The Next Wave of Meteorology: Evaluating the Impact of Deep Learning Models on Climate Forecast Accuracy

1. Project Summary:

The ever-changing and dynamic nature of weather and climate conditions adds significant variability in the planning and execution of both short and long-term activities in numerous domains, including agriculture, travel, transportation, outdoor robot deployment, and construction. Consequently, being able to predict weather and climate parameters plays a vital role reliably and accurately in scheduling, strategy, and preparation for the previously mentioned activities. While there exist several models to forecast weather conditions, the long-term, seasonal, and short-term trends, and fluctuations make it challenging to obtain reliable and accurate predictions. This work aims to forecast several weather features such as air pressure, humidity, precipitation, and CO_2 concentration.

2. What will we do?

Weather forecasting can be formulated as a time-series problem since the current and future outcomes for variables depend on past variable outcomes. With the advancement There currently exist numerous architectures such as RNNs, LSTMs, and transformers that deal with such time series data and are effective at forecasting. During this project, we aim to explore several such models and architectures used for weather forecasting. We also aim to study the effects of different hyperparameters and loss functions for each model. The models will aim to utilize all the parameters as the unput to predict the outputs.

We will:

(i) Predict:

- Air pressure
- Temperature
- Relative humidity
- Air density
- CO_2 concentration
- Precipitation

(ii) Explore several currently available models for weather forecasting (this includes SOTA architectures)

(iii) Analyze the effects of different loss functions.

As a final contribution, we aim to propose a novel architecture that captures the spatial and temporal dependencies of the data to complete the task of weather prediction reliably and accurately. We also hope to use the intuition developed through performing the model explored before to explain why the proposed model works the way it does.

4. Resources and Related Work

There is existing work performed on using Transformers.

SOTA: TENT: Tensorized Encoder Transformer for temperature forecasting incorporates tensor-based attention mechanisms, enabling it to leverage the spatiotemporal characteristics of meteorological data by processing it as multidimensional tensors.[1]

Climate Forecasting: Long Short-Term Memory Model using Global Temperature Data utilizes LSTM that is well-adapted for making predictions using time-series data. [2]

Temporal convolutional neural (TCN) network for an effective weather forecasting using time: This model serves as an effective tool for community-based localized weather forecasting and can operate independently on a personal computer.[3]

5. Dataset

The dataset being utilized for our task is the **weather time series dataset [4]** from the Max Planck Institute for Biogeochemistry.

This dataset comprises 14 distinct features including air temperature, atmospheric pressure, and humidity.

The data is recorded at a frequency of every 10 minutes, starting from the year 2003.

Weather Time Series Dataset Overview:

Symbol	**Unit**	**Variable**
Date Time	DD.MM.YYYY HH:MM (MEZ)	Date and time of the data record (end)
p	mbar	Air pressure
T	oC	Air temperature
Tpot	K	Potential temperature
Tdew	oC	Dew point temperature
rh	%	Relative humidity
VPmax	mbar	Saturation water vapor pressure
VPact	mbar	Actual water vapor pressure
VPdef	mbar	Water vapor pressure deficit
sh	g kg−1	Specific humidity
H2OC	mmol mol−1	Water vapor concentration
rho	g m−3	Air density
wv	m s−1	Wind velocity
Max. wv	m s−1	Maximum wind velocity
wd	o	Wind direction
rain	mm	Precipitation
raining	s	Duration of precipitation
SWDR	Wm−2	Shortwave downward radiation
PAR	µmol m−2 s−1	Photosynthetically active radiation
Max. PAR	µmol m−2 s−1	Maximum photosynthetically active radiation
Tlog	oC	Internal logger temperature
CO2	ppm	CO2-concentration of ambient air

6. Ethical Considerations

In our project to predict weather information through Machine Learning, we are committed to upholding the highest standards of ethical practice. The dataset we intend to use is sourced from the publicly accessible repository provided by the Max Planck Institute for Biogeochemistry. We will ensure that our use of this data is in strict compliance with any terms of service, usage guidelines, and ethical standards set forth by the data provider.

Furthermore, we acknowledge the importance of transparency and reproducibility in scientific research. As such, we will provide clear documentation of our methodologies and findings, allowing for peer review and constructive critique. By adhering to these ethical considerations, we aim to contribute positively to the field of meteorology and the broader scientific community, while respecting the integrity of the data and the rights of the data providers.

7. Group Members

Shashank Garikipati

Vaibhav Malhotra

Pranav Sharma

Koushik Karan Geetha Nagaraj

8. References

[1] Onur Bilgin, Paweł Mąka, Thomas Vergutz, Siamak Mehrkanoon,” TENT: Tensorized Encoder Transformer for Temperature Forecasting” ,https://doi.org/10.48550/arXiv.2202.11214, Feb 2022.

[2] P. Akhila, R. L. S. Anjana and M. Kavitha, "Climate Forecasting:Long short Term Memory Model using Global Temperature Data," 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2022, pp. 469-473, doi: 10.1109/ICCMC53470.2022.9753779.

[3] Hewage, P., Behera, A., Trovati, M. et al. Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station. Soft Comput 24, 16453–16482 (2020). https://doi.org/10.1007/s00500-020-04954-0.

[4] https://www.bgc-jena.mpg.de/wetter/