as part of their daily forecast releases. The China Meteorological Administration has launched a testing plan for the large-scale ML weather forecasting models (<http://aimfdp.nmc.cn/ai/dist/>, accessed on 1 December 2024), using real-time live analysis data as input to produce weather forecasts for the next 0 to 15 days, including high-altitude meteorological elements, ground meteorological elements, quantitative precipitation forecasts, etc.

However, more efforts are still needed to more effectively embed ML weather forecasting models into existing forecasting systems. Some of the aspects that can be improved are discussed here. ML methods can be employed to optimize data assimilation by detecting patterns in observational data, filling data gaps, improving initial conditions for NWP simulations, etc. For example, Xiang et al. [180] proposed an artificial intelligence-based data assimilation framework (ADAF) that generates high-quality kilometer-scale analyses, validated using diverse real-world observations, including sparse surface weather data and satellite imagery. Additionally, ML methods can serve as post-processing tools to refine NWP outputs, correct biases, enhance resolution, and tailor forecasts for localized or extreme weather events. For instance, Zhong et al. [113] adapted transformer-based models, including SwinIR and Uformer, to downscale and correct biases of near-surface temperature and wind-speed forecasts. ML methods can also support operational efficiency by streamlining computationally expensive components of NWP systems, such as parameterizing cloud microphysics or radiation processes, through surrogate models. In practice, ML models can be deployed alongside NWP workflows, providing real-time adjustments, uncertainty quantification, and scenario-specific optimizations. Future developments could focus on creating robust pipelines for integrating ML solutions into operational systems, ensuring compatibility with existing workflows, and providing interpretable results to improve forecast utility for decision-makers.

## 4. Conclusions

Weather forecasting as a complex scientific problem requiring extensive geoscience data has seen significant advancements through the integration of data-driven machine learning approaches. These methods offer promising opportunities to complement and enrich traditional numerical weather prediction systems. By leveraging the strengths of ML weather forecasting methods, such as its remarkable capabilities to process vast datasets and identify intricate patterns, the field is expected to achieve more robust and efficient forecasting capabilities.

This paper provides a comprehensive review of the application of machine learning techniques in weather forecasting, including traditional machine learning methods, simple neural networks, convolutional neural networks, recurrent neural networks, transformer-based neural networks, generative deep learning methods, physics-informed neural networks, and other deep learning methods. In addition, most of the existing datasets and applied metrics are detailed and summarized, facilitating further ML-based weather forecasting research. Through analyzing and discussing current technical challenges, this paper suggests some valuable aspects that can be focused on future work, e.g., improving explainability, enhancing the ability to predict rare weather events, and integrating them into

existing systems. Furthermore, exploring larger-scale, higher-resolution, and longer-term weather forecasts is also a trend, and the technical and hardware challenges faced in this research and development process are also worthy of attention.

Overall, this study supports researchers in advancing their work on machine learning-based weather forecasting, provides a comprehensive overview of current progress, highlights potential future directions, and contributes to the continued development of this emerging field.

**Author Contributions:** Conceptualization, H.Z.; methodology, H.Z. and Y.L.; investigation, H.Z., Y.L. and C.Z.; resources, N.L.; data curation, Y.L.; writing—original draft preparation, H.Z., Y.L. and C.Z.; writing—review and editing, N.L.; visualization, H.Z. and Y.L.; supervision, H.Z.; project administration, H.Z.; funding acquisition, H.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the research fund of China Huaneng Clean Energy Research Institute (QNYJJ24-02).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

## Abbreviations

The following abbreviations are used in this manuscript:

AAR	Average Annual Rainfall
ACC	Anomaly Correlation Coefficient
AEMET	Spanish Meteorological Agency
AI	Artificial Intelligence
ANN	Artificial Neural Network
APA	Portuguese Environmental Agency
ARIMA	Autoregressive Integrated Moving Average Model
CEH-GEAR	The Centre for Ecology & Hydrology—Gridded Estimates of Areal Rainfall Dataset
CMIP6	Coupled Model Intercomparison Project Phase 6
CNN	Convolutional Neural Network
ConvLSTM	Convolutional Long Short-Term Memory
CRPS	Continuous Ranked Probability Score
CSI	Critical Success Index
DL	Deep Learning
DNN	Deep Neural Network
DT	Decision Tree
ECMWF	European Centre for Medium-Range Weather Forecasts
ECMWF HRES	ECMWF's High Resolution Model
ECMWF IFS	ECMWF Integrated Forecasting System
ECMWF SEAS5	ECMWF's Fifth Seasonal Forecasting System
ERA5	The fifth generation ECMWF reanalysis for the global climate and weather
ESGCM	Earth System and Global Climate Model
FAR	False Alarm Rate
GAN	Generative Adversarial Network
GAT	Graph Attention Network

GCN	Graph Convolutional Network
GFS	Global Forecast System
GNN	Graph Neural Network
GNSS	Global Navigation Satellite System
GOES 13	Geostationary Satellite 13
GRU	Gated Recurrent Unit
HRRR	High-Resolution Rapid Refresh
HSS	Heidke Skill Score
IFS	Integrated Forecasting System
IFS ENS	IFS Ensemble System
IFS HRES	IFS High-resolution System
IPMA	Portuguese Institute for Sea and Atmosphere
KNN	K Nearest Neighbor
LightGBM	Light Gradient Boosting Machine
LR	Linear Regression
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MCS	Mesoscale Convective System
ML	Machine Learning
MLP	Multi-Layered Perceptron
MODIS	Moderate Resolution Imaging Spectroradiometer
MSE	Mean Square Error
NB	Naïve Bayes
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
PILOT	Wind report from pilot balloon
POD	Probability of Detection
POFD	Probability of False Detection
PredRNN	Predictive Recurrent Neural Network
PredRNN++	The improved Predictive Recurrent Neural Network
pv	Potential vorticity
q	Specific humidity
r	Relative humidity
RAR	Radar-AWS Rainrates
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SARIMA	Seasonal AutoRegressive Integrated Moving Average Model
SEEPS	Stable equitable error in probability space
SEVIRI	Spinning Enhanced Visible Infra-Red Imager
SMSC	Soil Moisture Storage Capacity
SSP	Shared Socioeconomic Pathway
STGNN	Spatial-Temporal Graph Neural Network
SVM	Support Vector Machine
SVR	Support Vector Regression
t	Temperature
u	u component of wind
UTC	Universal Coordinated Time
v	v component of wind
VAE	Variational Auto-Encoder
vo	Vorticity
WRF	Weather Research and Forecasting Model
XGBoost	EXtreme Gradient Boosting
z	Geopotential

## References

1. Dey, K.C.; Mishra, A.; Chowdhury, M. Potential of intelligent transportation systems in mitigating adverse weather impacts on road mobility: A review. *IEEE Trans. Intell. Transp. Syst.* **2014**, *16*, 1107–1119. [\[CrossRef\]](#)
2. Bendre, M.; Thool, R.; Thool, V. Big data in precision agriculture: Weather forecasting for future farming. In Proceedings of the 2015 1st International Conference on Next Generation Computing Technologies (NGCT), Dehradun, India, 4–5 September 2015; pp. 744–750. [\[CrossRef\]](#)
3. Papaefthymiou, G.; Dragoon, K. Towards 100% renewable energy systems: Uncapping power system flexibility. *Energy Policy* **2016**, *92*, 69–82. [\[CrossRef\]](#)
4. Morris, D.A.; Crawford, K.C.; Kloesel, K.A.; Wolfinbarger, J.M. OK-FIRST: A meteorological information system for public safety. *Bull. Am. Meteorol. Soc.* **2001**, *82*, 1911–1924. [\[CrossRef\]](#)
5. Meenal, R.; Binu, D.; Ramya, K.; Michael, P.A.; Vinoth Kumar, K.; Rajasekaran, E.; Sangeetha, B. Weather forecasting for renewable energy system: A review. *Arch. Comput. Methods Eng.* **2022**, *29*, 2875–2891. [\[CrossRef\]](#)
6. Saha, S.; Saha, A.; Das, M.; Saha, A.; Sarkar, R.; Das, A. Analyzing spatial relationship between land use/land cover (LULC) and land surface temperature (LST) of three urban agglomerations (UAs) of Eastern India. *Remote Sens. Appl. Soc. Environ.* **2021**, *22*, 100507. [\[CrossRef\]](#)
7. Khatun, S.; Saha, A.; Gogoi, P.; Saha, S.; Sarkar, R. Assessment of Landslide Vulnerability Using Statistical and Machine Learning Methods in Bageshwar District of Uttarakhand, India. In *Geomorphic Risk Reduction Using Geospatial Methods and Tools*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 99–116. [\[CrossRef\]](#)
8. Pal, S.; Saha, A.; Gogoi, P.; Saha, S. An Ensemble of J48 Decision Tree with AdaBoost and Bagging for Flood Susceptibility Mapping in the Sundarbans of West Bengal, India. In *Geomorphic Risk Reduction Using Geospatial Methods and Tools*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 117–133. [\[CrossRef\]](#)
9. Yan, X.; Jiao, J.; Li, M.; Qi, H.; Liang, Y.; Xu, Q.; Zhang, Z.; Jiang, X.; Li, J.; Zhang, Z.; et al. Lateral sediment connectivity of landslides occurred under a heavy rainstorm and its influence on sediment yield of slope-channel cascade on the loess plateau. *Catena* **2022**, *216*, 106378. [\[CrossRef\]](#)
10. Saima, H.; Jaafar, J.; Belhaouari, S.; Jillani, T. Intelligent methods for weather forecasting: A review. In Proceedings of the 2011 National Postgraduate Conference, Perak, Malaysia, 19–20 September 2011; pp. 1–6. [\[CrossRef\]](#)
11. Zhang, H.; Zhang, M.; Yi, R.; Liu, Y.; Wen, Q.H.; Meng, X. Growing Importance of Micro-Meteorology in the New Power System: Review, Analysis and Case Study. *Energies* **2024**, *17*, 1365. [\[CrossRef\]](#)
12. Linnenluecke, M.K.; Griffiths, A.; Winn, M. Extreme weather events and the critical importance of anticipatory adaptation and organizational resilience in responding to impacts. *Bus. Strategy Environ.* **2012**, *21*, 17–32. [\[CrossRef\]](#)
13. Day, J.W.; Hall, C.; Day, J.W.; Hall, C. Global climate change: A warmer and more unpredictable future. In *America's Most Sustainable Cities and Regions: Surviving the 21st Century Megatrends*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 137–166. [\[CrossRef\]](#)
14. Bauer, P.; Thorpe, A.; Brunet, G. The quiet revolution of numerical weather prediction. *Nature* **2015**, *525*, 47–55. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Ren, X.; Li, X.; Ren, K.; Song, J.; Xu, Z.; Deng, K.; Wang, X. Deep learning-based weather prediction: A survey. *Big Data Res.* **2021**, *23*, 100178. [\[CrossRef\]](#)
16. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, F. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Sunil, S.; Aparna, B.; Anik, S.; Tusar, K.H.; Biswajeet, P.; Abdullah, A. Deep learning algorithms based landslide vulnerability modeling in highly landslide prone areas of Tamil Nadu, India. *Geosci. J.* **2024**, *28*, 1013–1038. [\[CrossRef\]](#)
18. Dhilipkumar, D.; Bala, P.Y.; Yogeswaran, T.; Vanitha, K. Real Time Weather Prediction System using Ensemble Machine Learning. In Proceedings of the 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAIS), Trichy, India, 23–25 August 2023; pp. 477–484. [\[CrossRef\]](#)
19. Holmstrom, M.; Liu, D.; Vo, C. Machine learning applied to weather forecasting. *Meteorol. Appl.* **2016**, *10*, 1–5.
20. Narang, U.; Juneja, K.; Upadhyaya, P.; Salunke, P.; Chakraborty, T.; Behera, S.K.; Mishra, S.K.; Suresh, A.D. Artificial intelligence predicts normal summer monsoon rainfall for India in 2023. *Sci. Rep.* **2024**, *14*, 1495. [\[CrossRef\]](#) [\[PubMed\]](#)
21. Zhang, J.; Wang, B.; Hua, M.; Chen, Z.; Liang, S.; Kang, X. Spatiotemporal Meteorological Prediction Based on Fully Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4107818. [\[CrossRef\]](#)
22. Gómez-Chova, L.; Tuia, D.; Moser, G.; Camps-Valls, G. Multimodal classification of remote sensing images: A review and future directions. *Proc. IEEE* **2015**, *103*, 1560–1584. [\[CrossRef\]](#)
23. Camps-Valls, G.; Tuia, D.; Bruzzone, L.; Benediktsson, J.A. Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. *IEEE Signal Process. Mag.* **2013**, *31*, 45–54. [\[CrossRef\]](#)
24. Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random forests for land cover classification. *Pattern Recognit. Lett.* **2006**, *27*, 294–300. [\[CrossRef\]](#)

25. Biswas, M.; Dhoom, T.; Barua, S. Weather forecast prediction: An integrated approach for analyzing and measuring weather data. *Int. J. Comput. Appl.* **2018**, *182*, 20–24. [\[CrossRef\]](#)
26. Anderson, G.J.; Lucas, D.D. Machine learning predictions of a multiresolution climate model ensemble. *Geophys. Res. Lett.* **2018**, *45*, 4273–4280. [\[CrossRef\]](#)
27. Era, C.A.A.; Rahman, M.; Alvi, S.T. Short Term Weather Forecasting Comparison Based on Machine Learning Algorithms. In Proceedings of the 2023 Intelligent Methods, Systems, and Applications (IMSA), Giza, Egypt, 15–16 July 2023; pp. 369–374. [\[CrossRef\]](#)
28. Sun, L.; Lan, Y.; Sun, X.; Liang, X.; Wang, J.; Su, Y.; He, Y.; Xia, D. Deterministic forecasting and probabilistic post-processing of short-term wind speed using statistical methods. *J. Geophys. Res. Atmos.* **2024**, *129*, e2023JD040134. [\[CrossRef\]](#)
29. Lang, Z.; Wen, Q.H.; Yu, B.; Sang, L.; Wang, Y. Forecast of winter precipitation type based on machine learning method. *Entropy* **2023**, *25*, 138. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Kühnlein, M.; Appelhans, T.; Thies, B.; Nauss, T. Improving the accuracy of rainfall rates from optical satellite sensors with machine learning—A random forests-based approach applied to MSG SEVIRI. *Remote Sens. Environ.* **2014**, *141*, 129–143. [\[CrossRef\]](#)
31. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J.; et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [\[CrossRef\]](#)
32. Hong, D.; Zhang, B.; Li, H.; Li, Y.; Yao, J.; Li, C.; Werner, M.; Chanussot, J.; Zipf, A.; Zhu, X.X. Cross-city matters: A multimodal remote sensing benchmark dataset for cross-city semantic segmentation using high-resolution domain adaptation networks. *Remote Sens. Environ.* **2023**, *299*, 113856. [\[CrossRef\]](#)
33. Schizas, C.N.; Michaelides, S.; Pattichis, C.S.; Livesay, R. Artificial neural networks in forecasting minimum temperature (weather). In Proceedings of the 1991 Second International Conference on Artificial Neural Networks, Bournemouth, UK, 18–20 November 1991; pp. 112–114.
34. Baboo, S.S.; Shereef, I.K. An efficient weather forecasting system using artificial neural network. *Int. J. Environ. Sci. Dev.* **2010**, *1*, 321. [\[CrossRef\]](#)
35. Abhishek, K.; Singh, M.P.; Ghosh, S.; Anand, A. Weather forecasting model using artificial neural network. *Procedia Technol.* **2012**, *4*, 311–318. [\[CrossRef\]](#)
36. Hossain, M.; Rekabdar, B.; Louis, S.J.; Dascalu, S. Forecasting the weather of Nevada: A deep learning approach. In Proceedings of the 2015 International Joint Conference on Neural Networks (IJCNN), Killarney, Ireland, 12–17 July 2015; pp. 1–6. [\[CrossRef\]](#)
37. Liu, J.N.; Hu, Y.; He, Y.; Chan, P.W.; Lai, L. Deep neural network modeling for big data weather forecasting. *Inf. Granularity Big Data Comput. Intell.* **2015**, 389–408. [\[CrossRef\]](#)
38. Yonekura, K.; Hattori, H.; Suzuki, T. Short-term local weather forecast using dense weather station by deep neural network. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 1683–1690. [\[CrossRef\]](#)
39. Grimes, D.; Coppola, E.; Verdecchia, M.; Visconti, G. A neural network approach to real-time rainfall estimation for Africa using satellite data. *J. Hydrometeorol.* **2003**, *4*, 1119–1133. [\[CrossRef\]](#)
40. Capacci, D.; Conway, B. Delineation of precipitation areas from MODIS visible and infrared imagery with artificial neural networks. *Meteorol. Appl.* **2005**, *12*, 291–305. [\[CrossRef\]](#)
41. Rivolta, G.; Marzano, F.S.; Coppola, E.; Verdecchia, M. Artificial neural-network technique for precipitation nowcasting from satellite imagery. *Adv. Geosci.* **2006**, *7*, 97–103. [\[CrossRef\]](#)
42. Zhou, X.; Ren, H.; Zhang, T.; Mou, X.; He, Y.; Chan, C.Y. Prediction of pedestrian crossing behavior based on surveillance video. *Sensors* **2022**, *22*, 1467. [\[CrossRef\]](#)
43. Mathieu, M.; Couprie, C.; LeCun, Y. Deep multi-scale video prediction beyond mean square error. In Proceedings of the 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, 2–4 May 2016. [\[CrossRef\]](#)
44. Wang, J.; Zuo, L.; Cordente Martínez, C. Basketball technique action recognition using 3D convolutional neural networks. *Sci. Rep.* **2024**, *14*, 13156. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Feichtenhofer, C.; Fan, H.; Malik, J.; He, K. Slowfast networks for video recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6202–6211. [\[CrossRef\]](#)
46. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556. [\[CrossRef\]](#)
47. Liu, J.N.; Hu, Y.; You, J.J.; Chan, P.W. Deep neural network based feature representation for weather forecasting. In Proceedings of the International Conference on Artificial Intelligence (ICAI). The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), Las Vegas, NV, USA, 22–25 July 2024; p. 1.
48. Weyn, J.A.; Durran, D.R.; Caruana, R. Can machines learn to predict weather? Using deep learning to predict gridded 500-hPa geopotential height from historical weather data. *J. Adv. Model. Earth Syst.* **2019**, *11*, 2680–2693. [\[CrossRef\]](#)



49. Castro, R.; Souto, Y.M.; Ogasawara, E.; Porto, F.; Bezerra, E. STConvS2S: Spatiotemporal Convolutional Sequence to Sequence Network for weather forecasting. *Neurocomputing* **2021**, *426*, 285–298. [\[CrossRef\]](#)
50. Ali, J.; Cheng, L. Temperature forecasts for the continental United States: A deep learning approach using multidimensional features. *Front. Clim.* **2024**, *6*, 1289332. [\[CrossRef\]](#)
51. Gao, F.; Fei, J.; Ye, Y.; Liu, C. MSLKSTNet: Multi-Scale Large Kernel Spatiotemporal Prediction Neural Network for Air Temperature Prediction. *Atmosphere* **2024**, *15*, 1114. [\[CrossRef\]](#)
52. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Zhang, Z.; Lin, L.; Gao, S.; Wang, J.; Zhao, H. Wind speed prediction in China with fully-convolutional deep neural network. *Renew. Sustain. Energy Rev.* **2024**, *201*, 114623. [\[CrossRef\]](#)
54. Xu, B.; Wang, X.; Li, J.; Liu, C. Hierarchical U-net with re-parameterization technique for spatio-temporal weather forecasting. *Mach. Learn.* **2024**, *113*, 3399–3417. [\[CrossRef\]](#)
55. Qiu, M.; Zhao, P.; Zhang, K.; Huang, J.; Shi, X.; Wang, X.; Chu, W. A short-term rainfall prediction model using multi-task convolutional neural networks. In Proceedings of the 2017 IEEE International Conference on Data Mining (ICDM), New Orleans, LA, USA, 18–21 November 2017; pp. 395–404. [\[CrossRef\]](#)
56. Li, X.; Du, Z.; Song, G. A method of rainfall runoff forecasting based on deep convolution neural networks. In Proceedings of the 2018 Sixth International Conference on Advanced Cloud and Big Data (CBD), Lanzhou, China, 12–15 August 2018; pp. 304–310. [\[CrossRef\]](#)
57. Haidar, A.; Verma, B. Monthly rainfall forecasting using one-dimensional deep convolutional neural network. *IEEE Access* **2018**, *6*, 69053–69063. [\[CrossRef\]](#)
58. Singh, M.; Acharya, N.; Patel, P.; Jamshidi, S.; Yang, Z.L.; Kumar, B.; Rao, S.; Gill, S.S.; Chattopadhyay, R.; Nanjundiah, R.S.; et al. A modified deep learning weather prediction using cubed sphere for global precipitation. *Front. Clim.* **2023**, *4*, 1022624. [\[CrossRef\]](#)
59. Agrawal, S.; Barrington, L.; Bromberg, C.; Burge, J.; Gazen, C.; Hickey, J. Machine learning for precipitation nowcasting from radar images. *arXiv* **2019**, arXiv:1912.12132. [\[CrossRef\]](#)
60. Pan, B.; Hsu, K.; AghaKouchak, A.; Sorooshian, S. Improving precipitation estimation using convolutional neural network. *Water Resour. Res.* **2019**, *55*, 2301–2321. [\[CrossRef\]](#)
61. Barnes, A.P.; McCullen, N.; Kjeldsen, T.R. Forecasting seasonal to sub-seasonal rainfall in Great Britain using convolutional-neural networks. *Theor. Appl. Climatol.* **2023**, *151*, 421–432. [\[CrossRef\]](#)
62. Keller, V.; Tanguy, M.; Prosdociimi, I.; Terry, J.; Hitt, O.; Cole, S.; Fry, M.; Morris, D.; Dixon, H. CEH-GEAR: 1 km resolution daily and monthly areal rainfall estimates for the UK for hydrological and other applications. *Earth Syst. Sci. Data* **2015**, *7*, 143–155. [\[CrossRef\]](#)
63. Scher, S.; Messori, G. Spherical convolution and other forms of informed machine learning for deep neural network based weather forecasts. *arXiv* **2020**, arXiv:2008.13524v4. [\[CrossRef\]](#)
64. Rasp, S.; Dueben, P.D.; Scher, S.; Weyn, J.A.; Mouatadid, S.; Thuerey, N. WeatherBench: A benchmark data set for data-driven weather forecasting. *J. Adv. Model. Earth Syst.* **2020**, *12*, e2020MS002203. [\[CrossRef\]](#)
65. Clare, M.C.; Jamil, O.; Morcrette, C.J. Combining distribution-based neural networks to predict weather forecast probabilities. *Q. J. R. Meteorol. Soc.* **2021**, *147*, 4337–4357. [\[CrossRef\]](#)
66. Klein, B.; Wolf, L.; Afek, Y. A dynamic convolutional layer for short range weather prediction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4840–4848. [\[CrossRef\]](#)
67. Scher, S.; Messori, G. Predicting weather forecast uncertainty with machine learning. *Q. J. R. Meteorol. Soc.* **2018**, *144*, 2830–2841. [\[CrossRef\]](#)
68. Cho, D.; Im, J.; Jung, S. A new statistical downscaling approach for short-term forecasting of summer air temperatures through a fusion of deep learning and spatial interpolation. *Q. J. R. Meteorol. Soc.* **2024**, *150*, 1222–1242. [\[CrossRef\]](#)
69. Soares, P.M.; Johannsen, F.; Lima, D.C.; Lemos, G.; Bento, V.A.; Bushenkova, A. High-resolution downscaling of CMIP6 Earth system and global climate models using deep learning for Iberia. *Geosci. Model Dev.* **2024**, *17*, 229–259. [\[CrossRef\]](#)
70. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [\[CrossRef\]](#)
71. Herrera, S.; Cardoso, R.M.; Soares, P.M.; Espírito-Santo, F.; Viterbo, P.; Gutiérrez, J.M. Iberia01: A new gridded dataset of daily precipitation and temperatures over Iberia. *Earth Syst. Sci. Data* **2019**, *11*, 1947–1956. [\[CrossRef\]](#)
72. Xu, G.; Ng, M.K.; Ye, Y.; Li, X.; Song, G.; Zhang, B.; Huang, Z. TLS-MWP: A Tensor-based Long-and Short-range Convolution for Multiple Weather Prediction. *IEEE Trans. Circuits Syst. Video Technol.* **2024**, *34*, 8382–8397. [\[CrossRef\]](#)
73. Rodrigues, E.R.; Oliveira, I.; Cunha, R.; Netto, M. DeepDownscale: A deep learning strategy for high-resolution weather forecast. In Proceedings of the 2018 IEEE 14th International Conference on e-Science (e-Science), Amsterdam, The Netherlands, 29 October–1 November 2018; pp. 415–422. [\[CrossRef\]](#)

74. Racah, E.; Beckham, C.; Maharaj, T.; Ebrahimi Kahou, S.; Prabhat, M.; Pal, C. Extremeweather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [\[CrossRef\]](#)
75. Liu, Y.; Racah, E.; Correa, J.; Khosrowshahi, A.; Lavers, D.; Kunkel, K.; Wehner, M.; Collins, W. Application of deep convolutional neural networks for detecting extreme weather in climate datasets. *arXiv* **2016**, arXiv:1605.01156. [\[CrossRef\]](#)
76. Zhang, T.; Liu, J.; Gao, C.; Wang, P.; Leng, L.; Xiao, Y. Prior-Guided gated convolutional networks for rainstorm forecasting. *J. Hydrol.* **2024**, *633*, 130962. [\[CrossRef\]](#)
77. Cachay, S.R.; Erickson, E.; Buckner, A.F.C.; Pokropek, E.; Potosnak, W.; Osei, S.; Lütjens, B. Graph neural networks for improved El Nino forecasting. Thirty-Fourth Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 6–12 December 2020. [\[CrossRef\]](#)
78. Araujo, A.; Norris, W.; Sim, J. Computing receptive fields of convolutional neural networks. *Distill* **2019**, *4*, e21. [\[CrossRef\]](#)
79. Karevan, Z.; Suykens, J.A. Spatio-temporal stacked LSTM for temperature prediction in weather forecasting. *arXiv* **2018**, arXiv:1811.06341. [\[CrossRef\]](#)
80. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#) [\[PubMed\]](#)
81. Samy, V.; Thenkanidiyoor, V. Blizzard prediction in antarctic meteorological data using deep learning techniques. *J. Intell. Fuzzy Syst.* **2023**, Preprint, 1–16. [\[CrossRef\]](#)
82. Chkeir, S.; Anesiadou, A.; Mascitelli, A.; Biondi, R. Nowcasting extreme rain and extreme wind speed with machine learning techniques applied to different input datasets. *Atmos. Res.* **2023**, *282*, 106548. [\[CrossRef\]](#)
83. Zenkner, G.; Navarro-Martinez, S. A flexible and lightweight deep learning weather forecasting model. *Appl. Intell.* **2023**, *53*, 24991–25002. [\[CrossRef\]](#)
84. Yan, Y.; Li, G.; Li, Q.; Zhu, J. Enhancing Hydrological Variable Prediction through Multitask LSTM Models. *Water* **2024**, *16*, 2156. [\[CrossRef\]](#)
85. Li, X.; Zhang, Z.; Li, Q.; Zhu, J. Enhancing Soil Moisture Forecasting Accuracy with REDF-LSTM: Integrating Residual En-Decoding and Feature Attention Mechanisms. *Water* **2024**, *16*, 1376. [\[CrossRef\]](#)
86. Ding, T.; Wu, D.; Shen, L.; Liu, Q.; Zhang, X.; Li, Y. Prediction of significant wave height using a VMD-LSTM-rolling model in the South Sea of China. *Front. Mar. Sci.* **2024**, *11*, 1382248. [\[CrossRef\]](#)
87. 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*. [\[CrossRef\]](#)
88. 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. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [\[CrossRef\]](#)
89. Wang, Y.; Long, M.; Wang, J.; Gao, Z.; Yu, P.S. Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
90. Wang, Y.; Gao, Z.; Long, M.; Wang, J.; Philip, S.Y. Predrnn++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 5123–5132. [\[CrossRef\]](#)
91. Zhao, H.; Zhang, G.; Du, M.; Wang, X. Improving Global Precipitation in Numerical Weather Prediction Systems based on Deep Learning Techniques. In Proceedings of the 2023 IEEE International Conference on High Performance Computing & Communications, Data Science & Systems, Smart City & Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys), Melbourne, VC, Australia, 17–21 December 2023; pp. 333–339. [\[CrossRef\]](#)
92. Kim, H.W.; Lee, S.J.; Lee, Y.W. Comparison of deep learning-based models for forecasting precipitation using era-5 and radar images. In *Remote Sensing for Agriculture, Ecosystems, and Hydrology XXV*; SPIE: Paris, France, 2023; Volume 12727, pp. 11–13. [\[CrossRef\]](#)
93. Wang, C.; Hong, Y. Application of spatiotemporal predictive learning in precipitation nowcasting. AGU Fall Meeting Abstracts, Washington, DC, USA, 13 December 2018; Volume 2018, p. H31H-1988.
94. Ma, Z.; Zhang, H.; Liu, J. Preciplstm: A meteorological spatiotemporal lstm for precipitation nowcasting. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–8. [\[CrossRef\]](#)
95. Ma, Z.; Zhang, H.; Liu, J. Db-rnn: A rnn for precipitation nowcasting deblurring. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 5026–5041. [\[CrossRef\]](#)
96. Wang, B.; Lu, J.; Yan, Z.; Luo, H.; Li, T.; Zheng, Y.; Zhang, G. Deep uncertainty quantification: A machine learning approach for weather forecasting. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2087–2095. [\[CrossRef\]](#)
97. Naz, F.; She, L.; Sinan, M.; Shao, J. Enhancing Radar Echo Extrapolation by ConvLSTM2D for Precipitation Nowcasting. *Sensors* **2024**, *24*, 459. [\[CrossRef\]](#) [\[PubMed\]](#)
98. Tekin, S.F.; Fazla, A.; Kozat, S.S. Numerical Weather Forecasting using Convolutional-LSTM with Attention and Context Matcher Mechanisms. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4105313. [\[CrossRef\]](#)

99. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp. 6000–6010. [\[CrossRef\]](#)
100. Jin, Q.; Zhang, X.; Xiao, X.; Wang, Y.; Xiang, S.; Pan, C. Preformer: Simple and Efficient Design for Precipitation Nowcasting with Transformers. *IEEE Geosci. Remote Sens. Lett.* **2023**, *21*, 1000205. [\[CrossRef\]](#)
101. Bai, C.; Sun, F.; Zhang, J.; Song, Y.; Chen, S. Rainformer: Features extraction balanced network for radar-based precipitation nowcasting. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [\[CrossRef\]](#)
102. Gao, Z.; Shi, X.; Wang, H.; Zhu, Y.; Wang, Y.B.; Li, M.; Yeung, D.Y. Earthformer: Exploring space-time transformers for earth system forecasting. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 25390–25403. [\[CrossRef\]](#)
103. Liu, H.; Fung, J.C.; Lau, A.K.; Li, Z. Enhancing quantitative precipitation estimation of NWP model with fundamental meteorological variables and Transformer based deep learning model. *Earth Space Sci.* **2024**, *11*, e2023EA003234. [\[CrossRef\]](#)
104. Li, D.; Deng, K.; Zhang, D.; Liu, Y.; Leng, H.; Yin, F.; Ren, K.; Song, J. LPT-QPN: A lightweight physics-informed transformer for quantitative precipitation nowcasting. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4107119. [\[CrossRef\]](#)
105. Civitarese, D.S.; Szwarcman, D.; Zadrozny, B.; Watson, C. Extreme Precipitation Seasonal Forecast Using a Transformer Neural Network. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021. [\[CrossRef\]](#)
106. Bi, K.; Xie, L.; Zhang, H.; Chen, X.; Gu, X.; Tian, Q. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* **2023**, *619*, 533–538. [\[CrossRef\]](#)
107. Chen, K.; Han, T.; Gong, J.; Bai, L.; Ling, F.; Luo, J.J.; Chen, X.; Ma, L.; Zhang, T.; Su, R.; et al. Fengwu: Pushing the skillful global medium-range weather forecast beyond 10 days lead. *arXiv* **2023**, arXiv:2304.02948. [\[CrossRef\]](#)
108. Gan, L.; Man, X.; Li, C.; She, L.; Shao, J. W-MRI: A Multi-output Residual Integration Model for Global Weather Forecasting. In Proceedings of the Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data, Wuhan, China, 6–8 October 2023; pp. 209–222. [\[CrossRef\]](#)
109. Chen, L.; Zhong, X.; Zhang, F.; Cheng, Y.; Xu, Y.; Qi, Y.; Li, H. FuXi: A cascade machine learning forecasting system for 15-day global weather forecast. *NPJ Clim. Atmos. Sci.* **2023**, *6*, 190. [\[CrossRef\]](#)
110. Han, T.; Guo, S.; Ling, F.; Chen, K.; Gong, J.; Luo, J.; Gu, J.; Dai, K.; Ouyang, W.; Bai, L. Fengwu-ghr: Learning the kilometer-scale medium-range global weather forecasting. *arXiv* **2024**, arXiv:2402.00059. [\[CrossRef\]](#)
111. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1833–1844. [\[CrossRef\]](#)
112. Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; Li, H. Uformer: A general u-shaped transformer for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 17683–17693. [\[CrossRef\]](#)
113. Zhong, X.; Du, F.; Chen, L.; Wang, Z.; Li, H. Investigating transformer-based models for spatial downscaling and correcting biases of near-surface temperature and wind-speed forecasts. *Q. J. R. Meteorol. Soc.* **2024**, *150*, 275–289. [\[CrossRef\]](#)
114. Zhao, S.; Xiong, Z.; Zhu, X.X. Efficient Subseasonal Weather Forecast Using Teleconnection-Informed Transformers. In Proceedings of the IGARSS 2024 IEEE International Geoscience and Remote Sensing Symposium, Athens, Greece, 7–12 July 2024; pp. 2847–2852. [\[CrossRef\]](#)
115. Nguyen, T.; Brandstetter, J.; Kapoor, A.; Gupta, J.K.; Grover, A. ClimaX: A foundation model for weather and climate. *arXiv* **2023**, arXiv:2301.10343. [\[CrossRef\]](#)
116. Bodnar, C.; Bruinsma, W.P.; Lucic, A.; Stanley, M.; Brandstetter, J.; Garvan, P.; Riechert, M.; Weyn, J.; Dong, H.; Vaughan, A.; et al. Aurora: A foundation model of the atmosphere. *arXiv* **2024**, arXiv:2405.13063. [\[CrossRef\]](#)
117. Jaegle, A.; Gimeno, F.; Brock, A.; Vinyals, O.; Zisserman, A.; Carreira, J. Perceiver: General perception with iterative attention. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 4651–4664. [\[CrossRef\]](#)
118. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 11106–11115. [\[CrossRef\]](#)
119. Sanchez-Gonzalez, A.; Godwin, J.; Pfaff, T.; Ying, R.; Leskovec, J.; Battaglia, P. Learning to simulate complex physics with graph networks. In Proceedings of the International Conference on Machine Learning, Virtual, 13–18 July 2020; pp. 8459–8468. [\[CrossRef\]](#)
120. Pfaff, T.; Fortunato, M.; Sanchez-Gonzalez, A.; Battaglia, P. Learning Mesh-Based Simulation with Graph Networks. In Proceedings of the International Conference on Learning Representations, Vienna, Austria, 4 May 2021.
121. Alet, F.; Jeewajee, A.K.; Villalonga, M.B.; Rodriguez, A.; Lozano-Perez, T.; Kaelbling, L. Graph element networks: Adaptive, structured computation and memory. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 212–222. [\[CrossRef\]](#)

122. Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; Sun, M. Graph neural networks: A review of methods and applications. *AI Open* **2020**, *1*, 57–81. [\[CrossRef\]](#)
123. Keisler, R. Forecasting global weather with graph neural networks. *arXiv* **2022**, arXiv:2202.07575. [\[CrossRef\]](#)
124. Lam, R.; Sanchez-Gonzalez, A.; Willson, M.; Wirsberger, P.; Fortunato, M.; Alet, F.; Ravuri, S.; Ewalds, T.; Eaton-Rosen, Z.; Hu, W.; et al. Learning skillful medium-range global weather forecasting. *Science* **2023**, *382*, 1416–1421. [\[CrossRef\]](#) [\[PubMed\]](#)
125. Wei, X.; Hao, J.; Xu, Z.; Han, J. DCPGRN: A Dynamic Climate Pattern Graph Recurrent Network for Weather Forecasting. In Proceedings of the 2023 International Conference on Networks, Communications and Intelligent Computing (NCIC), Suzhou, China, 17–19 November 2023; pp. 41–45. [\[CrossRef\]](#)
126. Yu, C.; Yan, G.; Yu, C.; Mi, X. Attention mechanism is useful in spatio-temporal wind speed prediction: Evidence from China. *Appl. Soft Comput.* **2023**, *148*, 110864. [\[CrossRef\]](#)
127. Xu, Z.; Wei, X.; Hao, J.; Han, J.; Li, H.; Liu, C.; Li, Z.; Tian, D.; Zhang, N. DGFormer: A physics-guided station level weather forecasting model with dynamic spatial-temporal graph neural network. *GeoInformatica* **2024**, *28*, 499–533. [\[CrossRef\]](#)
128. Keriven, N. Not too little, not too much: A theoretical analysis of graph (over) smoothing. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 2268–2281. [\[CrossRef\]](#)
129. Kingma, D.P. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114. [\[CrossRef\]](#)
130. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [\[CrossRef\]](#)
131. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 10684–10695. [\[CrossRef\]](#)
132. Creswell, A.; White, T.; Dumoulin, V.; Arulkumaran, K.; Sengupta, B.; Bharath, A.A. Generative adversarial networks: An overview. *IEEE Signal Process. Mag.* **2018**, *35*, 53–65. [\[CrossRef\]](#)
133. Ravuri, S.; Lenc, K.; Willson, M.; Kangin, D.; Lam, R.; Mirowski, P.; Fitzsimons, M.; Athanassiadou, M.; Kashem, S.; Madge, S.; et al. Skillful precipitation nowcasting using deep generative models of radar. *Nature* **2021**, *597*, 672–677. [\[CrossRef\]](#)
134. Martinů, J.M.; Šimánek, P. Enhancing Weather Predictions: Super-Resolution via Deep Diffusion Models. In Proceedings of the International Conference on Artificial Neural Networks, Lugano, Switzerland, 17 September 2024; pp. 186–197. [\[CrossRef\]](#)
135. Zhang, Y.; Long, M.; Chen, K.; Xing, L.; Jin, R.; Jordan, M.I.; Wang, J. Skillful nowcasting of extreme precipitation with NowcastNet. *Nature* **2023**, *619*, 526–532. [\[CrossRef\]](#)
136. Zhong, X.; Chen, L.; Liu, J.; Lin, C.; Qi, Y.; Li, H. FuXi-Extreme: Improving extreme rainfall and wind forecasts with diffusion model. *Sci. China Earth Sci.* **2024**, *67*, 3696–3708. [\[CrossRef\]](#)
137. Chen, L.; Zhong, X.; Li, H.; Wu, J.; Lu, B.; Chen, D.; Xie, S.P.; Wu, L.; Chao, Q.; Lin, C.; et al. A machine learning model that outperforms conventional global subseasonal forecast models. *Nat. Commun.* **2024**, *15*, 6425. [\[CrossRef\]](#) [\[PubMed\]](#)
138. Zhong, X.; Chen, L.; Li, H.; Feng, J.; Lu, B. FuXi-ENS: A machine learning model for medium-range ensemble weather forecasting. *arXiv* **2024**, arXiv:2405.05925. [\[CrossRef\]](#)
139. Price, I.; Sanchez-Gonzalez, A.; Alet, F.; Andersson, R.T.; El-Kadi, A.; Masters, D.; Ewalds, T.; Stott, J.; Mohamed, S.; Battaglia, P.; et al. Probabilistic weather forecasting with machine learning. *Nature* **2025**, *637*, 84–90. [\[CrossRef\]](#) [\[PubMed\]](#)
140. Croitoru, F.A.; Hondru, V.; Ionescu, R.T.; Shah, M. Diffusion models in vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 10850–10869. [\[CrossRef\]](#)
141. Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X. Improved techniques for training gans. *Adv. Neural Inf. Process. Syst.* **2016**, *29*.
142. Song, Y.; Ermon, S. Generative modeling by estimating gradients of the data distribution. *Adv. Neural Inf. Process. Syst.* **2019**, *32*.
143. Volkovs, K.; Urtans, E.; Caune, V. Primed UNet-LSTM for Weather Forecasting. In Proceedings of the 2023 7th International Conference on Advances in Artificial Intelligence, Istanbul, Turkey, 13–15 October 2023; pp. 13–17. [\[CrossRef\]](#)
144. Sønderby, C.K.; Espeholt, L.; Heek, J.; Dehghani, M.; Oliver, A.; Salimans, T.; Agrawal, S.; Hickey, J.; Kalchbrenner, N. Metnet: A neural weather model for precipitation forecasting. *arXiv* **2020**, arXiv:2003.12140. [\[CrossRef\]](#)
145. Espeholt, L.; Agrawal, S.; Sønderby, C.; Kumar, M.; Heek, J.; Bromberg, C.; Gaze, C.; Carver, R.; Andrychowicz, M.; Hickey, J.; et al. Deep learning for twelve hour precipitation forecasts. *Nat. Commun.* **2022**, *13*, 1–10. [\[CrossRef\]](#)
146. Jin, Q.; Zhang, X.; Xiao, X.; Wang, Y.; Meng, G.; Xiang, S.; Pan, C. Spatiotemporal inference network for precipitation nowcasting with multimodal fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *17*, 1299–1314. [\[CrossRef\]](#)
147. Guastavino, S.; Piana, M.; Tizzi, M.; Cassola, F.; Iengo, A.; Sacchetti, D.; Solazzo, E.; Benvenuto, F. Prediction of severe thunderstorm events with ensemble deep learning and radar data. *Sci. Rep.* **2022**, *12*, 20049. [\[CrossRef\]](#) [\[PubMed\]](#)
148. Chen, R.; Wang, X.; Zhang, W.; Zhu, X.; Li, A.; Yang, C. A hybrid CNN-LSTM model for typhoon formation forecasting. *GeoInformatica* **2019**, *23*, 375–396. [\[CrossRef\]](#)



149. Lin, L.; Zhang, Z.; Yu, H.; Wang, J.; Gao, S.; Zhao, H.; Zhang, J. StHCFormer: A Multivariate Ocean Weather Predicting Method Based on Spatiotemporal Hybrid Convolutional Attention Networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 3600–3614. [\[CrossRef\]](#)
150. Hu, Y.; Chen, L.; Wang, Z.; Li, H. SwinVRNN: A Data-Driven Ensemble Forecasting Model via Learned Distribution Perturbation. *J. Adv. Model. Earth Syst.* **2023**, *15*, e2022MS003211. [\[CrossRef\]](#)
151. Wang, J.; Wang, X.; Guan, J.; Zhang, L.; Zhang, F.; Chang, T. STPF-Net: Short-Term Precipitation Forecast Based on a Recurrent Neural Network. *Remote Sens.* **2023**, *16*, 52. [\[CrossRef\]](#)
152. Zhang, P.; Zhang, L.; Leung, H.; Wang, J. A deep-learning based precipitation forecasting approach using multiple environmental factors. In Proceedings of the 2017 IEEE International Congress on Big Data (BigData Congress), Boston, MA, USA, 11–14 December 2017; pp. 193–200. [\[CrossRef\]](#)
153. Li, Z.; Kovachki, N.; Azizzadenesheli, K.; Liu, B.; Bhattacharya, K.; Stuart, A.; Anandkumar, A. Fourier neural operator for parametric partial differential equations. *arXiv* **2020**, arXiv:2010.08895. [\[CrossRef\]](#)
154. Pathak, J.; Subramanian, S.; Harrington, P.; Raja, S.; Chattopadhyay, A.; Mardani, M.; Kurth, T.; Hall, D.; Li, Z.; Azizzadenesheli, K.; et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv* **2022**, arXiv:2202.11214. [\[CrossRef\]](#)
155. Kurth, T.; Subramanian, S.; Harrington, P.; Pathak, J.; Mardani, M.; Hall, D.; Miele, A.; Kashinath, K.; Anandkumar, A. Fourcastnet: Accelerating global high-resolution weather forecasting using adaptive fourier neural operators. In Proceedings of the Platform for Advanced Scientific Computing Conference, Davos, Switzerland, 26–28 June 2023; pp. 1–11. [\[CrossRef\]](#)
156. Guibas, J.; Mardani, M.; Li, Z.; Tao, A.; Anandkumar, A.; Catanzaro, B. Efficient token mixing for transformers via adaptive fourier neural operators. In Proceedings of the International Conference on Learning Representations, Vienna, Austria, 4 May 2021. [\[CrossRef\]](#)
157. Bonev, B.; Kurth, T.; Hundt, C.; Pathak, J.; Baust, M.; Kashinath, K.; Anandkumar, A. Spherical fourier neural operators: Learning stable dynamics on the sphere. In Proceedings of the International Conference on Machine Learning, Zhuhai China, 17–20 February 2023; pp. 2806–2823. [\[CrossRef\]](#)
158. Lin, K.; Li, X.; Ye, Y.; Feng, S.; Zhang, B.; Xu, G.; Wang, Z. Spherical Neural Operator Network for Global Weather Prediction. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *34*, 4899–4913. [\[CrossRef\]](#)
159. Kochkov, D.; Yuval, J.; Langmore, I.; Norgaard, P.; Smith, J.; Mooers, G.; Klöwer, M.; Lottes, J.; Rasp, S.; Düben, P.; et al. Neural general circulation models for weather and climate. *Nature* **2024**, *632*, 1060–1066. [\[CrossRef\]](#) [\[PubMed\]](#)
160. Kashinath, K.; Mustafa, M.; Albert, A.; Wu, J.; Jiang, C.; Esmailzadeh, S.; Azizzadenesheli, K.; Wang, R.; Chattopadhyay, A.; Singh, A.; et al. Physics-informed machine learning: Case studies for weather and climate modelling. *Philos. Trans. R. Soc. A* **2021**, *379*, 20200093. [\[CrossRef\]](#) [\[PubMed\]](#)
161. Manepalli, A.; Albert, A.; Rhoades, A.; Feldman, D.; Jones, A.D. Emulating numeric hydroclimate models with physics-informed cGANs. In Proceedings of the AGU Fall Meeting, San Francisco, CA, USA, 9–13 December 2019.
162. Rasp, S.; Hoyer, S.; Merose, A.; Langmore, I.; Battaglia, P.; Russell, T.; Sanchez-Gonzalez, A.; Yang, V.; Carver, R.; Agrawal, S.; et al. WeatherBench 2: A benchmark for the next generation of data-driven global weather models. *J. Adv. Model. Earth Syst.* **2024**, *16*, e2023MS004019. [\[CrossRef\]](#)
163. Zhu, X.; Xiong, Y.; Wu, M.; Nie, G.; Zhang, B.; Yang, Z. Weather2K: A Multivariate Spatio-Temporal Benchmark Dataset for Meteorological Forecasting Based on Real-Time Observation Data from Ground Weather Stations. In Proceedings of the International Conference on Artificial Intelligence and Statistics, Valencia, Spain, 25–27 April 2023; pp. 2704–2722. [\[CrossRef\]](#)
164. Rasmussen, R.; Liu, C.; Prein, A.; Kim, J.; Schneider, T.; Dai, A.; Gochis, D.; Dugger, A.; Zhang, Y. CONUS404: The NCAR–USGS 4-km long-term regional hydroclimate reanalysis over the CONUS. *Bull. Am. Meteorol. Soc.* **2023**, *104*, E1382–E1408. [\[CrossRef\]](#)
165. Prein, A.F.; Rasmussen, R.M.; Ikeda, K.; Liu, C.; Clark, M.P.; Holland, G.J. The future intensification of hourly precipitation extremes. *Nat. Clim. Change* **2017**, *7*, 48–52. [\[CrossRef\]](#)
166. Li, Q.; Zhang, C.; Shanguan, W.; Wei, Z.; Yuan, H.; Zhu, J.; Li, X.; Li, L.; Li, G.; Liu, P.; et al. LandBench 1.0: A benchmark dataset and evaluation metrics for data-driven land surface variables prediction. *Expert Syst. Appl.* **2024**, *243*, 122917. [\[CrossRef\]](#)
167. Beck, H.E.; Pan, M.; Miralles, D.G.; Reichle, R.H.; Dorigo, W.A.; Hahn, S.; Sheffield, J.; Karthikeyan, L.; Balsamo, G.; Parinussa, R.M.; et al. Evaluation of 18 satellite-and model-based soil moisture products using in situ measurements from 826 sensors. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 17–40. [\[CrossRef\]](#)
168. Wang, C.C. On the calculation and correction of equitable threat score for model quantitative precipitation forecasts for small verification areas: The example of Taiwan. *Weather. Forecast.* **2014**, *29*, 788–798. [\[CrossRef\]](#)
169. Lyu, Y.; Zhu, S.; Zhi, X.; Ji, Y.; Fan, Y.; Dong, F. Improving subseasonal-to-seasonal prediction of summer extreme precipitation over southern China based on a deep learning method. *Geophys. Res. Lett.* **2023**, *50*, e2023GL106245. [\[CrossRef\]](#)
170. Haufe, S.; Meinecke, F.; Görgen, K.; Dähne, S.; Haynes, J.D.; Blankertz, B.; Bießmann, F. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* **2014**, *87*, 96–110. [\[CrossRef\]](#)



171. Samek, W.; Montavon, G.; Lapuschkin, S.; Anders, C.J.; Müller, K.R. Explaining deep neural networks and beyond: A review of methods and applications. *Proc. IEEE* **2021**, *109*, 247–278. [[CrossRef](#)]
172. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why should i trust you?” Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144. [[CrossRef](#)]
173. Zeiler, M.; Fergus, R. Visualizing and Understanding Convolutional Networks. In Proceedings of the 13th European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Volume 1311, pp. 818–833. [[CrossRef](#)]
174. Sundararajan, M.; Taly, A.; Yan, Q. Axiomatic attribution for deep networks. In Proceedings of the International Conference on Machine Learning, Sydney, NSW, Australia, 6–11 August 2017; pp. 3319–3328. [[CrossRef](#)]
175. Bach, S.; Binder, A.; Montavon, G.; Klauschen, F.; Müller, K.R.; Samek, W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS One* **2015**, *10*, e0130140. [[CrossRef](#)] [[PubMed](#)]
176. Ebert-Uphoff, I.; Hilburn, K. Evaluation, tuning and interpretation of neural networks for working with images in meteorological applications. *Bull. Am. Meteorol. Soc.* **2020**, *101*, E2149–E2170. [[CrossRef](#)]
177. Lee, Y.; Kummerow, C.D.; Ebert-Uphoff, I. Applying machine learning methods to detect convection using GOES-16 ABI data. *Atmos. Meas. Techn. Discuss* **2020**, *2020*, 1–28. [[CrossRef](#)]
178. Olivetti, L.; Messori, G. Advances and prospects of deep learning for medium-range extreme weather forecasting. *Geosci. Model Dev.* **2024**, *17*, 2347–2358. [[CrossRef](#)]
179. Watson, P.A. Machine learning applications for weather and climate need greater focus on extremes. *arXiv* **2022**, arXiv:2207.07390. [[CrossRef](#)]
180. Xiang, Y.; Jin, W.; Dong, H.; Bai, M.; Fang, Z.; Zhao, P.; Sun, H.; Thambiratnam, K.; Zhang, Q.; Huang, X. ADAP: An Artificial Intelligence Data Assimilation Framework for Weather Forecasting. *arXiv* **2024**, arXiv:2411.16807. [[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.