Project Introduction

Improving Precipitation Forecast Accuracy

This internship centers on developing a robust weather prediction pipeline with an emphasis on precise rainfall forecasting. Over two weeks, the intern will design, implement, and evaluate a series of models and verification tools using historical weather data:

- Week 1: The focus is on data collection, preprocessing, baseline forecast development (logistic regression for rain/no-rain, ARIMA/linear models for accumulation), and the implementation of core verification metrics including Threat Score (TS), RMSE, MAE, Precipitation Accuracy Score (PAS), and Fractional Skill Score (FSS) vlab.noaa.gov.
- Week 2: Advancing into machine learning (Random Forest, XGBoost, SVR) and deep learning approaches (LSTM, CNN-LSTM), the intern will apply self-supervised bias correction, ensemble strategies for uncertainty quantification, and threshold-focused validation. The goal is to optimize both event detection and rainfall volume accuracy—especially for extreme events—assessed via metrics like Brier Score, ROC-AUC, reliability diagrams, and ensemble spread analysis.

By the end of the internship, the candidate will have delivered:

- A structured modeling pipeline (baseline → ML → DL → ensemble)
- A comprehensive verification toolkit covering spatial, probabilistic, and volume-based metrics
- A final presentation showcasing performance comparisons and recommendations for enhancing operational forecast systems

Week 1 — Data & Modelling Foundations

Day 1 – Data Acquisition & Exploration

- **Objective:** Collect clean historical weather data.
- **Actions:**
    - Retrieve ~2 years of hourly data (rain amount, temperature, humidity, wind, pressure).
    - Analyze data: missing values, distributions, seasonality, and quality.

Day 2 – Preprocessing & Feature Engineering

- **Objective:** Prepare meaningful model inputs.
- **Actions:**
    - Impute missing data (linear interpolation, KNN).
    - Engineer features: lagged rainfall (past 1h, 3h, 6h), cumulative rainfall, humidity trends, and calendar attributes.
    - Normalize/standardize features for regression/ML models.

Day 3 – Baseline Precipitation Forecasts

- **Objective:** Set baseline for rain detection & amount.
- **Actions:**
    - Logistic regression for **rain/no-rain** and linear regression/ARIMA for **rain amount**.
    - Review evaluation metrics:
        - **Threat Score (TS)** for classification egusphere.copernicus.org+9wpc.ncep.noaa.gov+9wcrp-climate.org+9
        - **RMSE/MAE** for continuous forecasts

Day 4 – Precipitation Scoring Metrics

- **Objective:** Expand metric suite for event and spatial accuracy.

- **Actions:**
    - Compute **Precipitation Accuracy Score (PAS)** to capture both occurrence and volume researchgate.net
    - Implement **Fractional Skill Score (FSS)** for spatial consistency across neighborhoods wcrp-climate.org+15journals.ametsoc.org+15html.rhhz.net+15

Day 5 – Advanced Verification Tools

- **Objective:** Build probabilistic and spatial diagnostic capacity.
- **Actions:**
    - Build probabilistic classifiers or regression with thresholded events.
    - Score with **Brier Score**, **Reliability Diagrams**, **ROC-AUC**.
    - Compute **ESO metrics**: CSI, FAR, Bias, and use FSS to handle neighborhood displacement issues arxiv.orgscores.readthedocs.io+2html.rhhz.net+2researchgate.net+2wcrp-climate.org

Day 6 – Machine Learning Modeling

- **Objective:** Improve predictions with data-driven methods.
- **Actions:**
    - Train ML models (Random Forest, XGBoost, SVR) on engineered features to forecast amount and rain presence.
    - Evaluate using thresholds (e.g., ≥1 mm) with TS, PAS, FSS, MAE/RMSE.

Day 7 – Deep Learning Exploration

- **Objective:** Leverage temporal/spatial patterns for better forecasts.
- **Actions:**
    - Prototype LSTM or 1D-CNN for sequential temporal features.
    - If spatial grids available, explore CNN-LSTM.
    - Score models with PAS, FSS, continuous and probabilistic metrics.

Week 2 — Refinement, Ensembles & Reporting

Day 8 – Self-Supervised Post-Processing (SSLPDL)

- **Objective:** Enhance predictive reliability via self-supervised learning.
- **Actions:**
    - Fine-tune NWP-based forecasts using SSLPDL for rainfall probability estimation en.wikipedia.org+3learningweather.psu.edu+3arxiv.org+3wcrp-climate.org+1egusphere.copernicus.org+1arxiv.org+1journals.ametsoc.org+1
    - Compare improvements in RMSE, PAS, Brier Score.

Day 9 – Ensemble Forecasting & Uncertainty

- **Objective:** Combine models for robustness.
- **Actions:**
    - Generate ensembles (e.g., averaging across baseline, ML, DL models).
    - Evaluate ensemble spread, calibration, and reliability diagrams.
    - Use FSS to compare ensemble vs single-model spatial skill.

Day 10 – Spatial & Extreme Event Focus

- **Objective:** Fine-tune toward heavy rainfall accuracy.
- **Actions:**
    - Test models over thresholds (e.g., ≥10 mm).
    - Score with PAS across thresholds journals.ametsoc.org+4egusphere.copernicus.org+4wcrp-climate.org+4wired.com+1sciencedirect.com+1researchgate.net+11researchgate.net+11mausamjournal.imd.gov.in+11sciencedirect.com+5wcrp-climate.org+5nature.com+5
    - Analyze FSS at varying neighborhood scales for performance localization arxiv.org+3html.rhhz.net+3arxiv.org+3

Day 11 – Tooling & Efficiency

- **Objective:** Ensure computational efficiency and proper use of metrics.
- **Actions:**
    - Optimize FSS with summed-area table methods researchgate.net+1learningweather.psu.edu+1
    - Ensure scalable pipelines and avoid "double-penalty" effects in evaluations.

Day 12 – Final Model Selection & Validation

- **Objective:** Consolidate insights and pick top-performing model/ensemble.
- **Actions:**
    - Compare performance across models using RMSE, PAS, TS, FSS, Brier.
    - Evaluate on unseen test data period (e.g., last 6 months).
    - Identify best structured system for daily use.

Day 13 – Reporting & Presentation Prep

- **Objective:** Summarize learnings and propose next steps.
- **Actions:**
    - Create summary slides/report with:
        - Data sources and feature summaries
        - Evaluation metric definitions
        - Performance tables and charts
        - Reliability diagrams and FSS spatial maps
        - Precipitation threshold analysis
    - Present roadmap:
        - Incorporate spatial grids, NWP outputs, transformer models
        - Apply SSLPDL for real-time traffic
        - Operational deployment suggestions

Day 14 – Feedback, Documentation, & Handover

- **Objective:** Finalize deliverables and gather feedback.
- **Actions:**
    - Polish notebooks, codebase, and documentation (README, use guidelines).
    - Submit final code, presentations, and datasets.
    - Deliver final walk-through to mentor/team and seek improvement suggestions.

📊 **Metric Overview**

Metric	Type	Purpose
Threat Score (TS)/CSI	Binary event	Hit rate for thresholded rainfall (html.rhhz.net, arxiv.org, learningweather.psu.edu, researchgate.net, wpc.ncep.noaa.gov, arxiv.org, wired.com, journals.ametsoc.org)
RMSE / MAE	Continuous	Quantifies forecast error magnitude
Precipitation Accuracy Score (PAS)	Event + volume	Combines timing & volume accuracy
Fractional Skill Score (FSS)	Spatial/neighborhood	Mitigates location bias
Brier Score / ROC-AUC / Reliability	Probabilistic	Evaluates calibration of rain probability
Ensemble Stats	Spread & calibration	Quantifies uncertainty and forecast reach

Deliverables at Completion:

1. Organized code and notebooks
2. Final model pipelines (baseline, ML, DL, ensemble)
3. Complete scoring scripts for all metrics
4. Results documentation (tables & graphics)
5. Final presentation and recommendation deck