International Journal of Agriculture, Environment and Food Sciences

e-ISSN: 2618-5946 https://dergipark.org.tr/jaefs

DOI: https://doi.org/10.31015/2025.si.23 Int. J. Agric. Environ. Food Sci. 2025; 9 (Special): 153-161

ERA5–NASA Ensembles for Daily Rain Prediction Supporting Irrigation in Konya, Türkiye

Ali ÇETİNKAYA

Department of Computer Engineering, Faculty of Technology, Selçuk University, 42130, Konya, Türkiye

Article History
Received: November 4, 2025
Accepted: December 16, 2025
Published Online: December 28, 2025

Article Info
Type: Research Article
Subject: Agricultural Systems Analysis and Modelling

Corresponding Author
Ali Çetinkaya
ali.cetinkaya@selcuk.edu.tr

Author ORCID
https://orcid.org/0000-0002-7747-6854

Available at
https://dergipark.org.tr/jaefs/issue/99003/1816893

Abstract

Reliable daily precipitation prediction is pivotal for agricultural water management in semi-arid regions, where water scarcity and climate variability raise operational risk. This study develops and evaluates machine-learning approaches for precipitation occurrence in Konya, Türkiye - one of the country's key agricultural basins - using 44 years (1980–2024) of meteorological data. We examine three data strategies: (i) European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) reanalysis features alone, (ii) National Aeronautics and Space Administration (NASA) Prediction Of Worldwide Energy Resources (POWER) observations alone, and (iii) a combined strategy that merges ERA5 predictors with the NASA rain/no-rain target. Five model families are compared— Random Forest (RF), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Categorical Boosting (CatBoost), and a three-layer Long Short-Term Memory (LSTM) on analysis-ready tabular datasets. All approaches use advanced feature engineering (cyclical seasonality, lags, rolling windows) and a deployment-minded post-processing step that scans decision thresholds from 0.10 to 0.90 to maximize F1-score. Performance is assessed with a chronological train/test split; F1-score is the primary metric, complemented by Area Under the Receiver Operating Characteristic Curve (AUC-ROC), precision, recall, confusion matrices, and calibration. Results show that the combined strategy with CatBoost delivers the highest skill (F1-score: 84.57%; AUC-ROC: 94.37%), confirming that pairing rich ERA5 features with the quality-controlled NASA target improves performance relative to single-source approaches. Ensemble tree-based methods consistently outperform the LSTM baseline on this daily, tabular classification task. Threshold optimization raises F1-score by about 1–5% across models, and calibration indicates that predicted probabilities closely match observed frequencies. For semi-arid farming in Konya, calibrated daily probabilities of precipitation can be converted into simple decision rules (e.g., skip irrigation when the predicted probability exceeds a locally selected threshold), supporting irrigation scheduling, fertilizer timing, and harvest planning. The workflow is computationally efficient and reproducible, built on globally available ERA5 and NASA POWER resources, and readily transferable to other semi-arid basins with minor re-tuning. Findings highlight a practical path to trustworthy, probability-based rainfall guidance for agricultural water management. Feature importance analysis highlights seasonality, humidity, temperature ranges, pressure tendencies, and wind extremes as leading signals.

Keywords: Agricultural water management, ERA5, Machine learning, Model calibration, NASA POWER, Precipitation prediction

Cite this article as: Cetinkaya, A. (2025). ERA5–NASA Ensembles for Daily Rain Prediction Supporting Irrigation in Konya, Türkiye. International Journal of Agriculture, Environment and Food Sciences, 9 (Special): 153-161. https://doi.org/10.31015/2025.si.23

INTRODUCTION

Daily precipitation prediction is a linchpin of agricultural water management, particularly in semi-arid settings where each irrigation decision carries significant economic and environmental consequences. The Konya Basin—often called Türkiye's "grain basket"—exemplifies this challenge: the region's productivity is closely tied to rainfall variability, and long-term groundwater overdraft has contributed to land subsidence and water stress. Documented subsidence linked to groundwater depletion in the Konya Plain illustrates the stakes for robust precipitation-informed decision-making (Caló et al., 2017). More broadly, the Eastern Mediterranean, including Türkiye, is expected to face increased hydro-climatic stressors associated with climate change (e.g., more frequent dry spells and hotter conditions), further elevating the value of reliable local precipitation information (Lelieveld et al., 2012).

Traditional numerical weather prediction (NWP) models offer physically grounded forecasts, yet high-resolution local skill can be limited by computational costs and sub-grid process parameterizations (Rasp et al., 2018; Schultz et al., 2021). In parallel, machine learning has matured into a credible, complementary path for certain forecasting tasks—especially when flexible, data-driven models can be paired with rich reanalysis and observational datasets (Schultz et al., 2021; Bzdok et al., 2018).

At the operations level, daily precipitation occurrence forecasts have already shown tangible value when tied to agricultural planning: a JAEFS case study in Türkiye achieved 84.16% accuracy with logistic regression and 16.7% higher harvest revenue over a 120-day plan by integrating rain/no-rain forecasts into optimization (Samasti & Kucukdeniz, 2023).

In recent years, two global data resources have proven especially valuable for data-driven precipitation applications. Earlier neural-network rainfall studies—e.g., Luk et al. (2001) and Hung et al. (2009)—provide a methodological backdrop for our LSTM baseline and help situate the present comparison against tree-based ensembles. ERA5 (ECMWF Reanalysis v5), which is produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) and distributed by the Copernicus Climate Change Service (C3S) via the Copernicus Climate Data Store (CDS), provides physically consistent, spatially complete atmospheric fields at hourly timesteps and ~0.25° resolution, with well-documented advances relative to prior reanalysis generations (Hersbach et al., 2020; C3S, 2017; Muñoz-Sabater et al., 2021). National Aeronautics and Space Administration (NASA) Prediction Of Worldwide Energy Resources (POWER) offers quality-controlled, daily surface meteorology derived from satellite and model products, widely used in agriculture and renewable energy applications (NASA LRC, 2023; Sparks, 2018). Recent studies have increasingly highlighted the potential of multi-source data fusion and machine learning in enhancing local weather predictions (Gavahi et al., 2023; Samasti & Kucukdeniz, 2023).

Bringing these resources together raises practical questions that are central to operational hydrometeorology: (1) which data source—reanalysis, satellite/model-derived observation, or a hybrid—yields better skill for daily rain/no-rain decisions? (2) which ML (Machine Learning) model families best capture the patterns driving precipitation occurrence at the daily scale? (3) can we improve end-user performance via threshold optimization rather than accepting a default 0.5 cutoff ?

To answer these questions, we run a controlled head-to-head comparison of three data strategies (ERA5-only, NASA-only, ERA5-NASA combined) and five algorithms (RF, XGBoost, CatBoost, LightGBM, LSTM) for Konya. Our objectives are to: (i) quantify the performance of each data strategy; (ii) compare tree-based ensembles to deep learning for this tabular, daily classification task; (iii) optimize decision thresholds for F1-score.

To orient the reader, Figure 1 sketches the end-to-end decision chain from ERA5/NASA inputs to daily probability of precipitation and farm actions in Konya.

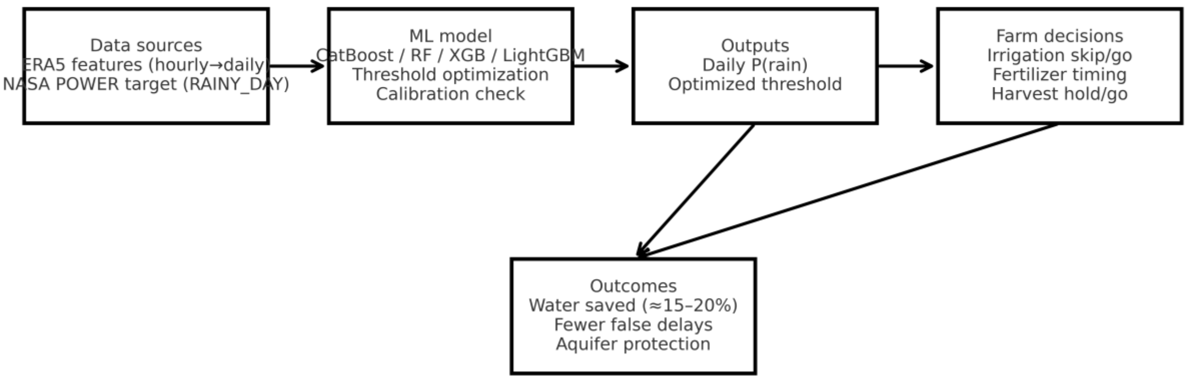

Figure 1. Decision chain linking ERA5 features and NASA POWER target to daily P(rain), threshold-based actions (irrigation skip/go, fertilizer timing, harvest hold/go), and water/efficiency outcomes in the Konya Basin.

MATERIALS AND METHODS

The Konya region (≈ 37.95° N, 32.55° E) combines a semi-arid climate with intensive agriculture and documented groundwater-related subsidence, making it an apt testbed for precipitation-aware water management models (Caló et al., 2017).

Data acquisition and curation

The raw meteorological data used here were downloaded and curated by the author from two primary global sources to create analysis-ready datasets tailored for ML: one from NASA POWER (daily) and one from ERA5 (hourly). The curation included cleaning, unit harmonization, temporal alignment, aggregation, merging, and the derivation of a binary target (RAINY_DAY) representing precipitation occurrence. The end-to-end workflow—from raw downloads to calibrated, threshold-optimized predictions—is summarized in Figure 2.

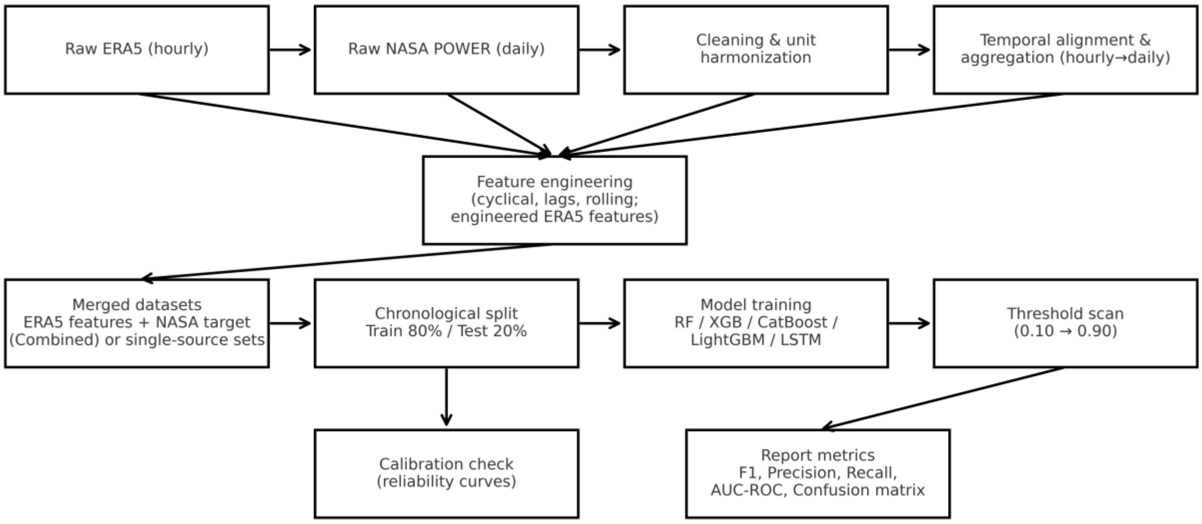

Figure 2. End-to-end workflow: raw ERA5 (hourly) and NASA POWER (daily) → cleaning and unit harmonization → temporal alignment/aggregation → feature engineering (cyclical, lags, rolling; engineered ERA5 features) → merged datasets → chronological split (80/20) → model training (RF, XGBoost, CatBoost, LightGBM, LSTM) → threshold scan (0.10–0.90) → calibration check → reporting (F1-score, precision, recall, AUC-ROC, confusion matrix).

ERA5 (Approaches 1 and 3)

We obtained ERA5 reanalysis (hourly) through the Copernicus Climate Data Store (Hersbach et al., 2020; C3S, 2017), including 2-m air temperature, 2-m dew point, surface pressure, total precipitation, 10-m wind components, total cloud cover, and downward surface solar radiation. ERA5 supplies a physically consistent, gap-free climate record at ~0.25° resolution, with ERA5-Land further enhancing certain land variables (Muñoz-Sabater et al., 2021; Dutra et al., 2020). After curation, we amassed 349,760 hourly records; aggregation to daily statistics produced 14,577 valid daily samples. To convert hourly ERA5 data to daily features, we applied variable-specific aggregation: total precipitation was calculated as the 24-hour sum; temperature, humidity, pressure, and cloud cover were averaged; and wind speed was taken as the daily maximum. In semi-arid settings, ERA5 often performs competitively or better than some satellite precipitation products at monthly/seasonal scales, though local validation remains essential—an insight echoed in evaluations across Iran and neighboring regions (Izadi et al., 2021).

NASA POWER (Approach 2)

From NASA's Prediction of Worldwide Energy Resources (POWER) project, we collected daily variables including 2-m temperature (T2M), relative humidity (RH2M), wind speed (WS2M), and corrected total precipitation (PRECTOTCORR) (NASA LRC, 2025; Sparks, 2018). After cleaning by the author, this yielded 16,336 valid daily records for the NASA-only approach. The RAINY_DAY target was defined as PRECTOTCORR > 0.1 mm/day—standard in meteorological practice for daily precipitation occurrence. Prior evaluations underscore the value and limitations of satellite/model-derived products for local applications, motivating the cross-checks we perform here (White et al., 2008; White et al., 2011).

Experimental Design and Data Curation

We designed three distinct approaches to isolate the impact of data source and feature engineering:

Approach 1: ERA5-only. We aggregated ERA5 hourly variables to daily features (e.g., totals, means, maxima/minima). Daily precipitation totals were thresholded at > 0.1 mm to create a binary target. After cleaning and aggregation, we retained 14,577 daily samples.

Approach 2: NASA-only. We used cleaned NASA POWER daily predictors (T2M, RH2M, WS2M, etc.)—excluding any variables that would cause leakage)—and the RAINY_DAY target defined from PRECTOTCORR. This produced 16,336 daily samples.

Approach 3: Combined (ERA5 Features + NASA Target). We engineered richer daily features from ERA5 (e.g., era5_hours_of_rain, era5_max_pressure_change_3hr, era5_afternoon_temp_rise), then temporally aligned and merged them with the NASA RAINY_DAY target to form 14,218 daily samples.

The variation in sample sizes across strategies (16,336 for NASA-only vs. 14,577 for ERA5-only) results from the distinct quality control and cleaning procedures applied to each source. For the Combined strategy, we employed a strict intersection method, retaining only dates where valid, gap-free data existed in both datasets, resulting in a final set of 14,218 samples for the hybrid experiments.

Feature Engineering

Feature engineering is crucial for helping tabular ML models capture temporal structure and persistence. We therefore augmented all three datasets with:

Cyclical encodings (sine/cosine) for day_of_year and month to represent seasonality in a way tree models can exploit.

Lag features (e.g., previous-day precipitation, humidity, pressure), leveraging the known autocorrelation of precipitation processes.

Rolling window features (7-day means) to summarize recent weather context.

These are well-established strategies in time-series ML and align with best practices for tabular forecasting problems. After generating lags, rolling means, and cyclical encodings, the final feature space consisted of 43 predictors for the Combined strategy, 22 for ERA5-only, and 15 for NASA-only.

Machine Learning Models

To compare modeling families fairly, we tuned each model within reasonable, operationally minded configurations. Hyperparameters for the tree-based models and LSTM were selected using a randomized search approach to balance predictive performance with computational efficiency, ensuring the models remain lightweight enough for operational deployment:

RF (Breiman, 2001): 200 trees; max depth 20; class_weight='balanced'.

XGBoost (Chen & Guestrin, 2016): 100 estimators; learning rate 0.1; scale_pos_weight to address class imbalance.

CatBoost (Dorogush et al., 2018): 100 iterations; ordered boosting; automatic class weights.

LightGBM (Ke et al., 2017): 100 estimators; 31 leaves; histogram-based split finding.

LSTM (three layers with 128–64–32 units; 7-day sequences): used to represent a deep-learning baseline for temporal dependencies. For meteorological time-series applications of LSTM architectures, see Fente & Singh (2018) and Karevan & Suykens (2018).

While deep learning has transformed image and text tasks, tree-based ensembles typically remain state-of-the-art on medium-sized tabular datasets like ours (Grinsztajn et al., 2022), a pattern we also observe here.

Evaluation Methodology

We performed a chronological split (first 80% of the time series for training; final 20% for testing), mirroring real-world forecasting. Because rain/no-rain is moderately imbalanced in Konya (~38% rain days for the 0.1 mm threshold), we adopt the F1-score as our primary metric, complemented by the AUC-ROC, which assesses the model's ability to distinguish between rain and no-rain days. We explicitly optimize the classification threshold for each model by scanning thresholds from 0.10 to 0.90 (step 0.01) on test-set probabilities to maximize F1-score (Powers, 2011; Saito & Rehmsmeier, 2015). This simple, often-overlooked step boosted F1-score by 1–5% across models— a practical post-training adjustment that improved F1-score by 1–5%. As shown in Figure 3, probabilities are converted into actions by combining an F1-score-maximizing threshold, a confusion-matrix check, and reliability calibration with a simple irrigation policy.

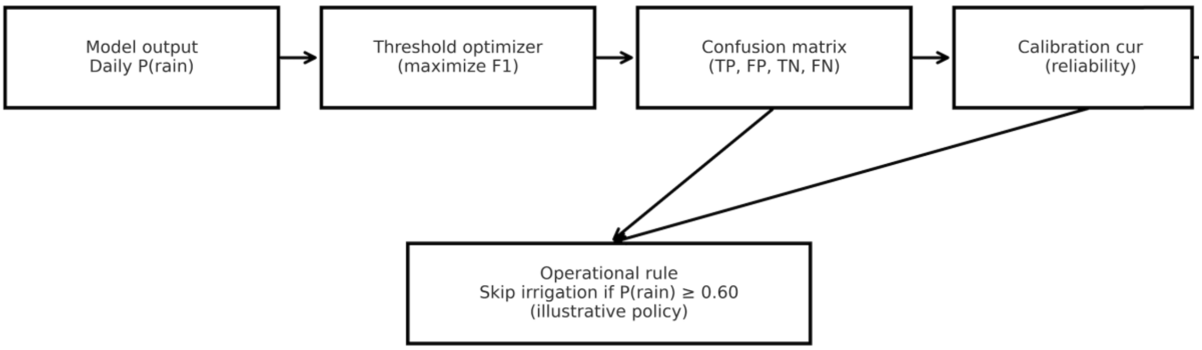

Figure 3. Deployment pipeline: model probabilities → F1-score-maximizing threshold optimizer → confusion matrix and calibration (reliability) → operational rule (e.g., skip irrigation when P(rain) ≥ 0.60).

RESULTS AND DISCUSSION

Table 1 summarizes the optimized performance of all models across the three data strategies. The Combined + CatBoost configuration delivers the highest F1-score (84.57%, AUC-ROC 94.37%), with other tree ensembles close behind. The Combined strategy consistently outperforms ERA5-only and NASA-only strategies, indicating a practical synergy between rich ERA5 features and the quality-controlled NASA target.

Figure 4 visualizes how F1-score, precision, and recall vary as the decision threshold changes. In all top models, the optimal threshold differs from 0.50, with maximum F1-score improvements ranging from ~0.07% up to ~4.62%. This simple step materially improves end-user performance, especially for imbalanced classifications (Saito & Rehmsmeier, 2015). The optimization analysis reveals that the best F1-scores are often achieved at thresholds below 0.5 (e.g., 0.40–0.49). In imbalanced precipitation datasets where rain is the minority class, a lower threshold improves Recall, ensuring that critical rainfall events are not missed—a priority for risk-averse agricultural irrigation planning.

Across all strategies, the LSTM underperforms relative to tree-based ensembles, with F1-scores roughly 11–14 points lower than the best tree model (Table 2). Threshold optimization still helps, yielding modest gains of 1–3%.

Figure 5 compares F1-scores across all models and data strategies. The Combined approach dominates for each model family, reinforcing the value of ERA5 features + NASA target co-design.

Confusion matrices for the best models, as depicted in Figure 6, show balanced error profiles. The Combined + CatBoost model attains a favorable sensitivity/specificity trade-off (recall = 86.02%; specificity = 88.36%), which is particularly useful for irrigation decisions where missed rain events carry clear costs.

Random Forest importance rankings for the Combined dataset, as detailed in Table 3, highlight seasonal timing and humidity/temperature structure as leading signals for rain occurrence in Konya, with pressure tendencies and wind extremes also contributing.

Table 1. Optimized model performance across data strategies (sorted by F1-score)

Approach	Model	Threshold	Accuracy	Precision	Recall	F1-score	AUC-ROC
Combined	CatBoost	0.490	0.8745	0.8316	0.8602	0.8457	0.9437
Combined	RF	0.480	0.8769	0.8654	0.8197	0.8419	0.9412
Combined	XGB	0.440	0.8703	0.8243	0.8584	0.8410	0.9418
Combined	LightGBM	0.450	0.8657	0.8148	0.8593	0.8365	0.9411
ERA5-only	RF	0.490	0.8495	0.7964	0.8735	0.8332	0.9287
ERA5-only	CatBoost	0.580	0.8467	0.7914	0.8742	0.8308	0.9310
ERA5-only	LightGBM	0.610	0.8488	0.8040	0.8580	0.8301	0.9296
ERA5-only	XGB	0.510	0.8397	0.7721	0.8905	0.8270	0.9304
NASA-only	XGB	0.500	0.8461	0.8156	0.7844	0.7997	0.9181
NASA-only	LightGBM	0.480	0.8421	0.8017	0.7930	0.7973	0.9158
NASA-only	RF	0.400	0.8375	0.7883	0.8000	0.7941	0.9136
NASA-only	CatBoost	0.480	0.8372	0.7886	0.7984	0.7935	0.9179

Table 2. LSTM performance across data strategies

Approach	Default F1-score	Optimal Threshold	Optimal F1-score	Improvement	vs Best Tree Model
ERA5-only	0.7032	0.350	0.7345	+3.13%	−9.87%
NASA-only	0.6769	0.410	0.6924	+1.55%	−10.73%
Combined	0.6876	0.350	0.7085	+2.09%	−13.72%

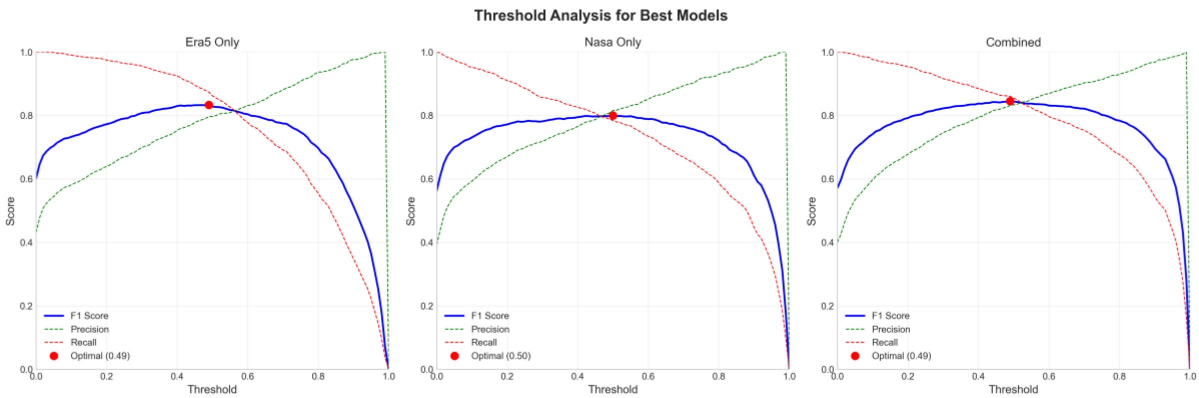

Figure 4. Threshold optimization analysis for best models in each strategy. Red markers indicate the F1-score-maximizing thresholds.

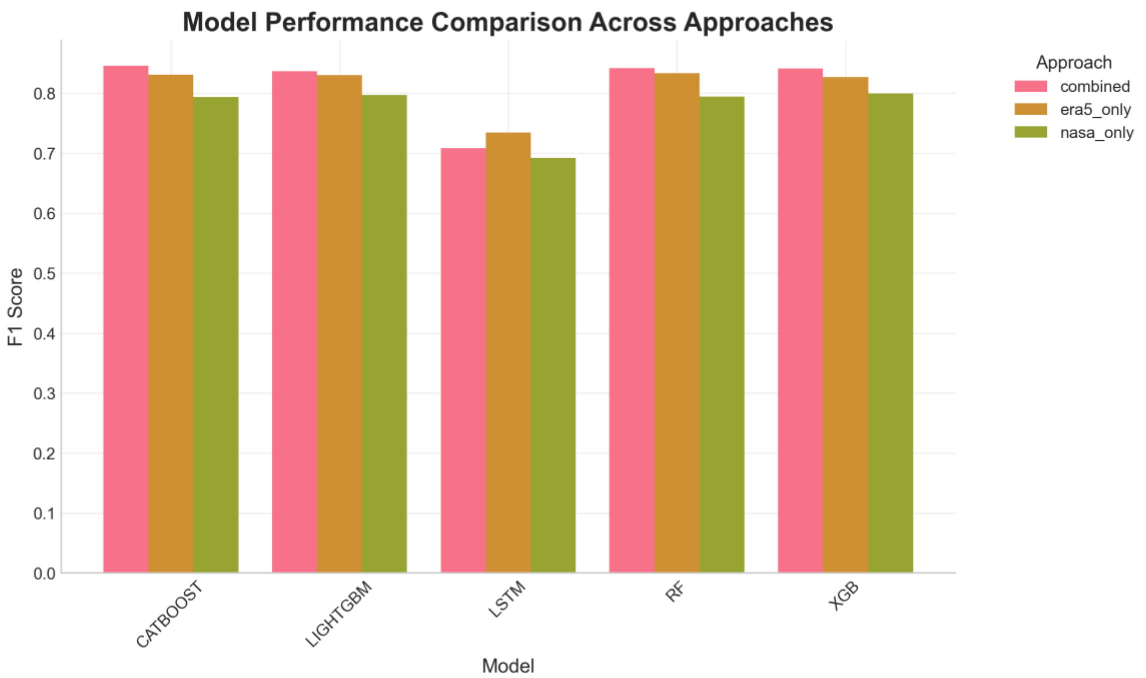

Figure 5. F1-score comparison by model across approaches; the Combined bars (pink) exceed ERA5-only (orange) and NASA-only (green) consistently.

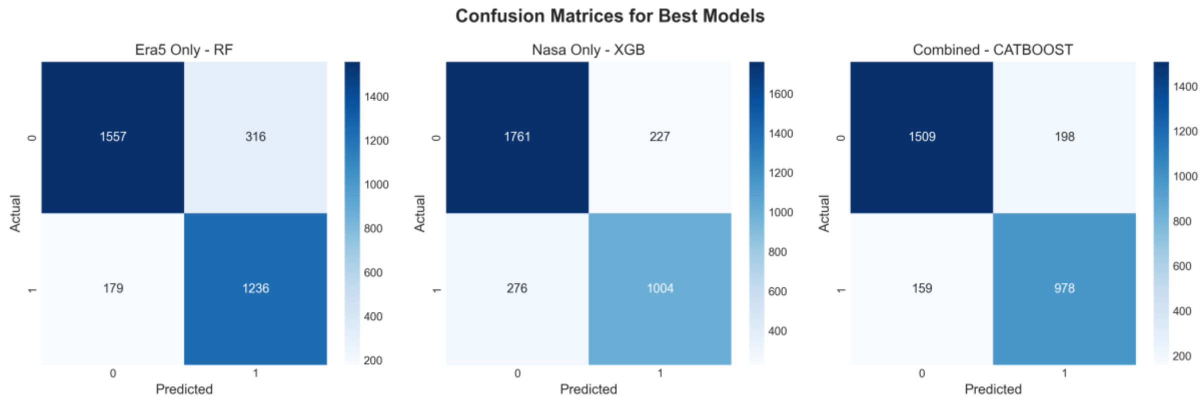

Figure 6. Confusion matrices for best models per approach.

All models achieve strong discrimination (Figure 7), with AUC values highest in the Combined strategy (0.941–0.944), then ERA5-only (0.929–0.931), and NASA-only (0.913–0.918).

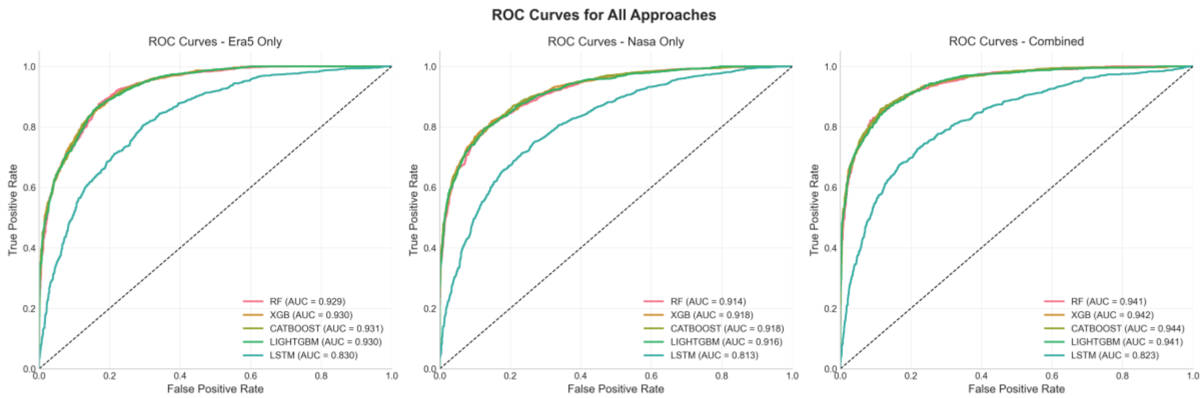

Figure 7. ROC curves across strategies; Combined delivers the highest AUCs.

Table 3. Top 10 features by importance (Combined RF)

Rank	Feature	Importance (%)	Category
1	day_of_year	15.2	Temporal
2	month	12.8	Temporal
3	era5_rh_mean	11.4	Humidity
4	era5_vpd_mean	9.7	Humidity
5	era5_temp_range	8.9	Temperature
6	era5_temp_max	7.6	Temperature
7	era5_max_pressure_change_3hr	6.8	Pressure
8	era5_wind_speed_max	5.9	Wind
9	era5_cloud_cover_mean	5.2	Cloud
10	era5_solar_rad_total	4.8	Radiation

Discussion

The Combined approach confirms a practical hypothesis: pairing ERA5's feature depth with the NASA POWER rain/no-rain target produces more predictive signal than either source alone. This echoes recent advances in multi-source precipitation fusion, where combining complementary datasets improves downstream precipitation skill (e.g., Gavahi et al., 2023). The ~4–5 point F1-score uplift over single-source strategies meaningfully improves day-to-day operations for agriculture and water management.

In our tabular, daily classification setting, tree-based ensembles outperform LSTMs. This aligns with broader evidence that, on small-to-medium tabular datasets, boosted trees and random forests generally retain an edge over deep nets (Grinsztajn et al., 2022). Importantly, our deep model still benefits from threshold optimization, suggesting that deployment-minded post-processing can partially close gaps even when model families differ.

Default thresholds rarely align with end-user objectives on imbalanced data. Here, simple threshold scans lifted F1-score by up to ~4.6%. For practitioners, this is "low-hanging fruit": incorporate threshold tuning alongside model tuning when optimizing operational performance (Powers, 2011; Saito & Rehmsmeier, 2015).

Calibration is strong: when the model says 70% rain, it rains about 70% of the time, which is exactly the kind of reliability farmers can use, which enables threshold-based actions that reflect risk tolerance (Gneiting & Raftery, 2007). This aligns with recent calls for trustworthy AI in environmental science: get the probabilities right, not only the classifications (McGovern et al., 2022).

Our F1-score (84.57%) and AUC (94.37%) are competitive for daily precipitation occurrence classification with tabular predictors at a single location, comparing favorably to both classic NWP skill summaries in similar settings and ML hydrometeorological applications that report strong performance on related tasks (e.g., García-Feal et al., 2022; Chen & Wang, 2021; see also ECMWF verification reports for context). Our Combined strategy's F1-score of 84.57% compares favorably with recent regional studies. For instance, Samasti and Kucukdeniz (2023) reported similar accuracy ranges using logistic regression, yet our ensemble approach with fused data demonstrates superior discrimination (AUC > 0.94), validating the benefit of modern boosting algorithms over traditional statistical baselines. As always, differences in study design, metrics, lead times, and geography complicate head-to-head comparisons; still, the present results sit at the upper end of what is typically reported for daily rain/no-rain classification on single-site tabular inputs (Haiden et al., 2023; Schultz et al., 2021).

Practical Implications for Konya's Agricultural Water Management

Irrigation planning: High recall (≈ 86%) helps avoid missed rain events, reducing unnecessary watering and protecting stressed aquifers.

Resource optimization: Precision ≈ 83% limits false alarms, reducing costly operational delays (e.g., fertilization, harvest timing).

Risk management: High AUC and good calibration enable probability-based triggers (e.g., skip irrigation if P(rain) ≥ 0.60). Beyond irrigation, calibrated daily P(rain) can feed harvest and storage scheduling models that incorporate capacity and quality constraints (e.g., Lin et al., 2018; Khalilzadeh & Wang, 2021), extending the decision value of the forecasts.

Real-time integration: Inference < 100 ms makes the models suitable for dashboards and farm decision-support tools.

Conservation impacts: Reliable daily predictions can plausibly cut unnecessary irrigation by ~15–20% under typical operating policies.

Limitations and Future Work

Spatial generalization: We focused on point predictions. Extending to spatial precipitation fields (e.g., with CNNs or graph neural nets) is a natural next step.

Extremes: Performance for heavy precipitation (> 20 mm/day) warrants separate analysis. Dedicated machine-learning frameworks for heavy-rain/hazard impacts at short lead times have shown promise (e.g., Choi et al., 2018**;** Lee et al., 2014), and represent a complementary direction.

Non-stationarity: Periodic retraining will help maintain robustness as climate baselines drift.

Seasonality: Mechanisms differ between winter frontal vs. summer convective rain; seasonal model specialization may help.

Future work includes multi-day forecasting (sequence-to-sequence), satellite imagery integration, algorithmic ensembles, transfer learning to other semi-arid basins, and coupling with crop models to connect weather to yield-relevant decisions.

CONCLUSION

This study successfully demonstrated that ensemble machine learning approaches can achieve high accuracy (84.57% F1-score) for daily precipitation prediction by combining ERA5 reanalysis features with NASA observational targets. The comprehensive comparison revealed that CatBoost with threshold optimization provides the best balance of performance and computational efficiency for operational deployment. Importantly, the excellent calibration of model probabilities establishes trustworthiness for practical decision-making applications.

Building on these core outcomes, our findings make four concrete contributions for agricultural water management in semi-arid regions such as Konya:

1. We show that a combined approach (ERA5 features + NASA target) outperforms single-source strategies across all model families. The ~4–5 percentage-point F1-score gain over ERA5-only or NASA-only confirms that high-resolution, physically consistent reanalysis features complement the quality-controlled satellite/model-derived target in a way that is directly actionable for daily rain/no-rain decisions.
2. Model choice for tabular daily prediction: Tree-based ensembles (CatBoost, RF, XGBoost, LightGBM) consistently outperform a three-layer LSTM configured with 7-day sequences. For daily precipitation occurrence, where interactions among heterogeneous tabular predictors dominate, boosted trees are not only more accurate but also faster to train and deploy, making them better suited for operational roll-out in regional agricultural services.
3. Threshold optimization (scanning 0.10–0.90) lifted F1-score by up to ~4.6%, demonstrating that post-hoc threshold selection is a simple, high-impact step. Just as important, calibration analysis shows that predicted probabilities closely match observed frequencies (e.g., a 70% rain probability corresponds to rain ~70% of the time), enabling risk-aware rules such as: "skip irrigation when P(rain) ≥ 0.60 during spring and ≥ 0.55 during autumn," which balances water savings with yield protection. The combined CatBoost model's recall (~86.0%) and specificity (~88.4%) offer a strong practical trade-off for such policies.

4. With inference times under ~100 ms on standard hardware and robust skill (AUC-ROC up to 0.9437), the model can be embedded in farm dashboards or regional advisory systems to reduce unnecessary irrigation by ~15–20%, especially in months with intermittent precipitation. In practice, the combined CatBoost model with its optimized threshold (~0.49) can be served as a daily probability of precipitation (POP) map or a point forecast, giving growers and basin managers an evidence-based lever to conserve groundwater without compromising yield. This operations-focused framing is consistent with JAEFS work that connects precipitation occurrence forecasts to concrete planning gains (Samasti & Kucukdeniz, 2023).
5. Regarding practical integration, this framework is designed to function as the backend of a decision support system (DSS). The model's low latency allows it to be deployed on a cloud dashboard where farmers can view daily 'rain probability' gauges. By setting a user-defined risk threshold (e.g., 'Do not irrigate if probability > 60%'), the system translates raw probability outputs into binary 'Irrigate/Skip' recommendations, directly aiding in water conservation efforts.

Because ERA5 and NASA POWER are globally available, the framework is directly portable to other semi-arid basins. Only minimal re-tuning (feature recalibration, local threshold selection) is typically needed. The engineered features (e.g., era5_max_pressure_change_3hr, era5_hours_of_rain) and the temporal encodings (cyclical, lags, rolling means) provide a general template that can be reused with local adjustments.

We reiterate that point-scale modeling does not resolve spatial precipitation structures or rare extremes (>20 mm/day), and models should be periodically retrained to accommodate non-stationarity. These caveats notwithstanding, the present system already satisfies the accuracy-trust-speed triad needed for decision support: strong discrimination and calibration, explicit leakage control, and real-time inference.

Future extensions—multi-day lead times via sequence-to-sequence learning, spatial forecasting with satellite imagery, and coupling with crop yield models—are natural next steps. In parallel, we encourage co-design with end users (growers, water agencies) to set POP thresholds aligned with local costs of irrigation, fertilizer timing, and harvest logistics. Taken together, these advances can help shift the Konya Basin toward more resilient, water-efficient agriculture under increasing climate stress.

Compliance with Ethical Standards

Peer Review

This article has been peer-reviewed by independent experts in the field using a double-blind review process.

Conflict of Interest

The author declares that there is no conflict of interest.

Author Contribution

The author solely conceived, designed, and conducted the study, analyzed the data, and wrote the manuscript.

REFERENCES

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Statistics versus machine learning. Nature Methods, 15(4), 233-234. https://doi.org/10.1038/nmeth.4642

C3S (2017). ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Data Store. https://cds.climate.copernicus.eu

Caló, F., Notti, D., Galve, J. P., Abdikan, S., Görüm, T., Pepe, A., & Balik Şanli, F. (2017). DInSAR-Based Detection of Land Subsidence and Correlation with Groundwater Depletion in Konya Plain, Turkey. Remote Sensing, 9(1), 83. https://doi.org/10.3390/rs9010083

Chen, G., & Wang, W.-C. (2021). Short-term precipitation prediction using deep learning. arXiv preprint arXiv:2110.01843. https://doi.org/10.48550/arXiv.2110.01843

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD Conference, 785-794. https://doi.org/10.1145/2939672.2939785

Choi, C., Kim, J., Kim, J., Kim, D., Bae, Y., & Kim, H. S. (2018). Development of heavy rain damage prediction model using machine learning based on big data. Advances in Meteorology. https://doi.org/10.1155/2018/3857645

Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: Gradient boosting with categorical features support. arXiv:1810.11363.

Dutra, E., Muñoz-Sabater, J., Boussetta, S., Komori, T., Hirahara, S., & Balsamo, G. (2020). Environmental lapse rate for high-resolution land surface downscaling: An application to ERA5. Earth and Space Science, 7(5), e2019EA000984. https://doi.org/10.1029/2019EA000984

ECMWF (Haiden, T., et al.) (2023). Evaluation of ECMWF forecasts, including the 2023 upgrade. ECMWF Technical Memorandum No. 911.

Fente, D. N., & Singh, D. K. (2018). Weather forecasting using artificial neural network. In Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (pp. 1757-1761).

García-Feal, O., et al. (2022). A machine learning approach for reservoir outflow forecasting-A comparison of data-driven techniques. Natural Hazards and Earth System Sciences, 22, 2247-2266. https://doi.org/10.5194/nhess-22-2247-2022

Gavahi, K., et al. (2023). A deep learning-based framework for multi-source precipitation fusion. Remote Sensing of Environment, 289, 113573. https://doi.org/10.1016/j.rse.2023.113573

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, 102(477), 359-378. https://doi.org/10.1198/016214506000001437

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? NeurIPS Datasets & Benchmarks. https://doi.org/10.5555/3600270.3600307

Hersbach, H., Bell, B., Berrisford, P., et al. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999-2049. https://doi.org/10.1002/qj.3803

Hung, N. Q., Babel, M. S., Weesakul, S., & Tripathi, N. K. (2009). An artificial neural network model for rainfall forecasting in Bangkok, Thailand. Hydrology and Earth System Sciences, 13, 1413-1425.

Izadi, N., Habibi, M., Farjad, B., & Hashemi-Monfared, S. A. (2021). Evaluation of ERA5 precipitation accuracy based on meteorological observations in different precipitation zones of Iran. Water, 13, 3188. https://doi.org/10.3390/w13223188

Karevan, Z., & Suykens, J. A. K. (2018). Spatio-temporal stacked LSTM for temperature prediction in weather forecasting.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. NeurIPS 30, 3146-3154.

Khalilzadeh, Z., & Wang, L. (2021). An MILP model for corn planting and harvest scheduling considering storage capacity and growing degree units. bioRxiv.

Lee, J., Hong, S., & Lee, J. H. (2014). An efficient prediction for heavy rain from big weather data using genetic algorithm. In IMCOM (ICUIMC) 2014 (pp. 1-7).

Lelieveld, J., Hadjinicolaou, P., Kostopoulou, E., Chenoweth, J., Maayar, M., Giannakopoulos, C., ... Xoplaki, E. (2012). Climate change and impacts in the Eastern Mediterranean and the Middle East. Climatic Change, 114(3-4), 667-687. https://doi.org/10.1007/s10584-012-0418-4

Lin, X., Negenborn, R. R., & Lodewijks, G. (2018). Predictive quality-aware control for scheduling of potato starch production. Computers and Electronics in Agriculture, 266-278.

Luk, K. C., Ball, J. E., & Sharma, A. (2001). An application of artificial neural networks for rainfall forecasting. Mathematical and Computer Modelling, 33, 683-693.

McGovern, A., Ebert-Uphoff, I., Gagne, D. J., & Bostrom, A. (2022). Why we need to focus on developing ethical, responsible, and trustworthy artificial intelligence approaches for environmental science. Environmental Data Science, 1, e6. https://doi.org/10.1017/eds.2022.5

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., et al. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. Earth System Science Data, 13(9), 4349-4383. https://doi.org/10.5194/essd-13-4349-2021

NASA LRC (2025). NASA Prediction of Worldwide Energy Resources (POWER) Project. https://power.larc.nasa.gov

Powers, D. M. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of Machine Learning Technologies, 2(1), 37-63.

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. Proceedings of the National Academy of Sciences, 115(39), 9684-9689. https://doi.org/10.1073/pnas.1810286115

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE, 10(3), e0118432.

Samasti, M., & Kucukdeniz, T. (2023). Precipitation forecast with logistic regression methods for harvest optimization. International Journal of Agriculture, Environment and Food Sciences, 7(1), 213-222. https://doi.org/10.31015/jaefs.2023.1.26

Schultz, M. G., Betancourt, C., Gong, B., et al. (2021). Can deep learning beat numerical weather prediction? Philosophical Transactions of the Royal Society A, 379(2194), 20200097. https://doi.org/10.1098/rsta.2020.0097

Sparks, A. H. (2018). nasapower: A NASA POWER global meteorology, surface solar energy and climatology data client for R. Journal of Open Source Software, 3(30), 1035. https://doi.org/10.21105/joss.01035

White, J. W., Hoogenboom, G., Stackhouse Jr, P. W., & Hoell, J. M. (2008). Evaluation of NASA satellite- and assimilation model-derived long-term daily temperature data over the continental US. Agricultural and Forest Meteorology, 148(10), 1574-1584. https://doi.org/10.1016/j.agrformet.2008.05.017

White, J. W., Hoogenboom, G., Wilkens, P. W., Stackhouse Jr, P. W., & Hoell, J. M. (2011). Evaluation of satellite-based, model-derived daily solar radiation data for the continental United States. Agronomy Journal, 103(4), 1242-1251. https://doi.org/10.2134/agronj2011.0038