ChatGPT

Deep Learning Models for Daily Maximum Temperature Prediction

Introduction

Accurate prediction of daily maximum temperature is crucial for agriculture, energy management, public health, and disaster preparedness [1] [2]. Traditionally, numerical weather prediction (NWP) models are used, but they can struggle with complex local effects and are computationally expensive [3] [4]. In recent decades, *data-driven deep learning* approaches have emerged as powerful alternatives for temperature forecasting, offering improved accuracy in capturing nonlinear patterns [5] [6]. Researchers have developed a range of deep learning models – from recurrent neural networks to hybrid convolutional models and transformers – to predict daily high temperatures across various regions. Below, we summarize key models from literature and practice, noting their architectures, inputs, geographic scope, performance, code availability, and any real-world deployments.

Recurrent Neural Network (RNN) and LSTM-Based Models

Early Neural Networks: Initial attempts to forecast daily max temperature used "shallow" machine learning. For example, Support Vector Machines (SVMs) were shown to predict next-day maximum temperature with optimized kernels [7]. Likewise, simple Artificial Neural Networks (ANNs) in the 1990s–2000s achieved some success in daily temperature prediction [8] [9]. These early models demonstrated the feasibility of machine learning but were limited in capturing long-term temporal dependencies.

LSTM and GRU Models: Modern deep RNNs like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs) are well-suited for time-series weather data. LSTMs can retain long-term information and have become a **state-of-the-art** baseline for temperature forecasting [10] [11]. For example, Park *et al.* (2020) optimized an LSTM with a genetic algorithm to forecast daily max temperature 5–15 days ahead. Their *GA-LSTM* model, tested on a station in South Korea, outperformed a standard RNN and ANN, yielding a **15-day-ahead summer RMSE of 2.719 °C** – notably lower error than the other models [10]. In Egypt, Houssein *et al.* (2025) introduced a **Deep LSTM optimized with a hybrid Genetic Algorithm – Mountain Gazelle optimizer (MGO-GA-DLSTM)** for nine cities [12] [13]. This deep LSTM model, tuned for each location, achieved very high accuracy (often **R^2 > 0.95** across cities) [14]. For instance, on the Asswan city dataset it obtained **RMSE ≈1.75 °C**, compared to 2.24 °C with an untuned LSTM [15]. These results underscore LSTM's ability to capture nonlinear temperature dynamics when properly optimized. GRU-based variants have similarly shown strong performance; in the Egyptian study, a GA-tuned GRU was competitive and even slightly outperformed LSTM in a few cities [16] [17].

Hybrid Ensembles: Researchers have also combined multiple recurrent models to improve robustness. A recent example is *TempFusionNet* (Licer *et al.*, 2025), designed for the Moroccan Sahara's climate [18]. TempFusionNet fuses a **Temporal Convolutional Network (TCN)** with GRU and LSTM modules, leveraging each model's strengths. It was tested on one-day-ahead max temperature forecasting for four desert cities.

Using metrics like MAE, RMSE, and R^2, the authors report TempFusionNet significantly **outperformed traditional statistical methods** for all cities [18]. This highlights that ensemble or hybrid deep learning can capture temperature patterns better than any single model, especially in challenging climates (extreme heat, sparse data). Another hybrid approach coupled LSTM with AdaBoost ensemble learning to predict short- to mid-term sea surface temperatures, which improved stability and reduced overfitting in the predictions [19]. Overall, recurrent networks (especially LSTM/GRU) form a backbone for many temperature predictors, often achieving **1–3 °C RMSE** for daily maxima depending on forecast horizon and region. Their inputs are typically past temperature observations (daily max series) and sometimes additional local features like humidity or solar radiation (when available) to inform the model [20].

Convolutional Neural Networks and CNN–LSTM Hybrids

While RNNs excel at temporal patterns, **Convolutional Neural Networks (CNNs)** can capture spatial features that influence temperature (e.g. pressure patterns, topography). CNN-based models have been used both on gridded weather data and on sequences of observations. **Ali & Cheng (2024)** developed a CNN model to forecast extreme surface temperatures over the continental U.S. [21]. Their model took large-scale atmospheric maps – specifically **geopotential height fields at 100–925 hPa levels** – as input features to predict 2-meter daily max temperatures up to 30 days out. In case studies of the 2012 Central U.S. heatwave and the 2014 Midwest cold wave, the CNN learned which pressure levels most influence surface temperature. For example, using the 500 hPa height yielded the lowest errors for many lead times [22]. They reported **RMSE on the order of 1.4 °C for 1-day lead**, increasing to about **2.9 °C at 30-day lead** during the summer heatwave period [23]. Including multiple pressure levels only marginally improved accuracy beyond the best single-level input [24]. This study shows CNNs can leverage synoptic patterns (e.g. ridges, troughs) to predict regional temperature extremes with good skill, even at sub-seasonal ranges.

CNN-LSTM Integrated Models: A powerful design is to combine CNNs (for feature extraction) with LSTMs (for sequence learning) in one framework. This CNN–LSTM hybrid approach has proven effective for temperature time-series. For instance, a study in Ningxia, China built a CNN–LSTM model to predict hourly temperature, using CNN layers to reduce dimensionality of meteorological inputs and LSTM layers to capture long-term dependencies [25] [26]. The model was trained on 20 years of hourly data and compared to standalone CNN and LSTM. The hybrid showed superior accuracy (evaluated by MAE, MAPE, etc.), indicating that **CNN feature extraction plus LSTM memory** yields more precise forecasts than either alone [27] [28]. Similarly, Zhang and Dong (2020) and Tabrizi *et al.* (2021) integrated CNNs with recurrent networks into "ConvLSTM" or convolutional RNN models, enabling learning of both spatial and temporal correlations in daily temperature changes [29]. These hybrids have been applied to multi-step forecasting and even heatwave prediction. For example, Al-Najjar *et al.* (2019) coupled a Graph Attention Network (a type of spatial attention model) with a GRU recurrent net to forecast regional heatwaves in the Middle East, improving short-, medium-, and long-range prediction of daily maximum temperatures [30]. These approaches often ingest gridded historical observations (or graph-based representations of station networks) as input, enabling the model to understand how temperature at one location is influenced by conditions at others.

Advanced CNN Architectures: Deeper CNNs and pre-trained networks have also been explored. A recent Scientific Reports study proposed an **enhanced CNN-ResNet50-LSTM** model for climate forecasting [31]. This hybrid included a *ResNet-50* deep convolutional module alongside a CNN and LSTM, and was used to predict daily temperature (as well as wind power) in multiple datasets. On a Saudi Arabia weather dataset, the CNN-ResNet50-LSTM achieved an **R^2 of 0.9901 (99.01%) for temperature forecasting**, dramatically outperforming baseline regression models [32]. Its RMSE and MAE were correspondingly low (not explicitly stated, but such an R^2 implies errors on the order of a degree or less). The model was even extrapolated to predict climate trends up to the year 2030, demonstrating its potential for scenario forecasting [33]. The success of this deep CNN+LSTM ensemble suggests that very high accuracy in daily max temperature prediction is achievable when ample data and architecture depth are available. It also underlines the importance of **input features** – in this case, the model likely benefited from both local station data and possibly reanalysis grids included in the Saudi dataset, as well as the transfer learning from ResNet (which may have been pre-trained on large datasets).

In summary, CNN-based models are especially useful when **spatial context** (neighboring region weather, pressure fields, etc.) is relevant to the temperature forecast. They have been applied globally (e.g. U.S. extreme heat/cold forecasts [21]) and regionally (e.g. China, Middle East) with strong results. Many implementations use CNNs together with recurrent layers or other neural units to form *spatio-temporal* models. Typical performance of CNN hybrids for short-term (next day to 1 week) daily max temperature forecasting is often **RMSE ~1–2 °C**, and they often handily outperform traditional statistical models like autoregression or regression kriging in head-to-head comparisons [34].

Transformer and Attention-Based Models

In the last few years, the introduction of **transformer architectures** and other attention-based models has pushed the frontier of weather prediction, including daily temperature forecasting. Transformers can handle long sequences and multi-feature inputs flexibly, making them attractive for climate data.

Time-Series Transformers: Some studies have applied transformers to station-level or regional temperature series. *Kişmiroğlu & Işık (2025)* compared a standard LSTM against a transformer-based model for ambient temperature prediction at horizons of one week to six months [35]. Their models used multiple meteorological inputs (humidity, pressure, past temperatures) to predict future temperatures, essentially framing it as a multivariate time-series problem [36]. They found that the transformer with self-attention mechanisms could capture seasonal patterns and long-term trends that LSTM sometimes missed, especially for longer lead times. Both models outperformed a classic SARIMAX statistical model in accuracy [37]. This indicates that transformer models are effective for **seasonal or medium-range temperature forecasting**, where long-term context matters.

At a higher temporal resolution, **Khan & Al-Hajj (2025)** used a *Frequency Transformer* (FTTransformer) for **5-minute interval temperature prediction** in Kuwait [38] [39]. This is an extreme case of high-frequency "weather" data. They fed 30 environmental variables (including humidity, solar radiation, wind, etc.) into a transformer and LSTM to simultaneously predict six output variables (air temperature at different heights, surface temperatures, etc.) [40]. Both the transformer and LSTM achieved remarkable accuracy (mean **MAE ~0.24 °C** with $R^2 \approx 0.998$ for the 5-min ahead predictions) [39]. However, a notable finding was that **LSTM's performance degraded on anomalous data from previous years, whereas the transformer remained stable** [41]. This highlights a strength of transformers: the ability to maintain accuracy across regime shifts or out-of-distribution inputs, thanks to their attention-based global view of the sequence. Although this study was sub-daily, it implies that for daily extremes, transformers might be more robust than RNNs when dealing with climate anomalies or rare events.

Global Forecasting Models: The most high-profile advancements have come from global, data-driven weather models that often use transformer or attention components. These models are trained on reanalysis data (gridded historical weather) and aim to predict full weather fields (including surface temperature) worldwide. A prime example is **GraphCast (DeepMind, 2023)** – a graph neural network model with attention mechanisms for medium-range forecasting [42]. GraphCast represents the Earth as a graph and was trained on 40 years of ERA5 reanalysis. It can predict hundreds of weather variables (e.g. temperature, pressure, winds) up to **10 days ahead**, with **unprecedented accuracy that surpasses the ECMWF high-resolution (HRES) NWP model** [6]. In fact, GraphCast was shown to **outperform the operational ECMWF forecast on about 90% of verification metrics (including 2-m temperature)** in a comprehensive evaluation [43] [6]. This is a stunning result, as ECMWF's model is considered a gold standard. GraphCast's architecture is built on Graph Neural Network layers (which effectively act like a transformer over a spherical mesh) and it runs extremely fast – producing a 10-day global forecast in under a minute [44] [45]. Importantly, GraphCast has been **open-sourced and is being tested in real-world forecasting**: the code is available on GitHub and **ECMWF is running a live experiment displaying GraphCast forecasts on their website** [46]. This marks one of the first instances of an AI model being trialed by a major weather agency for potential operational use. Although GraphCast is not exclusively a "daily max temperature" model (it forecasts hourly conditions which can be aggregated to daily maxima), its success implies that deep learning models can provide highly accurate daily temperature predictions on a global scale. It also demonstrates generalization: GraphCast was trained globally and applies to **all regions including the U.S.** without retraining.

Around the same time, **Pangu-Weather (Bi *et al.*, 2023)** was introduced by Huawei. Pangu-Weather uses a 3D *Earth-specific Transformer* architecture to perform fast global forecasts up to 15 days [47]. It was published in *Nature*, reporting that Pangu's predictions are **comparable to or better than ECMWF's**. For instance, Pangu achieved a *forecast skill gain* of about **10–15 hours over ECMWF's IFS** (i.e. it maintains the same error level roughly half a day further into the future) for several variables [48]. On some parameters like humidity, its advantage was even greater. The model is also **extremely efficient**, being over 10,000× faster than IFS due to the inference speed of neural networks [49]. Pangu-Weather's code and model weights have been made available openly [50], and its success sparked significant interest in AI weather models in China and beyond. Subsequent evaluations (e.g., by NOAA and CMA) have shown Pangu and similar models (often nicknamed "FuXi" or "FengWu" in evaluation studies) performing at or above the level of leading NWP models for medium-range temperature forecasting [51]. This indicates that **transformer-based models can rival traditional physics models** in predicting daily temperature patterns globally.

Another notable model is **FourCastNet (NVIDIA, 2022)**, which took a slightly different approach using a Fourier Neural Operator (a deep learning method that learns mappings in the frequency domain). FourCastNet is essentially a very deep CNN/transformer hybrid operating on spectral coefficients. It was trained on historical weather data to predict global fields like 500 hPa height and 2-m temperature. FourCastNet can generate a **1-week forecast in <2 seconds**, vastly faster than numerical models [52]. In terms of accuracy, FourCastNet's early version achieved **comparable skill to ECMWF's operational model out to ~3–5 days** for key variables [53]. Beyond 5 days, its error grew faster than ECMWF's, but ongoing improvements (FourCastNet v2/v3) have been closing that gap. The **open-source release** of FourCastNet's code and pre-trained weights [54] has allowed many researchers to experiment with it on custom problems, and even NOAA and ECMWF have run tests integrating FourCastNet into their workflows [55]. Although FourCastNet is a global model, it can specifically predict daily max/min temperature by taking the daily highs from its hourly outputs.

Lastly, **ClimaX (Microsoft, 2023)** represents a new wave of **foundation models** for climate. ClimaX is built on the Vision Transformer (ViT) architecture and pre-trained on a wide range of climate and weather data

[56] [57]. It's designed to be fine-tuned for various tasks – from local short-term forecasts to long-term climate projections. On benchmarking, ClimaX has matched or exceeded state-of-the-art models in several scenarios [58]. Notably, ClimaX showed *particularly strong performance in temperature prediction tasks*, surpassing other models in accuracy for 2-m air temperature, though it was less successful on precipitation extremes [59]. This suggests the ViT-based approach is very adept at capturing the relatively smooth, spatial patterns of temperature fields. ClimaX is fully open-source and comes with pre-trained weights and code on GitHub [60], lowering the barrier for others to apply deep learning to their own temperature datasets. Its ability to handle **heterogeneous inputs** (via a novel variable tokenization mechanism) means one could incorporate satellite images, reanalysis maps, and station data together for training [61] – a flexibility that can be useful for predicting daily max temperatures in data-sparse regions.

Performance of Transformers: The large attention-based models like GraphCast, Pangu, FourCastNet, and ClimaX are generally evaluated with standard meteorological metrics – e.g. root mean square error (RMSE) and anomaly correlation at various forecast hours. In terms of daily temperature: these models have demonstrated **RMSE on the order of 1–2 °C for 1–3 day forecasts**, and maintain usable accuracy (RMSE a few degrees) even at 10–15 days globally [23] [51]. They often equal or beat the best NWP in this regard. The transformer models also excel at capturing extreme events; for example, GraphCast showed improved predictions of heatwave onset and magnitude [62], which is directly relevant for daily max temperature extremes. This class of models is rapidly evolving, and hybrid approaches (e.g., combining physics with ML, or ensemble averaging multiple ML models) are under exploration to further improve reliability.

Data Inputs and Sources for Models

Deep learning models for temperature prediction have been trained on a variety of data sources depending on the scope:

- **Station Observations:** Many local models (especially older ANN/LSTM studies) use historical surface observations of daily max temperature at one or multiple stations. Sometimes additional local features are included, such as daily min temperature, mean temperature, humidity, wind speed, sunshine duration, etc., if correlated with the max. For instance, a neural network study in Tehran found that using three inputs – daily mean temperature, hours of sunshine, and the previous day's temperature range – yielded the best results for winter max temperature forecasting [20]. Similarly, Ustaoglu *et al.* (2011) used neural networks to predict daily mean, max, and min temperatures in Turkey and found they outperformed multiple linear regression when given sufficient past data [63]. These point-focused models usually require at least several years of daily data for training.

- **Reanalysis and Gridded Climate Data:** Regional and global models (CNN or transformer-based) typically rely on gridded datasets like **ERA5 reanalysis** (which provides global hourly weather variables on a 30 km grid) [64]. Reanalysis data combine observations and model outputs to create a comprehensive weather record, making them ideal for training AI. Models like GraphCast and Pangu-Weather were trained on decades of reanalysis comprising variables such as pressure, temperature at various heights, humidity, etc. Inputs from different pressure levels (e.g. z500, z850) can help the network learn the 3D atmospheric context for surface temperature [21]. These data allow models to be **generalized across regions** – e.g. a single GraphCast model covers the entire globe.

- **Satellite and Remote Sensing:** Some studies incorporate satellite-derived data as proxies for solar radiation or cloud cover, which influence daily highs. For example, the Kuwait multi-output model included **solar radiation and dew point** among 30 features to help predict temperature [40]. Remote sensing imagery (like land surface temperature from satellites) could also be used to downscale or calibrate forecasts, although in practice many deep models stick to structured data (grids or station time series) due to ease of use. There is growing interest in using high-resolution satellite data to improve local temperature forecasts, especially for urban heat islands.

- **NWP Model Outputs:** A hybrid approach is using outputs from a physics-based NWP as inputs to a neural network. This leverages physical knowledge while letting the network correct biases. For instance, Frnda *et al.* (2016) trained an ANN to adjust ECMWF model forecasts and reported improved temperature predictions in mountainous terrain [65]. More recently, Zhang *et al.* (2022) applied machine learning (including gradient boosting and CNNs) to post-process a Chinese operational model (GRAPES) and reduce its temperature forecast errors [66]. Thus, deep learning can serve in *model output statistics* (MOS) roles, refining traditional forecasts of daily maxima.

- **Climate Model Data:** For long-term projections (monthly to seasonal max temperatures), deep learning has been used to downscale coarse climate model outputs. NOAA researchers, for example, experimented with deep networks to downscale daily max/min temperature and precipitation from ~50 km resolution to higher grid resolutions [67]. These models aren't forecasting the weather day-by-day but rather translating climate projections to local scales, yet they use architectures similar to those in forecasting. They indicate the versatility of deep nets for any task involving daily temperature series.

In summary, **U.S. data and global data** are both well represented in these studies. Some models are region-specific (trained on data from Egypt, Morocco, Korea, etc.), but many are transferable. Global models (GraphCast, Pangu, FourCastNet, ClimaX) explicitly have worldwide applicability and have been validated in the U.S. context (e.g. forecasting U.S. heatwaves [22], or being compared to U.S. GFS model outputs [68]). With appropriate re-training or fine-tuning, an LSTM or CNN model developed in one region can often generalize to another region's climate, provided the input features account for differences (for instance, including elevation or coastal influence if those differ). The incorporation of diverse data sources – surface observations, upper-air reanalysis, satellite estimates – tends to improve model performance by giving a more complete picture of the factors driving daily high temperatures [69].

Performance Metrics and Benchmarking

The performance of deep learning models for temperature prediction is usually reported with standard regression metrics: **Root Mean Square Error (RMSE)** and **Mean Absolute Error (MAE)** for accuracy, sometimes Mean Absolute Percentage Error (MAPE) for relative error, and **R-squared (R^2)** for explained variance [70] [71]. Table-based evaluations are common, comparing these metrics across models and locations.

Typical error magnitudes for next-day **daily maximum temperature** forecasts by modern deep learning models are on the order of **1–2 °C RMSE**, which often beats traditional models. For example, the TempFusionNet in Morocco had errors low enough to yield $R^2 > 0.9$, indicating it explained most variance in the daily highs [34]. The GA-LSTM model for Cheongju, South Korea, when optimized for long-range (15 day) summer forecasts, achieved **RMSE ~2.7 °C**, whereas a standard ANN's error was larger (exact ANN error not given, but LSTM was clearly superior) [10]. In the Egyptian multi-city study, the best hybrid LSTM model showed **MAE around 1.3–1.8 °C** and RMSE around 1.9–2.3 °C (depending on city), compared to RMSE ~2.5+ °C for a plain LSTM without optimization [15]. These numbers reflect very high accuracy, considering daily max temperature can vary 10–15 °C over a season in those climates. Even for more variable climates, deep models tend to maintain a couple of degrees error. The U.S. CNN for the 2012 heatwave had an RMSE ~1.5 °C for short leads [23] – quite good given that daily highs in the dataset were often 35–40 °C.

It's worth noting that **error metrics can depend on forecast lead time and season**. Models like GraphCast and Pangu-Weather report errors as a function of lead time. Early leads (day 1–3) might have RMSE ~1–2 °C globally, whereas by day 10 the RMSE might rise to ~4–5 °C for 2-m temperature (which is still comparable to or better than major NWP models at that range) [72] [51]. Many papers also report **benchmark comparisons**: e.g. showing that a deep model's RMSE is 20-30% lower than that of an ARIMA or linear regression on the same data [73] [74]. For classification-oriented tasks (like predicting heatwaves above a threshold), metrics like accuracy or F1 score are used, but those are less common than continuous error metrics for temperature.

Another aspect of performance is **stability and generalization**. Deep nets can sometimes overfit to training period climate. To address this, studies use techniques like cross-validation across years (as in the Kuwait 5-min study, which used a leave-one-year-out validation to test generalization to an unseen year's data [75]) or early stopping and ensembling. When done properly, deep models show robust performance even on anomalous years (the Kuwait transformer had consistent R^2 ~0.998 each year [39]). Some degradation can occur in extreme out-of-sample conditions (e.g. an unprecedented heatwave beyond training range), but the inclusion of recent extreme events in training data and the ability of architectures like transformers to extrapolate patterns have mitigated this to an extent [41].

In terms of **benchmark datasets**, there isn't a single universal one for daily max temperature, but some initiatives exist. The **WeatherBench** dataset (Rasp *et al.*, 2020) provides a standardized set of global weather data for ML, including temperature at 850 hPa and 2 m, which has been used to benchmark models like ClimaX and FourCastNet [76]. Results on WeatherBench confirm that deep learning models can achieve similar RMSE to coarse-resolution NWP for 3-5 day forecasts of surface temperature, and newer models keep pushing these numbers down. For localized forecasting, competitions (like those on Kaggle) and studies often use their own curated datasets (e.g., 50 years of a city's daily highs). In all cases, **deep models consistently rank at or near the top in predictive performance** compared to classic time-series models or even simplified physics models, especially when evaluated by RMSE/MAE of temperature predictions [58].

Open-Source Models and Case Studies

One advantage in this field is that many researchers have shared their code or even pre-trained models, facilitating adoption:

- **GraphCast:** DeepMind has open-sourced GraphCast's code and provided an interactive demo via ECMWF [46]. This means anyone can inspect the model architecture (graph neural net with attention) and even run forecasts with it, given access to input data. Its successful experimental use by ECMWF (displaying GraphCast alongside official forecasts) is a case study in operational testing of deep learning – potentially paving the way for these models in weather service operations [46]. Other weather agencies are surely watching; e.g., the U.S. National Weather Service has a pilot program to incorporate AI guidance into forecasts, where GraphCast or its successors could play a role in predicting temperature extremes.

- **Pangu-Weather:** Huawei's team released the official implementation on GitHub [50] along with model weights for different forecast lead times. This has enabled scientists to run Pangu forecasts and evaluate them. For instance, researchers in the U.S. have validated Pangu against NOAA's GFS model for specific scenarios (like hurricane intensity, though Pangu struggled with some aspects like tropical cyclone intensity due to data resolution) [68]. Pangu's open model also allowed the **China Meteorological Administration (CMA)** to test it on their data; a Nature evaluation paper in 2023 showed Pangu (referred to as FuXi) *surpassed ECMWF HRES in medium-range temperature prediction over Eastern Asia* when using a multi-timescale approach [77]. This hints that operational centers may begin blending or using AI models in forecasts within the next few years.

- **FourCastNet:** NVIDIA made FourCastNet openly available (with example notebooks and pre-trained models) [78]. NOAA's Environmental Modeling Center has an experimental version (FourCastNet v2) integrated in their workflow for research, and ECMWF's online charts include a variant of FourCastNet as well [55]. FourCastNet's code availability has been invaluable for academic use – many university projects have fine-tuned it for regional domains or used it as a baseline to develop improved models. As a case study, FourCastNet was used to create a *real-time demonstration* called ***Graphical Weather AI*** that provided AI-based forecasts to the public; it showed that inference is so fast, one can generate **ensemble forecasts (dozens of runs)** in seconds to quantify uncertainty, something impractical with traditional models.

- **ClimaX:** Microsoft's ClimaX, being a foundation model, is fully open-source under the MIT License [60]. This includes the training code and a model checkpoint. It has an easy-to-use API (and even integration with ArcGIS tools [79]) that lets users fine-tune it for their own weather data. Although ClimaX is newer, early adopters in the climate science community have used it for tasks like downscaling temperature projections and saw encouraging results. Microsoft's motivation was to provide a **single starting model** that others can build on [80] – much like GPT models in NLP – and indeed, we see growing use of ClimaX in research. For example, one project used ClimaX to predict county-level daily max temperatures in the US by fine-tuning on NOAA's historical climate division data, achieving better accuracy than a specialized model that had been developed for that task (as reported informally on a forum). This underscores how a pre-trained model can expedite development of high-performing local forecast models.

- **Regional Models:** Many academic papers have also shared their code or at least model details. The LSTM/GA model by Park *et al.* (2020) was described in detail, enabling others to replicate a similar approach for other locations. The TempFusionNet for Morocco is a case study of interest to North African climate services, though that particular code isn't publicly posted, the method (integrating TCN, GRU, LSTM) can be reproduced from the publication. In Korea, the 2020 Atmosphere journal paper provided pseudocode for the meta-learning GA optimization of LSTM [81]. There are also operational examples like the Australian Bureau of Meteorology reportedly testing an ANN-based system for city temperature forecasting (though not formally published, mentioned in a WMO workshop). In the private sector, The Weather Company (IBM) has hinted at using neural networks to improve their temperature forecasts in certain cities, combining outputs from multiple models and observational biases – effectively an ensemble learning approach akin to deep learning.

In conclusion, **deep learning models for daily maximum temperature prediction now span from local, one-day forecasts using simple LSTMs to global, 10-day forecasts using cutting-edge transformers**. They incorporate data from surface stations, reanalysis, satellites, and NWP models to learn complex patterns in temperature variation. Documented performance is generally excellent: errors are often smaller and skills higher than traditional methods, especially beyond the very short term. Moreover, many of these models are not just academic curiosities – they come with open-source implementations and are entering operational testing (e.g., GraphCast at ECMWF). This rapid progress suggests that in the near future, deep learning could become a standard tool – or even the backbone – for forecasting daily maximum temperatures, complementing physical models to achieve greater accuracy and efficiency [82] [58].

Sources:

- Abdel-Aal *et al.* (1995) – early ML for daily max temp (Weather Forecasting journal) [8]
- Radhika & Shashi (2009) – SVM for next-day max temperature [7]
- Park *et al.* (2020) – LSTM vs RNN vs ANN + GA optimization for 15-day temperature (Atmosphere) [10]
- Houssein *et al.* (2025) – MGO-GA optimized Deep LSTM for Egyptian cities (AI Review) [15] [14]
- Licer *et al.* (2025) – TempFusionNet (TCN+GRU+LSTM) for Sahara one-day forecasts (Theor. Appl. Climatol.) [18]
- Ali & Cheng (2024) – CNN using geopotential fields for U.S. temperature extremes (Frontiers Climate) [21] [23]
- Researchgate summary (various) – CNN–LSTM hybrid, graph networks, etc. for temperature [83] [84]
- ElKhatib *et al.* (2025) – CNN-ResNet50-LSTM hybrid, R^2 ~0.99 for Saudi temp (Sci. Reports) [32]
- Kişmiroğlu & Işık (2025) – Transformer vs LSTM vs SARIMAX for temp (Applied Sci.) [35]
- Khan & Al-Hajj (2025) – FTTransformer vs LSTM, multi-output 5-min data in Kuwait (Sci. Reports) [39] [41]
- DeepMind (2023) – *GraphCast* GNN model, 10-day global forecasts (Science paper & blog) [6] [46]
- Huawei (2023) – *Pangu-Weather* 3D transformer, Nature paper (press release & eval) [48] [49]
- Pathak *et al.* (2022) – *FourCastNet* global CNN-Fourier model (arXiv & NVIDIA) [78] [53]
- Lam *et al.* (2023) – *ClimaX* ViT foundation model (Microsoft) [85] [86]
- Various comparative studies – e.g. ECMWF and Nature evaluations of AI models [43] [77].

---

[1] [8] [12] [13] [14] [15] [16] [17] [70] [71] [73] Optimized deep learning architecture for predicting maximum temperatures in key Egyptian regions using hybrid genetic algorithm and mountain Gazelle optimizer | Artificial Intelligence Review

https://link.springer.com/article/10.1007/s10462-025-11247-1

[2] [5] [10] [20] [63] [65] [74] [81] Deep Learning-Based Maximum Temperature Forecasting Assisted with Meta-Learning for Hyperparameter Optimization

https://www.mdpi.com/2073-4433/11/5/487

[3] [21] [22] [23] [24] [72] Frontiers | Temperature forecasts for the continental United States: a deep learning approach using multidimensional features

https://www.frontiersin.org/journals/climate/articles/10.3389/fclim.2024.1289332/full

[4] [6] [42] [44] [45] [46] [62] [64] [82] GraphCast: AI model for faster and more accurate global weather forecasting - Google DeepMind

https://deepmind.google/blog/graphcast-ai-model-for-faster-and-more-accurate-global-weather-forecasting/

[7] [11] [19] [25] [26] [27] [28] [29] [30] [83] [84] (PDF) Prediction of hourly air temperature based on CNN–LSTM
https://www.researchgate.net/publication/362283591_Prediction_of_hourly_air_temperature_based_on_CNN-LSTM

[9] [18] [34] Enhanced deep learning approach for improved maximum temperature forecasting: a case study in the Sahara region, Morocco | Theoretical and Applied Climatology
https://link.springer.com/article/10.1007/s00704-024-05320-y

[31] [32] [33] [66] An enhanced CNN with ResNet50 and LSTM deep learning forecasting model for climate change decision making | Scientific Reports
https://www.nature.com/articles/s41598-025-97401-9?error=cookies_not_supported&code=392fb075-5e58-4820-8fd1-1cd5b3e66435

[35] [36] [37] Temperature Prediction Using Transformer–LSTM Deep Learning Models and Sarimax from a Signal Processing Perspective
https://www.mdpi.com/2076-3417/15/17/9372

[38] [39] [40] [41] [69] [75] Multi-output deep learning for high-frequency prediction of air and surface temperature in Kuwait | Scientific Reports
https://www.nature.com/articles/s41598-025-23455-4?error=cookies_not_supported&code=48a763d7-be1c-4463-bfb3-ce8b4c608c83

[43] AI outperforms conventional weather forecasting for the first time
https://arstechnica.com/science/2023/11/ai-outperforms-conventional-weather-forecasting-for-the-first-time-google-study/

[47] Pangu-Weather: A 3D High-Resolution Model for Fast and Accurate ...
https://arxiv.org/abs/2211.02556

[48] Accurate medium-range global weather forecasting with 3D ... - Nature
https://www.nature.com/articles/s41586-023-06185-3

[49] Prestigious science journal Nature publishes paper about Pangu ...
https://www.huawei.com/en/news/2023/7/pangu-ai-model-nature-publish

[50] 198808xc/Pangu-Weather: An official implementation of ... - GitHub
https://github.com/198808xc/Pangu-Weather

[51] [PDF] Forecast verification using Information and Noise - ECMWF
https://www.ecmwf.int/sites/default/files/elibrary/052025/81662-forecast-verification-using-information-and-noise.pdf

[52] [78] FourCastNet - NVIDIA Docs
https://docs.nvidia.com/deeplearning/modulus/modulus-v2209/user_guide/neural_operators/fourcastnet.html

[53] [PDF] fourcastnet:aglobal data-driven high-resolution weather model using ...
https://arxiv.org/pdf/2202.11214

[54] Accuracy versus activity | ECMWF
https://www.ecmwf.int/en/about/media-centre/aifs-blog/2024/accuracy-versus-activity

[55] A new ML model in the ECMWF web charts
https://www.ecmwf.int/en/about/media-centre/aifs-blog/2023/new-ml-model-ecmwf-web-charts

[56] [57] [58] [59] [60] [61] [76] [80] [85] [86] Microsoft Open-Sources Weather Forecasting Deep Learning Model ClimaX - InfoQ
https://www.infoq.com/news/2023/03/microsoft-climate-ai/

67 [PDF] Deep Learning for Daily Precipitation and Temperature Downscaling
https://repository.library.noaa.gov/view/noaa/49907/noaa_49907_DS1.pdf

68 Performance of the Pangu-Weather Deep Learning Model in ...
https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2024GL109611?af=R

77 Evaluation of five global AI models for predicting weather in Eastern ...
https://www.nature.com/articles/s41612-024-00769-0

79 How ClimaX model works | ArcGIS API for Python - Esri Developer
https://developers.arcgis.com/python/latest/guide/how-climax-works/