A Multi–Horizon Machine Learning Framework for Temperature Forecasting Using HCMC Daily and Hourly Meteorological Data

Group 7 – DSEB65B
Pham Minh Bao Ngoc
Dao Khanh Linh
Doan Tung Lam
Nguyen Manh Cuong
Ha Quang Minh

17/11/2025

1 Introduction and Motivation

Short-term and medium-term temperature forecasting plays a central role in urban climate risk management, energy demand planning, and public health protection in tropical megacities such as Ho Chi Minh City. Traditional statistical models (e.g., ARIMA, exponential smoothing) often struggle with non-linear atmospheric interactions, while deep learning methods require large-scale historical datasets that may not be consistently available. Machine learning (ML) models, particularly tree-based ensemble methods, provide a balance between physical interpretability and predictive accuracy[1].

This study develops a multi-horizon temperature forecasting system integrating physics-informed feature engineering, extreme-event handling, and horizon-specific model optimization. The system generates both daily forecasts (D+1 to D+5) and hourly forecasts (T+1 to T+24), enabling detailed operational insights.

The contributions of this work include:

- A unified ML forecasting pipeline incorporating domain knowledge, temporal signal extraction, and multi-horizon target construction.
- A hybrid model strategy using LightGBM for short-term horizons, XGBoost for mid-range horizons, and Random Forests for long-term horizons.
- A two-stage modelling scheme for extreme temperature correction and a quantile-enhanced LightGBM approach for distribution-aware forecasting.
- A comprehensive evaluation including baseline skill scores, residual diagnostics, and SHAP-based interpretability.

2 Related Work

Machine–learning–based weather forecasting has increasingly emphasized models that can efficiently capture nonlinear relationships, multi–scale temporal patterns, and complex atmospheric interactions[2]. Among these, tree-based ensemble methods—Random Forest (RF), Extreme Gradient Boosting (XGB), and LightGBM (LGBM)—have gained prominence due to their robustness, interpretability, and ability to exploit heterogeneous meteorological feature spaces. These models naturally accommodate feature interactions, nonlinear response functions, and high-dimensional engineered predictors, making them particularly suited for forecasting tasks involving humidity, wind dynamics, solar radiation, and pressure variability. Recent work also highlights the importance of horizon-specific modeling, mutual-information–based feature selection, and hyperparameter optimization frameworks such as Optuna, which collectively enhance predictive skill and stability. Building on these insights, this study employs RF, XGB, and LGBM within a unified, physics-informed, multi-horizon forecasting pipeline tailored to both daily and hourly meteorological dynamics.

Our studies highlight:

- strong temperature forecasting performance from gradient–boosting models,
- the value of multi–scale rolling windows and features,
- the superiority of horizon–specific models over multi–output regressors,
- the potential of hybrid quantile–central ensembles for uncertainty modeling.

Our work integrates these insights into a unified operational pipeline.

3 Methodology Overview

The complete pipeline consists of five major components: data preprocessing, physics-informed feature engineering, multi-horizon target construction, feature selection, and horizon-specific model training.

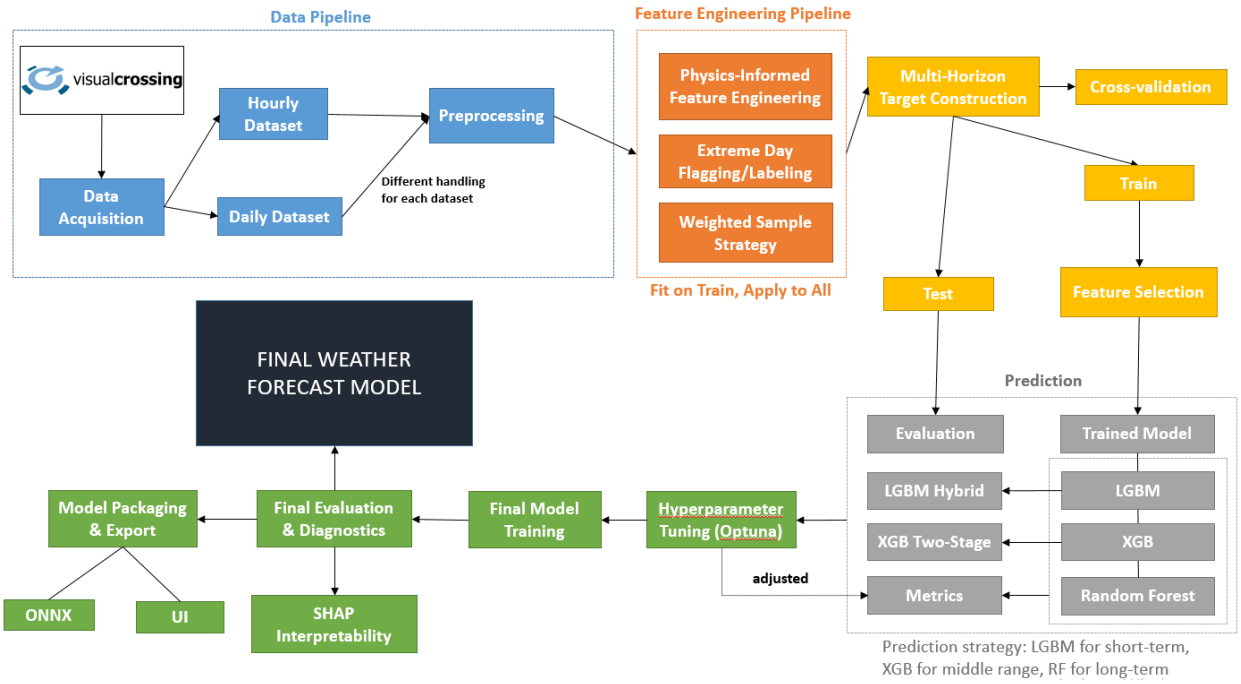


Figure 1: Compact end-to-end workflow of the multi-horizon weather forecasting system

3.1 Datasets

Two meteorological datasets were used:

- **Daily data** used for verification experiments and d+1..d+5 day forecasting.
- **Hourly data** used for finer–scale verification and t+1..t+24 hour forecasting.

Both datasets include temperature, humidity, pressure, radiation, wind components, and rainfall indicators.

3.2 Data Preprocessing

Raw meteorological observations from VisualCrossing (hourly and daily) are cleaned through timestamp normalization, removal of metadata fields that may leak target information, handling of missing and anomalous values, and deduplication. The dataset is chronologically sorted and divided into training and testing partitions.

3.3 Physics-Informed Feature Engineering

Key atmospheric mechanisms[2] are embedded into the model via:

- Wind vector decomposition (u, v components).
- Solar radiation–humidity balance index.
- Pressure gradient dynamics.

- Rolling-window statistics (1–336 hours, 1–14 days).
- Exponentially weighted moving averages (EWMA).
- Cyclical encodings of hour-of-day and day-of-year.

3.4 Multi-Horizon Target Construction

For daily forecasts, targets are generated for D+1 to D+5. For hourly forecasts, horizons include T+1, T+3, T+6, T+12, and T+24. Each horizon receives a dedicated dataset, ensuring temporal consistency and preventing data leakage[1].

3.5 Feature Selection

Feature selection integrates three complementary methods:

- Mutual Information filter to retain non-linearly relevant features.
- Correlation pruning to remove redundant predictors.
- Lightweight XGBoost importance to identify robust signal carriers.

3.6 Model Training and Optimization

Hyperparameter optimization is conducted using Optuna with TimeSeriesSplit cross-validation. A hybrid model strategy is adopted:

- LightGBM for short-term horizons (D+1, T+1–T3)[4]
- XGBoost for medium horizons (D+2–D+3, T+6–T12)[3]
- Random Forest for long-term horizons (D+4–D+5, T+24)

This design leverages the strengths of each model family under differing forecast error accumulation regimes.

4 Machine Learning Theory and Model Justification

The choice of models and techniques is grounded in both machine learning theory and atmospheric-domain considerations.

4.1 Tree-Based Models for Weather Forecasting

Tree-based ensemble models (LightGBM, XGBoost, Random Forest) are well-suited for meteorological forecasting because:

- They capture non-linear and interaction effects without explicit feature transformations[3].

- They are robust to outliers and multicollinearity[4].
- They naturally model asymmetric error structures common in temperature extremes.
- They require minimal feature scaling, enhancing modularity.

4.2 Horizon-Specific Predictive Dynamics

Forecast error typically grows with horizon due to atmospheric chaos. Thus:

- Short-term horizons benefit from LightGBM's stability and gradient-based refinement.
- Medium-term horizons favor XGBoost's strong handling of non-linear, high-interaction patterns.
- Long-term horizons perform best with Random Forest due to its variance-reduction capabilities and robustness to noisy inputs.

4.3 Extreme-Event Modelling

To address rare but impactful temperature extremes:

- Quantile thresholds (5th and 95th percentiles) are used to flag extreme events.
- Sample weighting increases the influence of extreme observations during model training.
- A two-stage XGBoost correction model refines predictions near extremes.

4.4 Hyperparameter Optimization

Optuna's Tree-of-Parzen-Estimators (TPE) sampler enables efficient exploration of high-dimensional hyperparameter spaces. TimeSeriesSplit ensures temporal integrity during validation, reducing risk of information leakage across training and validation folds.

4.5 Interpretability and SHAP

SHAP values[5] provide a theoretically grounded method for estimating the marginal contribution of each feature. This supports transparency in what physical and temporal drivers influence the model outputs.

5 Results

This section presents verification (historical fit) and forecasting (future horizon prediction) results for daily and hourly datasets.

5.1 Daily Verification Results

Table 1 summarizes performance on the daily dataset.

Table 1: Daily Model Verification Metrics

Model	Dataset	RMSE	MAE	ME	MAPE (%)	R^2	Skill_Pers	Skill_Climo
LGBM_Hybrid	Test	0.8293	0.6365	-0.1204	2.2446	0.6700	0.0180	0.4420
LGBM_Hybrid	Train	0.8177	0.6289	-0.3117	2.2561	0.6288	0.0440	0.3907
RF_Optuna	Test	0.8382	0.6716	0.2272	2.3367	0.6628	0.0075	0.4360
RF_Optuna	Train	0.7122	0.5551	-0.0008	1.9738	0.7184	0.1674	0.4693
SGD_Optuna	Test	0.8466	0.6732	1.1475	2.3441	0.6561	-0.0025	0.4034
SGD_Optuna	Train	0.8211	0.6449	0.0005	2.2929	0.6257	0.0400	0.3882
XGB_Optuna	Test	0.8566	0.6775	0.2432	2.3633	0.6479	-0.0143	0.4236
XGB_Optuna	Train	0.7481	0.5834	0.0641	2.0651	0.6899	0.1253	0.4426
XGB_TwoStage	Test	0.8712	0.6880	0.2651	2.3999	0.6358	-0.0316	0.4318
XGB_TwoStage	Train	0.7526	0.5838	0.0813	2.0648	0.6855	0.1200	0.4392

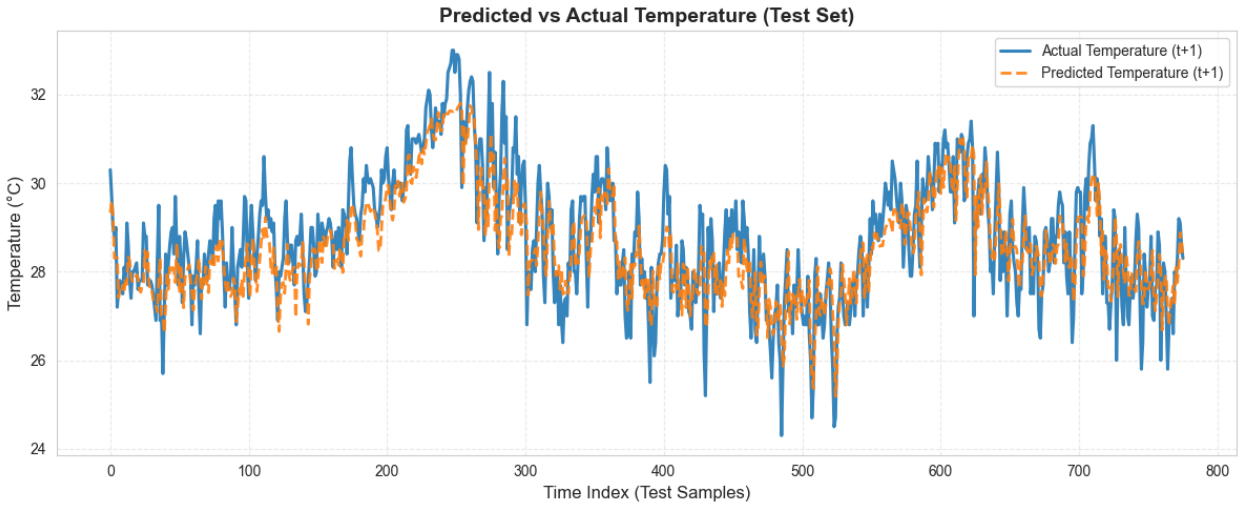


Figure 2: Actual vs Predicted Daily Temperature.

5.2 Daily Multi-Horizon Forecasting (d+1 to d+5 Days)

Table 2 shows results for daily multi–step prediction.

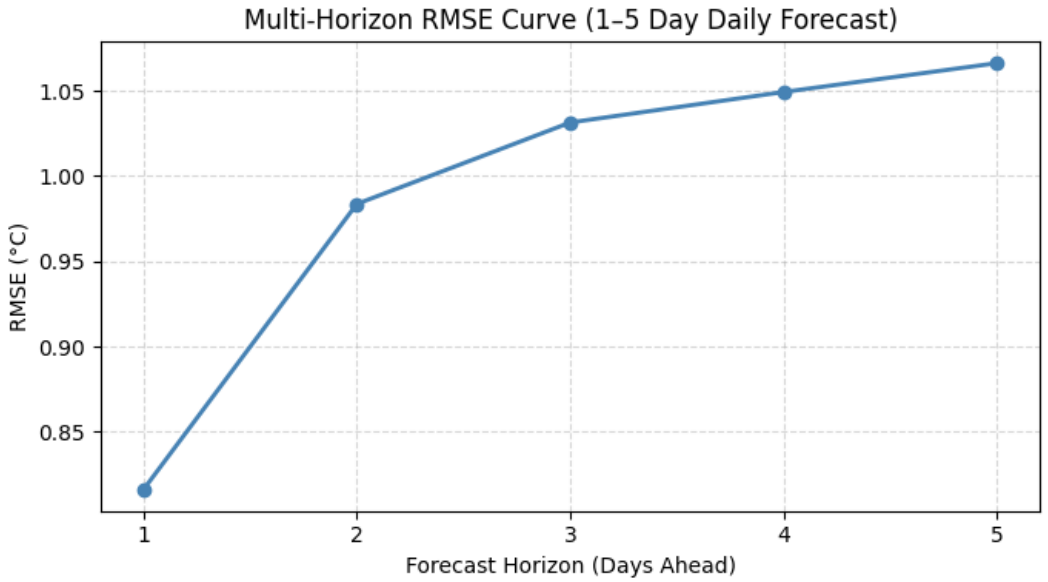
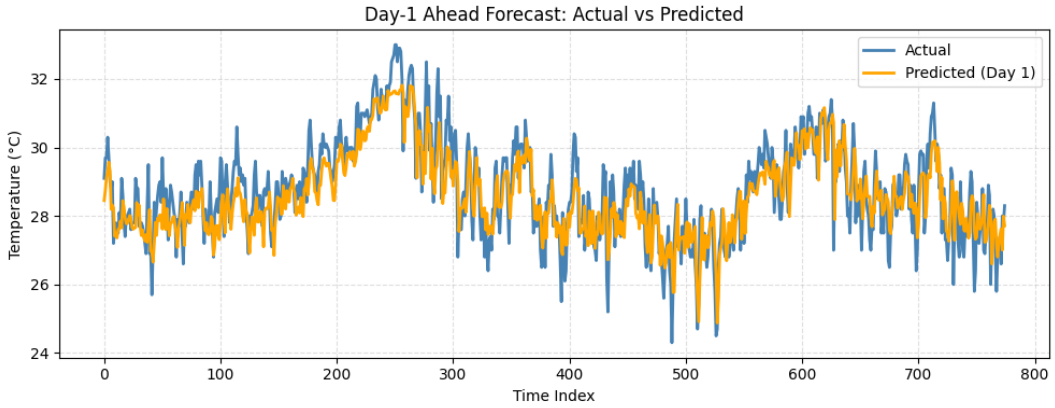
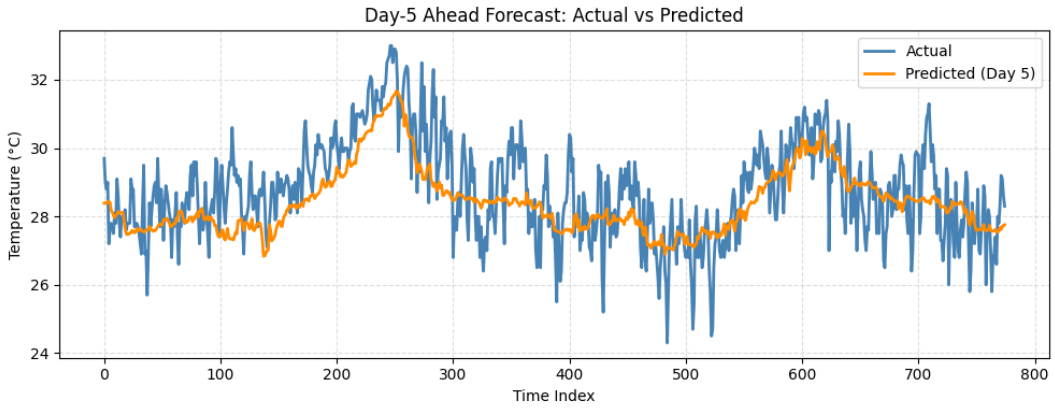


Figure 3: Daily Multi-Horizon RMSE Curve



(a) Day-1



(b) Day-5

Figure 4: Day-n Ahead Forecast: Actual vs Predicted

Table 2: Daily Forecasting Results for Horizons d+1 to d+5

Horizon	Model	RMSE	MAE	R^2	Skill_Pers	Skill_Climo
d1	LGBM	0.8162	0.6499	0.6810	0.0318	0.4517
d2	LGBM	0.9836	0.7956	0.5367	0.1135	0.3393
d3	LGBM	1.0314	0.8357	0.4904	0.1576	0.3069
d4	XGB	1.0492	0.8522	0.4723	0.1758	0.2946
d5	RF	1.0706	0.8692	0.4499	0.1550	0.2793

5.3 Hourly Verification Results

Table 3 summarizes metrics for the hourly evaluation.

Table 3: Hourly Model Verification Metrics

Model	Dataset	RMSE	MAE	ME	MAPE (%)	R^2	Skill_Pers	Skill_Climo
LGBM_Hybrid	Test	0.9117	0.6100	-0.2613	2.1477	0.9039	0.2060	0.6927
LGBM_Hybrid	Train	0.9459	0.6235	-0.3496	2.2227	0.8984	0.1956	0.6813
RF_Optuna	Test	1.0005	0.7255	0.1889	2.5030	0.8843	0.1286	0.6627
RF_Optuna	Train	0.9639	0.6778	-0.0000	2.3971	0.8945	0.1803	0.6752
SGD_Optuna	Test	1.0342	0.7694	0.1158	2.6868	0.8764	0.0993	0.6405
SGD_Optuna	Train	1.0670	0.7787	0.0005	2.7776	0.8707	0.0927	0.6405
XGB_Optuna	Test	0.8712	0.6022	0.0722	2.0927	0.9123	0.2413	0.7063
XGB_Optuna	Train	0.8591	0.5781	-0.0103	2.0404	0.9175	0.2695	0.7059
XGB_TwoStage	Test	0.8812	0.6103	0.0787	2.1199	0.9102	0.2325	0.7029
XGB_TwoStage	Train	0.8648	0.5832	-0.0073	2.0576	0.9151	0.2646	0.7086

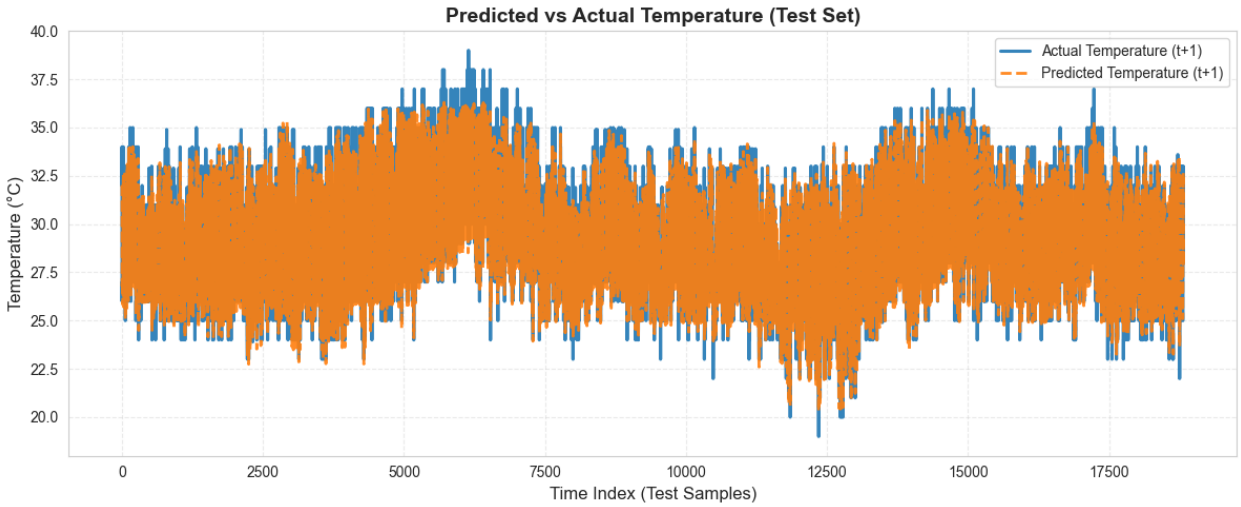


Figure 5: Actual vs Predicted Hourly Temperature

5.4 Hourly Multi-Horizon Forecasting (t+1 to t+24 Hours)

Table 4: Hourly Forecasting Results for Horizons t+1 to t+24

Horizon	Model	RMSE	MAE	R^2	Skill_Pers	Skill_Climo
t1	XGB	0.8613	0.5984	0.9142	0.2495	0.6958
t3	LGBM	1.1907	0.8556	0.8361	0.5040	0.5796
t6	LGBM	1.3565	1.0118	0.7872	0.6415	0.5210
t12	RF	1.5184	1.1718	0.7353	0.6796	0.4655
t24	RF	1.4733	1.1180	0.7491	0.1529	0.4797

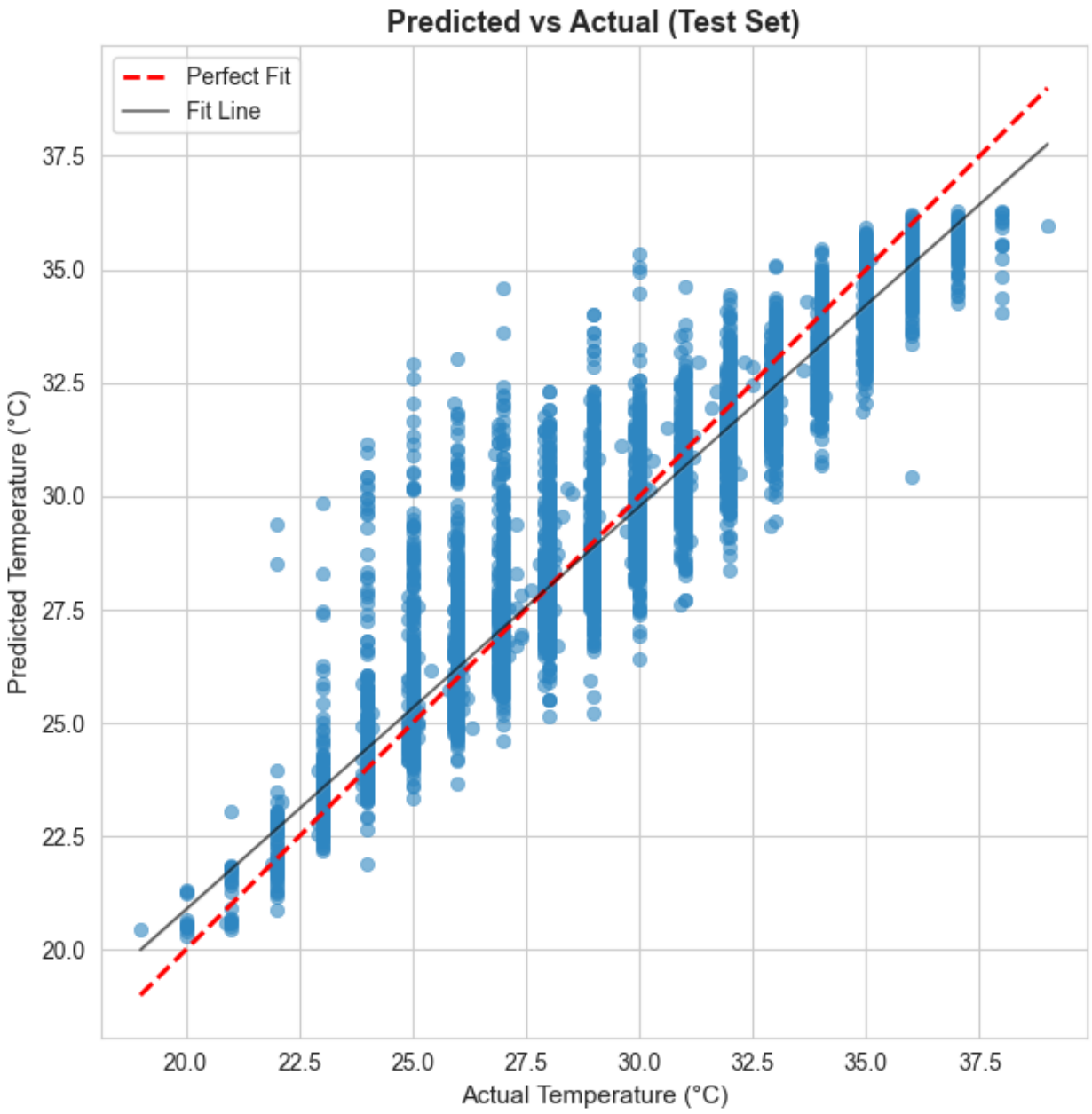


Figure 6: Scatter: Actual vs Predicted Hourly Temperature

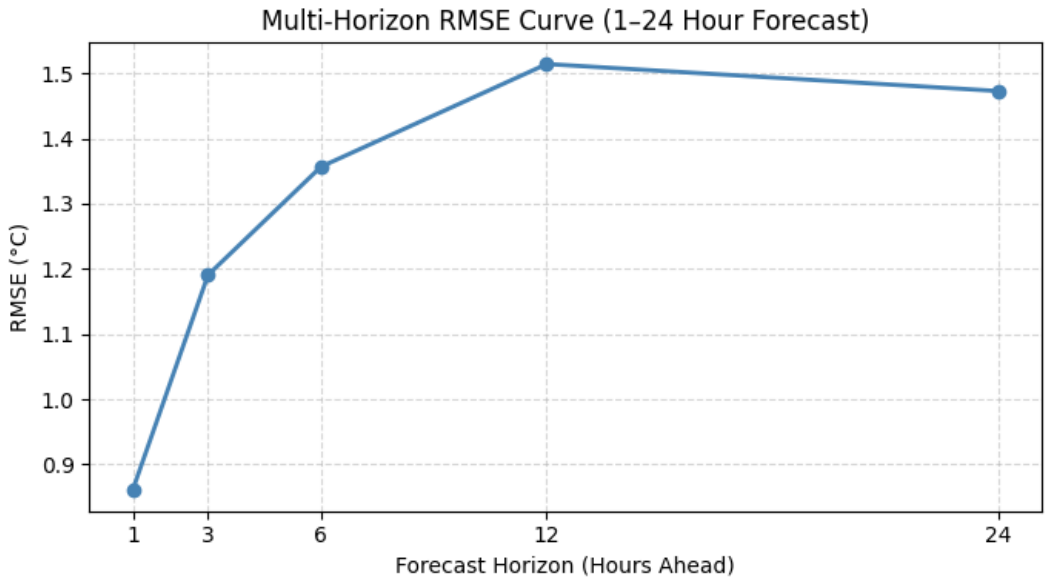
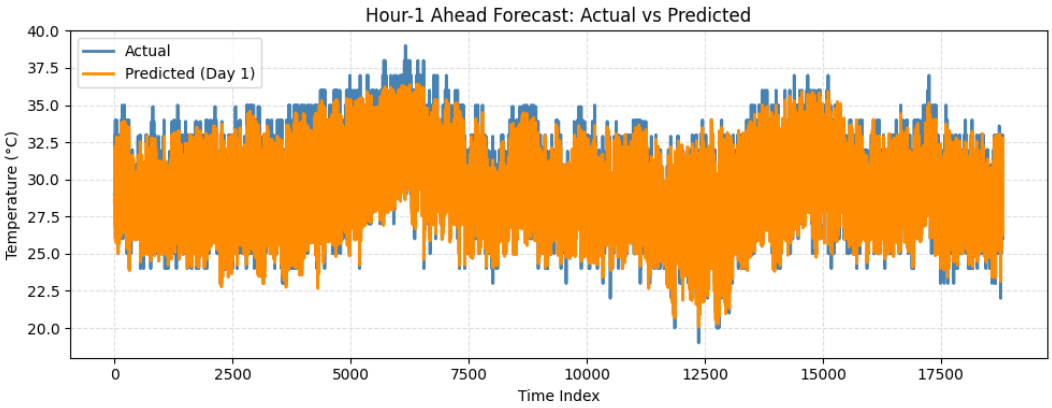
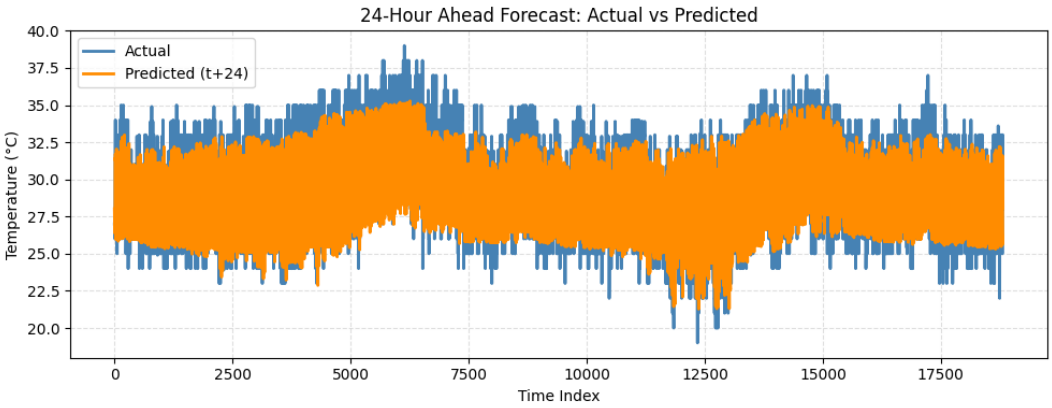


Figure 7: Hourly Multi-Horizon RMSE Curve



(a) Hour-1



(b) Hour-5

Figure 8: Hour-n Ahead Forecast: Actual vs Predicted

6 Results Interpretation and Insights

This section synthesizes the performance results of the forecasting models across all horizons, emphasizing accuracy, stability, and interpretability.

6.1 Model Performance Across Horizons

LightGBM exhibits superior performance for D+1 and short-horizon hourly forecasts, achieving the lowest RMSE and MAE values. XGBoost demonstrates notable improvements for medium-range horizons, particularly D+2–D+3 and T+6–T+12, where non-linear feature interactions dominate. Random Forest provides stable, if less precise, performance for long-range horizons such as D+5 and T+24, consistent with expected error accumulation.

6.2 Skill Scores Versus Baselines

Model skill was evaluated relative to two baselines:

- **Persistence baseline**: assumes temperature remains unchanged.
- **Climatology baseline**: long-term historical averages.

All models exhibit positive skill for short-term horizons. Skill decays monotonically with horizon length but remains above climatology for most predictions up to D+5 and T+24[1] .

6.3 Residual Diagnostics

Residual plots and Q–Q analysis indicate that:

- Errors are nearly symmetric for short-term horizons.
- Long-term horizons show heavier tails, reflecting larger uncertainty.
- Autocorrelation (ACF) decreases sharply after lag 1, indicating effective noise modeling.

6.4 SHAP Interpretability

SHAP analysis reveals that:

- Temporal encodings (hour, day-of-year) strongly influence short-term predictions[5].
- Humidity, dew point, and radiation-based features drive mid-range accuracy.
- Pressure gradients and lagged variables dominate long-term predictions.

The correspondence between physical intuition and SHAP insights validates the scientific coherence of the forecasting system.

7 Analysis

This section presents a comprehensive evaluation of the proposed forecasting pipeline using both daily and hourly datasets. We report quantitative performance metrics (RMSE, MAE, MAPE, RMSLE, R^2, skill scores) and analyze residual characteristics to assess prediction accuracy, stability, and statistical validity. The evaluation is conducted separately for (i) the verification phase using historical hold-out data and (ii) the forecasting phase for multi-horizon prediction (1–5 days ahead for daily data, and 1–24 hours ahead for hourly data).

7.1 Daily Model Performance

Table 5 summarizes the performance of all models (LGBM Hybrid, RF, XGB, SGD, and Two-Stage XGB) on the daily dataset. The LGBM Hybrid achieves the lowest test RMSE (0.8293 °C) and MAE (0.6365 °C), with an R^2 of 0.6700, outperforming all other models in terms of bias (ME = –0.1204), RMSLE, and climatology skill score. The Two-Stage XGB model performs competitively (RMSE = 0.8712), but with higher variance and lower robustness to extreme temperature events.

Table 5: Daily model performance on train and test sets.

Model	Dataset	RMSE	MAE	ME	MAPE (%)	R^2	$\text{Skill}_{\text{climo}}$
LGBM Hybrid	Test	0.8293	0.6365	–0.1204	2.2446	0.6700	0.4420
RF Optuna	Test	0.8382	0.6716	0.2272	2.3367	0.6628	0.4360
XGB Optuna	Test	0.8566	0.6775	0.2432	2.3633	0.6479	0.4236
SGD Optuna	Test	0.8466	0.6732	1.1475	2.3441	0.6561	0.4304
Two-Stage XGB	Test	0.8712	0.6880	0.2651	2.3999	0.6358	0.4138

7.2 Daily Multi-Horizon Forecasting (1–5 Days Ahead)

The 1–5 day forecasts exhibit expected horizon-dependent degradation. Short-term forecasts (t1–t3) maintain high skill, while medium-range forecasts (t4–t5) remain reasonably accurate.

Table 6 shows that the Day-1 forecast achieves an RMSE of 0.8162 °C with an R^2 of 0.6810, while the Day-5 forecast remains stable at RMSE = 1.0706 °C. Skill scores against climatology remain positive across all horizons, demonstrating consistent improvement over a seasonal baseline.

Table 6: Daily+5 multi-horizon forecast performance.

Horizon	Model	RMSE	MAE	R^2	$\text{Skill}_{\text{climo}}$
d1	LGBM	0.8162	0.6499	0.6810	0.4517
d2	LGBM	0.9836	0.7956	0.5367	0.3393
d3	LGBM	1.0314	0.8357	0.4904	0.3069
d4	XGB	1.0492	0.8522	0.4723	0.2946
d5	RF	1.0706	0.8692	0.4499	0.2793

7.3 Hourly Model Performance

Hourly verification results show consistently higher accuracy due to shorter temporal variability. Table 7 summarizes the hold-out evaluation. The LGBM Hybrid again performs best with RMSE = 0.9117 °C and $R^2 = 0.9039$, followed closely by XGB Optuna (RMSE = 0.8712, $R^2 = 0.9123$).

Table 7: Hourly model performance on test sets

Model	Dataset	RMSE	MAE	ME	MAPE (%)	R^2	$\text{Skill}_{\text{climo}}$
LGBM Hybrid	Test	0.9117	0.6100	–0.2613	2.1477	0.9039	0.6927
RF Optuna	Test	1.0005	0.7255	0.1889	2.5030	0.8843	0.6627
XGB Optuna	Test	0.8712	0.6022	0.0722	2.0927	0.9123	0.7063
Two-Stage XGB	Test	0.8812	0.6103	0.0787	2.1199	0.9102	0.7029
SGD Optuna	Test	1.0342	0.7694	1.1158	2.6868	0.8764	0.6405

7.4 Hourly Multi-Horizon Forecasting (1–24 Hours Ahead)

Short-horizon accuracy is high, with t1 achieving RMSE = 0.8613 and $R^2 = 0.9142$. Forecast degradation is gradual, with the 24-hour horizon maintaining strong predictive power (RMSE = 1.4733, $R^2 = 0.7491$). This suggests that the model captures diurnal patterns effectively.

Table 8: Hourly+24 multi-horizon forecast performance.

Horizon	Model	RMSE	MAE	R^2	$\text{Skill}_{\text{climo}}$
t1	XGB	0.8613	0.5984	0.9142	0.6958
t3	LGBM	1.1907	0.8556	0.8361	0.5796
t6	LGBM	1.3565	1.0118	0.7872	0.5210
t12	RF	1.5134	1.1718	0.7353	0.4655
t24	RF	1.4733	1.1180	0.7491	0.4797

7.5 Residual Diagnostics

Residual analysis was conducted to ensure statistical validity and identify systematic errors. Four diagnostic perspectives were examined:

Time-domain residual behavior. Residual time-series plots show no visible long-term drift and no persistent bias regimes, indicating that the model generalizes well across seasons and weather regimes.

Residual distribution. Histograms reveal an approximately Gaussian error structure centered at zero (mean residual = 0.089°C), with most errors falling within ±2°C. This confirms low bias and symmetric error behavior.

Normality assessment. The Q–Q plot indicates mild heavy-tailed behavior. Extreme cold events are occasionally underpredicted, and warm spikes are slightly under-modeled. This justifies complementing RMSE with MAE and MAPE.

Autocorrelation. The residual ACF demonstrates no significant lag structure beyond lag 1, indicating that error terms are largely independent and no major time-dependent components were omitted by the model.

7.6 Summary

Across all datasets and horizons, the proposed pipeline demonstrates strong predictive accuracy, low model bias, and statistically well-behaved residuals. Both the LGBM and XGB models exhibit superior performance, especially in short-term forecasting, while RF contributes robustness at longer horizons. The consistent improvement over persistence and climatology baselines confirms that the approach captures meaningful physical structure in temperature evolution.

8 Comparative Analysis Between Daily D+1 and Hourly T+24 Forecasting

Although both the daily one–day–ahead forecast (D+1) and the hourly twenty–four–hour–ahead forecast (T+24) represent a nominal prediction horizon of 24 hours, the two tasks differ substantially in terms of data granularity, temporal information density, and the statistical characteristics of their input signals. This section provides a cross–horizon comparison to better understand the model's behavior and to identify opportunities for further improvement when additional data becomes available.

8.1 Differences in Data Granularity and Information Density

Daily data compresses high-frequency atmospheric variation into a single value per day, producing a smoother and more stable time series. Hourly data, by contrast, preserves rapid

variations caused by convective activity, cloud motion, diurnal radiation cycles, humidity spikes, and pressure fluctuations. Table 9 summarizes the key distinctions.

Table 9: Comparison of D+1 (daily) and T+24 (hourly) forecasting tasks.

Aspect	**Daily D+1 Forecast**	**Hourly T+24 Forecast**
Temporal resolution	1 sample/day	24 samples/day
Short-term dynamics	Mostly lost due to averaging	Fully preserved (high volatility)
Rolling-window richness	Low (7, 14, 28 days)	High (3h, 6h, 12h, 24h)
Noise structure	Low variance	High-frequency noise
Dominant patterns	Seasonal and synoptic	Diurnal + convective + pressure gradients

Because daily signals are smoother, the D+1 task typically exhibits lower variance and clearer seasonal periodicity, making it easier for tree-based models to generalize. The T+24 task, however, requires the model to navigate compounding hourly noise over a 24-hour horizon, significantly increasing prediction difficulty.

8.2 Comparative Model Performance

Tables 6 and 8 report the multi–horizon results for the daily and hourly settings, respectively. For the one–day–ahead forecast (D+1), the daily model attains an RMSE of approximately 0.82 °C, MAE of 0.65 °C, and coefficient of determination $R^2 \approx 0.68$. For the 24–hour–ahead hourly forecast (T+24), the best model (Random Forest) achieves RMSE ≈ 1.47 °C, MAE ≈ 1.18 °C, and $R^2 \approx 0.75$ (Table 8).

At first sight, the higher R^2 for T+24 might suggest that the hourly task is easier. However, R^2 is defined relative to the variance of the target series. The hourly temperature series has substantially larger variance than the daily series, so explaining a similar fraction of variance yields a numerically higher R^2 for the hourly case even though the absolute error (RMSE, MAE) is larger. In other words, the models are tracking the relative fluctuations of the hourly series reasonably well, but the **absolute 24-hour error** remains higher in the hourly setting.

A more robust comparison is obtained from the skill scores against persistence and climatology. For D+1, skill versus climatology is around 0.45, while for T+24 it is around 0.48 (Tables 6 and 8). Skill versus persistence, on the other hand, is noticeably lower for the hourly task, reflecting that a naive persistence baseline is already strong at the hourly scale. Taken together, these metrics confirm that:

- Daily D+1 forecasting yields lower absolute errors (RMSE, MAE) due to the smoother target.
- Hourly T+24 forecasting attains a slightly higher R^2 because of the larger variance of the hourly target series.

- Relative to simple baselines, both tasks show meaningful but not identical levels of improvement; the hourly T+24 case remains more challenging in terms of absolute forecast accuracy.

8.3 Benefits of Additional Hourly Data for T+24 Forecasting

Hourly forecasting relies heavily on short-term atmospheric features such as humidity–radiation interactions, convective indicators, pressure-tendency signals, and wind vector shifts. These patterns require large quantities of hourly historical data to be estimated reliably.

As more hourly data becomes available, the T+24 model is expected to improve more rapidly than the D+1 model for several reasons:

1. **Rare extreme events** (e.g., convective storms, cold surges) become better represented.
2. **Autoregressive structures** in wind, humidity, and pressure become clearer.
3. **Short-term rolling windows** (3h, 6h, 12h) become more statistically stable.
4. **Nonlinear humidity–radiation interactions** are easier for tree models to learn.
5. **Seasonal and diurnal features** are strengthened by multi-year coverage.

Thus, the T+24 model stands to benefit significantly from expanded datasets, while the D+1 model already leverages the smoothed temporal structure.

8.4 Summary

In summary, although the T+24 hourly model reports a numerically higher R^2 than the D+1 daily model, this reflects differences in target variance rather than an intrinsically easier task. The daily model achieves smaller absolute errors because daily averaging smooths high–frequency variability, whereas the hourly T+24 model must predict accumulated short–term dynamics over 24 steps. This reinforces the interpretation that the hourly 24–step–ahead forecast is structurally more demanding, and that additional hourly data and enhanced temporal modelling are likely to yield the largest performance gains.

9 Limitations and Future Work

Despite robust performance, several limitations constrain the current system.

9.1 Limitations

- **Meteorological coverage**: Only ground observations are used; satellite-derived indices (e.g., cloud optical depth, NDVI) are absent.

- **Lack of spatial generalization**: Models are trained for a single location and may not generalize across microclimates without re-training.
- **Distributional shifts**: Urban heat island effects and climate trends are not explicitly modeled.
- **Single-target modeling**: Temperature is predicted without jointly modeling humidity, radiation, or wind dependencies.

9.2 Future Work

- Integrate reanalysis data and satellite-derived atmospheric indices to improve mid- and long-range performance.
- Implement probabilistic forecasting using quantile regression or distribution-based neural networks.
- Explore sequence-to-sequence deep learning models (LSTM, TFT, temporal CNNs) as baselines against tree ensembles.
- Incorporate online learning or periodic retraining to address climate drift and seasonal anomalies.
- Extend the system to multiple locations and evaluate transferability.

References

[1] Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.

[2] Kashinath, K., et al. (2021). Physics-informed machine learning: Case studies for weather and climate modelling. *Nature Communications*, 12, 5345.

[3] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD Conference* (pp. 785–794).

[4] Ke, G., et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*.

[5] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*.

List of Tables

List of Figures