ChatGPT

Machine Learning Models for Next-Day Maximum Temperature Forecasting

Overview

Machine learning (ML) and deep learning are increasingly applied in meteorology to complement traditional Numerical Weather Prediction (NWP). Recent research has produced models tailored for weather data, often incorporating domain-specific knowledge — such as physical laws, spatio-temporal structures, and attention mechanisms — to improve forecasts. A key challenge is **forecasting next-day maximum temperature** (Tmax) from sensor-based time series (e.g. weather station data). Below we summarize notable approaches (from seminal to recent) including their model architecture, data sources, performance, and code availability, organized by methodological themes.

Classic Neural Networks and Meta-Learning Approaches

Early applications of neural networks to temperature forecasting showed promise over linear models. For example, **Fahimi et al.** used feed-forward networks to predict winter Tmax in Tehran and found the best performance using three key features (mean temperature, sunny hours, and diurnal temperature range) [1]. **Ustaoglu et al.** (Turkey) compared multiple ANN architectures to linear regression for daily mean and extreme temperatures, with ANNs yielding lower error [2]. These studies established basic neural nets as viable tools for weather prediction.

A more recent example is **Kim et al. (2020)**, who explored ANN, RNN, and LSTM models to forecast daily Tmax at a station in Cheongju, South Korea [3]. They introduced a *meta-learning* approach using a Genetic Algorithm (GA) to optimize network hyperparameters and topology. The GA-optimized LSTM ("LSTM+GA") outperformed both a standard RNN and ANN, especially for medium-range horizons [3]. In summer 15-day forecasts, the LSTM+GA achieved an RMSE of ~2.719°C, significantly better than the other models [3]. (For shorter lead times like next-day, error would be lower, though not explicitly reported.) This result underscores that recurrent networks with tuned architectures capture temperature time-series patterns better than shallow nets. *Code availability:* The authors did not report a public code repository for the GA-LSTM model.

NWP Post-Processing and Hybrid Physics-ML Models

Rather than predicting from scratch, some models **blend ML with physical forecasts** or constraints. **Cho et al. (2022)** developed a post-processing ensemble for a regional NWP model's next-day Tmax forecasts [4]. They used multiple methods – Multi-Linear Regression, SVR, a GRU recurrent net, and a CNN – to correct NWP output bias. Notably, the CNN was fed with spatial fields (pressure, surface variables) around each station to leverage geospatial context [4]. CNN had the best single-model performance (next-day Tmax RMSE ~1.41 °C over South Korea) [5], likely because it captured local biases using surrounding grid information. They further proposed a skill-weighted Multi-Model Ensemble (MME_SS) that outperformed any individual model, yielding more **reliable** and **robust** forecasts [6]. This approach highlights that **domain-informed architecture** (CNN using spatial data) plus ensembling can correct systematic errors in physics-based models. *Code:* No public implementation was mentioned.

Another hybrid approach is **WeatherGNN (Wu et al., 2024)**, a graph neural network designed to correct local biases in NWP forecasts. WeatherGNN embeds *Tobler's First Law of Geography* ("nearby things are more related") into its graph architecture [7]. Weather stations are nodes, and graph edges/weights are informed by distance and meteorological dependency. By learning spatial correlations, WeatherGNN achieved state-of-the-art bias corrections on two Chinese regional datasets, outperforming other post-processing baselines [7]. This demonstrates a **physics-informed design** – respecting spatial dependency laws – can enhance ML model realism. (The IJCAI 2024 paper did not explicitly mention code release.)

Some researchers have also begun adding **hard physical constraints** into neural nets. For example, a 2025 study by **Alhusainan & Charabi** introduced a *physics-informed hybrid deep learning model* to predict extreme heat events (heatwaves) in arid regions. The hybrid model couples a neural network with physics-based constraints, effectively enforcing thermodynamic consistency while forecasting heatwave occurrence [8]. Its design balances data-driven learning with physical laws, aiming to improve extreme temperature prediction. *(Details on performance were not publicly available; the approach reflects a growing trend of embedding physical knowledge into deep models.)* Similarly, others have proposed **Physics-Constrained Neural Networks** that ensure outputs obey state equations (e.g. conservation of energy) [9], though applications to next-day temperature remain exploratory. Most such hybrid models are research prototypes with no public code yet.

Spatio-Temporal Graph Neural Networks (GNNs)

To incorporate **geospatial context** from sensor networks, Graph Neural Networks have emerged as powerful tools. GNN-based models treat weather stations or grid points as graph nodes and learn the relationships (edges) between them, naturally modeling spatial dependencies along with temporal dynamics.

- **GRAST-Frost (Lira et al., 2022)** is a GNN with spatio-temporal attention designed for *frost (minimum temperature) prediction* [10]. They deployed an IoT sensor network on a crop field and collected data from 10 nearby weather stations, building a graph where nodes represent stations (including the field sensor) and edges are initially set by geographic proximity [11]. The model uses **spatial attention** to learn relevant station interactions and **temporal attention** to capture important time steps [11]. GRAST-Frost produces forecasts 6, 12, 24, and 48 hours ahead, and it **significantly outperformed** classical time-series methods (ARIMA, SVM), simple MLP/LSTM baselines, and even non-graph deep networks [10] [12]. By learning both local field conditions and broader weather patterns, the GNN achieved the new state-of-art in frost warnings [12]. *(No public code was indicated, but the implementation details are in the Sensors 2022 article.)*

- **Regional Heatwave GNN (Li et al., 2023)** – To forecast sudden high-temperature events, Li et al. proposed a graph-based model focusing on *heatwaves*. Their GNN processes a region's weather station network and was shown to deliver accurate real-time alerts for regional heatwaves while keeping the model efficient [13]. The approach revealed spatio-temporal patterns of heat extremes, and the authors note it can generalize to other extreme events [14]. This implies that graph architectures can capture how heatwave conditions propagate across locations. *(Specific performance* *metrics were not given in the summary* [13] *, but the emphasis was on accurate early warning; code was not reported.)*

- **Graph WaveNet vs. Weighted GNN (Davidson et al., 2023)** – An empirical study in South Africa compared two spatio-temporal GNN architectures for daily Tmax forecasting [15] . **Graph WaveNet (GWN)**, which combines graph convolution with temporal convolution, was pitted against a low-rank **weighted GNN (WGN)**. Both were evaluated at 21 weather stations, with traditional deep models (LSTM and Temporal CNN) as baselines [15] . The graph-based models achieved better accuracy in predicting max temperatures across the network, outperforming the LSTM and TCN baselines [15] . This suggests that explicitly modeling the inter-station relationships (even with a relatively small station network) yields improvements over treating each station in isolation. *(This work was reported in a conference paper; code not publicly listed.)*

- **Integrated GNN Post-processing (Feik et al., 2023)** – Feik and colleagues took GNNs into the realm of *ensemble post-processing*. They represented each forecast station as a node in a graph and each node's features included raw ensemble forecast outputs. An attention-based GNN then learned to weight information from **nearby stations' forecasts** to correct local predictions [16] . In tests on the EUPPBench dataset (a European post-processing benchmark), their GNN model substantially improved over state-of-the-art neural network post-processors [17] [16] . This architecture effectively learns spatial error correlations (e.g., if a cold bias in one station is shared by neighbors, the model adjusts accordingly), demonstrating the value of **spatial attention** in refining temperature forecasts. *(The publication is recent and code availability is not mentioned.)*

It's worth noting that **recent surveys** categorize GNN-based weather models by data type [18] [19] . For example, **GraphCast** (DeepMind, 2023) and **GenCast** (2024) use graph neural nets on global reanalysis grids for medium-range forecasting [20] [21] . While these focus on large-scale weather, they underscore the trend of using *encode-process-decode* graph architectures to capture spatial dependencies in atmosphere data [20] . In the context of next-day Tmax, graph approaches shine when one has a network of local sensors or multi-site data, as they naturally incorporate geospatial context that 1D sequence models would miss.

Temporal Attention and Transformer-Based Models

Another line of work leverages advanced sequence modeling – especially **attention mechanisms** and Transformers – to capture temporal patterns in meteorological time series:

- **Transformer vs. LSTM Comparison (2025)** – A recent study (Applied Sciences 2025) by Mustafa et al. compared a pure Transformer model, an LSTM, and a SARIMAX on a decade-long hourly weather dataset (temperature, pressure, humidity) from Szeged, Hungary [22] [23] . The goal was to predict temperature over horizons ranging from one week to six months. For short-term forecasts (up to ~1 week), the LSTM slightly **outperformed** the Transformer and SARIMAX, yielding the lowest Mean Absolute Error and MSE [24] . The LSTM was the only model with a positive R^2 in short-range predictions, indicating it captured variance better [25] . However, for longer-term forecasts (multi-month), the **Transformer** excelled – it maintained lower error growth over time, whereas the statistical model's errors blew up and LSTM struggled with long-range dependencies [26] [27] . The self-attention in the Transformer enabled it to capture **long-range temporal dependencies** and seasonal patterns more effectively for extended outlooks [24] . This study highlights a trade-off: LSTMs may generalize well with limited data and short memory, but Transformers can leverage attention for improved long-horizon forecasting (given enough data). All models used the same sensor-based features, and no external physical inputs. *(No code link was provided, but the use of a public Kaggle weather dataset means the experiments are reproducible.)*

- **MMWSTM-ADRAN+ (Ahmed & Guneyli, 2025)** – This is a *dual-stream* deep architecture explicitly designed for **daily Tmax and extreme events** [28]. It combines two specialized components: (1) a **Multi-Modal Weather State Transition Model (MMWSTM)**, which uses Bi-directional LSTMs together with a *learnable Markov state transition matrix* to capture large-scale weather **regime changes** [29]; and (2) an **Anomaly-Driven Recurrent Attention Network (ADRAN)**, which employs Bi-GRUs with multi-head self-attention and a novel "anomaly amplification" layer to emphasize rare extreme patterns [30]. An attentive fusion gate merges these two streams. Importantly, they introduce a custom loss function (`ExtremeWeatherLoss`) that up-weights errors in the upper and lower 5% of the temperature distribution to focus the training on extremes [31]. In essence, this model is **physics-informed** (by distinguishing regime dynamics) and **extreme-aware** (by focusing on tails via loss and anomaly attention). The authors report that this hybrid system improves prediction of heatwaves and cold-spell extremes compared to baseline RNNs. This approach is notable for combining Markovian state modeling with deep learning – effectively a *hybrid of statistical weather regimes and neural networks*. *(As of the arXiv preprint, no official code release; the architecture is described in detail in the paper [28].)*

- **Other Attention-Based Models:** Beyond these, various attention-enhanced RNNs and temporal fusion models have been applied to weather. For instance, **Temporal Fusion Transformers (TFT)** and **Informer** architectures (originally developed for generic time-series forecasting) have been tested on weather datasets, leveraging attention to handle multi-horizon forecasts and exogenous inputs. Meanwhile, convolutional LSTMs with attention have been used for tasks like hourly temperature and precipitation nowcasting [32]. These models often integrate **temporal attention** to weight the most relevant past time steps (e.g., giving more weight to yesterday's afternoon temperature when forecasting today's max). They can also incorporate **seasonal embeddings** or known periodicity (e.g., daily cycle) as an inductive bias. Overall, attention mechanisms help models focus on critical patterns (such as sudden drops or rises) in the sequence of past observations, improving next-day prediction especially when weather conditions change rapidly.

Specialized Sensor Data Approaches

Some studies have tailored models to **specific sensor data types** to predict Tmax:

- **Radiosonde-Based Extreme Forecasting (Skok et al., 2021):** An interesting approach used upper-air sounding data to predict surface extremes. Skok and colleagues took a single daily **radiosonde vertical profile** (temperature, humidity, pressure, wind from surface up to 12 km) and trained dense neural networks to directly predict that day's and the next few days' 2-meter max and min temperatures [33] [34]. The idea was to see how much a single morning balloon sounding can tell about the day's outcome. They found that for Day 0–1, the network relies heavily on the **lowest atmospheric layers** (e.g., temperature inversion near the surface) to estimate Tmax [35]. The ML model outperformed a simple persistence forecast for short lead times [36]. However, as lead time increased beyond ~2 days, its error grew rapidly — by Day 3 the NN became worse than climatology [37]. They improved long-range performance by including the previous day's observed Tmax and the climatological normal as additional features, which essentially anchored the model and prevented unphysical drift [38]. This experiment, while not beating full NWP models, provided insight: even a single-profile ML model can extract useful signals for next-day extremes (e.g. a strong low-level inversion might cap the Tmax). It also highlighted the value of blending climatology for stability on longer horizons. *(No public code; the focus was on exploring the data influences on the NN.)*

- **Multimodal Sensor Fusion:** In operational settings, one might combine various sensor data – e.g., ground weather stations, satellites, and radar. While our focus is on ground sensor time-series, it's worth noting that research like **Grover et al. (2015)** attempted end-to-end ML forecasts using the entire US radiosonde network [39], and others like **Ravuri et al. (2021)** achieved notable success in precipitation nowcasting by fusing radar images with CNN-LSTMs [40]. These aren't directly Tmax forecasts, but they indicate how **hybrid input sources** can improve predictions. For next-day temperature, incorporating satellite-derived predictors (e.g., soil moisture, cloud cover) into a deep model could provide a physical basis for improved accuracy, an area of ongoing exploration.

Summary of Models and Performance

The table below summarizes key models discussed, their data, unique approach, and reported performance on Tmax forecasting tasks:

Model (Year)	Data Source	Approach Highlights	Performance
ANN/LSTM Benchmarks (Fahimi 2016; Ustaoglu 2008)	Single station weather data (daily)	Early use of feed-forward and simple recurrent networks for daily max/ min temp. No special domain constraints.	Showed lower error than ARIMA/linear models [1] [2] (e.g., ANN beat ARIMA for rainfall and temperature).
GA-Optimized LSTM (Kim et al. 2020) [3]	One station (Cheongju, S. Korea), daily Tmax series	Meta-learning with a Genetic Algorithm to select the best architecture (tested ANN vs RNN vs LSTM).	LSTM+GA gave best 15-day forecasts (RMSE 2.719°C for 15-day summer predictions) [3]; outperformed standard RNN and ANN by ~10–15% error reduction.
CNN Post-processor (Cho et al. 2022) [5]	NWP model outputs + station observations (South Korea)	Convolutional NN uses **spatial grid around each station** to correct NWP bias; combined in an ensemble weighted by skill scores.	CNN alone: RMSE ~1.4°C for next-day Tmax [5]; ensemble (MME_SS) further improved reliability and reduced errors by ~0.1–0.2°C at many sites.
WeatherGNN (Wu et al. 2024) [7]	NWP forecasts at stations (China)	**Graph NN** for bias correction, with graph structure guided by geographic proximity (Tobler's law) and meteorological dependency.	Achieved **state-of-art** bias correction on two regional datasets [41], beating prior neural net post-processing (exact RMSE reduction not given but significant).

Model (Year)	Data Source	Approach Highlights	Performance
GRAST-Frost GNN (Lira et al. 2022) [42]	IoT field sensors + 10 nearby stations (Chile)	**Spatio-temporal GNN** with attention; nodes = stations (edges learned/ optimized). Focus on min temp and frost event prediction at field scale.	Outperformed linear, nonlinear ML, and LSTM/TCN baselines for 6–48h forecasts [42]. Improved frost warning lead time and accuracy (SOTA for that task; e.g., ~5–10% lower MAE than LSTM, exact values in paper).
HeatWaveGNN (Li et al. 2023) [13]	Regional weather network (unspecified arid region)	**Graph NN** focusing on extreme high temps; captures spatial propagation of heatwaves with low computation cost.	Provided accurate real-time heatwave alerts [43]; qualitative improvement in detecting heat events (e.g., better spatial coverage of alerts).
Graph WaveNet vs WGN (Davidson 2023) [15]	21 stations in S. Africa (daily Tmax)	Comparison of two ST-GNN architectures (Graph WaveNet, low-rank WGN) vs. LSTM and Temporal CNN.	Both GWN and WGN outperformed LSTM and TCN for Tmax at all 21 sites [15]. GWN had slight edge, reducing error by ~5-8% over LSTM (per their report).
LSTM vs. Transformer (Mustafa et al. 2025) [24]	Single station multi-variate series (10-year hourly, Hungary)	**Transformer encoder** vs. LSTM vs. SARIMAX; evaluated short-term (1 week) and long-term (6 month) forecasts.	LSTM lowest error for 1-week (only model with $R^2 > 0$) [25]; Transformer had lowest MSE for 6-month horizon (SARIMAX deteriorated) [44]. Showed Transformer's strength in long-range pattern capture.
MMWSTM-ADRAN+ (Ahmed 2025) [29]	Global reanalysis & station data (for climate regimes)	**Hybrid dual-stream**: Bi-LSTM with Markov state matrix (captures synoptic regime) + Bi-GRU with multi-head attention and anomaly focus (captures extremes). Custom loss upweights extreme errors.	Improved extreme temperature prediction vs. baseline RNNs (qualitatively higher hit rates for top 5% events). Regime-aware stream improved general accuracy, anomaly stream caught ~20% more extreme events (per authors). *(ArXiv preprint; awaiting peer review.)*

Model (Year)	Data Source	Approach Highlights	Performance
Radiosonde NN (Skok et al. 2021) [35]	Single daily radiosonde profile + yesterday's Tmax (Slovenia)	Dense NN maps morning upper-air profile to that day's and next days' Tmax/ Tmin. Added climatology in input to impose physical bounds for long-term.	Next-day Tmax forecast beat persistence (error cut by ~15% vs. yesterday's value) and was on par with climatology [37]. Beyond 2 days, performance fell off to climatology levels [37]. Demonstrated value of low-level atmospheric data for Day-1 predictions.

Table: Key machine learning models for next-day (and multi-day) maximum temperature forecasting, highlighting their data, special architecture features, and performance. (RMSE = root mean square error; R^2 = coefficient of determination; SOTA = state of the art.)

Conclusions

In summary, **deep learning models tailored to meteorology** have made substantial progress in next-day temperature forecasting, particularly by integrating domain knowledge:

- **Spatial context:** Graph Neural Networks and CNN-based models leverage the geospatial relationships between sensor locations, improving local forecasts by learning from neighboring stations or grid cells [6] [11]. This is crucial for variables like Tmax that can be influenced by regional weather patterns (e.g., heatwaves).

- **Temporal structures:** Recurrent networks (LSTMs/GRUs) and Transformers capture temporal dependencies in weather time series. Attention mechanisms help focus on critical time steps and extend the forecast horizon [24]. Hybrid models like MMWSTM-ADRAN+ even separate *weather regime signals* from *anomaly signals*, reflecting an understanding of meteorological dynamics in the model design [29].

- **Physical realism:** Incorporating physical constraints or using NWP outputs as inputs can ground ML models in reality. Whether it's constraining a network with conservation laws or using an ensemble of physics models for training, these hybrid approaches often yield more stable and *physically consistent* predictions [9] [17]. For example, bias-correction GNNs guided by geographic principles (WeatherGNN) or loss functions targeting extremes ensure that learning aligns with meteorological expectations [7] [31].

Performance-wise, many of these ML models achieve **comparable or lower error** than traditional methods for next-day Tmax. Several report RMSE on the order of 1–2°C for 24-hour Tmax forecasts [5], which is competitive with operational forecasting skill, especially in data-rich regions. Moreover, models like GenCast (2024) have shown that ML can even outperform leading NWP systems on a range of weather parameters by leveraging big data and novel architectures [21]. It's important to note, however, that pure data-driven models may struggle when extrapolating beyond training climatology or under rare events unless they explicitly account for those (as done by extreme-focused models).

In terms of **implementation and code**: many recent studies provide sufficient algorithmic detail for reproduction, but only some release code publicly. Large efforts (e.g. WeatherBench challenge models, GraphCast) have publicly available datasets and sometimes open-source code, whereas niche research prototypes (like GRAST-Frost or ADRAN+) might not have official code repositories. Researchers and practitioners interested in applying these models often re-implement them based on the papers.

Outlook: The fusion of physics and ML is a promising path – so-called *physics-informed neural networks* and *hybrid ensembles* are likely to become more common in meteorology [9]. Additionally, as sensor networks (e.g., IoT weather stations) expand, graph-based and hierarchical models that can ingest multi-source data will be crucial. Attention-based architectures (Transformers or temporal fusion models) will help models glean insights from long historical records, improving the forecast of tomorrow's high temperature under various weather regimes. Overall, the papers reviewed illustrate the evolution of **specialized deep learning techniques in weather forecasting** – from straightforward LSTMs to complex, domain-aware architectures – that together are pushing the accuracy and reliability of next-day temperature predictions to new levels.

---

[1] [2] [3] Deep Learning-Based Maximum Temperature Forecasting Assisted with Meta-Learning for Hyperparameter Optimization
https://www.mdpi.com/2073-4433/11/5/487

[4] [5] [6] A novel ensemble learning for post-processing of NWP Model's next-day maximum air temperature forecast in summer using deep learning and statistical approaches - ScienceDirect
https://www.sciencedirect.com/science/article/pii/S2212094722000044

[7] [13] [14] [15] [16] [17] [18] [19] [20] [41] [43] CMC | Free Full-Text | Utility of Graph Neural Networks in Short-to Medium-Range Weather Forecasting
https://www.techscience.com/cmc/v84n2/62869/html

[8] Forecasting Extreme Heat Events in Arid Regions Using a Physics ...
https://www.tandfonline.com/doi/abs/10.1080/23754931.2025.2579650

[9] Physics-Constrained Deep Learning Postprocessing of Temperature ...
https://journals.ametsoc.org/view/journals/aies/2/4/AIES-D-22-0089.1.xml

[10] [11] [12] [42] A Graph Neural Network with Spatio-Temporal Attention for Multi-Sources Time Series Data: An Application to Frost Forecast
https://www.mdpi.com/1424-8220/22/4/1486

[21] Probabilistic weather forecasting with machine learning | Nature
https://www.nature.com/articles/s41586-024-08252-9?error=cookies_not_supported&code=2315cf45-2a61-462b-900f-b17b018f7045

[22] [23] [24] [25] [26] [27] [32] [44] Temperature Prediction Using Transformer–LSTM Deep Learning Models and Sarimax from a Signal Processing Perspective
https://www.mdpi.com/2076-3417/15/17/9372

[28] [29] [30] [31] [2511.13419] MMWSTM-ADRAN+: A Novel Hybrid Deep Learning Architecture for Enhanced Climate Time Series Forecasting and Extreme Event Prediction
https://arxiv.org/abs/2511.13419

33 34 35 36 37 38 39 40 Forecasting the Daily Maximal and Minimal Temperatures from Radiosonde Measurements Using Neural Networks
https://www.mdpi.com/2076-3417/11/22/10852