Exploring Long-Term Temperature Forecasts

CNNs to Attention-Enhanced ConvLSTMs

Dorian Boucher
49004563

Johannes Gungaard
49004571

Philipp Peter
49004549

Thomas Renaud
49004578

Abstract

This study investigates the application of deep neural networks in weather forecasting, focusing on the $2m$ temperature over South Korea for the next 24 months. Traditional Numerical Weather Prediction (NWP) models, despite their operational effectiveness, are computationally expensive. Motivated by the success of neural networks in video frame prediction, we employ a range of models, including a simple convolutional neural network, a ConvLSTM (recurrent neural network with convolutional filters), and an attention-enriched ConvLSTM for monthly forecasts. Leveraging 8 decades of ERA5 reanalysis data, with 56 years for training, 16 years for validation, and 8 years for testing, we use various meteorological parameters as predictors. The models, encompassing geopotential pressure and wind components, demonstrate superior predictive capabilities compared to persistence forecasts. This highlights the potential of advanced deep neural networks for achieving substantial forecast quality beyond short-term nowcasting range, relying solely on a data-driven approach.

1 Introduction

Weather forecasting is crucial for various sectors, impacting agriculture, energy, tourism, and more. Temperature, a key climate indicator, influences global phenomena like climate change and local weather conditions, affecting ecosystems and communities (1). With an increase of approximately $1\,^{\circ}\mathrm{C}$ in global average temperature since the 19th century (2), our focus is on predicting temperature changes in smaller regions to address the specific needs of local communities.

Weather forecasting and video prediction through deep learning share similarities in analyzing spatio-temporal patterns (3). However, they differ as video prediction often involves distinct objects with clear separations, whereas weather data lacks such separability and involve complex multiscale interactions in pattern evolution.

While data-driven neural networks continue to advance in computer vision, there is a growing interest in physics-informed neural networks (PINNs) (4). PINNs aim to incorporate physical laws expressed in partial differential equations, offering a promising framework for atmospheric dynamics. However, PINNs have been limited to simplified versions of the Navier–Stokes equation and may encounter convergence and accuracy issues for processes on multiple spatio-temporal scales, such as those present in the real atmosphere. Due to these challenges, this paper focuses on data-driven neural networks (3), exploring their potential to enhance predictive skills in Earth science applications by capturing non-linear relations in the data.

2 Methods

2.1 Dataset

Meteorological data was obtained from the ERA5 Reanalysis dataset (5) in a grib-format, provided by Copernicus Climate Change Service (C3S). It consists of hourly estimates of atmospheric, land and oceanic climate variables, ranging from 1st of January 1940 to present.
For each month, data variables from 5 evenly separated dates (1st, 7th, 13th, 19th, and 25th) at 00:00 and 12:00 were collected over South Korea, between latitude and longitude coordinates ranging [34°N, 38°N] and [125°W, 130°E], with a coordinate resolution of 0.25 degrees. Making it a total of 357 observation points per variable measurement.

2.2 Data Handling and Data Reduction Techniques

Based on the Exhaustive Feature Search (EFS) performed by Fister et al. (6), the 8 features listed in table 1 were selected. The paper propose a 9th feature, mean sea level pressure, however as this feature is limited to areas of water, it's desirable to remove for our geographical case.

To create a supervised time series dataset, we performed preprocessing by computing monthly averages, generating a single variable with 8 features for each coordinate in the range. The dataset was then split into training (70%), validation (20%), and testing (10%) sets. Sequential batches covering a 6-year timeframe were used as input, predicting the following 2 years. This process involved a stepwise moving window forward in time, resulting in a 5-dimensional dataset, as depicted in Figure 1.

Table 1: Meteorological variables used in the study.

No.	Variable	Unit
1.	10 m u-component of wind	$\frac{m}{s}$
2.	10 m v-component of wind	$\frac{m}{s}$
3.	100 m u-component of wind	$\frac{m}{s}$
4.	100 m v-component of wind	$\frac{m}{s}$
5.	mean sea level pressure	Pa
6.	volumetric soil water layer 1	$\frac{m^3}{m^3}$
7.	geopotential pressure level on 500 hPa	$\frac{m^2}{s^2}$
8.	air temperature (at 2m)	K

2.3 Model Architectures

In the following, we briefly introduce the three video prediction models probed in this study: a simple fully convolutional neural network (CNN), the convolutional LSTM (ConvLSTM) model, and the convolutional LSTM enriched with attention (ConvLSTM + Attention).

2.3.1 Convolutional Neural Network

Following up the study by Rasp et al. (7), we deploy a simple CNN as one of the baseline models. The CNN consists of one layers with 32 channels with a kernel size of 3 followed by 2 dense layers. The model was trained using Mean Squared Error as the loss function. It was trained for 20 epochs using the AdamW optimizer (8) with a learning rate of $lr = 0.001$ and batch size of 32.

2.3.2 ConvLSTM

The ConvLSTM model combines convolutional operations for spatial features and a gated LSTM for temporal coherence. It encodes the atmospheric state over the previous 6 years, and the forecasting network unfolds this state to generate a 2-year forecast using LSTM cell states and sequential predictions. We used a batch size of 32 and the model converged after 25 epochs. The model was trained using Mean Squared Error as the loss function and AdamW optimizer (8) with a learning rate of $lr = 0.001$.

2.3.3 ConvLSTM with Attention

The addition of the ProbSparse Attention Layer (9), used in the paper of Zhou et al. (10) to our ConvLSTM model is added in between the LSTM layer and the dense layers to enhances its spatial-temporal data analysis. Specifically designed to focus on crucial temporal features in complex

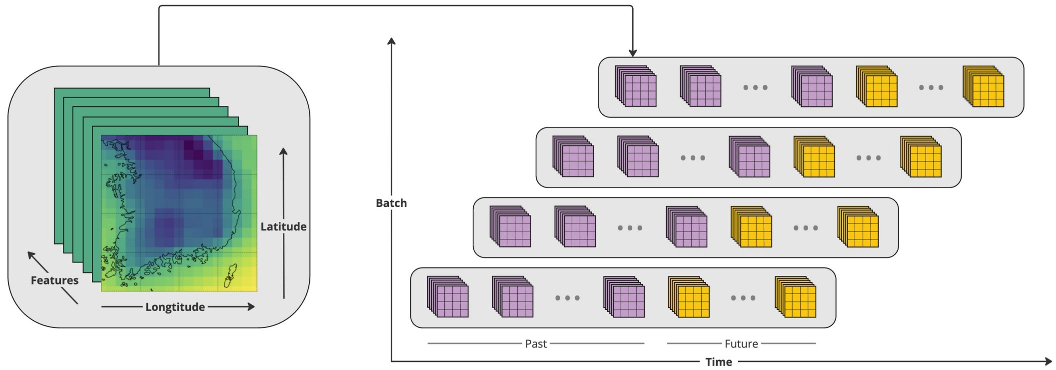

Figure 1: Left: a single variable showing a map of South Korea consisting of longitude, latitude and features. Right: The variables preprocessed in batches of past and future, moving each batch step-wise along in time.

datasets, this layer, with a sampling factor of 5, selectively processes key data points. By normalizing attention scores through the softmax function, it dynamically adjusts feature weighting, improving predictive accuracy. This integration addresses challenges in handling large spatial-temporal data, ensuring a more focused and accurate forecasting capability in the ConvLSTM model. Simarly to the CNN and the ConvLSTM model, the ConvLSTM+Attention has converged after 35 epochs, using the optimizer AdamW with a learning rate of $lr = 0.001$.

3 Results and Analysis

In the following, we evaluate the predictive skill of the Conv, ConvLSTM and ConvLSTM + Attention models for $2m$ temperature predictions up to a lead time of 2 years.

The performance metrics for different models are presented in Table 2. The Convolutional Neural Network (CNN) with Long Short-Term Memory (LSTM) demonstrates superior performance in generating long-term temperature forecasts. Its inherent ability to capture and retain information over extended time sequences is evident in the achieved metrics.

Table 2: Performance Metrics for Different Models

Model	MAE	MAPE	RMSE
Conv	0.2383	1.5355	0.2948
ConvLSTM	**0.2003**	**1.4153**	**0.2492**
ConvLSTM + Attention	0.3281	1.8888	0.2668

While the Conv LSTM outperforms the basic convolutional model in terms of MAE, MAPE, and RMSE, the introduction of an attention mechanism in the Conv LSTM leads to a setback in performance. The attention mechanism's underperformance suggests a delicate balance is needed when incorporating sophisticated components into a model.

4 Conclusion

In this study, we explored the utilization of Convolutional Long Short-Term Memory (ConvLSTM) models in weather forecasting, revealing a reduction in inference time compared to traditional Numerical Weather Prediction (NWP) methods. This efficiency not only enhances accessibility but also delivers precise forecasts for practical applications. The ConvLSTM model proves to be an efficient and effective approach for understanding and predicting climatic trends over extended periods, showcasing its potential for widespread application and decision support.

References

[1] DeWeber, J. T. ; Wagner, T.: Probabilistic measures of climate change vulnerability, adaptation action benefits, and related uncertainty from maximum temperature metric selection. In: *Global Change Biology* 24 (2018), S. 2735–2748. http://dx.doi.org/10.1111/gcb.14101. – DOI 10.1111/gcb.14101

[2] Lindsey, R. ; Dahlman, L.: *Climate Change: Global Temperature | NOAA Climate.gov*. http://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature

[3] Gong, B. ; Langguth, M. ; Ji, Y. ; Mozaffari, A. ; Stadtler, S. ; Mache, K. ; Schultz, M. G.: Temperature forecasting by deep learning methods. In: *Geoscientific Model Development* 15 (2022), Nr. 23, 8931–8956. http://dx.doi.org/10.5194/gmd-15-8931-2022. – DOI 10.5194/gmd–15–8931–2022

[4] Karniadakis, George E. ; Kevrekidis, Ioannis G. ; Lu, Lu ; Perdikaris, Paris ; Wang, Sifan ; Yang, Liu: Physics-informed machine learning. In: *Nature Reviews Physics* 3 (2021), Jun, Nr. 6, 422-440. http://dx.doi.org/10.1038/s42254-021-00314-5. – DOI 10.1038/s42254–021–00314–5. – ISSN 2522–5820

[5] *ECMWF Reanalysis v5 (ERA5)*. https://www.ecmwf.int/, . – Accessed on 12-08-2023

[6] Fister, D. ; Pérez-Aracil, J. ; Peláez-Rodríguez, C. ; Del Ser, J. ; Salcedo-Sanz, S.: Accurate long-term air temperature prediction with Machine Learning models and data reduction techniques. In: *Applied Soft Computing* 136 (2023), 110118. http://dx.doi.org/https://doi.org/10.1016/j.asoc.2023.110118. – DOI https://doi.org/10.1016/j.asoc.2023.110118. – ISSN 1568–4946

[7] Rasp, Stephan ; Dueben, Peter D. ; Scher, Sebastian ; Weyn, Jonathan A. ; Mouatadid, Soukayna ; Thuerey, Nils: WeatherBench: A Benchmark Data Set for Data-Driven Weather Forecasting. In: *Journal of Advances in Modeling Earth Systems* 12 (2020), Nr. 11, e2020MS002203. http://dx.doi.org/https://doi.org/10.1029/2020MS002203. – DOI https://doi.org/10.1029/2020MS002203. – e2020MS002203 10.1029/2020MS002203

[8] Loshchilov, Ilya ; Hutter, Frank: *Decoupled Weight Decay Regularization*. 2019

[9] Wang, Xiong ; Sun, Sining ; Xie, Lei ; Ma, Long: *Efficient Conformer with Prob-Sparse Attention Mechanism for End-to-EndSpeech Recognition*. 2021

[10] Zhou, Haoyi ; Zhang, Shanghang ; Peng, Jieqi ; Zhang, Shuai ; Li, Jianxin ; Xiong, Hui ; Zhang, Wancai: *Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting*. 2021