PHYSICALLY INFORMED DEEP LEARNING FOR PREDICTING TROPICAL CYCLONE RISK IN A WARMING CLIMATE

Hyunjeong Kim
Georgia Institute of Technology
Atlanta, GA, United States of America
hkim3257@gatech.edu

ABSTRACT

Predicting extreme tropical cyclone (TC) seasons remains challenging due to sparse data and the tendency of models to regress toward climatological averages. We propose a physics-informed CNN–Transformer trained on seven dynamically relevant ERA5 predictors, using a novel Hybrid Peak Loss that scales penalties with event magnitude to prioritize hyper-active years. Results on 45 years of data show a Mean Absolute Error (MAE) of 0.94 major hurricanes per year. Our framework demonstrates improved prediction of high-impact seasons, providing a robust baseline for climate risk assessment.

1 INTRODUCTION

Tropical cyclones (TCs) are among the most destructive geophysical phenomena, with their frequency and intensity modulated by complex interactions between local thermodynamics and large-scale climate oscillations (Emanuel, 1987; 2003). As the climate warms, evidence suggests a global increase in the probability of major TCs (Kossin et al., 2020; Wehner & Kossin, 2024). However, seasonal forecasting remains difficult because standard machine learning objectives, such as Mean Squared Error (MSE), treat all years equally. This leads to a systematic under-prediction of "hyper-active" seasons as models opt for the safer climatological mean to minimize global error.

In this work, we present a deep learning framework that combines spatial feature extraction via Convolutional Neural Networks (CNNs) with the long-range temporal reasoning of Transformers. Our primary contribution is the **Hybrid Peak Loss**, a cost-sensitive objective function designed to penalize under-predictions of extreme seasons more heavily than errors in average years, mitigating the "regression-to-the-mean" effect.

2 METHODOLOGY

2.1 DATA AND PHYSICS-INFORMED PREDICTORS

We utilize monthly ERA5 reanalysis data (Hersbach et al., 2020) spanning 1980–2024. Based on the genesis potential indices established in previous literature (Emanuel, 2000; Camargo & Barnston, 2009), we select seven dynamically relevant predictors: Potential Intensity (PI), Saturation Deficit, Specific Humidity, and U/V wind components at 850hPa and 200hPa to resolve Vertical Wind Shear.

Targets are derived from the International Best Track Archive for Climate Stewardship (IBTrACS) (Knapp et al., 2010). We aggregate counts into three categories: Total Frequency, Major Hurricanes (>96 kts), and Landfall Count. To address the class imbalance inherent in extreme event datasets, we apply weighted random sampling to oversample the top 5% of active years, a technique analogous to SMOTE in regression contexts (Chawla et al., 2002).

2.2 CNN-TRANSFORMER ARCHITECTURE

The model utilizes a hybrid architecture to process the spatio-temporal nature of climate data.

1. **Spatial Encoder:** A Time-Distributed CNN extracts spatial features from the ERA5 patches. Unlike standard Vision Transformers (ViT) which may struggle with small datasets, the CNN provides an inductive bias for local meteorological patterns (Hilburn, 2023).
2. **Temporal Backbone:** A 6-layer Transformer Encoder with sinusoidal positional encodings (Irani & Metsis, 2025) processes the 12-month sequence. The 6-layer depth was chosen because deeper models suffered optimization instability, while shallower models underfit long-range teleconnections and exhibited higher MAE on Major Hurricanes.

To prevent overfitting given the limited sample size ($N \approx 540$ months), we implement Elastic Net regularization ($\lambda_1 = 10^{-5}, \lambda_2 = 10^{-4}$) on all kernels (Zou & Hastie, 2005).

2.3 Hybrid Peak Loss for Extreme Events

Standard regression objectives, such as Mean Squared Error (MSE), often result in "conservative" predictions that gravitate toward the climatological mean. To mitigate this, we propose a Hybrid Peak Loss ($\mathcal{L}_{HPL}$) that incorporates cost-sensitive learning to prioritize hyper-active seasons. The loss is defined as a convex combination of Log-Cosh and a weighted MSE:

$$\mathcal{L}_{HPL} = \alpha \cdot \log(\cosh(\hat{y} - y)) + (1 - \alpha) \cdot [(\hat{y} - y)^2 \cdot (1 + \lambda y)] \tag{1}$$

where y is the observed storm count and $\hat{y}$ is the prediction. The term $(1 + \lambda y)$ acts as a magnitude-dependent penalty, where the hyperparameter λ (set to 5.0) scales the error based on the intensity of the season, and $\alpha = 0.7$ controls the balance between Log-Cosh smoothness and peak sensitivity. The hyperparameters $\alpha = 0.7$ and $\lambda = 5.0$ were chosen based on a small grid search over plausible λ and α values on the validation set. This ensures that under-predicting a major hurricane season is penalized more heavily than errors in inactive periods, effectively forcing the model to capture the tail of the TC distribution.

3 Experiments and Results

3.1 Experimental Setup

The dataset spans 1980–2024 ($N \approx 540$ monthly samples). We trained the proposed CNN–Transformer on the 1980–2015 period and evaluated performance on an unseen test set (2016–2024). This preserves temporal ordering and prevents data leakage.

3.2 Quantitative Performance and Baseline Comparison

Both models are trained with early stopping (patience=30) and per-model learning rate reduction (ReduceLROnPlateau, factor=0.5, patience=10), with fresh callback instances per run to avoid state carryover between the Hybrid and MSE training runs.

To provide contextualization, we compare our model against two baselines: (1) a **Statistical Climatology** baseline, representing the historical monthly mean of the training period, and (2) a **Baseline MSE** version of the same architecture.

As shown in Table 1, the Proposed Hybrid model achieves a Major Hurricane MAE of 0.940, outperforming the Statistical Climatology baseline (MAE = 1.025) and the MSE ablation (MAE = 0.963).

Table 1: Comparative Performance (Major Hurricanes MAE)

Method	**Architecture/Loss**	**MAE**
Statistical Climatology	Historical Mean	1.025
Baseline MSE	CNN–Transformer (MSE)	0.963
Proposed Hybrid	**CNN–Transformer (Hybrid)**	**0.940**

3.3 Multi-Target Ablation Study

We further evaluated the impact of the Hybrid Peak Loss across three TC metrics. Table 2 provides a comparative breakdown of the Mean Absolute Error (MAE) for the Hybrid and MSE configurations.

Table 2: Test Set MAE Performance (2016–2024)

Target Variable	**Proposed Hybrid**	**Baseline MSE**
Total Frequency	2.329	2.211
Major Hurricanes	**0.940**	0.963
Landfall Count	1.405	1.290

3.4 Qualitative Analysis

Regularization was critical to the framework's stability; without the implemented Elastic Net constraints, the Transformer backbone exhibited significant overfitting, resulting in negative R^2 values. Furthermore, the Hybrid Peak Loss demonstrated improved predictive capability for peak seasons, such as 2018 and 2020. By scaling penalties with event magnitude, the model successfully resisted the tendency to regress toward the climatological mean, better capturing high-activity periods compared to the standard MSE baseline. As expected from a cost-sensitive objective, this gain on Major Hurricanes involves a deliberate trade-off: the MSE variant achieves lower MAE on Total Frequency (2.211 vs. 2.329) and Landfall Count (1.290 vs. 1.405), where the uniform loss encourages predictions closer to the mean. This behaviour is consistent with the design intent of $\mathcal{L}_{HPL}$ — prioritizing rare, high-impact seasons over average-year accuracy — and reflects the asymmetric societal cost of under-predicting major hurricane activity (Mestre & Hallegatte, 2009).

4 Discussion

Permutation feature importance identifies Potential Intensity (PI) and Vertical Wind Shear as the dominant predictors, consistent with established thermodynamic theory (Emanuel, 1987). The Hybrid Peak Loss reduces regression-to-the-mean during hyper-active seasons by scaling penalties with event magnitude, allowing the model to capture tail outcomes that standard MSE objectives tend to suppress.

5 Conclusion

The results confirm that the CNN-Transformer successfully integrates multi-scale climate drivers. The Hybrid Peak Loss addresses the "conservative" nature of traditional ML regressors by explicitly weighting high-magnitude events. This is particularly vital for Major Hurricanes, where the societal cost of under-prediction far outweighs that of over-prediction (Mestre & Hallegatte, 2009).

Future work will involve benchmarking our results against AI-generated TC tracks (Weng & Gori, 2025) and downscaling CMIP6 projections (Emanuel, 2021) to assess future shifts in extreme seasons. In conclusion, our findings suggest that physics-informed predictors combined with magnitude-sensitive loss functions offer a promising path for reliable extreme event forecasting in a warming climate.

References

S J Camargo and A G Barnston. Experimental seasonal dynamical forecasts of tropical cyclone activity. *Weather and Forecasting*, 24(2):472–491, 2009.

N V Chawla et al. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.

K Emanuel. Tropical cyclones. *Annual Review of Earth and Planetary Sciences*, 31:75–104, 2003.

Kerry A Emanuel. The dependence of hurricane intensity on climate. *Nature*, 326(6112):483–485, 1987.

Kerry A Emanuel. A statistical analysis of tropical cyclone intensity. *Monthly Weather Review*, 128 (4):1139–1152, 2000.

Kerry A Emanuel. Response of global tropical cyclone activity to increasing CO2: Results from downscaling CMIP6 models. *Journal of Climate*, 34(1):57–70, 2021. doi: 10.1175/JCLI-D-20-0367.1.

H Hersbach et al. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.

Kyle A Hilburn. Understanding spatial context in convolutional neural networks using explainable methods: Application to interpretable GREMLIN. *Artificial Intelligence for the Earth Systems*, 2023.

Habib Irani and Vangelis Metsis. Positional encoding in Transformer-based time series models: A survey. *arXiv preprint arXiv:2502.12370*, 2025.

K R Knapp et al. The international best track archive for climate stewardship (IBTrACS). *Bulletin of the American Meteorological Society*, 91(3):363–376, 2010.

James P Kossin, Kenneth R Knapp, Timothy L Olander, and Christopher S Velden. Global increase in major tropical cyclone exceedance probability over the past four decades. *Proceedings of the National Academy of Sciences*, 117(22):11975–11980, 2020.

O Mestre and S Hallegatte. Predictors of tropical cyclone numbers and extreme hurricane activities in a warming climate. *Earth's Future*, 3(1):52–66, 2009.

Michael F Wehner and James P Kossin. The growing inadequacy of an open-ended saffir–simpson hurricane wind scale in a warming world. *Proceedings of the National Academy of Sciences*, 121 (7):e2308901121, 2024.

Yanmo Weng and Avantika Gori. Climatological benchmarking of AI-generated tropical cyclones. *arXiv preprint arXiv:2511.21792*, 2025.

H Zou and T Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B*, 67(2):301–320, 2005.

A APPENDIX

A.1 ERA5 PREDICTOR SELECTION

The input features were extracted from the ERA5 global reanalysis dataset (Hersbach et al., 2020). Each monthly sample consists of a 7-channel atmospheric tensor with a spatial grid size of 60×120 pixels, corresponding to a $1.0^\circ \times 1.0^\circ$ resolution covering the primary tropical cyclogenesis basins. Following the physical drivers identified in Emanuel (2003), the variables selected represent the thermodynamic and dynamic state of the atmosphere (Table 3).

Table 3: ERA5 Input Features and Physical Relevance

Variable	**Physical Significance**
Potential Intensity (PI)	Maximum achievable TC wind speed
Saturation Deficit	Mid-level moisture availability
Specific Humidity (700hPa)	Essential for convective initiation
U-Wind (850hPa & 200hPa)	Resolves Vertical Wind Shear
V-Wind (850hPa & 200hPa)	Resolves Vertical Wind Shear

A.2 HYPERPARAMETER CONFIGURATION

To ensure the reproducibility of the results (MAE = 0.94 for Major Hurricanes), the model was trained using the following configuration:

- **Sequence Length (T):** 12 months (Sliding window).
- **CNN Backbone:** 2 layers (32, 64 filters) with Global Average Pooling.
- **Transformer Blocks:** 6 layers, 2 attention heads per layer.
- **Dropout Rate:** 0.3 (applied to Dense layers and Attention).
- **Optimization:** AdamW with Weight Decay (10^{-3}) and a Learning Rate of 5×10^{-5}.
- **Batch Size:** 16.
- **Early Stopping:** Monitor `val_loss` with patience = 30 epochs.