Project Report: Predicting Precipitation Using DA-RNN on ERA5 Dataset

1. Introduction

Accurate precipitation prediction is crucial for managing flood risks, optimizing water resources, and improving weather-responsive planning in agriculture and infrastructure. Traditional forecasting techniques, such as statistical models or numerical weather prediction systems, often struggle with capturing the nonlinear temporal dependencies and spatial variability inherent in precipitation data.

In this project, we explore the application of the Dual-Stage Attention-Based Recurrent Neural Network (DA-RNN) [1] to predict precipitation levels using the ERA5 dataset [2]. The DA-RNN architecture is specifically designed to address two major limitations of classical time series models: (i) their inability to select the most relevant input features, and (ii) their weakness in capturing long-range temporal dependencies, which are critical for modeling complex weather patterns.

The DA-RNN model has already demonstrated state-of-the-art performance on financial (NASDAQ 100 Stock dataset) and environmental datasets (SML 2010 dataset) by capturing long-term dependencies and filtering relevant features through input and temporal attention mechanisms. Thus, this makes DA-RNN suitable for multi-variate, exogenous time series data such as ERA5, where the goal is to predict a target variable (precipitation) based on multiple atmospheric drivers.

2. Literature Review

Traditional time series forecasting models such as *ARMA* [3] and *ARIMA* [4] have been widely used for weather prediction. However, these models assume linear relationships and struggle with multi-input, nonlinear systems. To address this, *Nonlinear Autoregressive Exogenous (NARX)* models were introduced [5], allowing predictions based on both past values of the target and multiple exogenous variables. While effective in many domains, NARX models often lack the capacity to handle long-range dependencies and decide which input features are the most important.

Neural networks have been successfully used to implement NARX architectures, allowing for better handling of nonlinearities in the data. Recurrent Neural Networks (RNNs), including their variants like *LSTM* [6] and *GRU* [7], were later adopted for time series forecasting due to their ability to retain memory across sequences. Despite their success, standard RNNs and even LSTM-based encoder-decoder models suffer performance degradation when faced with long input sequences, a scenario common in climate modeling.

To address the sequence length limitation, attention mechanisms have been introduced in encoder-decoder frameworks [8], enabling the model to selectively focus on relevant portions of the input. However, these models primarily focus on temporal attention and often ignore the relevance of individual input variables (features).

In recent years, deep learning has been applied to precipitation forecasting with encouraging results, However, few models explicitly address both *feature selection and temporal dependency modeling*, which are vital for precipitation due to the multivariate and dynamic nature of atmospheric processes.

To solve the dual challenge of feature relevance and long-term temporal dependency, Qin et al. (2017) proposed the ***DA-RNN*** model [1]. This architecture integrates two attention stages: *Input Attention:* Selects relevant exogenous (driving) variables at each timestep by conditioning on the encoder's hidden state. This allows the model to dynamically filter out irrelevant features. *Temporal Attention:* Selects influential encoder hidden states over time, helping the decoder focus on the most critical moments in the sequence.

This design allows DA-RNN to outperform classical NARX models, vanilla LSTM models, and attention-only models on complex, real-world datasets.

3. Methodology

3.1 Dataset gathering:

To construct the dataset for precipitation forecasting, I accessed hourly meteorological data from the Copernicus Climate Data Store (CDS). Since ERA5 data is hosted remotely, each monthly or bi-monthly chunk of data had to be retrieved via individual API requests. I incrementally gathered and merged data spanning from January 2023 to December 2024, resulting in a complete, timely dataset.

The ERA5 dataset offers hourly gridded estimates of numerous atmospheric variables across a global spatial grid (latitude × longitude × variables × time). For this project, the geographical location was fixed to Munich, Germany during preprocessing to reduce the spatial dimension. This allowed the transformation of the data into a multivariate time series, structurally aligned with how the datasets were used in the original DA-RNN paper.

3.2 Dataset preparation:

In addition to the target variable total precipitation (tp), the following 10 driving (exogenous) variables were selected from ERA5, as they are commonly associated with precipitation dynamics: e (evaporation), es (surface evaporation stress), avg_rorwe (runoff from rainfall over land), avg_esrwe (soil evaporation from surface), u10 (10m u-component of wind), v10 (10m v-component of wind), t2m (2m temperature), msl (mean sea-level pressure), sp (surface pressure), tcc (total cloud cover).

To enhance the temporal representation of the time series, I also incorporated cyclical time features for both the time of day and the day of the year. These features were encoded using sine and cosine functions to capture the periodic nature of daily and seasonal cycles:

$$\text{hour\_sin} = \sin\left(\frac{2\pi \cdot \text{hour}}{24}\right), \quad \text{hour\_cos} = \cos\left(\frac{2\pi \cdot \text{hour}}{24}\right)$$

$$\text{doy\_sin} = \sin\left(\frac{2\pi \cdot \text{day\_of\_year}}{365}\right), \quad \text{doy\_cos} = \cos\left(\frac{2\pi \cdot \text{day\_of\_year}}{365}\right)$$

3.3 Models:
To evaluate the effectiveness of different models for precipitation forecasting, three architectures were implemented and compared: DA-RNN, LSTM, and a baseline Linear Regression model. All models were trained using a consistent configuration:

B*atch size: 32, Epochs: 100, Optimizer: Adam, Learning rate: 0.001 , Sliding window size (T): 10 hours* (Each sample consisted of 10 hours of meteorological data, used to predict the next-hour total precipitation)

DARNN - The Dual-Stage Attention-based Recurrent Neural Network (DA-RNN) closely follows the architecture proposed by Qin et al. (2017). A key aspect of DA-RNN is its use of both the exogenous input sequence and the historical target series *(y_history)* as inputs to the decoder, enabling it to model both feature interactions and temporal dependencies more effectively.

LSTM - The LSTM model is a single-layer recurrent neural network with 64 hidden units. It processes the input multivariate time series over a fixed window of 10 hours and outputs a single-step precipitation forecast. Only the final hidden state from the LSTM is passed through a fully connected layer to produce the prediction. Unlike DA-RNN, it does not explicitly incorporate a separate attention mechanism or a second input sequence of historical target values.

Linear Regression - For the linear regression baseline, the multivariate time series input across 10 hours was flattened into a single vector per instance. This converts the input into a tabular form, allowing the model to learn a direct linear mapping from past feature values to the next precipitation value.

3.4 Evaluation:
To assess model performance, we used two standard error metrics: *Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE)*. These were computed on the test set for all models to ensure a fair comparison.
Additionally, both LSTM and DA-RNN were used for recursive multi-step forecasting, predicting precipitation hourly over a **24-hour future window.**

4. Results

The hourly precipitation data from January 2023 to December 2024 was used to train and evaluate all models. This dataset was split into 14,035 samples for training, 1,754 for validation, and 1,754 for testing.

The results on the Testing set can be seen below:

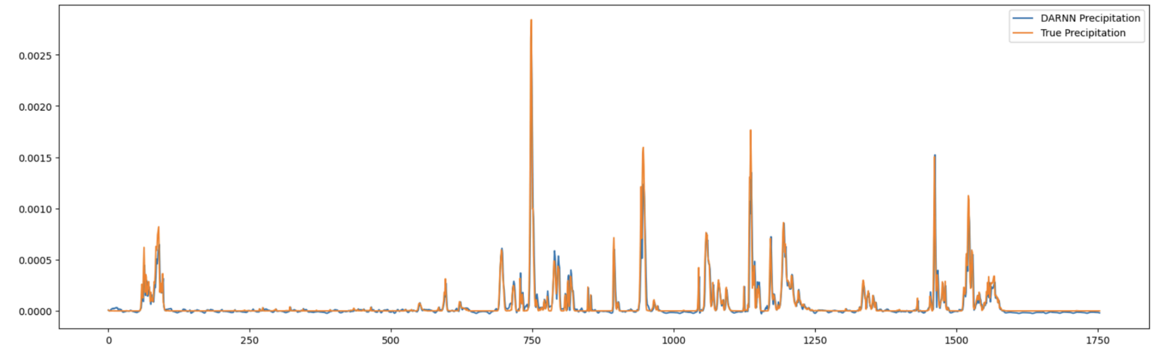

Figure 1: DARNN's prediction on the testing data

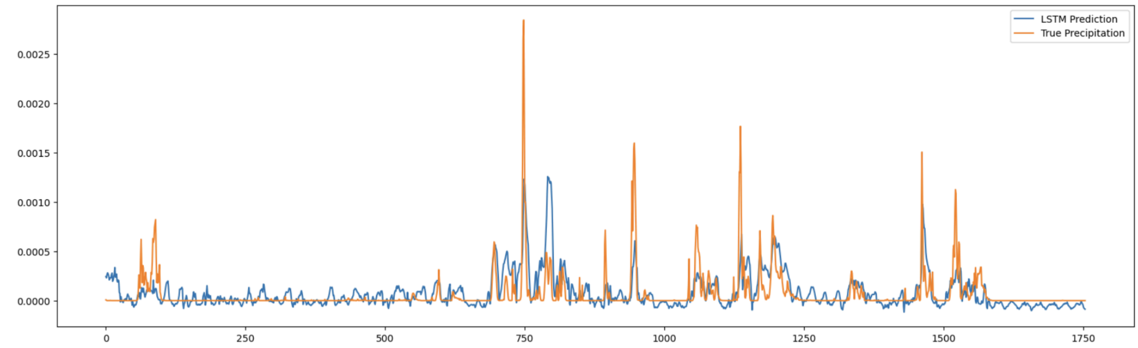

Figure 2: LSTM's prediction on the testing data

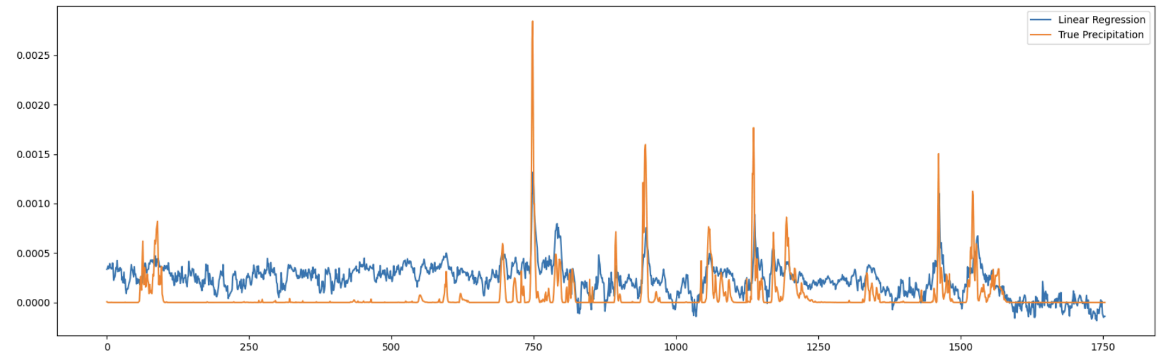

Figure 3: LR's prediction on the testing data

Both DA-RNN and LSTM models forecasted zero precipitation for the next 24 hours, which aligns accurately with the ERA5 dataset, where precipitation on January 1st, 2025, remained zero across all hourly intervals.

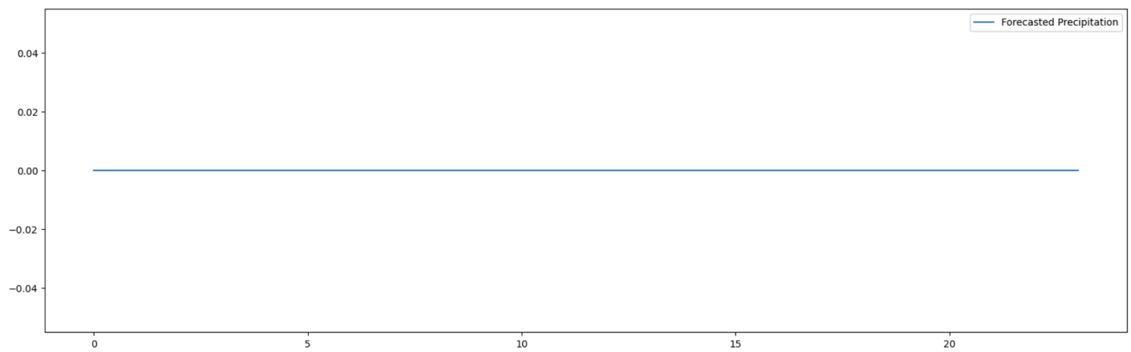

Figure 4: Forecasted precipitation for next 24 hours by DARNN & LSTM

Comparison of the models using *Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE)* as evaluation metrics can be seen below:

Model	MAE	RMSE
Linear Regression	0.000208	0.000248
LSTM	0.000097	0.000179
DA-RNN	**0.000033**	**0.000082**

The **DA-RNN model outperformed** both the LSTM and Linear Regression models, achieving the lowest MAE and RMSE values, indicating superior accuracy in precipitation prediction. The LSTM model also demonstrated strong performance, notably better than Linear Regression, which struggled to capture the nonlinear dependencies in the data.

Additionaly:

- Github repo: https://github.com/laibaq18/Time-series-prediction
- Link to notebook: Percipitation_ERA5.ipynb

References

[1] Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., & Cottrell, G. (2017). A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction. arXiv preprint arXiv:1704.02971. https://arxiv.org/pdf/1704.02971

[2] ERA5 Climate Reanalysis Dataset. Copernicus Climate Data Store. https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=overview

[3] M. Xu, J. Zhang and Z. Li, "A novel Lasso-ARMA model for time series prediction," 2018 Chinese Automation Congress (CAC), Xi'an, China, 2018, pp. 1698-1703, doi: 10.1109/CAC.2018.8623506.

[4] Gulnara, Bektemyssova & Ahmad, Abdul & Mirzakulova, Sharafat & Ибраева, Жанар. (2022). TIME SERIES FORECASTING BY THE ARIMA METHOD. Scientific Journal of Astana IT University. 14-23. 10.37943/HFCH4395.

[5] Yan, Linjun & Elgamal, Ahmed-Waeil & Cottrell, Garrison. (2013). Substructure Vibration NARX Neural Network Approach for Statistical Damage Inference. Journal of Engineering Mechanics. 139. 737-747. 10.1061/(ASCE)EM.1943-7889.0000363.

[6] Elsworth, Steven, and Stefan Güttel. "Time Series Forecasting Using LSTM Networks: A Symbolic Approach." arXiv preprint arXiv:2003.05672 (2020).

[7] Arokodare, Oluwatomisin. (2024). Leveraging Machine Learning for Enhanced Predictive Accuracy in Time Series Forecasting: A Comparative Analysis of LSTM and GRU Models. 10.13140/RG.2.2.30014.80965.

[8] X. Wen and W. Li, "Time Series Prediction Based on LSTM-Attention-LSTM Model," in IEEE Access, vol. 11, pp. 48322-48331, 2023, doi: 10.1109/ACCESS.2023.3276628.