

Weather Prediction Using Artificial Intelligence Approaches, Challenges, and Future Directions

Mahak Yadav¹ and Anmol Tanwar²

Department of Computer Science

Dronacharya College of Engineering, Gurgaon

mahakyadav@example.ac.in

Abstract: *Weather forecasting has always been one of the most technically demanding areas in applied science, requiring the simultaneous processing of vast, multi-dimensional datasets while accounting for the chaotic and non-linear dynamics that govern atmospheric systems. Traditional numerical weather prediction (NWP) methods, though reliable to a considerable extent, are computationally intensive and struggle to capture fine-grained, localized weather phenomena with adequate precision. Over the past decade, artificial intelligence — and more specifically, deep learning — has begun to reshape the forecasting landscape in fundamental ways. This paper examines the role of AI techniques, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and hybrid transformer-based architectures, in improving the accuracy, efficiency, and spatial resolution of weather prediction. Drawing on recent advances such as Google DeepMind's GraphCast and NVIDIA's FourCastNet, we explore how these models are trained on decades of reanalysis data and how they compare against established operational forecasting systems. The paper also addresses key challenges — including interpretability, data quality, and model generalization — and outlines promising directions for future research. Our findings suggest that AI-driven forecasting, particularly when integrated with physical modeling, offers a compelling and increasingly practical path toward more reliable and accessible weather prediction.*

Keywords: weather prediction, artificial intelligence, deep learning, LSTM, convolutional neural networks, numerical weather prediction, FourCastNet, GraphCast, atmosphere modeling

I. INTRODUCTION

Accurate weather forecasting is not merely a matter of scientific curiosity — it is a public safety imperative. Decisions about evacuating coastal areas ahead of a cyclone, planning irrigation schedules for an upcoming dry season, or managing air traffic during a storm system all hinge on the reliability of weather predictions. Yet, despite more than a century of organized meteorological research, the atmosphere remains stubbornly difficult to predict beyond a window of roughly ten days with high confidence.

Historically, forecasters relied on physical models — sets of differential equations encoding the laws of fluid dynamics and thermodynamics — to simulate atmospheric evolution. These numerical weather prediction (NWP) models, run on supercomputers around the world, have steadily improved since their inception in the 1950s. However, they remain computationally expensive, slow to update, and inherently limited by the resolution at which they can practically be run. Even with modern high-performance computing, running a global NWP model at one-kilometre resolution remains infeasible in real time.

Into this landscape, artificial intelligence has entered not as a replacement for physical understanding, but as a powerful complement. Machine learning models — particularly those built on deep neural networks — can learn complex, high-dimensional mappings from historical weather data with remarkable efficiency. Once trained, they can produce forecasts in a matter of seconds, compared to hours for a full NWP simulation. Equally importantly, they can capture statistical regularities in the data that physically-based models might miss or oversimplify.



The goal of this paper is to provide a structured review of how AI is being applied to weather forecasting, what results have been achieved so far, where the major technical and practical obstacles lie, and what the field might look like in the coming decade. We aim to write for an audience with a foundational background in either computer science or earth science, though we have tried to make the core ideas accessible across both disciplines.

II. BACKGROUND AND RELATED WORK

2.1 Traditional Numerical Weather Prediction

Numerical weather prediction traces its conceptual origins to Lewis Fry Richardson, who in 1922 proposed the idea of solving atmospheric equations by hand using a large team of human calculators. While impractical at the time, the idea anticipated modern computational meteorology. The first successful NWP computation was made in 1950 on ENIAC, one of the earliest electronic computers, by a team led by Jule Charney.

Modern NWP systems such as the Global Forecast System (GFS) operated by the U.S. National Weather Service and the European Centre for Medium-Range Weather Forecasts (ECMWF) model work by dividing the atmosphere into a three-dimensional grid and numerically integrating the governing fluid dynamics equations forward in time. These models assimilate observations from weather stations, radiosondes, satellites, and ocean buoys to initialize each forecast cycle. They are sophisticated, well-validated, and trusted operationally worldwide.

Despite their strengths, NWP models have inherent limitations. The atmosphere is a chaotic system, meaning small errors in initial conditions amplify over time — a phenomenon famously described by the butterfly effect metaphor. This places a hard theoretical ceiling on deterministic forecast skill, typically around two weeks. Additionally, resolving features smaller than the model grid size requires parameterization schemes that introduce their own sources of uncertainty.

2.2 Early Machine Learning Approaches in Meteorology

Machine learning was applied to meteorological problems well before the deep learning era. Traditional approaches included artificial neural networks with one or two hidden layers used for tasks such as precipitation classification, short-range temperature forecasting, and downscaling of coarse model output. Statistical post-processing techniques like model output statistics (MOS) have long employed regression-based methods to correct systematic biases in NWP output.

These early approaches were limited both by computational resources and by the relatively shallow representational capacity of the models available at the time. They were also heavily dependent on handcrafted feature engineering, requiring domain experts to extract meaningful predictors from raw data before feeding them into learning algorithms.

2.3 The Deep Learning Revolution

The emergence of deep learning in the 2010s fundamentally changed what was computationally tractable. With the availability of large datasets, powerful GPUs, and architectural innovations such as convolutional neural networks (CNNs) and recurrent networks, neural models began to achieve superhuman performance on image recognition, speech transcription, and natural language tasks. Researchers quickly recognized that these same architectural breakthroughs could be adapted to spatiotemporal prediction problems — including weather forecasting.

A pivotal early result was the demonstration that CNNs could effectively extract spatial patterns from gridded weather data, treating atmospheric fields (temperature, pressure, wind velocity) similarly to how image classifiers treat pixel arrays. Subsequent work incorporated recurrent architectures to capture temporal dependencies across forecast lead times, and attention mechanisms to weight the relative importance of different spatial and temporal locations.



III. AI METHODS APPLIED TO WEATHER PREDICTION

3.1 Convolutional Neural Networks (CNNs)

CNNs are particularly well-suited to gridded, spatial data because their convolutional filters learn to detect local patterns in a translation-invariant way. When applied to weather data, CNNs can identify synoptic-scale features — fronts, troughs, ridges — as well as mesoscale structures like convective cells without requiring explicit feature engineering.

In precipitation nowcasting, CNNs have been used to predict the motion and evolution of radar reflectivity fields over horizons of one to six hours. DeepMind's MetNet series, for instance, uses a CNN-based architecture to generate probabilistic precipitation forecasts across the contiguous United States at high spatial and temporal resolution. These models operate directly on radar and satellite observations, bypassing the need for physical simulation entirely.

Researchers have also explored the use of CNNs for statistical downscaling — that is, transforming low-resolution NWP output into higher-resolution fields that capture local terrain effects. Super-resolution convolutional networks, originally developed for image enhancement, have been adapted to produce fine-scale wind and temperature fields from coarser model grids, yielding improvements over traditional bilinear interpolation.

3.2 Recurrent Neural Networks and LSTMs

While CNNs excel at capturing spatial structure, weather prediction is fundamentally a sequence modeling problem — what the atmosphere looks like now is strongly conditioned on its recent history. Recurrent neural networks (RNNs) are designed precisely for this: they maintain a hidden state that encodes information about past time steps, allowing the network to learn temporal dynamics.

Long short-term memory networks (LSTMs), introduced by Hochreiter and Schmidhuber in 1997, addressed the vanishing gradient problem that plagued earlier RNNs, enabling them to learn dependencies over longer time horizons. In meteorological applications, LSTMs have been employed for multi-day temperature and rainfall forecasting, as well as for predicting the track and intensity of tropical cyclones.

A notable characteristic of LSTM-based forecasters is their ability to model seasonal cycles and multi-week climate anomalies when trained on sufficiently long time series. Studies have shown that LSTMs can achieve competitive skill against statistical climate models for tasks like ENSO prediction, where the relevant timescales stretch to months or years.

3.3 Hybrid and Encoder-Decoder Architectures

Combining the spatial processing strengths of CNNs with the temporal learning of RNNs, convolutional LSTM (ConvLSTM) architectures were introduced specifically for spatiotemporal prediction. In ConvLSTM, the matrix multiplications within standard LSTM cells are replaced by convolution operations, allowing the network to process entire spatial grids at each time step while retaining sequential memory.

Encoder-decoder architectures, popularized by the U-Net design in biomedical image segmentation, have also found application in weather prediction. These networks compress input data into a compact latent representation (encoding) and then progressively reconstruct an output forecast (decoding), with skip connections preserving fine-scale spatial detail. They have proven effective for tasks such as probabilistic precipitation forecasting, where the output is a spatial field rather than a single value.

3.4 Transformer Models and Attention Mechanisms

The transformer architecture, which relies entirely on self-attention mechanisms rather than convolution or recurrence, has achieved dominant performance across natural language processing and vision tasks. Researchers have begun adapting transformers to atmospheric prediction with considerable success.

Pangu-Weather, developed by Huawei Research and published in Nature (2023), uses a 3D Earth-specific transformer to produce global weather forecasts at approximately 0.25-degree resolution across multiple pressure levels. Trained on



40 years of ECMWF ERA5 reanalysis data, Pangu-Weather matched or outperformed the ECMWF high-resolution (HRES) forecast system on multiple verification metrics for forecast lead times of one to seven days — a landmark result that attracted significant attention from both the AI and meteorological communities.

Similarly, GraphCast by Google DeepMind employs a graph neural network architecture that represents the Earth's surface as an irregular mesh, allowing for variable resolution and efficient information propagation across hemispheric distances. GraphCast demonstrated that a trained AI model could generate a 10-day global forecast in less than one minute on a single Google TPU, compared to several hours of supercomputer time for equivalent NWP runs. The efficiency gain is potentially transformative for applications requiring rapid ensemble generation.

3.5 Generative and Ensemble Approaches

A persistent limitation of deterministic AI weather models is their tendency to produce overly smooth forecasts that lack the fine-scale variability present in real atmospheric fields — an artifact of training with mean squared error loss functions, which penalize large deviations. To address this, researchers have explored the use of generative models, including generative adversarial networks (GANs) and diffusion models, which can produce physically plausible realizations with realistic spatial texture.

ECMWF's AIFS (Artificial Intelligence Forecast System) and Nvidia's CorrDiff model are recent examples of generative AI applied to ensemble weather prediction. These approaches hold the promise of producing probabilistic forecast distributions that better represent forecast uncertainty, which is critical for decision-making in risk-sensitive contexts such as flood management and energy grid balancing.

IV. PERFORMANCE COMPARISON: AI V/S TRADITIONAL METHODS

Rigorous comparison between AI-based and NWP-based forecasting is a complex undertaking, not least because the verification metrics, geographic domains, and initialization datasets used vary across studies. Nevertheless, a number of broad patterns have emerged from the recent literature.

For medium-range forecasting (roughly three to ten days ahead), newer AI models have consistently approached or matched the skill of ECMWF's HRES system on standard metrics such as root mean square error (RMSE) and anomaly correlation coefficient (ACC) for fields like 500 hPa geopotential height and 850 hPa temperature. Pangu-Weather and GraphCast, in particular, have demonstrated competitive performance, with some studies showing AI models outperforming NWP at certain lead times and vertical levels.

For short-range and nowcasting applications (zero to six hours), AI models have shown particularly strong results, especially for precipitation, where radar-based deep learning systems consistently outperform NWP on spatial sharpness and object-based metrics. This is partly because NWP models struggle to initialize convective-scale phenomena, while data-driven models can leverage the rich spatial and temporal structure in real-time radar composites.

It is worth noting, however, that AI models currently rely on NWP output or reanalysis data for their initial conditions — they do not replace the data assimilation process. This means their skill ultimately depends, to some degree, on the accuracy of the underlying NWP-generated initial states. Additionally, AI models trained on historical data may exhibit reduced skill during rare or unprecedented weather events that fall outside the distribution of training examples, a concern that remains an active area of investigation.

V. CHALLENGES AND LIMITATIONS

5.1 Data Quality and Coverage

The performance of supervised learning models is fundamentally constrained by the quality and representativeness of the data on which they are trained. For global weather AI models, the primary training resource has been ECMWF's ERA5 reanalysis dataset, which provides gridded atmospheric fields from 1940 to near-present at approximately 31-



kilometer resolution. While ERA5 is the most comprehensive reanalysis product available, it is a model-derived product and inherits the biases and assumptions of the underlying assimilation system.

In regions with sparse observing networks — much of the tropics, the oceans, and the polar regions — reanalysis quality is lower, and AI models trained on such data may underperform in these areas. The global disparity in surface observation density means that AI models developed primarily on ERA5 may carry forward geographic biases, potentially performing less reliably over data-sparse regions precisely where improved forecasting would have the greatest humanitarian benefit.

5.2 Interpretability and Physical Consistency

One of the most frequently cited concerns about AI weather models is the so-called black box problem: large neural networks make predictions through non-linear transformations that are difficult for humans to interpret. In operational meteorology, forecasters need to understand why a model is predicting a certain outcome, not just what it predicts. This interpretability gap complicates the adoption of AI models in institutions where forecast accountability and process understanding are central values.

Related to interpretability is the question of physical consistency. Standard neural networks are not constrained to obey conservation laws, and it is possible for them to generate forecast fields that are meteorologically implausible in subtle ways — violating mass or energy conservation, for instance. Some research groups have explored physics-informed neural networks (PINNs) and architecture designs that incorporate hard or soft physical constraints, but this remains an open research challenge.

5.3 Handling Extreme Events

Extreme weather events — major hurricanes, catastrophic floods, record-breaking heatwaves — are precisely the scenarios where accurate forecasting matters most, and precisely where AI models face their greatest challenges. These events are rare in historical data, creating a class imbalance problem, and they often involve physical processes at the limits of normal atmospheric behaviour. A model trained to minimize average prediction error over the full training distribution may effectively deprioritize accuracy on rare extremes in favour of routine conditions.

Several approaches have been proposed to address this, including oversampling of extreme events in training, use of loss functions that weight tail events more heavily, and explicit incorporation of climate change projections into training pipelines to account for shifting frequency distributions of extremes. These remain active research areas without fully satisfactory solutions as of the time of this writing.

5.4 Generalization and Robustness

Machine learning models are known to sometimes perform well on the data distributions they were trained on but fail to generalize to out-of-distribution examples. For weather models, this could mean degraded performance during atypical seasons, in regions underrepresented in the training data, or under future climate conditions that differ systematically from the historical period used for training. Ensuring that AI weather models are robust and well-calibrated across a wide range of conditions is an important prerequisite for operational deployment.

VI. FUTURE DIRECTIONS

6.1 Physics-AI Hybrid Systems

Perhaps the most promising direction in near-term weather AI is the development of hybrid systems that tightly couple data-driven machine learning with physics-based modelling. Rather than asking whether AI should replace NWP, this framing asks how the strengths of each approach — the physical rigor of NWP and the learning efficiency of neural networks — can be made to complement one another. Approaches under active exploration include AI-based parameterization schemes within NWP models, learned bias correction layers applied to NWP output, and emulators that replace computationally expensive components of NWP with fast neural surrogates.



6.2 Foundation Models for Weather

The success of large language models in natural language processing has inspired interest in training foundation models for weather — large, general-purpose neural networks pre-trained on broad atmospheric datasets that can be fine-tuned for a wide variety of downstream forecasting tasks. Such models could, in principle, be adapted with modest additional training to produce specialized forecasts for agriculture, energy, aviation, or disaster management, democratizing access to high-quality prediction tools for applications where custom model development is currently impractical.

6.3 Integration of Novel Data Sources

The proliferation of new observational platforms — including commercial satellite constellations, IoT-based weather station networks, aircraft-based sensing, and even smartphone barometers — is creating new data streams that were not available when historical reanalysis products were developed. AI models, with their capacity to assimilate heterogeneous, high-volume data sources, may be well positioned to exploit these emerging observational resources in ways that traditional NWP data assimilation pipelines cannot easily accommodate.

6.4 Climate Projections and Seasonal Forecasting

While much of the current AI weather literature focuses on day-to-week forecasting, there is growing interest in applying AI to the longer timescales of seasonal and decadal climate prediction. Neural models have shown skill in predicting the El Niño Southern Oscillation (ENSO) months in advance — a problem with profound implications for agricultural planning, water resource management, and disaster preparedness across the tropics. Extending AI skill further into climate timescales, and ensuring that such models faithfully represent the physical mechanisms that govern low-frequency climate variability, represents a significant and consequential research frontier.

VII. CONCLUSION

Weather prediction has always been a problem at the intersection of science, mathematics, and engineering, demanding the best tools each field can offer. Artificial intelligence, and deep learning in particular, has brought a new class of tools to this problem — ones that learn from data rather than first principles, and that can produce forecasts with a speed and efficiency that traditional numerical methods cannot match.

The progress over the past five years has been genuinely remarkable. Models like GraphCast, Pangu-Weather, and FourCastNet have demonstrated that neural networks can achieve medium-range forecast skill comparable to the world's best operational NWP systems, at a fraction of the computational cost. For precipitation nowcasting, AI has become the method of choice in several operational centres. The trajectory of improvement shows no sign of slowing.

At the same time, the challenges are real and should not be minimized. Questions of interpretability, physical consistency, extreme event performance, and equitable geographic coverage all require sustained research attention. The most productive path forward is likely neither pure data-driven AI nor unmodified physics-based simulation, but a thoughtful integration of both — systems that bring physical understanding to bear as a constraint and prior, while leveraging AI's capacity for pattern recognition and computational efficiency.

As AI tools become more accessible and as the volume and variety of atmospheric observations continues to grow, the opportunity to develop more accurate, more timely, and more widely available weather forecasts has never been greater. Meeting this opportunity — and ensuring its benefits are shared broadly — is one of the important scientific and societal challenges of the coming decade.

REFERENCES

1. Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619(7970), 533–538. <https://doi.org/10.1038/s41586-023-06185-3>



2. Lam, R., Sanchez-Gonzalez, A., Willson, M., et al. (2023). Learning skillful medium-range global weather forecasting. *Science*, 382(6677), 1416–1421. <https://doi.org/10.1126/science.adi2336>
3. Pathak, J., Subramanian, S., Harrington, P., et al. (2022). FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. arXiv preprint arXiv:2202.11214.
4. Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, 28, 802–810.
5. Sønderby, C. K., Espenholt, L., Heek, J., et al. (2020). MetNet: A neural weather model for precipitation forecasting. arXiv preprint arXiv:2003.12140.
6. Ham, Y. G., Kim, J. H., & Luo, J. J. (2019). Deep learning for multi-year ENSO forecasts. *Nature*, 573(7775), 568–572. <https://doi.org/10.1038/s41586-019-1559-7>
7. Espenholt, L., Agrawal, S., Sønderby, C., et al. (2022). Deep learning for twelve hour precipitation forecasts. *Nature Communications*, 13(1), 5166.
8. Rasp, S., Dueben, P. D., Scher, S., Weyn, J. A., Mouatadid, S., & Thuerey, N. (2020). WeatherBench: A benchmark data set for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, 12(11), e2020MS002203.
9. Schultz, M. G., Betancourt, C., Gong, B., et al. (2021). Can deep learning beat numerical weather prediction? *Philosophical Transactions of the Royal Society A*, 379(2194), 20200097.
10. Weyn, J. A., Durran, D. R., & Caruana, R. (2020). Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *Journal of Advances in Modeling Earth Systems*, 12(9), e2020MS002109.
11. Hersbach, H., Bell, B., Berrisford, P., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049.
12. Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378, 686–707.

