ML Lab Research Proposal

Hybrid-Weather: Combining RNN/LSTM and Gradient Boosting for Extreme Weather Prediction

Team: Team B - Anubhav Chatterjee & Arya Addagarla

Date: April 19, 2026

Project: Hybrid-Weather: Combining RNN/LSTM and Gradient Boosting for Extreme Weather Prediction

Target Compute: 1x AMD Instinct MI300X GPU, approx. 2 weeks

---

1. PROBLEM STATEMENT

Accurately predicting extreme weather events (e.g., heatwaves, storms) remains difficult due to the separation of short-term sensor data and long-term climate patterns in existing models. This project aims to improve prediction accuracy by combining both data types into a unified AI model. Better forecasts can reduce economic damage and improve public safety.

2. PROPOSED APPROACH

We will develop a hybrid model that combines a recurrent neural network (RNN/LSTM) for short-term temporal patterns with a gradient boosting model for long-term climate trends. The outputs will be fused using a meta-model (ensemble method).

This approach is chosen because LSTMs handle time-series dependencies well, while boosting models capture structured, non-linear relationships. The ensemble allows leveraging the unique strengths of both paradigms over traditional single-architecture setups.

3. DATASET

We will use publicly available weather datasets such as NOAA or Kaggle weather data, including temperature, humidity, wind speed, and precipitation over time. The dataset will include multiple years of hourly or daily observations.

Preprocessing

- Handling missing data points via statistical imputation or backward filling.
- Standard normalization across continuous feature columns.
- Feature engineering, including rolling averages, historical lag variables, and cyclic time feature encoding.

4. EVALUATION METRICS

Experiment Category	Metric	Target / Interpretation
Prediction Accuracy	Mean Absolute Error (MAE) & Root Mean Squared Error (RMSE)	Evaluates the continuous track precision of the model
Extreme Event Detection	Precision, Recall, and F1-score	Ensures high sensitivity for identifying rare, high-impact weather anomalies

5. COMPUTE REQUIREMENTS

Training is highly efficient and optimized for standard multi-core setups, utilizing high-performance acceleration where necessary.

Phase	Hardware	Estimated Wall-Clock	Notes
Model Architecture Training & Tuning	1x AMD Instinct MI300X GPU	1-3 hours	Fits comfortably inside a fraction of the HBM memory buffer

Software Dependencies

- Python 3.10+, PyTorch or TensorFlow framework core
- scikit-learn, pandas, NumPy, and gradient boosting bindings
- ROCm compatibility setup for execution on AMD hardware infrastructure

6. EXPECTED DELIVERABLES

- A fully trained and optimized hybrid LSTM-Gradient Boosting model.
- A short research report summarizing data preparation methods, architectural choices, and evaluation results.
- A simple demo notebook containing prediction visualizations overlaid on top of genuine test data sequences.

7. RISK & MITIGATION

Risk	Likelihood	Mitigation
Insufficient improvement over baseline architectures due to limited data or time constraints	Medium	Implement a simpler baseline model (e.g., linear regression or single ML model) for immediate comparison. If the hybrid model underperforms, we will refine engineering features or reduce model complexity to ensure stable results.