

StormSens.AI

Rain Hazard Detection: Backtest Validation Results

July 2026 (cross-reference to lightning detection rate updated October 2026)

Executive Summary

StormSens.AI's rain hazard detection system has been independently validated against **178 real, historically-verified rain events** and **519 certified rain-free windows**, drawn from the same five climatologically distinct U.S. regions used to validate the platform's lightning detection engine. This is a companion validation, not a replacement — rain and lightning are scored, composed, and reported as independent hazard channels.

The headline result: when rain reaches an intensity meaningfully above ambient conditions (the WARNING tier, triggered above 7.6 mm/hr at the venue or 50 mm/hr within 10 miles), the system correctly identified **113 of 117 verified events (96.6%)**, while never triggering **WARNING incorrectly on any of the 519 certified negatives (0%)**. Unlike a separate 100% figure reported later in this paper, this result is not a mathematical certainty of the validation design — it is a genuine, measured outcome.

The system's composition logic — which combines rain and lightning into a single reported status — was checked against all **15,645 real historical evaluation points** in this dataset. In every case, the composed status was confirmed to never rank below lightning's own independently-computed status. Rain can add urgency to a report; it cannot reduce a lightning warning.

Methodology

Data source

Rain intensity is measured using NOAA's Multi-Radar Multi-Sensor (MRMS) Surface Precipitation Rate product — a radar-derived estimate updated every two minutes at approximately 1-kilometer resolution, accessed via the same AWS public-data infrastructure already used for this platform's lightning detection engine. Both live and historical MRMS data were verified against independent sources before use in this validation.

Ground truth construction

This validation reuses the platform's existing lightning ground-truth dataset (295 independently verified severe-weather events and 759 certified storm-free windows) as its starting point, then applies a dedicated, pre-registered re-verification pass specific to rain: for each event, historical MRMS data was pulled at the venue's exact coordinates and checked against real precipitation thresholds, rather than assuming a lightning-verified event automatically implies meaningful rain at that location.

A material limitation is disclosed here directly, not left implicit: AWS's MRMS archive begins in October 2020. Approximately 40% of the original positive events and 31% of the original negative windows predate this boundary and could not be evaluated for rain — not because of any detection failure, but because no radar data exists for that period. These events are excluded from the rain-specific dataset, not counted as failures. The effective validated sample is **178 usable positive events** (of which 117 were confirmed to have produced measurable rain at the venue) and **519 effective usable negative windows** (518 certified rain-free, 1 failed).

¹ **The single failed window:** Monroe County, PA, 2025-06-16 21:39–22:39 UTC. It failed the radar rain-rate check, not a data-availability gap — with full 25-of-25 usable radar samples, the archive records a brief passing shower peaking at 4.9 mm/hr (three consecutive samples at or above the 2.5 mm/hr certification threshold), rising from and returning to 0.0 mm/hr within the hour. The window had passed the storm-event, warning, and thunder screens — light rain without lightning — so the radar archive itself shows this “rain-free” candidate was not rain-free. It is deliberately retained in the 519-window scoring denominator: the system's WATCH-tier response to it counts against the reported caution rate even though the measured rain shows that response was correct — the conservative direction.

Indeterminate-data handling

Radar coverage gaps are treated as a distinct, honest outcome — neither a pass nor a fail — mirroring the same principle already applied to lightning sensor blind spots elsewhere in this platform. For positive events, a gap only excludes an event if radar coverage is missing for the entire evaluation window. For negative windows, any single contiguous coverage gap of 25 minutes or longer voids certification for that window — a threshold grounded in published meteorological research on the typical lifecycle of an isolated convective rain cell (approximately 20–30 minutes of active precipitation), not an arbitrary cutoff.

Replay methodology

Detection performance was scored using a two-stage process: first, historical rain data was fetched for every evaluation point across the full ground-truth dataset; second, a dedicated replay harness walked each event in 5-minute increments, independently reconstructing both the lightning and rain model assessment at each point using the platform’s live weather detection logic — the same code path used in the deployed mobile app, not a simplified or reimplemented approximation.

Results

Detection performance

At the WATCH tier (any measurable rain at the venue), the system detected **117 of 117 verified events (100%)**.

This figure should always be read together with its explanation: it is a mathematical consequence of how the ground truth was defined, not an independent measure of system quality. An event only qualifies as a verified positive if it crossed a rate threshold (2.5 mm/hr) at some point; the WATCH trigger itself fires on any rate above 0 mm/hr — a strictly looser bar. Every qualifying event was therefore guaranteed to cross the WATCH threshold by construction.

The WARNING-tier result reported in the executive summary above — **113 of 117 (96.6%)** — does not share this property. The WARNING threshold (7.6 mm/hr at the venue, or 50 mm/hr within 10 miles) sits well above the 2.5 mm/hr bar used to define a positive event, so this result had to be genuinely earned.

False-positive analysis

Across all 519 certified rain-free windows, the WATCH-tier trigger fired incorrectly on **11 occasions (2.12%)**. Every case was individually investigated rather than accepted as an aggregate statistic:

- 10 cases were real, measured trace rainfall (0.0–2.0 mm/hr) at the venue — genuine precipitation below the 2.5 mm/hr threshold used to certify a window as rain-free, correctly detected by a trigger intentionally more sensitive than the certification standard.
- 1 case was a real, measured storm cell 8.6 miles from the venue, with the venue itself completely dry — the correct, intended behavior of the system’s “heavy rain nearby” warning path.

At the WARNING tier — the threshold that actually changes user-facing urgency — the false-positive rate was **zero (0 of 519)**.

Composition safety

StormSens.AI reports a single composed hazard status when both rain and lightning conditions are present, using the higher-ranked of the two. Rain is deliberately capped below the platform’s DANGER tier, which remains reserved exclusively for lightning proximity — preserving the specific, validated meaning that tier already carries. Across all 15,645 real historical evaluation points in this dataset, the composed status never once ranked below lightning’s own independently-computed status.

Results by region

Region	Detected (WATCH)	False positives (WATCH)	Steps checked	FP rate
--------	------------------	-------------------------	---------------	---------

Monroe, PA	30/30	4/128	3,725	3.13%
Citrus, FL	4/4	0/20	575	0%
Palm Beach, FL	25/25	2/72	2,280	2.78%
Travis, TX	24/24	1/130	3,760	0.77%
El Paso, CO	34/34	4/169	5,305	2.37%

Per-region variance shown above is not statistically meaningful at these sample sizes; 95% confidence intervals overlap substantially across all five regions.

² **Citrus, FL rests on the smallest sample of any region:** its perfect 4/4 detection and 0/20 false-positive figures are genuine and reproducible, but should be read with that in mind rather than as evidence the system performs uniformly better in this region. The small sample is a reporting-density artifact, not a sign of infrequent storms. The underlying NCEI storm-report base for Citrus was already the smallest of the five regions (16 of 295 pooled lightning positives), because NCEI report density tracks population and spotter/damage-assessment coverage, not actual meteorology. Citrus's reports also skew older than the other regions' (69% predate the MRMS archive's October 2020 start, versus ~38% pooled), so the archive-availability limitation already disclosed in this paper reduces Citrus's usable sample more than any other region: 16 seeded positives → 5 in-archive → 4 rain-verified.

Known Limitations

Effective sample size. The pre-October-2020 MRMS archive boundary reduced the usable validation sample below the original lightning ground truth (178 positives / 519 negatives, versus 295 / 759). This is a genuine data-availability constraint, disclosed directly rather than adjusted around.

Forecast magnitude accuracy is not validated by this work. StormSens.AI's rain forecasting can estimate expected accumulation over the following 45 minutes. This capability is explicitly outside the scope of this validation, which addresses hazard classification only. Magnitude-forecast accuracy requires a separate, dedicated backtest with its own no-lookahead methodology, not yet conducted.

An open product question. The current WATCH-tier trigger fires on any measurable rain at the venue, deliberately more sensitive than the 2.5 mm/hr certification standard. Raising this trigger to match the certification threshold would reduce WATCH-tier false positives from 11 to 4 while holding detection at 117/117 — a legitimate tradeoff between informational sensitivity and alert frequency, not yet decided.

Relationship to Lightning Detection

Rain hazard detection is an independently validated, structurally separate capability — not an extension or reinterpretation of the platform's existing lightning detection guarantees. The lightning system's own validated results — 72.2% detection, <0.40% false-positive rate (lightning white paper revised October 2, 2026) — are unaffected by this work.

³ The lightning detection rate has moved since this paper's original publication in July 2026, from the previously reported 70.5% to the current 72.2% (lightning white paper revised October 2, 2026), following a September 2026 correction to the detection engine and an October 2026 widening of the satellite near-band, both described in the lightning white paper's revision history. Those changes are unrelated to rain detection and do not touch anything in this document — it is noted here only so the two papers do not disagree.