Flood Prediction — Feature Engineering & Hybrid Stacking Report

Dhinna Tretarnthip · Caio Appolinario  
2025-08-12 · Engineering

# 1. Methodology Overview

This document summarizes the engineer-focused pipeline using per-station models with feature engineering and hybrid stacking to predict hourly flood risk.

## Key Settings

• Flood label: Per-station 82nd percentile of discharge

• Lead time: t+2 hours

• Split: 70/30 stratified (per station)

• Thresholds: D08A071=0.50, D08A084=0.533, D08A115=0.932

• Objective: Prioritize catching floods (recall) over false alarms (precision)

# 2. Overall Results

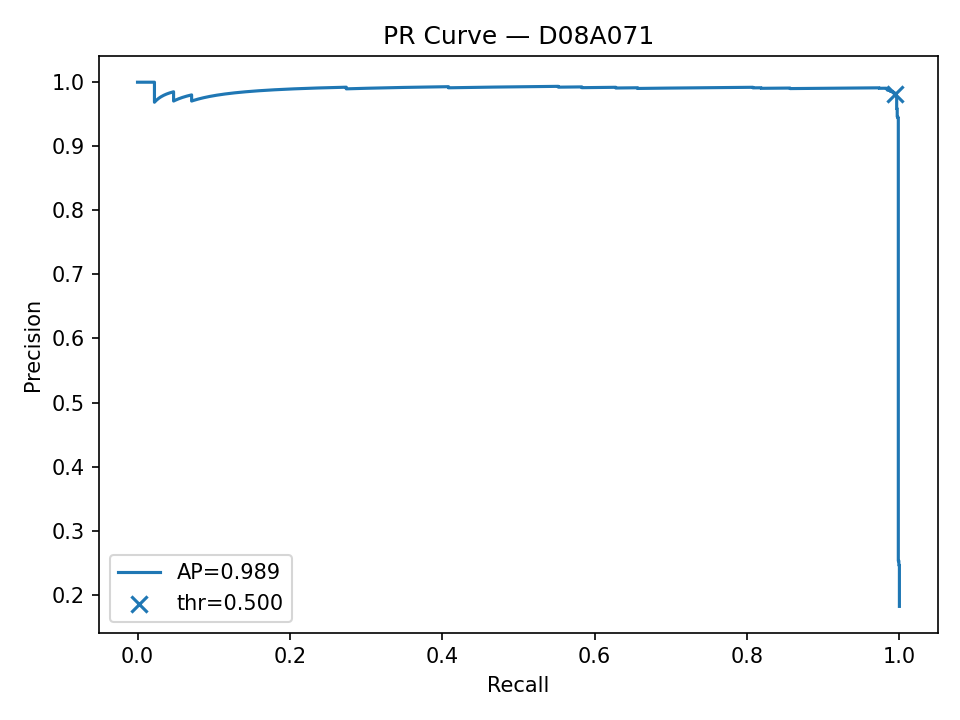
Overall (fixed thresholds): Acc=0.993, Prec=0.967, Rec=0.988

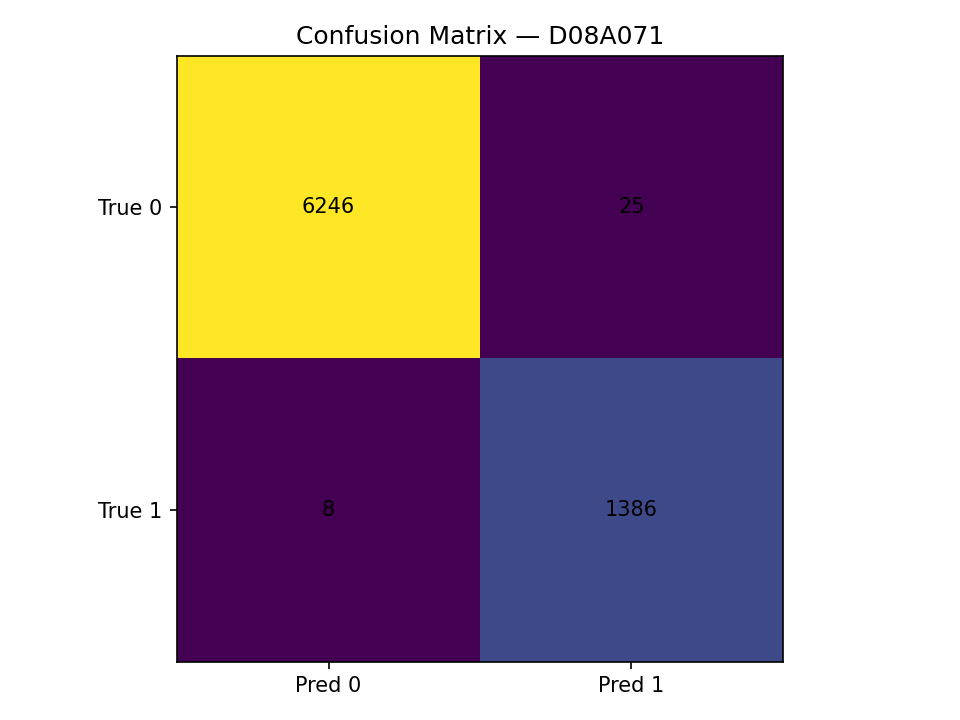
|  |  |  |
| --- | --- | --- |
|  | Pred 0 | Pred 1 |
| True 0 | 20176 | 121 |
| True 1 | 41 | 3499 |

# 3. Station D08A071

## PR Curve & Confusion Matrix

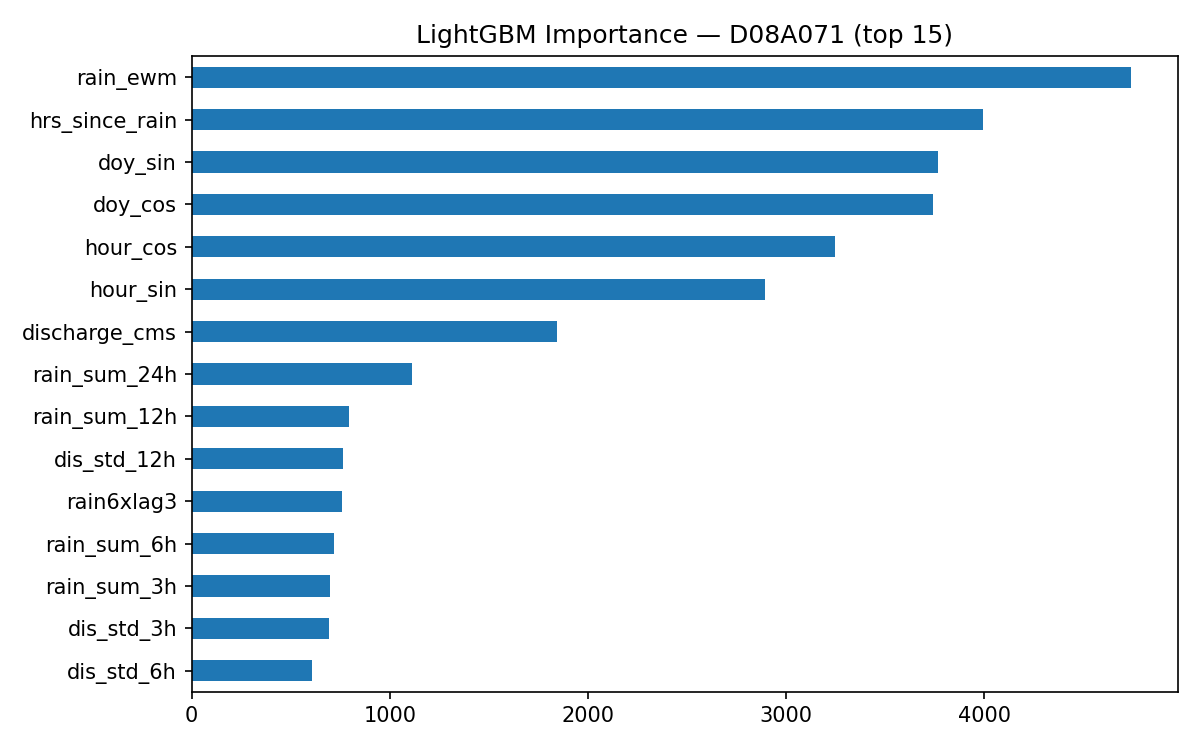
Precision–Recall curve with the chosen operating threshold.





## Feature Importance (LightGBM) & Notes

Top 15 features by LightGBM impurity importance.



1) Recent discharge lags dominate (e.g., dis\_lag\_4h, dis\_lag\_1h, dis\_lag\_6h).

2) Short-horizon rainfall sums contribute (rain\_sum\_6h, rain\_sum\_3h); EWMA captures persistence.

3) Rate-of-change features (dis\_rate\_3h, dis\_rate\_1h) help detect surges.

4) Range/variability (dis\_std\_12h, dis\_std\_3h) adds volatility context.

5) Mild seasonal effects present (doy\_sin, doy\_cos).

## Ablation (XGB → +Stack → +LGBM → +LSTM)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| variant | ap | recall | precision | acc | FP | FN |
| XGB only | 0.996 | 0.993 | 0.986 | 0.996 | 20.000 | 10.000 |
| +Stack (no LGBM/LSTM) | 0.991 | 0.994 | 0.983 | 0.996 | 24.000 | 9.000 |
| +LGBM | 0.991 | 0.994 | 0.983 | 0.996 | 24.000 | 8.000 |
| +LSTM | 0.989 | 0.994 | 0.982 | 0.996 | 25.000 | 8.000 |

## Error Analysis (Top FP / Top FN)

Top False Positives (highest predicted probability):

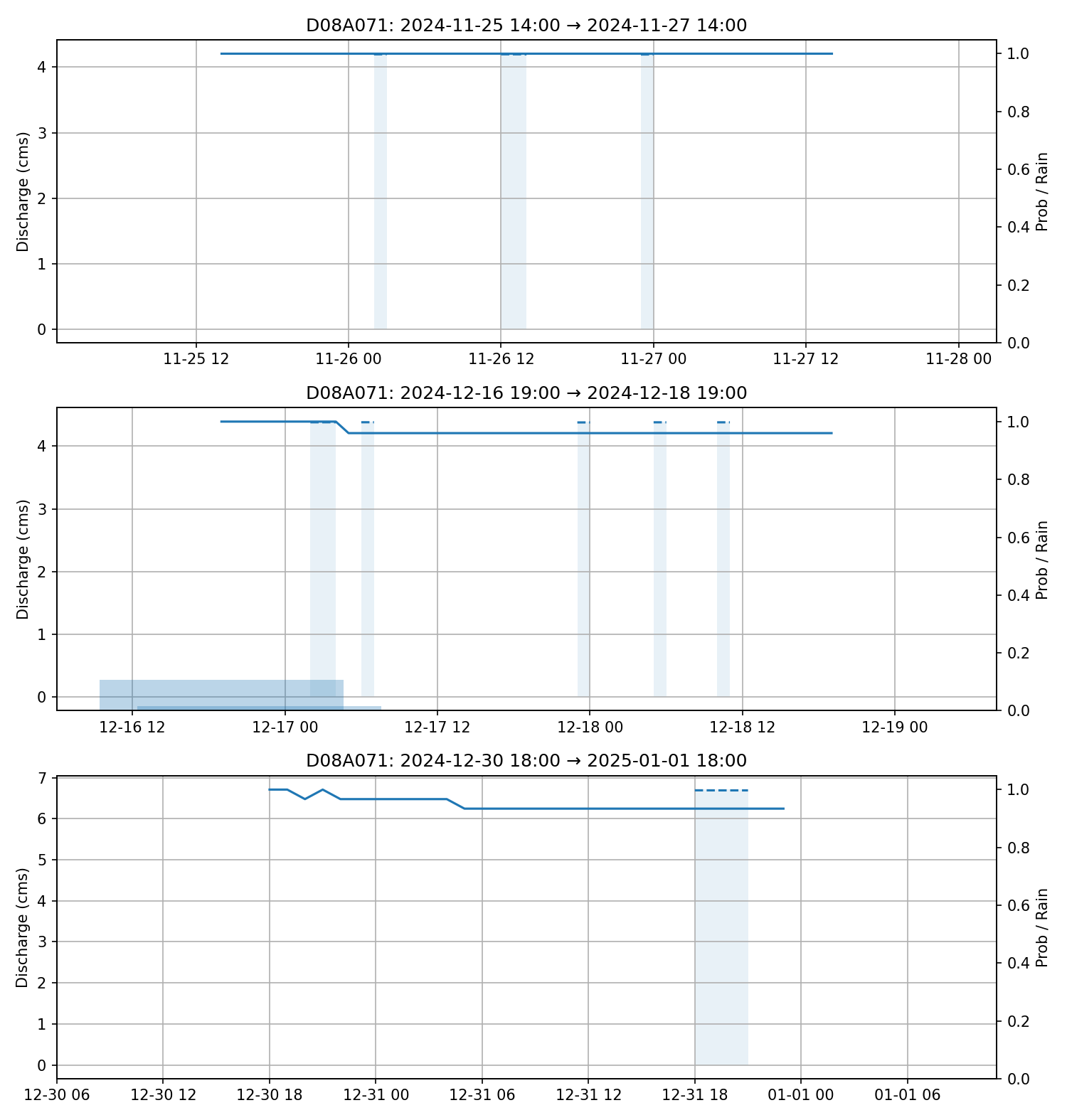
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| proba | y\_true | pred | discharge\_cms | rain\_mm | rain\_sum\_12h | dis\_lag\_1h | dis\_rate\_1h |
| 1.000 | 0.000 | 1.000 | 2.176 | 0.000 | 0.000 | 2.176 | 0.000 |
| 1.000 | 0.000 | 1.000 | 3.380 | 0.000 | 92.400 | 3.044 | 0.336 |
| 0.999 | 0.000 | 1.000 | 3.380 | 0.000 | 16.200 | 3.380 | 0.000 |
| 0.999 | 0.000 | 1.000 | 3.380 | 0.000 | 5.000 | 3.380 | 0.000 |
| 0.999 | 0.000 | 1.000 | 4.392 | 0.893 | 7.156 | 4.392 | 0.000 |
|  |  |  |  |  |  |  |  |

Top False Negatives (lowest predicted probability):

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| proba | y\_true | pred | discharge\_cms | rain\_mm | rain\_sum\_12h | dis\_lag\_1h | dis\_rate\_1h |
| 0.002 | 1.000 | 0.000 | 0.389 | 0.000 | 0.000 | 0.389 | 0.000 |
| 0.002 | 1.000 | 0.000 | 0.389 | 0.000 | 0.000 | 0.389 | 0.000 |
| 0.011 | 1.000 | 0.000 | 1.910 | 0.400 | 24.600 | 1.910 | 0.000 |
| 0.011 | 1.000 | 0.000 | 3.090 | 0.000 | 0.000 | 3.262 | -0.172 |
| 0.019 | 1.000 | 0.000 | 2.312 | 1.200 | 22.000 | 1.910 | 0.402 |
|  |  |  |  |  |  |  |  |

## Timeline Panels (Recent Events)

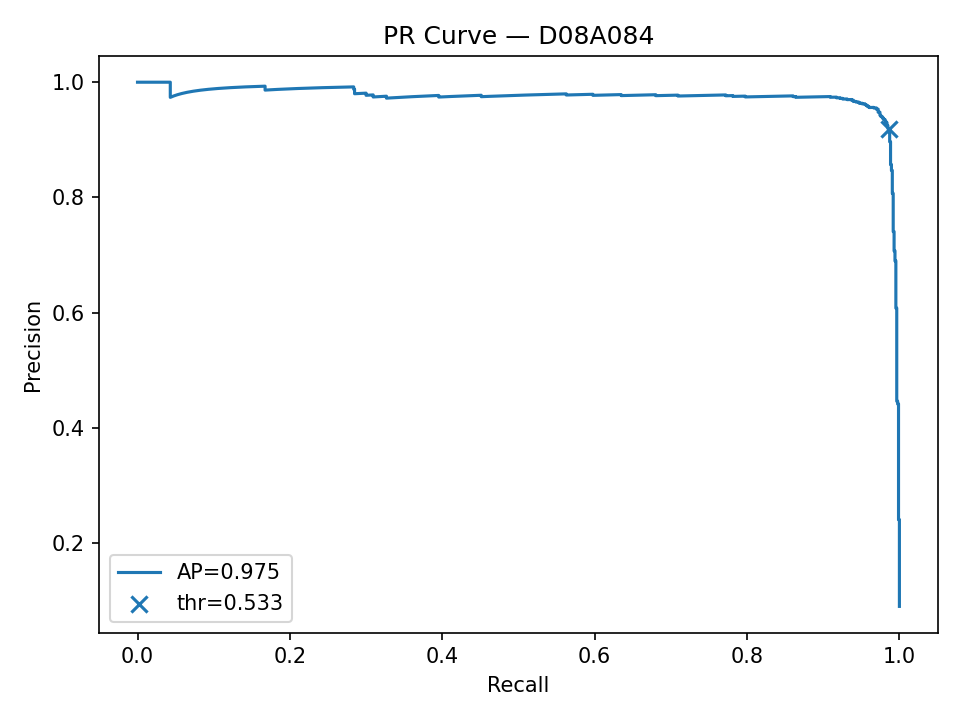
Discharge (cms), rain (mm), and predicted probability for recent events.

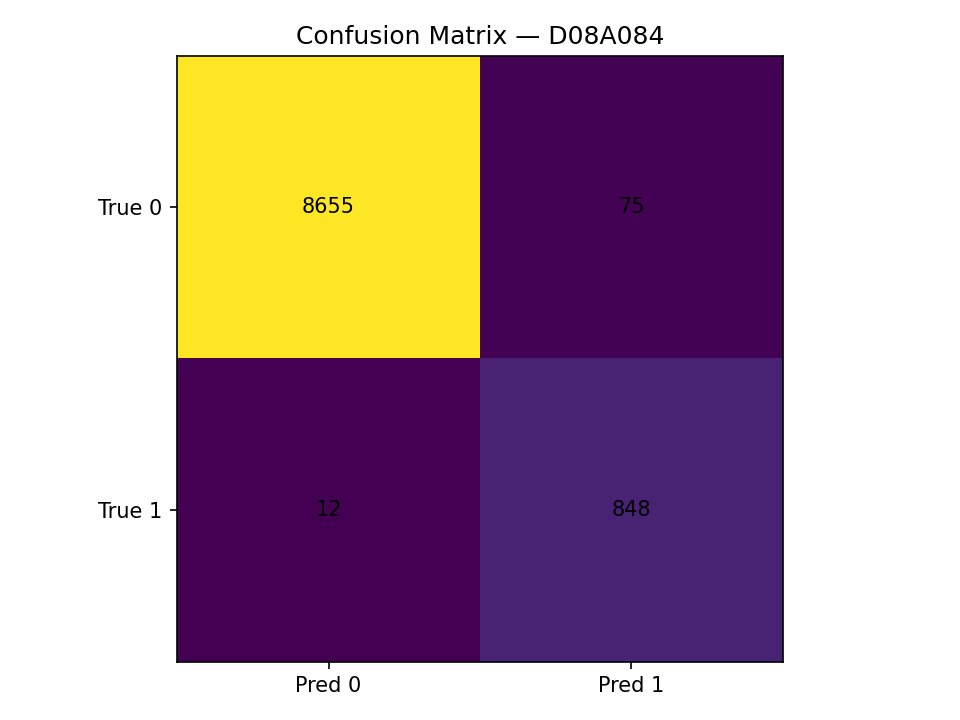


# 3. Station D08A084

## PR Curve & Confusion Matrix

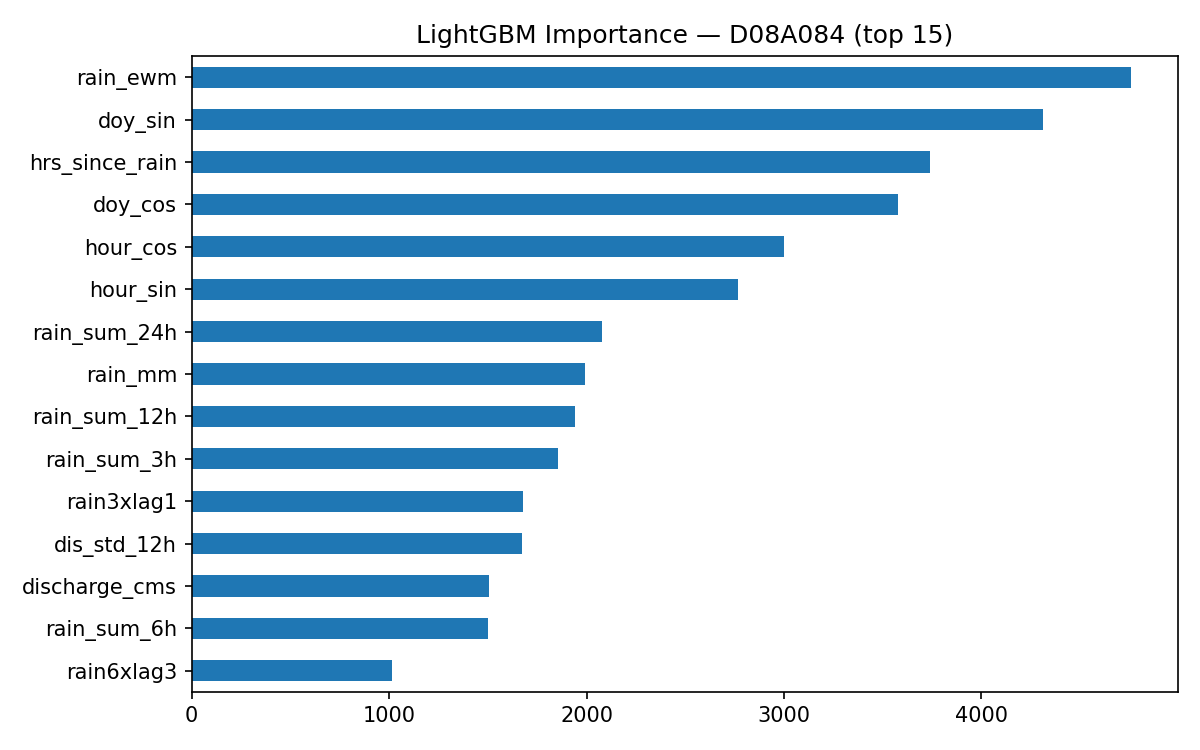
Precision–Recall curve with the chosen operating threshold.





## Feature Importance (LightGBM) & Notes

Top 15 features by LightGBM impurity importance.



1) Recent discharge lags dominate (e.g., dis\_lag\_12h, dis\_lag\_11h, dis\_lag\_7h).

2) Short-horizon rainfall sums contribute (rain\_sum\_3h, rain\_sum\_6h); EWMA captures persistence.

3) Rate-of-change features (dis\_rate\_3h, dis\_rate\_1h) help detect surges.

4) Range/variability (dis\_std\_12h, dis\_std\_6h) adds volatility context.

5) Mild seasonal effects present (doy\_sin, doy\_cos).

## Ablation (XGB → +Stack → +LGBM → +LSTM)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| variant | ap | recall | precision | acc | FP | FN |
| XGB only | 0.994 | 0.966 | 0.965 | 0.994 | 30.000 | 29.000 |
| +Stack (no LGBM/LSTM) | 0.975 | 0.985 | 0.918 | 0.991 | 76.000 | 13.000 |
| +LGBM | 0.974 | 0.985 | 0.919 | 0.991 | 75.000 | 13.000 |
| +LSTM | 0.975 | 0.986 | 0.919 | 0.991 | 75.000 | 12.000 |

## Error Analysis (Top FP / Top FN)

Top False Positives (highest predicted probability):

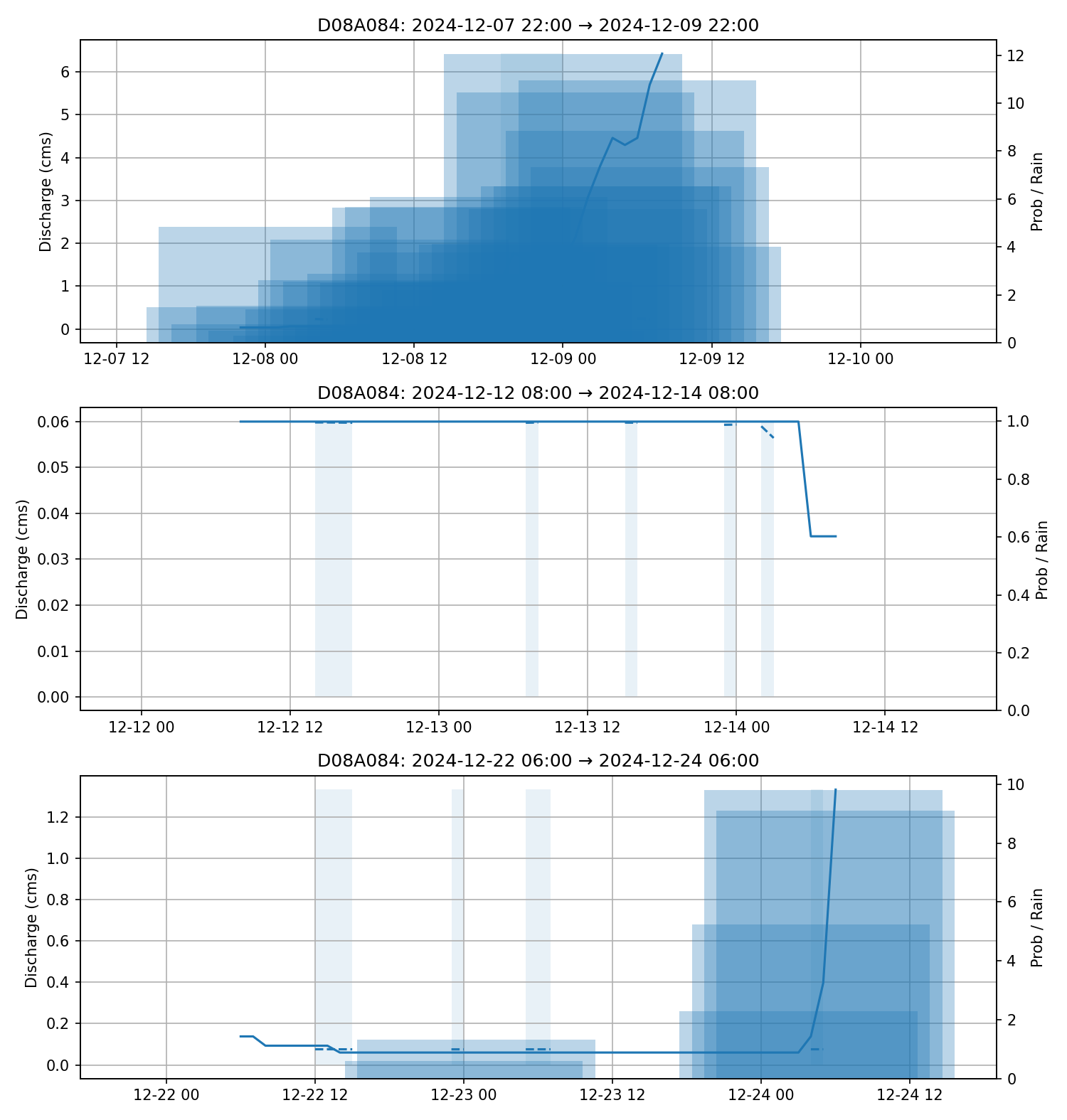
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| proba | y\_true | pred | discharge\_cms | rain\_mm | rain\_sum\_12h | dis\_lag\_1h | dis\_rate\_1h |
| 1.000 | 0.000 | 1.000 | 0.232 | 0.000 | 0.000 | 0.232 | 0.000 |
| 1.000 | 0.000 | 1.000 | 0.006 | 4.582 | 43.894 | 0.006 | 0.000 |
| 0.999 | 0.000 | 1.000 | 0.093 | 0.200 | 0.286 | 0.093 | 0.000 |
| 0.999 | 0.000 | 1.000 | 0.006 | 25.238 | 39.311 | 0.006 | 0.000 |
| 0.999 | 0.000 | 1.000 | 0.140 | 2.686 | 6.906 | 0.140 | 0.000 |
|  |  |  |  |  |  |  |  |

Top False Negatives (lowest predicted probability):

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| proba | y\_true | pred | discharge\_cms | rain\_mm | rain\_sum\_12h | dis\_lag\_1h | dis\_rate\_1h |
| 0.013 | 1.000 | 0.000 | 0.015 | 0.000 | 0.054 | 0.041 | -0.026 |
| 0.029 | 1.000 | 0.000 | 0.020 | 0.000 | 4.282 | 0.020 | 0.000 |
| 0.029 | 1.000 | 0.000 | 0.020 | 0.000 | 4.282 | 0.020 | 0.000 |
| 0.042 | 1.000 | 0.000 | 0.000 | 0.363 | 5.075 | 0.000 | 0.000 |
| 0.050 | 1.000 | 0.000 | 0.020 | 0.805 | 5.056 | 0.020 | 0.000 |
|  |  |  |  |  |  |  |  |

## Timeline Panels (Recent Events)

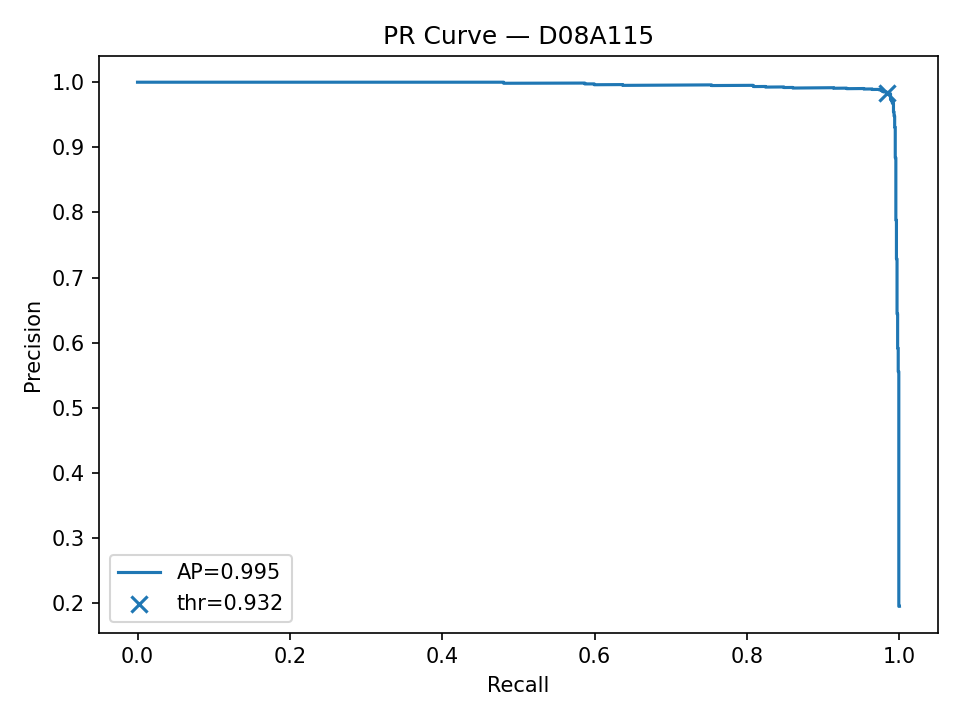
Discharge (cms), rain (mm), and predicted probability for recent events.

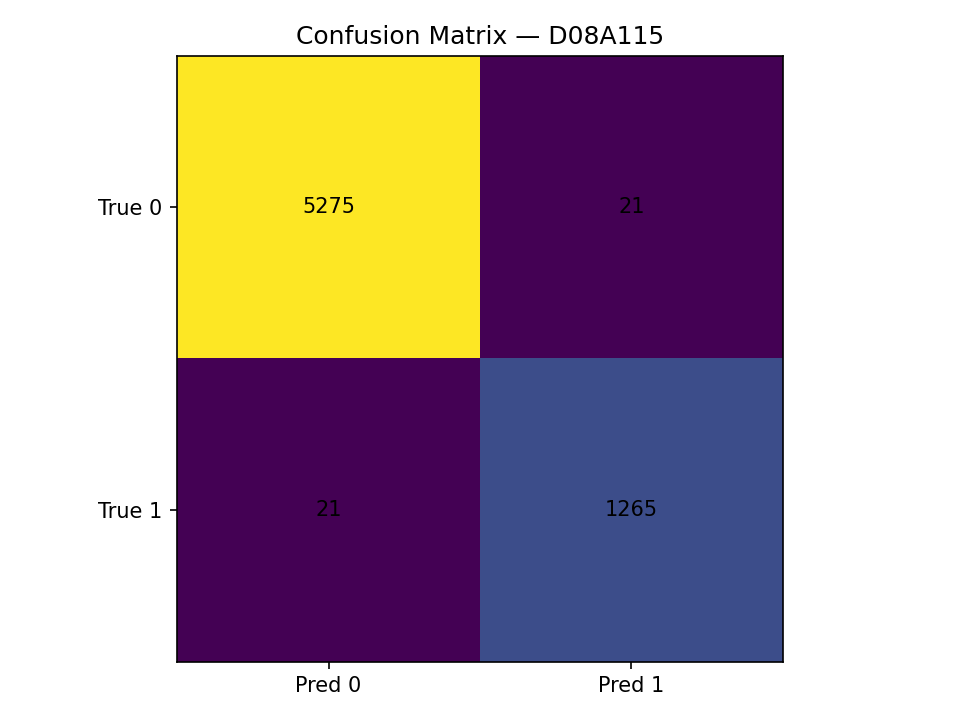


# 3. Station D08A115

## PR Curve & Confusion Matrix

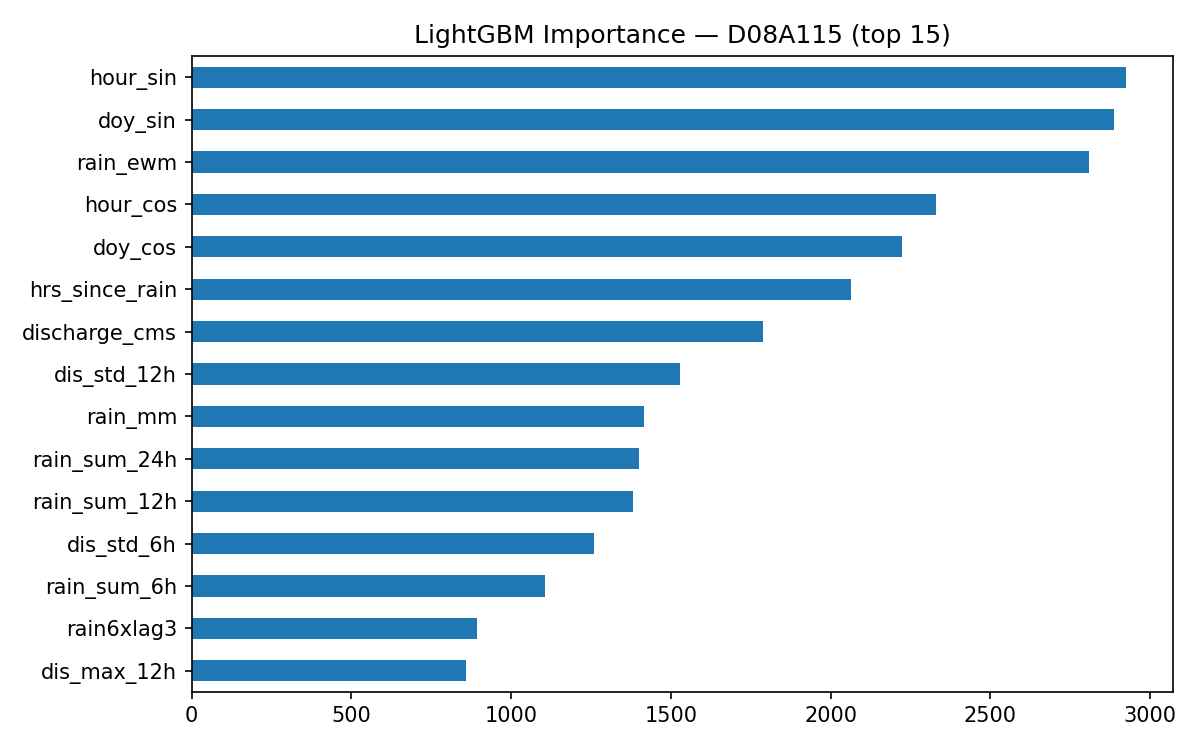
Precision–Recall curve with the chosen operating threshold.





## Feature Importance (LightGBM) & Notes

Top 15 features by LightGBM impurity importance.



1) Recent discharge lags dominate (e.g., dis\_lag\_1h, dis\_lag\_5h, dis\_lag\_12h).

2) Short-horizon rainfall sums contribute (rain\_sum\_6h, rain\_sum\_3h); EWMA captures persistence.

3) Rate-of-change features (dis\_rate\_3h, dis\_rate\_1h) help detect surges.

4) Range/variability (dis\_std\_12h, dis\_std\_6h) adds volatility context.

5) Mild seasonal effects present (hour\_sin, doy\_sin).

## Ablation (XGB → +Stack → +LGBM → +LSTM)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| variant | ap | recall | precision | acc | FP | FN |
| XGB only | 0.998 | 0.967 | 0.993 | 0.992 | 9.000 | 43.000 |
| +Stack (no LGBM/LSTM) | 0.995 | 0.984 | 0.983 | 0.993 | 22.000 | 21.000 |
| +LGBM | 0.995 | 0.984 | 0.984 | 0.994 | 21.000 | 21.000 |
| +LSTM | 0.995 | 0.984 | 0.984 | 0.994 | 21.000 | 21.000 |

## Error Analysis (Top FP / Top FN)

Top False Positives (highest predicted probability):

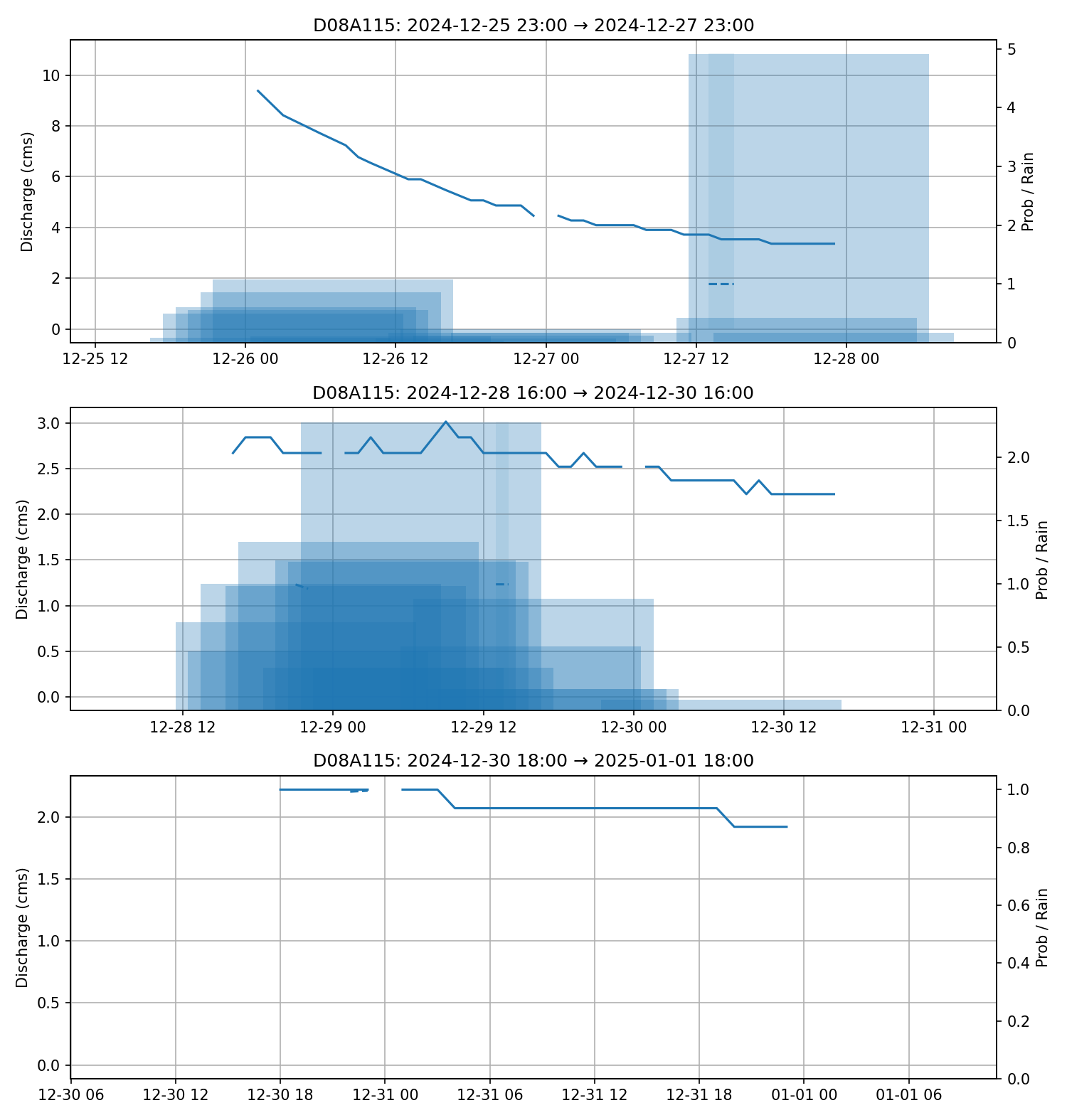
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| proba | y\_true | pred | discharge\_cms | rain\_mm | rain\_sum\_12h | dis\_lag\_1h | dis\_rate\_1h |
| 0.998 | 0.000 | 1.000 | 1.860 | 0.000 | 2.485 | 2.000 | -0.140 |
| 0.997 | 0.000 | 1.000 | 11.160 | 1.385 | 22.239 | 10.920 | 0.240 |
| 0.997 | 0.000 | 1.000 | 2.000 | 0.000 | 1.631 | 2.000 | 0.000 |
| 0.997 | 0.000 | 1.000 | 1.860 | 1.077 | 1.715 | 1.860 | 0.000 |
| 0.996 | 0.000 | 1.000 | 2.060 | 12.377 | 29.323 | 2.288 | -0.228 |
|  |  |  |  |  |  |  |  |

Top False Negatives (lowest predicted probability):

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| proba | y\_true | pred | discharge\_cms | rain\_mm | rain\_sum\_12h | dis\_lag\_1h | dis\_rate\_1h |
| 0.000 | 1.000 | 0.000 | 28.400 | 1.938 | 77.078 | 38.860 | -10.460 |
| 0.008 | 1.000 | 0.000 | 0.759 | 0.000 | 7.754 | 0.759 | 0.000 |
| 0.009 | 1.000 | 0.000 | 0.424 | 0.685 | 7.592 | 0.424 | 0.000 |
| 0.010 | 1.000 | 0.000 | 0.807 | 0.000 | 0.000 | 0.744 | 0.063 |
| 0.012 | 1.000 | 0.000 | 1.204 | 0.000 | 0.000 | 1.204 | 0.000 |
|  |  |  |  |  |  |  |  |

## Timeline Panels (Recent Events)

Discharge (cms), rain (mm), and predicted probability for recent events.



# 4. Recommended Feature Engineering Improvements & Next Steps

Focus: boost recall while keeping false alarms acceptable, using only rainfall and discharge data (no external sources).

Feature engineering (add incrementally):

• Multi-scale rainfall: rolling sums at 1/2/3/6/12/24/48/72h; exponential (EWMA) with multiple halflives.

• Storm segmentation: event-based totals, peak intensity, time since peak, inter-event dry gaps.

• Discharge dynamics: higher-order lags (up to 24–48h), rolling max/min/median/quantiles, rolling std & IQR, multi-step rates (1/3/6h).

• Cross-terms: interactions (rain×lagged flow), (intensity×dry-gap), (API×current rain).

• Seasonality: hour-of-day/day-of-year sin–cos pairs already included; consider weekday/weekend flag if operational.

Modeling stack:

• Keep per-station stacking (XGB/RF/GB/LGBM/LR/MLP + optional LSTM).

• Calibrate probabilities (Platt/Isotonic) per station before thresholding.

• Tune per-station thresholds against FP/FN trade-off; keep your nudged results as defaults.

• Optional: cost-sensitive loss or class-weight sweeps around current operating point.

# 5. Reproducibility Notes

Random seed: 42; Per-station 70/30 stratified split; Fixed thresholds as noted. Libraries: scikit-learn, xgboost, lightgbm, tensorflow (LSTM via scikeras).